Signatures of selection and introgression in the genus *Gallus*

Raman Akinyanju LAWAL, B. Agric (Hons)

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

December 2017

Cells, Organisms and Molecular Genetics Division
School of Life Sciences



Declaration

I, Raman Akinyanju LAWAL, hereby declare that this thesis was the original work of myself except where explicitly stated and no part of it has been previously presented or submitted for examination anywhere else. All references to the work of others were duly acknowledged wherever necessary.

December 2017

Dedication

I dedicate this thesis in appreciation to the effort of my wonderful mother Olufunmilayo OGUNDEYI.

Acknowledgements

Special gratitude to the University of Nottingham for paying my full tuition fee through the Vice Chancellor's scholarship (International) programme.

There are several people worth mentioning for contributing to the success of this thesis. But first, my deepest appreciation goes to my supervisor, Professor Olivier Hanotte for his selflessness and positive ways of looking at situations. Every moment spent with him was always about new idea and creative thinking. His intensity of mentoring, the patience, the encouragement, the trust, and the support throughout my PhD study has helped me transform into a more better person in all aspects. He gave me freely, numerous opportunities to acquire the right set of skills, great mindset, ability to explore, to make mistakes, to self-correct, to grow and to continue contributing to the progress and survival of human race. Thank you!

To my internal assessor, Professor John Armour, I couldn't have wished for less mentor. Your wonderful suggestions and constructive criticisms indeed have shaped this thesis for the best. I should mention that the haplotype phylogenetic tree analysis which was used to re-assess the evidence and source of introgression in this study was one of the greatest suggestions by John.

The following collaborators have contributed data which were analysed in this thesis: Dr Takele Taye Desta (contributed Ethiopian chicken blood samples), Professor Raed Mahmod Al-Atiyat of Mutah University, Jordan and Professor Riyadh Salah Aljumaah of King Saud University, Saudi Arabia (contributed Saudi Arabian chicken sequences), Professor Pradeepa Silva of Peradeniya University, Sri Lanka (contributed Sri Lanka chicken and Ceylon junglefowl blood samples), Mr Koen Vanmechelen of Open University of Diversity, Belgium and Dr Addie Vereijken of Hendrix Genetics, Netherland (contributed fancy chicken, red, grey, Ceylon and green junglefowls blood and sequences). Dr Tatsuhiko Goto of Obihiro University, Japan facilitated the collection of the published sequences of Indonesian Kedu Hitam chicken, Javan red junglefowl and green junglefowl while the remaining red junglefowl sequences were provided by Professor Dong-Dong Wu of Kunming Institute, China.

Dr David Wragg (Roslin Institute) deserves special mention for supporting me during my wet lab experience and was always the first point of call when I ran into bioinformatic issues. The now widely used mapping and variants calling scripts for whole-genome sequences in our lab was first prepared by Dr David Wragg and later improved by myself. Dr Simon H. Martin (Cambridge University) provided extensive support with the interpretation of both the ABBA-BABA and topology weighting results. A big thanks to Iyelola Oluwaseun Turner for reading and correcting grammatical errors in some of my chapters. To my colleagues in the lab, Abdulfatai Tijjani, Abulgasim Ahbara, Akil Alshawi, Ahmed Al-Jumaili, Vallejo Trujillo Adriana, and Akinci Ibrahim, you guys are wonderful.

To my two sweetest, Bimbo and Adeshewa Lawal, I cannot thank you both enough for your patience and understanding. You guys are the champions for the success of this thesis. Finally, a big thank you to my parents and siblings and to Mr Akintayo Adejare alongside three of my uncles (Mr Olufemi Akinlawon, Hon Abiodun Akinlade and Mr Fatai Lawal) for supporting me during my student visa application and flight processes.

Published papers

Wragg, D., Mason A. S., Yu, L., Kuo, R., **Lawal**, R. A., Desta, T. T., Mwacharo, J. M., Cho, C-Y., Kemp, S., Burt, D. W., and Hanotte O. (2015). Genome-wide analysis reveals the extent of EAV-HP integration in domestic chicken. *BMC genomics*, 16:784. doi:10.1186/s12864-015-1954-x.

Hassaballah, K., Zeuh, V., **Lawal R. A**., Hanotte, O. and Sembene, M. (2015). Diversity and Origin of indigenous village chickens (*Gallus gallus*) from Chad, Central Africa. *Advances in bioscience and biotechnology*, 6, 592-600. doi:10.4236/abb.2015.69062.

Desta, T., **Lawal, R. A**., and Hanotte, O. (2015). Genetic diversity of village chickens In: Third report on chicken genes and chromosomes 2015. *Cytogenetic and genome research, 145*(2), 130-133.

Selected conferences

Lawal, R. A., Silva, P., Al-Atiyat, R., Hanotte O. (2017). Genome-wide analysis for signature of selection in domestic chicken and red jungle fowl. 36th International society for animal genetics conference, Dublin, Ireland.

Lawal, R. A., Wragg, D., Silva, P., Vanmechelen, K., Vereijken, A., Wu, D-D., Al-Atiyat, R., Hanotte O. (2016). Genetic introgression in domestic chicken, insight from whole genome sequences. 35th International society for animal genetics conference. Salt Lake City, United State.

Lawal, R. A., Desta, T. T., and Hanotte O. (2015). Genome-wide analysis of signature of selection within and between two African indigenous chicken populations. Genetic society meeting, breeding for bacon, beer and biofuels, Roslin Institute, University of Edinburg, UK

Lawal, R. A., Adebambo, A. O., Desta, T. T., Wragg D., and Hanotte O. (2014). Genetic differentiation of Ethiopian and Nigerian village chicken. Population genetic group meeting 47, Bath, UK.

Thesis objectives

This thesis has two main objectives. First, I aim to contribute to understanding the process of chicken domestication and their subsequent adaptation to diverse agro-ecologies by identifying and characterising signatures of positive selection in three indigenous chicken populations *Gallus gallus domesticus*, red junglefowl *Gallus gallus*, Ceylon Junglefowl *G. lafayettii* and the green junglefowl *G. varius*. Secondly, I aim to contribute to understanding the origin of the genetic diversity of the domestic chicken by investigating if the grey junglefowl *G. sonneratii*, the Ceylon junglefowl *G. lafayettii* and the Javan red junglefowl *G. g. bankiva* have contributed to the genetic pool of indigenous domestic chicken populations.

Accordingly, the thesis is divided into the following chapters:

Chapter 1 provides an introduction about the genus *Gallus*, the origin of domestic chicken and its dispersion across the world.

Chapter 2 assesses the presence of signatures of selection in red junglefowl and domestic chicken populations from Africa (Ethiopia), Saudi Arabia (Middle East), and Sri Lanka (Asia) in relation to the domestication of the species and its adaptation to local agro-ecologies.

Chapter 3 examines the presence of signatures of positive selection in two wild junglefowl species, the Ceylon and the green junglefowls. This will allow me to address whether some of these signatures originate from the genus *Gallus* rather than the red junglefowl or domestic chicken specifically.

Chapter 4 examines the autosomal contributions of the grey junglefowl and the Ceylon junglefowl in the indigenous domestic chicken populations from Africa (Ethiopia), Middle East (Saudi Arabia), Asia (Sri Lanka) and Fancy birds (Europe).

Chapter 5 assesses if the Javan red junglefowl (a red junglefowl subspecies restricted to the Indonesian Islands of Sumatra and Java) with no evidence so far of a maternal genetic contribution to the domestic chicken genome, has contributed to the gene pool of indigenous chicken population from Indonesia, Ethiopia, Saudi Arabia, Sri Lanka, and Fancy birds.

C**hapter 6** synthesises the findings of the previous chapters and suggests possible routes of further investigation in order to better understand the diversity and adaptation of the domestic chicken.

Thesis summary

Here I investigate, using autosomal whole-genome sequence data, the signature of positive selection and/or introgression in the indigenous domestic village chickens from three countries (Ethiopia, Saudi Arabia, Sri Lanka), three fancy birds, three junglefowl species (red junglefowl *Gallus gallus*, grey junglefowl *G. sonneratii*, Ceylon Junglefowl *G. lafayettii*) and the Javan red junglefowl *G. g. bankiva*. All the new sequencing data were obtained from Illumina HiSeq 2000/2500 DNA sequencers with an individual bird depth of genome coverage ranging from 10 X to 30 X. The analyses in this thesis have been completed using the reference genome *Galgal* 4.0.

For the detection of signatures of positive selection, this analysis excluded the three fancy birds and the grey junglefowl due to small sample size. Using the pool heterozygosity and SweeD composite likelihood selection signature methods, I identified two candidate selected regions shared between all the three indigenous domestic village chicken populations and the red junglefowl (chapter 2). These regions contain genes that are associated with the development of the central nervous system and adaptation to hypoxic environments. Five candidate regions were shared among the three indigenous village chicken populations, and they represent candidate domestication regions. Unique regions in each domestic chicken population were also identified. Functional genes have not been assigned to most of these regions but in those where the genes have been annotated, the gene function may be related to production and reproductive traits as well as adaptation to cold/hot temperatures and hypoxia.

In chapter 3, I analysed only the Ceylon and green junglefowl whole genome sequences for the detection of candidate signatures of positive selection using both the pool heterozygosity and Tajima's D. In both species, I identified candidate selected regions that contained genes which may be linked to adaptation to different environmental challenges e.g disease resistance, stress, thermoregulation and hypoxia. In the genome of green junglefowl, candidate selected regions associated with skeletal formation and ovarian follicle development were significantly detected.

In chapter 4, I identified introgressed candidate regions from the grey and Ceylon junglefowls in domestic chicken (including the three indigenous chicken populations and fancy birds) using the ABBA – BABA four taxon method. Our result shows that, domestic

chickens shared 75.8% of their genome with the red junglefowl, 4% with the grey junglefowl and 1.1% with the Ceylon junglefowl. I observed introgression in both directions, namely from the grey/Ceylon junglefowls into domestic chickens and *vice versa*. While from the grey junglefowl, introgression was present in all the domestic chicken populations as well as interestingly in the red junglefowl, for the Ceylon junglefowl, introgression was more restricted to the domestic chicken from Sri Lanka. From the ABBA – BABA analysis between the grey junglefowl and the domestic chicken, I also identified a single candidate introgressed region from the green junglefowl *G. varius* in two domestic birds from Sri Lanka. Future study should therefore consider investigating the genome-wide analysis of introgression from the green junglefowl into the domestic chicken.

In chapter 5, I ended our introgression study by investigating if a distantly related subspecies of red junglefowl, the Javan red junglefowl *Gallus gallus bankiva*, has contributed to the gene pool of the domestic chicken. Alongside the three indigenous domestic chicken populations, I included the genome sequences of three domestic chickens sampled from the Java island of Indonesia. Our result shows a significant 10.6% genome admixture between the domestic chicken and the Javan red junglefowl.

Overall, our results indicate that the genetic make-up of the domestic chicken is rather complex with multiple species and subspecies influences. These introgression events have contributed to the genetic diversity of these domesticates. Our results also support the geographic difference of introgression and indicate that these introgression events may have contributed to the adaptive traits of the domestic chicken. However, this requires further investigation.

Contents

Chapter 1 1
The genus *Gallus* 2
Green junglefowl *Gallus varius* 2
Ceylon junglefowl *Gallus lafayettii* 3
Grey junglefowl *Gallus sonneratii* 4
Red junglefowl *Gallus gallus* 6
Cochin-Chinese red junglefowl *Gallus gallus gallus* 7
Burmese red junglefowl *Gallus gallus spadiceus* 8
Tonkinese red junglefowl *Gallus gallus jabouillei* 9
Indian red junglefowl *Gallus gallus murghi* 9
Javan red junglefowl *Gallus gallus bankiva* 10
Genealogy of domestic chicken: Molecular evidence 11
Chicken domestication centres: Archaeological evidence 12
Chicken dispersal: Archaeological and linguistic evidence 13
Asia: Archaeological and linguistic evidence 13
Europe: Archaeological and linguistic evidence 14
Africa: Archaeological and linguistic evidence 15
America: Archaeological and linguistic evidence. 16
Genetic features, diversity and selection in domestic chicken 17
The indigenous domestic chicken or village chicken 17
Commercial and experimental chicken lines 17
Standardized breeds 19
References 19

Chapter 2: Signature of positive selection in red junglefowl and domestic chicken 24
Abstract 25
Introduction 25
Materials and methods 27
Sampling and sequencing 27
Sequence alignment and variants calling 28
Population structure and relationships analyses 29
Mitochondrial DNA analysis 29
Selective sweeps detection 30
Test for homozygosity 31

Gene ontology and pathway analyses 31
Results 31
Sequencing and SNPs identification 31
Population structure 32
Mitochondrial phylogenetic relationships 35
Mean genome heterozygosity 36
Genome-wide SNPs restriction for Hp sweep detection 37
Selective sweeps detection in domestic chicken populations 37
Selective sweeps detection in red junglefowl 40
Sweep detection using SweeD 41
Homozygosity test 49
Functional annotations for the enriched genes within sweep regions 52
Discussion 54
Signature of selection in relation to domestication 54
Signature of selection in relation to environmental adaptations 55
Common genome regions selected in both domestic and red junglefowl 57
Conclusions 58
References 58

Chapter 3: Whole-genome sequences reveals candidate signature of positive selection for environmental pressures in Ceylon (*Gallus lafayettii*) and green (*G. varius*) junglefowl 65
Abstract 66
Introduction 66
Materials and methods 68
Sampling and sequencing 68
Mappings and variants calling protocol 68
Genome-wide nucleotide diversity 69
Selective sweeps detection 69
Functional annotation 71
Results 71
Sequencing and read mapping 71
Polymorphism detection 71
Consequences of the variants detected 73
Within population diversity 74
Sweep detection in Ceylon junglefowl using Tajima's D 74
Sweep detection in Ceylon junglefowl using Hp 75

Sweep detection in green junglefowl using Tajima's D 77
Sweep detection in green junglefowl using Hp 78
Common selected regions between species 80
Functional annotation for genes within sweep regions 80
Discussion 81
Selection for skeletal structure and reproduction traits in green junglefowl 86
Conclusions 87
References 87

Chapter 4: The autosomal genome ancestry of domestic chicken: Introgression from grey and Ceylon junglefowls 94
Abstract 95
Introduction 95
Materials and Methods 97
Samples and DNA extraction 97
Raw sequence quality check 98
Sequence mapping and variants calling 98
Population structure and samples relationships 99
Introgression – admixture analysis 99
Haplotype analysis 101
Size of introgressed regions 101
Remapping Galgal 4.0 introgressed regions to Galgal 5.0 102
Genome-wide topology weighting estimation 102
Gene ontology analysis 102
Overlaps between introgressed regions and chicken QTLs 102
Results 103
Sequence polymorphisms and variants statistics 103
Population and species differentiation 104
Population and species relationships 107
Mitochondrial DNA analysis 107
Autosomal SNPs analysis 107
Genome-wide autosomal introgression 109
Introgression between the grey junglefowl and domestic chicken 110
Introgression between the Ceylon junglefowl and domestic chicken 119
Topology weighting estimation 128
Genes function analyses at the introgressed regions (Galgal 4.0) 130

Introgression of grey junglefowl into domestic chicken 130
Introgression of Ceylon junglefowl into domestic chicken 131
Introgression of domestic chicken into grey junglefowl 131
Overlaps between introgressed region and domestic chicken QTLs 131
Overlaps between introgressed regions and signature of selection 131
Discussion 134
Conclusions 139
References 139

Chapter 5: Introgression of Javan red junglefowl *Gallus g. bankiva* to domestic chicken: insights from autosomal whole-genome sequences 145
Introduction 146
Materials and methods 148
Samples and DNA extraction 148
Sequence mapping and variants calling procedures 148
Genetic differentiation 149
Population structure and samples relationships 149
Introgression analysis 149
Haplotype analysis 150
Results 151
Sequence reads, mappings, and SNPs identification 151
Pairwise genetic differentiation 152
Population and species relationships 152
Introgression landscape between domestic chicken and Javan red junglefowl 156
Overlaps with signatures of selection 163
Discussion 165
Conclusions 167
References 167

Chapter 6 171
General discussion 171
Conclusions, future studies and recommendations 174
References 175

Chapter 1

The genus *Gallus*

The genus *Gallus*

The *Gallus* species may have originated from an Indian refuge forest species during the late Pliocene and early Pleistocene (Harrison, 1978). However, their fossil remains are scarce. Only a single fossil species *Gallus aesculapii* has been described, from the lower Pliocene in Greece and the Black Sea region (Harrison, 1978). Their present day geographic distribution largely depends on weather conditions and competition with other species. For instance, the absence of *Gallus* in Eastern Asia may be attributed to the presence of pheasant species in that region while today weather conditions in Europe are unsuitable for a tropical species (Harrison, 1978).

The domestic chicken belongs to the genus *Gallus* also known as junglefowl where four species inhabiting the Asian continent have been described. Starting with the most distantly related species, I have the green junglefowl *Gallus varius,* Ceylon junglefowl *Gallus lafayettii*, grey junglefowl *Gallus sonneratii*, and red junglefowl *Gallus gallus.* Speciation within the genus likely followed geographical isolation within a large region including the South of the Himalayas (from Kashmir to Malaysia), Indo-China and Indonesia (Harrison, 1978). All the four species are easily distinguishable by their plumage, voice, behaviour, and other morphological characteristics while the features such as comb and wattles distinguished the genus *Gallus* from other members in the Phasianidae family (Crawford, 1990). All species behave differently under captive environments and the level of fertile hybrids when crossed with domestic chicken differs (Delacour, 1977).

Here, I present an overview of their characteristics.

Green junglefowl *Gallus varius*

Having been isolated to Indonesia following the ancestral species dispersion, the green junglefowl (**Figure 1**) is popularly found in Java and its neighbouring islands including Bali, Sumbawa, Lombok, Bawean, Madura, Kangean, Flores and Alor (Delacour, 1977). Unlike other junglefowls, they live in low elevation usually close to the coast around seashore and low-lying valleys but could also be found in an altitude of up to 3,000 m from the sea-level in shrub and arable land (Del Hoyo *et al.*, 2017a). If found in the forest, they are known to forage on the dry and rocky scrub where they move in small numbers (Delacour, 1977). In its geographic distribution which overlaps with the domestics, F1

crosses called Bekisar are produced. They are highly prized for their voices and are supposedly infertile (Delacour, 1977).

The green junglefowl is the most distant relative of the domestic chicken within the *Gallus*. They also exhibit unique genetic and phenotypic distinction compared to other junglefowl species. Morphologically, the comb is unserrated, and there is just one median wattle (Morejohn, 1968b). There is presence of 16 tail feathers instead of 14 as for other species and the shape of the neck hackles is truncated rather than the pointed hackle feathers found in other species (Morejohn, 1968b). They possess green plumage colour, a clutch size of six to ten eggs, buffy white and spotless in colour (Delacour, 1977). They incubate their eggs for a period of 21 days. They are very delicate and easily susceptible to infectious diseases in captivity when compared to other species and in winter, usually require artificial heat (Delacour, 1977).



Figure 1. Green junglefowl *Gallus varius* (credit: NaturePixels.org and BESG)

Ceylon junglefowl *Gallus lafayettii*

Ceylon junglefowl (**Figure 2)** is geographically confined to the island in Sri Lanka where they spend most of their life in the jungle and mostly shun cultivation (Delacour, 1977). They scavenge in the forest through the dry coastal scrub and can seek cover for up to an

elevation of 2,300 m above the sea-level (Del Hoyo *et al.*, 2017b). Like the red junglefowl, they are hardy, prolific except that they are not as social. Hybridization with the sister species, grey junglefowl has been reported (Nishibori *et al.*, 2005), while the hybridization with the red junglefowl is believed to result in sterile offspring (Delacour, 1977). After green junglefowl, they are with the grey junglefowl, the next distantly related species to the domestic chicken (Nishibori *et al.*, 2005). Apart from the pointed and fringed breast feathers which differed from the female-like feather of red junglefowl, the male and female plumage for both the Ceylon and red junglefowls are similar (Morejohn, 1968b). There is presence of bluish-purple feather colour at the upper breast close to the lower neck (Morejohn, 1968b). They lay between two to four eggs, buff or pink tint colour which could be purple/brownish spotted or unspotted with an incubation period of 20-21 days (Delacour, 1977).



Figure 2. Ceylon junglefowl *Gallus lafayettii* (credit: Steve Garvie, flickr)

Grey junglefowl *Gallus sonneratii*

The grey junglefowl (**Figure 3**) is endemic to the Western and Southern part of India and can also be found in the central India and Rajputana (Delacour, 1977). Though, the samples collected from Northern India are usually lighter in colour than those from the

South, no subspecies have been defined (Delacour, 1977). The subspecies "*wangyeli"* from the Mount Abu in the North-West of India, was never recognised despite a significant difference in the voices when the cock crows (Ali and Ripley, 1983; Storer, 1988). Grey junglefowl roam the forest and bamboo groves and feed on seeds, insects, and termites. They are sometimes found in an elevation of up to 1,500 m above the sea-level. They may gather in large numbers when there is an abundance of food, otherwise, they appear either single or occasionally in pairs (Delacour, 1977). Hybridization with the red junglefowl *Gallus g. murghi* may be occurring in the wild especially in an area where the two species live in sympatry along the northern border of the grey junglefowl geographic distribution (Rasmussen and Anderton, 2005). Genetic exchanges with domestic chicken have also been reported (Morejohn, 1968a; Eriksson *et al.*, 2008). The grey and Ceylon junglefowl are two sister species (Fumihito *et al.*, 1996; Nishibori *et al.*, 2005) (also see chapter 4). The grey junglefowl required little care in captivity. The chicks are not as delicate as in the green junglefowl but not as robust as for the red and Ceylon junglefowls (Delacour, 1977). They are hunted for making artificial flies used in trout fishing which has impacted negatively on their population sizes in some areas (Wayre, 1976). The main distinguishing features is their sealing wax-like spots found on some of the feathers which are expanded and flattened portions of the rachis (Morejohn, 1968b). At the tail of the feather, the flattened rachis becomes shredded and has yellow spots. The body feather including the wing and tail are mostly black complemented with white shaft and grey border. The male grey and Ceylon junglefowls have lesser and median wing coverts (Morejohn, 1968b). The female plumage of both the red and grey junglefowls are similar with the exception to the breast feathers in grey junglefowl which is white with broad black or brown borders (Morejohn, 1968b). Like the Ceylon junglefowl, the incubation period is between 20 and 21 days but they lay more eggs between 4 and 8 that are usually unspotted white to rosy and buff (Delacour, 1977).



Figure 3. Grey junglefowl *Gallus sonneratii* (credit: www.purelypoultry.com)

Red junglefowl *Gallus gallus*

The red junglefowl formerly referred to as the *Gallus bankiva* or *Gallus ferrugineus* is the primary ancestor of modern chicken with fully fertile offspring always produced when crossed with domestic chicken (Darwin, 1868; Morejohn, 1968a). Plumage colours and pattern in some domestic chicken are identical to the red junglefowl ones (Jaap and Hollander, 1954). Domestic and red junglefowl have similar morphology (Delacour, 1977; Crawford, 1990). Both the male and female red junglefowls differ by the comb, wattles, and body size. There is presence of red or white ear patches on the sides of the head. Plumage variations alongside the geographical locations distinguished the subspecies of red junglefowl (**Figures 4 to 8**). Generally, they are more social and may live in large group (Delacour, 1977). The form and colour of neck, back and rump feathers are similar between the red junglefowl and Ceylon junglefowl while the female of both the red junglefowl and the grey junglefowl have uniform light-brown plumage on the back, tail, rump and wing coverts (Morejohn, 1968b). The clutch size is generally larger than in the grey and Ceylon junglefowls but smaller compared to the green junglefowl (Delacour, 1977).

The five recognised subspecies of red junglefowl are: Cochin-Chinese red junglefowl *G. g. gallus*, Burmese red junglefowl *G. g. spadiceus*, Tonkinese red junglefowl *G. g. jabouillei*, Indian red junglefowl *G. g. murghi*, and Javan red junglefowl G. *g. bankiva* (Delacour, 1977).

Cochin-Chinese red junglefowl *Gallus gallus gallus*

Located in South Vietnam, Cambodia, Thailand, and Laos, the Cochin-Chinese red junglefowl differs from other subspecies by its plumage, large dented comb, and the presence of large milk-white ear lobes (**Figure 4**) (Delacour, 1977). They inhabit the dry and damp forest and are usually found in an elevation of up to 1,200 m above sea-level (Delacour, 1977). Cochin-Chinese red junglefowl has previously been reported as the single matriarch ancestor from which all chicken breeds evolved (Fumihito *et al.*, 1994), but a more recent study has suggested that genetic contributions from other subspecies likely occurred (Liu *et al.*, 2006). Although, the geographic distribution overlaps with the Tonkinese red junglefowl in the Northern Annam (Delacour, 1977), they are believed to be genetically closer to the Burmese red junglefowl (Fumihito *et al.*, 1994).



Figure 4. Cochin-Chinese red junglefowl *G. g. gallus* (credit: Peter Ericsson of internet bird collection)

Burmese red junglefowl *Gallus gallus spadiceus*

They are found in Myanmar, Thailand (except East), South-western Yunnan (China), Burma, Northern Laos, Malaysia, and Northern Sumatra. They are characterized by short ornamental feathers, red ear lobes which are smaller than the Cochin-Chinese red junglefowl ones (**Figure 5**). Both Cochin-Chinese and Burmese red junglefowl share similar habitat. The Burmese red junglefowl may have contributed to the genetic background of the domestic chicken gene pool (Liu *et al.*, 2006). They live in sympatry with the Tonkinese red junglefowl in the North-Western part of Indo-China, where hybridization is likely to occur.



Figure 5. Burmese red junglefowl *G. g. spadiceus* (credit: Trevor Hardaker of (http://www.hardaker.co.za)

Tonkinese red junglefowl *Gallus gallus jabouillei*

It is found in the extreme North of Annam and extreme South-East of Yunnan and Hainan (China), and Northern Vietnam. They have a darker red plumage, small comb, and red ear lobes (**Figure 6**). They prefer the colder and damper part of the country. The Tonkinese red junglefowl has also likely contributed to the gene pool of the domestic chicken (Liu *et al.*, 2006).



Figure 6. Tonkinese red junglefowl *Gallus g. jabouillei* (credit: BioLib.cz)

Indian red junglefowl *Gallus gallus murghi*

They are endemic to the Northern and North-Eastern part of India, Nepal, Bhutan, Bengal, Gamjam, and Bangladesh. They can easily be distinguished by the white but small ear lobes (**Figure 7**). They played a prominent role in chicken domestication with the Indian subcontinent likely one of the centres of domestication. Hybridization with the grey junglefowl in the central provinces near Jubhulpore is likely (Delacour, 1977).



Figure 7. Indian red junglefowl *Gallus g. murghi* (credit: Kuldeep Kharade of Oriental bird images database)

Javan red junglefowl *Gallus gallus bankiva*

They are found in Indonesia particularly in the Southern part of Sumatra, Java, and Bali. They have red ear lobes (**Figure 8**) like the Burmese red junglefowl but are easily distinguished by the neck hackles which are broader and almost rounded at the tip (Delacour, 1977). They share same geographical location with the green junglefowl (Delacour, 1977). Mitochondrial DNA analysis has revealed no evidence of contribution of Javan red junglefowl to the domestication of the chicken (Fumihito *et al.*, 1994; Liu *et al.*, 2006).



Figure 8. Javan red junglefowl *G. g. bankiva* (credit: Ekapob Trouatphon of National Geographic)

Genealogy of domestic chicken: Molecular evidence

The domestic chicken genealogy has been an intense subject of debate since Darwin proposed the red junglefowl as its sole ancestor (Darwin, 1868). Archaeological materials have also been unreliable because the remains of domestic chicken obtained from archaeological sites could not be easily associated with any of the four junglefowl species due to similarity in the skeletal morphology (Crawford, 1990). Evidence therefore depends on breeding experiments, phenotypic observations, and molecular studies. As reported in Crawford (1990), Punnett recommended crosses between each of the four species and domestic chicken to settle the origin of the domestic chicken (Punnett, 1923). Though F_1 hybrids were subsequently produced, backcrossing and *inter se* mating of the hybrids usually lead to fatality (Crawford, 1990), further supporting the red junglefowl as the main ancestor.

With careful experimental procedures and using birds held in captivity for a long time, Morejohn (1968) crossed successfully red and grey junglefowls and could produce F_2 birds as well as backcrosses with both the grey junglefowl and the red junglefowl. This

represents first breakthrough in the introgression barriers beyond the production of F_1 hybrids.

Analysing a small fragment from the mitochondrial DNA (mtDNA) sequence, the first molecular study conducted in 1994 concluded that all domestic chicken breeds originated from a single subspecies, the Cochin-Chinese red junglefowl (Fumihito *et al.*, 1994). Three other subspecies of red junglefowls (Cochin-Chinese, Burmese and Javan) and three wild junglefowls (grey, Ceylon, and green) were subsequently analysed with the result further supporting a single maternal ancestral origin, possibly a red junglefowl population living in Thailand (Fumihito *et al.*, 1996). Also, it has been proposed that both the Cochin-Chinese and Burmese red junglefowls should be included in a single subspecies (Niu *et al.*, 2002). In more recent and exhaustive studies, Liu *et al.*, (2006) have concluded that four out of the five subspecies of red junglefowls may have contributed to the domestication of the chicken, the only exception being the Javan red junglefowl *G. g. bankiva*.

The debate became even more interesting in 2005 when two segments of nuclear genome and the full mtDNA genome were analysed with the results showing introgression from domestic chicken/red junglefowl in grey junglefowl (Nishibori *et al.*, 2005), thus suggesting that hybridization between the two species may have occurred in the past, therefore, opening the door for a possible grey junglefowl contribution to the domestic chicken gene pool. This was confirmed in 2008, with the demonstration that the yellow skin phenotype, observed in several breeds of domestic chicken, likely originated from the grey junglefowl (Eriksson *et al.*, 2008).

Chicken domestication centres: Archaeological evidence

The geographic origin and the timing of chicken domestication are still uncertain. The initial interpretations relied on the fossil materials excavated from several archaeological sites such as the discovery of bones in Mohenjo-Daro and Harappa in Indus Valley dated to 2,500 – 2,100 BC (Zeuner, 1963; Crawford, 1984) or earlier around 3,200 BC to 2,000 BC (Wood-Gush, 1959; Zeuner, 1963). Particularly, seals which depicts fighting cocks and clay figures of chicken at the Mohenjo-Daro led earlier writers to believed it as the staging point from which chicken subsequently dispersed to East and through Mesopotamia to Greece and Europe during the Iron age (Zeuner, 1963). The early

advancement in Harappan culture civilization could have possibly played prominent roles in chicken domestication.

In recent years, archaeological remains from 16 Neolithic sites in Northern China suggested the presence of chicken as early as 6,000 BC (West and Zhou, 1988), which predates the remains found at the Mohenjo-Daro. Because of the characteristics of the early Holocene climate in Northern China which likely have prevented the survival of red junglefowl at these sites, domestication was proposed to have occurred first in South-East Asia, after which chicken subsequently diffused to North to become established in China about 6,000 BC (West and Zhou, 1988). Using latest sequencing methodologies, ancient chicken DNA obtained from archaeological sites in China was recently analysed (Xiang *et al.*, 2014). The authors concluded that the chicken domestication first occurred in Northern China around 10,000 years ago (Xiang *et al.*, 2014). However, the outcome of the study was challenged. Particularly, it has been argued that the wild red junglefowl could not have survived at the studied archaeological sites due to unfavourable climatic conditions at the time (West and Zhou, 1988), and that the species identification of the bones is inconclusive. Moreover, one of the reported DNA polymorphisms originated from the primer sequence, undermining the confidence of the ancient DNA part of the study (Peng *et al.*, 2015; Peters *et al.*, 2015).

Chicken dispersal: Archaeological and linguistic evidence

Chicken dispersal from domestication centres to different parts of the world is through trading and human migration.

Asia: Archaeological and linguistic evidence

Chicken may have been initially domesticated in Asia for recreational purpose rather than for food. Around 1,000 BC, it became a religious symbol for the Aryans where their consumption was completely forbidden following their introduction from the Indus Valley to India (Zeuner, 1963). Written law obtained from the sacred book in the institute of Manu allowed only the wild red junglefowl to be eaten and not the tamed one (Carter, 1923). From the Indus Valley, chicken subsequently dispersed to Persia (Iran). A single bone discovery in Iran at the site of Tepe Yahya was dated to 3,900 – 3,800 BC (West and Zhou, 1988). Chicken then spread westward to Mesopotamia and Asia Minor where

it was also used for both religious activities and cockfighting (Zeuner, 1963). The word "egg-bird" obtained from the omen-texts may suggest that chicken was already known in Babylonia in the early Sumerian period and before 2,500 BC, where they were being kept for religious and traditional purposes (Carter, 1923). Domestic chicken was probably introduced into Babylonia from Persia and from there, they dispersed to Syria (Carter, 1923). In South Arabia, chicken remains found at the site of Al-Hamid in Yemen dated to about 800 – 600 BC. A strong trading network existed between the South Arabia and South Asia including the Harappan civilization, thus, supporting maritime introduction to the Arabian Peninsula (Fuller, 2003; McIntosh, 2008).

Chicken dispersal from the Indian subcontinent to South-East and East Asia, while possible, is difficult to assess given the likely presence of domestication centres in these later areas (Crawford, 1990; Liu *et al.*, 2006). It is however possible that chicken from these regions have both Chinese and Indian subcontinental origins. Japanese chickens may have been introduced from China through Korea around 300 BC – AD 300 (West and Zhou, 1988), while the diffusion to the Pacific Islands may have originated from China with linguistic evidence suggesting that chicken migrated to the Pacific following the Austronesian expansion that carried them alongside dog and pig (Bellwood, 1995). It is also speculated that chicken spread North of the Himalayas from India and then diffused to North-East Central Asia (Nishi, 1990).

Europe: Archaeological and linguistic evidences

The dispersal of chicken to Europe was likely through trading routes originating from Persia (Iran) (Carter, 1923). Aristophanes, an Athenian poet (448 – 385 BC), called the chicken the "Persian birds". The dispersion was *via* the Mediterranean Sea through trading across the Greeks colonies, including in Southern and Northern Italy. The Roman empire played a major role in the spread of chicken. Also, dispersion occurred *via* the Black Sea and Scythia where domestic chicken entered Central Europe through the South of today's Russia (Peters, 1913; Brown, 1929). The Mediterranean chicken may represent the ancestral type of chicken in Europe. Six chicken bones were discovered at the archaeological sites of the Kalythies Cave in Greece; dated to around 4,000 – 3,000 BC, they predates the Harappan civilization dates (West and Zhou, 1988).

Though chicken was already present in Europe during the late Neolithic and Bronze Age, the main time of dispersal to Europe was during the Iron Age (West and Zhou, 1988). In 700 BC, chicken were common in Greece and Persia, being primarily used for religious and/or cultural activities (Zeuner, 1963). Already, they displayed morphological characteristics still found in the breeds today such as the rose comb, beard phenotypes, muff and cockfighting characters etc (Pollard, 1977).

Once established in Europe, chicken became gradually a major agricultural commodity. In the Medieval period, beginning around 1,100 years ago, during the time of Christian's fasting where the consumption of four legged animals was forbidden, poultry breeding was given priority attention (Loog *et al.*, 2017).

Africa: Archaeological and linguistic evidence

Information about chicken dispersal to Africa is still being debated and the time of spread to the continent also remain largely unknown (Gifford-Gonzalez and Hanotte, 2011). Though multiple arrival of chicken on the continent is likely (Mwacharo *et al.*, 2013) two routes of dispersal have so far been reported. According to Kiple and Ornelas (2000), arrival was followed by the Berbers and the Phoenician expansion along the North of Africa across the Sahara and secondly, through the maritime introduction along the East African and the Red Sea coasts. It is believed that African chicken likely originated from the Indian subcontinent following the trading network with the Fertile Crescent and with the East African coast, including during the Greco-Roman period in around 100 AD to 500 (Carter, 1971). In fact, the black feathers, meat, and bones described for Mozambique chicken resembled those of India (Crawford, 1990).

Chicken became well established in Egypt following contact between the Egyptian civilization, Persia and Greece (Coltherd, 1966). MacDonald (1992) reported that the earliest known osteological remains in Africa were found in Egypt. They date to about the 18th dynasty between 1,567 and 1,320 BC (Darby *et al.*, 1977). Some of the earliest archaeological artefacts include an ostracon with an artistic design of a chicken from a painted limestone from the royal excavation sites of Ramesses IX dated around 1,425 – 1,123 BC (Carter, 1923). More so, an ostracon from a rooster, engraved on a silver bowl likely from the reign of Seti II, dated to 1,200 – 1,194 BC, was also found (Coltherd, 1966). Of more interest is the pictorial evidence depicting that of rooster's head in a mural

obtained from the tomb of Rekhmara, vizier of Thutmose III dated to about 1,504 – 1,450 BC. It resembles the conventional pea comb type and facial features of modern Asiatic breeds (Lowe, 1934; Crawford, 1984).

Chicken was probably used for recreational purposes in 18th dynasty and later formed an element of the Egyptian farmyard during Ptolemaic time (Darby *et al.*, 1977). Later, the Egyptian chickens suddenly declined after political and economic turmoil and then disappeared from the graphic records for more than five centuries until around 650 BC when they reappeared in Nubia (Carter, 1923; Coltherd, 1966).

Introduction to sub-Saharan Africa was also likely through Egypt, between the mid-second and late first millennium BC and then spreads southwards along the Nile valley to Nubia and along the Sudano-Sahelian belt to West Africa (Macdonald and Edwards, 1993).

Evidence of chicken in East Africa around 4,000 – 3,000 BC (Chami, 2007) has now been disputed (Dueppen, 2011). In Northern Ethiopia, the radiocarbon dates of chicken bones from archaeological excavation obtained from the Mezber site indicate the presence of chicken as early as 1055 - 825 BC (D'Andrea *et al.*, 2011). The presence of these chicken remains in Horn of Africa support ancient contact with South Asia and the Middle East (Woldekiros and D'Andrea, 2016).

America: Archaeological and linguistic evidence

The origin of chicken in South America remained unclear. The first chicken may have arrived before the Spanish Conquest through pre-Columbian transoceanic contact from Polynesia in the West (Pacific Islands) (Carter, 1971; Storey *et al.*, 2007). The identification of chicken bones in the archaeological sites of the Pacific Reef and Santa Cruz islands, dated to around 1,400 – 900 BC are supportive of possible introduction of chicken to South America from the Pacific (Storey *et al.*, 2010). Archaeological chicken remains at the El-Arenal site in Chile have been proposed as a pre-Columbian evidence of chicken on the South American continent (Storey *et al.*, 2007). However, the evidence was subsequently disputed (Gongora *et al.*, 2008). Undoubtedly Chicken were introduced during the Spanish Conquest diffusing through South and Central America from which they likely reached the eastern part of North America.

Genetic features, diversity and selection in domestic chicken

Chicken represents the most populous livestock species and are globally accepted irrespective of cultural, traditional and religious affiliations (Al-Nasser *et al.*, 2007; Dessie *et al.*, 2011). Domestic chicken exhibits wide range of variations in plumage colour as against the uniform plumage observed in the wild species. Domestic chicken can be categorised based on their selection history:

The indigenous domestic chicken or village chicken

There is no specific breed for village chicken and they are rather defined as ecotype (Duguma, 2006) (**Figure 9**). They emerged following domestication and are known to exhibit within population, different morphological diversities such as the comb size and morphology, wattle, plumage colour etc (Desta *et al.*, 2015). They are well adapted to different environmental pressures, very hardy, and possess natural immunity against infectious diseases (Dessie *et al.*, 2011). They are subjected to various sociocultural practises such as for spiritual offerings during traditional and religious festivals and as a result, selection for colour and other morphological variants may be high (Crawford, 1990; Dessie *et al.*, 2011). Chickens showing special visual appeals are usually subjected to intensive selection by smallholder farmers leading to increase in the frequencies of desirable traits in the population (Dana *et al.*, 2010). With little or no input, they can survive and reproduce but their productivity is very low compared to commercial chicken breeds. Mating occurs at random and is uncontrolled (Wragg *et al.*, 2012). They constitute most of the stock in the backyard farming and play a vital role in the livelihood of poor families.

Commercial and experimental chicken lines

With little or no phenotypic external diversity, they may be referred to as a breed (**Figure 10**). The development of these breeds for either experimental or for commercial purpose started in the first half of the 20th century in parallel to the emergence and development of the quantitative genetic theory with two major lines of commercial chicken established, one for meat production and the second for egg production (Tixier-Boichard *et al.*, 2011). There are now numerous breeds of chicken selected for eggs and meat production such as the white egg layers (e.g. white leghorns) that involved three-way or four-way crosses

of intensively selected grandparents, the brown egg layers which have a broader genetic spectrum and the broilers whose grandparents were derived in part from the white Cornish and White Plymouth Rock (Crawford, 1990).

Generally, these chicken types are less hardy and cannot cope under harsh environmental conditions compared to the village chicken. Furthermore, the genome diversity is lower compared to their indigenous counterparts.



Figure 9. Ethiopian Horro indigenous village chicken (credit: Desta T. T.)



Figure 10. White Leghorn commercial chicken line

Standardized breeds

These are populations selected for specific attributes often by fancy chicken breeders (**Figure 11**). They are mostly distinguished by differences in morphology such as the comb type, plumage colour and pattern, size etc (Crawford, 1990). A typical example is the Assendelfter Bantam developed around 1980s with slate-blue shanks, large rose comb, and characterised by gold and silver colour varieties of the 'ears of wheat' (Scrivener, 2009). There are other tens if not hundreds of fancy breeds. (see for examples Scrivener (2009) and Scrivener (2006)).



Figure 11. Polish fancy chicken (credit: My pet chicken blog)

References

Al-Nasser, A., Al-Khalaifa, H., Al-Saffar, A., Khalil, F., Albahouh, M., Ragheb, G., *et al.* (2007). Overview of chicken taxonomy and domestication. *World's Poultry Science Journal, 63*(02), 285-300.

Ali, S., and Ripley, S. D. (1983). *Handbook of the birds of India and Pakistan (compact edition)* (Vol. 32): Delhi Oxford University Press.

Bellwood, P. (1995). *Austronesian prehistory in Southeast Asia: homeland, expansion and transformation*: Canberra: ANU E Press, Australia.

Brown, E. (1929). *Poultry Breeding and Production* (Vol. 1 and 2): Ernest Benn Ltd., London.

936 Carter, G. F. (1971). Pre-Columbian chickens in America. In C. Riley, J. Kelley, C.
937 Pennington, and R. Rands (Eds.), *Man across the sea: problems of Pre-*
938 *Columbian contacts*: University of Texas Press, Austin.

939 Carter, H. (1923). An ostracon depicting a red junglefowl.(the earliest known drawing of
940 the domestic cock). *The Journal of Egyptian Archaeology, 9*(1/2), 1-4.

941 Chami, F. A. (2007). Diffusion in the studies of the African past: Reflections from new
942 archaeological findings. *African Archaeological Review, 24*(1-2), 1-14.

943 Coltherd, J. (1966). The domestic fowl in ancient Egypt. *Ibis, 108*(2), 217-223.

944 Crawford, R. (1984). Domestic fowl. In *Evolution of domesticated animals,* : ed L.L.
945 Mason,. London.

946 Crawford, R. (1990). *Poultry Breeding and Genetics* (Vol. 1): Elsevier Amsterdam.

947 D'Andrea, A. C., Richards, M. P., Pavlish, L. A., Wood, S., Manzo, A., and Wolde-Kiros,
948 H. (2011). Stable isotopic analysis of human and animal diets from two pre-
949 Aksumite/Proto-Aksumite archaeological sites in northern Ethiopia. *Journal of*
950 *Archaeological Science, 38*(2), 367-374.

951 Dana, N., Dessie, T., van der Waaij, L. H., and van Arendonk, J. A. (2010).
952 Morphological features of indigenous chicken populations of Ethiopia. *Animal*
953 *GeneticGresources, 46*, 11-23.

954 Darby, W. J., Ghalioungui, P., and Grivetti, L. (1977). *Food: the gift of Osiris* (Vol. 1):
955 Academic Press London.

956 Darwin, C. (1868). *The variation of animals and plants under domestication* (2 ed. Vol.
957 2): John Murray, United Kingdom.

958 Del Hoyo, J., Elliott, A., Sargatal, J., Christie, D., and de Juana, E. (2017a). Green
959 junglefowl (*Gallus varius*). In P. McGowan, G. Kirwan, and P. Boesman (Eds.),
960 *Handbook of the birds of the world alive*. Lynx Edicions, Barcelona.

961 Del Hoyo, J., Elliott, A., Sargatal, J., Christie, D., and de Juana, E. (2017b). Sri Lanka
962 junglefowl (*Gallus lafayettii*). In P. McGowan, G. Kirwan, and P. Boesman
963 (Eds.), *Handbook of the birds of the world alive*. Lynx Edicions, Barcelona.

964 Delacour, J. (1977). *The pheasant of the world* (2 ed.): Saiga Publishing Co. Ltd. Surr
965 GU26 GTD. England.

966 Dessie, T., Taye, T., Dana, N., Ayalew, W., and Hanotte, O. (2011). Current state of
967 knowledge on phenotypic characteristics of indigenous chickens in the tropics.
968 *World's Poultry Science Journal, 67*(03), 507-516.

969 Desta, T., Lawal, R., and Hanotte, O. (2015). Genetic diversity of village chickens In:
970 Third Report on chicken genes and chromosomes 2015. *Cytogenetic and Genome*
971 *Research, 145*(2), 130-133.

Dueppen, S. A. (2011). Early evidence for chickens at Iron Age Kirikongo (c. AD 100–1450), Burkina Faso. *Antiquity, 85*(327), 142-157.

Duguma, R. (2006). Phenotypic characterization of some indigenous chicken ecotypes of Ethiopia. *Livestock Research For Rural Development, 19*(29.9), 21-25.

Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt, L., *et al.* (2008). Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genetics, 4*(2). doi.org/10.1371/journal.pgen.1000010.

Fuller, D. Q. (2003). African crops in prehistoric South Asia: a critical review. *Food, Fuel and Fields. Africa Praehistorica, 15*, 239-227.

Fumihito, A., Miyake, T., Sumi, S., Takada, M., Ohno, S., and Kondo, N. (1994). One subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic ancestor of all domestic breeds. *Proceedings of The National Academy of Sciences, 91*(26), 12505-12509.

Fumihito, A., Miyake, T., Takada, M., Shingu, R., Endo, T., Gojobori, T., *et al.* (1996). Monophyletic origin and unique dispersal patterns of domestic fowls. *Proceedings of The National Academy of Sciences, 93*(13), 6792-6795.

Gifford-Gonzalez, D., and Hanotte, O. (2011). Domesticating animals in Africa: Implications of genetic and archaeological findings. *Journal of World Prehistory, 24*(1), 1-23.

Gongora, J., Rawlence, N. J., Mobegi, V. A., Jianlin, H., Alcalde, J. A., Matus, J. T., *et al.* (2008). Indo-European and Asian origins for Chilean and Pacific chickens revealed by mtDNA. *Proceedings of The National Academy of Sciences, 105*(30), 10308-10313.

Harrison, C. (1978). A new junglefowl from the Pleistocene of Europe. *Journal of Archaeological Science, 5*(4), 373-376.

Jaap, R. G., and Hollander, W. (1954). Wild type as standard in poultry genetics. *Poultry Science, 33*(1), 94-100.

Kiple, K. F., and Ornelas, K. (2000). *The Cambridge world history of food* (Vol. 2): Cambridge University Press.

Liu, Y.-P., Wu, G.-S., Yao, Y.-G., Miao, Y.-W., Luikart, G., Baig, M., *et al.* (2006). Multiple maternal origins of chickens: Out of the Asian jungles. *Molecular Phylogenetics and Evolution, 38*(1), 12-19.

Loog, L., Thomas, M. G., Barnett, R., Allen, R., Sykes, N., Paxinos, P. D., *et al.* (2017). Inferring allele frequency trajectories from ancient DNA indicates that selection on a chicken gene coincided with changes in medieval husbandry practices. *Molecular Biology and Evolution*. doi:10.1093/molbev/msx142

Lowe, P. R. (1934). XXIII.—A further Note bearing on the Date when the Domestic Fowl was first known to the Ancient Egyptians. *Ibis, 76*(2), 378-382.

MacDonald, K. C. (1992). The domestic chicken (*Gallus gallus*) in sub-Saharan Africa: A background to its introduction and its osteological differentiation from indigenous fowls (*Numidinae* and *Francolinus sp*.). *Journal of Archaeological Science, 19*(3), 303-318.

Macdonald, K. C., and Edwards, D. N. (1993). Chickens in Africa: the importance of Qasr Ibrim. *Antiquity, 67*(256), 584-590.

McIntosh, J. (2008). *The ancient Indus Valley: New perspectives*: Abc-Clio, California.

Morejohn, G. V. (1968a). Breakdown of isolation mechanisms in two species of captive junglefowl (*Gallus gallus* and *Gallus sonneratii*). *Evolution, 22*(3), 576-582.

Morejohn, G. V. (1968b). Study of plumage of the four species of the genus *Gallus. The Condor, 70*(1), 56-65.

Mwacharo, J. M., Bjornstad, G., Han, J. L., and Hanotte, O. (2013). The History of African village chickens: An archaeological and molecular perspective. *African archaeological review, 30*(1), 97-114.

Nishi, Y. (1990). Can fowls fly hundreds of miles over the Himalayas? *Asian Languages and General Linguistics*, 55-77.

Nishibori, M., Shimogiri, T., Hayashi, T., and Yasue, H. (2005). Molecular evidence for hybridization of species in the genus *Gallus* except for *Gallus varius*. *Animal Genetics, 36*(5), 367-375.

Niu, D., Fu, Y., Luo, J., Ruan, H., Yu, X.-P., Chen, G., *et al.* (2002). The origin and genetic diversity of Chinese native chicken breeds. *Biochemical genetics, 40*(5), 163-174.

Peng, M.-S., Shi, N.-N., Yao, Y.-G., and Zhang, Y.-P. (2015). Caveats about interpretation of ancient chicken mtDNAs from northern China. *Proceedings of The National Academy of Sciences, 112*(16), E1970–E1971.

Peters, J. (1913). The cock *Journal of American Oriental Society, 33*, 363-401.

Peters, J., Lebrasseur, O., Best, J., Miller, H., Fothergill, T., Dobney, K., *et al.* (2015). Questioning new answers regarding Holocene chicken domestication in China. *Proceedings of The National Academy of Sciences*, *12*(112), E2415.

Pollard, J. (1977). *Birds in Greek life and myth*: thames & Hudson, London.

Punnett, R. C. (1923). *Heredity in poultry*: Macmillan & Co. Ltd., London.

Rasmussen, P. C., and Anderton, J. C. (2005). *Birds of south Asia: the Ripley guide* (Vol. 2): Smithsonian Institution and Lynx Edicions.

Scrivener, D. (2006). *Rare poultry breeds* (1 ed.). : Crowood Press Ltd, Wiltshire.

Scrivener, D. (2009). *Popular poultry breeds* (1 ed.). : Crowood Press Ltd, Wiltshire.

1046 Storer, R. W. (1988). *Type specimens of birds in the collections of the University of*
1047 *Michigan Museum of Zoology*.

1048 Storey, A. A., Ramírez, J. M., Quiroz, D., Burley, D. V., Addison, D. J., Walter, R., *et al.*
1049 (2007). Radiocarbon and DNA evidence for a pre-Columbian introduction of
1050 Polynesian chickens to Chile. *Proceedings of The National Academy of Sciences,*
1051 *104*(25), 10335-10339.

1052 Storey, A. A., Spriggs, M., Bedford, S., Hawkins, S. C., Robins, J. H., Huynen, L., *et al.*
1053 (2010). Mitochondrial DNA from 3000-year old chickens at the Teouma site,
1054 Vanuatu. *Journal of Archaeological Science, 37*(10), 2459-2468.

1055 Wayre, P. (1976). Sonnerat's – a junglefowl threatened by fishermen. *Newsletter for*
1056 *Birdwatchers, 16*(5), 1-3.

1057 West, B., and Zhou, B.-X. (1988). Did chickens go north? New evidence for
1058 domestication. *Journal of Archaeological Science, 15*(5), 515-533.

1059 Woldekiros, H., and D'Andrea, A. (2016). Early evidence for domestic chickens (*Gallus*
1060 *gallus domesticus*) in the Horn of Africa. *International Journal of*
1061 *Osteoarchaeology, 27*(3), 329–341.

1062 Wood-Gush, D. (1959). A history of the domestic chicken from antiquity to the 19th
1063 century. *Poultry Science, 38*(2), 321-326.

1064 Wragg, D., Mwacharo, J., Alcalde, J., Hocking, P., and Hanotte, O. (2012). Analysis of
1065 genome-wide structure, diversity and fine mapping of Mendelian traits in
1066 traditional and village chickens. *Heredity, 109*(1), 6-18.

1067 Xiang, H., Gao, J., Yu, B., Zhou, H., Cai, D., Zhang, Y., *et al.* (2014). Early Holocene
1068 chicken domestication in northern China. *Proceedings of The National Academy*
1069 *of Sciences, 111*(49), 17564-17569.

1070 Zeuner, F. E. (1963). *A history of domesticated animals*: Hutchinson a Co. (Publishers)
1071 Ltd. London.

1072

Chapter 2

Signature of positive selection in red junglefowl and domestic chicken

Abstract

The red junglefowl *Gallus gallus* is the main progenitor of modern chicken having been domesticated around 6,000 BC. Environmental and human selection pressures have shaped their genome, making them one of the most diverse livestock species with a global population of more than 50 billion. Three main categories of domestic chicken have been defined, the commercial, the indigenous, and the fancy chicken populations. The genetic control for production traits has been relatively well studied in commercial chicken but the genetic mechanisms underlying the adaptation to production in the local environment remains largely unknown in the indigenous counterpart. Here, I report the analyses of indigenous chicken populations from Ethiopia, Saudi Arabia and Sri Lanka, alongside the red junglefowl, for signatures of positive selection across the autosomes using the pool heterozygosity (Hp) and SweeD composite likelihood methods. Only two candidate regions located on chromosomes 7 and 23, out of the 191 found in the red junglefowl, were shared with all the domestic chicken populations. They are significantly associated with the development of the central nervous system and response to hypoxia. Five candidate sweep regions, shared and unique to all indigenous domestic chicken populations, were found, while candidate regions unique to each domestic chicken population with significant functions in adaptation to temperature gradient, production, reproduction and immunity were identified. Our results provide new insights on the selection pressures which may have followed the domestication of the species as well as its subsequent adaptation to different agro-ecologies, which may have direct application for genetic improvement of productivity in breeding programmes.

Introduction

Chicken *Gallus gallus domesticus* was domesticated from four out of the five subspecies of red junglefowl (Fumihito *et al.*, 1994; Liu *et al.*, 2006). Since Charles Darwin proposed a monophyletic ancestry for the species, its origin as monophyletic or polyphyletic has been a subject of intense debate (Darwin, 1868; Beebe, 1918; Danforth, 1958; Morejohn, 1968; Fumihito *et al.*, 1994). Recently, introgression studies have shown genetic contributions from other junglefowls (Eriksson *et al.*, 2008) (see also chapter 4 of this thesis).

Amongst the earliest scientists and to date, there was controversial evidence as to the timing and places where domestication first occurred (Zeuner, 1963; Crawford, 1984; West and Zhou, 1988; Fumihito *et al.*, 1996; Xiang *et al.*, 2014; Peters *et al.*, 2015; Xiang *et al.*, 2015). More recently, studies from mitochondrial DNA suggested multiple centres for chicken domestication (Liu *et al.*, 2006). From there, chicken dispersed to different parts of the world through trading and human migration. They entered North Africa, the Middle East and Sri Lanka from the Indian subcontinent, while maritime introduction, originating from the South-East Asia, occurred along the coast of East Africa (Silva *et al.*, 2009; Mwacharo *et al.*, 2011). Following these migration events, natural and human selection pressures have shaped the domestic chicken genome landscape resulting in a spectrum of breeds and chicken ecotypes.

Aside from the fancy breeds, domestic chickens primarily come under two major categories; the commercial and indigenous populations. In developing countries, indigenous domestic chickens play a prominent role in the livelihood of the poor families. Being locally adapted to the environmental conditions and under the custody of women and children, they are particularly kept for eggs, meat or for dual-purpose. Also, indigenous domestic chickens showing special visual appeal such as comb type, skin and feather colours may have been selected by smallholder farmers, thereby increasing the frequencies of desirable phenotypic traits (Dana *et al.*, 2010; Desta *et al.*, 2013). Extensive phenotypic variations such as the plumage colour and other morphological characteristics, behavioural, and production traits, which are present in domestic chicken but absent in red junglefowl, are the result of domestication, adaptation to various agro-ecosystems and recent strong selection for production and/or aesthetic values (Schütz *et al.*, 2001; Keeling *et al.*, 2004; Tixier-Boichard *et al.*, 2011).

In commercial chicken lines, the genetic factors controlling growth, development, reproduction, and production traits have been well studied (Rubin *et al.*, 2010; Fu *et al.*, 2016). Meanwhile, the genetic mechanisms underlying unique adaptation to environmental pressures of the tropics and productivity remain poorly characterized in indigenous chicken. Likewise, in red junglefowl, little is known about the genetic control responsible for its adaptation and survivability under its natural habitat. Here, I analysed using whole-genome sequence data for signature of positive selection in the genome of red junglefowl and several domestic chicken populations.

Materials and methods

Sampling and sequencing

A total of 27 indigenous domestic chickens were sampled and then grouped into three populations based on countries of collection. They include, Ethiopian domestic chickens from two districts, Horro (n = 6, altitude around 2,320 m) and Jarso (n = 5, altitude around 1,870 m), Saudi Arabian domestic chickens from three villages Al Qurin (n = 2, altitude around 130 m), Goligglah (n = 2, altitude around 130 m) and Al Oyoun (n = 1, altitude around 110 m) in the Eastern Province, and Sri Lankan domestic chickens from Puttalam district (n = 11, altitude around 60 m). The agro-ecology of Horro is sub-humid with an annual rainfall of 1,685 mm and an average temperature around 19°C while the agro-ecology of Jarso is semi-arid with an average temperature of 21°C, and an annual rainfall of 700 mm (Desta *et al.*, 2013). The Eastern Province of Saudi Arabia has periods of relatively low to very high temperature ranging from 21.2^{o}C to 50.8^{o}C across the year. Puttalam district of Sri Lanka has an average annual rainfall of < 1,000 mm and a temperature of 27°C.

Collection of blood samples was done through the wing vein and the genomic DNA was extracted using the ammonium acetate precipitation protocol (Bruford *et al.*, 1998). A minimum amount of 3 µg at 30 ng/µl concentration was used for the whole genome re-sequencing. Samples were sequenced at the Beijing Genomic Institute (BGI) or at Novogene on a HiSeq 2000/2500 Illumina platform, 500 bp paired-end insert size libraries with read length that ranged from 90 - 100 bp and genome coverage of 10 X – 30 X (**supplementary Table S1**). Adapter pollutions from the raw reads as well as the sequences with quality values ≤ 5 were deleted at the source (BGI/Novogene).

For the six red junglefowls, one whole-genome sequence (15 X genome coverage) of a captive bird (Cosmopolitan Chicken Research Project (CC®P)) (http://www.ccrp.be/) and five whole genome sequences (12 X – 36 X genome coverage) from Wang *et al.* (2015) study were included in the analyses (**supplementary Table S1**). The five birds were sampled in Yunnan (altitude ~ 3,000 m) and Hainan (altitude ~ 1,840 m) provinces of China. Yunnan is a subtropical highland or humid tropical zone with an average annual rainfall that ranges from 600 mm to 2,300 mm and annual temperature that ranges from

8°C to 27°C. As for the humid tropical Hainan province, the average annual rainfall is about 2,000 mm and temperature ranges from 16°C to 29°C.

Sequence alignment and variants calling

All the 33 whole-genome sequences were independently aligned to the chicken genome of reference *Galgal* 4.0 using Burrows-Wheeler Aligner (BWA) version 0.7.5a (Li and Durbin, 2010). Sorting of the alignment files into coordinate order, marking of duplicate reads and indexing of the binary alignment map (bam) files were processed using Picard tools version 1.105 (http://picard.sourceforge.net). Using the genome analysis toolkit (GATK) version 3.4.0 (McKenna *et al.*, 2010; DePristo *et al.*, 2011; Auwera *et al.*, 2013), I ran a two-steps protocol for local realignment around insertion and deletions (indels) to clean up artifacts due to misalignments from the initial mapping steps. Finally, I applied a quality score recalibration step for each base call to remove any errors carried over during the sequencing.

To call variants, I ran "HaplotypeCaller" from GATK for each sample bam file to create a single-sample "gVCF" using the "-emitRefConfidence GVC" option. I then followed the multi-sample aggregation approach which jointly genotyped variants by merging together, records of all genome data belonging to each population. Using the "-selectType SNP" option along with the "SelectVariants" from GATK, I extracted the SNPs from the raw genotype file before filtering the extracted SNPs using "VariantFiltration". Further analyses were restricted to the bi-allelic single nucleotide polymorphisms (SNPs) using bcftools version 1.2 (Li *et al.*, 2009), autosomes (chromosomes 1 – 28), the full mitochondrial DNA (mtDNA) and the hypervariable region of the mtDNA D-loop.

The mapping metrics including the percentage of read pairs that properly mapped to the same chromosome, mean depth coverage, total reads mapped, percentage of the genome with bases covered by at least 5, 10 and 20 reads were calculated using SAMtools version 0.1.19 (Li *et al.*, 2009). Using the Ensembl "VEP" version 85 (Aken *et al.*, 2016), I predicted the consequences of the variants while the total number of SNPs in each sample/population were identified using VCFtools version 0.1.11 (Danecek *et al.*, 2011). The R based "VennDiagram" package (Chen and Boutros, 2011) was used to plot the unique and shared SNPs between the domestic chicken and the red junglefowl populations.

Population structure and relationships analyses

I removed SNPs in linkage disequilibrium to establish the genetic structure of each population and the relationships between samples using PLINK version 1.9 (https://www.cog-genomics.org/plink2). I used the following parameters: "indep-pairwise 50 10 0.3", "bp-space 1000", "maf 0.1" and "geno 0.1". I then assessed the structure of each population unsupervised using ADMIXTURE version 1.3.0 (Alexander *et al.*, 2009). Using the default (folds = 5) for cross-validation, I ran analysis for 10 clusters (*K*). For the principal component analysis (PCA), I ran the smartpca program in eigenstrat version 6.0.1 (Price *et al.*, 2006). The proportion of variance as explained by the eigenvector was calculated by dividing the corresponding eigenvalue by the sum of all the eigenvalues.

At the genome level, nucleotide diversity (π) and genetic differentiation (F_{ST}) were calculated within and between population(s) respectively in 20 kb windows with 10 kb slides using VCFtools version 0.1.11 (Danecek *et al.*, 2011). For the F_{ST}, the pairwise analysis was conducted between each domestic chicken population and the red junglefowl.

Mitochondrial DNA analysis

The full mitochondrial consensus sequence was extracted from the whole genome sequence of each of the 33 samples using "consensus" option in bcftools version 1.2 (Li *et al.*, 2009). Multiple sequence alignment was conducted for the 33 full mtDNA using ClustalX version 2.1 (Thompson *et al.*, 1997). To identify the best-fit model for the nucleotide substitution analysis, I ran the jModeltest version 2.1.7 (Darriba *et al.*, 2012). The HKY+I+G (Hasegawa, Kishino and Yano) model was selected by the Akaike Information Criterion (AIC) out of the 24 candidate models included in the procedure. It was then used to construct an unrooted maximum likelihood tree using phyml 3.0 (Guindon and Gascuel, 2003). The tree file was plotted using MEGA 7.0 (Kumar *et al.*, 2016).

To assess the haplogroup (clade) at each mtDNA sequence, I extracted the first 397 bp hypervariable region (HVR) of the D-loop from the full mitochondrial sequences using the reference sequence that has the accession number "AB098668". I then used as

reference, six haplogroup sequences (sensus (Mwacharo *et al.*, 2011)) (**supplementary Table S3**) to establish the migration source of the studied domestic chicken populations.

A haplotype data file including all the 40 HVR D-loop sequences was generated using DnaSP version 5.1 (Librado and Rozas, 2009) from which the median-joining network was constructed using network 5.0.0.1 (http://www.fluxus-engineering.com/sharenet.htm).

Selective sweeps detection

To detect the putative sweeps under positive selection, I used two methods. The first involved the pool heterozygosity (Hp) method as described in (Rubin *et al.*, 2010). It was calculated for 20 kb window size with a 10-kb sliding step using the equation:

$$\mathrm{Hp} = \frac{2\sum n_{MAJ}\sum n_{MIN}}{\left(\sum n_{MAJ}+\sum n_{MIN}\right)^2}$$

Where $\sum n_{MAJ}$ and $\sum n_{MIN}$ are the sums of major and minor allele frequencies respectively for all the SNPs in the 20 kb windows. The values for the Hp calculated for each window size were then subsequently Z-transformed using the equation:

$$\mathrm{Z(Hp)} = \frac{\mathrm{Hp} - \overline{X}(\mathrm{Hp})}{\sigma\,(\mathrm{Hp})}$$

Where $\overline{X}$ is the mean and σ is the standard deviation of the Hp.

Following Rubin *et al.* (2010), a genome-wide threshold score of Z(Hp) ≤ - 4.0 was considered as significant (**supplementary Figures S3A** and **S3B**). The size of each candidate selective sweep region was calculated by adding the number of overlapping adjacent window size above the significant threshold.

For the composite likelihood ratio method, I used SweeD (Pavlidis *et al.*, 2013) with the "folded SFS" option. The choice of the grid size varies across the genome and it depends on the length of the chromosome for which the likelihood scores were evaluated at every 100 polymorphic sites. Peaks detected using this method were used for comparison purpose with the regions detected in the Hp method. The actual fragment size of the putative sweeps was determined using the Z(Hp).

Test for homozygosity

The proportion of homozygous SNPs at a candidate selected region was calculated using the "stat" option of the VCFtools version 0.1.11 (Danecek *et al.*, 2011). Our expectation is that, regions under a selective sweep will contain a larger number of homozygous variations compared to non-selected regions. I defined a nearly or fixed candidate sweep regions as those containing at least 80% to 100% homozygous SNPs in ≥ 70% of the birds in a population.

Gene ontology and pathway analyses

To establish the biological significance of the genes found within each candidate selected region, I performed gene ontology and pathways analyses using Database for Annotation, Visualization, and Integrated Discovery (DAVID version 6.8, https://david.ncifcrf.gov/) and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (KOBAS version 3.0, http://kobas.cbi.pku.edu.cn/). The two analyses restricted over-represented genes to the Fisher exact *P* value < 0.05 default threshold.

Results

Sequencing and SNPs identification

With the reference genome size *Galgal* 4.0 at 1.07 G (Hillier *et al.*, 2004; Aken *et al.*, 2016), the clean sequence reads for each domestic chicken sample after being filtered for quality check and adapter pollutions ranged from 108.8 million to 408.9 million depending on the genome sequence coverage (10 X – 30 X coverage) (see **supplementary Table S1)**. In each domestic chicken genome, the number of nucleotides with quality score > 20 (Q20) ranged from 94% to 96% while the GC content was ~ 41%.

More than 90% of read pairs in all samples were properly mapped to the same chromosome except for red jungleowl1 sample (67.51%). Except for the red junglefowl_koen sample with 94.69% of mapped read, ≥ 97% of all reads were mapped to the reference genome. On average ≥ 97% of the bases were covered by at least 5 reads and ≥ 89% on average have their bases covered by at least 10 reads (**supplementary Table S1**).

The intermediate genomic variants generated for individual sample using the "HaplotypeCaller" from GATK (Auwera *et al.*, 2013) was used to jointly genotype all samples belonging to a population into a single variants file. Excluding the multi-allelic sites and with the exception of a few red junglefowl samples having ≥ 7 million SNPs, the average number of SNPs in each sample is ~ 6 million (~ 6 SNPs/kb). Except for JB1A25B, JB2A04B and Saudi Arabia1 having ~ 45% of their SNPs being heterozygote in comparison to the reference genome, all other samples have ~ 60% (**supplementary Table S1)**. At the population level, I identified 13.07 million SNPs in Ethiopian domestic chicken, 10.23 million SNPs in Saudi Arabian domestic chicken, 14.46 million SNPs in Sri Lankan domestic chicken and 15.31 million SNPs in red junglefowl (**supplementary Table S2**). Eleven million and five hundred thousand SNPs were shared between domestic chicken and red junglefowl, whereas 5.4 million and 3.8 million SNPs were unique to domestic chicken and red junglefowl populations respectively (**supplementary Figure S1**).

I identified 1.76 million (13% of the total SNPs number) novel SNPs in Ethiopian domestic chicken, 1.03 million (10%) in Saudi Arabian domestic chicken, 2.33 million (16%) in Sri Lankan domestic chicken and 4.45 million (29%) in red junglefowl. The distribution of the variants shows that more than 54% of all SNPs are found within intron, 30% in the intergenic region, 5.7% are upstream gene variants, 4.3% are downstream gene variants while 3' and 5' UTR variants account for 1.8% and 0.4% of the SNPs respectively. The non-coding transcript exon variants account for 1% of the SNPs, while the coding sequence variants are less than 0.00002% of the total number of SNPs (**supplementary Table S2)**.

Population structure

Population structure at autosomal level was examined using principal component (PC) (**Figure 1**) and admixture analyses (**Figure 2**). PC1 and PC2 separate all the domestic population and the red junglefowl, a result also obtained at K = 4. The other admixture plots of $5 \leq K \leq 10$ obtained from 10 runs can be found at the **supplementary Figure S2.** Comparing **Figures 1** and **2** shows that except for Saudi Arabia, where I observed that the admixed "SaudiArabia4" sample in **Figure 2** is not separated from the other Saudi birds in **Figure 1**, all other admixed birds are separated in the PC analysis. For examples, "JB1B16A" (Ethiopian Jarso) shows evidence of genetic admixture in **Figure 2.** Also,

both the admixture and PC results distinguished four of the Sri Lankan domestic chicken samples (SriLanka1, 3, 10 and 13). Similarly, the non-admixed red junglefowl3, 4 and 5 (**Figure 2**) also group together in the PC analysis.

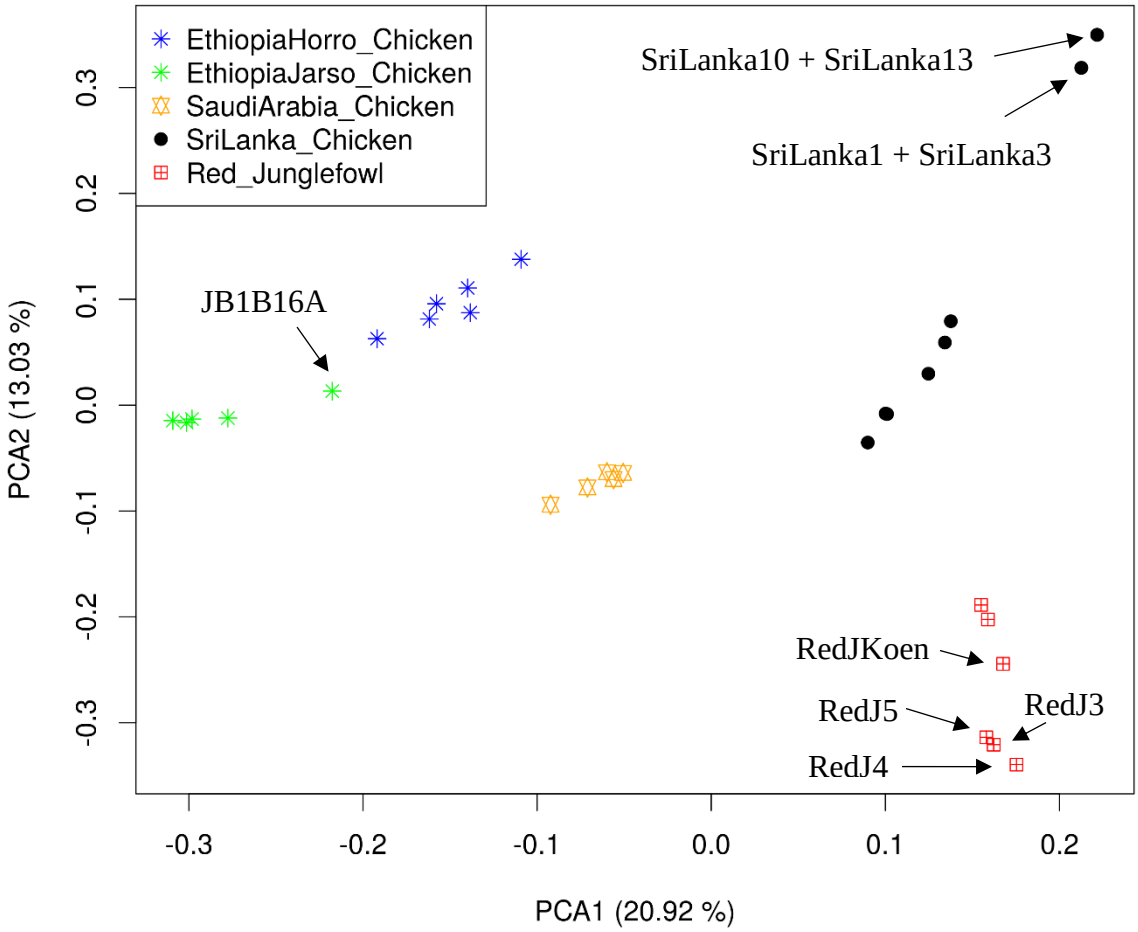


Figure 1. Principal component analysis. The top left label defined colours for each population. Individuals with name annotation have been uniquely identified for comparison purpose with **Figure 2** and **Figure 3.** Six out of the eleven samples from the Sri Lanka population overlap on each other due to low genetic variation between them. The proportion of variance explained by the eigenvector in the x- and y-axes are denoted beside the PCA1 and PCA2.

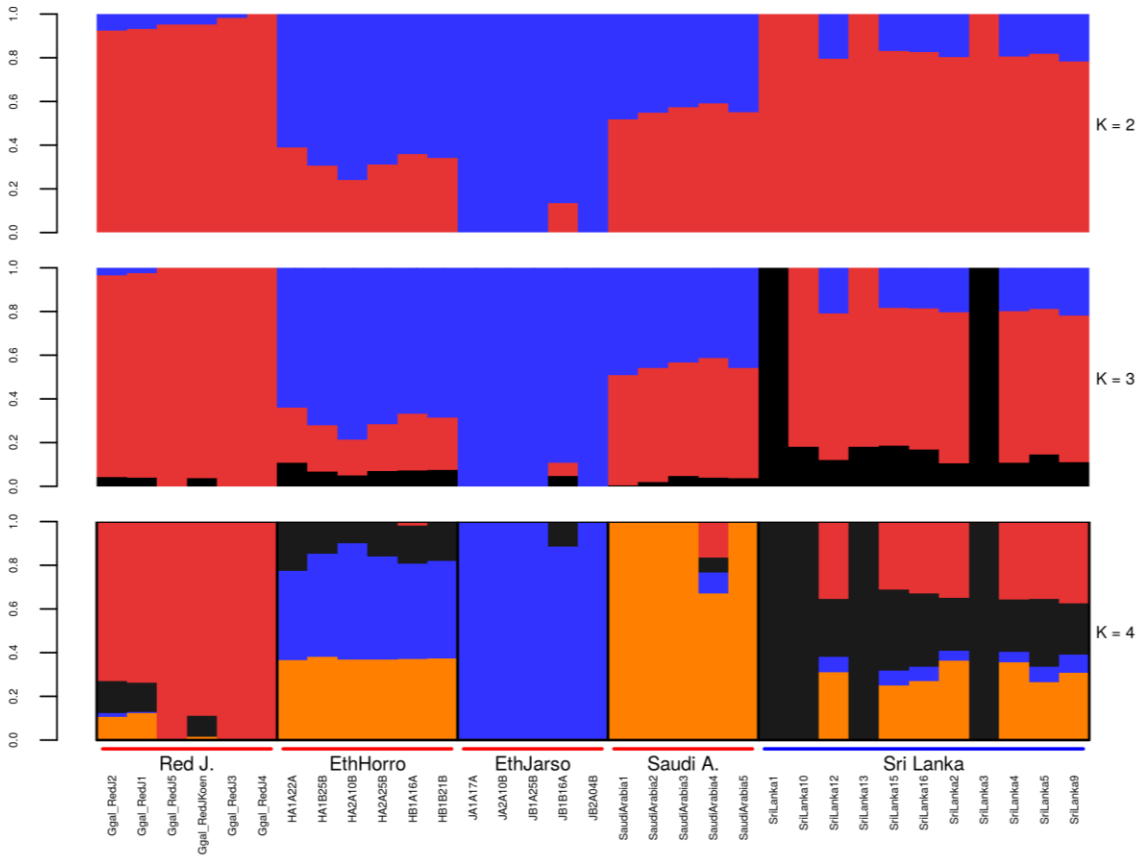


Figure 2. Autosomal admixture plots. The labels on the x-axis; Red J. (red junglefowl), EthHorro (Ethiopian Horro) and EthJarso (Ethiopian Jarso), Saudi A. (Saudi Arabia Chicken), and Sri Lanka (Sri Lanka chicken). Each population is demarcated with border lines and each admixture bar is annotated with the sample names within their respective populations. Under the Red J. label, the sample names with prefix Ggal_RedJ (1, 2, 3, 4, 5, koen) correspond to red junglefowl (1, 2, 3, 4, 5 and koen) samples in **supplementary Table S1**.

Diversity and genetic differentiation

Sri Lankan domestic chicken has the highest diversity (π = 0.0046), followed by the Ethiopian Horro domestic chicken (π = 0.0040), Saudi Arabia (π = 0.0039) and Ethiopian Jarso domestic chicken (π = 0.0036). Across populations, I observed the highest π (0.0052) in red junglefowl.

For the pairwise F_{ST} analysis, I measured the genetic distance between the red junglefowl and each of the domestic chicken population. The Jarso domestic chicken has the highest F_{ST} value (0.148), followed by Horro domestic chicken (F_{ST} = 0.113), Saudi Arabian domestic chicken (F_{ST} = 0.095), while the lowest autosomal genome difference was with Sri Lankan domestic chicken (F_{ST} = 0.062).

Mitochondrial phylogenetic relationships

The full mitochondrial sequences from the 33 individual genomes were used to construct an unrooted maximum likelihood tree using phyml 3.0 (Guindon and Gascuel, 2003) (**Figure 3**). Sri Lankan domestic chicken samples were divided into two major clusters. The first cluster belongs to the same lineage where the Ethiopian Horro and Saudi Arabian chicken were found. The second cluster was closer to the red junglefowl than to Ethiopian Jarso chicken, even though, they all belong to same lineage.

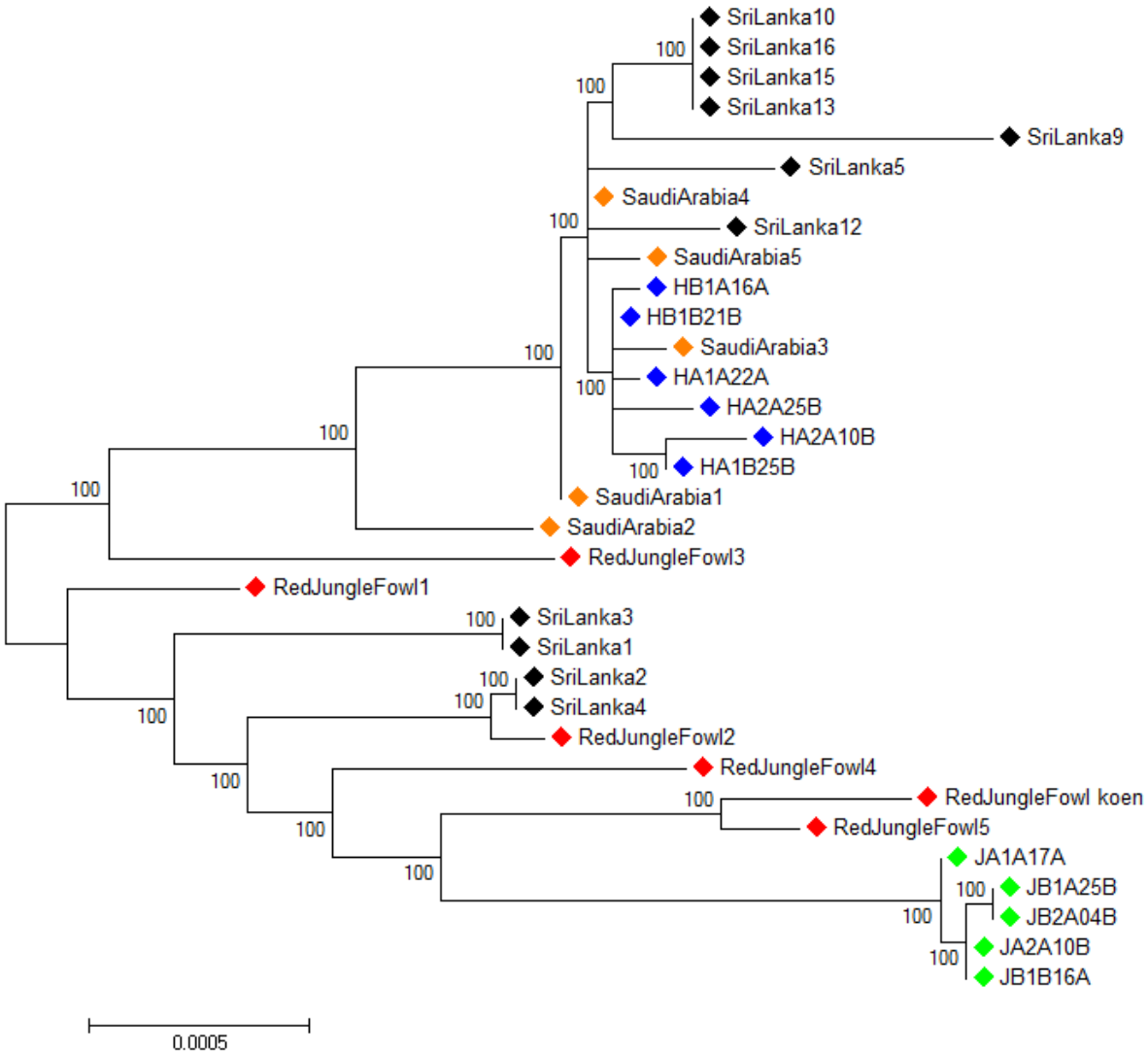


Figure 3. Unrooted maximum likelihood phylogenetic analysis for 33 full mitochondrial DNA sequences. ◆: Ethiopian Horro chicken, ◆: Ethiopian Jarso chicken, ◆: Saudi Arabian chicken, ◆: Sri Lankan chicken, ◆: Red junglefowl

To assess the maternal diversity origin of our indigenous domestic chicken populations, I extracted the hypervariable region (spanning the first 397 bp) of the mitochondrial D-

loop sequence. I included haplotypes from six haplogroups of reference (**supplementary Table S3**). Haplogroups A, B, C, and D were observed in our dataset (**Figure 4**). Within a single mutation, all Horro, four Saudi Arabian and two Sri Lankan domestic chicken haplotypes were linked to haplogroup D. Four Sri Lankan haplotypes separated by three mutations are also close to haplogroup D. Other Sri Lankan domestic chicken haplotypes (n = 5) form a clade with haplogroups B and C and a single Saudi Arabian4 haplotype is also close to haplogroup B. A single haplotype, not directly linked to any of the reference haplotypes, was observed in Jarso chicken.

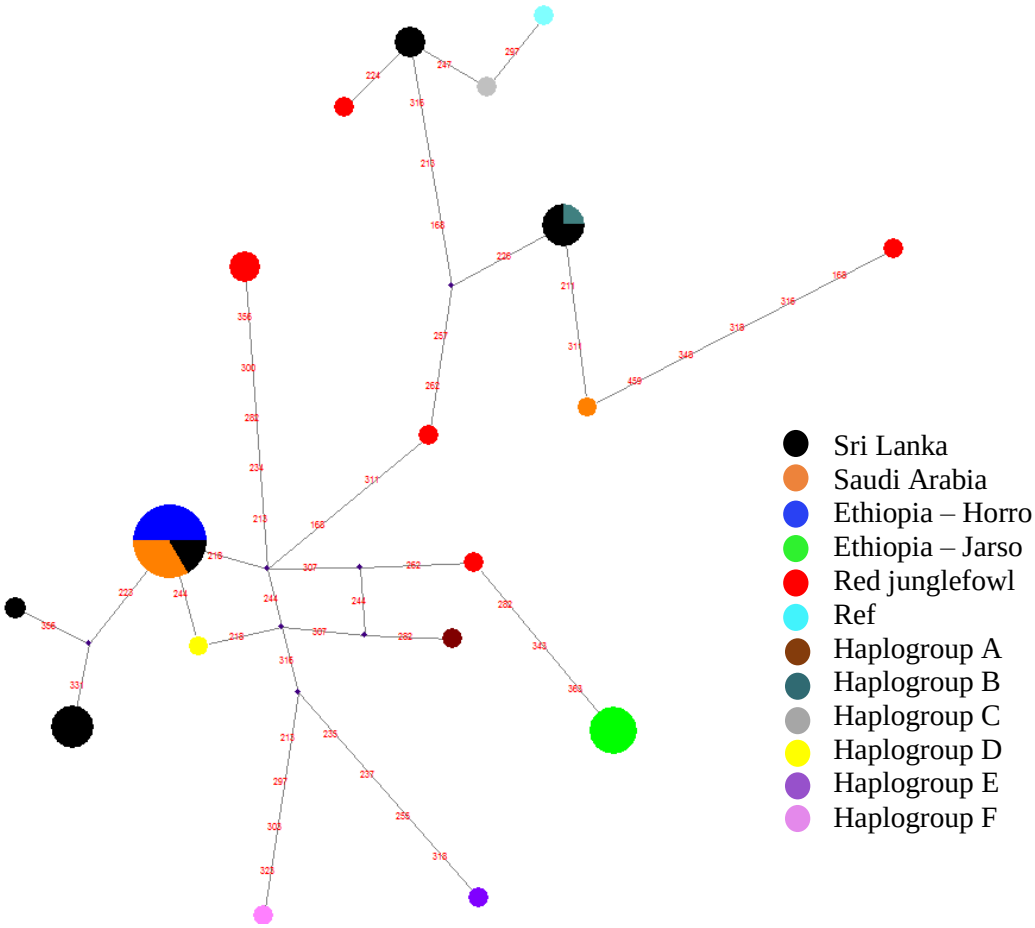


Figure 4. Haplotype median-joining network analysis for the hypervariable D-loop region of the mitochondrial DNA. The red number on the lines represents the mutation sites. Each colour represents a population or a haplotype of reference from previously described haplogroups (Mwacharo *et al*., 2011).

Mean genome heterozygosity

Across the populations, I calculated the average level of genome heterozygosity using the result obtained from the Hp method. Among the domestic chicken populations, chickens

from Ethiopia showed the lowest level of overall genome heterozygosity (mean 0.31 ± 0.051, n = 11) followed by chickens from Sri Lanka (0.32 ± 0.039, n = 11), while the Saudi Arabian chickens have the highest level of heterozygosity (0.36 ± 0.048, n = 5). For the red junglefowl, the mean genome heterozygosity is 0.32 ± 0.028 (n = 6) (**Table 1**).

Genome-wide SNPs restriction for Hp sweep detection

Since accuracy of sweep detection will depend on the number of SNPs in each window and considering the high polymorphisms identified within populations, I selected windows with minimum of 50 SNPs. Following this criterion, 52, 103, 56 and 39 windows with SNPs < 50 were excluded from the Ethiopian, Saudi Arabian, Sri Lankan and red junglefowl populations respectively.

Selective sweeps detection in domestic chicken populations

From the total number of 89,443 windows analysed in the Ethiopian domestic chicken population, 247 windows passed the genome-wide threshold of ≤ -4. They defined 85 candidates sweep regions (**Table 1**). The 50-kb candidate region found on chromosome 5 (40060000 - 40110000 bp) has the lowest Z(Hp) score (-5.8 ± 0.289) in the genome (**supplementary Table S4A**). It harbours the *TSHR* locus (Rubin *et al.*, 2010). Genome-wide, the largest size candidate sweep region is located on chromosome 8 (210 kb size, between 420000 and 630000 bp) with Z(Hp) score of -4.29 ± 0.055 (**Figure 5a** and **supplementary Table S4A**). Three other candidate regions have a fragment sizes > 100 kb, two on chromosome 3 with a size of 110 kb (position 102360000 - 102470000 bp) and 150 kb (position 102720000 - 102870000 bp) and one 110 kb region on chromosome 8 (position 160000 - 270000 bp) (**supplementary Table S4A)**. Analysing the fragment size of each sweep region above the genome-wide threshold of ≤ -4 shows that 37 candidate regions have a 20-kb size each, 13 candidate regions have a 30-kb size each, nine candidate regions each have a 40-kb size each, ten candidate regions each have a 50-kb size, and 16 candidate regions each have sizes of 60 kb and above (**supplementary Table S4A**). No significant peaks were observed on chromosomes 6, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 25 26, 27 and 28 (**Figure 5a)**.

1435 **Table 1**. Genome-wide pool heterozygosity statistics

Population	**Sample No**	**Pool heterozygosity (Hp) statistics**			
		Total windows	Genome mean (Hp)	Z(Hp) ≤ - 4.0[a]	Sweep regions identified
Ethiopian domestic chicken	11	89,443	0.31 ± 0.051	247	85
Saudi Arabian domestic chicken	5	87,646	0.36 ± 0.048	565	213
Sri Lankan domestic chicken	11	89,701	0.32 ± 0.039	299	127
Red junglefowl	6	90,170	0.32 ± 0.028	434	191

1436 [a] Total number of windows that passed the genome-wide threshold

For the Saudi Arabian domestic chicken population, I identified a total of 87,646 windows and only 565 of them passed the genome-wide threshold, defining 213 candidate sweep regions (**Table 1**). The peak with the lowest Z(Hp) score (-7.27 ± 0.087) is a 30-kb region on chromosome 8 (position 200000 - 230000 bp) (**supplementary Table S4B**). As for the Ethiopian domestic chicken population, the largest size sweep region in the genome is located on chromosome 8 (210 kb size, between 420000 and 630000 bp) with a Z(Hp) score of -4.78 ± 0.272 (**Figure 6a** and **supplementary Table S4B)**. Five other candidate selected regions have fragment sizes > 100 kb including two regions on chromosome 2 (a 140 kb at position 75140000 - 75280000 bp) and a 110-kb region (position 146690000 - 146800000 bp), one region on chromosome 4 (120 kb length, position 28120000 - 28240000 bp) and two regions on chromosome 8 (a 190-kb length at position 8730000 - 8920000 bp and a 110-kb region at position 9020000 - 9130000 bp) (**supplementary Table S4B).** Analysing the selective sweep fragment sizes show that 81 candidate regions out of the total of 213 have a 20-kb fragment size each, 55 candidate regions have a size of 30 kb each, 27 candidate regions have a 40-kb size each, 15 candidate regions each have a 50-kb size and 35 candidate regions each have a 60-kb size and above (**supplementary Table S4B**). No significant peaks were observed on chromosomes 12, 13, 16, 17, 19, 20, 21, 22, 24, 25 26, 27 and 28 (**Figure 6a).**

In Sri Lankan domestic chicken populations, a total of 89,701 windows from which 299 windows passed the genome-wide threshold resulting in 127 candidates sweep regions were identified (**Table 1**). Like the Ethiopian domestic chicken population, the lowest Z(Hp) score (-6.32 ± 1.634) in the genome is a 50-kb region located on chromosome 5 (40060000 - 40110000 bp) (**Figure 7a** and **supplementary Table S4C)**. The candidate region with the largest fragment size in the genome is a 290-kb located on chromosome 2 (81950000 - 82240000 bp) with a Z(Hp) score of - 4.65 ± 0.454. Two other candidate regions have fragment sizes > 100 kb. It includes a 130-kb region (chromosome 3, position 110160000 – 110290000 bp) and a 220-kb region (chromosome 5, position 21670000 – 21890000 bp). Analysing the 127 total candidate regions show that 63 regions have a 20-kb fragment size each, 30 candidate regions have a 30-kb size each, 15 candidate regions have a 40-kb size each, five candidate regions each have a 50-kb size and 14 candidate regions each have a size of 60 kb and above (**supplementary Table S4C**). No significant peaks were observed on chromosomes 13, 14, 15, 16, 17, 19, 20, 21, 24, 25 and 28 (**Figure 7a)**.

Selective sweeps detection in red junglefowl

A total of 434 windows passed the genome-wide threshold out of 90,170 total windows analysed resulting in 191 candidate sweep regions (**Table 1**). At the genome level, a single window of 20 kb located on chromosome 5 (51720000 – 51740000 bp) has the lowest Z(Hp) score (-5.93) (**Figure 8a** and **supplementary Table S5**). The region with the largest fragment size in the genome, a 210-kb region at position 2350000 - 2560000 bp is located on chromosome 22 with a Z(Hp) score of - 4.63 ± 0.653 (**supplementary Table S5**). Two other candidate regions with a fragment size > 100 kb are also present. It includes a 110 kb (chromosome 2, position 30000 – 140000 bp) and a 150 kb (chromosome 22, position 570000 – 720000 bp) regions. Ninety-one candidate sweep regions out of the total of 191 have fragment size of 20 kb each, 44 candidate regions each have size of 30 kb, 14 candidate regions each have size of 40 kb, 17 candidate regions each have sizes of 50 kb and 25 candidate regions each have sizes of 60 kb and above (**supplementary Table S5**). No significant peaks were observed on chromosomes 14, 16, 20, 21, 24, 25, 27 and 28 (**Figure 8a**).

Overlapping sweep regions across populations

At the genome level, only two sweep regions were shared between all domestic chicken and red junglefowl populations. They include a 20-kb candidate region located on chromosome 7. The closest gene to this region, within a 186-kb distance in the reverse strand is *TMEFF2*. The second region is 60 kb in size and it is located on chromosome 23. It includes seven functional genes (*NT5C1A, HPCAL4, TRIT1, MYCL, MFSD2A, CAP1* and *NIPAL3*) (**Table 2**). This region on chromosome 23 overlaps with an introgressed region in chapter 5.

Comparing the candidate selected regions in each domestic chicken population with the red junglefowl ones, I observed three overlaps with the Ethiopian domestic chicken, ten with the Saudi Arabian domestic chicken and 12 with Sri Lankan domestic chicken population (**Table 2**).

Five candidate sweep regions were commonly found across the three domestic chicken populations. They are located on chromosomes 1, 2 and 5. The 50-kb chromosomal region on chromosome 5 (position: 40060000 - 40110000 bp) includes two genes *TSHR* and *GTF2A1* (**Table 2**).

Between domestic chicken populations, 19 candidate sweep regions out of the total of 71 were uniquely shared between the Ethiopian and Saudi Arabian domestic chicken populations (**Table 2**). Four of these candidate regions overlap with database annotated genes; *HMGN3* (chromosome 3), *LCORL* (chromosome 4), *C14orf37* (chromosome 5) and *GK5* (chromosome 9) (**Table 2**). Two out of the six uniquely shared candidate regions between the Ethiopian and Sri Lankan domestic chicken populations have been assigned genes and include *TACR3* (chromosome 4) and *PLOD5* (chromosome 9). The genes present on the 13 uniquely shared candidates sweep regions between Saudi Arabian and Sri Lankan domestic chicken populations are *SEMA3A* (chromosome 1), *5S_rRNA* and *KCNQ3* (chromosome 2), *RIMS1* (chromosome 3), *ARHGAP22* (chromosome 6), *BAZ2B, Mar-07* and *NAA20* (chromosome 7) (**Table 2**).

Sweep detection using SweeD

I also used the site frequency spectra patterns of SNPs using SweeD (Pavlidis *et al.*, 2013) to identify candidate polymorphisms that may be under strong selection. I then compared the candidate regions obtained with this method with the result of the Hp analysis. **Figures 5b, 6b, 7b** and **8b** show the genome-wide plots obtained from SweeD analysis. Overlapping regions between the two methods were found only for a 50 kb (chromosome 9, position 9740000 -9790000 bp) region within Ethiopian domestic chicken and three regions shared in the genome of the Ethiopian and Saudi Arabian domestic chicken populations (chromosome 3, position 102720000 - 102870000 bp 150 kb in length; chromosome 4, position 75400000 – 75420000 bp, 20 kb in length; chromosome 8, position 420000 - 630000 bp, 210 kb in length).

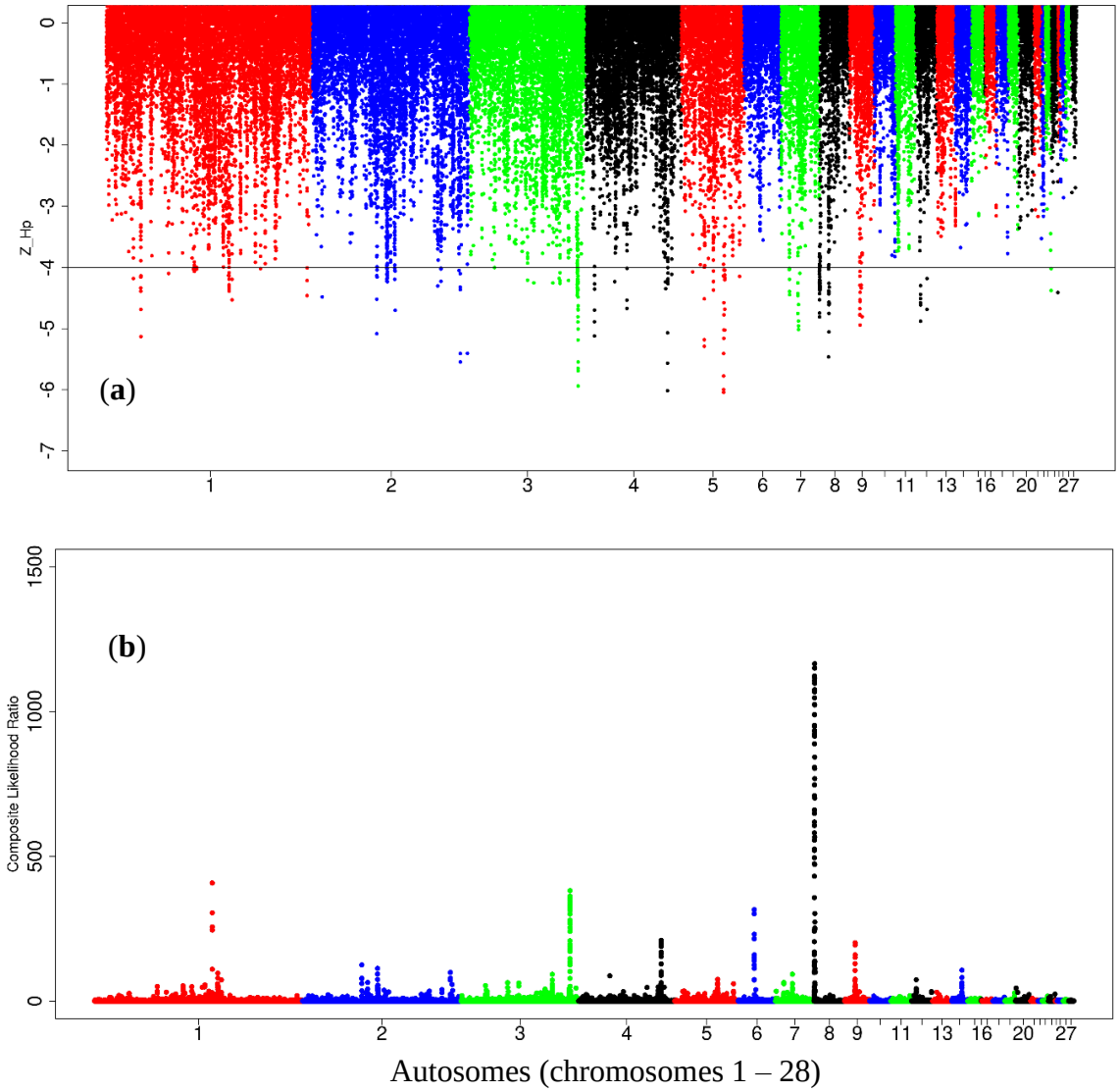


Figure 5. Candidate signatures of positive selection in Ethiopian domestic chicken population using two different comparison methods: (**a**) the standardized pool heterozygosity Z(Hp), (**b**) SweeD composite likelihood ratio test.

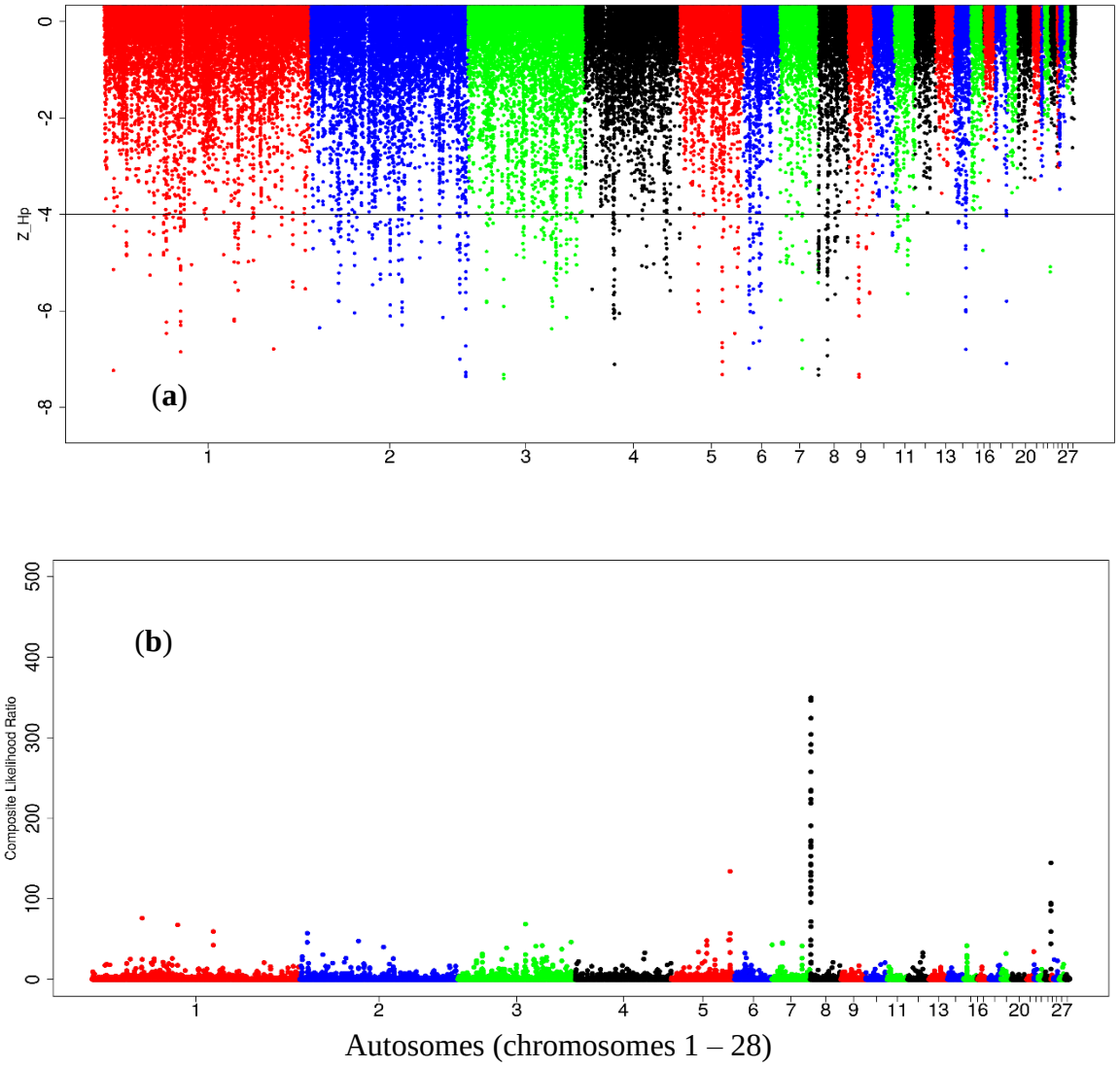


Figure 6. Candidate signatures of positive selection in Saudi Arabian domestic chicken population using two different methods: (**a**) the standardized pool heterozygosity Z(Hp), (**b**) SweeD composite likelihood ratio test.

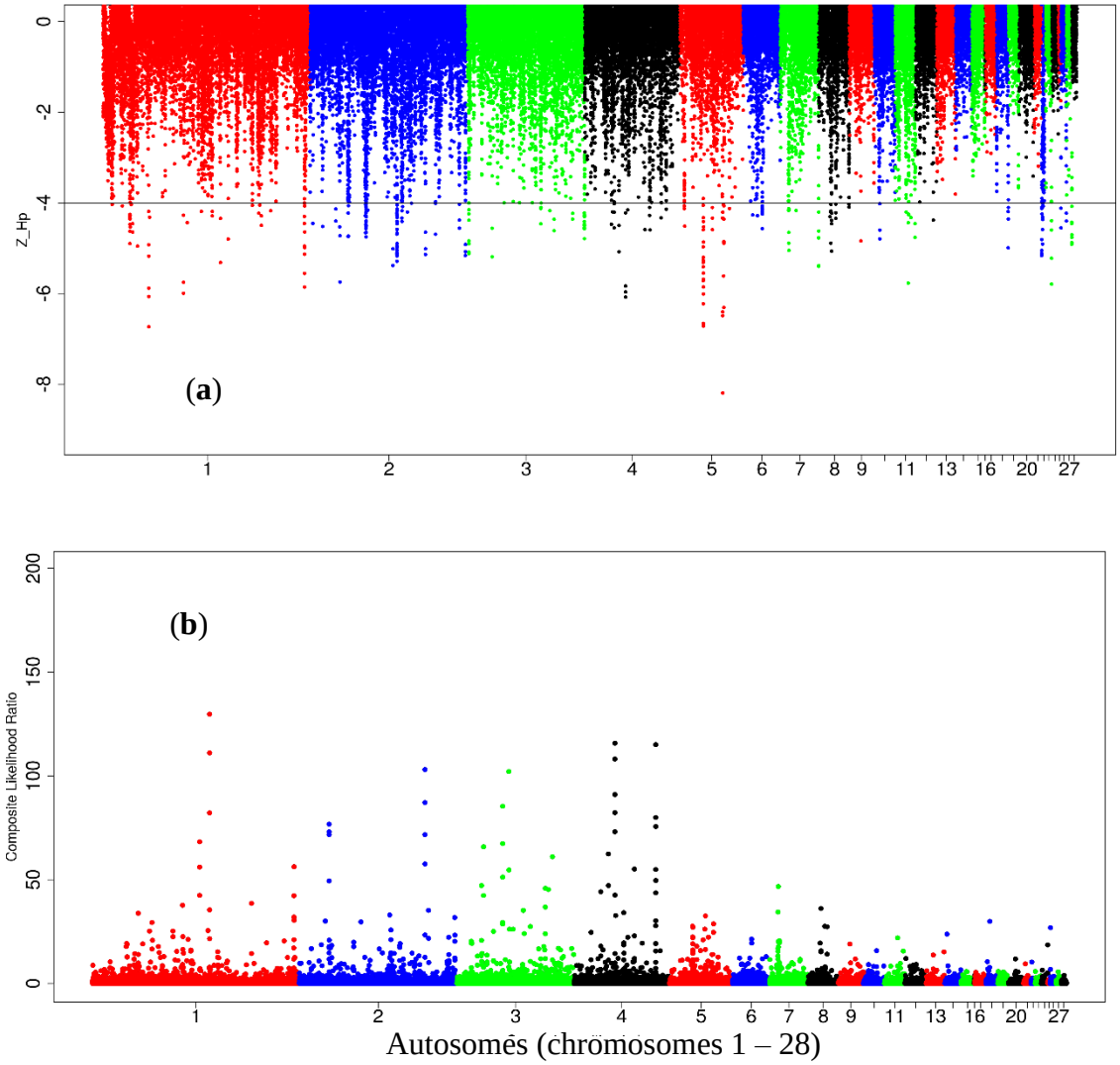


Figure 7. Candidate signatures of positive selection in Sri Lankan domestic chicken population using two different methods: (**a**) the standardize pool heterozygosity Z(Hp), (**b**) SweeD composite likelihood ratio test.

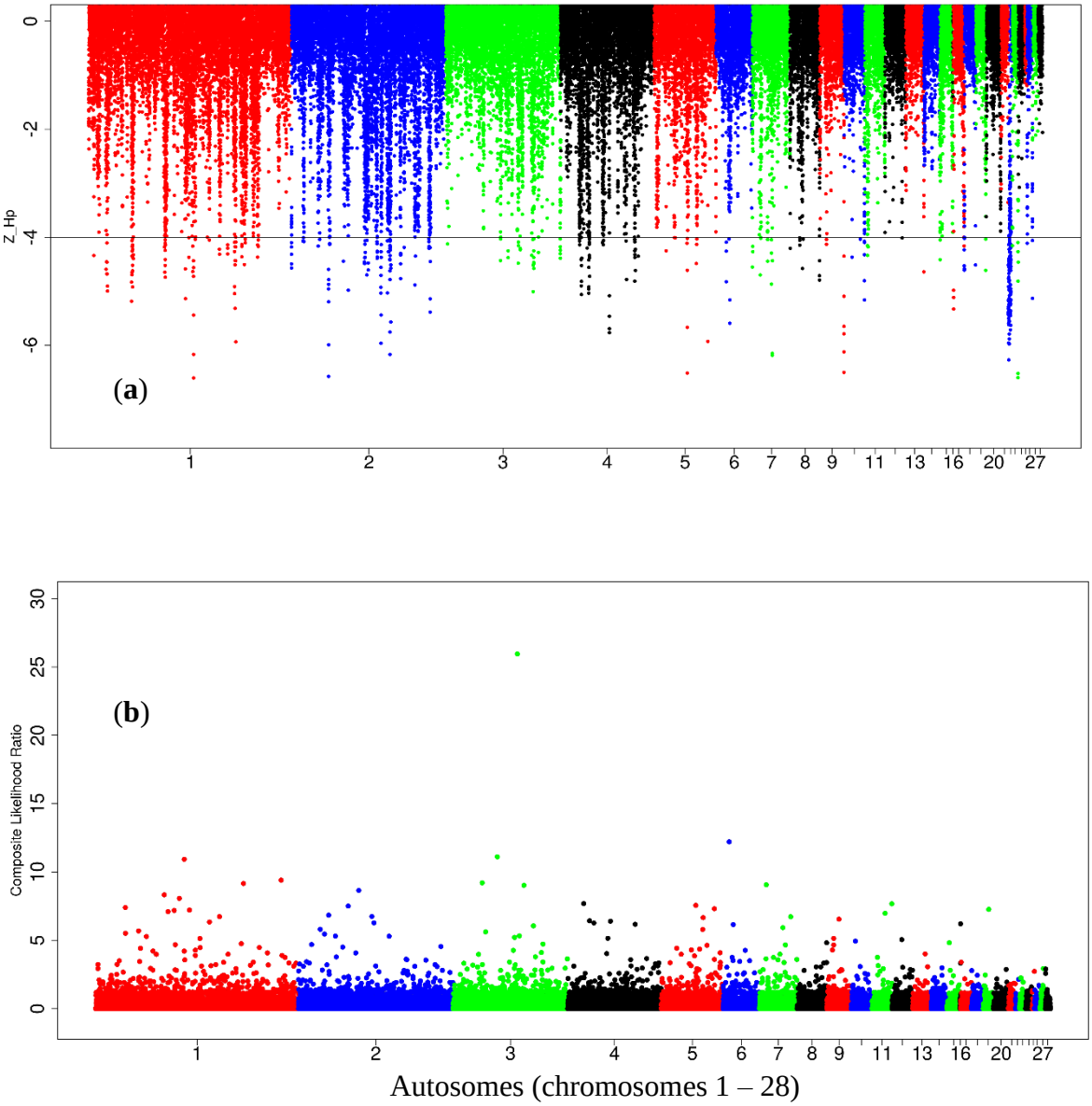


Figure 8. Candidate signatures of positive selection regions in red junglefowl population using two different comparison methods: (**a**) the standardize pool heterozygosity Z(Hp), (**b**) SweeD composite likelihood ratio test.

Table 2. Overlapping candidates sweep regions detected by the Hp method between populations. The (**x**) means that the region is present in that population. *****Domestic chicken populations. **[a]**Overlapping regions among all domestic populations. [b]Overlapping regions between all domestic chicken populations and red junglefowl.

Chr	Start	Stop	Length (kb)	Ethiopian*	Saudi Arabian*	Sri Lankan*	Red Junglefowl	Genes
1	8510000	8540000	30		x	x		*SEMA3A*
1	25440000	25460000	20	x		x		
1	32430000	32450000	20	x		x		
1	58890000	58910000	20	x	x			
1	82370000	82390000	20	x	x			
1	140990000	141050000	60		x		x	
1[a]	190070000	190110000	40	x	x	x		
2	28250000	28290000	40		x	x		*5S_rRNA*
2	35840000	35920000	80		x		x	
2	60840000	60880000	40	x	x			
2	70370000	70400000	30	x			x	
2	70640000	70670000	30	x			x	
2	70840000	70880000	40	x	x		x	
2	78310000	78340000	30	x		x		
2	81900000	81920000	20		x	x		
2	86210000	86280000	70		x		x	
2	86350000	86380000	30		x		x	
2	92550000	92570000	20		x		x	*MIR1803*
2	95750000	95790000	40		x	x	x	
2	138730000	138750000	20	x	x			
2	141000000	141070000	70		x	x		*KCNQ3*
2	146610000	146630000	20		x	x		
2[a]	146720000	146740000	20	x	x	x		
3	52820000	52840000	20		x		x	

3	78890000	78910000	20	x	x			*HMGN3*
3	81610000	81630000	20		x	x		*RIMS1*
3	84040000	84070000	30		x		x	
3	102720000	102870000	150	x	x			
3	110220000	110280000	60			x	x	*EVA1A*
4	26410000	26430000	20		x	x		
4	27110000	27130000	20	x	x			
4	27860000	27910000	50		x	x		
4	38620000	38660000	40	x		x		*TACR3*
4	41190000	41210000	20		x		x	
4	75400000	75420000	20	x	x			*LCORL*
4	77130000	77180000	50	x	x			
4	77410000	77490000	80	x	x			
5	21810000	21850000	40	x		x		
5 [a]	39870000	39930000	60	x	x	x		
5 [a]	40060000	40110000	50	x	x	x		*TSHR, GTF2A1*
5 [a]	41050000	41140000	90	x	x	x		
5	41060000	41080000	20	x			x	
5	51720000	51740000	20		x		x	*AHNAK2*
5	55190000	55210000	20	x	x			*C14orf37*
6	13300000	13330000	30			x	x	*KCNMA1*
6	17860000	17890000	30		x	x		*ARHGAP22*
7	470000	500000	30		x		x	*COL5A2*
7	7860000	7920000	60	x		x		
7 [b]	8070000	8090000	20	x	x	x	x	
7	14780000	14830000	50			x	x	*TTN*
7	36160000	36190000	30		x	x		*BAZ2B, Mar-07*
7	36200000	36230000	30		x	x		*NAA20, Mar-07*
8	160000	270000	110	x	x			

8	420000	630000	210	x	x			
8	8740000	8760000	20	x	x			
8	8810000	8830000	20	x	x			
8	8900000	8920000	20	x	x			
8	9050000	9130000	80	x	x			
8	9420000	9450000	30	x	x			
8	16040000	16120000	80		x	x		
8	28540000	28570000	30			x	x	*TYW3*
9	9870000	9900000	30	x	x			*GK5*
9	10960000	10990000	30	x		x		*PLOD5*
18	10910000	10930000	20			x	x	*SLC16A5, ARMC7*
22	230000	310000	80			x	x	*ANTXR1, GKN2, BMP10, ARHGAP25*
22	570000	720000	150			x	x	*PPP2R2A, EBF2*
22	1060000	1090000	30			x	x	
22	1090000	1180000	90			x	x	*STC1, SLC25A37*
22	1200000	1250000	50			x	x	*ENTPD4, LOXL2, R3HCC1*
22	1510000	1610000	100			x	x	*LZTS1, ATP6V1B2, SLC18A1, MAK16, TTI2, RNF122*
23 [b]	5450000	5510000	60	x	x	x	x	*NT5C1A, HPCAL4, TRIT1, MYCL, MFSD2A, CAP1, NIPAL3*
26	5190000	5240000	50			x	x	*PRELP, OPTC*

1556

Homozygosity test

I selected two candidate regions that were shared in all the domestic chicken, two candidate regions shared between all the domestic chicken and the red junglefowl and the two largest candidate selected regions (150 kb and 210 kb) detected by the two selection signature methods within the genome of Ethiopian and Saudi Arabian chicken for the estimation of the homozygote SNPs proportion and I assess their levels of homozygosity. (**Table 3).**

For the 50 kb *TSHR* and *GTF2A1* loci, the proportion of homozygote SNPs for all the indigenous domestic chicken samples ranges from 79.35% to 100% (**Table 3**). With the exception of two "red junglefowl5" and "red junglefowl_koen" samples, where the observed SNPs homozygosity was 35.72% and 31.78% respectively, all other red junglefowl samples display at least 80% homozygote SNPs for the *TSHR* and *GTF2A1* loci.

The shared 20 kb candidate sweep region on chromosome 7 between all the domestic chicken and red junglefowl shows that most samples have ≥ 82% homozygous SNPs, while few samples fall a little below the threshold but with ~ ≥ 70% homozygous SNPs. Except "HA2A10B" and "Saudi Arabia4" samples, more than 98% homozygous SNPs were observed for the 210-kb region on chromosome 8 which has the largest fragment size in the genome of Ethiopian and Saudi Arabian chickens as detected by the two methods. Only two samples in Sri Lankan domestic chicken passed the cut off for this region, which is consistent with the lack of strong peak in **Figures 7a** and **7b.** I detail results obtained for other regions in **Table 3.**

1587 **Table 3.** Homozygosity for some selected sweep regions. Values in italics falls below the arbitrary 80% cut off threshold.

Sweep regions	**Chr1: 190.07 – 190.11 Mb**		**Chr5: 40.06 - 40.11 Mb**		**Chr7: 8.07 - 8.09 Mb**		**Chr23: 5.45 -5.51 Mb**		**Chr3: 102.72 – 102.87 Mb**		**Chr8: 0.42 – 0.63 Mb**	
Size	**40 kb[a]**		**50 kb[a]**		**20 kb[b]**		**60 kb[b]**		**150 kb[c]**		**210 kb[c]**	
Samples/	**∑SNPs**	**% Hom**	**∑SNPs**	**% Hom**	**∑SNPs**	**% Hom**	**∑SNPs**	**% Hom**	**∑SNPs**	**% Hom**	**∑SNPs**	**% Hom**
HA2A25B - Ethiopia	152	98.68	357	99.44	70	*74.29*	90	*67.78*	461	99.35	1059	99.34
HA1B25B - Ethiopia	152	99.34	355	100	60	96.67	114	*44.34*	464	98.70	1057	99.62
HB1A16A - Ethiopia	162	88.27	357	99.44	69	*73.91*	82	*69.51*	464	98.92	1060	99.43
HB1B21B - Ethiopia	153	98.04	361	93.35	60	96.67	80	*66.25*	459	99.56	1057	99.62
HA1A22A - Ethiopia	161	88.20	363	92.84	60	96.67	72	100	463	100	1068	98.41
HA2A10B - Ethiopia	161	88.20	362	99.04	60	98.33	77	*76.62*	466	98.01	1062	*0.38*
JA2A10B - Ethiopia	150	100	355	100	59	98.31	69	98.55	459	99.78	1056	99.62
JB1B16A - Ethiopia	151	99.34	355	100	61	96.72	76	*76.32*	522	81.99	1058	98.96
JB1A25B - Ethiopia	288	*30.56*	356	99.72	67	82.09	91	*68.13*	549	*73.22*	1060	98.96
JB2A04B - Ethiopia	275	*33.45*	358	100	61	92.71	84	*67.86*	471	91.93	1056	100
JA1A17A- Ethiopia	148	100	355	100	60	98.33	90	*68.89*	499	83.37	1060	98.96
Saudi Arabia1	147	100	380	87.89	66	*68.18*	81	*74.07*	548	95.07	1045	99.62
Saudi Arabia2	257	*35.80*	374	93.05	57	100	87	*67.82*	562	*50.71*	1059	98.77
Saudi Arabia3	141	100	374	86.36	58	100	77	*59.74*	477	86.58	1021	99.80
Saudi Arabia4	198	*55.05*	356	99.72	65	*78.46*	104	*57.69*	686	*56.27*	1047	*32.95*
Saudi Arabia5	146	98.63	386	82.38	59	93.22	70	*67.14*	468	92.74	1043	99.23
Sri Lanka1	155	97.42	379	87.60	62	90.32	93	*62.37*	754	*23.74*	1042	*0.58*
Sri Lanka2	166	87.95	361	97.78	60	98.33	95	*63.16*	31	100	1098	*35.25*
Sri Lanka3	155	97.41	380	87.63	62	90.32	96	*61.46*	758	*24.01*	1038	*0.48*
Sri Lanka4	160	88.75	382	90.31	66	84.85	74	*63.51*	457	*7.44*	1063	98.59
Sri Lanka5	291	*27.49*	358	99.16	69	*75.36*	121	*47.11*	679	*34.17*	1087	*0.46*
Sri Lanka9	158	96.20	361	93.35	69	*26.09*	83	*74.70*	504	83.53	1061	*0.75*
Sri Lanka10	167	88.62	361	96.95	60	98.33	72	*62.50*	676	*62.13*	1055	*0.76*

Sri Lanka12	234	*35.47*	359	100	60	91.67	147	*37.41*	666	*50.00*	16	*25*
Sri Lanka13	162	88.27	360	97.22	60	98.33	74	*59.46*	677	*37.37*	1054	*0.66*
Sri Lanka15	162	88.88	363	100	60	98.33	118	*39.83*	616	*19.48*	1060	*1.23*
Sri Lanka16	154	98.70	402	*79.35*	61	90.16	80	*77.50*	679	*50.22*	1056	99.53
Red junglefowl1	151	98.01	396	80.05	58	100	77	100	600	*48.17*	1101	98.64
Red junglefowl2	153	98.69	360	96.94	66	*74.24*	96	*62.50*	739	*33.42*	38	*21.05*
Red junglefowl3	222	*53.60*	319	85.60	70	*68.57*	98	*63.27*	327	98.78	995	*26.63*
Red junglefowl4	259	*24.35*	362	87.29	61	*8.20*	76	*77.63*	694	*36.74*	971	*0.82*
Red junglefowl5	284	*26.76*	459	*35.72*	72	*72.22*	161	*37.89*	802	*16.58*	1346	*44.80*
Red junglefowl_koen	284	*32.04*	472	*31.78*	63	82.54	158	*40.51*	638	*58.78*	1305	*51.88*

[a]Candidate regions shared in all domestic chicken. [b]Candidate regions shared between all domestic chicken and red junglefowl. [c]The two regions with largest fragment size in the genome of Ethiopian and Saudi Arabian chickens as detected by both the Hp and SweeD methods. ∑**SNPs** – sum of all SNPs within the genomic candidate region. **% Hom -** Percentage homozygosity. **Mb** – Megabase pair

Functional annotations for the enriched genes within sweep regions

To identify cluster functions of key genes that may play significant roles in the adaptation to production environment and domestication process, I performed enrichment analysis for individual populations using the Ensembl genes obtained from Ensembl "BioMart" 85.0 (Zerbino *et al.*, 2015). Only classes of genes with default Fisher exact *P*-value < 0.05 were considered overrepresented for the gene ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway analysis

In Ethiopian domestic chicken, I identified one GO term, "regulation of vasoconstriction" (two genes: *HRH1* and *AGTR1*) and three KEGG pathways, "neuroactive ligand-receptor interaction" (four genes: *AGTR1, TACR3, HRH1* and *TSHR*), "calcium signaling" (three genes: *AGTR1, TACR3* and *HRH1*) and "adrenergic signaling in cardiomyocytes" (two genes: *AGTR1* and *ATP1B3*) (**supplementary Table S6A** and **S7A**).

In Saudi Arabian domestic chicken, I identified 23 GO terms: "blood circulation" (eight genes: *PRKG1, PDE5A, SEMA3A, PDE3A, HDAC4, CTNNA3, CACNA1G* and *GCLC*), "regulation of heart contraction" (five genes: *PDE5A, SEMA3A, HDAC4, CTNNA3* and *CACNA1G*), "regulation of muscle system process" (five genes: *PRKG1, PDE5A, MTPN, HDAC4* and *CTNNA3*), "positive regulation of developmental process" (13 genes: *TSHR, PDE5A, RHOB, SIRT1, PROX1, FLRT2, SEMA3A, RIMS1, PDE3A, IMPACT, MAP2K1, TMEM30A* and *HIF1AN*), "intracellular protein transport" (five genes: *RHOB, SEC24C, RABGAP1L, MYO6,* and *RIMS1*), "regulation of cell development" (11 genes*: PDE5A, PROX1, FLRT2, SEMA3A, RIMS1, PDE3A, HDAC4, IMPACT, MYCL, MAP2K1,* and *TMEM30A*), "cellular response to hydrogen peroxide" (three genes: *RHOB, SIRT1* and *IMPACT*), "bile acid metabolic process" (three genes: *AKR1D1, SIRT1* and *PROX1*), "negative regulation of transcription DNA-templated" (11 genes: *SIRT1, TFDP2, PHF14, FRK, PROX1, DACH1, HDAC4, IMPACT, AICDA, HIF1AN,* and *GCLC*), "negative regulation of RNA biosynthetic process" (11 genes: *SIRT1, TFDP2, PHF14, FRK, PROX1, DACH1, HDAC4, IMPACT, AICDA, HIF1AN* and *GCLC*), "regulation of muscle adaptation" (three genes: *PDE5A, MTPN,* and *HDAC4*), "adherens junction" (eight genes: *RHOB, TNS3, ARHGAP22, TLN2, FLRT2, RSU1, CTNNA3,* and *RPL4*), "regulation of membrane potential" (six genes: *RIMS1, NETO1, KCNMA1, CTNNA3, CACNA1G,* and *GCLC*), "positive regulation of reproductive process" (three genes:

PDE5A, SEMA3A, and *PDE3A*), "regulation of cell differentiation" (15 genes: *SIRT1, GLG1, PROX1, SEMA3A, RIMS1, MYCL, MAP2K1, TMEM30A, HIF1AN, PDE5A, FLRT2, HDAC4, PDE3A, IMPACT,* and *COL5A2*), "regulation of cardiac muscle contraction" (three genes: *PDE5A, HDAC4,* and *CTNNA3*), "single-organism catabolic process" (nine genes: *RP11-240B13.2, PDE5A, GK5, AKR1D1, ABHD5, LPIN1, PDE3A, HDAC4,* and *USP13*), "nucleotide metabolic process" (eight genes: *TSHR, PDE5A, PRPS2, RP11-411B6.6, GRM8, MDH1B, PDE3A,*and *HDAC4*), "Cgmp-mediated signaling" (two genes: *PRKG1,* and *PDE3A*), "peptidyl-lysine deacetylation" (two genes: *SIRT1,* and *HDAC4*) "ribose phosphate metabolic process" (seven genes: *TSHR, PDE5A, PRPS2, RP11-411B6.6, GRM8, PDE3A,* and *HDAC4*), "microtubule associated complex" (four genes: *DYNC1I1, SNX4, KIF26B,* and *BORCS5*) and "small molecule metabolic process", (16 genes: *GK5, PRPS2, SIRT1, LARS2, PROX1, GRM8, LPIN1, ASNSD1, MDH1B, GCLC, PDE5A, TSHR, AKR1D1, RP11-411B6.6, HDAC4,* and *PDE3A*) (**supplementary Table S6B**) and two KEGG pathways chicken "toll-like receptor signaling pathway" (three genes: *TLR7, MAP2K1* and *MAP3K7*) and "vascular smooth muscle contraction" (three genes: *KCNMA1, PRKG1* and *MAP2K1*) (**supplementary Table S7B**)

In Sri Lankan domestic chicken, I identified one GO term "protein-hormone receptor activity" (two genes: *TSHR* and *MCHR1*) and two KEGG pathways, "adrenergic signaling in cardiomyocytes" and "calcium signaling" (three genes: *ATP2B4, TACR3* and *EGFR*) (three genes: *ATP2B4, MAPK13* and *PPP2R2A*) (**supplementary Table S6C** and S**7C**).

In red junglefowl, I identified seven GO terms and they include, "regulation of biological process" (24 genes: *ADCY1, AL158801.1, EBF2, KCNA5, PPP2R2A, CCND2, NAB1, NEFM, FURIN, BMP10, GNRH1, ARHGAP25, MYCL, NPY2R, RHOBTB2, PDIA3, ZNF302, BCAR3, NKX2-6, ADAM9, ZFHX4, KCNMA1, RCAN3* and *PAFAH1B1*) and "regulation of protein metabolic process" (five genes: *AL158801.1, CCND2, ADAM9, FURIN* and *BMP10*), "transport" (14 genes: *SCN2A, ATP6V1B2, KCNA5, GRIK2, MFSD2A, SLC25A37, NEFM, KCNA1, FURIN, KCTD9, SLC16A5, KCNMA1, SLC26A8* and *PAFAH1B1*), "monovalent inorganic cation transport" (five genes: *ATP6V1B2, KCNA5, KCNA1, KCTD9* and *KCNMA1***),** "gated channel activity" (five genes: *KCNA5, GRIK2, KCNA1, KCTD9* and *KCNMA1*), "axon cargo transport" (two genes: *NEFM* and

PAFAH1B1) and "response to calcium ion" (two genes: *ADAM9* and *KCNMA1*) (**supplementary Table S6D**) and three KEGG pathways, which include "purine metabolism" (four genes: *ADCY1, ENTPD4, GUCY1A3* and *NT5C1A*), "progesterone-mediated oocyte maturation" (three genes: *RPS6KA1, ADCY1* and *CDC23*) and "vascular smooth muscle contraction" (three genes: *KCNMA1, GUCY1A3* and *ADCY1*) (**supplementary Table S7D**).

Discussion

The autosomal genetic background and adaptation to local production environments of three populations of indigenous domestic chicken were analysed alongside the wild progenitor, the red junglefowl, using whole-genome re-sequencing data. Our objectives are to identify candidate regions in their genomes that may be linked to the domestication process, as well as through the inclusion of indigenous populations living in different agro-ecologies, regions in the genome underlying adaptation to environmental challenges. Also, the analysis of both the domestic chicken and the wild ancestor may reveal common genomic regions selected between them.

It should be noted that our Ethiopian population consisted of two subpopulations, Horro and Jarso. Our mitochondrial DNA, admixture and to a less extent PCA results indicate that these two populations have witnessed different selection history. However, both are living in the Ethiopian highlands and to some extent do witness similar environmental conditions (Desta, 2015). I therefore decided to combine these two subpopulations in a single population for our signature of selection analysis.

Signature of selection in relation to domestication

Depending on the selection pressures during the domestication process, selected "domesticated genome regions" are expected to be present in most if not all the domestic chicken population examined. Compared to fancy chicken breeds and commercial chicken lines characterized by smaller effective population size and heavily selected by human, the indigenous village chicken with large effective population size, mating largely uncontrolled and not as heavily selected by human, may represent a better model for the identification of "domesticated" red junglefowl genome regions.

I identified five candidate genome regions positively selected in all the domestic chicken population but not in the red junglefowl (**Table 2**). Excluding one region on chromosome 2, these regions have all been previously identified in commercial broiler and layer chicken lines (Rubin *et al.*, 2010) supporting their status as "selected domesticated" regions. For the excluded region on chromosome 2, Johnsson *et al.* (2016) in their study of genomic regions that involves the chicken feralization have reported a selected candidate domesticated region (chromosome 2, position 146660000 - 146700000 bp) which falls at a 20-kb distance from our region (chromosome 2, position 146720000 - 146740000 bp), suggesting that this region might have indeed be also selected during the domestication process.

Amongst the five domesticated regions identified, candidate genes were present only in one of them (chromosome 5: 40060000 – 40110000 bp). It includes the *TSHR*, a previously reported locus with known functions in metabolic regulation and reproduction process (Yoshimura *et al.*, 2003; Hanon *et al.*, 2008; Rubin *et al.*, 2010), and *GTF2A1* known to be involved in the production of eggs in birds (Yuan *et al.*, 2015).

Looking at the SNP polymorphisms at the selected 50 kb region of the *TSHR* and *GTF2A1* loci, more than 80% are homozygous within in all the domestic chicken. It was also the case for four out of the six of red junglefowl samples. These red junglefowl samples may either represent evidence for feralization or the effect of admixture with local chicken breeds.

Signature of selection in relation to environmental adaptations

Response to selection is environmentally driven either naturally or artificially (Oleksyk *et al.*, 2010). The ancestral species of domestic chicken, the red junglefowl has a very large geographic distribution (see chapter 1) and while different wild subspecies and populations may be witnessing different environmental challenges (e.g. altitudes), all are living in region characterized by rather a warm climate but also substantial rainfall which however may varied considerably during the year (e.g. monsoon cycles). Accordingly, signature of selection related to heat tolerance, humidity or thermotolerance in general would be expected in domestic as well as the red junglefowl.

In the case of Ethiopian chicken, I identified two overrepresented genes (*HRH1* and *AGTR1*) associated with "vasoconstriction regulation" which might play important roles in thermoregulation, muscle relaxation and reproduction processes (Collier and Collier, 2011; Su *et al.*, 2011). The reduction in the evaporative heat loss or stress through decrease in the cutaneous blood flow has been reported previously in cattle and birds (Collier and Collier, 2011; Klotz *et al.*, 2016). In relation to thermoregulation and muscle relaxation, vasoconstriction has been linked to reduction in blood flow leading to increase in internal body temperature (Sessler *et al.*, 1990). Compared to the average chicken body temperature of 41°C (Bolzani *et al.*, 1979), the ambient temperature of both the Ethiopian Horro and Jarso districts are relatively low and the two selected candidate genes may have played important roles in adaptation to their local environments.

At the opposite, the climatic condition in Saudi Arabia is very dry with extreme heat during the day which could be more than 50°C in July/August. Here, I identified, several candidate gene ontologies such as "blood circulation", "regulation of heart contraction", "regulation of muscle system process", "regulation of muscle adaptation", "regulation of cardiac muscle contraction" which has been linked to the control of blood flow and evaporative heat loss (Klotz *et al.*, 2016). Other studies have associated some of the identified gene ontologies to oxygen deprivation response due to high elevation (Li *et al.*, 2013; Wang *et al.*, 2015), but this causative explanation is however unlikely here because the Saudi Arabian chicken samples were collected at an altitude of about 100 meters above sea-level. Therefore, the selected gene ontologies may rather favour adaptation to heat response. The significantly enriched cellular response to hydrogen peroxide and the toll-like receptor signalling pathways observed in Saudi Arabian domestic chickens may also suggest strong selection for innate immune response of host defense following pathogen infection (Medzhitov, 2001; Stone and Yang, 2006).

In the genome of Saudi Arabian and Sri Lankan chicken alongside the red junglefowl, I uncovered evidence for common signature of selection that may be linked to hypoxia response. Several candidates related genes involved in hypoxia have been reported in pigs and domestic chickens living in high altitudes in Tibet (Li *et al.*, 2013; Wang *et al.*, 2015). Positive selection for gene ontologies and pathways enriched in calcium ion, development of heart and blood vessels, heart size and rate, ductus arteriosus, cardiomyocyte, smooth muscle contraction, muscle development and neurogenesis have significant importance

in adaptation to hypoxic environments (Neuhaus *et al.*, 1999; Seals and Courtneidge, 2003; Olson *et al.*, 2006; Li *et al.*, 2013; Wang *et al.*, 2015). Of significance here is *KCNMA1*, a gene associated with the regulation of smooth muscle contraction through the activation of calcium ion (Williams *et al.*, 2004). Increase in calcium ion stimulates the activities of hypoxia-inducible factor-1 (Hui *et al.*, 2006). However, the biological roles that this gene plays in the three populations may be different. While in the two domestic chicken populations, it may be related rather to heat tolerance because the sampling locations were of low elevation, in red junglefowl however, it may rather play a role in adaptation to high altitude. Both the Yunnan (altitude ~ 3,000 m) and Hainan (altitude ~ 1,840 m) provinces, the two sampling locations of the red junglefowls, are mountainous. High elevation is associated with decrease in arterial oxygen content (Simonson *et al.*, 2010). Another gene, *ADAM9,* detected in red junglefowl, which is known to play a role in the development of the cardiorespiratory system has also been proposed to be involved in the adaptation to high-altitude in Tibetan chicken (Zhang *et al.*, 2016).

Both *KCNMA1* and *ADAM9* genes were not detected in our signature of selection in Ethiopian chicken, even though, the two Ethiopian subpopulations are living in an altitude of around 2,000 meters. Perhaps, neither the climatic and/or altitude characteristics where Horro and Jarso subpopulations lives may have led to strong selection pressure in their genome regions. Analysis of Ethiopian chicken populations, living in a much higher altitude may confirm the role that *KCNMA1* and *ADAM9* play in adaptation to altitude.

Common genome regions selected in both domestic and red junglefowl

I identified two shared regions between domestic and red junglefowl populations. One of the regions located on chromosome 7 has no gene with the closest *TMEFF2* within a 186-kb distance. It is thought to play a role in the regulation of cell adhesion (Chen *et al.*, 2014). For the second region on chromosome 23, seven candidate genes were present. In particular, one of them, *NT5C1A* has been reported as being within a selected candidate region in layer chicken lines (Rubin *et al.*, 2010). It is known to be involved in the regulation of heart adenosine level during hypoxia and ischemia especially when blood supply becomes inadequate in some part of the body (Hunsucker *et al.*, 2001). This may or may not be related to environmental adaptation. However, it is well documented that extreme exercise related activities may induce hypoxia (Springer *et al.*, 1991; Lindholm

and Rundqvist, 2016). Wild and domestic cocks are often violent in nature with the latter having a long history of being selected for fighting games (Delacour, 1977). Within the same region, I also identified a *MFSD2A* gene known to play role in the regulation of blood-brain barrier activities for effective functioning of central nervous system, just like the *HPCAL4* (Kobayashi *et al.*, 1998; Ben-Zvi *et al.*, 2014). Selection at these three genes may be therefore related to behaviour rather than to environmental challenges.

Conclusions

In this study, I assess the significance of signature of positive selection in three indigenous domestic chicken populations and red junglefowl in relation to their survival in production environment. I identified several candidate regions associated with domestication and environmental adaptation. The different numbers of birds included in each population, ranging from 5 to 11, might have influenced the results obtained here with for example the possibility of false negatives (e.g. region selected in the Saudi Arabian chicken population but not reaching the significant threshold). Increasing the number of birds to a similar size across each population may address to some extent these challenges, but even so, it may not remove the consequence of population demographic factors. Also, in the case of the wild red junglefowl our studied population is far from ideal as it does not include birds from a single population. Unfortunately, there is currently no such single red junglefowl population available for analysis.

Our findings were based on the analysis of the autosomal genomes only and it would be interesting to further investigate the selection pressures on the Z chromosome. Also, while performing this analysis, a new reference genome *Galgal* 5.0 was released, it will be equally important to reassess our results regarding the new assembly, in particular for regions with new functional annotations between *Galgal* 4.0 and *Galgal* 5.0.

References

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., *et al.* (2016). The ensembl gene annotation system. *The Journal of Biological Databases and Curation, 2016*. doi:10.1093/database/baw093

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research, 19*(9), 1655-1664.

Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., *et al.* (2013). From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics, 43*, 11.10.11-33. doi:10.1002/0471250953.bi1110s43

Beebe, W. (1918). *Monograph of the Pheasants* (1 ed.): New York Zoological Society.

Ben-Zvi, A., Lacoste, B., Kur, E., Andreone, B. J., Mayshar, Y., Yan, H., *et al.* (2014). Mfsd2a is critical for the formation and function of the blood-brain barrier. *Nature, 509*(7501), 507-511.

Bolzani, R., Ruggeri, F., and Olivo, O. (1979). Average normal temperature of the chicken in the morning and after 1-2 days of fasting. *Bollettino Della Societa Italiana Di Biologia Sperimentale, 55*(16), 1618-1622.

Bruford, M., Hanotte, O., Brookfield, J., and Burke, T. (1998). *Single-locus and multilocus DNA Fingerprinting* (2 ed.): Hoelzel AR, IRL Press, Oxford, UK.

Chen, H., and Boutros, P. C. (2011). VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics, 12*(1), 1. doi: 10.1186/1471-2105-12-35.

Chen, X., Corbin, J. M., Tipton, G. J., Yang, L. V., Asch, A. S., and Ruiz-Echevarría, M. J. (2014). The TMEFF2 tumor suppressor modulates integrin expression, RhoA activation and migration of prostate cancer cells. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 1843*(6), 1216-1224.

Collier, R. J., and Collier, J. (2011). *Environmental physiology of livestock*: John Wiley and Sons.

Crawford, R. (1984). Domestic fowl. In*: Evolution of domesticated animals,* : L.L. Mason,. London.

Dana, N., Dessie, T., van der Waaij, L. H., and van Arendonk, J. A. (2010). Morphological features of indigenous chicken populations of Ethiopia. *Animal Genetic Resources, 46*, 11-23.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., *et al.* (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.

Danforth, C. (1958). *Gallus sonnerati* and the domestic fowl. *Journal of Heredity, 49*(4), 167-170.

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods, 9*(8), 772-772.

Darwin, C. (1868). *The variation of animals and plants under domestication* (2 ed. Vol. 2): John Murray, United Kingdom.

Delacour, J. (1977). *The pheasant of the world* (2 ed.): Saiga Publishing Co. Ltd. Surr GU26 GTD. England.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., *et al.* (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics, 43*(5), 491-498.

Desta, T. (2015). *Phenomic and genomic landscape of Ethiopian village chickens.* (Doctoral degree thesis), University of Nottingham, Nottingham.

Desta, T., Dessie, T., Bettridge, J., Lynch, S., Melese, K., Collins, M., *et al.* (2013). Signature of artificial selection and ecological landscape on morphological structures of Ethiopian village chickens. *Animal Genetic Resources, 52*, 17-29.

Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt, L., *et al.* (2008). Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS genetics, 4*(2). doi.org/10.1371/journal.pgen.1000010

Fu, W., Lee, W. R., and Abasht, B. (2016). Detection of genomic signatures of recent selection in commercial broiler chickens. *BMC Genetics, 17*(1), 122.

Fumihito, A., Miyake, T., Sumi, S., Takada, M., Ohno, S., and Kondo, N. (1994). One subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic ancestor of all domestic breeds. *Proceedings of The National Academy of Sciences, 91*(26), 12505-12509.

Fumihito, A., Miyake, T., Takada, M., Shingu, R., Endo, T., Gojobori, T., *et al.* (1996). Monophyletic origin and unique dispersal patterns of domestic fowls. *Proceedings of The National Academy of Sciences, 93*(13), 6792-6795.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology, 52*(5), 696-704.

Hanon, E. A., Lincoln, G. A., Fustin, J.-M., Dardente, H., Masson-Pévet, M., Morgan, P. J., *et al.* (2008). Ancestral *TSH* mechanism signals summer in a photoperiodic mammal. *Current Biology, 18*(15), 1147-1152.

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., *et al.* (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature, 432*(7018), 695-716.

Hui, A. S., Bauer, A. L., Striet, J. B., Schnell, P. O., and Czyzyk-Krzeska, M. F. (2006). Calcium signaling stimulates translation of HIF-α during hypoxia. *The FASEB Journal, 20*(3), 466-475.

Hunsucker, S. A., Spychala, J., and Mitchell, B. S. (2001). Human Cytosolic 5′-Nucleotidase I characterization and role in nucleoside analog resistance. *Journal of Biological Chemistry, 276*(13), 10498-10504.

Johnsson, M., Gering, E., Willis, P., Lopez, S., Van Dorp, L., Hellenthal, G., *et al.* (2016). Feralisation targets different genomic loci to domestication in the chicken. *Nature Communications, 7*, 12950. doi:10.1038/ncomms12950.

1880 Keeling, L., Andersson, L., Schütz, K. E., Kerje, S., Fredriksson, R., Carlborg, Ö., *et al.*
1881 (2004). Chicken genomics: feather-pecking and victim pigmentation. *Nature,*
1882 *431*(7009), 645-646.

1883 Klotz, J. L., Aiken, G. E., Bussard, J. R., Foote, A. P., Harmon, D. L., Goff, B. M., *et al.*
1884 (2016). Vasoactivity and vasoconstriction changes in cattle related to time off
1885 toxic endophyte-infected tall fescue. *Toxins, 8*(10), 271.

1886 Kobayashi, M., Sakai, E., Furuta, Y., and Takamatsu, K. (1998). Isolation of two human
1887 cDNAs, *HLP3* and *HLP4*, homologous to the neuron-specific calcium-binding
1888 protein genes. *DNA Sequence, 9*(3), 171-176.

1889 Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary
1890 Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and*
1891 *Evolution, 33*(7), 1870-1874.

1892 Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–
1893 Wheeler transform. *Bioinformatics, 26*(5), 589-595.

1894 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009). The
1895 sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078-
1896 2079.

1897 Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y., *et al.* (2013). Genomic analyses
1898 identify distinct patterns of selection in domesticated pigs and Tibetan wild boars.
1899 *Nature Genetics, 45*(12), 1431-1438.

1900 Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of
1901 DNA polymorphism data. *Bioinformatics, 25*(11), 1451-1452.

1902 Lindholm, M. E., and Rundqvist, H. (2016). Skeletal muscle hypoxia-inducible factor-1
1903 and exercise. *Experimental Physiology, 101*(1), 28-32.

1904 Liu, Y.-P., Wu, G.-S., Yao, Y.-G., Miao, Y.-W., Luikart, G., Baig, M., *et al.* (2006).
1905 Multiple maternal origins of chickens: out of the Asian jungles. *Molecular*
1906 *Phylogenetics and Evolution, 38*(1), 12-19.

1907 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., *et al.*
1908 (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing
1909 next-generation DNA sequencing data. *Genome Research, 20*(9), 1297-1303.

1910 Medzhitov, R. (2001). Toll-like receptors and innate immunity. *Nature Reviews*
1911 *Immunology, 1*(2), 135-145.

1912 Morejohn, G. V. (1968). Breakdown of isolation mechanisms in two species of captive
1913 junglefowl (*Gallus gallus* and *Gallus sonneratii*). *Evolution, 22*(3), 576-582.

1914 Mwacharo, J., Bjørnstad, G., Mobegi, V., Nomura, K., Hanada, H., Amano, T., *et al.*
1915 (2011). Mitochondrial DNA reveals multiple introductions of domestic chicken
1916 in East Africa. *Molecular Phylogenetics and Evolution, 58*(2), 374-382.

Neuhaus, H., Rosen, V., and Thies, R. S. (1999). Heart specific expression of mouse *BMP-10* a novel member of the *TGF-β* superfamily. *Mechanisms of Development, 80*(2), 181-184.

Oleksyk, T. K., Smith, M. W., and O'Brien, S. J. (2010). Genome-wide scans for footprints of natural selection. *Philosophical transactions of the royal society B: Biological Sciences, 365*(1537), 185-205.

Olson, T. M., Alekseev, A. E., Liu, X. K., Park, S., Zingman, L. V., Bienengraeber, M., *et al.* (2006). Kv1. 5 channelopathy due to *KCNA5* loss-of-function mutation causes human atrial fibrillation. *Human Molecular Genetics, 15*(14), 2185-2191.

Pavlidis, P., Živković, D., Stamatakis, A., and Alachiotis, N. (2013). SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution, 30*(9), 2224-2234.

Peters, J., Lebrasseur, O., Best, J., Miller, H., Fothergill, T., Dobney, K., *et al.* (2015). Questioning new answers regarding Holocene chicken domestication in China. *Proceedings of The National Academy of Sciences, 12*(112), E2415.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics, 38*(8), 904-909.

Rubin, C.-J., Zody, M. C., Eriksson, J., Meadows, J. R., Sherwood, E., Webster, M. T., *et al.* (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature, 464*(7288), 587-591.

Schütz, K. E., Forkman, B., and Jensen, P. (2001). Domestication effects on foraging strategy, social behaviour and different fear responses: A comparison between the red junglefowl (*Gallus gallus*) and a modern layer strain. *Applied Aanimal Behaviour Science, 74*(1), 1-14.

Seals, D. F., and Courtneidge, S. A. (2003). The *ADAMs* family of metalloproteases: multidomain proteins with multiple functions. *Genes and Development, 17*(1), 7-30.

Sessler, D. I., Moayeri, A., Støen, R., Glosten, B., Hynson, J., and McGuire, J. (1990). Thermoregulatory vasoconstriction decreases cutaneous heat loss. *Anesthesiology, 73*(4), 656-660.

Silva, P., Guan, X., Ho-Shing, O., Jones, J., Xu, J., Hui, D., *et al.* (2009). Mitochondrial DNA-based analysis of genetic variation and relatedness among Sri Lankan indigenous chickens and the Ceylon junglefowl (*Gallus lafayetti*). *Animal Genetics, 40*(1), 1-9.

Simonson, T. S., Yang, Y., Huff, C. D., Yun, H., Qin, G., Witherspoon, D. J., *et al.* (2010). Genetic evidence for high-altitude adaptation in Tibet. *Science, 329*(5987), 72-75.

Springer, C., Barstow, T. J., Wasserman, K., and Cooper, D. M. (1991). Oxygen uptake and heart rate responses during hypoxic exercise in children and adults. *Medicine and Science in Sports and Exercise, 23*(1), 71-79.

Stone, J. R., and Yang, S. (2006). Hydrogen peroxide: a signaling messenger. *Antioxidants and Redox Signaling, 8*(3-4), 243-270.

Su, M.-T., Lin, S.-H., and Chen, Y.-C. (2011). Genetic association studies of angiogenesis-and vasoconstriction-related genes in women with recurrent pregnancy loss: a systematic review and meta-analysis. *Human Reproduction Update, 17*(6), 803-812.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research, 25*(24), 4876-4882.

Tixier-Boichard, M., Bed'hom, B., and Rognon, X. (2011). Chicken domestication: From archeology to genomics. *Comptes Rendus Biologies, 334*(3), 197-204.

Wang, M.-S., Li, Y., Peng, M.-S., Zhong, L., Wang, Z.-J., Li, Q.-Y., *et al.* (2015). Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. *Molecular Biology and Evolution, 32*(7), 1880-1889.

West, B., and Zhou, B.-X. (1988). Did chickens go north? New evidence for domestication. *Journal of Archaeological Science, 15*(5), 515-533.

Williams, S. E., Wootton, P., Mason, H. S., Bould, J., Iles, D. E., Riccardi, D., *et al.* (2004). Hemoxygenase-2 is an oxygen sensor for a calcium-sensitive potassium channel. *Science, 306*(5704), 2093-2097.

Xiang, H., Gao, J., Yu, B., Hofreiter, M., and Zhao, X. (2015). Reply to Peters et al.: Further discussions confirm early Holocene chicken domestication in northern China. *Proceedings of The National Academy of Sciences, 112*(19), E2416-E2416.

Xiang, H., Gao, J., Yu, B., Zhou, H., Cai, D., Zhang, Y., *et al.* (2014). Early Holocene chicken domestication in northern China. *Proceedings of The National Academy of Sciences, 111*(49), 17564-17569.

Yoshimura, T., Yasuo, S., Watanabe, M., Iigo, M., Yamamura, T., Hirunagi, K., *et al.* (2003). Light-induced hormone conversion of T4 to T3 regulates photoperiodic response of gonads in birds. *Nature, 426*(6963), 178-181.

Yuan, J., Sun, C., Dou, T., Yi, G., Qu, L., Qu, L., *et al.* (2015). Identification of promising mutants associated with egg production traits revealed by genome-wide association study. *PloS One, 10*(10), doi.org/10.1371/journal.pone.0140615.

Zerbino, D., Wilder, S., Johnson, N., Juettemann, T., and Flicek, P. (2015). The Ensembl regulatory build. *Genome Biology, 16*(56). doi.org/10.1186/s13059-015-0621-5.

Zeuner, F. E. (1963). *A history of domesticated animals*: Hutchinson & Co. (Publishers) Ltd. London.

Zhang, Q., Gou, W., Wang, X., Zhang, Y., Ma, J., Zhang, H., *et al.* (2016). Genome resequencing identifies unique adaptations of tibetan chickens to hypoxia and

High-Dose ultraviolet radiation in high-altitude environments. *Genome Biology and Evolution, 8*(3), 765-776.

Chapter 3

Whole-genome sequencing reveals candidate signature of positive selection for environmental pressures in Ceylon (*Gallus lafayettii*) and green (*G. varius*) junglefowls

Abstract

Here, I assess through whole-genome autosomal sequence analysis, the presence of candidate signature of positive selection in Ceylon (*Gallus lafayettii*) and green (*G. varius*) junglefowls. These two *Gallus* species have a geographic distribution more restricted than the red junglefowl with the Ceylon junglefowl present in Sri Lanka only and the green junglefowl in the Indonesian Islands. Selective sweep analyses using Tajima's D and Hp methods identifies genomic regions associated with adaptation to disease challenges, environmental stress, thermoregulation, and hypoxia in both species. Selection for skeletal formation and a single locus associated with ovarian follicle development were detected in green junglefowl only. These regions may provide genome landmarks for the subsequent identification of signature of introgression in domestic chicken.

Introduction

Junglefowls (*Gallus* species) are important wild species due to their genetic contributions to the domestication of the chicken either as the main ancestral species (red junglefowl) or following possible subsequent introgression (Delacour, 1977; Fumihito *et al.*, 1994; Liu *et al.*, 2006; Eriksson *et al.*, 2008) (see chapter 4 of this thesis). Domestic chicken constitute one of the most diverse livestock species supporting human survival through the production of meat and eggs (Bell and Weaver, 2002). Among the four known species of junglefowls, the Ceylon junglefowl *Gallus lafayettii* and green junglefowl *Gallus varius* have the most restricted geographic distributions, living in Sri Lanka, where the red junglefowl is absent, and on the Indonesian Islands respectively where its geographic distribution partially overlaps with *G. gallus bankiva*.

Ceylon junglefowl are known to live from sea level to an altitude of up to 2,300 m above sea-level (Delacour, 1977; Del Hoyo *et al.*, 2017b). Though, morphologically, they resemble the red junglefowl, wattles and combs are less developed in both sexes, and they have red legs and yellow eyes. Although their mating is unknown they are not as social as the polygamous red junglefowl (Del Hoyo *et al.*, 2017b). In Sri Lanka, the Ceylon junglefowl cocks are being crossed with domestic chicken for cockfighting purposes, a practice not officially allowed by the authority. Ceylon junglefowl are known to travel a large distance but they are rarely observed in flight (Delacour, 1977; Al-Nasser *et al.*,

2007; Del Hoyo *et al.*, 2017b). The natural feeding behaviour includes scavenging on small fruits, live insects etc found around the dry coastal scrub and forest (Delacour, 1977; Del Hoyo *et al.*, 2017b). They breed throughout the year but more particularly between February and May (Del Hoyo *et al.*, 2017b) and sometimes between April and June (Delacour, 1977). They lay an average of 2 – 6 eggs under both the natural and captive conditions (Delacour, 1977; Al-Nasser *et al.*, 2007; Del Hoyo *et al.*, 2017b).

The green junglefowl, endemic to Java and other part of Indonesia, are generally birds of the coast and low-lying valleys (Delacour, 1977), but they may inhabit areas with altitude of up to 3,000 m above sea-level (Del Hoyo *et al.*, 2017a). They are social birds, able to sustain flight for long distance over open water. Their food includes copepods, small crabs and marine insects (Al-Nasser *et al.*, 2007). Their reproductive ability is higher compared to other *Gallus* species, including the red junglefowl, being able to lay more eggs throughout the breeding season (Delacour, 1977). The captive green junglefowl are frequently crossed with domestic hens with the hybrids known as "Bekissar", prized for their loud, long-drawn, and monosyllabic voices (Delacour, 1977). Unlike the Ceylon junglefowl, the captive green junglefowl are less hardy, and easily susceptible to infectious diseases and usually require artificial intervention to increase body temperature during winter (Delacour, 1977).

Populations from both species are suspected to be stable and far from the vulnerability threshold. They are included in the category as "Not globally threatened (Least Concern)" according to *BirdLife International (2017) Species factsheet* website (http://www.birdlife.org/), Del Hoyo *et al.* (2017a) and Del Hoyo *et al.* (2017b).

Recent advances in sequencing technologies have helped in the detection of candidate genome regions playing crucial roles in the evolution of adaptive traits in poultry. It includes candidate regions selected following domestication (Rubin *et al.*, 2010), adaptation to high altitude (Li *et al.*, 2013) and more recently, feralization (Johnsson *et al.*, 2016) and behavioural traits such as cockfighting (Guo *et al.*, 2016). Using the Illumina sequencing platform, I report here, the autosomal genome scanning of the Ceylon and green junglefowls, for positive selection signatures, with the aim to identify environmentally selected candidate regions in these two wild *Gallus* species as well as to provide baseline information for our subsequent analysis of possible introgression between these species and the domestic chicken.

Materials and methods

Sampling and sequencing

The genome of eight Ceylon junglefowls were analysed. Whole blood samples of five wild Ceylon junglefowls kept at conservation centres (Gannoruwa (n = 1) and Wayamba (n = 4) in Uva province of Sri Lanka were collected by Pradeepa Silva (University of Peradeniya, Sri Lanka). Two Ceylon junglefowls blood samples, part of the Cosmopolitan Chicken Research Project (CC®P) (http://www.ccrp.be/), were collected by Olivier Hanotte (University of Nottingham). One unpublished genome sequence (~ 14 X coverage) was also obtained from the Cosmopolitan Chicken Research Project (CC®P).

Twelve green junglefowl genomes were analysed. It includes nine publicly available green junglefowl genome sequences (Ulfah *et al.,* 2016) downloaded from the DNA Data Bank Japan Sequence Read Archive (Accession No. DRA003951) with a ~ 6 X to ~ 8 X genome coverage (**supplementary Table S1**). Blood samples from two green junglefowls were obtained from the CC®P project. As for the Ceylon junglefowl, one unpublished genome sequence (~ 14 X coverage) was obtained from CC®P project.

The DNA preparation and sequencing is as described in chapter 2 Materials and Methods section. Sequencing of new samples was done at ~ 30 X genome coverage.

Mappings and variants calling protocol

Since reference genomes for the two wild species are not available in the public domain, I aligned each sample against its closest relative, the red junglefowl *Galgal* 4.0 reference genome (Hillier *et al.*, 2004; Aken *et al.*, 2016). From the mapping of reads stage to variants calling steps, the protocol is the same as described in chapter 2 Materials and Methods section.

The mapping metrics including the percentage of both mates of read pair that properly mapped to same chromosome, percentage of mapped reads, sequence coverage and percentage of the genome with bases covered by at least 5, 10 and 20 reads, were calculated using SAMtools version 0.1.19 (Li *et al.,* 2009). Using the Ensembl "VEP" version 85 (Aken *et al.,* 2016), I predicted the consequences of the variants while the total number of SNPs in each sample/species was identified using VCFtools version 0.1.11

(Danecek *et al.*, 2011). The R based "VennDiagram" package (Chen and Boutros, 2011) was used to plot the unique and shared SNPs between species.

Genome-wide nucleotide diversity

I estimated the average genome nucleotide diversity (π) for each population on 20 kb windows over a 10 kb sliding step using VCFtools version 0.1.11 (Danecek *et al.*, 2011).

Selective sweeps detection

Sweep detection was carried out using two methods:

I ran Tajima's D, a neutrality test that detects selection from nucleotide sequences (Tajima, 1989) using VCFtools version 0.1.11 (Danecek *et al.*, 2011) in a non-overlapping bin size of 20 kb (since overlaps are not supported for this tool version). The approach involves comparing the average number of nucleotide differences from pairwise DNA sequences and the number of segregating sites.

The analysis relies on the following equation:

$$\text{Tajima's D} = \frac{\frac{\sum_{i<j} d_{ij}}{n(n-1)/2} - \frac{S}{\sum_{i=1}^{n-1}\frac{1}{i}}}{\sqrt{\hat{V}\left(\frac{\sum_{i<j} d_{ij}}{n(n-1)/2} - \frac{S}{\sum_{i=1}^{n-1}\frac{1}{i}}\right)}} \quad equ.(1)$$

Where d_{ij} is the number of nucleotide differences between pairwise sequences, i and j and *n* is given by the number of sequences to compare, *S* is the number of segregating sites in each sample (Tajima, 1989). Under the neutrality hypothesis where selection and recombination do not occur, both the nucleotide differences and segregating sites tends to zero. To detect selective sweep regions, only windows with Tajima's D index values below zero where further analysed as that correspond to positive selection. Using the formula in *equ. (3)*, the value of Tajima's D for each window was subsequently Z-transformed.

For the pool heterozygosity (Hp) method (Rubin *et al.*, 2010), I measured the level of heterozygosity in the genome at a window of 20 kb. Since the script required specification for step size, I ran the analysis in a 20-kb window over a 20-kb step size therefore allowing comparison with the Tajima's D 20 kb non-step size calculation.

The calculation of the pool heterozygosity (Hp) method relies on the following equation:

$$\mathrm{Hp} = \frac{2\sum n_{MAJ}\sum n_{MIN}}{\left(\sum n_{MAJ}+\sum n_{MIN}\right)^2} \quad equ.(2)$$

Where $\sum n_{MAJ}$ and $\sum n_{MIN}$ are the sums of major and minor allele frequencies respectively for all the SNPs within each the 20-kb window. The values for the Hp calculated for each window size were then subsequently Z-transformed using the equation:

$$\mathrm{ZHp} = \frac{\mathrm{Hp} - \overline{X}(\mathrm{Hp})}{\sigma\,(\mathrm{Hp})} \quad equ.(3)$$

Where $\overline{X}$ is the mean and σ is the standard deviation of the Hp. Windows with large number of heterozygote SNPs will show values above zero, they may reflect balancing selection signature. Here, only windows that falls below zero (windows with many homozygote SNPs) and candidate genome widows under positive selection were further analysed.

For both the Tajima's D and Hp methods, the arbitrary negative 0.2% values of the windows at the extreme of the distribution were considered as the empirical significant threshold, therefore resulting in same number of candidate windows used in further analysis for both methods. For each window, I also set a minimum threshold of 30 SNPs. The putative size of the candidate selective sweep regions was calculated by merging together all continuous but significant windows. This approach accounts for windows containing sweep SNPs which were inadvertently split into adjacent or separate windows due to sequential allocation of 20 kb window size. For candidate regions that include more than a single window, I sum all SNPs within such region and the average values for the Tajima's D and Hp were calculated (**supplementary Table S2 – S5**). Genes within each region were then considered candidate genes under positive selection.

I identified all candidate regions which overlapped between methods and between species using bedtools "intersect" version 2.17.0 (Quinlan and Hall, 2010).

Functional annotation

As explained in chapter 2.

Results

Sequencing and read mapping

The analyses included the genome sequences of eight Ceylon junglefowls and 12 green junglefowls. With the exception of nine publicly downloaded green junglefowl sequences (Ulfah *et al.*, 2016) with genome coverage ranging from ~ 6 X to ~ 8 X (see **materials and methods**) and two sequences (Ceylon junglefowlk3 and Green junglefowlk3, see **supplementary Table S1**) obtained from CC®P project with genome coverage of ~ 14 X, all other samples were newly sequenced at ~ 30 X coverage in comparison to the 1.07 G size of the *Galgal* 4.0 reference genome (Hillier *et al.*, 2004). After the quality control step, the total number of sequence reads per sample for Ceylon junglefowl ranged from 158,521,405 to 346,513,098 and for the green junglefowl from 58,763,807 to 348,251,083 (**supplementary Table S1**).

Across all samples, I identified between 85.22% and 95.84% of both mates of a read pair properly mapped to the same chromosome resulting in more than 97% uniquely mapped reads (**supplementary Table S1**). The number of reads supporting each nucleotide across the autosomes depends on sample depth of sequence coverage. For instance, samples with sequence coverage of ~ 30 X has more than 95% of their bases covered by at least 5 to 10 reads when compared to those having an overall ~ 6 X sequence coverage in which ~ 50% of their bases were only covered by at least 5 reads (**supplementary Table S1**).

Polymorphism detection

Variants for single nucleotide polymorphisms (SNPs) in each sample were jointly called following the multi-sample aggregation approach in GATK version 3.4.0 as detailed in the Materials and Methods section of chapter 2. Analyses were restricted only to the bi-

allelic SNPs and the autosomes. Including the homozygous and the heterozygous, the number of autosomal SNPs identified in each sample varies and depends on how deep the sequence coverage was (**supplementary Table S1**). Samples with ~ 30 X sequence coverage have higher number of SNPs than those having ~ 6 X coverage. For Ceylon junglefowl, it ranges from 10,883,174 to 11,801,876 SNPs (~ 11 SNPs/kb), while for the green junglefowl, it ranges from 9,042,763 to 12,826,733 SNPs (~ 12 SNPs/kb) (**supplementary Table S1**). At the species level, a total of 17,316,240 SNPs and 15,205,220 SNPs were detected in Ceylon and green junglefowl species respectively (**Table 1**). At the chromosome level, the highest number of SNPs for each species were detected on chromosome 1, while the fragmented sequence assembly of chromosome 16 (based on information from reference *Galgal* 4.0) shows the lowest number of SNPs (**Figure 1**). Between the two species, only 5,878,712 SNPs were shared while 11,437,528 and 9,326,508 SNPs were unique to Ceylon and green junglefowl populations respectively (**Figure 2**).

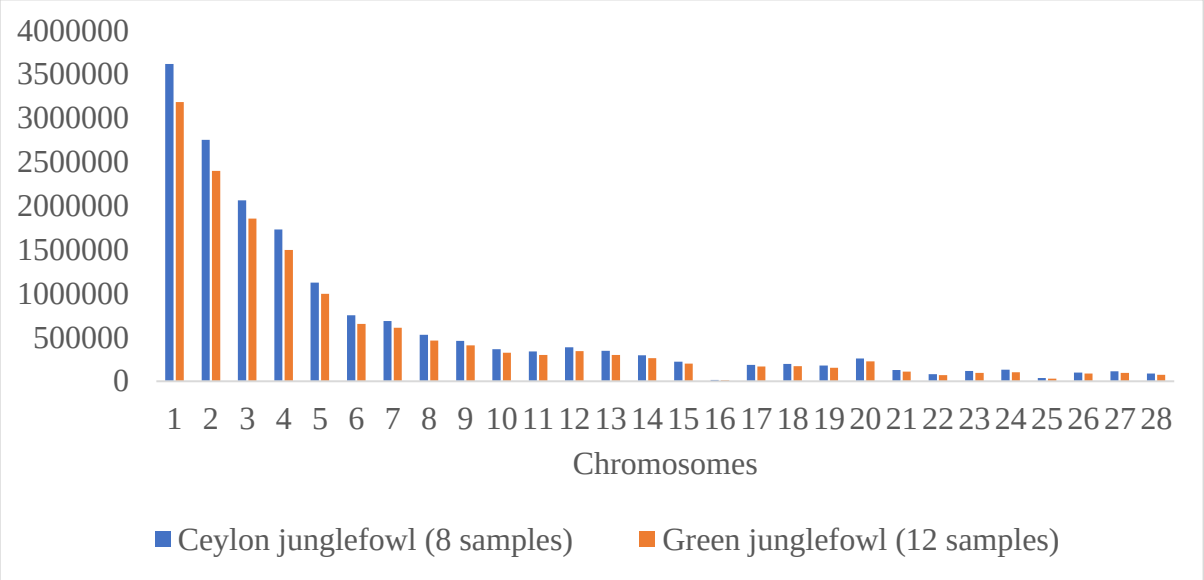


Figure 1. Histogram plot comparing the number of SNPs per chromosome between Ceylon and green junglefowls.

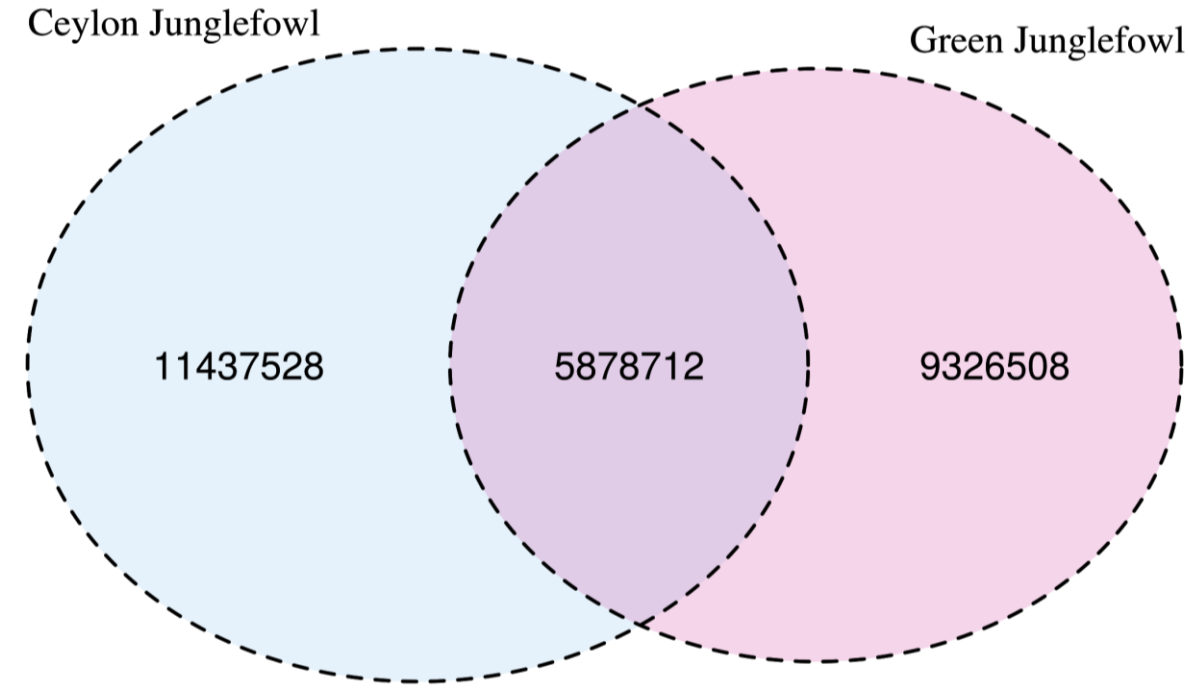


Figure 2. Venn diagram showing the number of shared and unique SNPs between Ceylon and green junglefowls.

Consequences of the variants detected

I compared the results of the SNPs detected to the one available on the Ensembl version 85 database (Aken *et al.*, 2016), only 30.6% of the SNPs in Ceylon junglefowl have been previously reported while 69.4% are novel. In the green junglefowl, 31.5% of the total number of SNPs have been previously reported and 68.5% are novel (**Table 1**). The Ensembl variants effect predictor tools shows that 52.9% of the SNPs in Ceylon junglefowl are found in intron, followed by 32.1% in intergenic variants, 5.7% in upstream gene variants, 4.2% are downstream gene variants, while polymorphisms at annotated coding sequence are observed at very low frequency (0.00006%) (**Table 1**).

In green junglefowl, 44.9% of the total SNPs are found within intergenic region, followed by the 42.4% in intron, 5.8% are upstream gene variants, 4.2% are downstream gene variants while most of other annotations occurred at very low frequencies with the coding sequence variants having the lowest proportion (0.00008%) (**Table 1**).

Within population diversity

I estimated the level of diversity within each population by calculating the nucleotide diversity (π) in 20 kb windows at 10 kb slides. The average genome nucleotide diversity for Ceylon junglefowl is 0.0034 ± 0.0013 and 0.0016 ± 0.0007 in green junglefowl.

Table 1. Species SNPs statistic and annotation

Variant consequences	**Ceylon junglefowl**	**Green junglefowl**
Sample size (genome)	8	12
Genome sequence coverage (X)	~ 14 to ~ 30	~ 5 to ~ 30
Total number of bi-allelic SNPs	17,316,240	15,205,220
Novel variants	12,019,614	10,418,860
Existing variants	5,296,626	4,786,360
Overlapped genes	39,930	14,882
Overlapped transcript	47,665	15,649
Splice donor variants	791	379
Splice acceptor variants	899	362
Stop gained	1571	706
Stop lost	419	59
Start lost	513	221
Missense variants	99,949	61,470
Splice region variants	33,848	24,604
Stop retained variants	126	71
Synonymous variants	160,411	126,554
Coding sequence variants	11	12
Mature miRNA variants	242	226
5 prime UTR variants	68,764	28,147
3 prime UTR variants	323,714	155,210
Non-coding transcript exon variants	177,987	2,092
Intron variants	9,167,140	6,452,855
Upstream gene variants	981,457	887,601
Downstream gene variants	735,887	641,994
Intergenic variants	5,562,507	6,822,652

Sweep detection in Ceylon junglefowl using Tajima's D

From the total number of 45,548 windows analysed, only 175 were excluded following the > 30 minimum SNPs criteria, leaving us with 45,373 windows for further study. The 90 windows of the extreme 0.2% (i.e $Z(D) \leq -3.897$) of the Z-transformed Tajima's D (Z(D)) window value distribution defines 32 Tajima's D candidate selective sweep regions (**supplementary Table S2**). Across the genome, selective sweeps were identified on chromosomes 1, 2, 3, 4, 6, 7, 8, 9, 11, 14 and 24 with most of candidate regions

detected on chromosomes 2 (10 sweeps) and 7 (7 sweeps) (**supplementary Table S2** and **Figure 3a**). The most significant peak in the genome is a single 20 kb region located on chromosome 24 (6,280,000 – 6,300,000 bp), which harbours the *VWA5A* gene with Z(D) score of - 4.984. Two large candidate regions were found on chromosome 7, a 420-kb region (position 6,800,000 – 7,220,000 bp, Z(D) score of - 4.491) and a 360-kb region (position 7,240,000 – 7,600,000 bp, Z(D) score of - 4.631). They are separated by a single 20 kb window (position 7220000 and 7240000 bp). The Z(D) score for this 20-kb window is -3.851, just below the 0.2% genome-wide threshold (Z(D) ≤ -3.897). The proportion of homozygous SNPs for this 20-kb region for all the Ceylon junglefowl samples is more than 85%. Together, these results then suggest a single large candidate signature of positive selection region of at least 800 kb in size. Within the 800-kb region, I identified 18 candidate genes out of a total of 39 genes detected across the 32 initial regions. These are: *COL6A2, FTCD, MCM3AP, YBEY, POFUT2, LSS, S100B, TUBA1C, PCNT, C21orf58, KMO, ITGB2, FAM207A, ADARB1, GLS, STAT1, STAT4* and *MYO1B.*

As for the other regions, 21 out of the 30 candidate regions are defined by one window above the threshold (20 kb length), five regions are defined by two windows (40 kb length each), two regions by three windows (60 kb length each, one region by four windows (80 kb length) and one other region is 200 kb in size (10 windows) (**supplementary Table S2 and Figure 3a**). The total size of the selective sweeps (31 candidate regions) is 1.80 megabases (Mb) and the median is 20 kb.

Sweep detection in Ceylon junglefowl using Hp

For the Hp analyses, I estimated an average of 0.206 ± 0.0394 genome-wide level of heterozygosity. I then searched the genome along a 20-kb window for regions showing excess homozygosity. From 45,552 windows analysed, only 71 were excluded following our 30 SNPs minimum criteria leaving 45,481 windows for further analysis. The 90 windows within the 0.2% threshold criteria have Z(Hp) value of ≤ -3.947 (**Figure 3b**). These 90 widows define 50 candidate selective sweep regions. Only chromosomes 1, 2, 4, 5, 6, 7, 8, 9, 11, 17, 20, 26, 27 and 28 show significant peaks (**Figure 3b**). Chromosome 2 has the highest number of sweep regions (15 sweeps), followed by chromosome 5 (8 sweeps) and chromosome 1 (6 sweeps) (**supplementary Table S3 and Figure 3b**). Across the genome, the strongest peak is a single 20 kb region located on chromosome 11 (19,340,000 – 19,360,000 bp) with a Z(Hp) score of - 4.561 and three genes *ADAT1,*

GABARAPL2 and *RP11-77K12.7*. The region with the largest fragment size is 200 kb in length, it is located on chromosome 7 at position 36,000,000 and 36,200,000 bp with a Z(Hp) score of - 4.338. It contains four annotated genes *TANC1, WDSUB1, BAZ2B* and *Mar-07*. Out of the 50 candidate regions identified across the genome, 34 are defined by a single window (20 kb size each), eight regions by two windows (40 kb size each), two regions by three windows (60 kb size each), two regions by 5 windows (100 kb size each) while the largest regions are 140 kb and 200 kb in sizes (**supplementary Table S3**). The total size of the selective sweep for the 50 candidate regions is 1.86 Mb with a median of 20 kb.

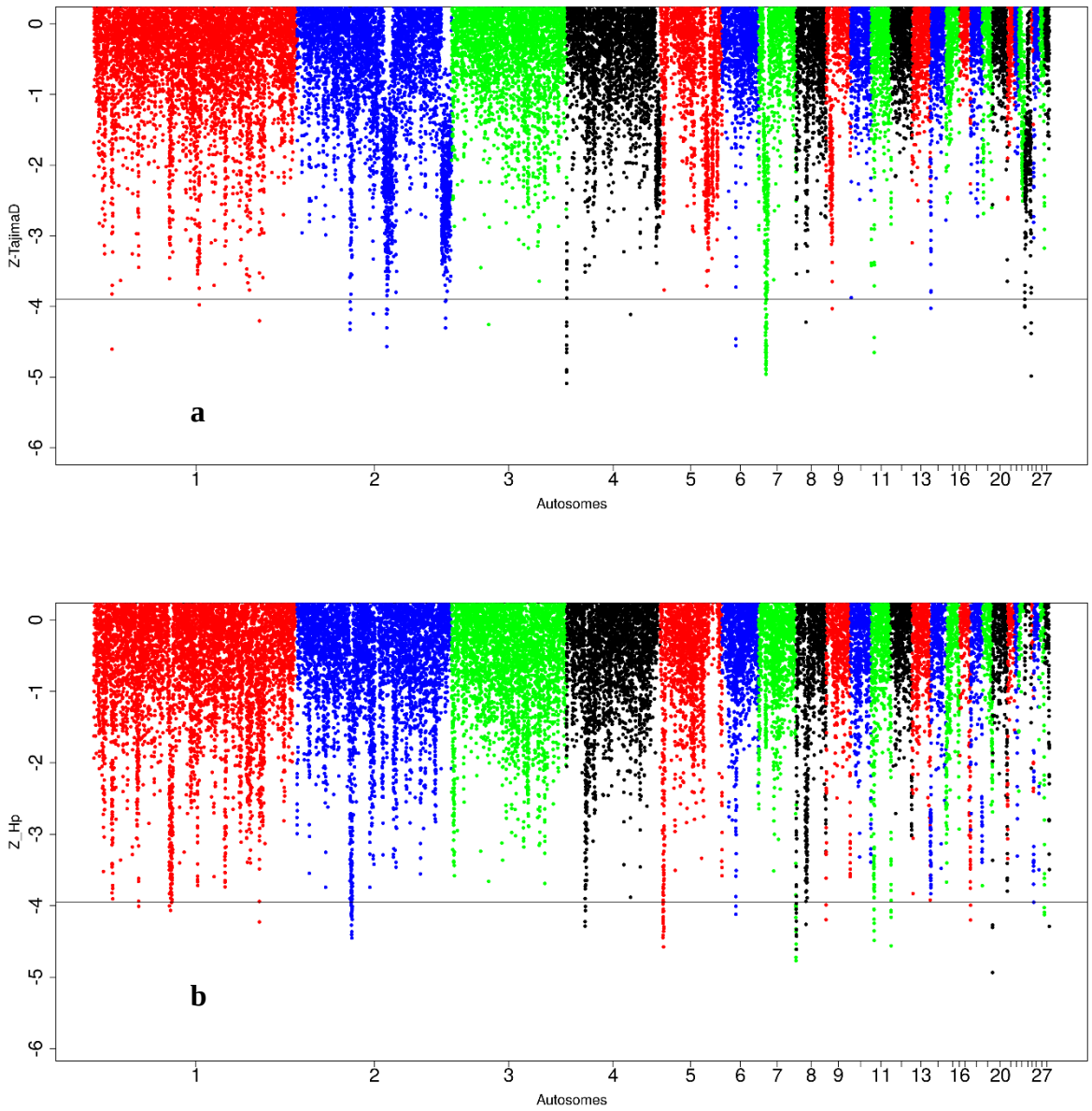


Figure 3. Signature of selection test using (**a**) Tajima's D and (**b**) Hp for Ceylon junglefowl. The horizontal lines corresponding to Z(D) ≤ -3.897 (**a**) or Z(Hp) ≤ -3.947 (**b**), indicate the arbitrary significant threshold (0.2% of the extreme negative end of the total numbers of windows).

Sweep detection in green junglefowl using Tajima's D

From the total number of 45,536 windows, only 235 were excluded following the minimum 30 SNPs criteria, leaving 45,321 windows for further analyses. After the Z-transformation of Tajima's D, the genome-wide threshold value was Z(D) ≤ -3.414 (90 windows, 0.2% window threshold). They define 66 candidate sweep regions. Thirty-three candidate regions were identified on chromosome 1, five on chromosomes 3, 5 and 8, four candidate regions on chromosome 4, two candidate regions on chromosomes 2, 11,

15 and 20, and one candidate region on chromosomes 6, 7, 9, 12, 14 and 17 (**Figure 4a**). Across the genome, a single 20 kb window located on chromosome 1 (72,940,000 – 72,960,000 bp), with the *FAR2* locus, has the lowest Z(D) score of -4.339. The candidate region with the largest fragment size is 80 kb in length. It is located on chromosome 1 (73,140,000 – 73,220,000 bp) and it contains two genes, *RAD51AP1* and *C12orf4* (**supplementary Table S4**). In terms of the distribution of the fragment sizes, 52 out of the 66 candidate regions have 20 kb size each, eight regions have 40 kb size each, five regions have 60 kb size each, while only one region is 80 kb in size (**supplementary Table S4**). The total size of the selective sweep for the 66 candidate regions is 1.74 Mb and the median is 20 kb.

Sweep detection in green junglefowl using Hp

Genome-wide, the average window heterozygosity was 0.128 ± 0.0368. Only 70 windows from a total of 45,550 were removed following the 30 SNPs minimum criteria, leaving 45,480 windows for subsequent analysis. The 0.2% genome-wide cut-off threshold corresponds to 90 windows with Z(Hp) values ≤ -2.735. They define 42 candidates sweep regions. Thirty-five out of the 42 candidate regions were found on chromosome 1. On the other chromosomes, three candidate regions are present on chromosome 4 and one on chromosomes 5, 8, 9 and 20 (**supplementary Table S5** and **Figure 4b**). The most significant region (Z(Hp) = -3.082) was defined by a single 20 kb window on chromosome 1 (73,340,000 – 73,360,000 bp), it contains a single gene (*CCND2*). Chromosome 1 has also the largest region with a fragment size of 220 kb (9000000 – 9220000 bp). This region does not include any annotated gene. For the distribution of fragment sizes, 28 candidate regions have 20 kb size each, four regions have 40 kb size each, three regions have 60 kb each, three regions 80 kb each and the remaining four regions are 100 kb, 140 kb, 200 kb and 220 kb in sizes. The total size of the selective sweep for the 42 candidate regions is 1.80 Mb and the median is 20 kb.

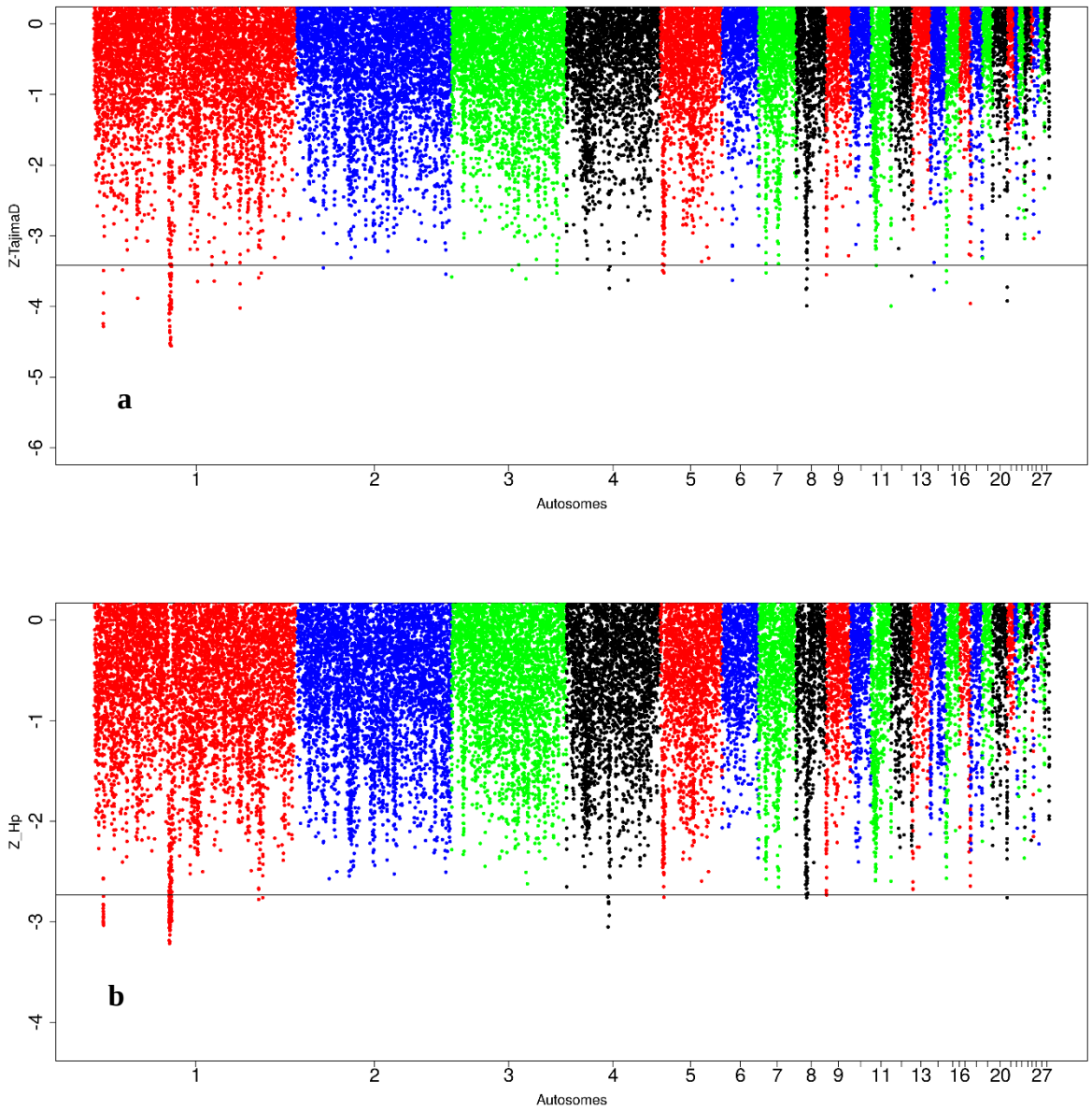


Figure 4. Signature of selection test using (**a**) Tajima's D and (**b**) Hp for green junglefowl. The horizontal lines, corresponding to Z(D) ≤ -3.414 (**a**) or Z(Hp) ≤ -2.735 (**b**), indicate the arbitrary significant threshold (0.2% of the extreme negative end of the total numbers of windows).

Overlaps between sweep detection methods

In Ceylon junglefowl, three candidate sweep regions located on chromosomes 6, 8 and 11 were detected both by the Tajima's D and Hp methods (**supplementary Table S6**). Two genes are present within these regions, *KCNMA1* and *SLC12A4* on chromosomes 6 and 11 respectively.

In green junglefowl, out of a total of 66 candidate sweep regions detected by Tajima's D and 42 sweep regions detected by Hp, 28 candidates sweep regions were commonly

identified with both approaches. Twenty-four are located on chromosome 1, two are on chromosome 4 and one on chromosomes 5, 8 and 20. Genes within these overlapping regions are *MRPS35, MANSC4, PTHLH, FAR2, RAD51AP1, C12orf4, FGF23, TIGAR, CCND2, RP11-234B24.6, KCNA1, KCNA5, NTF3, ANO2, VWF, NOBOX* and *PRMT8* (**supplementary Table S7**).

Common selected regions between species

I investigated the genome of the two species for common regions under selection. I compared all the sweep regions identified between species but within methods and the overlaps were then considered. One candidate region on chromosome 7 including six genes *FAM207A, ADARB1, GLS, STAT1, STAT4* and *MYO1B* was commonly found using Tajima's D, while three candidate regions on chromosome 1 including the *PTHLH, KCNA1,* and *PRMT8* genes were commonly identified using Hp method (**supplementary Table S8**). Among these regions, one region on chromosome 1 at position 73,660,000 – 73,680,000 bp, also came significant in the analysis of signature of positive selection in red junglefowl (Chapter 2). The *KCNA1* is found within this region. This region was not detected as a significant region or below the threshold in any of the domestic chicken populations analysed (see chapter 2**).**

Functional annotation for genes within sweep regions

To identify the functional clusters of genes located within the sweep regions, I performed the enrichment analysis for gene ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway using DAVID 6.8 for GO (Huang *et al.*, 2009) and KOBAS 3.0 for KEGG (Xie *et al.*, 2011) for a total of 83 and 39 genes identified in Ceylon junglefowl and green junglefowl, respectively. Only clusters of genes showing the default *P* value of < 0.05 were included in the analysis.

Enrichment test for the Ceylon junglefowl shows that only 11 GO terms were significant. These are: "deaminase activity" (three genes), "glutamate metabolic process" (three genes), "alpha-amino acid catabolic process" (three genes), "modified amino acid binding" (three genes), "transmembrane transporter activity" (nine genes), "regulation of peptidyl-tyrosine phosphorylation" (two genes), "membrane-bounded vesicle" (16 genes), "anion transmembrane transporter activity" (four genes), "response to osmotic

stress" (two genes), "response to abiotic stimulus" (seven genes) and "secondary active transmembrane transporter activity" (four genes). Only three KEGG pathways were observed and they include, "dorso-ventral axis formation" (two genes), "regulation of autophagy" (two genes) and "foxO signaling pathway" (three genes) (**supplementary Table S9**).

For the green junglefowl, only eight GO terms were identified, "regulation of phosphorylation" (six genes), "regulation of neuron death" (three genes), "cellular response to endogenous stimulus" (five genes), "transmembrane receptor protein tyrosine kinase signaling pathway" (four genes), "regulation of phosphorus metabolic process" (seven genes), "osteoblast differentiation" (four genes), "developmental process" (12 genes), "protein homooligomerization" (three genes) and one KEGG pathway, "MAPK signaling pathway" (three genes) was significantly enriched (**supplementary Table S10**).

Discussion

This study represents the first to investigate the signature of selection and to provide insight on the genomic regions that may have played roles in the adaptation and survival of Ceylon and green junglefowls in their natural habitats.

The regions detected and interpretation of the identified selective sweeps rely on the red junglefowl *Galgal* 4.0 reference genome since the reference genome annotation for the two species are not available. Further investigation using the species reference genome sequences will be of interest. In particular, it could lead to the detection of new regions under positive selection, regions not covered following, e.g., chromosomal rearrangements. However, the use of a single genome of reference allows us to assess if the candidate region of signature of selection identified are unique or present as well in red junglefowl and/or domestic chicken population (see chapter 2).

The use of the red junglefowl as a reference population likely explains the larger number of SNPs identified in the two populations and particularly, the larger number of novel SNPs compared to the domestic and red junglefowl populations examined in chapter 2.

The nucleotide diversity and average genome heterozygosity in Ceylon junglefowl and green junglefowls were lower compared to the one observed in domestic chicken and red junglefowl (see chapter 2). This is somewhat surprising considering that the samples analysed here are from wild population. Perhaps, this may be a consequence of the complex genetic history of the domestic chicken and wild red junglefowl examined here, which may have witnessed several episodes of admixture.

The nucleotide diversity and average genome heterozygosity was also approximately twice in Ceylon junglefowl compared to the green junglefowl. Ceylon junglefowls are hard to tame and thus exhibit wilder behaviour in captivity than the green junglefowl, which are often kept as pets (Delacour, 1977). Our results may suggests a remarkable loss in the genetic diversity of green junglefowl, an observation also consistent with previous breeding practices (Delacour, 1977).

Due to the power of different selection tests (Hancock and Di Rienzo, 2008), the results presented here, obtained from both the Tajima's D and Hp should be considered as complementary. See materials and methods for the description of the two tests. A nearly equal number of windows were analysed in both populations, I also detected a similar number of candidates' signature of positive selection in green junglefowl compared to Ceylon junglefowl (80 *versus* 78, both test combined and considering the overlapping regions). Only four of these regions were commonly found in both species, so, most of the candidate regions for positive selection detected here are species specific. This is not surprising as both species are morphologically different, and they likely witnessed different environmental pressures and/or they may have responded differently to the same environmental challenges. Analyses of the genes within the candidate regions under positive selection may provide us here with some further insight.

Several genes identified within the sweep regions might play a role in the adaptation of the two junglefowls species to disease challenges. In Ceylon junglefowl, the detection of "deaminase activity" gene ontology (GO) is strongly linked to T-cell responses (Marone *et al.*, 1980; Martinez-Navio *et al.*, 2012). Also, I identified a GO term, "glutamate metabolic" process known to increase resistance to oxidative stress (Shanware *et al.*, 2011) and to control macrophages efficiency and proliferation of lymphocytes and cytokines (Dutta *et al.*, 2013). I also identified a GO described for alpha-amino acid catabolic process which is known to regulate the innate and adaptive immune cells

(McGaha *et al.*, 2012). Amongst the two overrepresented genes associated with the regulation of "peptidyl-tyrosine phosphorylation" ontology, *EGFR* has been previously detected in commercial broiler (Rubin *et al.*, 2010) and in Sri Lanka indigenous chicken population (see chapter 2). One of its several functions involves adaptation to stress in general and ultra violet-A radiation (He *et al.*, 2008) while the second gene, *ITGB2* is associated with the immune response (Kragstrup *et al.*, 2016). For the KEGG pathway analysis, I identified the "dorso-ventral axis formation" pathway known to play a role in immune response for cytokines inflammation inducement in insects and mammals (Khush and Lemaitre, 2000; Imler and Zheng, 2004). More so, among the eight candidate genes detected by both the Tajima's D and Hp, two of them, *STAT1* and *STAT4*, are known to play significant roles during the immune response (Gil *et al.*, 2012).

In contrast, in green junglefowl, fewer GO, potentially linked to diseases resistance were identified. An exception is the "regulation of phosphorylation" GO term known to regulate proteins expression, e.g. cytokines, in response to immune challenges *(*Johnson, 2009*)*. Generally, the smaller number of functional terms associated with disease resistance in green junglefowl compared to Ceylon junglefowl may explain why the former is more susceptible to contagious and infectious diseases when kept in captivity compared to the Ceylon junglefowl, particularly if in close contact with domestic chicken (Delacour, 1977). Also, in the wild, the green junglefowl is a more solitary bird compared to the Ceylon junglefowl and therefore infectious diseases transmission across birds within population may be less an issue but this require further investigation.

I also identified other signatures of positive selection which may be linked to environmental stresses. In Ceylon junglefowl, one of these is the "response to osmotic stress". Changes in the osmotic concentration of the body and lack of response for the immediate compensation of the solute gradient is known to cause abnormal physiological processes (Brocker *et al.*, 2012). Aquatic animals and yeast, *Saccharomyces cerevisiae* represent an example of the known species with significant adaptation to high/low level of solutes concentration (Fiol and Kültz, 2007; Saito and Posas, 2012). *MAPK13*, one of the enriched genes associated with osmotic stress, is known to be induced during osmotic challenge in mammals (Saito and Posas, 2012). Functional analysis also identifies "response to abiotic stimulus" which may signify a possible tolerance toward extreme high or low temperature, drought, ion excess, salinity and heavy metals (Bohnert *et al.*,

2006). Ceylon junglefowl is known to live in the dry coastal scrub (Delacour, 1977), an environment where optimal response to osmotic stress will be advantageous.

Starvation is one of the main stressors that occur within animal ecology leading to competition, predation and then population reduction. For instance, competition for food during winter has been reported as one of the cause of death in Mute swan (Nielsen *et al.*, 1981) and population decline is known to occur in Deer during winter when vegetables become scarce (Gilbert, 2003). In Whitetail Deer, adaptive strategies during this extreme condition includes reduction in metabolic process and movement restriction (Gilbert, 2003). Unravelling the genetic mechanisms that shaped the genome of these junglefowls especially during the period of starvation is key to our evolutionary understanding and may provide useful information about the management of these populations.

In Ceylon junglefowl, I identified key functional KEGG pathway "regulation of autophagy" known to be induced during nutrient starvation where cytoplasmic materials are degraded for protein to synthesize in order to supply energy required by the cell through ATP production (Levine and Yuan, 2005). During child birth especially when the interaction of nutrients between mother and neonates across the placenta becomes suddenly depleted (Kuma *et al.*, 2004), autophagy is known to play a key role. Alongside the nutritional factor, autophagy is also known to play role in the innate immune defense against foreign attack (Deretic and Levine, 2009). In some cases, unneeded proteins are excessively synthesized causing the death of many cells (Levine and Yuan, 2005), therefore, autophagy regulation is important in such perspective. I also observed a "foxO signaling pathway" known to play role in response to metabolic stress during starvation and adaptation to environmental changes (Eijkelenboom and Burgering, 2013). Also in green junglefowl, a single gene, *TIGAR* known to play role in autophagy (Xie *et al.*, 2014) was detected.

Overall, positive selection signals toward stresses were less common in green junglefowl. As mentioned before, the green junglefowl is known to be less hardy, delicate and in winter requires artificial heat when compare to other junglefowls (Delacour, 1977). Further analysis of the biological importance of some of the genes in green junglefowl shows that *MRPS35* plays a role in the detection of DNA damage (Ishiguchi *et al.*, 2004) with *RAD51AP1* known to be involved in DNA repair mechanisms in mammalian cells (Kovalenko *et al.*, 1997), while *TIGAR* which was previously described in relation to

autophagy, and also plays a role in the regulation of DNA damage (Yu *et al.*, 2015). I also found within the green junglefowl, candidate selection signature *VWF*, which plays a role in wound healing and platelet activation that reduces excessive bleeding during injury (Savage *et al.*, 1996).

Adaptation to high altitude where oxygen level is low is common in livestock (e.g. chicken, sheep and pig) (Li *et al.*, 2013; Wang *et al.*, 2015; Wei *et al.*, 2016) and in humans (Lorenzo *et al.*, 2014). The identification of mutation in *EGLN1* which prevents thickness of the blood cell was one of the observed genetic mechanisms that allowed successful adaptation of humans living under hypoxic conditions (Lorenzo *et al.*, 2014), while several candidate genes involving calcium-signaling pathways, blood vessels etc. have been implicated in the adaption to low oxygen responses in chicken (Wang *et al.*, 2015), sheep (Wei *et al.*, 2016) and pig (Li *et al.*, 2013). Hypoxia is also a condition in malignant tissues that prevents adequate blood and oxygen supply due to insufficient vascular systems (Giatromanolaki *et al.*, 2004).

Ceylon and green junglefowls are known to successfully inhabit areas of more than 2,300 m above the sea-level, even though, the latter have been reported to frequent along the sea-shore, costal area and low elevation (Delacour, 1977). I therefore investigated the genome landscape of these two species for evidence of selection for hypoxia.

In Ceylon junglefowl, I identified four candidate genes known to play role in response to oxygen deprivation. One of which is *KCNMA1* previously described in Tibetan domesticated pig (Li *et al.*, 2013) and known to play an active role during hypoxia (Williams *et al.*, 2004). Other genes identified which involved in the response to anaerobic metabolic conditions and angiogenesis regulated by hypoxia-inducible factor 1alpha (*HIF1A*) includes *FMN2* known to increase its expression level during hypoxia activation (Yamada *et al.*, 2013), *STAT1* known to decrease its expression during hypoxia (Ivanov *et al.*, 2007), while overexpression of *BECN1* is known to decrease hypoxia-induced angiogenesis and in association with autophagy activity (Giatromanolaki *et al.*, 2011; Lee *et al.*, 2011). During oxygen depletion, *HIFA* plays a role in the regulation of autophagy (Bellot *et al.*, 2009).

In green junglefowl, three candidate genes with known functional relationship with response to oxygen deprivation were identified. They include *FGF6*, a member of

fibroblast growth factor linked to angiogenesis involving regeneration process of muscular system (Folkman and Klagsbrun, 1987), *KCNA5* associated with smooth muscle cells that creates hypoxic pulmonary vasoconstriction response allowing adequate oxygen supply in the pulmonary artery (Platoshyn *et al.*, 2006), while the survival of cells during hypoxia is mediated by the activity of *TIGAR* (Cheung *et al.*, 2012).

Regarding the genes present in the candidate positive selected regions in red junglefowl or domestic chicken identified in chapter 2, *KCNMA1* was found in both the red and Ceylon junglefowls and *KCNA5* in red and green junglefowls. I found no evidence for selection of any of the *HIFA* genes in previous studies in chicken with exception of the identification of *KCNMA1* in Saudi Arabia and Sri Lanka indigenous domestic chicken populations analysed in chapter 2.

Male junglefowls are aggressive by nature and are frequently crossed with domestic chickens to improve cockfighting behavioural characteristics beside aesthetic values. Cockfighting is a gaming exercise that stimulates the heart rate and increases the flow of oxygenated blood to different part of working muscles (Springer *et al.*, 1991; Lindholm and Rundqvist, 2016). It is however possible that beside high elevation, the detection of overrepresented genes linked to hypoxia may also be associated with the behavioural characteristic of the genus *Gallus*.

Selection for skeletal structure and reproduction traits in green junglefowl

In green junglefowl, I observed several GO terms of importance in bone formation, such as the *"*regulation of phosphorus" associated with energy and lipid metabolism as well as skeletal structures and other cellular processes (Berndt and Kumar, 2009). As an essential mineral constituents known for tissue repair, more than 80% of phosphorus is found in bones and teeth (Cheng *et al.*, 2011), in which its excess production can cause significant damage to the liver and intestines (Guo, 2007) while deficiency of phosphorus results in abnormal muscle and bone function such as osteomalacia (Shaikh *et al.*, 2008; Berndt and Kumar, 2009). The detection of "osteoblast differentiation" is also known to plays a role in bone formation from mesenchymal stem cells (Komori, 2006). *FGF23* is associated with the regulation of metabolic activities of vitamin D (Shimada *et al.*, 2004). In addition, I identified *PTHLH* and *KCNA1* in a 20-kb region which is also present in Ceylon junglefowl. *PTHLH* is known to play an important role in skeletal system

development (Amling *et al.*, 1997), through the regulation of the body's calcium and phosphate level (Hall and Guyton, 2015). This gene product is also known to occur at high concentration in the amniotic fluid when compared to the maternal and foetal blood circulation (Farrugia *et al.*, 2000), thereby suggesting an important developmental role during pregnancy. Magnesium is an important mineral mediating energy transfer through the activation of adenosine triphosphate. Magnesium deficiency results in bone loss and skeletal fragility in rats, pigs and human (Rude and Gruber, 2004). *KCNA1* is known to plays a role in the reabsorption of magnesium ion in the kidney (Glaudemans *et al.*, 2009), while *KCNA5* is associated with response to hypoxia and also involved in insulin regulation through the control of blood glucose level (Philipson *et al.*, 1991).

Only one gene, *NOBOX*, present within a specific green junglefowl candidate signature of positive selection, is known to play a role in the female reproduction process. This oocyte-specific homeobox gene is expressed during the development of ovarian follicles in humans (Qin *et al.*, 2007). This green junglefowl specific signature of selection might be therefore linked to the higher reproductive ability (clutch size) when compared to other *Gallus* species (Delacour, 1977).

Conclusions

The identification of signature of positive selection in the genome of the Ceylon and green junglefowl not only provide us with important genome landmarks for the study between *Gallus* species introgression events (the focus of chapters 4 and 5) but also, they provide us also with an insight about the environmental selection pressures that have shaped the genome of these two species and by extension the environmental characteristics to which these species are adapted. Several genes were identified but not all were examined in detail. Further investigations including all candidate genes may further reveal the effect of the selection pressures on these two wild species and its link to environmental adaptation.

References

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., *et al.* (2016). The Ensembl gene annotation system. *The Journal of Biological Databases and Curation, 2016*. doi:10.1093/database/baw093

Al-Nasser, A., Al-Khalaifa, H., Al-Saffar, A., Khalil, F., Albahouh, M., Ragheb, G., *et al.* (2007). Overview of chicken taxonomy and domestication. *World's Poultry Science Journal, 63*(02), 285-300.

Amling, M., Neff, L., Tanaka, S., Inoue, D., Kuida, K., Weir, E., *et al.* (1997). Bcl-2 lies downstream of parathyroid hormone–related peptide in a signaling pathway that regulates chondrocyte maturation during skeletal development. *The Journal of Cell Biology, 136*(1), 205-213.

Bell, D. D., and Weaver, W. D. (2002). *Commercial chicken meat and egg production* (5 ed.): Springer Science and Business Media.

Bellot, G., Garcia-Medina, R., Gounon, P., Chiche, J., Roux, D., Pouysségur, J., *et al.* (2009). Hypoxia-induced autophagy is mediated through hypoxia-inducible factor induction of *BNIP3* and *BNIP3L* via their *BH3* domains. *Molecular and Cellular Biology, 29*(10), 2570-2581.

Berndt, T., and Kumar, R. (2009). Novel mechanisms in the regulation of phosphorus homeostasis. *Physiology, 24*(1), 17-25.

Bohnert, H. J., Gong, Q., Li, P., and Ma, S. (2006). Unraveling abiotic stress tolerance mechanisms–getting genomics going. *Current Opinion in Plant Biology, 9*(2), 180-188.

Brocker, C., Thompson, D. C., and Vasiliou, V. (2012). The role of hyperosmotic stress in inflammation and disease. *Biomolecular Concepts, 3*(4), 345-364.

Chen, H., and Boutros, P. C. (2011). VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics, 12*(1), 1.

Cheng, C.-Y., Kuro-o, M., and Razzaque, M. S. (2011). Molecular regulation of phosphate metabolism by fibroblast growth factor-23–Klotho system. *Advances in Chronic Kidney Disease, 18*(2), 91-97.

Cheung, E. C., Ludwig, R. L., and Vousden, K. H. (2012). Mitochondrial localization of *TIGAR* under hypoxia stimulates *HK2* and lowers *ROS* and cell death. *Proceedings of the National Academy of Sciences, 109*(50), 20491-20496.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., *et al.* (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.

Del Hoyo, J., Elliott, A., Sargatal, J., Christie, D., and de Juana, E. (2017a). Green junglefowl (*Gallus varius*). In P. McGowan, G. Kirwan, and P. Boesman (Eds.), *Handbook of the birds of the world alive*. Lynx Edicions, Barcelona.

Del Hoyo, J., Elliott, A., Sargatal, J., Christie, D., and de Juana, E. (2017b). Sri Lanka Junglefowl (*Gallus lafayettii*). In P. McGowan, G. Kirwan, and P. Boesman (Eds.), *Handbook of the birds of the world alive*. Lynx Edicions, Barcelona.

Delacour, J. (1977). *The pheasant of the world* (2 ed.): Saiga Publishing Co. Ltd. Surr GU26 GTD. England.

Deretic, V., and Levine, B. (2009). Autophagy, immunity, and microbial adaptations. *Cell Host and Microbe, 5*(6), 527-549.

Dutta, S., Ray, S., and Nagarajan, K. (2013). Glutamic acid as anticancer agent: An overview. *Saudi Pharmaceutical Journal, 21*(4), 337-343.

Eijkelenboom, A., and Burgering, B. M. (2013). *FOXOs*: signalling integrators for homeostasis maintenance. *Nature Reviews Molecular Cell Biology, 14*(2), 83-97.

Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt, L., *et al.* (2008). Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genetics, 4*(2). doi.org/10.1371/journal.pgen.1000010

Farrugia, W., Ho, P., Rice, G., Moseley, J., Permezel, M., and Wlodek, M. (2000). Parathyroid hormone-related protein (1-34) in gestational fluids and release from human gestational tissues. *Journal of Endocrinology, 165*(3), 657-662.

Fiol, D. F., and Kültz, D. (2007). Osmotic stress sensing and signaling in fishes. *Febs Journal, 274*(22), 5790-5798.

Folkman, J., and Klagsbrun, M. (1987). Angiogenic factors. *Science, 235*(4787), 442-448.

Fumihito, A., Miyake, T., Sumi, S., Takada, M., Ohno, S., and Kondo, N. (1994). One subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic ancestor of all domestic breeds. *Proceedings of the National Academy of Sciences, 91*(26), 12505-12509.

Giatromanolaki, A., Sivridis, E., and Koukourakis, M. I. (2004). Tumour angiogenesis: vascular growth and survival. *Apmis, 112*(7-8), 431-440.

Giatromanolaki, A. N., St Charitoudis, G., Bechrakis, N. E., Kozobolis, V. P., Koukourakis, M. I., Foerster, M. H., *et al.* (2011). Autophagy patterns and prognosis in uveal melanomas. *Modern Pathology, 24*(8), 1036-1045.

Gil, M. P., Ploquin, M. J., Watford, W. T., Lee, S.-H., Kim, K., Wang, X., *et al.* (2012). Regulating type 1 *IFN* effects in *CD8* T cells during viral infections: changing *STAT4* and *STAT1* expression for function. *Blood, 120*(18), 3718-3728.

Gilbert, R. (2003). *A to Z guide to White-tailed Deer and Deer hunting*. : Woods N' Water, Inc, Bellvale, NY.

Glaudemans, B., van der Wijst, J., Scola, R. H., Lorenzoni, P. J., Heister, A., van der Kemp, A. W., *et al.* (2009). A missense mutation in the Kv1. 1 voltage-gated potassium channel–encoding gene *KCNA1* is linked to human autosomal dominant hypomagnesemia. *The Journal of Clinical Investigation, 119*(4), 936-942.

Guo, L. (2007). Doing battle with the green monster of Taihu Lake. *Science, 317*(5842), 1166-1166.

2664 Guo, X., Fang, Q., Ma, C., Zhou, B., Wan, Y., and Jiang, R. (2016). Whole-genome
2665 resequencing of Xishuangbanna fighting chicken to identify signatures of
2666 selection. *Genetics Selection Evolution, 48*(1), 62.

2667 Hall, A. C., and Guyton, J. E. (2015). *Textbook of medical physiology* (11 ed.): W.B.
2668 Saunders, Philadelphia.

2669 Hancock, A. M., and Di Rienzo, A. (2008). Detecting the genetic signature of natural
2670 selection in human populations: models, methods, and data. *Annual Review of*
2671 *Anthropology, 37*, 197-217.

2672 He, Y.-Y., Council, S. E., Feng, L., and Chignell, C. F. (2008). UVA-induced cell cycle
2673 progression is mediated by a disintegrin and metalloprotease/epidermal growth
2674 factor receptor/AKT/Cyclin D1 pathways in keratinocytes. *Cancer Research,*
2675 *68*(10), 3752-3758.

2676 Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., *et al.*
2677 (2004). Sequence and comparative analysis of the chicken genome provide unique
2678 perspectives on vertebrate evolution. *Nature, 432*(7018), 695-716.

2679 Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative
2680 analysis of large gene lists using DAVID bioinformatics resources. *Nature*
2681 *Protocols, 4*(1), 44-57.

2682 Imler, J.-L., and Zheng, L. (2004). Biology of Toll receptors: lessons from insects and
2683 mammals. *Journal of Leukocyte Biology, 75*(1), 18-26.

2684 Ishiguchi, H., Izumi, H., Torigoe, T., Yoshida, Y., Kubota, H., Tsuji, S., *et al.* (2004).
2685 *ZNF143* activates gene expression in response to DNA damage and binds to
2686 cisplatin-modified DNA. *International Journal of Cancer, 111*(6), 900-909.

2687 Ivanov, S., Salnikow, K., Ivanova, A., Bai, L., and Lerman, M. (2007). Hypoxic
2688 repression of *STAT1* and its downstream genes by a p*VHL/HIF-1* target
2689 *DEC1/STRA13*. *Oncogene, 26*(6), 802-812.

2690 Johnson, L. N. (2009). The regulation of protein phosphorylation. *Biochemical Society*
2691 *Transactions, 37*(4), 627-641.

2692 Johnsson, M., Gering, E., Willis, P., Lopez, S., Van Dorp, L., Hellenthal, G., *et al.* (2016).
2693 Feralisation targets different genomic loci to domestication in the chicken. *Nature*
2694 *Communications, 7*, 12950. doi: 10.1038/ncomms12950.

2695 Khush, R. S., and Lemaitre, B. (2000). Genes that fight infection: what the Drosophila
2696 genome says about animal immunity. *Trends in Genetics, 16*(10), 442-449.

2697 Komori, T. (2006). Regulation of osteoblast differentiation by transcription factors.
2698 *Journal of Cellular Biochemistry, 99*(5), 1233-1239.

2699 Kovalenko, O. V., Golub, E. I., Bray-Ward, P., Ward, D. C., and Radding, C. M. (1997).
2700 A novel nucleic acid-binding protein that interacts with human *rad51*
2701 recombinase. *Nucleic Acids Research, 25*(24), 4946-4953.

Kragstrup, T. W., Jalilian, B., Keller, K. K., Zhang, X., Laustsen, J. K., Stengaard-Pedersen, K., *et al.* (2016). Changes in Soluble *CD18* in murine autoimmune arthritis and rheumatoid arthritis reflect disease establishment and treatment response. *PloS One, 11*(2). doi: 10.1371/journal.pone.0148486.

Kuma, A., Hatano, M., Matsui, M., Yamamoto, A., Nakaya, H., Yoshimori, T., *et al.* (2004). The role of autophagy during the early neonatal starvation period. *Nature, 432*(7020), 1032-1036.

Lee, S.-J., Kim, H.-P., Jin, Y., Choi, A. M., and Ryter, S. W. (2011). Beclin 1 deficiency is associated with increased hypoxia-induced angiogenesis. *Autophagy, 7*(8), 829-839.

Levine, B., and Yuan, J. (2005). Autophagy in cell death: an innocent convict? *The journal of Clinical Investigation, 115*(10), 2679-2688.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.

Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y., *et al.* (2013). Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nature Genetics, 45*(12), 1431-1438.

Lindholm, M. E., and Rundqvist, H. (2016). Skeletal muscle hypoxia-inducible factor-1 and exercise. *Experimental Physiology, 101*(1), 28-32.

Liu, Y.-P., Wu, G.-S., Yao, Y.-G., Miao, Y.-W., Luikart, G., Baig, M., *et al.* (2006). Multiple maternal origins of chickens: out of the Asian jungles. *Molecular Phylogenetics and Evolution, 38*(1), 12-19.

Lorenzo, F. R., Huff, C., Myllymäki, M., Olenchock, B., Swierczek, S., Tashi, T., *et al.* (2014). A genetic mechanism for Tibetan high-altitude adaptation. *Nature Genetics, 46*(9), 951-956.

Marone, G., Plaut, M., and Lichtenstein, L. M. (1980). The role of adenosine in the control of immune function. *International journal of clinical and laboratory research, 10*(2), 303-312.

Martinez-Navio, J. M., Climent, N., Gallart, T., Lluis, C., and Franco, R. (2012). An old enzyme for current needs: adenosine deaminase and a dendritic cell vaccine for HIV. *Immunology and Cell Biology, 90*(6), 594-600.

McGaha, T. L., Huang, L., Lemos, H., Metz, R., Mautino, M., Prendergast, G. C., *et al.* (2012). Amino acid catabolism: a pivotal regulator of innate and adaptive immunity. *Immunological Reviews, 249*(1), 135-157.

Nielsen, B. B., Clausen, B., and Elvestad, K. (1981). The incidence of salmonella bacteria in wild-living animals from Denmark and in imported animals. *Nordisk Veterinaermedicin, 33*(9-11), 427-433.

Philipson, L. H., Hice, R. E., Schaefer, K., LaMendola, J., Bell, G. I., Nelson, D. J., *et al.* (1991). Sequence and functional expression in Xenopus oocytes of a human insulinoma and islet potassium channel. *Proceedings of the National Academy of Sciences, 88*(1), 53-57.

Platoshyn, O., Brevnova, E. E., Burg, E. D., Yu, Y., Remillard, C. V., and Yuan, J. X.-J. (2006). Acute hypoxia selectively inhibits *KCNA5* channels in pulmonary artery smooth muscle cells. *American Journal of Physiology-Cell Physiology, 290*(3), C907-C916.

Qin, Y., Choi, Y., Zhao, H., Simpson, J. L., Chen, Z.-J., and Rajkovic, A. (2007). *NOBOX* homeobox mutation causes premature ovarian failure. *The American Journal of Human Genetics, 81*(3), 576-581.

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841-842.

Rubin, C.-J., Zody, M. C., Eriksson, J., Meadows, J. R., Sherwood, E., Webster, M. T., *et al.* (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature, 464*(7288), 587-591.

Rude, R. K., and Gruber, H. E. (2004). Magnesium deficiency and osteoporosis: animal and human observations. *The Journal of Nutritional Biochemistry, 15*(12), 710-716.

Saito, H., and Posas, F. (2012). Response to hyperosmotic stress. *Genetics, 192*(2), 289-318.

Savage, B., Saldívar, E., and Ruggeri, Z. M. (1996). Initiation of platelet adhesion by arrest onto fibrinogen or translocation on von Willebrand factor. *Cell, 84*(2), 289-297.

Shaikh, A., Berndt, T., and Kumar, R. (2008). Regulation of phosphate homeostasis by the phosphatonins and other novel mediators. *Pediatric Nephrology, 23*(8), 1203-1210.

Shanware, N. P., Mullen, A. R., DeBerardinis, R. J., and Abraham, R. T. (2011). Glutamine: pleiotropic roles in tumor growth and stress resistance. *Journal of Molecular Medicine, 89*(3), 229-236.

Shimada, T., Hasegawa, H., Yamazaki, Y., Muto, T., Hino, R., Takeuchi, Y., *et al.* (2004). *FGF-23* is a potent regulator of vitamin D metabolism and phosphate homeostasis. *Journal of Bone and Mineral Research, 19*(3), 429-435.

Springer, C., Barstow, T. J., Wasserman, K., and Cooper, D. M. (1991). Oxygen uptake and heart rate responses during hypoxic exercise in children and adults. *Medicine and Science in Sports and Exercise, 23*(1), 71-79.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123*(3), 585-595.

2778 Ulfah, M., Kawahara-Miki, R., Farajalllah, A., Muladno, M., Dorshorst, B., Martin, A.,
2779 *et al.* (2016). Genetic features of red and green junglefowls and relationship with
2780 Indonesian native chickens Sumatera and Kedu Hitam. *BMC Genomics, 17*(1), 1.

2781 Wang, M.-S., Li, Y., Peng, M.-S., Zhong, L., Wang, Z.-J., Li, Q.-Y., *et al.* (2015).
2782 Genomic analyses reveal potential independent adaptation to high altitude in
2783 Tibetan chickens. *Molecular Biology and Evolution, 32*(7), 1880-1889.

2784 Wei, C., Wang, H., Liu, G., Zhao, F., Kijas, J. W., Ma, Y., *et al.* (2016). Genome-wide
2785 analysis reveals adaptation to high altitudes in Tibetan sheep. *Scientific reports,*
2786 *6*. doi:10.1038/srep26770

2787 Williams, S. E., Wootton, P., Mason, H. S., Bould, J., Iles, D. E., Riccardi, D., *et al.*
2788 (2004). Hemoxygenase-2 is an oxygen sensor for a calcium-sensitive potassium
2789 channel. *Science, 306*(5704), 2093-2097.

2790 Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., *et al.* (2011). KOBAS 2.0: a web
2791 server for annotation and identification of enriched pathways and diseases.
2792 *Nucleic Acids Research, 39*, W316-322. doi: 10.1093/nar/gkr483.

2793 Xie, J.-M., Li, B., Yu, H.-P., Gao, Q.-G., Li, W., Wu, H.-R., *et al.* (2014). *TIGAR* has a
2794 dual role in cancer cell survival through regulating apoptosis and autophagy.
2795 *Cancer Research, 74*(18), 5127-5138.

2796 Yamada, K., Ono, M., Perkins, N. D., Rocha, S., and Lamond, A. I. (2013). Identification
2797 and functional characterization of FMN2, a regulator of the cyclin-dependent
2798 kinase inhibitor p21. *Molecular Cell, 49*(5), 922-933.

2799 Yu, H.-P., Xie, J.-M., Li, B., Sun, Y.-H., Gao, Q.-G., Ding, Z.-H., *et al.* (2015). *TIGAR*
2800 regulates DNA damage and repair through pentosephosphate pathway and *Cdk5-*
2801 *ATM* pathway. *Scientific Reports, 5*, 9853. doi:10.1038/srep09853

2802

Chapter 4

The autosomal genome ancestry of domestic chicken: Introgression from grey and Ceylon junglefowls

Abstract

In this chapter, I assess the autosomal genome ancestry of the domestic chicken using full genome re-sequencing data. Our results show that the red junglefowl *Gallus gallus* is the main ancestor of domestic chicken but with substantial contribution from the grey *G. sonneratii* and the Ceylon *G. lafayettii* junglefowls. Introgression from the grey junglefowl was found in all the domestic chicken populations studied, while Ceylon junglefowl introgression was mainly present in domestic chickens from Sri Lanka. I also reveal extensive introgression from domestic chicken or red junglefowl into grey junglefowl, introgression of domestic chicken into Ceylon junglefowl and, likely through domestic chicken, introgression of grey junglefowl into wild red junglefowl from China. None of the introgressed regions were found to be fixed or close to fixation. Multiple introgression from a single, but also from several *Gallus* species were commonly observed at a single genome region, including for a genome region showing evidence of domestic chicken introgression into grey junglefowl, and green junglefowl *G. varius* introgression into two domestic birds from Sri Lanka. Our results indicate that introgression and/or hybridization episodes with wild *Gallus* relatives have contributed substantially to the modern diversity and genetic make-up of domestic chicken population.

Introduction

Today, four species belonging to the genus *Gallus* have been described. In South Asia, the Ceylon junglefowl *Gallus lafayettii* is found in Sri Lanka and the grey junglefowl *G. sonneratii* in South-West India. In South-East Asia, the green junglefowl *G. varius*, a bird of the seashore and low-lying valleys, is endemic to the Java and neighbouring islands (Delacour, 1977), while the red junglefowl *G. gallus* is widely distributed across the South and South-East Asia with five recognized subspecies (Delacour, 1977). In the wild, hybridizations between the red junglefowl and the grey junglefowl have been reported on the Indian subcontinent where the range of the two species overlaps (Delacour, 1977).

The red junglefowl is the recognized maternal ancestor of domestic chicken with the assumption of multiple domestication centres (Liu *et al.*, 2006), and possible maternal contribution from four of its subspecies (Fumihito *et al.*, 1994; Fumihito *et al.*, 1996; Liu

et al., 2006). No mitochondrial DNA (mtDNA) haplotypes from other *Gallus* species as well as from the *G. gallus bankiva*, a red junglefowl subspecies with a geographic distribution restricted to Java and Sumatra in the Indonesian islands, have been found so far in domestic birds (Fumihito *et al.*, 1994; Fumihito *et al.*, 1996).

Inter-species hybridization and introgression is an evolutionary process that plays a major role in the genetic history of species and their adaptation (Arnold, 1997; Barton, 2001). It may occur in the wild when species live in sympatry or in captivity following human intervention. The unravelling of the mechanism of successful hybridization and introgression is central to our understanding of the speciation process in wild species (Mayr, 1963). Inter-species hybridization is commonly practised in agricultural plants and mammals leading to new environmental adaptation and/or improvement of productivity (Hufford *et al.*, 2013; Mina-Vargas *et al.*, 2016). Hybridization and introgression are relatively common in wild birds, including in the Galliformes (Shaklee and Knox, 1954; Asmundson and Lorenz, 1957; Olsen, 1960; Wilcox and Clark, 1961; Morejohn, 1968; Skjervold and Mjelstad, 1992). For example, the genetic integrity of the rock partridge *Alectoris graeca* is being threatened following the introduction and hybridization in its natural habitat of the red-legged partridge *A. rufa* (Barilani *et al.*, 2007). The presence of Japanese quail alleles in the wild migratory common quail *Coturnix coturnix* reveals hybridization between domestic quail and wild-type species (Barton, 2001; Chazara *et al.*, 2010). Mitochondrial DNA and nuclear microsatellite analysis indicate gene flow between silver pheasant *Lophura nycthemera* and kalij pheasant *L. leucomelanos* (Dong *et al.*, 2013).

Also, hybridization in captivity between *Gallus* species has been previously reported (Danforth, 1958; Morejohn, 1968). Morejohn (1968) successfully produced F1 hybrids between male grey junglefowl and female red junglefowl. He showed that these F1 are fertile in subsequent backcrossing with both species. Also, red junglefowl/domestic chicken mtDNA has been found in captive grey junglefowls (Nishibori *et al.*, 2005) and the yellow skin phenotype has been shown to likely originate from the introgression of a chromosomal fragment including the *BCDO2* locus from the grey junglefowl into domestic chicken (Eriksson *et al.*, 2008). Captive F1 hybrids between domestic hen and male green junglefowl are commonly bred in Indonesia where they are known as bekisar or kakok (Beebe, 1918; Delacour, 1977). They are prized for their plumage colour and

voice (Beebe, 1918; Delacour, 1977). Across genus, unfertile F1 hybrids between the common pheasant *Phasianus colchicus* and domestic chicken have also been reported in captivity (Castillo *et al.*, 2010; Castillo *et al.*, 2012).

In this chapter, I analyse the autosomal genome of geographically distinct populations of indigenous domestic chicken to assess their wild species ancestry. I report that the main ancestor is the red junglefowl but with grey junglefowl, Ceylon junglefowl and green junglefowl introgression as well as extensive introgression of domestic and/or red junglefowl into the grey junglefowl.

Materials and Methods

Samples and DNA extraction

Detailed samples information including location, country and categories are available at **supplementary Table S1**.

Blood samples were collected from the wing vein from 27 indigenous domestic chickens from four domestic populations: Ethiopia, (Horro n = 6 and Jarso n = 5) (Desta *et al.*, 2013), Sri Lanka, (n = 11), Saudi Arabia (n = 5) and two grey junglefowls, two green junglefowls and two Ceylon junglefowls all sampled from Koen Vanmechelen captive bird collection (http://www.koenvanmechelen.be/). Five wild Ceylon junglefowls originating from the Uva province in Sri Lanka were also sampled. Genomic DNA was extracted following the procedure described in Bruford *et al.* (1998) and was sent to the Beijing Genomic Institute (BGI) or Novogene for full genome sequencing on a HiSeq 2000/2500 Illumina platform. A minimum quantity of 3 µg of genomic DNA at 30 ng/µl concentration was used for the construction of 500 bp paired-end insert size library. The read length varies from 90 - 100 bp and 10 X - 30 X genome coverage (**supplementary Table S2**).

This dataset was complemented with the genome sequences (10 X – 36 X genome coverage) obtained from other studies including sequences from three domestic birds from the Cosmopolitan Chicken Research Project (CC®P) (the parental birds (Poule de Bresse, and Mechelse Koekoe) and the Mechelse Styrian, a 16th generation crossbred bird), one captive red, grey, Ceylon and green junglefowl and five red junglefowls, sampled in the Yunnan or Hainan Provinces (People's Republic of China) (Wang *et al.*, 2015).

I grouped the 50 genomes into five populations: domestic chicken (n = 30), red junglefowl (n = 6), grey junglefowl (n = 3), Ceylon junglefowl (n = 8) and green junglefowl (n = 3). Domestic chickens were further separated into five subpopulations: Europe (CC®P parent n = 2, 16th generation crossbred, n = 1), Horro (n = 6, Ethiopia), Jarso (n = 5, Ethiopia), Saudi Arabia (n = 5), Sri Lanka (n = 11).

For mitochondrial DNA sequences analysis, I completed the above dataset with 20 full mitochondrial DNA sequences downloaded from NCBI, including 10 domestic chicken from China, India, and Japan and as well, 8 non-red junglefowl birds (*G. sonneratii* (n = 4), *G. lafayettii* (n = 2), *G. varius* (n = 2)) (**supplementary Table S3**).

Raw sequence quality check

The adapter pollutions from the raw reads were deleted at the source (BGI/Novogene) and reads that contained more than 50% low quality bases (quality value ≤ 5) were removed.

Sequence mapping and variants calling

For detail regarding mapping and variants calling procedure, please refer to materials and methods section of Chapter 2.

I report here the analysis of the autosomal chromosomes (1 – 28) and the mitochondrial DNA only.

The consequences of the autosomal variants were predicted using the Ensembl 'VEP' version 85 (Aken *et al.*, 2016). The number of SNPs per sample and per population were identified using VCFtools version 0.1.11 (Danecek *et al.*, 2011). Percentage of read pairs mapped to the same chromosome, mean depth coverage, total reads mapped, and percentage of the genome with bases covered by at least 5, 10 and 20 reads were calculated using SAMtools version 0.1.19 (Li *et al.*, 2009). Unique and shared SNPs among populations were calculated and plotted in R using the package 'VennDiagram' (Chen and Boutros, 2011).

Population structure and samples relationships

Principal component analysis was used to visualise the genetic similarity among the populations using smartpca program in Eigenstrat version 6.0.1 (Price *et al.*, 2006). All the ~ 40 million autosomal SNPs identified across the 50 samples were used for this analysis.

I then randomly subset 80,906 SNPs corresponding to 0.2% of all jointly called autosomal SNPs in the 50 genomes. Best-fit model was predicted using jModeltest version 2.1.7 (Darriba *et al.*, 2012) and a maximum likelihood relationship tree across the samples was built using phyml version 3.0 (Guindon and Gascuel, 2003).

Nucleotide diversity (π) across the autosomes and within population was estimated using 20 kb windows with a 5 kb sliding step using VCFtools version 0.1.11 (Danecek *et al.*, 2011). Pairwise genetic differentiation F_{ST} and nucleotide divergence (d_{XY}) between populations were further performed using VCFtools-F_{ST} and Martin's d_{XY} script (Martin *et al.*, 2015) at 20 kb windows with a 5 kb sliding step.

For the mitochondrial DNA (mtDNA) analysis, as the full mtDNA length varies across *Gallus* species, I mapped the reads of all the non-red junglefowl to their corresponding mtDNA reference species (**supplementary Table S3**). Sequences were then aligned using ClustalX version 2.1 (Thompson *et al.*, 1997) and all gaps or missing positions were excluded from the analysis. Following jModeltest best-fit model prediction, neighbour-joining and maximum likelihood trees were built with the MEGA 7.0 (Kumar *et al.*, 2016) and phyml 3.0 (Guindon and Gascuel, 2003) software respectively.

Introgression – admixture analysis

To detect gene flow between domestic chicken and the non-red junglefowl species, I used the four-taxon admixture method (ABBA/BABA) from which the Patterson's D (Durand *et al.*, 2011) and the Martin's *fd* (Martin *et al.*, 2015) statistics were calculated. Analyses were performed across each autosome on a 20-kb window size with 5 kb sliding step. Only windows with a minimum of 100 SNPs were considered. It excluded 195 windows for the domestic chicken – grey junglefowl comparison and 185 windows for the domestic chicken – Ceylon junglefowl comparison (**supplementary Figure S1**). These two statistics detect regions having an excess of shared derived allele between any two in-

groups by comparing the allelic state of the third in-group with the fourth assumed to be the ancestral out-group population (Green *et al.*, 2010; Durand *et al.*, 2011). Considering the three in-groups, P_1 (red junglefowl), P_2 (domestic chicken) and P_3 (Ceylon or grey junglefowl), and an out-group *O* (green junglefowl), the expected phylogeny is (((*P1*, *P2*), *P3*), *O*). The admixture test can then be computed by comparing the two groups P_1 - P_3 (BABA) and P_2 - P_3 (ABBA). I defined ABBA as any given sites where the derived allele 'B' is shared between the domestic chicken 'P_2' and the Ceylon or grey junglefowl 'P_3', while the red junglefowl 'P_1' shared the ancestral allele 'A' with the green junglefowl '*O*'. In BABA, the red junglefowl 'P_1' shared the derived allele 'B' with 'P_3' while the domestic chicken 'P_2' shared the same ancestral state with the outgroup '*O*'. Under the assumption of no gene flow or neutral coalescent model, both the ABBA and BABA should result in same count and D value tends to zero.

The Martin's *fd* statistics was calculated using the equation:

$$fd = \frac{S(P1,P2,P3,O)}{S(P1,PD,PD,O)}$$ (Martin *et al.*, 2015)

Where *S* is the difference between the sums of ABBAs and BABAs. *P1*, *P2*, *P3* represent the in-groups and *O* the out-group. The *PD* is the donor population for independent sites. Since *fd* detect bi-directional introgression, the *PD* could either be *P*2 or *P*3 as in the case of ABBAs.

The values of the *fd* were then Z-transformed using the equation:

$$Zfd = \frac{fd - \overline{X}(fd)}{\sigma(fd)}$$

Where $\overline{X}$ is the mean and σ is the standard deviation of the *fd*

I defined arbitrarily, the candidate region of introgression from windows that fall within the genome-wide *Zfd* score of ≥ 6.0 for the domestic chicken - Ceylon junglefowl comparison and *Zfd* ≥ 4.0 for the domestic chicken - grey junglefowl comparison, following visual examination of the Manhattan plots.

The overall admixture proportion at genome and chromosome levels within population was estimated by taking the average of D statistics (Durand *et al.*, 2011) and topology weighting (Martin and Van Belleghem, 2017)..

Haplotype analysis

To infer relationship between haplotypes at each introgressed region and to predict possible direction of gene flow, I followed a haplotype phylogenetic tree approach. The expectation is that introgressed regions in a population will show haplotypes closely linked to or embedded within the donor's species rather than within the recipient species haplotypes.

First, I phased our data using the read aware phasing method of shapeit (Delaneau *et al.*, 2013) that extracts directly the phase informative read from the sequence BAM file. The phasing output was then converted into a variant call format (vcf) file and subsequently formatted in Plink 1.9 with the 'beagle recode' option (https://www.cog-genomics.org/plink2) which was then used to create, using a customized script, a fasta file. Multiple sequences alignment was done using Clustalx 2.1 (Thompson *et al.*, 1997). The evolutionary model was predicted using jModeltest 2.1.7 (Darriba *et al.*, 2012) based on the Akaike information criterion (AIC). Phyml 3.0 (Guindon and Gascuel, 2003) was used to compute the approximate likelihood ratio score for each branch using the best predicted model.

Frequency of the introgressed haplotypes in each chicken population was calculated by dividing the number of introgressed haplotypes with the total number of haplotype in each population.

Size of introgressed regions

Given our chosen arbitrary significant threshold of significance, adding the size of significant adjacent windows will most likely provide an underestimation of the size of the introgressed regions. For the introgression of Ceylon junglefowl and green junglefowl into domestic chicken, I also assess adjacent windows next to the significant introgressed ones for evidence of introgression by building-up individual window haplotype tree for these non-significant windows. The size of the introgressed region was then calculated by adding the size of each adjacent window showing evidence of haplotype introgression whether it reached or not the significant threshold. Introgressed regions on the same chromosomes were considered distinct if separated by at least one 20 kb window. Evidence of introgression in grey junglefowl into domestic chicken and from domestic

and red junglefowl into grey junglefowl were only assessed for windows above the significant thresholds given the large number of candidate regions detected.

Remapping *Galgal* 4.0 introgressed regions to *Galgal* 5.0

Following the release of new reference genome *Galgal* 5.0 after the mapping and variant calling using *Galgal* 4.0 reference genome, I remapped the identified *Galgal* 4.0 introgressed regions to the corresponding *Galgal* 5.0 coordinates using NCBI remapper (February 2017 release). All remapping options were set to default threshold. All changes in the genes annotation between the two reference genomes were reported, except for the *Galgal* 5.0 genes of domestic introgression into grey junglefowl.

Genome-wide topology weighting estimation

I estimated the relationship between the five taxa (domestic, red, grey, Ceylon and green junglefowls) using non-overlapping windows of a fixed number of 50 SNPs across the genome using *Twisst* (Martin and Van Belleghem, 2017). At each window, trees were inferred using Phyml 3.0 (Guindon and Gascuel, 2003) with the optimization of maximum likelihood for the topology and branch lengths. The result for the 15 possible topologies rooted around the green junglefowl were plotted in R using the package ape version 4.1 (Paradis *et al.*, 2004). Our data were stacked plotted with a Loess smoothing (span = 0.005).

Gene ontology analysis

see chapter 2.

Overlaps between introgressed regions and chicken QTLs

The Quantitative Trait Locus database (QTLdb) release 31 (https://www.animalgenome.org/cgi-bin/QTLdb/GG/index) was used to identify possible overlaps between introgressed regions and previously reported chicken QTL by downloading the reference *Galgal* 4.0 chicken QTLdb positions available in BED format.

Results

Sequence polymorphisms and variants statistics

An average mapping rate of 98% was obtained with at least 90% of both mates of a read pair mapped to the same autosome in domestic chicken. In red junglefowl, it ranges from 90.0% to 95.33%, except for redjunglefowl1 with 67.51% of read pairs mapped to the same autosome. For the grey junglefowl, it ranges from 87.11% to 94.68%, 88.22% to 94.27% for the Ceylon junglefowl, and 85.22% to 94.38% for the green junglefowl (**supplementary Table S2**). The percentage of each sample variant supported by at least 10 reads ranged from 43.2% to 97.3% depending on the genome coverage (**supplementary Table S2**). To produce the correct genotype likelihood, VCF files from all 50 genomes were jointly analysed and after quality control, 40,377,927 autosomal bi-allelic SNPs (26,798,977 SNPs novel and 13,578,950 SNPs previously reported) (Aken *et al.*, 2016) were identified. The autosomal SNPs distribution per chromosome is presented at **supplementary Figure S2.** As expected, the number of SNPs is proportional to the chromosome length with the notable exception of chromosome 16 for which the genome assembly is fragmentary (**supplementary Figure S2**).

The numbers of unique and shared SNPs across population are presented at **supplementary Figures S3.1 – S3.3.** At the individual bird level, it ranges from ~ 5.8 million to 6.7 million for domestic chicken, 6.0 to 7.1 million SNPs for red junglefowl, ~ 10 million for grey junglefowl, ~ 11.0 – 11.8 million for Ceylon junglefowl and 12.8 million for green junglefowl. These numbers include both the heterozygous and homozygous SNPs (**supplementary Table S2**). As expected all non-red junglefowl samples have higher number of SNPs since they are more genetically distant from the *Galgal* 4.0 reference genome.

At the population/species level, ~ 17.5 million SNPs were found in domestic chicken (n = 30), ~ 15.3 million in red junglefowl (n = 6), ~ 11.9 million in grey junglefowl (n = 3), ~ 17.3 million in Ceylon junglefowl (n = 8) and ~ 13.7 million SNPs in green junglefowl (n = 3) (**supplementary Figure S3.3)**. Amongst these, ~ 11.7 million SNPs were shared between domestic and red junglefowl, ~ 5.8 million are unique to domestic chicken and ~ 3.6 million are unique to red junglefowl (**supplementary Figure S3.2**). In the non-red

junglefowl species, ~ 3.0 million, ~ 8.0 million and ~ 6.6 million SNPs are unique to grey, Ceylon and green junglefowl, respectively (**supplementary Figure S3.3**).

Around 5.5 million SNPs are common to domestic and grey junglefowl, including 287,841 SNPs uniquely found in these two species, ~ 5.6 million SNPs are common to domestic chicken and Ceylon junglefowl, including 250,526 unique SNPs, and ~ 4.8 million SNPs are common to domestic and green junglefowl of which 162,082 were uniquely shared (**supplementary Figure S3.3**).

The variants statistics for the SNPs as predicted by Ensembl version 85 (Aken *et al.*, 2016) across the five populations are presented at **supplementary Table S4**. Polymorphisms were identified in more than 39,000 genes and ~ 47,000 transcripts. Intronic variants in each population are more than 5 million while intergenic variants range from ~ 3 million to ~ 5.5 million.

Population and species differentiation

I assessed the overall genetic variation among and within the populations using the ~ 40 million autosomal SNPs identified across the 50 genomes**.** I performed two PCA plots. First, I analysed all the five populations together (**Figure 1A)**. Domestic chicken and red junglefowl cluster tightly together, and they are well separated from the grey and Ceylon junglefowl, while the green junglefowl is the most distinct. Then, I analysed the domestic and red junglefowl birds only (**Figure 1B)**. Domestic chicken birds are now clearly distinct from the red junglefowl (PC1), with the Saudi Arabian chicken positioned closer to the Ethiopian population than to the domestic chicken from Sri Lanka. The fancy chicken group is embedded within a group of eight Sri Lankan domestic, separated by PC2 from two other domestic birds from Sri Lanka. The five red junglefowl are well separated in three groups.

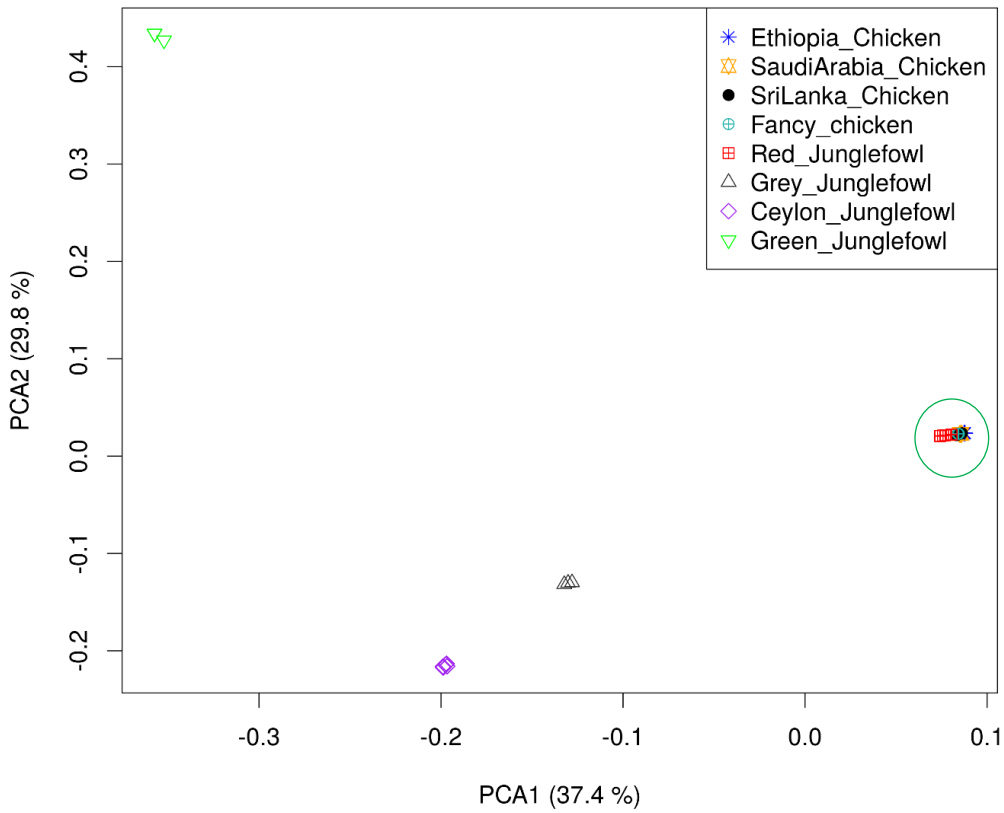


Figure 1A. Principal component analysis for the 50 chicken genomes (~ 40 million SNPs). The samples within the green circle include all the red junglefowl and domestic chicken.

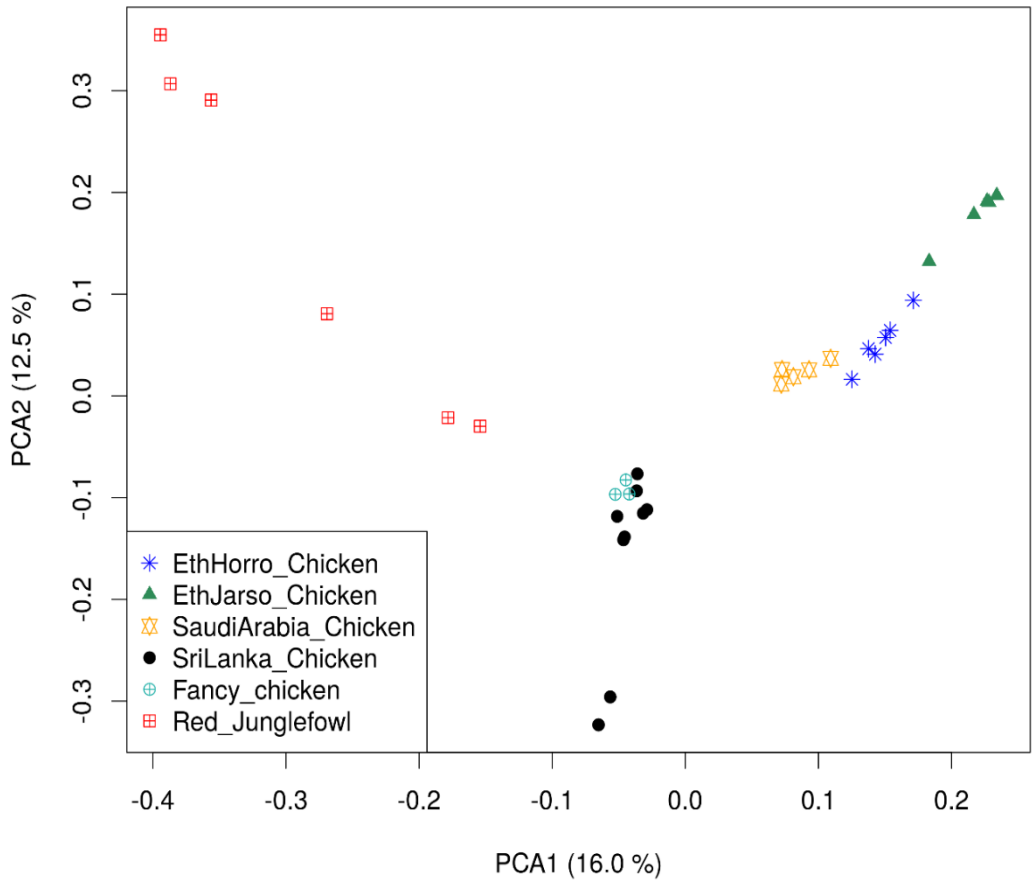


Figure 1B. Principal component analysis for domestic chickens (n = 30) and red junglefowl (n = 6), ~ 20,218,465 autosomal million SNPs.

I estimated the average genome nucleotide diversity for each population. Within the domestic chicken populations, the highest π was found in Sri Lanka domestic chickens (π = 0.0046 ± 0.0018, n = 11) and the lowest average genome nucleotide diversity was observed in Saudi Arabian domestic chicken (π = 0.0039 ± 0.0018, n = 5). Among the wild junglefowl population, π was highest in red junglefowl (0.0052 ± 0.0019, n = 6) and the lowest in the green junglefowl (0.0014 ± 0.0008, n = 3) (**supplementary Table S5**).

I measured the level of autosomal genetic differentiation between populations using F_{ST} (Danecek *et al.*, 2011) and mean absolute genetic divergence (d_{XY}) (Martin *et al.*, 2015). Between species, F_{ST} ranged from 0.596 (red – grey junglefowls pairwise comparison) to 0.856 (grey – green junglefowls pairwise comparison). The comparison of each of the three domestic chicken populations (Ethiopia, Saudi Arabia, Sri Lanka domestic chicken) with the red junglefowl show an average F_{ST} value less than 0.100 ± 0.019. Comparisons of each of the domestic population with the non-red junglefowl show F_{ST} values ranging

from 0.621 (between domestic chicken from Sri Lanka and grey junglefowl) to 0.779 (between the fancy birds and the green junglefowl) (**supplementary Table S6**).

The lowest genetic divergence (d_{XY}) was observed between domestic chicken and red junglefowl (0.110 ± 0.025), while divergence with other junglefowl species was at least twice as high. Between junglefowl species, it ranges from 0.185 ± 0.039 (Ceylon and grey junglefowl) to 0.287 ± 0.035 (red and green junglefowl) (**supplementary Table S7**).

Population and species relationships

Mitochondrial DNA analysis

The 50 mitochondrial DNA sequences together with the mitochondrial genomes of *Galgal* 4.0 and 20 reference sequences were used to build-up a maximum likelihood (**Figure 2A)** and a neighbour-joining (**supplementary Figure S4)** trees. Following the AIC prediction, GTR+I+G was selected as the best fit-model for both trees (**supplementary Table S8.1).** Both trees show red junglefowl mtDNA embedded within the domestic chicken. I also find within the same lineage the mitochondrial DNA of five out of seven grey junglefowls examined, indicative of domestic or red junglefowl maternal introgression into the captive breed junglefowl studied here (Sonneratii 1k – 3k). At the opposite, I find no evidence of maternal introgression from the red junglefowl and/or domestic chicken into the Ceylon and green junglefowl. The tree also indicates a common maternal ancestry between the grey and Ceylon junglefowl. These two species link to the red junglefowl – domestic chicken lineages and the green junglefowl is the outgroup, with two distinct lineages amongst the birds examined.

Autosomal SNPs analysis

I randomly subset 80,906 SNPs across the genome, corresponding to 0.2% of our dataset. This subset of SNPs was used to construct an un-rooted maximum likelihood tree (**Figure 2.2**, **supplementary Table S8.2**).

All nodes show a 100% bootstrap support. Samples from the same population or species grouped together. The red junglefowl birds are positioned at the root of the domestic chicken lineages. Among the domestic chickens, birds from Ethiopia and Saudi Arabia cluster together. Grey and Ceylon junglefowls are more closely related to each other than

to any other populations or species. The ancestral node of these two species links to the domestic chicken and red junglefowl lineages. The green junglefowl lineage is the outgroup.

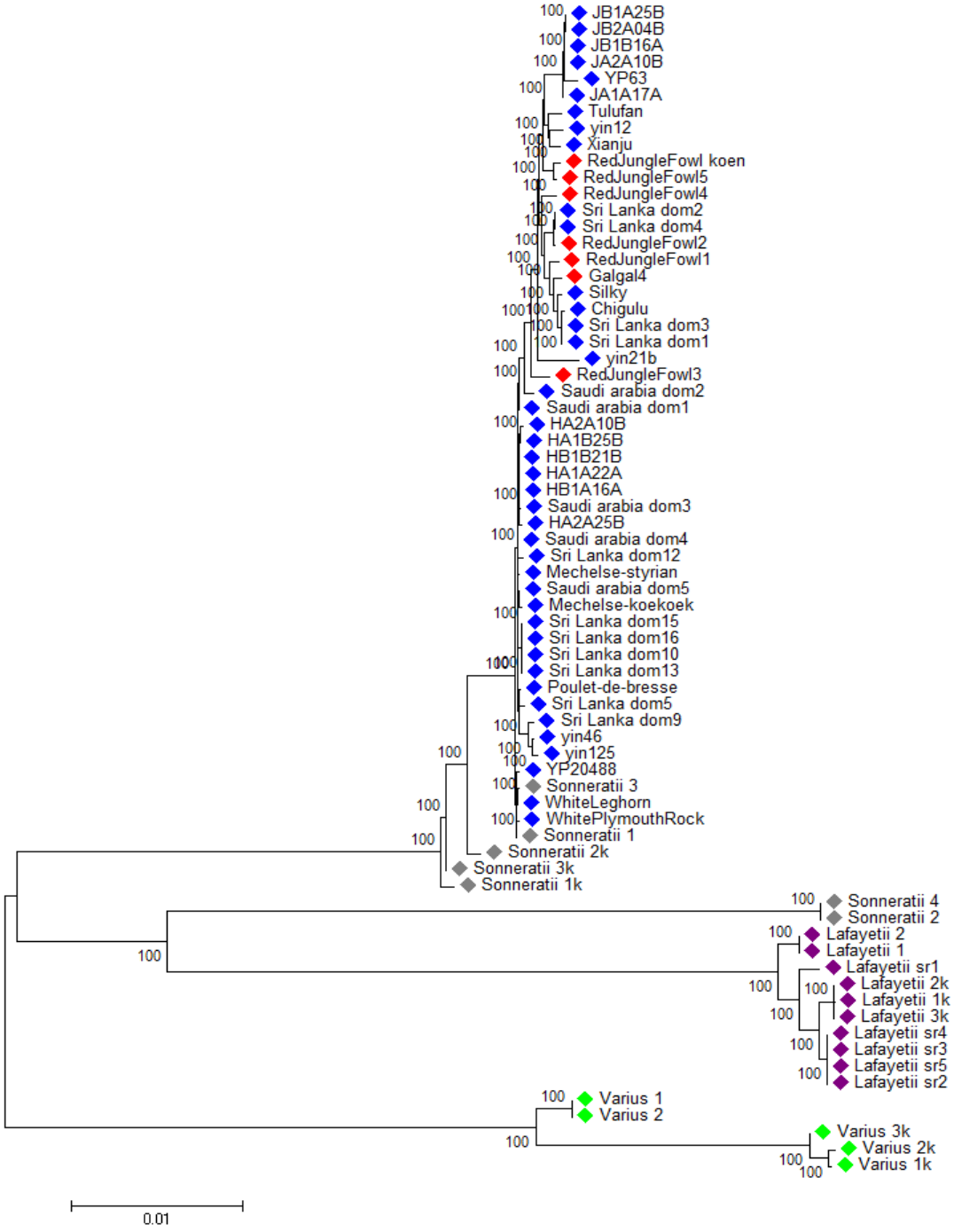


Figure 2A. Full mitochondrial DNA maximum likelihood tree. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. Besides the 50 studied birds, the tree includes the full mtDNA of 11 birds of reference (**supplementary Table S3**). Numbers above the nodes are bootstraps values. The tree was constructed based on the predicted GTR+I+G model by the Akaike information criterion

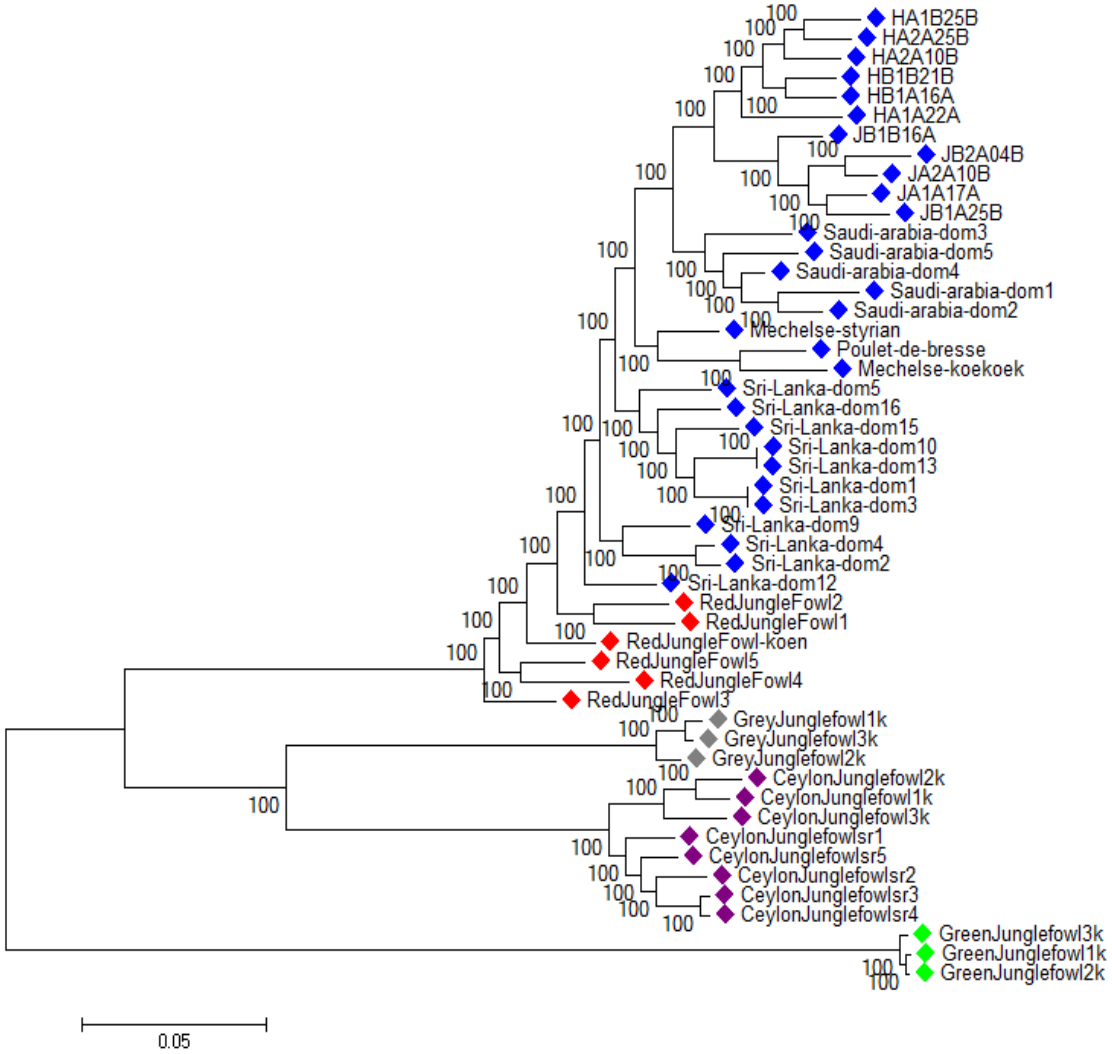


Figure 2B. Maximum likelihood tree based on 80900 (0.2%) randomly selected autosomal SNPs. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. Numbers above the nodes are bootstraps values (see **supplementary Table S1** for further details regarding the samples).

Genome-wide autosomal introgression

Using the Patterson D statistics (Durand *et al.*, 2011), which ranges from + 1 (complete admixture between P_2 and P_3) to -1 (complete shared variation between the P_1 and P_3), I estimated the proportion of introgression across the autosomal genome. The population mean genome-wide D statistics are 0.758 ± 0.192 between domestic chicken and red junglefowl, 0.040 ± 0.215 between domestic chicken and grey junglefowl, 0.011 ± 0.197 between domestic chicken and Ceylon junglefowl. **Supplementary Table S9** indicates the average D statistic values per chromosome.

Locus specific introgression

Introgression between the grey junglefowl and domestic chicken

Comparing the genome of domestic chicken with the grey junglefowl, 2337 windows (1.29%) out of 181,340 windows reached the genome threshold of *Zfd* score ≥ 4.0 of the ABBA/BABA test. These defined 375 distinct candidate introgressed regions. No candidate introgressed region between these two species was found at this significant threshold on chromosomes 5, 9, 10, 12, 14, 15, 16, 22, 23, 25, 27 and 28 (**Figure 3).** Details regarding the introgressed regions are provided at **Table 1A, 1B, Tables S10A** and **S10B**.

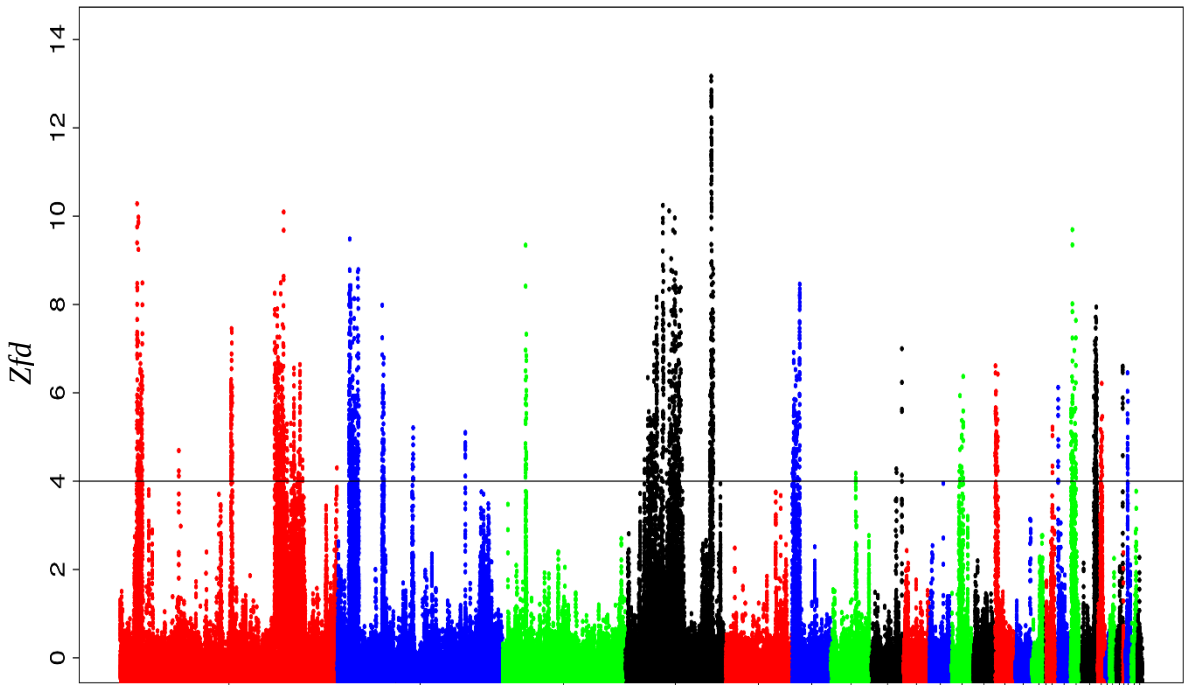


Figure 3. ABBA/BABA introgression plots. Introgression between domestic chicken and grey junglefowl. The *Zfd* is the transformed value obtained from the Martin *fd* statistics. The horizontal line corresponds to the *Zfd* score of 4.

Haplotype tree analyses show that 41 of these candidate regions correspond to introgression of grey junglefowl into domestic chicken, including 32 regions which also show evidence of introgression from domestic chicken into grey junglefowl (**Table 1A**, **supplementaries Figures S5.1 - S5.2, S5.7 – S5.11, S5.14 – S5.38**), 317 candidate regions show introgression of domestic chicken into the grey junglefowl (**supplementary Table S10A**). Two candidate regions show both the introgression of domestic chicken

into grey junglefowl, and introgression of green junglefowl into domestic chicken (**Table 1B**, **Figures 4B** and **S6.1**). For the windows above the significant threshold of introgression from grey junglefowl into domestic chicken, the size of the introgressed fragments ranges from 20 kb to 100 kb. Mean and median sizes are 24 kb and 30 kb respectively (**Table 1A**).

The size of the two regions of introgression from green junglefowl into domestic chicken were then further estimated by building-up the haplotype trees of the adjacent window to the significant ones. It shows that the two candidate regions of introgression of green junglefowl positioned close to each other on chromosome 2 and separated by only a window of 20 kb is a single region of introgression of around 220 kb (*Galgal* 4.0 chromosome 2, position 18235001 – 18455001 bp) in length (**Figures 4B** and **supplementary Figure S6.1** and **S6.2, Table 1B)**.

I observed the highest number of candidate grey junglefowl introgressed regions in Sri Lanka (n = 33), then Ethiopia (Jarso (n = 16), Horro (n = 14)) and Saudi Arabia (n = 14)). Three candidate introgressed regions are present in all the CC®P birds, while three are unique to Mechelse Koekoe and four are unique to Mechelse Styrian. Population frequencies of the grey introgressed haplotypes were higher in Ethiopia (Horro (50.6%) and Jarso (38.8%)) and Saudi Arabia (37.9%) when compared to Sri Lanka (15.5%) and red junglefowl (15.5%) (see **supplementary Table S11A** for details). The single candidate introgressed region of green junglefowl origin is found in two Sri Lanka birds (population frequency 11%) (**supplementary Table S11B**).

For 15 regions, the haplotype trees do not allow to identify the direction of introgression (**supplementary Figures S8** and **supplementary Table S10B**).

An example of haplotype trees showing introgression from grey junglefowl into domestic chicken is shown at **Figure 4A**. The other trees (n = 40) are presented at **supplementary Figures S5.1 – S5.40**. Haplotype trees (32 out of 41) showing introgression from grey junglefowl into domestic as well as domestic chicken into grey junglefowl are shown at (**supplementary Figures S5.1 - S5.2**, **S5.7 - S5.11, S5.14 - S5.27, S5.29 - S5.39)**. **Supplementary Figure S5.39** includes the *BCDO2* locus region which showed previously the introgression in domestic chicken from the grey junglefowl (Eriksson *et al.*, 2008). This 30-kb region shows multiple introgression of grey junglefowl into

domestic chicken. Six introgressed candidate regions of grey junglefowl origin were also observed in three red junglefowl samples (red junglefowl1, red junglefowl2, and red junglefowl3) (**Figures 4A, supplementary Figures S5.4, S5.6, S5.18 - S5.20, S5.29, supplementary Table S11**).

The size of the 317 candidate introgressed regions from domestic chicken into grey junglefowl, excluding those showing bi-directional introgression, ranges from 20 to 380 kb, with a mean fragment size of 39 kb and a 30-kb median size. A few regions (n = 31) have size larger than 70 kb. The largest one (380 kb) is located on chromosome 4 at position 76905001 - 77285001 bp (**Figure 4C)**. At this region, the average D statistics value is 0.58 ± 0.167 and the *Zfd* score 8.84 ± 3.077. The other trees are presented at **supplementary Figures S7.1 - S7. 317.**

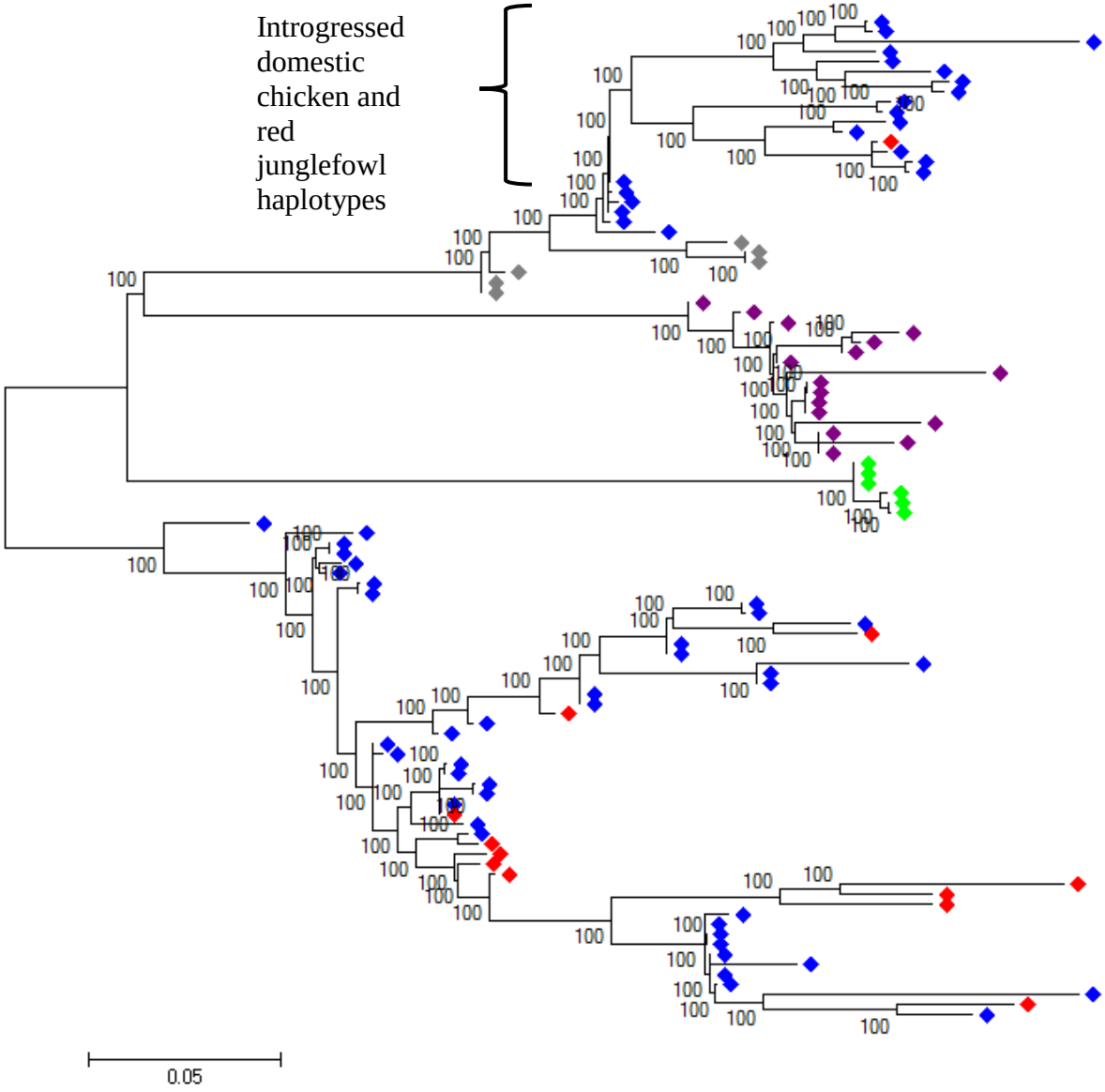


Figure 4A. Haplotype based maximum likelihood tree illustrating introgression of grey junglefowl into domestic chicken. ABBA – BABA test grey junglefowl – domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is on chromosome 1 (99665001 to 99740001 bp) and it is 75 kb in length. The six grey junglefowl haplotypes are found at the root of 21 domestic chicken and one red junglefowl haplotypes.

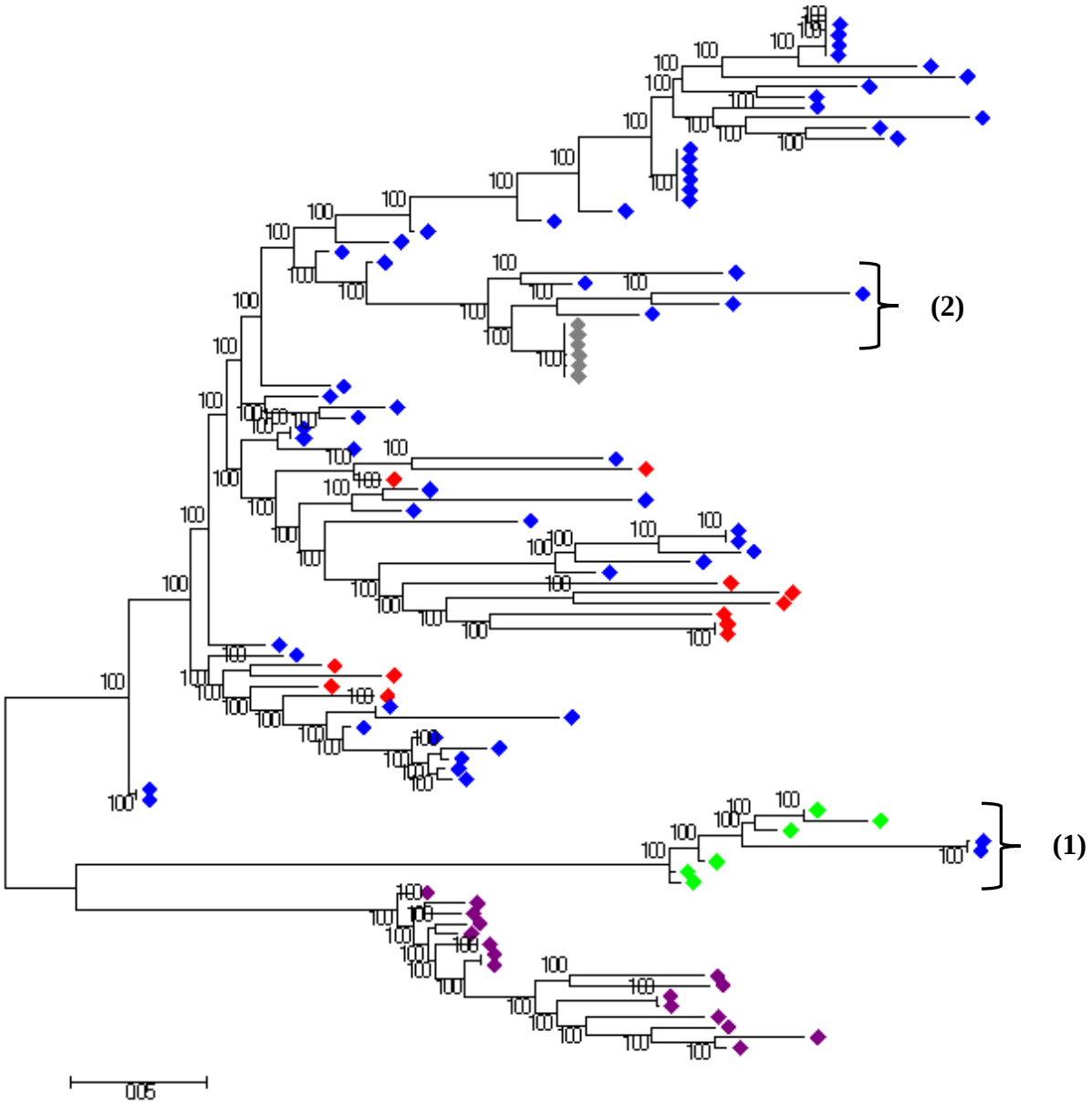


Figure 4B. Haplotype based maximum likelihood tree illustrating introgression of green junglefowl into domestic chicken. ABBA – BABA test grey junglefowl – domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The tree shows two patterns of introgression. The region is on chromosome 2 (18235001 – 18455001 bp) and is 220 kb fragment in length. **(1)** Introgression of a green junglefowl haplotype into Sri Lanka1 and 3. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

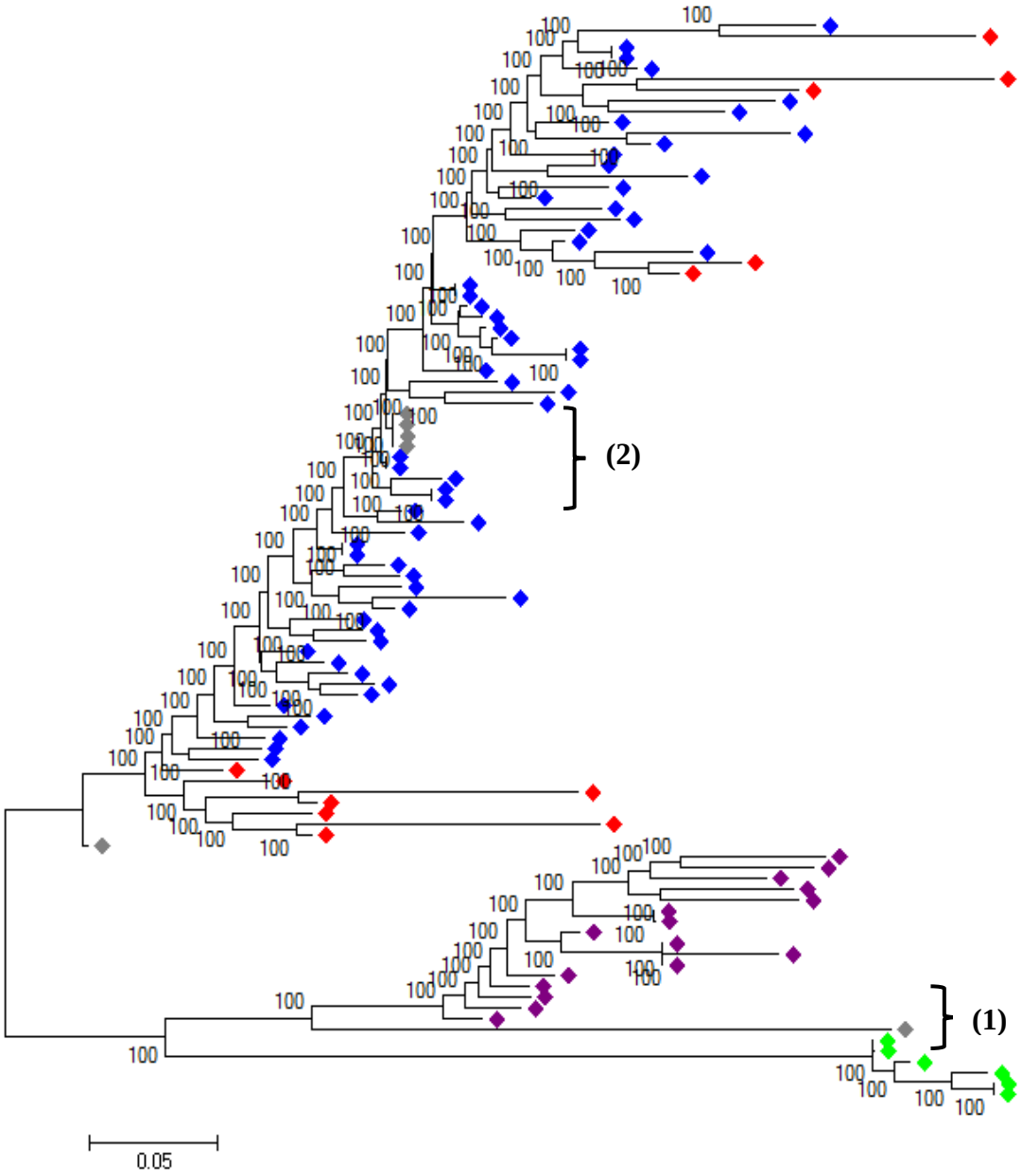


Figure 4C. Haplotype based maximum likelihood tree illustrating introgression of domestic chicken – red junglefowl into grey junglefowl. ABBA – BABA test grey junglefowl – domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The region is on chromosome 4 (76905001 - 77285001 bp) and it is 380 kb in length. **(1)** A single grey junglefowl haplotype was observed. **(2)** All others but one grey haplotype is nested within the domestic chicken haplotypes.

Table 1A. ABBA – BABA test grey junglefowl – domestic chicken, introgressed grey junglefowls into domestic chicken regions (n = 41). The **Ggal** is the reference genome (version 4 or 5) and **Ln4** and **Ln5** are the length in kilobase for the **Ggal4** and **Ggal5** coordinates respectively. **Ggal4_μ(D)** and **Ggal4_ μ(*Zfd*)** are mean of D and *Zfd* statistics based on the *Galgal* 4.0 reference genome.

Chr	Ggal4_start	Ggal4_stop	Ln4	Ggal4_μ(D)	Ggal4_μ(*Zfd*)	Ggal4_Genes	Ggal5_start	Ggal5_stop	Ln5	Ggal5_Genes
1	15180001	15255000	75	0.30 ± 0.057	4.68 ± 0.957	*KIF21A,*	15380995	15456146	75	*KIF21A*
1*	18335001	18355001	20	0.62	6.24	-	18536701	18556706	20	-
1	90920001	90940000	20	0.39	6.03	*RCSD1*	91555311	91575316	20	*RCSD1*
1	99665001	99740000	75	0.67 ± 0.017	4.47 ± 0.326	-	100356010	100431002	75	-
1	99915001	99935000	20	0.66	6.47	-	100605752	100625751	20	-
1*	100195001	100240001	45	0.84 ± 0.014	6.47 ± 0.367	*NCAM2*	100885910	100930911	45	*NCAM2*
1	100785001	100885000	100	0.56 ± 0.074	4.34 ± 1.166	-	101449355	101549388	100	-
1*	139220001	139245000	25	0.31 ± 0.045	4.72 ± 0.656	-	139891008	139916184	25	*ENSGALG00000042334 (LincRNA)*
1*	141700001	141755000	55	0.47 ± 0.068	4.81 ± 0.651	-	142447972	142503079	55	-
1*	145600001	145630000	30	0.40 ± 0.051	4.91 ± 0.721	-	146369482	146399480	30	-
1*	147760001	147810001	50	0.53 ± 0.077	4.58 ± 0.395	-	148530506	148578338	48	*ENSGALG00000045327 (LincRNA)*
1*	147640001	147660000	20	0.68	4.03	-	148410494	148430492	20	*ENSGALG00000038439 (LincRNA)*
1	185245001	185300000	55	0.34 ± 0.05	6.97 ± 0.392	*SLC36A4*	186055645	186110602	55	*SLC36A4*
1	185435001	185485000	50	0.57 ± 0.045	6.81 ± 0.523	*MTNR1B*	186245604	186295602	50	*MTNR1B*
4*	17015001	17035000	20	0.53	4.89	*IL1RAPL2*	17123967	17143964	20	*IL1RAPL2*
4*	33685001	33720000	35	0.39 ± 0.034	5.20 ± 0.427	*INTU*	34501238	34536233	35	*INTU*
4*	37045001	37070000	25	0.42 ± 0.005	4.43 ± 0.378	*COL25A1*	37874775	37899780	25	*COL25A1*
4*	41895001	41930000	35	0.27 ± 0.031	5.27 ± 0.842	-	42736892	42771892	35	-
4*	41960001	41985001	25	0.38 ± 0.004	6.36 ± 0.420	-	42801899	42826897	25	-
4*	42020001	42040001	20	0.37	5.83	-	42861898	42881898	20	-
4*	42085001	42115000	30	0.28 ± 0.014	4.20 ± 0.659	-	42926902	42956900	30	-

4*	42395001	42430001	35	0.39 ± 0.063	7.22 ± 0.717	*GALNTL6*	43236814	43271814	35	-
4*	45220001	45255000	35	0.33 ± 0.045	5.46 ± 0.407	*PTPN13*	46120392	46155390	35	*PTPN13*
4*	45950001	45970000	20	0.3	6.05	-	46850403	46870402	20	-
4*	46970001	47005000	35	0.35 ± 0.041	4.30 ± 0.517	-	47877081	47912685	36	-
6*	1260001	1280000	20	0.32	4.04	-	1759046	1779044	20	*GRID1*
6*	1370001	1400000	30	0.35 ± 0.067	4.89 ± 0.578	*GRID1*	1869048	1899047	30	*GRID1*
6*	1415001	1445001	30	0.41 ± 0.051	6.35 ± 0.114	*GRID1*	1914051	1944050	30	*GRID1*
7	22200001	22245000	45	0.67 ± 0.033	3.96 ± 0.185	-	22825998	22870995	45	-
3*	2165001	2190001	25	0.31 ± 0.015	4.55 ± 0.321	*ETF1*	2794531	2819530	25	*ETF1*
3*	2395001	2415001	20	0.29	6.02	-	3024953	3044953	20	-
9*	1635001	1680001	45	0.52 ± 0.049	7.63 ± 1.092	-	1613175	1658172	45	*AUTS2*
20*	13345001	13370001	25	0.49 ± 0.028	6.38 ± 0.060	*ENSGALG00000020877*	13217720	13242708	25	*ZFP64*
20*	13505001	13540001	35	0.59 ± 0.055	6.05 ± 0.011	*ATP9A, NFATC2*	13377552	13412401	35	*NFATC2, ENSGALG00000034664 (LincRNA)*
21*	3005001	3030001	25	0.33 ± 0.009	4.22 ± 0.207	-	3001214	3026217	25	-
21*	3230001	3250001	20	0.31	4.65	*SLC2A5*	3226367	3246366	20	-
21*	3335001	3360001	25	0.43 ± 0.052	4.42 ± 0.088	*SPSB1*	3331357	3356357	25	*SPSB1*
21*	3495001	3520001	25	0.28 ± 0.019	4.66 ± 0.348	*CLSTN1*	3491257	3516257	25	*CLSTN1*
21*	3570001	3595001	25	0.36 ± 0.045	4.77 ± 0.918	*LZIC, NMNAT1, RBP7*	3566258	3591258	25	*LZIC, NMNAT1, RBP7*
24*	6130001	6160000	30	0.68 ± 0.035	6.53 ± 0.073	*PTS, BCDO2*	6107102	6137115	30	*PTS, BCDO2*
28	2110001	2130000	20	0.29	5.07	*CSNK1G2, SCAMP4*	2181105	2201006	20	*SCAMP4*

*Bidirectional introgression: regions with evidence from grey junglefowl introgression into domestic chicken and *vice versa*.

Table 1B. ABBA – BABA test grey junglefowl – domestic chicken, introgressed green junglefowl regions into domestic chicken (n = 1). The **Ggal** is the reference genome for 4 or 5 version and **Ln4** and **Ln5** are the length in kilobase for the **Ggal4** and **5** coordinates respectively. **Ggal4_μ(D)** and **Ggal4_ μ*(Zfd)*** are mean of D and *Zfd* statistics based on *Galgal* 4.0 reference genome respectively

Chr	Ggal4_start	Ggal4_stop	Ln4	Ggal4_μ(D)	Ggal4_ μ*(Zfd)*	Ggal4_Genes	Ggal5_start	Ggal5_stop	Ln5	Ggal5_Genes
2*	18310001	18345001	35	0.42 ± 0.023	4.73 ± 0.413	-	18443544	18478545	35	-
2*	18365001	18395001	30	0.43 ± 0.063	4.93 ± 0.909	-	18498547	18528547	30	-

*Bidirectional introgression: regions with evidence of introgression from green junglefowl into domestic chicken and from domestic chicken into grey junglefowl.

Introgression between the Ceylon junglefowl and domestic chicken

A total of 187 windows (0.11%) out of 181,489 are above the significant genome threshold of *Zfd* ≥ 6.0 of the ABBA/BABA test. They defined 49 candidate regions of introgression between the Ceylon junglefowl and domestic chicken. No significant window was detected on chromosomes 6, 8, 9, 10, 12, 13, 15, 16, 18, 20, 21, 26, 27 and 28 (**Figure 5**). Details regarding the introgressed regions are provided at **Table 1C, 1D, supplementary Tables S12A** and **S12B**.

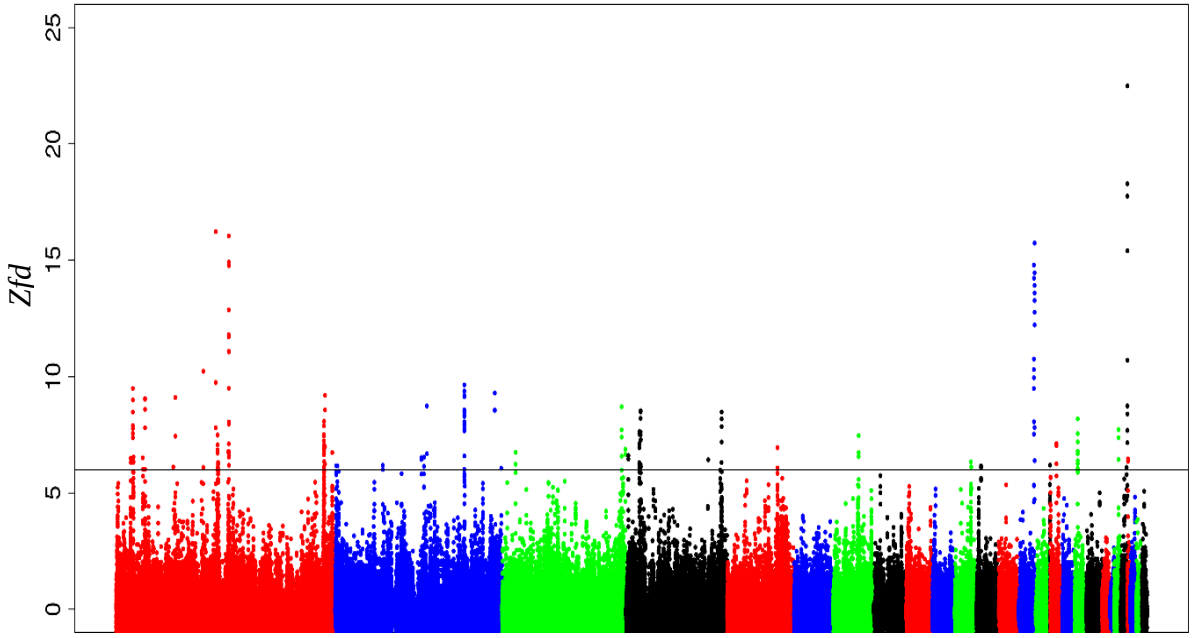


Figure 5. ABBA/BABA introgression plots. Introgression between domestic chicken and Ceylon junglefowl. The *Zfd* is the transformed value obtained from the Martin *fd* statistics. The horizontal line corresponds to the *Zfd* score of 6.

From a total number of 49 regions identified as candidate region of introgression of Ceylon junglefowl into domestic chicken, only 15 were supported by the haplotype trees (**Table 1C**). They are all found on chromosome 1 with the exception of one region on chromosome 4. Except for one candidate region, they also show introgression of domestic chicken into grey junglefowl (**Table 1C**). An example of Ceylon junglefowl introgression into domestic chicken is presented at **Figure 6**. The other haplotype trees that correspond to introgression from Ceylon junglefowl into domestic chicken are presented at **supplementary Figures S9.1 – 9.14.** Haplotype trees showing bi-directional introgression are shown at **supplementary Figures S9.1, S9.3 – S9.14**.

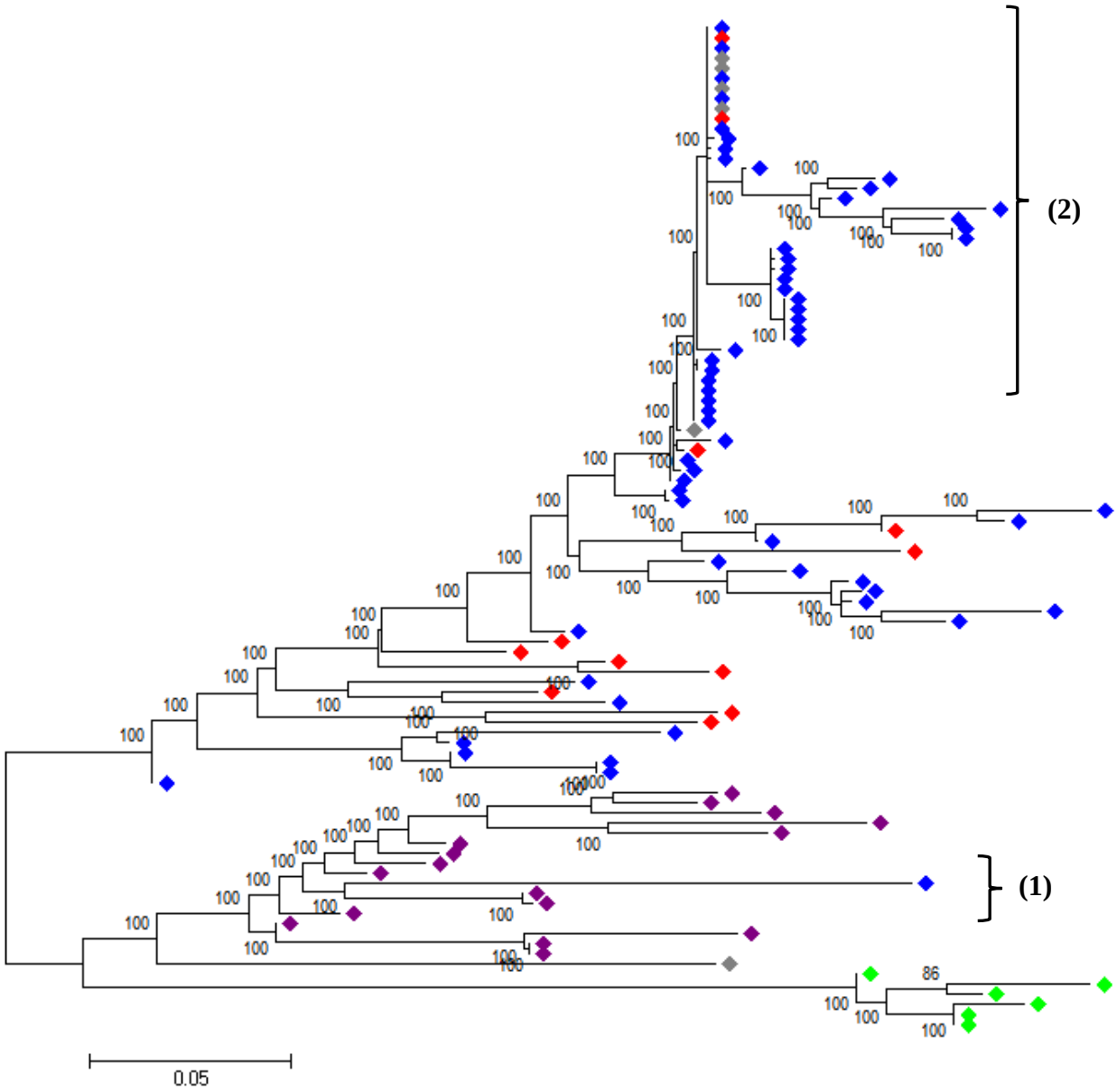


Figure 6. Haplotype based maximum likelihood tree illustrating introgression of Ceylon junglefowl into domestic chicken. ABBA – BABA test Ceylon junglefowl – domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. 35 kb candidate region of introgression, position 147120001 to 147155001 bp on chromosome 1. **(1)** Introgression from Ceylon junglefowl to domestic chicken sample (Sri Lanka16). **(2)** Introgression of domestic chicken into grey junglefowl.

A 30-kb region on chromosome 1 (15525001 to 15555001 bp) shows evidence of three distinct introgression events: (i) Ceylon junglefowl into domestic chicken, likely followed by (ii) introgression of domestic chicken into grey junglefowl, and (iii) introgression from domestic chicken into grey junglefowl (**supplementary Figure S9.1**). Also, a 35-kb length region on chromosome 1 (25800001 – 25835001 bp), with *Zfd* score of 8.23 ±

3362 0.559, shows introgression of both Ceylon and grey junglefowl into domestic chicken
3363 (**supplementary Figure S9.2, Table 1C**).

Table 1C. ABBA – BABA test Ceylon junglefowl – domestic chicken, introgressed Ceylon junglefowl regions into domestic chicken (n = 15). The **Ggal** is the reference genome for 4 or 5 version and **Ln4** and **Ln5** are the length in kilobase for the **Ggal4** and **5** coordinates respectively. **Ggal4_μ(D)** and **Ggal4_ μ(*Zfd*)** are mean of D and *Zfd* statistics based on *Galgal* 4.0 reference genome respectively.

Chr	Ggal4_start	Ggal4_stop	Ln4	Ggal4_μ(D)	Ggal4_ μ(*Zfd*)	Ggal4_Genes	Ggal5_start	Ggal5_stop	Ln5	Ggal5_Genes
1*	15525001	15555000	30	0.39 ± 0.009	6.12 ± 0.117	*TRMU, CELSR1*	15726206	15756207	30	*TRMU, CELSR1*
1**	25800001	25835000	35	0.56 ± 0.037	8.23 ± 0.559	-	25862538	25897536	35	-
1*	145735001	145810000	75	0.38 ± 0.061	4.63 ± 0.647	*DCT, GPC6*	146504328	146579330	75	*DCT, GPC6*
1*	146075001	146095000	20	0.65	5.9	*GPC6*	146844190	146864187	20	*GPC6*
1*	146175001	146195000	20	0.36	7.06	*GPC6*	146944181	146964179	20	*GPC6*
1*	146225001	146245000	20	0.41	6.17	*GPC6*	146994182	147014701	21	*GPC6*
1*	146460001	146485000	25	0.40 ± 0.0510	4.36 ± 0.094	-	147230082	147255074	25	-
1*	146505001	146525000	20	0.35	4.04	-	147274995	147294994	20	-
1*	146705001	146730000	25	0.52 ± 0.0008	6.64 ± 0.025	-	147474416	147499417	25	-
1*	146765001	146805000	40	0.48 ± 0.032	4.17 ± 0.116	-	147534418	147574455	40	-
1*	146965001	147005000	40	0.719 ± 0.029	6.742 ± 0.50	*GPC5*	147734469	147774463	40	*GPC5*
1*	147060001	147080000	20	0.37	6.6	*GPC5*	147830121	147850118	20	*GPC5*
1*	147605001	147635000	30	0.52 ± 0.007	4.84 ± 0.075	-	148375495	148405492	30	*ENSGALG00000038439 (LincRNA)*
1*	147120001	147155000	35	0.46 ± 0.054	8.66 ± 0.714	*GPC5*	147890122	147925134	35	*GPC5*
4*	23965001	23985000	20	0.45	4.16	*SPOCK3*	24691925	24711924	20	*SPOCK3*

*Bidirectional introgression, from Ceylon junglefowl into domestic chicken and from domestic chicken into grey junglefowl. ** Bidirectional introgression, from Ceylon junglefowl into domestic chicken and from grey junglefowl into domestic chicken

Haplotype tree analyses also show that 15 out of the 49 of these candidate regions correspond to introgression of grey junglefowl into domestic chicken and not Ceylon junglefowl into domestic chicken (**Table 1D**). These genome regions did not reach our significant threshold in our domestic chicken - grey junglefowl introgression analysis, and therefore were not considered previously. Three of these regions also show introgression of domestic chicken into grey junglefowl. An example of grey junglefowl introgression into domestic chicken in this category is presented at **Figure 7** and other haplotype trees are presented at **supplementary Figure S10.1 – S10.14.** Haplotype trees showing bidirectional introgression are shown at **supplementary Figure S10.2, S10.10** and **Figure 7**.

Three trees display evidence for introgression from domestic chicken into Ceylon junglefowl; a 20-kb region on chromosome 5 which shows bidirectional introgression from domestic into one haplotype of Ceylon junglefowl and then grey junglefowl into domestic chicken (**Figure 7**), a 30-kb region on chromosome 2 with two identical Ceylon junglefowl haplotypes at the root of the domestic red junglefowl haplotypes (**supplementary Figure S11.1**), and a 70-kb region (6120001 - 6190001 bp) on chromosome 24 (including the *BCDO2* locus, **supplementary Figure S11.2**).

For 17 regions, the haplotype analysis could not resolve the direction of introgression (**supplementary Figures S12.1 – S12.17** and **Supplementary Table S12B)**.

The size of the regions of introgression of Ceylon junglefowl into domestic chicken were then further estimated by building-up the haplotype trees of the adjacent window to the significant ones. It reduces the number of Ceylon junglefowl introgression into domestic from 15 to four (55 kb, 65 kb, 75 kb and 4495 kb) (**Table 1E**). The Ceylon junglefowl introgressed regions are generally small with the longest, a 4495-kb length region found on chromosome 1 (position: 24985001 – 29480001 bp) after the regions were collapsed from 15 to four (**Table 1C** and **1E**). The average fragment size is 30 kb and the median size 25 kb (**Table 1C**).

Ceylon junglefowl introgressed regions were found mainly in Sri Lanka domestic chicken (n = 4), followed by the chicken from Ethiopia (Horro (n = 2)) and one region each for Ethiopian Jarso and Saudi Arabian chickens. The frequencies of Ceylon junglefowl introgressed haplotypes varied across population. The highest within population for

average proportion of introgressed haplotypes were found in Ethiopia (Horro (18.8%) and Jarso (12.5%), followed by Saudi Arabia (5%) and then Sri Lanka (4.5%) (**supplementary Table S13B)**. None of the Ceylon junglefowl specific introgressed regions were found in the red junglefowl and CC®P birds (**supplementary Table S13B**).

The grey junglefowl introgressed regions have sizes that range from 20 kb to 115 kb (**Table 1D**). The average fragment size is 37 kb and the median size is 30 kb. The grey introgressed regions were mainly found in domestic chicken from Ethiopia (Horro (n = 14) and Jarso (n = 13)), followed by Sri Lanka (n = 13), and then Saudi Arabia (n = 11) (**supplementary Table S13A**). The frequency of the introgressed haplotypes is highest in domestic chicken from Ethiopia (Jarso (43.8%) and Horro (39.9%)), followed by Saudi Arabia (31.8%) and then Sri Lanka (15.4%) (**supplementary Table S13A**).

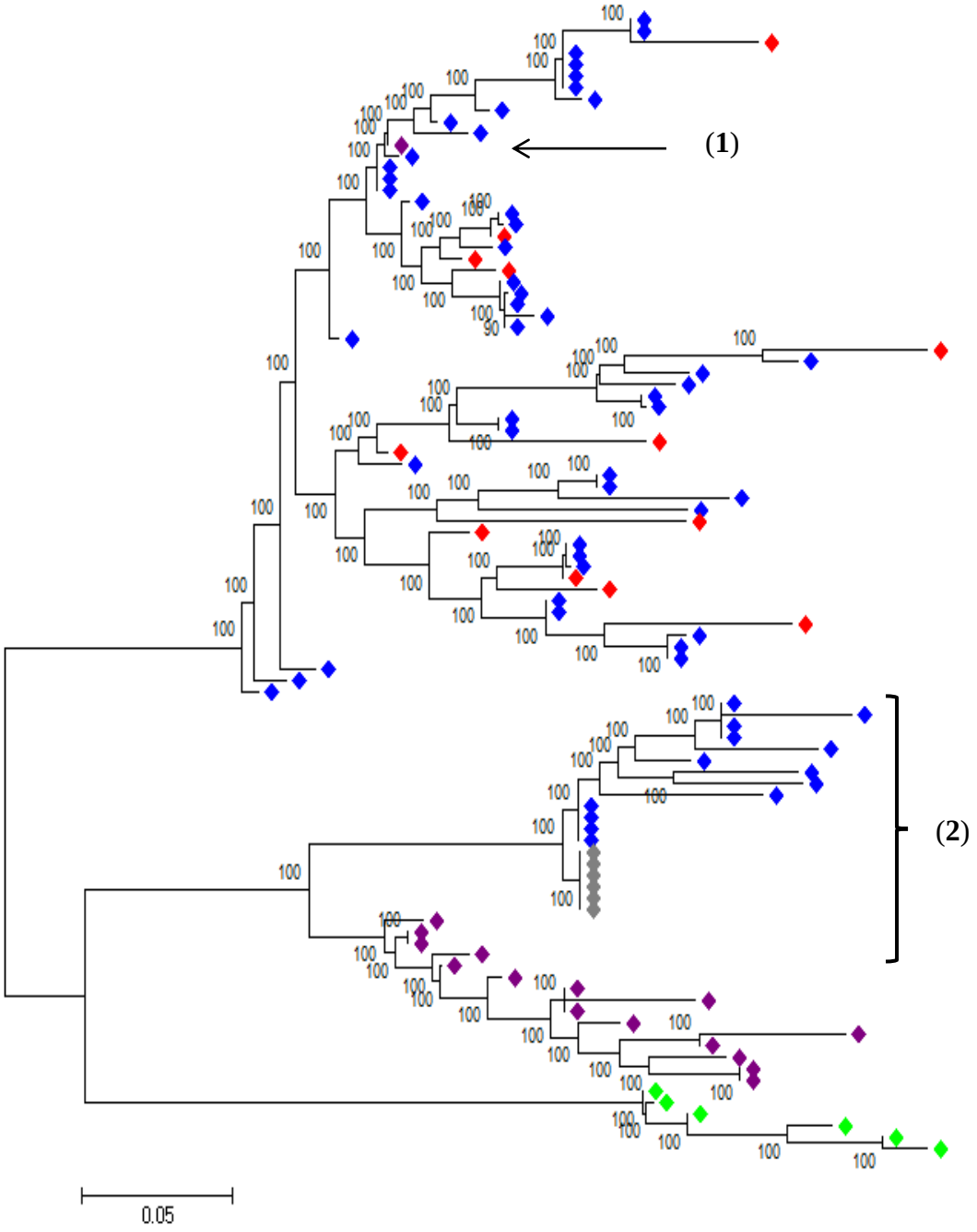


Figure 7. Haplotype based maximum likelihood tree illustrating introgression of domestic chicken into Ceylon junglefowl and grey junglefowl into domestic chicken. ABBA – BABA test Ceylon junglefowl – domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The region is on chromosome 5 and it is 20 kb in length (44955001 to 44975001 bp). **(1)** introgression of domestic chicken into Ceylon junglefowl, **(2)** the same region also shows introgression of grey junglefowl haplotypes introgression into domestic chicken.

Table 1D. ABBA – BABA test Ceylon junglefowl - domestic chicken, introgressed grey junglefowl regions into domestic chicken (n = 15). The **Ggal** is the reference genome for 4 or 5 version and **Ln4** and **Ln5** are the length in kilobase for the **Ggal4** and **5** coordinates respectively. **Ggal4_μ(D)** and **Ggal4_ μ(*Zfd*)** are mean of D and *Zfd* statistics based on *Galgal* 4.0 reference genome respectively

Chr	Ggal4_start	Ggal4_stop	Ln4	Ggal4_μ(D)	Ggal4_ μ(*Zfd*)	Ggal4_Genes	Ggal5_start	Ggal5_stop	Ln5	Ggal5_Genes
1	13070001	13095000	25	0.44 ± 0.07	6.11 ± 0.120	*RELN*	13259981	13284978	25	*RELN*
1*	14950001	15005000	55	0.50 ± 0.24	7.89 ± 0.89	*SLC2A13*	15151020	15206078	55	*SLC2A13*
1	52835001	52865000	30	0.81 ± 0.045	4.37 ± 0.304	-	53039791	53069789	30	-
1	90490001	90530000	40	0.38 ± 0.019	6.71 ± 0.323	*EPHA6*	91116693	91156693	40	*EPHA6*
1	192750001	192770000	20	0.4	6.42	*GUCY2F*	193614735	193634732	20	*GUCY2F*
2	79185001	79205000	20	0.46	6.23	*MTRR, FASTKD3, C5orf49*	79463745	79483773	20	*MTRR, FASTKD3, C5orf49*
2	115180001	115295000	115	0.67 ± 0.134	3.90 ± 0.930	*CPA6*	115882710	115997246	115	*CPA6*
3	106610001	106645001	35	0.41 ± 0.048	7.25 ± 0.838	*ENSGALT00000042 806*	107439757	107474757	35	*XKR6*
4*	2195001	2225000	30	0.31 ± 0.029	6.26 ± 0.082	*OGT, TAF1*	2231066	2261063	30	*OGT, TAF1*
4	12010001	12060000	50	0.53 ± 0.089	6.78 ± 0.507	*ENSGALG00000007 656*	12103683	12153680	50	*ENSGALG0000000765 6*
4	13130001	13180000	50	0.43 ± 0.033	7.66 ± 0.520	*ALG13*	13222088	13272086	50	*ALG13, ENSGALG0000004454 2 (LincRNA)*
4	13265001	13285000	20	0.59	6.97	*CAPN6*	13357087	13377083	20	*ENSGALG0000004542 6 (Pseudogene)*
5*	44955001	44975000	20	0.721	6.63	*SERPINA12*	45769256	45789253	20	-
7	22795001	22825000	30	0.48 ± 0.046	6.61 ± 0.454	*IGFBP2*	23420987	23450928	30	*IGFBP2*
11	14570001	14590000	20	0.45	6.05	*DYNLRB2*	15311635	15331663	20	*DYNLRB2*

*Bidirectional introgression, from grey junglefowl into domestic chicken and from domestic chicken into grey junglefowl.

Table 1E. Ceylon junglefowl introgressed regions into domestic chicken. It was based on *Galgal* 4.0. The Old_start and stop are based on the significant regions from ABBA-BABA (**Table 1D**) and the New_start and stop were obtained after building-up the haplotype trees of the adjacent window to the significant ones.

Chr	Old_start	Old_stop	kb	New_start	New_stop	kb
1	15525001	15555000	30	15515001	15580001	65
1	25800001	25835000	35	24985001	29480001	4495
1	145735001	145810000	75	145735001	147155001	75
4	23965001	23985000	20	23930001	23985000	55

Topology weighting estimation

I complemented the introgressed signal detected from the ABBA/BABA test by estimating the weight for each topology across the five taxa, assuming the total weight for the fifteen possible topologies equal 1 (**Figure 8A**). I observe slight deviation in the sum (≥ 1 ≤) of all topologies due to approximation. An estimated genome wide level of 1.24% weight topology for the three topologies (topo1, topo2, topo3) showing sister relationship between domestic chicken and Ceylon junglefowl was obtained, 80.30% for the topologies (topo4, topo5, topo6) showing sister relationship between domestic chicken and red junglefowl, and 2.00% for the topologies (topo7, topo8, topo9) showing sister relationship between domestic chicken and grey junglefowl). The other topologies (topo10, topo11, topo12, topo13, topo14, topo15) have a combined estimate percentage of 16.88% topology weight. Topology weights per chromosome are presented at **supplementary Table S14**.

A visualized example of the result is presented in detail for chromosome 24 in **Figure 8B – 8C,** which includes the yellow skin *BCDO2* locus. The most consistent topologies, which carry the combined weight estimate of 70.98% are topo4 (purple), topo5 (black) and topo6 (dark green). It illustrates the sister relationship between domestic chicken and red junglefowl (**Figure 8A**). Other weights estimation on the chromosome 24 genome are 3.74% for topo1, topo2 and topo3 (Ceylon and domestic), 4.95% for topo7, topo8, topo9 (grey and domestic). **supplementary Table S14** includes the values for the observed percentage of all sub-trees that matches each taxon topology.

I further zoomed into the region of the 30-kb encoding the yellow skin *BCDO2* allele (see **Figure 8C**). I observed a mixed species signal with three most prominent topologies, topo1 (pink colour), topo2 (blue colour) and topo7 (light green colour). At this region, the weights estimation for the three topologies showing sister relationship between domestic chicken and grey junglefowl is 29.3%, while the weights estimation for the three topologies showing sister relationship between domestic chicken and Ceylon junglefowl is 23.1%.

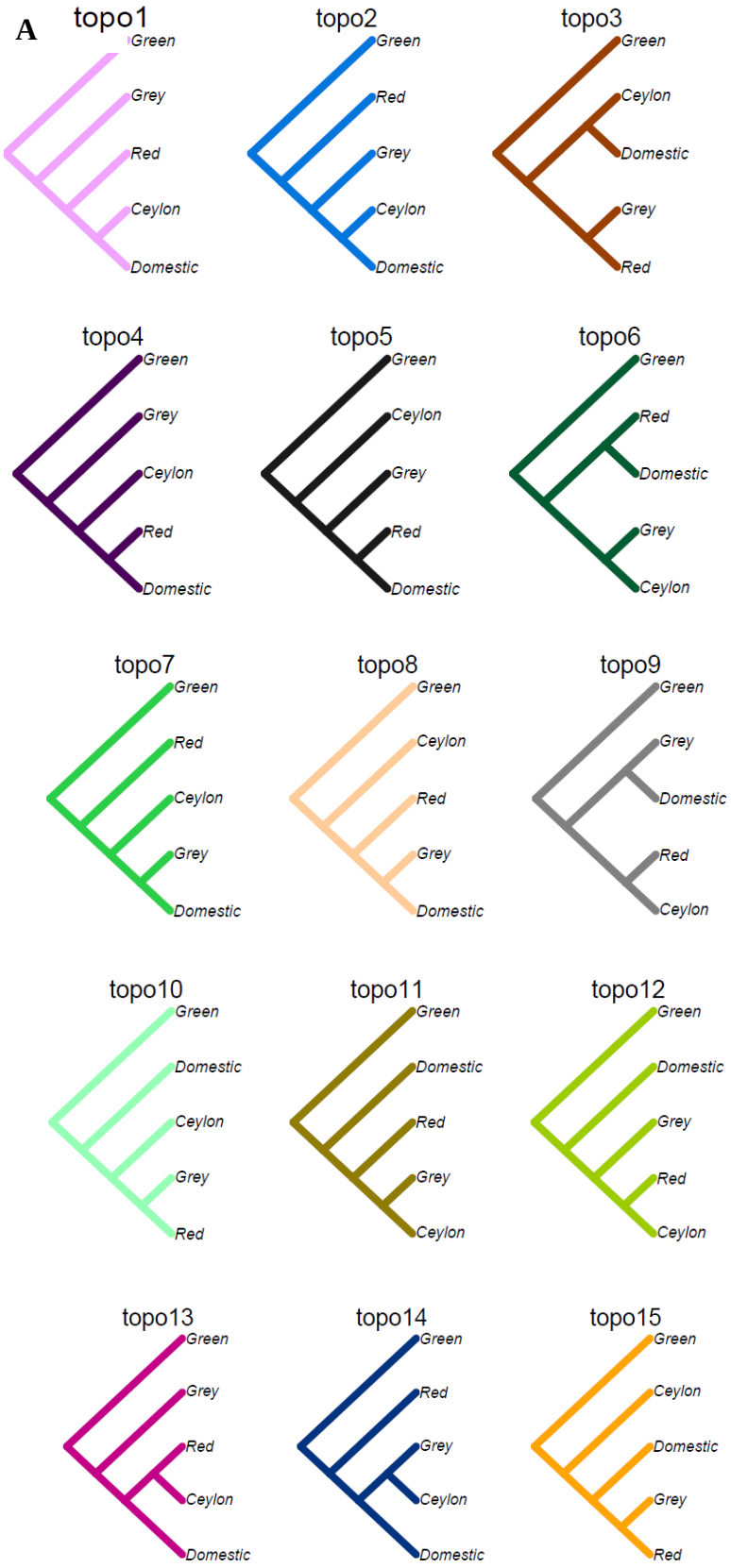


Figure 8A. The 15 possible different topologies obtained from the autosomal topology weighting analysis

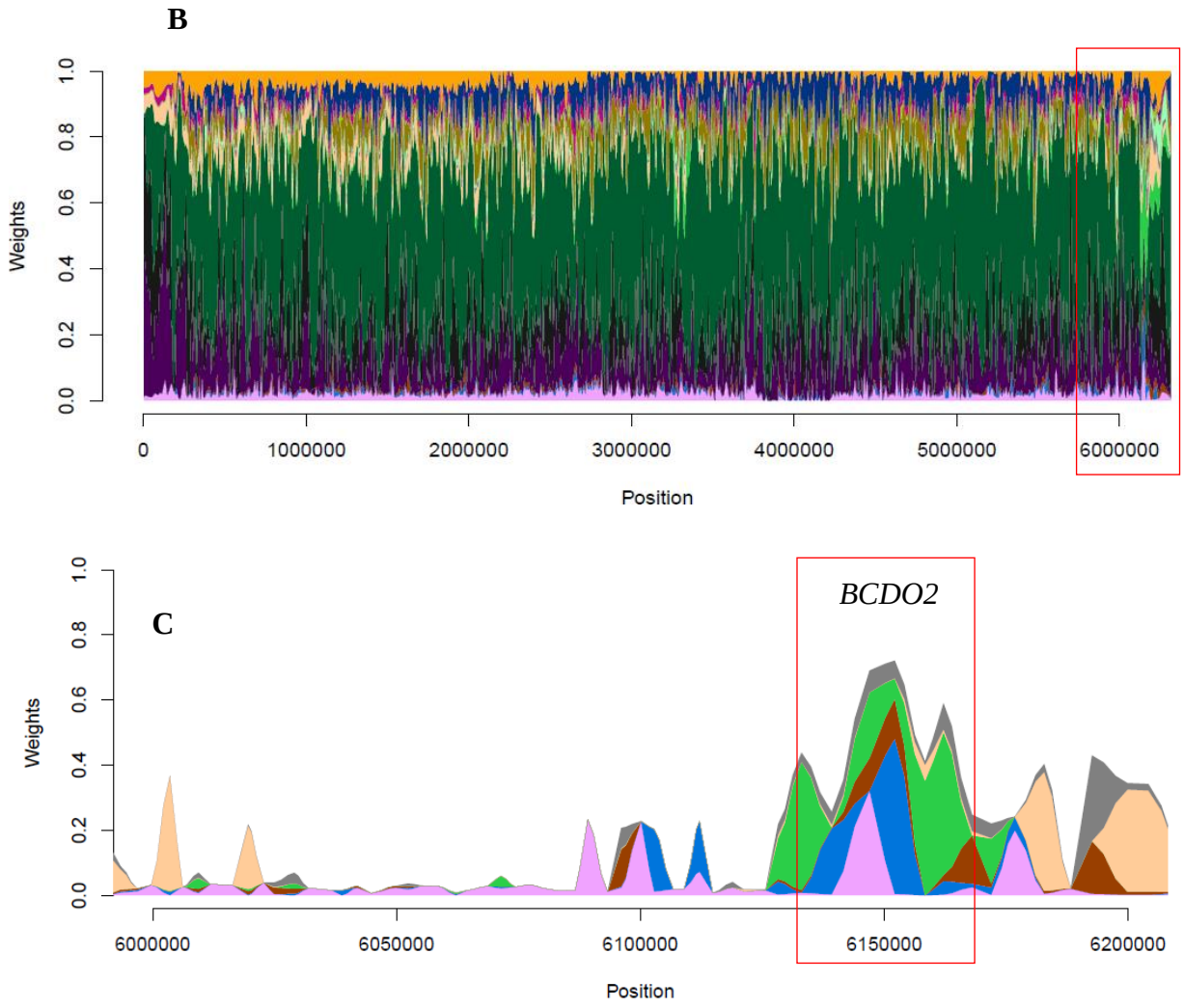


Figure 8. Topology Weighting Analysis. (**B**) Weighting plots for the 15 topologies for chromosome 24 plotted with Loess span = 0.005. (**C**) Zoomed region of 30 kb harbouring the *BCDO2* locus (delimited by the red square in A) plotted with reduced Loess span = 0.0005

Genes function analyses at the introgressed regions (*Galgal* 4.0)

Introgression of grey junglefowl into domestic chicken

For the gene ontology enrichment analysis, I combined all genes within or overlapping the grey junglefowl introgressed regions into domestic chicken, whether these were identified following the grey junglefowl – domestic chicken or the Ceylon junglefowl – domestic ABBA/BABA analysis (**Table 1A** or **1D)**. It includes a total of 39 candidate genes, from which one GO terms was identifies as significant, "protein kinase activity" (4 genes: *FASTKD3, RELN, EPHA6* and *GUCY2F*) (**supplementary Table S15**).

Introgression of Ceylon junglefowl into domestic chicken

Six candidate genes were found in the introgressed regions of Ceylon junglefowl into domestic chicken. GO analysis did not identify functional annotation enrichment.

Introgression of domestic chicken into grey junglefowl

From the total of 195 genes present in introgressed regions from domestic chicken into grey junglefowl, 27 GO terms and one KEGG pathway were identified (**supplementary Table S16)**. Some of the GO terms are: "lymph vessel development" (four genes), "nervous system development" (24 genes), "kinase activity" (13 genes), "cardiocyte differentiation" (four genes), "lens fiber cell morphogenesis" (two genes), "cytosol" (16 genes), "cell-cell adhesion" (three genes), "growth factor receptor binding" (three genes), "venous blood vessel morphogenesis" (two genes) and "cell chemotaxis" (three genes). The only described KEGG pathway is the “Erbb signaling pathway” (six genes).

Overlaps between introgressed region and domestic chicken QTLs

I compared the candidate regions introgressed in the domestic chicken from both the grey and Ceylon junglefowls with the chicken QTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/GG/index). I found several overlaps associated with body weight, pectoralis major weight, fear-stand latency, abdominal fat weight, shank length, egg shell stiffness, egg weight, residual feed intake, insulin-like growth factor level etc (see **supplementary Table S17A - S17D** for details).

Overlaps between introgressed regions and signature of selection

To investigate if any candidate introgressed region is under positive selection in domestic chicken and red junglefowl, I compared each of the grey and Ceylon junglefowls introgressed region to those detected from signature of selection in indigenous domestic chicken/red junglefowl in chapter 2.

Combining all regions corresponding to introgression of grey junglefowl into domestic chicken (**Tables 1A** and **1D**), three candidate introgressed regions overlapped with regions under positive selection (**Table 2**). These include two regions on chromosome 1 positively selected in red junglefowl and one region on chromosome 24 positively selected in Ethiopian chicken.

For the Ceylon junglefowl introgressed region into domestic chicken (**Table 1C**), an overlap was found only on chromosome 1 (**Table 2**) for which selection signature was present in Sri Lanka domestic chicken. This region is not identified as being under selection in Ceylon junglefowl (chapter 3).

These introgressed regions overlapping with regions selected in domestic and/or red junglefowl represent candidate adaptive introgressed regions.

Table 2. Candidate introgressed regions from grey and Ceylon junglefowls that overlapped with signature of selection regions in domestic chicken/red junglefowl. This table is based on *Galgal* 4.0 reference genome.

Chr	Ggal4_start	Ggal4_stop	Introgressed origin	Ggal4_gene	Ethiopian	Saudi Arabian	Sri Lankan	Red junglefowl
1	146965000	147005000	Ceylon junglefowl	*GPC5*			✓	
1	100195000	100240001	Grey junglefowl	*NCAM2*				✓
1	141700000	141755000	Grey junglefowl	-				✓
24	6130000	6160000	Grey junglefowl	*PTS, BCDO2*	✓			

Discussion

This study is the first to assess the autosomal genome ancestry of indigenous village domestic chicken using full genome re-sequencing data. I investigated the relative contribution of the red junglefowl *G. gallus*, grey junglefowl *G. sonneratii* and Ceylon junglefowl *G. lafayettii* to the autosomal gene pool of the domestic chicken. Our results clearly show that the main ancestor of the domestic chicken is the red junglefowl *G. gallus,* in agreement with Charles Darwin's initial suggestion (Darwin, 1875). Indeed, both shared the largest number of newly identified SNPs with low genetic differentiation between the two. Autosomal and mitochondrial genetic relationships include domestic chicken and the red junglefowl within the same lineage or with the red junglefowl at the root of the tree. Our mitochondrial as well as our autosomal relationship analysis support also a common ancestry for the grey and Ceylon junglefowl. These two *Gallus* species are more closely related to each other than to the red and the green junglefowls.

At the autosomal genome level, an overall average proportion of 75.8% admixture (D statistic) and 80.3% (topology weighting) was estimated for the red junglefowl and domestic chicken comparison, a 4% average genome admixture (D statistics) and 2% (topology weighting) were obtained for the grey junglefowl - domestic chicken comparison, and 1.1% (D statistics) and 1.24% (topology weighting) observed for the Ceylon junglefowl – domestic chicken comparison. It should be noted that our red junglefowl samples did not include all subspecies, particularly *G. g. murghi* from the Indian subcontinent and from the Indonesian islands *G. g. bankiva*. The autosomal level of genetic differentiation of *G. g. murghi* compared to other red junglefowl subspecies remains unknown. However, considering the geographic distribution of the domestic chicken population studied (Ethiopia, Saudi Arabia, and Sri Lanka) close to the Indian subcontinent, as well as mitochondrial DNA evidences (Silva *et al.*, 2009; Mwacharo *et al.*, 2011), a *G. g. murghi* ancestry is likely. The possible contribution to the domestic chicken gene pool of the Javan red junglefowl *G. g. bankiva* is assessed in chapter 5.

Interestingly, I do find here also one evidence of introgression from the green junglefowl *G. varius* into domestic chicken. The green junglefowl is the most distantly related species from the domestic chicken but it is frequently crossed with the domestic from Java to produce a hybrid "bekisar" (Delacour, 1977). Our results therefore indicate that the green

junglefowl has also contributed to the gene pool of domestic chicken. Genome sequence of an outgroup species from the genus *Gallus*, e.g the common pheasant *Phasianus colchicus* is required for further investigation regarding the genome contribution of the green junglefowl to the genetic make-up of the domestic chicken.

Most of the introgression events detected here were from domestic chicken into grey junglefowl, followed by the introgression events from grey junglefowl into domestic chicken, Ceylon junglefowl into domestic chicken and domestic chicken into Ceylon junglefowl. In particular, haplotype trees analysis identifies 3 to 4 more introgressed regions from grey junglefowl into domestic than Ceylon junglefowl introgression into the domestic birds analysed here. Since the grey and Ceylon junglefowls are sister species and are more closely related to each other than to the red junglefowl/domestic chicken clade, these admixture differences are likely not the result of reduced hybrid fitness between the types of crosses. A sampling artefact remains possible as the number of domestic chicken populations examined was small. However, I believe that these different introgression levels might rather be explained by the geographic distribution of the grey and Ceylon junglefowl and by the domestication history of the red junglefowl. Indeed, the grey junglefowl geographic distribution overlaps with the red junglefowl one in India, while the latter is not present in Sri Lanka (Delacour, 1977). The northern part of the Indian subcontinent has also been proposed as a centre of domestication for the species where *G. gallus murghi* is found (Delacour, 1977; Rasmussen and Anderton, 2005; Liu *et al.*, 2006). So, the grey junglefowl introgression into domestic chicken, might have originated from the admixture already present in the wild red junglefowl *G. gallus murghi* prior domestication and/or from the long history of sympatry between domestic chicken and grey junglefowl following domestication. At the opposite, the presence of introgressed Ceylon junglefowl regions into domestic chicken could only have followed the arrival of domestic chicken in Sri Lanka, arrival only possible after the domestication of the species on the Indian subcontinent.

Our results at autosomal level combined with mitochondrial DNA information support rather a male mediated process of introgression of grey junglefowl into domestic chicken than a female mediated one. At the opposite, our captive grey junglefowls from the CC®P project and two of the grey junglefowl mtDNA sequence downloaded from genebank illustrates a successful female mediated introgression process of domestic – red

junglefowl into grey junglefowl. Interestingly in the former, the three captive grey junglefowl sequences were positioned at the roots of the domestic – red junglefowl lineages and therefore they do represent haplotypes not present in the domestic – red junglefowl samples analysed here. It is tempting to hypothesise here that they might be the legacy of introgression from *G. g. murghi*, the red junglefowl subspecies present on the Indian subcontinent, into grey junglefowl. However, this requires confirmation.

Crosses between the red junglefowl *G. g. murghi* and grey junglefowl are believed to occur frequently in the wild in their zones of contact (Delacour, 1977). While in captivity, fertile hybrids have been obtained between male grey junglefowl and female domestic chicken (Danforth, 1958; Morejohn, 1968). The introgression observed here in the three captive held grey junglefowl examined could have therefore resulted from natural mating between male or female red junglefowl/domestic chicken with grey junglefowl as well as of male grey junglefowl into domestic hen. Alternatively, they could have occurred during breeding with the captive grey junglefowl. Interestingly, amongst the introgressed regions in the grey junglefowl, I do find several regions including genes (e.g. *RAB28* on an introgressed 380 kb fragment on chromosome 4) previously reported as candidate signature of chicken domestication (Rubin *et al.*, 2010). It supports that some of these regions are the legacy of genetic introgression of domestic chicken, rather than red junglefowl, into grey junglefowl.

None of the Ceylon junglefowl examined here showed a domestic/red junglefowl mitochondrial DNA but I do reveal three cases of likely introgression of domestic chicken into Ceylon junglefowl (**Figure 7, supplementary Figure S11.1** and **S11.2**). Ceylon junglefowl has been reported to reproduce freely in captivity with both *G. gallus* and *G. sonneratii* and to produce fertile hybrids (Delacour, 1977). There are also anecdotal evidences of a human mediated crosses between male Ceylon junglefowl with domestic hen in Sri Lanka (Pradeepa Silva personal communication) for increasing the vigour of cockfighting chicken.

Interestingly, I do find evidence of grey junglefowl introgression in all domestic chicken population examined. This result clearly supports the Indian subcontinent as a major centre of origin and dispersion of domestic chicken for Ethiopia, Saudi Arabia, Sri Lanka and Europe. But more surprisingly, I do find also evidence of grey junglefowl

introgression into three of the wild red junglefowls examined here. These red junglefowl samples originated from the Yunnan Province of China, well outside the geographic distribution of the grey junglefowl, which is confined to India. Most likely, these signatures of introgression are the results of crossbreeding between domestic chicken and the local wild red junglefowl. Introgression between domestic chicken and wild red junglefowl populations have been earlier reported using microsatellite loci in Vietnam (Berthouly *et al.*, 2009). By extension, these results support a movement of domestic chicken from a centre of origin on the Indian subcontinent towards East to South East Asia across the geographic range of the wild ancestor, a result supported by mitochondrial DNA analysis indicative of the presence at low frequency of a mitochondrial DNA haplogroup in East Asia which may have likely originated from the Indian subcontinent (Mwacharo *et al.*, 2011).

Introgression from Ceylon junglefowl was more commonly observed among domestic chicken in Sri Lanka than in the Ethiopian and Saudi Arabian chicken examined. It was not detected in the European chicken breeds neither in the red junglefowl. Its presence in both the populations of Ethiopia as well as in Saudi Arabia may be a legacy of the Indian Ocean maritime trading routes.

There is increasing evidence in the literature from "adaptive" cross-species introgression amongst the mammalian domesticates (Ai *et al.*, 2015) as well as in humans (Racimo *et al.*, 2015). In chicken, a previous study had revealed that the yellow skin phenotype is the consequence of the introgression of a grey junglefowl genome region, including the *BCDO2* locus into domestic chicken (Eriksson *et al.*, 2008), a phenotype favoured by some chicken breeders and now often found fixed in fancy breeds and commercial breeds (Eriksson *et al.*, 2008; Rubin *et al.*, 2010). Here, I analysed indigenous chicken population characterized by a high level of phenotypic diversity, plumage and/or morphology. Introgressed regions were never found fixed or closed to fixation in any of the indigenous village chicken population examined. The most frequent introgressed region (42.4%) was identified in Horro chicken subpopulation from Ethiopia, but often introgressed haplotypes were present at low frequency.

Annotation and functional analyses of genes present within the introgressed regions did not reveal introgression which may be directly connected to the plumage colour, pattern, and diversity of the species. Nevertheless, these introgression events may be contributing

to the functional diversity of the genome of domestic chicken. For example, among several genes that introgressed domestic chicken from grey junglefowl is *MTNR1B* associated with the regulation of circadian rhythm (Beesley *et al.*, 2015). Circadian rhythm is known to be involved in the metabolic sleep-wake cycle and pathogenesis of a disease (Arendt, 2006; Maywood *et al.*, 2006). During this process of 24-hour cycle, the retina can adjust to changes in light and dark, thereby stimulating production of signals to aid visual perception. This process is mediated by one of the introgressed gene *GUCY2F*, known to function as a photoreceptor (Stiebel-Kalish *et al.*, 2012). I also identified *GRID1* gene associated with social behaviour (Yadav *et al.*, 2012) alongside *SCAMP4* known to play role in sexual behaviour (Krebs and Pfaff, 2001).

The identification of the enriched protein kinase activity in the introgressed domestic chicken genome region originating from the grey junglefowl suggest a significant boost in the genome of domestic chicken in terms of survivability and metabolic process (Manning *et al.*, 2002). Protein kinases are one of the member of largest family of genes in eukaryotes including human, and aberrant regulation in their functions are likely to affect several cellular and biological processes including disease development (Manning *et al.*, 2002). Interestingly, I identify also another cluster of 11 overrepresented genes associated with protein kinase activity in the introgression of domestic chicken back into grey junglefowl, thus indicating that both direction of introgression may play significant role in the regulation of physiological processes in the introgressed birds (Cheng *et al.*, 2011).

Interestingly, the *GPC5* locus found within the introgressed region from Ceylon junglefowl in Sri Lanka domestic chicken is under positive selection in Sri Lanka domestic chicken (chapter 2). However, the region was only introgressed in one domestic bird. The *GPC5* locus is involved in the immune response with interaction of heparan sulfate proteoglycan (*GPC5* and *GPC6*) which is found in the extracellular matrix of the cell surface (Sarrazin *et al.*, 2011). Introgression from the Ceylon junglefowl, a locally adapted wild species may therefore confer to domestic introgressed birds some fitness advantages.

Furthermore, in the introgression of Ceylon junglefowl to domestic chicken, I identified *CELSR1*, a gene with significant function in wound healing (Caddy *et al.*, 2010) alongside

SPOCK3 implicated in disease development (Bradshaw, 2012). Similarly, I identified a *DCT* gene known to associate with oxidation-reduction and melanin biosynthetic process. *DCT* play significant role in the photo-protective pigment of the skin through the isomerisation of the orange-red dopachrome precursors which after oxidization protect the skin against the ultraviolet radiation (Ainger *et al.*, 2014). Again, these introgression may confer some selective advantages.

Conclusions

In conclusion, our study supports the polyphyletic origin of domestic chicken with major contributions from the red junglefowl but also introgression from the grey, Ceylon and green junglefowls to the gene pool of this domesticate. Unlike the yellow skin phenotype, most of the introgressed genes in domestic chicken do not show strong evidence for observable phenotype, but nevertheless, they have undoubtedly contributed to genome diversity of the domestic chicken. Farmer's visual appeal and economic factors are major selection pressures in chicken (Desta *et al.*, 2013). Introgressed regions may play a role as regulatory sequences of gene expression for which a causative link with a specific phenotype is more difficult to establish. Considering the modern geographic repartition of the non-red junglefowls, it may be expected that their genetic contribution to the domestic gene pool will vary from one geographic region to the others. This will require further investigations through the analysis of more populations in a widespread geographic distribution. If non-red junglefowl species have contributed to the gene of domestic chicken, it may be also expected that all subspecies of red junglefowl could have contributed to the gene pool of the domestic chicken, in particular *G. gallus murghi* and *G. gallus bankiva* not studied here. The latter will be the focused of chapter 5.

References

Ai, H., Fang, X., Yang, B., Huang, Z., Chen, H., Mao, L., *et al.* (2015). Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nature Genetics, 47*(3), 217-225.

Ainger, S. A., Yong, X. L., Wong, S. S., Skalamera, D., Gabrielli, B., Leonard, J. H., *et al.* (2014). *DCT* protects human melanocytic cells from UVR and ROS damage and increases cell viability. *Experimental Dermatology, 23*(12), 916-921.

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., *et al.* (2016). The ensembl gene annotation system. *The Journal of Biological Databases and Curation, 2016*. doi:10.1093/database/baw093.

Arendt, J. (2006). Melatonin and human rhythms. *Chronobiology International, 23*(1-2), 21-37.

Arnold, M. L. (1997). *Natural hybridization and evolution*: Oxford University Press.

Asmundson, V., and Lorenz, F. (1957). Hybrids of ring-necked pheasants, turkeys, and domestic fowl. *Poultry Science, 36*(6), 1323-1334.

Barilani, M., Bernard-Laurent, A., Mucci, N., Tabarroni, C., Kark, S., Garrido, J. A. P., *et al.* (2007). Hybridisation with introduced chukars (*Alectoris chukar*) threatens the gene pool integrity of native rock (*A. graeca*) and red-legged (*A. rufa*) partridge populations. *Biological Conservation, 137*(1), 57-69.

Barton, N. (2001). The role of hybridization in evolution. *Molecular Ecology, 10*(3), 551-568.

Beebe, W. (1918). *Monograph of the Pheasants* (1 ed.): New York Zoological Society.

Beesley, S., Lee, J., and Olcese, J. (2015). Circadian clock regulation of melatonin *MTNR1B* receptor expression in human myometrial smooth muscle cells. *Molecular Human Reproduction, 21*(8), 662-671.

Berthouly, C., Leroy, G., Van, T. N., Thanh, H. H., Bed'Hom, B., Nguyen, B. T., *et al.* (2009). Genetic analysis of local Vietnamese chickens provides evidence of gene flow from wild to domestic populations. *BMC Genetics, 10*(1), 1.

Bradshaw, A. D. (2012). Diverse biological functions of the *SPARC* family of proteins. *The International Journal of Biochemistry and Cell Biology, 44*(3), 480-488.

Bruford, M., Hanotte, O., Brookfield, J., and Burke, T. (1998). *Single-locus and multilocus DNA fingerprinting* (2 ed.): Hoelzel AR, IRL Press, Oxford, UK.

Caddy, J., Wilanowski, T., Darido, C., Dworkin, S., Ting, S. B., Zhao, Q., *et al.* (2010). Epidermal wound repair is regulated by the planar cell polarity signaling pathway. *Developmental Cell, 19*(1), 138-147.

Castillo, A., Marzoni, M., Pirone, A., and Romboli, I. (2012). Histological observations in testes of hybrids of *Gallus gallus* x *Phasianuns colchicus*. *Avian Biology Research, 5*(1), 21-30.

Castillo, A., Romboli, I., Marzoni, M., and Stanyon, R. (2010). Karyotype of chicken–pheasant hybrids. *Avian Biology Research, 3*(1), 17-21.

Chazara, O., Minvielle, F., Roux, D., Bed'hom, B., Feve, K., Coville, J.-L., *et al.* (2010). Evidence for introgressive hybridization of wild common quail (*Coturnix coturnix*) by domesticated Japanese quail (*Coturnix japonica*) in France. *Conservation Genetics, 11*(3), 1051-1062.

Chen, H., and Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics, 12*(1), 1.

3757 Cheng, H.-C., Qi, R. Z., Paudel, H., and Zhu, H.-J. (2011). Regulation and function of
3758 protein kinases and phosphatases. *Enzyme Research, 2011*.
3759 doi.org/10.4061/2011/794089.

3760 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., *et al.*
3761 (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.

3762 Danforth, C. (1958). *Gallus sonnerati* and the domestic fowl. *Journal of heredity, 49*(4),
3763 167-170.

3764 Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more
3765 models, new heuristics and parallel computing. *Nature Methods, 9*(8), 772-772.

3766 Darwin, C. (1875). *The variation of animals and plants under domestication* (2 ed. Vol.
3767 2): The Echo Library , Middlesex, UK.

3768 Delacour, J. (1977). *The pheasant of the world* (2 ed.): Saiga Publishing Co. Ltd. Surr
3769 GU26 GTD. England.

3770 Delaneau, O., Howie, B., Cox, A. J., Zagury, J.-F., and Marchini, J. (2013). Haplotype
3771 estimation using sequencing reads. *The American Journal of Human Genetics,*
3772 *93*(4), 687-696.

3773 Desta, T., Dessie, T., Bettridge, J., Lynch, S., Melese, K., Collins, M., *et al.* (2013).
3774 Signature of artificial selection and ecological landscape on morphological
3775 structures of Ethiopian village chickens. *Animal Genetic Resources, 52*, 17-29.

3776 Dong, L., Heckel, G., Liang, W., and Zhang, Y. (2013). Phylogeography of Silver
3777 Pheasant (*Lophura nycthemera L.*) across China: aggregate effects of refugia,
3778 introgression and riverine barriers. *Molecular Ecology, 22*(12), 3376-3390.

3779 Durand, E. Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient
3780 admixture between closely related populations. *Molecular Biology and Evolution,*
3781 *28*(8), 2239-2252.

3782 Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt,
3783 L., *et al.* (2008). Identification of the yellow skin gene reveals a hybrid origin of
3784 the domestic chicken. *PLoS Genetics, 4*(2).
3785 doi.org/10.1371/journal.pgen.1000010.

3786 Fumihito, A., Miyake, T., Sumi, S., Takada, M., Ohno, S., and Kondo, N. (1994). One
3787 subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic
3788 ancestor of all domestic breeds. *Proceedings of the National Academy of Sciences,*
3789 *91*(26), 12505-12509.

3790 Fumihito, A., Miyake, T., Takada, M., Shingu, R., Endo, T., Gojobori, T., *et al.* (1996).
3791 Monophyletic origin and unique dispersal patterns of domestic fowls.
3792 *Proceedings of the National Academy of Sciences, 93*(13), 6792-6795.

3793 Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., *et al.* (2010).
3794 A draft sequence of the Neandertal genome. *Science, 328*(5979), 710-722.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology, 52*(5), 696-704.

Hufford, M. B., Lubinksy, P., Pyhäjärvi, T., Devengenzo, M. T., Ellstrand, N. C., and Ross-Ibarra, J. (2013). The genomic signature of crop-wild introgression in maize. *PLoS Genetics, 9*(5). doi.org/10.1371/journal.pgen.1003477.

Krebs, C. J., and Pfaff, D. W. (2001). Expression of the *SCAMP-4* gene, a new member of the secretory carrier membrane protein family, is repressed by progesterone in brain regions associated with female sexual behavior. *Molecular Brain Research, 88*(1), 144-154.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution, 33*(7), 1870-1874.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.

Liu, Y.-P., Wu, G.-S., Yao, Y.-G., Miao, Y.-W., Luikart, G., Baig, M., *et al.* (2006). Multiple maternal origins of chickens: out of the Asian jungles. *Molecular Phylogenetics and Evolution, 38*(1), 12-19.

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science, 298*(5600), 1912-1934.

Martin, S. H., Davey, J. W., and Jiggins, C. D. (2015). Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Molecular Biology and Evolution, 32*(1), 244-257.

Martin, S. H., and Van Belleghem, S. M. (2017). Exploring evolutionary relationships across the genome using topology weighting. *Genetics, 205*(4). doi.org/10.1534/genetics.116.194720.

Mayr, E. (1963). *Animal species and evolution* (Vol. 797): Belknap Press of Harvard University Press Cambridge, Massachusetts.

Maywood, E. S., O'Neill, J., Wong, G. K., Reddy, A. B., and Hastings, M. H. (2006). Circadian timing in health and disease. *Progress in Brain Research, 153*, 253-269.

Mina-Vargas, A. M., McKeown, P. C., Flanagan, N. S., Debouck, D. G., Kilian, A., Hodkinson, T. R., *et al.* (2016). Origin of year-long bean (*Phaseolus dumosus Macfady, Fabaceae*) from reticulated hybridization events between multiple Phaseolus species. *Annals of Botany, 118*(5), 957-969.

Morejohn, G. V. (1968). Breakdown of isolation mechanisms in two species of captive junglefowl (*Gallus gallus* and *Gallus sonneratii*). *Evolution, 22*(3), 576-582.

Mwacharo, J., Bjørnstad, G., Mobegi, V., Nomura, K., Hanada, H., Amano, T., *et al.* (2011). Mitochondrial DNA reveals multiple introductions of domestic chicken in East Africa. *Molecular Phylogenetics and Evolution, 58*(2), 374-382.

Nishibori, M., Shimogiri, T., Hayashi, T., and Yasue, H. (2005). Molecular evidence for hybridization of species in the genus *Gallus* except for *Gallus varius*. *Animal Genetics, 36*(5), 367-375.

Olsen, M. (1960). Turkey-chicken hybrids. *Journal of Heredity, 51*, 69-73.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics, 20*(2), 289-290.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics, 38*(8), 904-909.

Racimo, F., Sankararaman, S., Nielsen, R., and Huerta-Sánchez, E. (2015). Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics, 16*(6), 359-371.

Rasmussen, P. C., and Anderton, J. C. (2005). *Birds of south Asia:* the Ripley guide (Vol. 2): Smithsonian Institution and Lynx Edicions.

Rubin, C.-J., Zody, M. C., Eriksson, J., Meadows, J. R., Sherwood, E., Webster, M. T., *et al.* (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature, 464*(7288), 587-591.

Shaklee, W. E., and Knox, C. (1954). Hybridization of the pheasant and fowl. *Journal of Heredity, 45*(4), 183-190.

Silva, P., Guan, X., Ho-Shing, O., Jones, J., Xu, J., Hui, D., *et al.* (2009). Mitochondrial DNA-based analysis of genetic variation and relatedness among Sri Lankan indigenous chickens and the Ceylon junglefowl (*Gallus lafayetti*). *Animal Genetics, 40*(1), 1-9.

Skjervold, H., and Mjelstad, H. (1992). Capercaillie-chicken hybrids. *Journal of Animal Breeding and Genetics, 109*(1-6), 149-152.

Stiebel-Kalish, H., Reich, E., Rainy, N., Vatine, G., Nisgav, Y., Tovar, A., *et al.* (2012). Gucy2f zebrafish knockdown–a model for Gucy2d-related leber congenital amaurosis. *European Journal of Human Genetics, 20*(8), 884-889.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research, 25*(24), 4876-4882.

Wang, M.-S., Li, Y., Peng, M.-S., Zhong, L., Wang, Z.-J., Li, Q.-Y., *et al.* (2015). Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. *Molecular Biology and Evolution, 32*(7), 1880-1889.

Wilcox, F., and Clark, C. E. (1961). Chicken-quail hybrids. *Journal of Heredity, 52*(4), 167-170.

Yadav, R., Gupta, S. C., Hillman, B. G., Bhatt, J. M., Stairs, D. J., and Dravid, S. M. (2012). Deletion of glutamate delta-1 receptor in mouse leads to aberrant emotional and social behaviors. *PloS One, 7*(3). doi:10.1371/journal.pone.0032969.

Chapter 5

Introgression of Javan red junglefowl *Gallus g. bankiva* to domestic chicken: Insights from autosomal whole-genome sequences

Abstract

Mitochondrial DNA sequence analyses suggest that four out of the five subspecies of red junglefowl *G. gallus* may have contributed to the gene pool of domestic chicken, with no evidence to support a maternal contribution from the Javan red junglefowl *G. gallus bankiva.* Here, I investigate using whole-genome autosomal sequence analysis, the possible introgression of Javan red junglefowl into domestic chicken using the four-taxon admixture method (ABBA/BABA). The PCA results show genetic differentiation between the Javan red junglefowl and domestic chicken – red junglefowl. Similarly, the autosomal maximum likelihood phylogenetic analysis positions the Javan red junglefowl subspecies at the root of the red junglefowl - domestic chicken lineage. A whole-genome Patterson's D statistic of 0.106 ± 0.130 was obtained, supporting genetic exchanges between the Javan red junglefowl and domestic chicken. Using a threshold of $Zfd > 6$, 64 candidate introgressed regions were identified across the genome and were further analysed. Haplotype relationship analysis identifies a minimum of 14 introgressed regions from the Javan red junglefowl into domestic chicken. Our results indicate for the first time that the Javan Red Junglefowl *G. gallus bankiva* has likely contributed substantially to the genetic pool and diversity of the domestic chicken. These introgressed regions may provide new genetic markers to study ancient dispersion of domestic chicken from the Indonesia Islands across Asia and beyond.

Introduction

Domestic chickens *Gallus gallus domesticus* represent the most diverse livestock species and are globally accepted irrespective of cultural, traditional, and religious affiliations. They serve as primary sources of protein to people and contribute substantially to nations' gross domestic product. Domestic chicken is also used as a model organism e.g. to study the developmental stages of the embryo (Bahr, 2008), infection transmission (Pasteur, 1880) and the immune system (Stern, 2005). Following human-mediated domestication events (Fumihito *et al.*, 1994; Liu *et al.*, 2006) and subsequent selection pressures, hundreds of breeds/strains/ecotypes of chicken have emerged (Hoffmann, 2009), all part of the diversity of the species. The commercial breeds have been intensively selected by humans, producing more meat and eggs than the more diverse indigenous domestic chicken populations adapted to the local production systems (Havenstein *et al.*, 2003; Muir *et al.*, 2008; Desta *et al.*, 2015).

Despite initial controversies (Darwin, 1868; Fumihito *et al.*, 1994; Fumihito *et al.*, 1996), molecular evidence is increasingly supporting multiple domestication events, of which the red junglefowl played a primary role and genetic contributions from more than one *Gallus* ancestor is also likely (Hutt, 1949; Nishibori *et al.*, 2005; Eriksson *et al.*, 2008) (see also Chapter 4).

The red junglefowl *Gallus gallus* comprises five subspecies including *Gallus gallus gallus* found in Indo-China, a geographic region within the cultural influence of India and China and *Gallus g. spadiceus* found in Burma (Myanmar) and adjacent countries (Delacour, 1977) with the former earlier proposed as the direct ancestor of all chicken breeds (Fumihito *et al.*, 1994; Fumihito *et al.*, 1996). The other three subspecies are, *Gallus g. jabouillei* endemic to Vietnam, *Gallus g. murghi* endemic to the Indian subcontinent particularly the North of India and Bangladesh and *Gallus gallus bankiva* also known as the Javan red junglefowl with a native geographic distribution in the Indonesian Islands of Sumatra (south), Java and Bali (Delacour, 1977). Improving on previous studies, Fumihito *et al.* (1996) and Liu *et al.* (2006) analysed the hypervariable region of mitochondrial sequence from a larger number of birds and together, their study suggest that with the exception of the Javan red junglefowl, all four other subspecies of red junglefowl may have contributed to the domestic chicken gene pool.

Recently, whole genome sequence analysis has been used to unravel introgression events in domestic species. It includes, for example, introgression related to production and reproduction traits in pigs (Bosse *et al.*, 2014), hybridization between yaks and cattle (Medugorac *et al.*, 2017), adaptation to altitude in dogs (Gou *et al.*, 2014) etc. Within the genus *Gallus*, there have been so far, no genome-wide analyses on introgression (except in Chapter 4 of this thesis). Here, I investigate for the first time using whole-genome autosomal sequence data, the possible contribution of the Javan red junglefowl subspecies to chicken domestic genome.

Materials and methods

Samples and DNA extraction

Details regarding 50 out of the 56 samples included in this study can be found at **supplementary table S1** of chapters 2, 3 and 4. The DNA extraction and sequencing information are described under the materials and methods section of those chapters.

For the other six samples, sequence reads of three domestic Kedu Hitam chickens and three Javan red junglefowls, both from the island of Java (Indonesia) were retrieved from Ulfah *et al.* (2016). The sequence coverage for each of the six samples was ~ 7 X (**supplementary Table S1**).

Combining all samples, the 56 genomes were grouped into six populations: domestic chicken (n = 33), red junglefowl (n = 6), Javan red junglefowl (n = 3), grey junglefowl (n = 3), Ceylon junglefowl (n = 8) and green junglefowl (n = 3). The domestic chicken was further grouped into six subpopulations: Kedu Hitam (n = 3), Horro (n = 6, Ethiopia), Jarso (n = 5, Ethiopia), Saudi Arabia (n = 5), Sri Lanka (n = 11) and Europe (Cosmopolitan Chicken Research Project (CC®P) parent n = 2 and 16th generation crossbred n = 1).

Sequence mapping and variants calling procedures

Sequence reads for each of the 56 genomes were independently mapped to the *Galgal* 4.0 reference genome using the Burrows-Wheeler Aligner (BWA) version 0.7.5a (Li and Durbin, 2010). See materials and methods section of chapter 2 for the procedures followed to generate the binary alignment map and genotype variant calling format files. Results for autosomes 1 to 28 are only reported here.

The number of SNPs per sample and for the jointly called 56 genomes were identified using VCFtools version 0.1.11 (Danecek *et al.*, 2011). For the six new genome sequences the percentage mate of read pairs mapped to the same chromosome, the mean sequence coverage, the total reads mapped, and the percentage of the genome with bases covered by at least 5, 10 and 20 reads were calculated using SAMtools version 0.1.19 (Li *et al.*, 2009). For the other 50 genomes, these statistics can be found at **supplementary Table S1** of chapters 2 and 3, and **supplementary Table S2** of chapter 4.

Genetic differentiation

Pairwise genetic differentiation F_{ST} was estimated using VCFtools-F_{ST} at a 20kb window with a 5-kb sliding step.

Population structure and samples relationships

Principal component analysis was used to visualise the genetic similarity among the populations using smartpca program in Eigenstrat version 6.0.1 (Price *et al.*, 2006). All the ~ 45 million autosomal SNPs identified across the 56 samples were used for this analysis.

To construct the evolutionary tree for the autosomal genome, I randomly subset 89,696 SNPs corresponding to 0.2 % of all jointly called autosomal SNPs in the 56 genomes. Best-fit model was predicted using jModeltest version 2.1.7 (Darriba *et al.*, 2012) and a maximum likelihood relationship tree across the samples was built using phyml version 3.0 (Guindon and Gascuel, 2003).

Introgression analysis

To detect gene flow between domestic chicken and Javan red junglefowl, I used the four-taxon admixture method (ABBA/BABA) from which the Patterson's D (Durand *et al.*, 2011) and the Martin's *fd* (Martin *et al.*, 2015) statistics were calculated. Analyses were performed across each autosome on a 20-kb window size with 5 kb sliding step. Only windows with a minimum of 100 SNPs were considered. It excluded 170 windows (**supplementary figure S1**). These two statistics detect regions having an excess of shared derived allele between any two in-groups by comparing the allelic state of the third in-group with the fourth assumed to be the ancestral out-group population (Green *et al.*, 2010; Durand *et al.*, 2011). Considering three in-groups, P_1 (red junglefowl), P_2 (domestic chicken) and P_3 (Javan red junglefowl), and an out-group O (green junglefowl), the expected phylogeny is (((*P1*, *P2*), *P3*), *O*). The admixture test can then be computed by comparing any two groups P_1 and P_3 (BABA) or P_2 and P_3 (ABBA). I defined ABBA as any given site where the derived allele 'B' is shared between the domestic chicken 'P_2' and the Javan red junglefowl 'P_3', while the red junglefowl 'P_1' shared the ancestral allele 'A' with the green junglefowl 'O'. In BABA, the red junglefowl 'P_1' shared the derived allele 'B' with 'P_3' while the domestic chicken 'P_2' shared the same ancestral state with the

outgroup '*O*'. Under the assumption of no gene flow and neutral coalescent model, both the ABBA and BABA should result in same count and D value tends to zero.

The Martin's *fd* statistics was calculated using the equation:

$$fd = \frac{S(P1,P2,P3,O)}{S(P1,PD,PD,O)}$$ ((Martin *et al.*, 2015))

Where *S* is the difference between the sums of ABBAs and BABAs. *P1*, *P2*, *P3* represents the in-groups and *O* the out-group. The *PD* is the donor population for independent sites. Since *fd* detects bi-directional introgression, the *PD* could either be *P*2 or *P*3 as in the case of ABBAs.

The values of the *fd* were then Z-transformed using the equation:

$$Zfd = \frac{fd - \overline{X}(fd)}{\sigma(fd)}$$

Where $\overline{X}$ is the mean and σ is the standard deviation of the *fd.*

The genome-wide *Zfd* score of 6.0 or more was defined arbitrarily after the examination of the Manhattan plot as the significant threshold. To define the size of the candidate introgressed regions, I included overlapping or immediately adjacent windows in the same candidate introgressed region provided each reached the significant genome-wide threshold. Introgressed regions on the same chromosomes were considered distinct if separated by at least one 20 kb window.

The genome admixture proportion was estimated by taking the average of the D statistics (Durand *et al.*, 2011). Frequency of the introgressed haplotypes in each chicken population was calculated by dividing the number of introgressed haplotypes with the total number of haplotypes in each population.

Haplotype analysis

To infer relationships between haplotypes at each introgressed region and to predict possible direction of gene flow, I follow a haplotype phylogenetic tree approach. The expectation is that introgressed regions in a population will show haplotypes closely linked to or embedded within the donor's species rather than within the recipient species haplotypes.

First, I phased our data using the read aware phasing method of shapeit (Delaneau *et al.*, 2013) that extracts directly the phase informative read from the sequence BAM file. The phasing output was then converted into a variant call format (vcf) file and subsequently formatted in Plink 1.9 with the 'beagle recode' option (https://www.cog-genomics.org/plink2), which was then used to create, using a customized script, a fasta file. The multiple sequences alignment was done using Clustalx 2.1 (Thompson *et al.*, 1997). The evolutionary model was predicted using jModeltest 2.1.7 (Darriba *et al.*, 2012) based on the Akaike information criterion (AIC). Phyml 3.0 (Guindon and Gascuel, 2003) was used to compute the approximate likelihood ratio score for each branch using the best predicted model.

Results

Sequence reads, mappings, and SNPs identification

Details regarding the sequence information for 50 out of the 56 genomes used in this study can be found at chapters 2, 3 and 4. I downloaded 79.6 Mb to 83.2 Mb of raw sequences read for the three Kedu Hitam domestic chickens and for the three Javan red junglefowls, the raw sequences read varies from 79.0 Mb to 84.0 Mb (**supplementary Table S1**). Across the six samples, more than 96% of both mates of a read pair were properly mapped to the same chromosome and more than 98% of the reads for each sample were mapped to the reference *Galgal* 4.0 genome. Compared to the 50 genomes previously analysed (see **supplementary Table S1** of chapters 2 and 3 and **supplementary Table S2** of chapter 4), the mean sequence coverage for each of the six genomes was low and ranged from 7.21 X to 7.78 X (**supplementary Table S1**). Similarly, the percentage of bases covered by at least 5 reads ranged from 76.8% to 80.9% (**supplementary Table S1**) when compared to 87.4% to 98.7% in previous chapters (see **supplementary Table S1** of chapter 2 and 3 and **supplementary Table S2** of chapter 4).

Excluding the multi-allelic sites, the total number of single nucleotide polymorphisms (SNPs) for Kedu Hitam domestic chicken ranges from 5,334,991 to 5,689,161 (**supplementary Table S1**), (5,487,603 to 6,715,400 in other domestic chicken, see **supplementary Table S1** of chapter 2). For the Javan red junglefowl, it ranges from 6,852,820 to 6,986,401 (**supplementary Table S1**), (6,098,075 to 7,147,729 in other red junglefowl subspecies, see **supplementary Table S1** of chapter 2). The proportion of

homozygous SNPs varies from 42% to 53% in Kedu Hitam (**supplementary Table S1**), (35% to 55% in other domestic chicken, see **supplementary Table S1** of chapter 2). It was around 59 % in Javan red junglefowl samples (**supplementary Table S1**), (33.5% to 43.9% detected in other red junglefowls subspecies, see **supplementary Table S1** of chapter two).

Pairwise genetic differentiation

As in the chapter 4, I measured the level of pairwise genetic distance (F_{ST}) between the Javan red junglefowl and each of the other species/populations reported here. Among the domestic chicken populations, the lowest differentiation was with Sri Lankan (F_{ST} = 0.342), followed by the fancy birds (F_{ST} = 0.356), Kedu Hitam (F_{ST} = 0.364), while the highest F_{ST} was found with the Ethiopian Horro, (F_{ST} = 0.393) and Jarso (F_{ST} = 0.437) (**Table 1**).

Compared to the junglefowl species, the lowest genetic distance was observed with the red junglefowl (F_{ST} = 0.274), then the Ceylon junglefowl (F_{ST} = 0.668), grey junglefowl (F_{ST} = 0.695) and the green junglefowl (F_{ST} = 0.789) (**Table 1**).

Population and species relationships

I assessed the overall genetic variation among the populations using the ~ 45 million autosomal SNPs identified across the 56 genomes using PC analysis. While the domestic chicken and red junglefowl still clustered tightly together, PC1 clearly separates from them the Javan red junglefowl subspecies (**Figure 1**).

For the phylogenetic relationships, I selected randomly, a 0.2% subset of the ~ 45 million SNPs across the autosomes (**Figure 2**). The maximum likelihood unrooted evolutionary trees show the Javan red junglefowl at the root of the red junglefowl and domestic chicken lineage. Other observed relationships are consistent with the findings discussed in chapter 4.

Table 1. Pairwise F_{ST} values between populations. The domestic chicken populations are **Kedu Hitam, Ethiopian** (Horro and Jarso), **Saudi Arabian, Sri Lankan,** and **Fancy birds**. The junglefowl species are **Red, Javan** (subspecies of red junglefowl)**, Grey, Ceylon** and **Green junglefowls**.

Pairwise F_{ST}	**Kedu Hitam**	**EthiopianJarso**	**EthiopianHorro**	**Saudi Arabia**	**Sri Lanka**	**Fancy**	**Red**	**Javan**	**Grey**	**Ceylon**	**Green**
Kedu Hitam	0										
EthiopianJarso	0.239	0									
EthiopianHorro	0.194	0.102	0								
Saudi Arabia	0.174	0.129	0.092	0							
Sri Lanka	0.119	0.138	0.089	0.082	0						
Fancy	0.133	0.176	0.126	0.097	0.07	0					
Red	0.072	0.148	0.113	0.095	0.062	0.063	0				
Javan	**0.364**	**0.437**	**0.393**	**0.378**	**0.342**	**0.356**	**0.274**	0			
Grey	0.691	0.670	0.670	0.666	0.621	0.676	0.596	**0.695**	0		
Ceylon	0.669	0.682	0.666	0.662	0.633	0.662	0.615	**0.668**	0.625	0	
Green	0.790	0.786	0.761	0.762	0.714	0.779	0.699	**0.789**	0.856	0.776	0

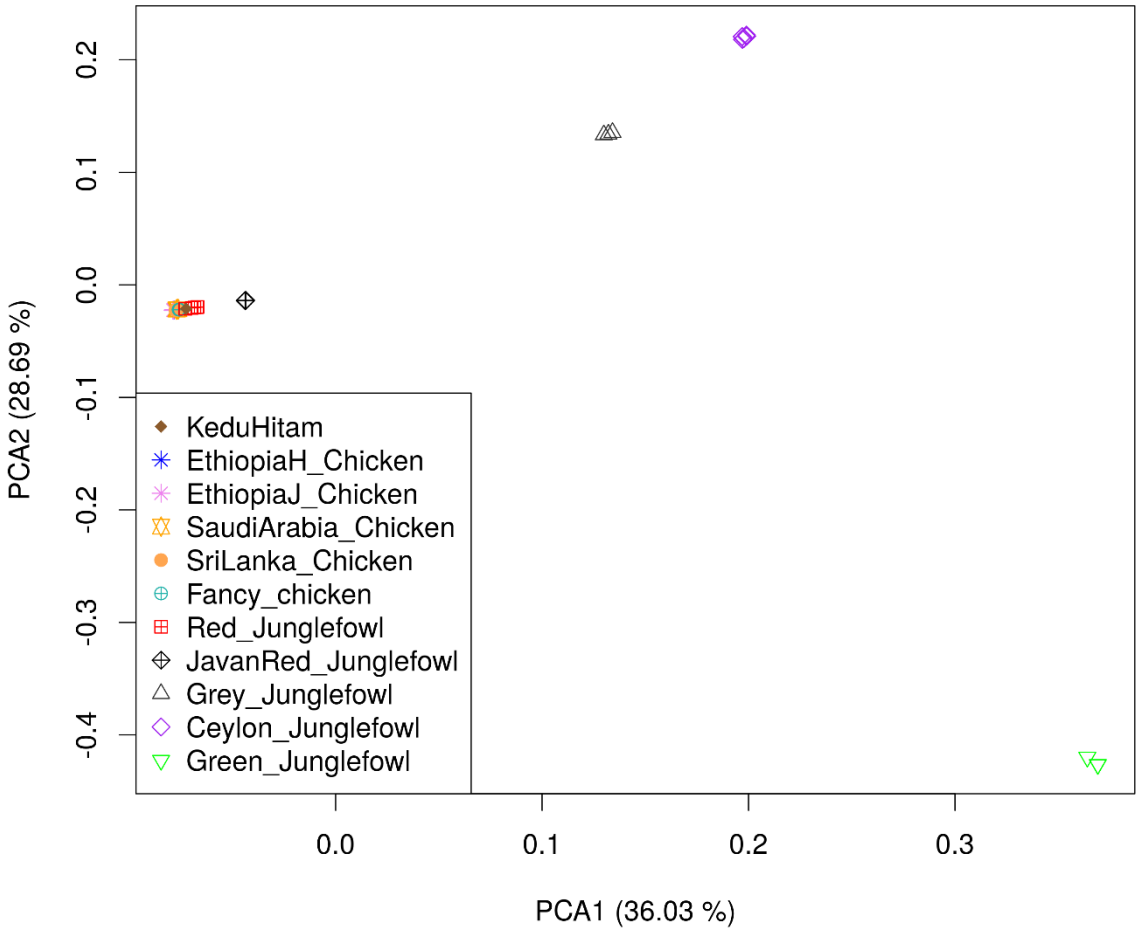


Figure 1. Principal component analysis for the 56 chicken genomes (~ 45 million SNPs).

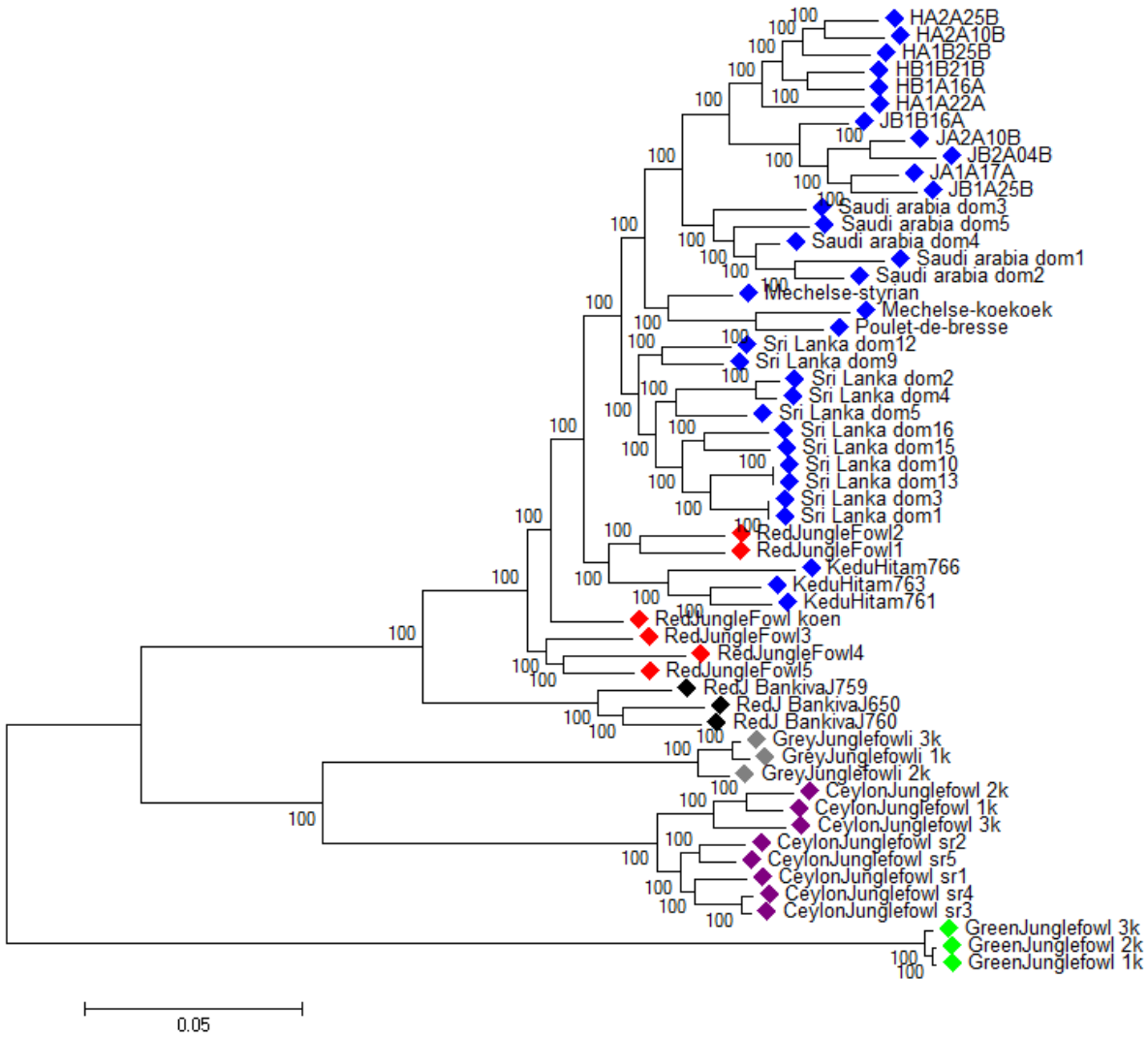


Figure 2. Maximum likelihood tree based on 89,696 (0.2%) randomly selected autosomal SNPs. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. Numbers above the nodes are bootstrap values (see **supplementary Table S1** for further details regarding the samples).

Introgression landscape between domestic chicken and Javan red junglefowl

To assess if Javan red junglefowl contributed to the gene pool of domestic chicken, I used the ABBA - BABA statistics previously used for our introgression study in chapter 4. The D statistics range from + 1 (complete admixture between P_2 and P_3) to -1 (complete shared variation between the P_1 and P_3). I estimated the mean genome-wide D statistics as 0.106 ± 0.130. I then searched for those candidate windows carrying derived allele shared between domestic chicken and Javan red junglefowl from a total of 181,923 windows analysed across the genome. Using the *Zfd* arbitrary threshold of 6, 166 (0.09%) candidate windows were identified which define 64 candidate introgressed regions. No significant window was found on chromosomes 9, 10, 12, 13, 15, 16, 18, 22, 24, 25, 26 and 28 (**Figure 3**). Details regarding the candidate introgressed regions are provided at **Table 2** and **3, supplementary table S2** and **S3.**

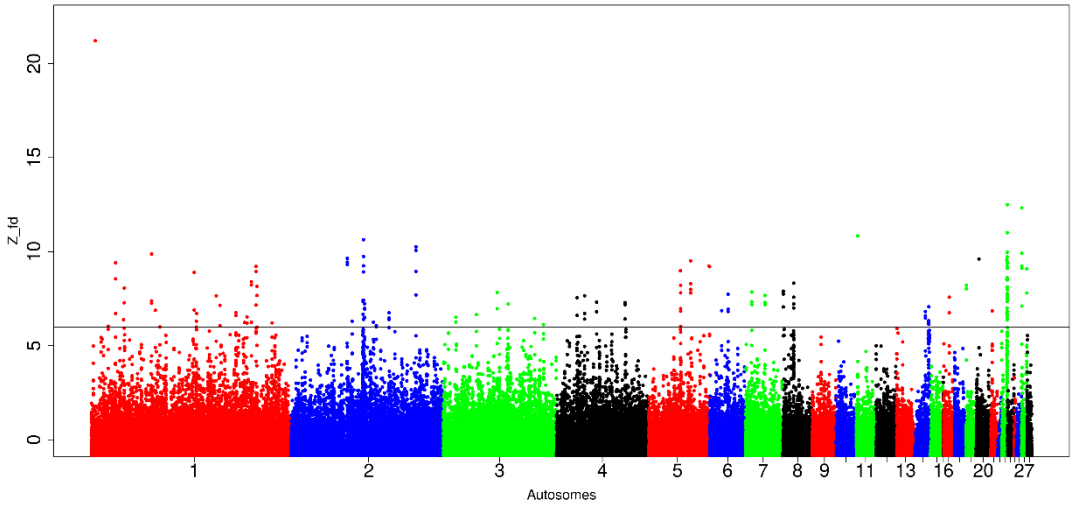


Figure 3. Plot of ABBA-BABA analysis (chromosomes 1 – 28) illustrating the introgression between the domestic chicken and Javan red junglefowl. The *Z_fd* is the transformed value obtained from the Martin's *fd* statistics. The horizontal line corresponds to the *Z_fd* score of 6.

I then performed haplotype relationship analysis to support and assess the direction of introgression. Only trees with Javan red junglefowl at the root of the red junglefowl – domestic chicken lineages were analysed further, with domestic haplotypes embedded within the *G. g, bankiva* ones considered as introgressed from this subspecies. For the introgression of domestic chicken into Javan red junglefowl, the Javan red junglefowl haplotypes will be embedded inside the donor red junglefowl – domestic chicken lineages.

Haplotype tree analyses show that 14 of these 64 candidate regions likely correspond to introgression of Javan red junglefowl into domestic chicken/red junglefowl (**Table 2**), one candidate region supports introgression of grey junglefowl into domestic chicken, one candidate region supports introgression of grey junglefowl into red junglefowl (**Table 3**) and 39 candidate regions support introgression of domestic chicken/red junglefowl into Javan red junglefowl (**supplementary table S2**). For nine regions, the direction of introgression could not be determined (**supplementary table S3**).

For the 14 haplotype trees showing introgression from Javan red junglefowl into domestic chicken/red junglefowl, examples are shown at **Figures 4** and **5.** The other 12 haplotype trees including those showing more than one directional introgression are presented at **supplementary Figures S2.1 – S2.12.** Six regions on chromosomes 1, 2, 3, and 23 show as well, introgression of domestic chicken/red junglefowl into Javan red junglefowl (see **Figure 5, supplementary Figures S2.1, S2.4, S2.5, S2.7, S2.8** and **Table 2**), two candidate regions located on chromosomes 1 and 4 also show introgression of domestic chicken into grey junglefowl (**supplementary Figures S2.1, S2.9** and **Table 2**) and one candidate region located on chromosome 23 also shows introgression of domestic chicken into Ceylon junglefowl (**Figure 5** and **Table 2**).

The fragment sizes of the significant windows of Javan red junglefowl candidate introgressed origins ranged from 20 kb to 115 kb with a mean of 40 kb and a median of 32 kb. For the 115-kb region, the longest candidate introgressed region above the threshold, the D statistics value was 0.29 ± 0.062 with the *Zfd* score of 6.37 ± 1.546. This region harbours four candidate genes, *HN1L*, *LA16c-431H6.6*, *IFT140* and *TMEM204* based on the Ensembl version 85 annotation (Aken *et al.*, 2016) (**Table 2**).

The 20-kb haplotype tree for each of the two candidate regions of introgression from grey junglefowl to domestic chicken and red junglefowl are shown at **Figure 6** and **supplementary Figure S3.1** respectively. The two trees also suggest introgression of domestic chicken into Javan red junglefowl while the **supplementary Figure S3.1** also shows introgression from domestic chicken into grey junglefowl.

The haplotype trees that correspond to the 39 candidate introgressed regions from domestic chicken/red junglefowl into Javan red junglefowl are presented at **supplementary Figures S4.1 – S4.39.** From these trees, nine candidate regions located

on chromosomes 1, 4 and 27 also show introgression into grey junglefowl (**supplementary Figures S4.9 - S4.14, S4.22, S4.23** and **S4.39**) and six candidate regions located on chromosome 23 also shows introgression into Ceylon junglefowl (**supplementary Figures S4.32 – S4.37**). The fragment size for introgression from domestic chicken ranged from 20 kb to 85 kb with a mean of 28 kb and median of 25 kb. For the 85-kb region, the longest region with window above the significant threshold, the D statistics value is 0.32 ± 0.13 with the *Zfd* score of 6.77 ± 1.956. This region harbours eight candidate genes which include, *KIAA1522, YARS, FNDC5, HPCA, RNF19B, AK2, AZIN2* and *NDUFS5* based on the Ensembl version 85 annotation (Aken *et al.*, 2016) (**supplementary Table S2**).

The **supplementary Figures S5.1 - S5.9** includes all candidate regions whose direction of introgression could not be determined. One of the regions located on chromosome 3 also shows introgression of domestic chicken into grey junglefowl (**supplementary Figure S5.3**).

The population distribution analysis of the 14 candidate Javan red junglefowl introgressed origin showed that nine are present in Sri Lanka domestic chicken and in red junglefowl, 8 in Kedu Hitam chicken and seven in Jarso, Horro and Saudi Arabian chicken (**supplementary Table S4**). For the CC®P parent birds, five candidate introgressed regions were observed in Poule de Bresse and seven regions in Mechelse Koekoe, while in the hybrid Mechelse Styrian, only five regions were identified. Within the indigenous domestic populations, the highest frequency of introgressed haplotypes from Javan red junglefowl was observed in Ethiopia (Jarso (62.8%)), followed by Saudi Arabia (61.4%), Sri Lanka (52%), Ethiopia (Horro (41.7%), Kedu Hitam (39.6%) and red junglefowl (34.3%) (**supplementary Table S4**).

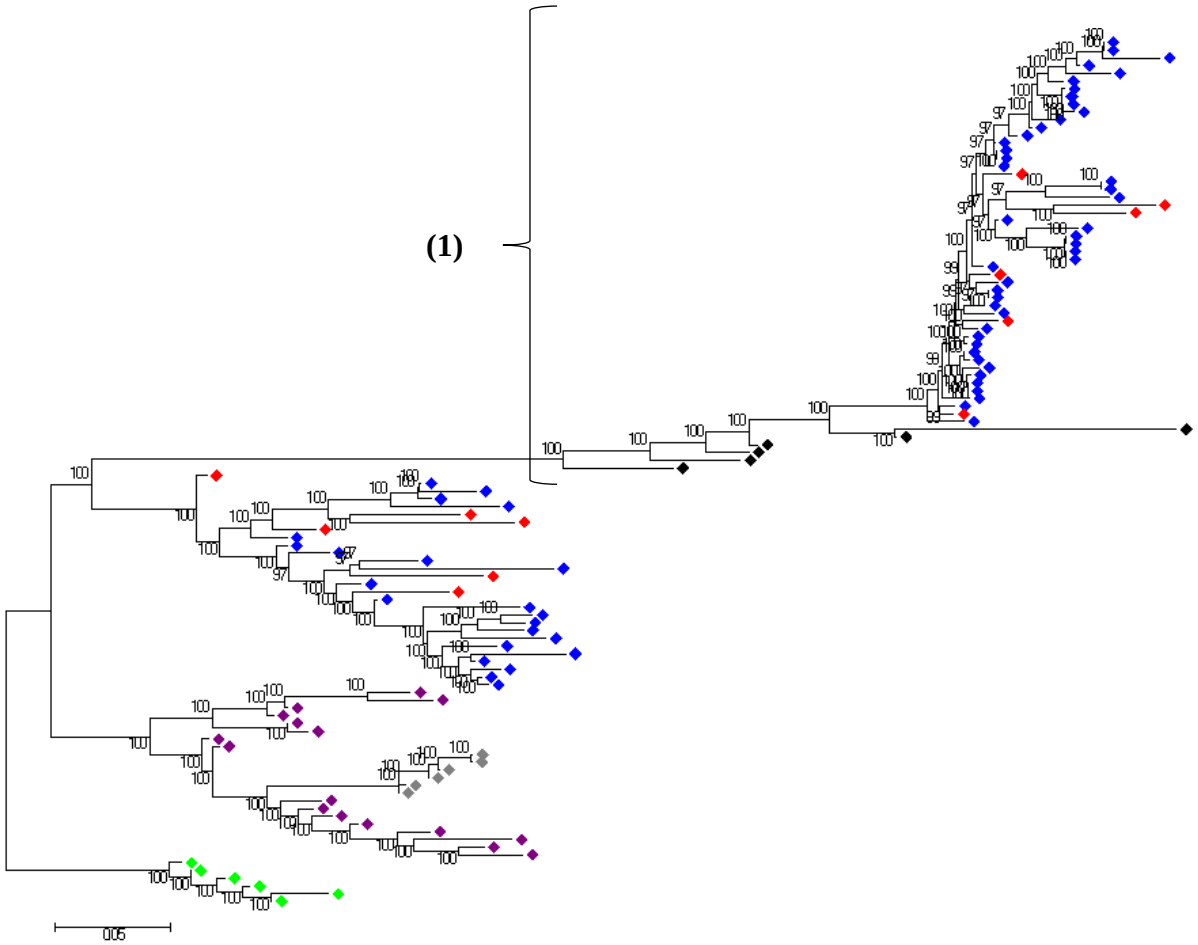


Figure 4. Haplotype based maximum likelihood tree for a region of introgression from (**1**) Javan red junglefowl into domestic chicken and red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 115-kb fragment size that extends from 13655000 to 13770000 bp on chromosome 14. It is the longest introgressed region identified here.

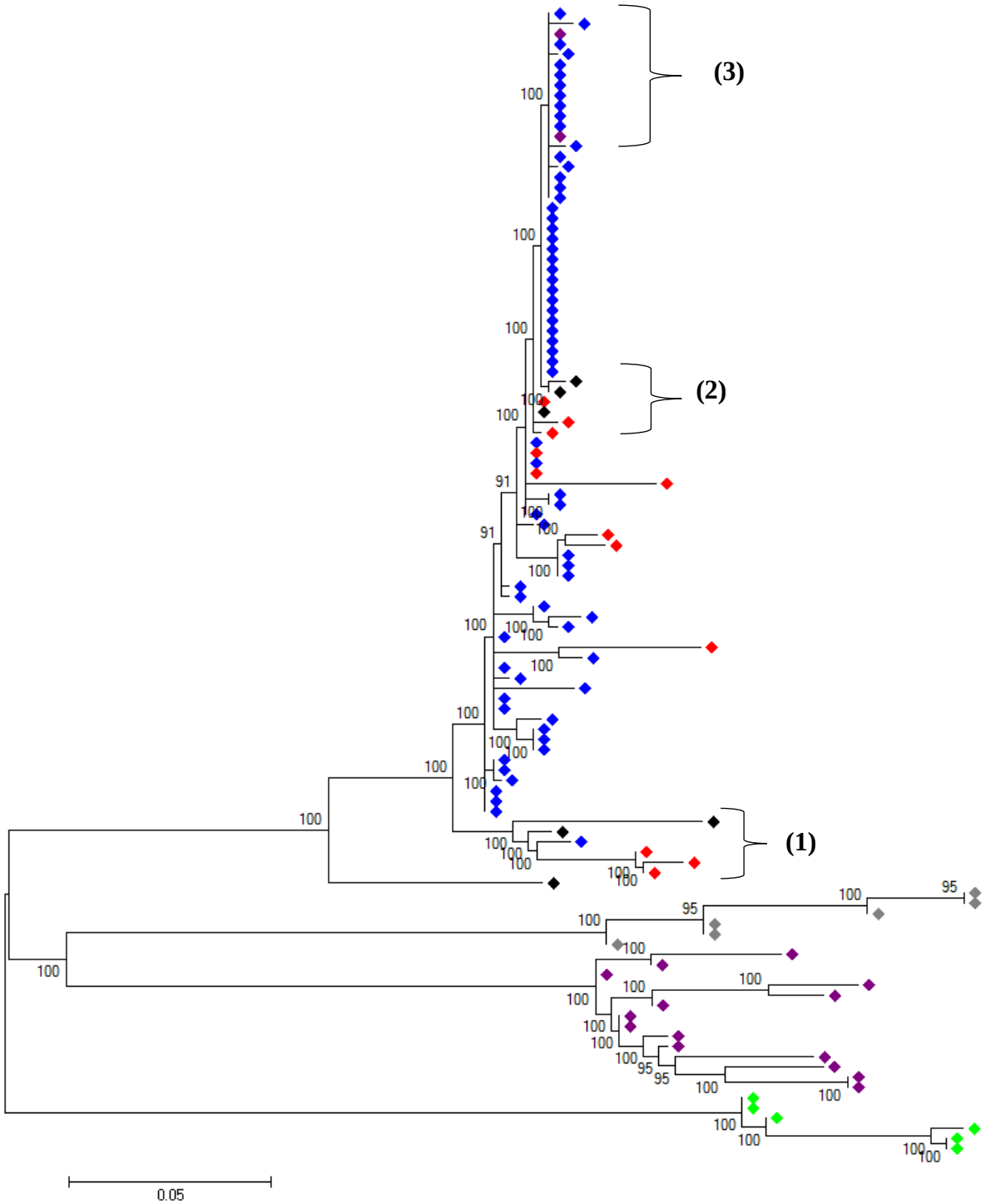


Figure 5. Haplotype based maximum likelihood tree for candidate regions of introgression of (**1**) Javan red junglefowl haplotypes into Sri Lanka5, red junglefowl1 and red junglefowl2, (**2**) domestic chicken/red junglefowl haplotypes into Javan red junglefowl, **(3)** domestic chicken haplotypes into Lafayetii2k sample. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 70-kb fragment size that extends from 5415000 to 5485000 bp on chromosome 23.

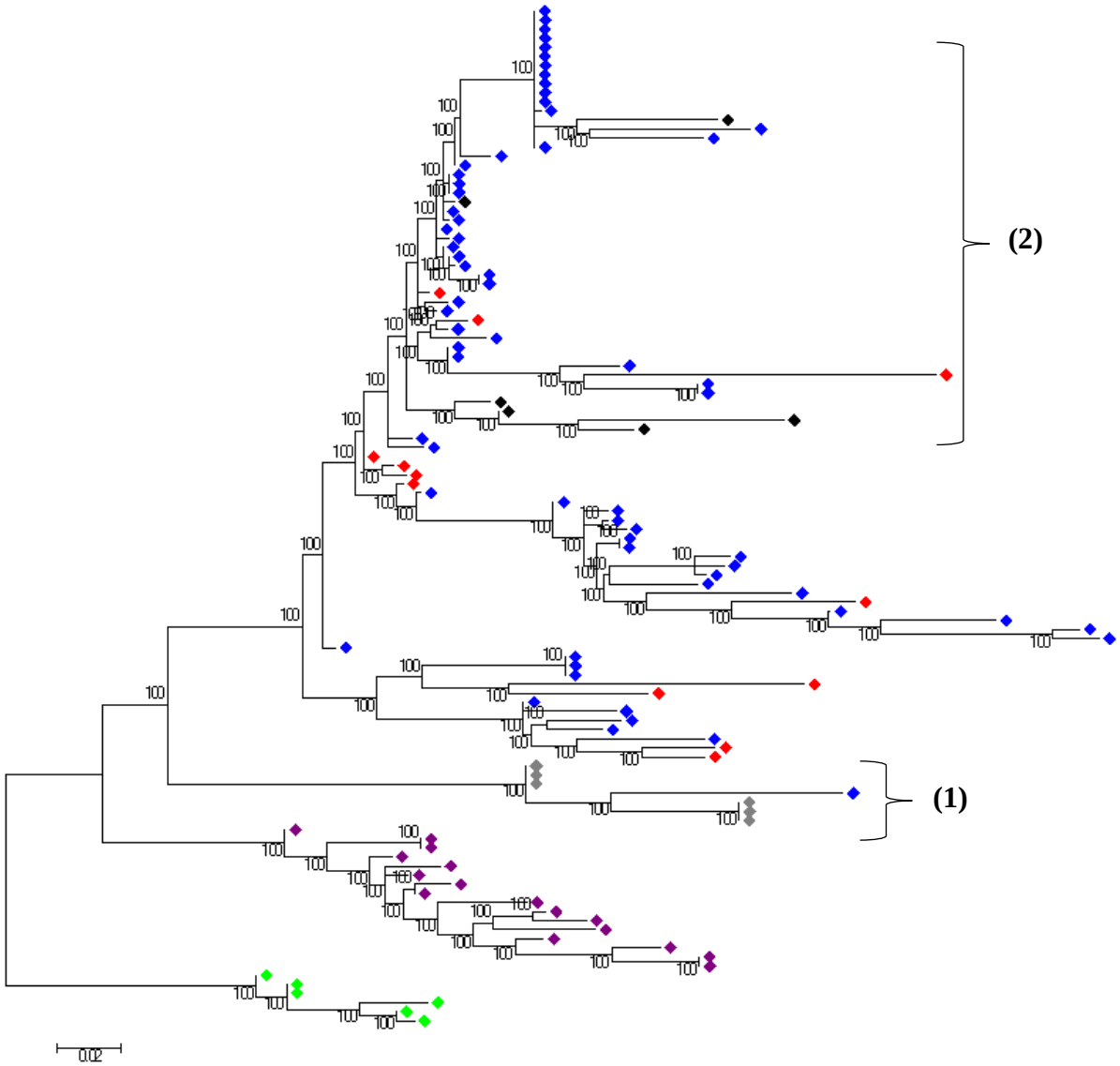


Figure 6. Haplotype based maximum likelihood tree for two candidate patterns of introgression based on the 20-kb fragment size that extends from 80375000 to 80395000 bp on chromosome 2. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of grey junglefowl haplotype into domestic chicken JA2A10B. **(2)** Introgression of domestic chicken/red junglefowl haplotypes to Javan red junglefowl into.

Table 2. Characteristics of the candidate introgressed Javan red junglefowl into domestic chicken regions (n = 14). ABBA – BABA test between Javan red junglefowl – domestic chicken.

Chr	Start	Stop	kb[c]	D (mean)	*fd* (mean)	*Zfd* (mean)	Overlapping genes
1*, [a]	16520000	16540000	20	0.56	0.32	6.03	
1	23745000	23775000	30	0.58 ± 0.05	0.41 ± 0.061	8.23 ± 1.377	
1	32250000	32275000	25	0.7 ± 0.023	0.39 ± 0.025	7.67 ± 0.551	
1*	176660000	176680000	20	0.23	0.32	6.22	*MIPEP*
2*	70260000	70295000	35	0.51 ± 0.042	0.35 ± 0.035	6.79 ± 0.781	
2	70675000	70720000	45	0.81 ± 0.038	0.44 ± 0.069	8.73 ± 1.555	
2*	71755000	71790000	35	0.57 ± 0.016	0.35 ± 0.014	6.91 ± 0.305	
3*	63290000	63310000	20	0.27	0.37	7.22	*TRAPPC3L, FAM26E, FAM26F, DSE*
4[a]	27440000	27470000	30	0.58 ± 0.048	0.36 ± 0.029	6.93 ± 0.648	
4	67000000	67030000	30	0.63 ± 0.049	0.36 ± 0.021	6.96 ± 0.473	
5	40880000	40915000	35	0.66 ± 0.013	0.42 ± 0.034	8.4 ± 0.767	
6	17915000	17945000	30	0.53 ± 0.068	0.37 ± 0.022	7.16 ± 0.499	
14	13655000	13770000	115	0.29 ± 0.062	0.33 ± 0.069	6.37 ± 1.546	*HN1L, LA16c-431H6.6, IFT140, TMEM204*
23*, [b]	5415000	5485000	70	0.63 ± 0.076	0.42 ± 0.066	8.4 ± 1.484	*PPIE, PABPC4, HEYL, NT5C1A, HPCAL4, TRIT1, MYCL*

*Regions also show introgression of domestic chicken into Javan red junglefowl. [a]Region also show introgression of domestic chicken into grey junglefowl. [b]This region also shows introgression of domestic chicken into Ceylon junglefowl. [c] Length in kilobase

Table 3. Characteristics of the grey junglefowl (*Gallus sonneratii*) introgressed region into domestic chicken regions. ABBA – BABA test between Javan red junglefowl – domestic chicken.

Chr	Start	Stop	kb[c]	D (mean)	*fd* (mean)	*Zfd* (mean)	Overlapping genes
2*	80375000	80395000	20	0.35	0.33	6.25	
21[a, b]	1485000	1505000	20	0.58	0.35	6.85	*PLCH2*

* This region also shows introgression of domestic chicken into Javan red junglefowl. [a]This region also shows introgression of grey junglefowl into red junglefowl. [b]This region also shows introgression of domestic into grey junglefowl. [c] Length in kilobase

Overlaps with signatures of selection

I compared each of the introgressed regions to those detected from signature of selection in indigenous domestic chicken/red junglefowl in chapter 2. From the total of 14 candidate introgressed regions from Javan red junglefowl, five of them overlapped with regions under positive selection (**Table 4**). They include introgressed region on chromosome 2 positively selected in Ethiopian domestic chicken, introgressed region on chromosome 4 positively selected in Saudi Arabian chicken and introgressed regions on chromosomes 5 and 6 positively selected in Sr Lanka domestic chicken (**Table 4**). The region on chromosome 23 (position 5415000 - 5485000 bp, **Figure 5**) for which the signature of positive selection signal was detected in Ethiopian, Saudi Arabia, Sri Lanka chicken and red junglefowl (see **Table 2** of chapter 2) provides a possible example of adaptive introgression in these populations.

Table 4. Candidate introgressed regions from *G. gallus bankiva* that overlapped with signature of selection regions in domestic and/or red junglefowl.

.

Chr	Start	Stop	Introgressed origin	Ethiopian	Saudi Arabian	Sri Lankan	Red junglefowl
2	70675000	70720000	Javan red junglefowl	✓			
4	27440000	27470000	Javan red junglefowl		✓		
5	40880000	40915000	Javan red junglefowl			✓	
6	17915000	17945000	Javan red junglefowl			✓	
23	5415000	5485000	Javan red junglefowl	✓	✓	✓	✓

Discussion

I investigated here using whole-genome sequences, the contribution of the Javan red junglefowl *G. g. bankiva* into the gene pool of domestic chicken.

Our analysis also illustrates the genetic relationships between the Javan red junglefowl subspecies, non-Javan red junglefowl subspecies, domestic chicken and other species within the genus *Gallus*. The presence of ~ 20% more homozygous SNPs in the Javan red junglefowl in comparison to the other non-Javan red junglefowl subspecies and the relatively large genetic distance between the Javan red junglefowl and domestic chicken support previous mitochondrial DNA analysis which suggested that the Javan red junglefowl among the other red junglefowl subspecies is the most distantly related one (Fumihito *et al.*, 1996). It should be noted that the subspecies status of the red junglefowl subspecies analysed here is not mentioned in the publication of reference (Wang *et al.*, 2015). However, as the birds have been sampled in the Yunnan and Hainan Province of China, they may be tentatively assigned to Burmese red junglefowl *G. gallus spadiceus* or Tonkinese red junglefowl *G. g. jabouillei*. Further analysis using well-defined subspecies samples of red junglefowl could ascertain the autosomal relationships between them.

I observed candidate introgression peaks on nearly all the chromosomes (**Figure 3)** but as I followed a stringent significant threshold criterion, 12 chromosomes were excluded from subsequent analysis. Genome-wide, I estimated 10.6% admixture between the Javan red junglefowl and domestic chicken and ~ 22% candidate introgressed regions into domestic chicken (i.e 14 out of the 64 total introgression originating from Javan red junglefowl), indicating that the autosomal genetic contributions of the Javan red junglefowl to the domestic chicken gene pool examined is significant. These results clearly disagree with previous D-loop segment of mitochondrial genome based studies as well as our mitochondrial results presented in this chapter, which excluded the Javan red junglefowl from the chicken maternal ancestry (Fumihito *et al.*, 1996; Liu *et al.*, 2006). This observation suggests a male mediated process of Javan red junglefowl introgression into domestic chicken population rather than a local domestication from Javan red junglefowl. However, it should be noted that the number of Indonesian domestic birds examined here was very small (n =3) and the inclusion of more domestic birds from different population may change the interpretation. A male mediated introgression pattern

is common in domestic as well as in wild species. For example, this scenario has been proposed for zebu introgression into taurine cattle in Africa (Hanotte *et al.*, 2000), in wild birds (Bellemain *et al.*, 2008), wild mammals (Melo-Ferreira *et al.*, 2009), wild amphibians (Zieliński *et al.*, 2013) etc.

Interestingly, our results are providing some insights regarding the dispersion pattern of domestic chicken after their arrival on the Indonesian islands. Javan red junglefowl autosomal haplotypes were identified in all the domestic chicken populations analysed. In particular, the Ethiopian Jarso, Saudi and Sri Lankan chicken populations have about 50% or more frequency of introgressed haplotypes, thus, suggesting an ancient domestic chicken movement through the Indian ocean trade network from the Indonesian islands. With the lower genetic differentiation between the Javan and Sri Lankan chicken and with highest frequency of Javan introgressed regions in the latter, domestic chicken arrival from Indonesia to Sri Lanka was likely before the dispersion of such birds to the Middle East/East Africa. Reconstructing this dispersal route and the history of Javan red junglefowl introgressed domestic chicken will obviously require the analysis of more populations.

As for the Javan red junglefowls introgressed regions from domestic chickens, there are two possibilities: first, an ancient hybridization between subspecies, i.e. with *Gallus gallus spadiceus* and *G. g. bankiva* since both are found in Sumatra, or, as more likely, hybridization could have occurred following the feralization of domestic chicken from Indian subcontinent as seen in the grey junglefowl introgression to red junglefowl1, 2, and 3 samples in chapter 4. The autosomal phylogenetic tree also clustered the red junglefowl1 and 2 with the domestic chicken from Indonesia (**Figure 2**). More so, most of the Javan introgressed haplotypes were present in red junglefowl1, 2 and 3 samples, thus, further supporting feralization or domestic chicken introgression.

Introgression from the Javan red junglefowl has undoubtedly contributed to the diversity of domestic chicken and perhaps also to its adaptive plasticity. Amongst several candidate genes found on the 115 kb longest Javan introgressed region is *IFT140* which is known to play a significant role in the development of the brain, renal, retinal and skeletal systems (Perrault *et al.*, 2012). It is also involved in the proper development and function of ciliated cells. Other gene include *FAM26E*, a member of *FAM26* family which plays a critical role in neuronal growth and sensory development (Ma *et al.*, 2016). Also, five

candidate Javan introgressed regions are under positive selection in some indigenous domestic chicken populations analysed in chapter 2. One of the introgressed regions on chromosome 23, a 70-kb region was found positively selected in all the domestic chickens and in the red junglefowl. One of the candidate introgressed genes *NT5C1A* present within this region have been previously reported as selected in layer chicken lines (Rubin *et al.*, 2010) and extensively described in chapter 2 to play a role in the control of hypoxia (Hunsucker *et al.*, 2001).

Conclusions

This study provides new insight on the evolutionary history of domestic chicken. It shows for the first time that the Javan red junglefowl has contributed substantially to the gene pool of the domestic chicken. Combined with the results of chapter 4, it shows that the domestic chicken populations should not be considered as a single lineage but rather constitute an ensemble of populations with different level of introgression from different taxa. The analysis of a wider range of chicken populations will be necessary to unravel how these introgressions have shaped the diversity landscape of chicken globally. Also, as supported in chapter 4, hybridization with green junglefowl might have also shaped the genetic make-up of the Indonesian domestic chicken studied here.

References

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., *et al.* (2016). The Ensembl gene annotation system. *The Journal of Biological Databases and Curation, 2016*. doi:10.1093/database/baw093

Bahr, J. M. (2008). The chicken as a model organism In: *Sourcebook of models for biomedical research* (pp. 161-167): Springer.

Bellemain, E., Bermingham, E., and Ricklefs, R. E. (2008). The dynamic evolutionary history of the bananaquit (*Coereba flaveola*) in the Caribbean revealed by a multigene analysis. *BMC Evolutionary Biology, 8*(1), 240.

Bosse, M., Megens, H.-J., Frantz, L. A., Madsen, O., Larson, G., Paudel, Y., *et al.* (2014). Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nature Communications, 5*. doi:10.1038/ncomms5392

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., *et al.* (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.

4369 Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: More
4370 models, new heuristics and parallel computing. *Nature Methods, 9*(8), 772-772.

4371 Darwin, C. (1868). *The variation of animals and plants under domestication* (2 ed. Vol.
4372 2): John Murray, United Kingdom.

4373 Delacour, J. (1977). *The pheasant of the world* (2 ed.): Saiga Publishing Co. Ltd. Surr
4374 GU26 GTD. England.

4375 Delaneau, O., Howie, B., Cox, A. J., Zagury, J.-F., and Marchini, J. (2013). Haplotype
4376 estimation using sequencing reads. *The American Journal of Human Genetics,*
4377 *93*(4), 687-696.

4378 Desta, T., Lawal, R., and Hanotte, O. (2015). Genetic diversity of village chickens In:
4379 Third report on chicken genes and chromosomes 2015. *Cytogenetic and Genome*
4380 *Research, 145*(2), 130-133.

4381 Durand, E. Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient
4382 admixture between closely related populations. *Molecular Biology and Evolution,*
4383 *28*(8), 2239-2252.

4384 Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt,
4385 L., *et al.* (2008). Identification of the yellow skin gene reveals a hybrid origin of
4386 the domestic chicken. *PLoS Genetics, 4*(2).
4387 doi.org/10.1371/journal.pgen.1000010

4388 Fumihito, A., Miyake, T., Sumi, S., Takada, M., Ohno, S., and Kondo, N. (1994). One
4389 subspecies of the red junglefowl (*Gallus gallus gallus*) suffices as the matriarchic
4390 ancestor of all domestic breeds. *Proceedings of the National Academy of Sciences,*
4391 *91*(26), 12505-12509.

4392 Fumihito, A., Miyake, T., Takada, M., Shingu, R., Endo, T., Gojobori, T., *et al.* (1996).
4393 Monophyletic origin and unique dispersal patterns of domestic fowls.
4394 *Proceedings of the National Academy of Sciences, 93*(13), 6792-6795.

4395 Gou, X., Wang, Z., Li, N., Qiu, F., Xu, Z., Yan, D., *et al.* (2014). Whole-genome
4396 sequencing of six dog breeds from continuous altitudes reveals adaptation to high-
4397 altitude hypoxia. *Genome Research, 24*(8), 1308-1315.

4398 Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., *et al.* (2010).
4399 A draft sequence of the Neandertal genome. *Science, 328*(5979), 710-722.

4400 Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate
4401 large phylogenies by maximum likelihood. *Systematic Biology, 52*(5), 696-704.

4402 Hanotte, O., Tawah, C., Bradley, D., Okomo, M., Verjee, Y., Ochieng, J., *et al.* (2000).
4403 Geographic distribution and frequency of a taurine Bos taurus and an indicine Bos
4404 indicus Y specific allele amongst sub-Saharan African cattle breeds. *Molecular*
4405 *Ecology, 9*(4), 387-396.

4406 Havenstein, G., Ferket, P., and Qureshi, M. (2003). Growth, livability, and feed
4407 conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001
4408 broiler diets. *Poultry Science, 82*(10), 1500-1508.

4409 Hoffmann, I. (2009). The global plan of action for animal genetic resources and the
4410 conservation of poultry genetic resources. *World's Poultry Science Journal,*
4411 *65*(02), 286-297.

4412 Hunsucker, S. A., Spychala, J., and Mitchell, B. S. (2001). Human Cytosolic 5′-
4413 Nucleotidase I characterization and role in nucleoside analog resistance. *Journal*
4414 *of Biological Chemistry, 276*(13), 10498-10504.

4415 Hutt, F. (1949). *Genetics of the fowl*: McGraw-Hill Book Company Inc., New York, NY.

4416 Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–
4417 Wheeler transform. *Bioinformatics, 26*(5), 589-595.

4418 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009). The
4419 sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078-
4420 2079.

4421 Liu, Y.-P., Wu, G.-S., Yao, Y.-G., Miao, Y.-W., Luikart, G., Baig, M., *et al.* (2006).
4422 Multiple maternal origins of chickens: out of the Asian jungles. *Molecular*
4423 *Phylogenetics and Evolution, 38*(1), 12-19.

4424 Ma, Z., Tanis, J. E., Taruno, A., and Foskett, J. K. (2016). Calcium homeostasis
4425 modulator (CALHM) ion channels. *Pflügers Archiv-European Journal of*
4426 *Physiology, 468*(3), 395-403.

4427 Martin, S. H., Davey, J. W., and Jiggins, C. D. (2015). Evaluating the use of ABBA–
4428 BABA statistics to locate introgressed loci. *Molecular Biology and Evolution,*
4429 *32*(1), 244-257.

4430 Medugorac, I., Graf, A., Grohs, C., Rothammer, S., Zagdsuren, Y., Gladyr, E., *et al.*
4431 (2017). Whole-genome analysis of introgressive hybridization and
4432 characterization of the bovine legacy of Mongolian yaks. *Nature Genetics, 49,*
4433 470–475.

4434 Melo-Ferreira, J., Alves, P., Freitas, H., Ferrand, N., and Boursot, P. (2009). The genomic
4435 legacy from the extinct Lepus timidus to the three hare species of Iberia: contrast
4436 between mtDNA, sex chromosomes and autosomes. *Molecular Ecology, 18*(12),
4437 2643-2658.

4438 Muir, W. M., Wong, G. K.-S., Zhang, Y., Wang, J., Groenen, M. A., Crooijmans, R. P.,
4439 *et al.* (2008). Genome-wide assessment of worldwide chicken SNP genetic
4440 diversity indicates significant absence of rare alleles in commercial breeds.
4441 *Proceedings of the National Academy of Sciences, 105*(45), 17312-17317.

4442 Nishibori, M., Shimogiri, T., Hayashi, T., and Yasue, H. (2005). Molecular evidence for
4443 hybridization of species in the genus *Gallus* except for *Gallus varius*. *Animal*
4444 *Genetics, 36*(5), 367-375.

Pasteur, L. (1880). Sur les maladies virulentes: et en particulier sur la maladie appelée vulgairement choléra des poules. *Comptes Rendus de l'académie des Sciences, 90*, 242–248.

Perrault, I., Saunier, S., Hanein, S., Filhol, E., Bizet, A. A., Collins, F., *et al.* (2012). Mainzer-Saldino syndrome is a ciliopathy caused by IFT140 mutations. *The American Journal of Human Genetics, 90*(5), 864-870.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics, 38*(8), 904-909.

Rubin, C.-J., Zody, M. C., Eriksson, J., Meadows, J. R., Sherwood, E., Webster, M. T., *et al.* (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature, 464*(7288), 587-591.

Stern, C. D. (2005). The chick: a great model system becomes even greater. *Developmental cell, 8*(1), 9-17.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research, 25*(24), 4876-4882.

Ulfah, M., Kawahara-Miki, R., Farajalllah, A., Muladno, M., Dorshorst, B., Martin, A., *et al.* (2016). Genetic features of red and green junglefowls and relationship with Indonesian native chickens Sumatera and Kedu Hitam. *BMC Genomics, 17*(1), 1.

Wang, M.-S., Li, Y., Peng, M.-S., Zhong, L., Wang, Z.-J., Li, Q.-Y., *et al.* (2015). Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. *Molecular Biology and Evolution, 32*(7), 1880-1889.

Zieliński, P., Nadachowska-Brzyska, K., Wielstra, B., Szkotak, R., Covaciu-Marcov, S., Cogălniceanu, D., *et al.* (2013). No evidence for nuclear introgression despite complete mtDNA replacement in the Carpathian newt (*Lissotriton montandoni*). *Molecular Ecology, 22*(7), 1884-1903.

Chapter 6

General discussion

The genetic changes that shaped the diversity in domesticated animals started at the time of domestication events (Zeuner, 1963; Sossinka, 1982; Mason, 1984) and were followed by the alteration in gene frequencies resulting from the influence of evolutionary forces, e.g. human selection pressure, genetic drift, founder effects etc (Crawford, 1990). Increasingly, it appears that the genomes of our livestock domesticates were subsequently also shaped by post-domestication species introgression such as the ones revealed here for the domestic chicken (see chapters 4 and 5).

Unlike the commercial chicken lines, the indigenous village chicken, mostly found in developing countries, are largely unmanaged, allowed to scavenge all day under harsh weather and at night finding shelter on the village trees and bushes. Their productivity is low and mortality is high. However, given the extensive system of management for the indigenous village chickens, their genome has remained more diverse compared to their commercial counterparts. This is illustrated here by the higher number of SNPs, ranging from 5.5 to 6.7 million identified in each of our indigenous domestic village chicken samples when compared to the previously published results on commercial and experimental chicken lines where 3.6 to 5 million SNPs were detected on average following the analysis of 10 birds per population (Gheyas *et al.*, 2015).

By 2050, the world human population is projected to reach 9.7 billion, a 2.5 billion increase in the current population estimates (United Nations, 2015 report). Most of the population increase will be expected to take place in developing countries, thus, implying that the demand for food will significantly go up. In Africa for instance, indigenous chicken served as sources of food for the poor and contributes a significant proportion to the national and continental poultry flocks (Scanes, 2007). To meet the future food demand, it is important to improve the productivity of the local breeds. Taking advantage of their genetic diversity and local adaptation will help to develop sustainable breeding improvement strategies.

In chapter 2, I identified several candidate selected regions containing genes that play roles in adaptation to production environments. For instance, in Ethiopian chicken, the regulation of vasoconstriction as detected by the gene network analysis may suggest the ability to adapt to an environment where the temperature is lower than the average normal body temperature. The two sampling locations in Ethiopia are known to have temperature

average around 21ºC which is lower than 41ºC normal chicken body temperature. Also, candidate selected regions with genes linked to the regulation of heat loss was found in Saudi Arabian chicken. This agrees with the climatic conditions of the sampling locations and in general of Saudi Arabia, a dry country with peaks of temperature during the day and of more than 50ºC in July/August. These findings are of relevance in the context of climatic changes and its impact on agricultural productivity.

The gene ontology analysis shows the presence of fewer candidate genes associated with production and reproduction traits in Ethiopian and Sri Lankan chickens when compared to the Saudi Arabian chicken. Historically, the sampling locations in Saudi Arabia have been under geographic isolation long before the establishment of the kingdom of Saudi Arabia. During this time, interaction with people from outside the community was limited and food production was locally sourced. This may have led to stronger selection for productivity traits in the Saudi Arabian population compared to the other populations studied here. Similar evidence in other livestock species living in this environment may provide insight on this historic practice. Signatures of selection for productivity traits are of course important in the context of breeding programmes for improvement purposes.

The signature of selection identified in chapter 3 shows candidate regions containing genes that may be linked to adaptation to environmental stresses other than the climatic one e.g. immune response, hypoxia and starvation. These selection pressures are stronger in Ceylon junglefowl than in green junglefowl, thus agreeing with previous observations that the green junglefowl is more susceptible to infectious diseases, less hardy overall, delicate and in winter requires artificial heat compared to the Ceylon junlefowl (Delacour, 1977). However, candidate selected regions containing genes associated with bone and muscle development were more common in green junglefowl than in red and Ceylon junglefowls. The identification of *NOBOX* gene which plays a role in the development of ovarian follicle in green junglefowl is consistent with earlier report that their reproductive ability is higher than in other *Gallus* species (Delacour, 1977). Together with our introgression data indicating that cross-species hybridization within the genus *Gallus* is indeed common, some of these regions may represent important introgressed regions in domestic chicken populations and may be involved in the enhancement of their adaptabilities to environmental challenges.

Darwin believed that domestication, which likely took place around 6000 BC for the domestic chicken, involved only the red junglefowl because in the knowledge of the time only crosses with domestic chicken produced fertile offspring when compared to other junglefowl species (Darwin, 1868; Crawford, 1990). The wild-type standard (plumage, morphology etc.) found in domestic chicken has also been proposed to originate from the red junglefowl (Jaap and Hollander, 1954). In chapters 4 and 5, I studied the genome ancestry of domestic chicken and in particular the possible introgression from grey junglefowl, Ceylon junglefowl and the subspecies Javan red junglefowl. Our result showed that the chicken genome contained the signature of four junglefowl species alongside the Javan red junglefowl. These genetic contributions to chicken gene pool could have occurred either during domestication or rather likely after the dispersion of the domestic chicken. These different introgression episodes have played significant roles in the diversity that I observed in today domestic chicken.

Conclusions, future studies and recommendations

The work in this thesis provides knowledges on the genomic regions adapted to different agro-ecological environments in different geographically distributed domestic chicken, the red, Ceylon and green junglefowls populations. It also provides for the first time knowledge on the selected regions that may be genus *Gallus* specific. This thesis is also the first to report introgression from the four junglefowl species to the chicken genome. It supports the polyphyletic origin of domestic chicken. These selected and introgressed candidate regions help us understand further, the evolutionary history of the domestic chicken and the different adaptive mechanisms to their natural habitats.

Obviously, one of the limitations in our study is the number of chicken populations examined and in particular, from East Asia. There is still much more to discover and in a sense, it may be argued that I only present here, the tip of the iceberg.

I performed independently our signature of selection and introgression analyses and then examined the overlaps between the two analyses. Considering our results which indicate the relatively common introgression across the genome, an alternative approach would have been to assess any candidate signature of selection regions for evidence of introgression. This will be worth pursuing and it will address to some extent, the use of arbitrary thresholds in the two analyses.

While performing the analyses in this thesis, a new reference genome *Galgal* 5.0 was released which contained additional gene annotation when compared to the *Galgal* 4.0 reference. Except for chapter 4 where the introgressed regions detected using *Galgal* 4.0 were also re-examined using *Galgal* 5.0, candidate regions from other chapters have not been reassessed using the new reference. As the new reference genome may contain addition gene annotation, it is recommended that future study should consider using the latest reference genome with attention to candidate selected or introgressed regions that have not been assigned to genes in *Galgal* 4.0.

References

Crawford, R. (1990). *Poultry breeding and genetics* (Vol. 1): Elsevier Amsterdam.

Darwin, C. (1868). *The variation of animals and plants under domestication* (2 ed. Vol. 2): John Murray, United Kingdom.

Delacour, J. (1977). *The pheasant of the world* (2 ed.): Saiga Publishing Co. Ltd. Surr GU26 GTD. England.

Gheyas, A. A., Boschiero, C., Eory, L., Ralph, H., Kuo, R., Woolliams, J. A., *et al.* (2015). Functional classification of 15 million SNPs detected from diverse chicken populations. *DNA Research, 22*(3), 205-217.

Jaap, R. G., and Hollander, W. (1954). Wild type as standard in poultry genetics. *Poultry Science, 33*(1), 94-100.

Mason, I. L. (1984). *Evolution of domesticated animals* (Vol. 89): Longman London.

Scanes, C. (2007). *Contribution of poultry to quality of life and economic development in the developing world*: Oxford University Press.

Sossinka, R. (1982). *Domestication in birds* (Vol. 6): Academic Press, New York.

United Nations. Department of Economic and Social Affairs. World Population Prospects: The 2015 Revision.

Zeuner, F. E. (1963). *A history of domesticated animals*: Hutchinson & Co. (Publishers) Ltd. London.

Supplementary Tables

Chapter 2

Signature of positive selection in red junglefowl and domestic chicken

4637 **Table S1.** Summary sequencing statistics and numbers of SNPs

Sample	Reads	PPR (%)	MR (%)	MD (X)	5_b (%)	10_b (%)	20_b (%)	Hom	Het	Total SNPs
HA2A25B[a]	404491652	93.75	99.27	32.38	97.9	96.7	91.3	2663926	3625010	6288936
HA1B25B[a]	375265404	93.68	98.77	31.05	98.1	96.8	89.7	2731530	3490701	6222231
HB1A16A[a]	402727784	93.55	98.79	31.93	98.2	97	90.8	2663242	3678469	6341711
HB1B21B[a]	406780167	93.92	98.87	32.51	98.1	96.9	91.1	2683493	3611342	6294835
HA1A22A[a]	386714341	94.48	98.84	31.52	98.1	96.9	90.4	2978357	3010187	5988544
HA2A10B[a]	408859746	92.63	98.84	32.81	98.2	97.1	91.4	2773136	3440448	6213584
JA2A10B[a]	399745882	93.38	98.79	32.04	98.1	96.9	90.7	2865428	3210674	6076102
JB1B16A[a]	363393410	94.61	99.36	29.31	98.4	96.2	83.1	2658765	3617612	6276377
JB1A25B[a]	382829596	93.76	98.88	32.04	98.1	96.9	90.7	3081092	2741341	5822433
JB2A04B[a]	401918477	93.25	98.31	31.85	98.1	96.8	90.3	3185080	2589279	5774359
JA1A17A[a]	396334439	91.52	97.61	31.55	98	96.7	89.5	2822391	3291155	6113546
Saudi Arabia1[b]	108937623	91.83	98.87	13.75	96.1	78.2	8.7	3011627	2475976	5487603
Saudi Arabia2[b]	109127614	91.97	98.64	13.79	95.7	75.5	10	2563278	3487603	6050881
Saudi Arabia3[b]	108810954	92.71	98.02	10.22	87.4	43.2	4.4	2626554	3051262	5677816
Saudi Arabia4[b]	109038659	94.2	98.73	13.87	96.3	78.8	9	1768821	4878947	6647768
Saudi Arabia5[b]	109394273	92.41	98.53	10.5	91.5	50.9	2.1	2523799	3350890	5874689
Sri Lanka1[c]	272007288	92.75	98.74	31.46	97.9	96.4	87.7	2427647	4167743	6595390
Sri Lanka2[c]	269162107	92.84	98.65	31.03	98	96.4	87.7	2380427	4236500	6616927
Sri Lanka3[c]	262738537	92.9	98.77	30.45	98	96.3	86.4	2428299	4166738	6595037
Sri Lanka4[c]	269326187	93.02	98.57	31.05	98	96.4	86.9	2365620	4349780	6715400
Sri Lanka5[c]	281834982	92.6	98.82	32.41	98.5	96.5	85.3	2395630	4285906	6681536
Sri Lanka9[c]	267558558	92.73	98.78	30.9	98.4	96.1	83	2333190	4347940	6681130
Sri Lanka10[c]	276733672	92.68	98.65	31.89	98.5	96.4	84.2	2462458	4110133	6572591
Sri Lanka12[c]	269404630	92.48	98.8	31.23	98	96.5	87.2	2365932	4326208	6692140
Sri Lanka13[c]	249093732	91.34	98.72	28.73	98.3	95.4	79.3	2461960	4099930	6561890
Sri Lanka15[c]	229370558	92.89	98.99	26.8	97.3	93.6	74.7	2453733	4150327	6604060
Sri Lanka16[c]	251649616	93.18	98.8	29.2	97.9	96.1	84.5	2428773	4249479	6678252
Red junglefowl1[d]	154287870	67.51	97.72	12.96	93.4	70.2	9.8	2524724	3573351	6098075

Red junglefowl2[d]	286656521	95.33	98.93	25.6	98.7	95.8	76.2	2561487	4115394	6676881
Red junglefowl3[d]	186743667	94.21	99.14	15.9	95.1	77.1	27.7	2633783	4364117	6997900
Red junglefowl4[d]	224231071	95.5	99.07	19.09	97.1	92.8	43.2	2974376	3800042	6774418
Red junglefowl5[d]	403120509	94.8	98.91	36.05	98.2	97.3	93.4	2540845	4919100	7459945
Red junglefowl_koen[e]	149692501	90.08	94.69	15.97	96.6	83.5	20.8	2393695	4754034	7147729

[a] Population from Ethiopia including Horro samples (name start with H) and Jarso samples (name start with J)
[b] Population from Saudi Arabia
[c] Population from Sri Lanka
[d] Red junglefowl sequences retrieved from "Wang, M.-S., Li, Y., Peng, M.-S., Zhong, L., Wang, Z.-J., Li, Q.-Y. *et al.* Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. *Molecular biology and evolution* **32**, 1880-1889 (2015)"
[e] Red junglefowl sequence from Cosmopolitan Chicken Research Project in Belgium (http://www.ccrp.be/#/0)

PPR: Percentage of both mates of a read pair that were properly mapped to the same chromosome.
MR: Total reads mapped to the *Galgal* 4.0 reference genome.
MD: The mean sequence depth or the genome sequence coverage.
5_b: Percentage of the genome with bases covered by at least 5 reads.
10_b: Percentage of the genome with bases covered by at least 10 reads.
20_b: Percentage of the genome with bases covered by at least 20 reads.
Hom: Homozygous SNPs compare to the reference.
Het: Heterozygous SNPs compare to the reference.
The **Total SNPs** includes the bi-allelic **Hom** + **Het.**

Table S2. Variants SNPs statistics and annotations at population level

Variant consequences	Ethiopian Domestic	Saudi Arabian Domestic	Sri Lankan Domestic	Red Junglefowl
Total number of sample	11	5	11	6
Total number of bi-allelic SNPs	13,075,004	10,231,837	14,455,255	15,317,274
Novel Variants	1,765,053	1,029,993	2,325,291	4,451,313
Existing Variants	11,309,951	9,201,844	12,129,964	10,865,961
Overlapped genes	39,892	39,868	39,901	39,922
Overlapped transcript	47,659	47,647	47,659	47,677
Splice donor variant	643	444	693	730
Splice acceptor variant	570	496	603	653
Stop gained	1154	886	1302	1281
Stop lost	307	265	345	371
Start lost	365	303	416	428
Missense variant	74,488	57,637	83,438	87,272
Splice region variant	26,015	20,522	28,811	30,350
Stop retained variant	79	59	80	92
Synonymous variant	123,041	98,486	138,077	145,705
Coding sequence variant	3	1	4	1
Mature miRNA variant	170	133	197	214
5 prime UTR variant	51,655	40,869	57,787	60,366
3 prime UTR variant	246,270	193,040	273,686	289,436
Non-coding transcript exon variant	133,818	10,6657	148,262	158,914
Intron variant	7,058,272	554,4587	7,791,825	8,244,547
Upstream gene variant	758,660	604,830	838,568	885,235
Downstream gene variant	568,176	451,103	626,089	664,720
Intergenic variant	4,031,318	3,111,519	4,465,072	4,746,959

Table S3. Reference haplogroup sequences for the D-loop mitochondrial DNA analysis.

Haplogroup	D-loop sequence
A	AATTTTATTTTTTAACCTAACTCCCCTACTAAGTGTACCCCCCCTTTCCCCCCCCAGGGGGGGTATACTATGCATAATCGTGCATACATTTATATACCAC ATATATTATGGTACCGGTAATATATACTATATATGTACTAAACCCATTATATGTATACGGGCATTAATCTATATTCCACATTTCTCCCAATGTCCATTCT ATGCATGATCCAGGACATACTCATTCACCCTCCCCATAGACAGCTCCAAACCACTACCAAGTCACCTAACTATGAATGGTTGCAGGACATAAATCTCA CTCTCATGCTCTTCCCCCAACAAGTCACCTAACTATGAATGGTTACAGGACATACATTTAACTACCATGTTCTAACCCATTTGGTTATGCTCGCCGTATC
B	AATTTTATTTTTTAACCTAACTCCCCTACTAAGTGTACCCCCCCTTTCCCCCCCCAGGGGGGGTATACTATGCATAATCGTGCATACATTTATATACCACAT ATATTATGGTACCGGTAATATATACTATATATGTACTAAACCCATTATATGTATACGGGCATTAACCTATATTCCACATTTCTCCCAATGTCCATTCTATGC ATGATCCAGGACATACTCATTTACCCTCCCCATAGACAGTTCCAAACCACTATCAAGCCACCTAACTATGAATGGTTACAGGACATAAATCTCACTCTCA TGTTCTCCCCCCAACAAGTCACCTAACTATGAATGGTTACAGGACATACATTTAACTACCATGTTCTAACCCATTTGGTTATGCTCGCCGTATC
C	AATTTTATTTTTTAACCTAACTCCCCTACTAAGTGTACCCCCCCTTTCCCCCCCCAGGGGGGGTATACTATGCATAATCGTGCATACATTTATATACCACAT ATATTATGGTACCGGTAATATATACTATATATGTACTAAACCCATTATATGTATACGGGCATTAATCTATATTCCACATTTCTCCCAATGTCCATTCTATGC ATGATCCAAGACATACTCATTCACCCTCCCCATAGACAGTTCTAAACCACTATCAAGCCACCTAACTATGAATGGTTACAGGACATAAATCTCACTCTCA TGTTCTCCCCCTAACAAGTCACCTAACTATGAATGGTTACAGGACATACATTTAACTACCATGTTCTAACCCATTTGGTTATGCTCGCCGTATC
D	AATTTTATTTTTTAACCTAACTCCCCTACTAAGTGTACCCCCCCTTTCCCCCCCCAGGGGGGGTATACTATGCATAATCGTGCATACATTTATATACCACAT ATATTATGGTACCGGTAATATATACTATATATGTACTAAACCCATTATATGTATACGGGCATTAATCTATATTCCACATTTCTCCCAATGTCCATTCTATGC ATGATCCAGGACACACTCATTCACCCTCCCCATAGACAGCTCCAAACCACTACCAAGTCACCTAACTATGAATGGTTACAGGACATAAATCTCACTCTCA TGTTCTTCCCCCAACAAGTCACCTAACTATGAATGGTTACAGGACATACATTTAACTACCATGTTCTAACCCATTTGGTTATGCTCGCCGTATC
E	AATTTTATTTTTTAACCTAACTCCCCTACTAAGTGTACCCCCCCTTTCCCCCCCCAGGGGGGGTATACTATGCATAATCGTGCATACATTTATATACCACAT ATATTATGGTACCGGTAATATATACTATATATGTACTAAACCCATTATATGTATACGGGCATTAATCTATATTCCACATTTCTCCCAATGTCCATTCTATGC ATGATCCAGGACATACTCATTCACCCTCCCTACAGACAGCTCCAAACCACCACCAAGTCACCTAACTATGAATGGTTACAGGACATAAATCTCACTCTCA TGTTCTTCCCCTACCAAGTCACCTAACTATGAATGGTTACAGGACATACATTTAACTACCATGTTCTAACCCATTTGGTTATGCTCGCCGTATC
F	AATTTTATTTTTTAACCTAACTCCCCTACTAAGTGTACCCCCCCTTTCCCCCCCCAGGGGGGGTATACTATGCATAATCGTGCATACATTTATATACCACA TATATTATGGTACCGGTAATATATACTATATATGTACTAAACCCATTATATGTATACGGGCATTAATCTATATTCCACATTTCTCCCAATGTCCATTCTATG CATGATCCAAGACATACTCATTCACCCTCCCCATAGACAGCTCCAAACCACTACCAAGTCACCTAACTATGAATGGTTACAGGACATAAATCTAACTCTT ATGTTCTTCCCCTAACAAGCCACCTAACTATGAATGGTTACAGGACATACATTTAACTACCATGTTCTAACCCATTTGGTTATGCTCGCCGTATC

The nomenclature of the haplogroup can be found in Mwacharo, J., Bjørnstad, G., Mobegi, V., Nomura, K., Hanada, H., Amano, T., *et al.* (2011). Mitochondrial DNA reveals multiple introductions of domestic chicken in East Africa. *Molecular phylogenetics and evolution, 58*(2), 374-382.

Table S4A. Candidate selective sweep regions detected in Ethiopian domestic chicken population using Hp. **Chr**: Chromosome. **Ln (kb)**: Fragment size for the sweep regions in kilobase pair. **nWd**: Number of windows analysed for the sweep region. **SNP**: Average number of SNPs present in the sweep region. **u*ZH*p:** Mean number of Z-transformed pool heterozygosity.

Chr	**Start**	**Stop**	**Ln (kb)**	**nWd**	**SNP**	**u*H*p**	**u*ZH*p**	**Gene**
1	25440000	25460000	20	1	78	0.1	-4.18	
1	32430000	32450000	20	1	146	0.09	-4.34	
1	32480000	32510000	30	2	130	0.09 ± 0.008	-4.27 ± 0.164	
1	32580000	32620000	40	3	112	0.07 ± 0.025	-4.65 ± 0.497	
1	58890000	58910000	20	1	194	0.1	-4.1	
1	82370000	82390000	20	1	121	0.1	-4.03	
1	83290000	83310000	20	1	252	0.1	-4.07	
1	85220000	85240000	20	1	268	0.1	-4.03	
1	115620000	115640000	20	1	202	0.1	-4.13	*IL1RAPL1*
1	116020000	116120000	100	8	64	0.09 ± 0.007	-4.25 ± 0.14	
1	116200000	116240000	40	3	81	0.1 ± 0.008	-4.2 ± 0.161	
1	118940000	118970000	30	2	110	0.09 ± 0.012	-4.36 ± 0.239	
1	145950000	145970000	20	1	136	0.1	-4.02	*GPC6*
1	190070000	190110000	40	3	205	0.09 ± 0.012	-4.23 ± 0.227	
2	9110000	9130000	20	1	145	0.08	-4.48	*PTPRN2*
2	60840000	60880000	40	3	173	0.08 ± 0.024	-4.58 ± 0.477	
2	61100000	61130000	30	2	228	0.1 ± 0.004	-4.1 ± 0.081	
2	70360000	70380000	20	1	133	0.11	-4	
2	70450000	70500000	50	4	138	0.1 ± 0.005	-4.14 ± 0.092	
2	70650000	70680000	30	2	196	0.1 ± 0.003	-4.06 ± 0.06	
2	70730000	70750000	20	1	78	0.1	-4.13	
2	70840000	70880000	40	3	131	0.1 ± 0.004	-4.16 ± 0.083	
2	77810000	77860000	50	4	116	0.1 ± 0.005	-4.05 ± 0.1	*CTNND2*
2	78310000	78340000	30	2	117	0.09 ± 0.023	-4.38 ± 0.455	
2	118570000	118590000	20	1	98	0.09	-4.3	
2	121660000	121700000	40	3	319	0.1 ± 0.006	-4.09 ± 0.121	

2	138730000	138750000	20	1	124	0.1	-4.05	
2	139930000	139950000	20	1	105	0.09	-4.32	
2	139970000	140020000	50	4	141	0.06 ± 0.037	-4.85 ± 0.728	
2	146720000	146740000	20	1	109	0.03	-5.4	
3	23760000	23780000	20	1	86	0.11	-4	
3	54860000	54880000	20	1	99	0.1	-4.21	*AHI1*
3	60820000	60840000	20	1	108	0.09	-4.25	
3	78890000	78910000	20	1	280	0.09	-4.26	*HMGN3*
3	84850000	84870000	20	1	58	0.1	-4.18	*EYS*
3	85680000	85720000	40	3	129	0.1 ± 0.005	-4.19 ± 0.09	*KHDRBS2*
3	93670000	93690000	20	1	162	0.09	-4.21	
3	101900000	101950000	50	4	211	0.09 ± 0.008	-4.29 ± 0.156	*APOB*
3	102150000	102210000	60	5	86	0.09 ± 0.004	-4.35 ± 0.087	
3	102260000	102290000	30	2	123	0.1 ± 0.001	-4.08 ± 0.028	
3	102360000	102470000	110	10	83	0.1 ± 0.007	-4.17 ± 0.129	*gga-mir-6678*
3	102490000	102520000	30	2	67	0.09 ± 0.006	-4.28 ± 0.126	
3	102570000	102660000	90	8	124	0.08 ± 0.015	-4.45 ± 0.299	
3	102720000	102870000	150	12	91	0.06 ± 0.033	-4.96 ± 0.651	
4	7890000	7970000	80	7	105	0.08 ± 0.02	-4.51 ± 0.392	
4	27110000	27130000	20	1	129	0.09	-4.23	
4	38620000	38660000	40	3	213	0.08 ± 0.018	-4.41 ± 0.346	*TACR3*
4	75220000	75240000	20	1	107	0.09	-4.24	
4	75400000	75420000	20	1	108	0.09	-4.35	*LCORL*
4	77130000	77180000	50	4	154	0.04 ± 0.04	-5.21 ± 0.778	
4	77410000	77490000	80	7	131	0.1 ± 0.008	-4.07 ± 0.163	
4	81040000	81060000	20	1	205	0.1	-4.11	
5	10780000	10800000	20	1	123	0.1	-4.07	*SOX6*
5	21810000	21850000	40	3	66	0.05 ± 0.021	-4.99 ± 0.42	
5	30440000	30490000	50	4	81	0.09 ± 0.007	-4.22 ± 0.127	
5	39900000	39920000	20	1	131	0.1	-4.14	
5	40060000	40110000	50	4	166	0.01 ± 0.015	-5.8 ± 0.289	*TSHR, GTF2A1*

5	40250000	40310000	60	5	166	0.07 ± 0.022	-4.71 ± 0.438	
5	41050000	41140000	90	8	92	0.1 ± 0.027	-4.15 ± 0.537	
5	55190000	55210000	20	1	338	0.1	-4.15	*C14orf37*
7	7770000	7790000	20	1	95	0.1	-4.04	*TMEFF2*
7	7860000	7920000	60	5	96	0.13 ± 0.054	-4.0 ± 1.051	
7	8070000	8090000	20	1	100	0.08	-4.45	
7	15860000	15920000	60	5	260	0.08 ± 0.01	-4.43 ± 0.204	*KIAA1715*
7	16480000	16530000	50	4	183	0.06 ± 0.009	-4.86 ± 0.18	*OLA1*
7	17880000	17900000	20	1	229	0.1	-4.1	*MYO3B*
8	10000	70000	60	5	100	0.1 ± 0.005	-4.16 ± 0.09	*AMY2A*
8	110000	140000	30	2	53	0.1 ± 0.001	-4.03 ± 0.021	
8	160000	270000	110	10	70	0.09 ± 0.015	-4.34 ± 0.287	
8	370000	390000	20	1	53	0.1	-4.04	
8	420000	630000	210	19	116	0.09 ± 0.003	-4.29 ± 0.055	
8	8740000	8760000	20	1	221	0.03	-5.46	
8	8810000	8830000	20	1	299	0.1	-4.03	
8	8900000	8920000	20	1	292	0.09	-4.38	
8	9050000	9130000	80	7	125	0.08 ± 0.013	-4.58 ± 0.262	
8	9420000	9450000	30	2	150	0.08 ± 0.01	-4.45 ± 0.188	
9	9630000	9680000	50	4	144	0.07 ± 0.012	-4.63 ± 0.23	*SLC16A14*
9	9740000	9790000	50	4	185	0.09 ± 0.013	-4.32 ± 0.255	*RNF7, GRK7, ATP1B3*
9	9870000	9900000	30	2	213	0.07 ± 0.019	-4.67 ± 0.379	*GK5*
9	10960000	10990000	30	2	194	0.09 ± 0.002	-4.31 ± 0.035	*PLOD5*
9	11860000	11880000	20	1	135	0.06	-4.8	*AGTR1*
12	4350000	4420000	70	6	161	0.08 ± 0.01	-4.55 ± 0.195	*ATG7, HRH1*
12	10490000	10520000	30	2	209	0.08 ± 0.018	-4.43 ± 0.357	*SLC41A3*
23	5460000	5490000	30	2	67	0.1 ± 0.013	-4.2 ± 0.254	*HPCAL4, TRIT1, MYCL*
24	6140000	6160000	20	1	199	0.08	-4.41	*BCDO2*

Table S4B. Candidate selective sweep regions detected in Saudi Arabian domestic chicken population using Hp. **Chr**: Chromosome. **Ln (kb)**: Fragment size for the sweep regions in kilobase pair. **nWd**: Number of windows analysed for the sweep region. **SNP**: Average number of SNPs present in the sweep region. **u*ZH*p:** Mean number of Z-transformed pool heterozygosity.

Chr	**Start**	**Stop**	**Ln (kb)**	**nWd**	**SNP**	**u*H*p**	**u*ZH*p**	**Gene**
1	8510000	8540000	30	2	54	0.06 ± 0.072	-6.19 ± 1.48	*SEMA3A*
1	9050000	9070000	20	1	76	0.15	-4.24	
1	20820000	20850000	30	2	125	0.14 ± 0.008	-4.47 ± 0.171	*GRM8*
1	20960000	21000000	40	3	101	0.13 ± 0.005	-4.77 ± 0.107	*GRM8*
1	42470000	42490000	20	1	50	0.15	-4.36	
1	43140000	43170000	30	2	136	0.11 ± 0.014	-5.05 ± 0.296	
1	46610000	46640000	30	2	176	0.15 ± 0.012	-4.23 ± 0.247	*ANKS1B*
1	55950000	55990000	40	3	147	0.14 ± 0.017	-4.46 ± 0.343	*TBXAS1*
1	56010000	56040000	30	2	167	0.14 ± 0.02	-4.5 ± 0.415	*TBXAS1*
1	56730000	56750000	20	1	140	0.16	-4.01	*AKR1D1*
1	58390000	58430000	40	3	172	0.14 ± 0.001	-4.52 ± 0.03	*MTPN*
1	58680000	58730000	50	4	124	0.1 ± 0.06	-5.28 ± 1.244	
1	58890000	58910000	20	1	187	0.13	-4.74	
1	65060000	65080000	20	1	108	0.16	-4.05	*PDE3A*
1	66600000	66620000	20	1	198	0.15	-4.28	*ETNK1*
1	70150000	70180000	30	2	167	0.14 ± 0.029	-4.5 ± 0.593	*PHF21B*
1	71980000	72020000	40	3	238	0.12 ± 0.027	-4.97 ± 0.565	*MANSC1*
1	72040000	72110000	70	6	138	0.13 ± 0.05	-4.79 ± 1.04	*BORCS5, DUSP16*
1	72130000	72210000	80	7	132	0.11 ± 0.054	-5.1 ± 1.121	*CREBL2, GPR19*
1	75410000	75430000	20	1	81	0.16	-4.16	*AICDA, MFAP5*
1	75460000	75500000	40	3	78	0.16 ± 0.003	-4.19 ± 0.068	*RIMKLB*
1	82380000	82400000	20	1	94	0.11	-5.04	*DRD3*
1	96210000	96230000	20	1	50	0.16	-4.09	
1	112840000	112860000	20	1	142	0.12	-4.83	
1	122810000	122880000	70	6	156	0.13 ± 0.061	-4.75 ± 1.252	*TLR7, PRPS2*
1	123700000	123770000	70	6	113	0.15 ± 0.036	-4.25 ± 0.742	*ARHGAP6*

1	126500000	126580000	80	6	72	0.13 ± 0.022	-4.79 ± 0.447	
1	126910000	126940000	30	2	106	0.15 ± 0.003	-4.22 ± 0.053	
1	127140000	127160000	20	1	95	0.16	-4.19	
1	136170000	136190000	20	1	136	0.12	-4.98	*Sep-10*
1	139970000	139990000	20	1	53	0.16	-4.08	
1	140980000	141020000	40	3	112	0.14 ± 0.022	-4.46 ± 0.453	
1	141160000	141180000	20	1	86	0.16	-4.04	
1	156100000	156120000	20	1	164	0.14	-4.45	*DACH1*
1	160120000	160140000	20	1	67	0.03	-6.79	
1	167780000	167830000	50	4	212	0.17 ± 0.016	-4.0 ± 0.321	
1	167930000	167950000	20	1	208	0.16	-4.06	*SUCLA2*
1	178290000	178360000	70	6	128	0.13 ± 0.031	-4.67 ± 0.65	*ZMYM2*
1	190030000	190100000	70	6	114	0.16 ± 0.029	-4.02 ± 0.59	
1	190120000	190140000	20	1	70	0.09	-5.54	
2	2250000	2280000	30	2	260	0.14 ± 0.019	-4.42 ± 0.4	*OBSCN*
2	8230000	8260000	30	2	82	0.1 ± 0.069	-5.34 ± 1.426	
2	18920000	18940000	20	1	96	0.15	-4.28	
2	19950000	19970000	20	1	167	0.14	-4.53	*RSU1*
2	21680000	21710000	30	2	250	0.14 ± 0.002	-4.58 ± 0.037	*CDK14*
2	24200000	24220000	20	1	152	0.16	-4.03	*DYNC1I1*
2	26210000	26270000	60	5	111	0.12 ± 0.032	-4.93 ± 0.661	*PHF14*
2	26530000	26580000	50	4	91	0.12 ± 0.038	-4.85 ± 0.796	*THSD7A*
2	28190000	28220000	30	2	74	0.15 ± 0.015	-4.24 ± 0.312	*MEOX2*
2	28250000	28290000	40	3	91	0.14 ± 0.023	-4.5 ± 0.483	*5S_rRNA*
2	35850000	35890000	40	3	56	0.16 ± 0.01	-4.19 ± 0.209	
2	35910000	35930000	20	1	108	0.14	-4.52	
2	39300000	39320000	20	1	141	0.15	-4.25	
2	39360000	39380000	20	1	178	0.14	-4.56	*CMC1*
2	41420000	41460000	40	3	134	0.1 ± 0.032	-5.3 ± 0.653	*ABHD5*
2	43490000	43520000	30	2	320	0.16 ± 0.009	-4.19 ± 0.189	*LARS2*
2	55240000	55260000	20	1	136	0.15	-4.37	*TNS3*

2	58240000	58270000	30	2	128	0.14 ± 0.003	-4.46 ± 0.053	
2	58560000	58580000	20	1	103	0.09	-5.46	
2	60860000	60900000	40	3	171	0.14 ± 0.02	-4.45 ± 0.405	
2	63110000	63140000	30	2	203	0.15 ± 0.003	-4.25 ± 0.063	*C1orf145*
2	65860000	65900000	40	3	148	0.12 ± 0.032	-4.91 ± 0.656	
2	70850000	70870000	20	1	56	0.15	-4.22	
2	74290000	74310000	20	1	97	0.16	-4.18	*CDH18*
2	75140000	75280000	140	13	159	0.13 ± 0.029	-4.8 ± 0.605	
2	81900000	81920000	20	1	51	0.15	-4.25	
2	83160000	83200000	40	3	120	0.13 ± 0.028	-4.77 ± 0.588	*FHOD3*
2	83460000	83490000	30	2	82	0.14 ± 0.007	-4.53 ± 0.141	*FHOD3*
2	86240000	86260000	20	1	128	0.15	-4.22	
2	86340000	86440000	100	9	106	0.1 ± 0.036	-5.36 ± 0.752	
2	92550000	92580000	30	2	144	0.15 ± 0.007	-4.29 ± 0.145	*NETO1, MIR1803*
2	95760000	95780000	20	1	74	0.16	-4.03	
2	103040000	103060000	20	1	268	0.12	-4.9	*IMPACT*
2	104610000	104640000	30	2	144	0.15 ± 0.017	-4.32 ± 0.344	
2	121820000	121840000	20	1	245	0.15	-4.24	
2	122030000	122090000	60	5	249	0.17 ± 0.033	-4.0 ± 0.687	
2	125030000	125050000	20	1	96	0.06	-6.14	
2	138730000	138760000	30	2	155	0.14 ± 0.009	-4.59 ± 0.191	
2	139130000	139150000	20	1	180	0.14	-4.42	
2	140870000	140970000	100	8	180	0.11 ± 0.018	-5.08 ± 0.382	*EFR3A*
2	141000000	141070000	70	6	123	0.12 ± 0.057	-4.97 ± 1.182	*KCNQ3*
2	146610000	146630000	20	1	85	0.15	-4.38	
2	146690000	146800000	110	10	101	0.07 ± 0.052	-6.03 ± 1.067	
2	148640000	148690000	50	4	208	0.14 ± 0.019	-4.51 ± 0.384	*SCRIB*
3	17530000	17590000	60	5	163	0.11 ± 0.034	-5.14 ± 0.698	
3	18920000	18940000	20	1	75	0.15	-4.23	
3	20610000	20630000	20	1	84	0.16	-4.1	*PROX1*
3	23740000	23760000	20	1	187	0.16	-4.18	

3	33710000	33760000	50	4	100	0.04 ± 0.05	-6.49 ± 1.027	*KIF26B*
3	45690000	45720000	30	2	135	0.14 ± 0.027	-4.44 ± 0.564	
3	51940000	51970000	30	2	166	0.14 ± 0.008	-4.48 ± 0.166	
3	52710000	52800000	90	5	64	0.14 ± 0.017	-4.45 ± 0.351	
3	52830000	52850000	20	1	56	0.15	-4.25	
3	52910000	52930000	20	1	163	0.14	-4.43	
3	60860000	60880000	20	1	58	0.16	-4.04	
3	63510000	63530000	20	1	81	0.14	-4.47	*FRK*
3	69200000	69220000	20	1	118	0.15	-4.32	
3	71210000	71250000	40	3	160	0.15 ± 0.011	-4.29 ± 0.221	
3	71470000	71490000	20	1	124	0.15	-4.37	
3	72460000	72480000	20	1	68	0.12	-4.99	
3	72500000	72530000	30	2	126	0.15 ± 0.011	-4.27 ± 0.235	
3	74500000	74520000	20	1	225	0.16	-4.12	*MAP3K7*
3	78360000	78380000	20	1	95	0.14	-4.45	
3	78890000	78910000	20	1	211	0.16	-4.1	*HMGN3*
3	79180000	79230000	50	4	124	0.1 ± 0.045	-5.28 ± 0.932	
3	79910000	79950000	40	3	217	0.1 ± 0.05	-5.25 ± 1.041	*MYO6*
3	80180000	80200000	20	1	210	0.15	-4.39	*TMEM30A, COX7A2*
3	80740000	80760000	20	1	81	0.15	-4.36	
3	81610000	81630000	20	1	260	0.1	-5.32	*RIMS1*
3	83220000	83320000	100	8	98	0.13 ± 0.009	-4.63 ± 0.177	
3	83370000	83450000	80	7	76	0.12 ± 0.01	-4.98 ± 0.216	
3	83500000	83520000	20	1	80	0.15	-4.2	
3	83590000	83650000	60	5	73	0.11 ± 0.009	-5.19 ± 0.192	
3	84050000	84070000	20	1	122	0.16	-4.03	
3	87580000	87630000	50	4	202	0.14 ± 0.023	-4.53 ± 0.476	*GCLC*
3	93120000	93140000	20	1	160	0.16	-4.1	
3	93160000	93240000	80	6	132	0.11 ± 0.028	-5.07 ± 0.581	*TSSC1, TRAPPC12*
3	97010000	97050000	40	3	151	0.15 ± 0.013	-4.2 ± 0.276	*LPIN1*
3	101360000	101390000	30	2	154	0.14 ± 0.026	-4.46 ± 0.532	*RHOB*

3	102690000	102760000	70	6	118	0.14 ± 0.036	-4.41 ± 0.749	
4	6960000	6990000	30	2	120	0.13 ± 0.052	-4.79 ± 1.074	
4	26410000	26430000	20	1	239	0.16	-4.17	
4	27110000	27140000	30	2	123	0.14 ± 0.01	-4.55 ± 0.21	
4	27160000	27190000	30	2	80	0.16 ± 0.002	-4.05 ± 0.043	
4	27380000	27470000	90	7	99	0.09 ± 0.032	-5.45 ± 0.656	
4	27860000	27910000	50	4	123	0.15 ± 0.016	-4.29 ± 0.331	
4	28120000	28240000	120	11	123	0.12 ± 0.084	-4.93 ± 1.743	*PCDH18*
4	32790000	32820000	30	2	163	0.11 ± 0.059	-5.19 ± 1.214	
4	41190000	41210000	20	1	53	0.16	-4.02	
4	54050000	54070000	20	1	185	0.11	-5.06	*PDE5A*
4	54640000	54660000	20	1	101	0.16	-4.09	
4	58500000	58530000	30	2	198	0.12 ± 0.015	-4.87 ± 0.31	
4	65280000	65300000	20	1	220	0.12	-5.02	*LNX1*
4	75380000	75410000	30	2	154	0.13 ± 0.021	-4.63 ± 0.427	*LCORL*
4	77110000	77170000	60	5	155	0.14 ± 0.03	-4.55 ± 0.625	
4	77410000	77440000	30	2	132	0.16 ± 0.001	-4.13 ± 0.014	
4	77600000	77630000	30	2	133	0.12 ± 0.011	-4.87 ± 0.229	*HS3ST1*
4	80840000	80900000	60	3	60	0.11 ± 0.035	-5.03 ± 0.73	
4	89900000	89930000	30	2	154	0.14 ± 0.004	-4.44 ± 0.073	
5	13300000	13320000	20	1	127	0.15	-4.27	
5	16940000	16980000	40	3	117	0.09 ± 0.02	-5.49 ± 0.423	
5	18260000	18300000	40	3	89	0.12 ± 0.049	-4.91 ± 1.017	*CD44*
5	39870000	39930000	60	5	72	0.06 ± 0.049	-6.2 ± 1.006	
5	40060000	40110000	50	4	197	0.08 ± 0.051	-5.72 ± 1.059	*TSHR, GTF2A1*
5	41100000	41150000	50	4	94	0.15 ± 0.007	-4.2 ± 0.146	
5	41450000	41470000	20	1	99	0.16	-4.07	*FLRT2*
5	47950000	47970000	20	1	195	0.15	-4.25	*CCDC85C*
5	51720000	51750000	30	2	101	0.09 ± 0.06	-5.59 ± 1.24	*AHNAK2*
5	54080000	54120000	40	3	156	0.12 ± 0.028	-4.98 ± 0.577	
5	55190000	55230000	40	3	267	0.14 ± 0.014	-4.59 ± 0.285	*C14orf37*

6	5550000	5590000	40	3	328	0.11 ± 0.028	-5.12 ± 0.581	
6	5610000	5640000	30	2	279	0.05 ± 0.058	-6.34 ± 1.205	
6	6070000	6100000	30	2	314	0.1 ± 0.012	-5.41 ± 0.246	*SIRT1*
6	6480000	6510000	30	2	180	0.14 ± 0.03	-4.52 ± 0.621	*CTNNA3*
6	6860000	6920000	60	5	174	0.11 ± 0.04	-5.07 ± 0.821	
6	9640000	9690000	50	4	168	0.14 ± 0.01	-4.6 ± 0.216	
6	9700000	9750000	50	4	96	0.08 ± 0.046	-5.66 ± 0.957	*PRKG1*
6	12910000	12940000	30	2	61	0.15 ± 0.008	-4.28 ± 0.17	*KCNMA1*
6	15480000	15530000	50	4	129	0.1 ± 0.047	-5.38 ± 0.964	*CHCHD1, FUT11, SEC24C*
6	15550000	15590000	40	3	111	0.13 ± 0.02	-4.65 ± 0.42	*SYNPO2L, MYOZ1*
6	15610000	15640000	30	2	198	0.12 ± 0.003	-4.85 ± 0.068	
6	16960000	17020000	60	5	204	0.1 ± 0.033	-5.36 ± 0.685	*HIF1AN, RP11-411B6.1, SEC31B, WNT8B*
6	17860000	17890000	30	2	182	0.15 ± 0.003	-4.25 ± 0.064	*ARHGAP22*
7	290000	320000	30	2	126	0.13 ± 0.021	-4.68 ± 0.431	*OSGEPL1, ASNSD1*
7	430000	450000	20	1	58	0.14	-4.5	*COL5A2*
7	490000	510000	20	1	65	0.08	-5.78	
7	6090000	6110000	20	1	74	0.16	-4.03	
7	6370000	6420000	50	4	102	0.12 ± 0.006	-4.88 ± 0.125	*HDAC4*
7	8070000	8090000	20	1	74	0.12	-4.87	
7	11130000	11150000	20	1	142	0.16	-4.09	
7	11890000	11920000	30	2	183	0.12 ± 0.013	-4.84 ± 0.268	*FASTKD2, MDH1B, DYTN*
7	21060000	21100000	40	3	104	0.07 ± 0.081	-5.94 ± 1.68	
7	21240000	21260000	20	1	102	0.11	-5.2	*RBMS1*
7	24870000	24890000	20	1	157	0.14	-4.56	*PTPN4*
7	27580000	27600000	20	1	125	0.13	-4.8	*SNX4*
7	36160000	36190000	30	2	52	0.12 ± 0.013	-4.95 ± 0.275	*BAZ2B, Mar-07*
7	36200000	36230000	30	2	78	0.1 ± 0.013	-5.23 ± 0.262	*NAA20, Mar-07*
8	1	30000	30	2	80	0.13 ± 0.007	-4.71 ± 0.155	
8	200000	230000	30	2	60	0.01 ± 0.004	-7.27 ± 0.087	
8	420000	630000	210	19	113	0.13 ± 0.013	-4.78 ± 0.272	
8	7240000	7270000	30	2	174	0.16 ± 0.01	-4.16 ± 0.208	*RABGAP1L*

8	8730000	8920000	190	18	271	0.11 ± 0.036	-5.2 ± 0.743	
8	9020000	9130000	110	10	129	0.15 ± 0.014	-4.33 ± 0.295	
8	9330000	9350000	20	1	103	0.15	-4.28	
8	9420000	9450000	30	2	143	0.13 ± 0.012	-4.73 ± 0.25	
8	10880000	10900000	20	1	81	0.13	-4.75	*TRMT13, SASS6, HIAT1*
8	16040000	16120000	80	7	115	0.13 ± 0.024	-4.66 ± 0.486	
8	17920000	17950000	30	2	140	0.15 ± 0.008	-4.35 ± 0.16	*ZZZ3*
8	27370000	27390000	20	1	214	0.1	-5.32	
8	27810000	27850000	40	3	130	0.14 ± 0.002	-4.51 ± 0.048	*LRRC7*
9	9000000	9030000	30	2	100	0.13 ± 0.003	-4.63 ± 0.054	*SLC19A3, CCL20*
9	9190000	9220000	30	2	124	0.11 ± 0.045	-5.1 ± 0.933	
9	9810000	9910000	100	9	128	0.07 ± 0.042	-5.9 ± 0.874	*TFDP2, GK5, XRN1*
9	16860000	16900000	40	3	133	0.14 ± 0.019	-4.46 ± 0.392	*USP13*
9	19760000	19790000	30	2	90	0.09 ± 0.001	-5.63 ± 0.016	
9	22840000	22860000	20	1	109	0.15	-4.4	*MBNL1*
10	3900000	3920000	20	1	192	0.16	-4.01	*TLN2*
10	17980000	18010000	30	2	206	0.14 ± 0.002	-4.41 ± 0.036	*MAP2K1, SNAPC5, RPL4, ZWILCH*
11	1800000	1820000	20	1	227	0.13	-4.77	*GLG1*
11	3310000	3340000	30	2	86	0.16 ± 0.008	-4.14 ± 0.165	
11	3450000	3500000	50	4	68	0.15 ± 0.019	-4.34 ± 0.385	
11	5270000	5290000	20	1	108	0.15	-4.24	
11	9180000	9200000	20	1	138	0.16	-4.11	
11	12460000	12550000	90	7	65	0.13 ± 0.028	-4.72 ± 0.589	
11	12570000	12610000	40	3	72	0.14 ± 0.018	-4.49 ± 0.37	
11	14210000	14230000	20	1	84	0.14	-4.54	
14	3580000	3610000	30	2	213	0.16 ± 0.001	-4.16 ± 0.024	*SDK1*
14	10290000	10310000	40	2	101	0.13 ± 0.001	-4.72 ± 0.034	
14	10560000	10580000	80	6	132	0.16	-4.03	*CACNA1G, MIR163*
14	10760000	10790000	30	2	92	0.13 ± 0.029	-4.69 ± 0.592	
15	11550000	11570000	20	1	86	0.13	-4.75	
18	10500000	10540000	40	3	126	0.09 ± 0.074	-5.64 ± 1.538	

23	5460000	5490000	30	2	51	0.11 ± 0.004	-5.14 ± 0.075	*HPCAL4, TRIT1, MYCL*

Table S4C. Candidate selective sweep regions detected in Sri Lankan domestic chicken population using Hp. **Chr**: Chromosome. **Ln (kb)**: Fragment size for the sweep regions in kilobase pair. **nWd**: Number of windows analysed for the sweep region. **SNP**: Average number of SNPs present in the sweep region. **u*ZH*p:** Mean number of Z-transformed pool heterozygosity.

Chr	**Start**	**Stop**	**Ln (kb)**	**nWd**	**SNP**	**u*H*p**	**u*ZH*p**	**Gene**
1	8520000	8540000	20	1	274	0.16	-4.03	*SEMA3A*
1	18110000	18130000	20	1	145	0.16	-4.07	
1	25290000	25350000	60	5	216	0.14 ± 0.01	-4.51 ± 0.262	*TFEC*
1	25400000	25430000	30	2	243	0.15 ± 0	-4.35 ± 0.008	
1	25450000	25470000	20	1	227	0.15	-4.28	
1	25580000	25600000	20	1	321	0.14	-4.53	
1	26450000	26470000	20	1	270	0.16	-4.04	
1	27840000	27870000	30	2	319	0.14 ± 0.005	-4.55 ± 0.131	
1	32430000	32450000	20	1	186	0.12	-4.95	
1	42950000	42980000	30	2	146	0.14 ± 0.02	-4.55 ± 0.518	
1	43060000	43110000	50	4	84	0.09 ± 0.025	-5.96 ± 0.64	
1	43840000	43860000	20	1	103	0.15	-4.32	*KERA*
1	75840000	75880000	40	3	108	0.11 ± 0.036	-5.34 ± 0.932	*OVSTL, gga-mir-6606*
1	79430000	79450000	20	1	242	0.14	-4.43	
1	92710000	92730000	20	1	180	0.15	-4.19	*CADM2*
1	103010000	103040000	30	2	376	0.15 ± 0.003	-4.24 ± 0.069	
1	110720000	110750000	30	2	161	0.13 ± 0.027	-4.83 ± 0.687	
1	118020000	118040000	20	1	292	0.13	-4.8	
1	140850000	140870000	20	1	119	0.16	-4.05	
1	146980000	147000000	20	1	263	0.15	-4.23	*GPC5*
1	147930000	147950000	20	1	99	0.16	-4.05	
1	149380000	149410000	30	2	96	0.15 ± 0.006	-4.39 ± 0.148	
1	158080000	158100000	20	1	85	0.15	-4.17	
1	189930000	189990000	60	5	181	0.15 ± 0.016	-4.32 ± 0.415	
1	190070000	190110000	40	3	199	0.1 ± 0.014	-5.51 ± 0.364	
1	190200000	190240000	40	3	202	0.14 ± 0.018	-4.56 ± 0.463	

2	2640000	2660000	20	1	246	0.14	-4.45	
2	11890000	11910000	20	1	116	0.14	-4.52	
2	24350000	24380000	30	2	355	0.15 ± 0.007	-4.27 ± 0.188	
2	27990000	28010000	20	1	109	0.14	-4.53	*AGMO*
2	28260000	28290000	30	2	102	0.11 ± 0.028	-5.23 ± 0.722	*5S_rRNA*
2	36010000	36040000	30	2	240	0.14 ± 0.014	-4.48 ± 0.365	*KCNH8*
2	36170000	36190000	20	1	281	0.13	-4.73	
2	36200000	36220000	20	1	273	0.16	-4.06	
2	52310000	52330000	20	1	83	0.15	-4.28	*VOPP1*
2	52410000	52430000	20	1	67	0.14	-4.66	*EGFR*
2	52500000	52540000	40	3	82	0.15 ± 0.01	-4.18 ± 0.252	*EGFR*
2	52610000	52630000	20	1	116	0.15	-4.28	
2	52650000	52670000	20	1	70	0.16	-4.08	
2	52690000	52720000	30	2	103	0.15 ± 0.009	-4.39 ± 0.245	*SEC61G*
2	52870000	52890000	20	1	76	0.16	-4.03	*TPK1*
2	52980000	53050000	70	6	148	0.15 ± 0.01	-4.38 ± 0.252	*TPK1*
2	53090000	53180000	90	8	116	0.16 ± 0.012	-4.15 ± 0.323	*TPK1*
2	54720000	54740000	20	1	244	0.16	-4.16	
2	78310000	78340000	30	2	168	0.12 ± 0.01	-5.2 ± 0.259	
2	81900000	81920000	20	1	143	0.15	-4.23	
2	81950000	82240000	290	27	125	0.14 ± 0.018	-4.65 ± 0.454	
2	86710000	86730000	20	1	113	0.15	-4.33	
2	86790000	86820000	30	2	122	0.15 ± 0.006	-4.29 ± 0.166	
2	87440000	87460000	20	1	86	0.16	-4.08	
2	91940000	91960000	20	1	306	0.15	-4.2	*CYB5A*
2	95760000	95780000	20	1	108	0.16	-4.01	
2	108940000	108980000	40	3	199	0.13 ± 0.01	-4.93 ± 0.255	
2	118490000	118520000	30	2	126	0.15 ± 0.019	-4.35 ± 0.482	
2	140970000	141010000	40	3	199	0.15 ± 0.01	-4.32 ± 0.268	*HHLA1, KCNQ3*
2	146620000	146650000	30	2	175	0.12 ± 0.002	-5.12 ± 0.061	
2	146720000	146740000	20	1	199	0.13	-4.92	

3	1120000	1210000	90	8	111	0.14 ± 0.017	-4.59 ± 0.445	
3	23050000	23080000	30	2	166	0.13 ± 0.017	-4.87 ± 0.445	*EML4*
3	69020000	69060000	40	3	159	0.15 ± 0.008	-4.23 ± 0.2	
3	69250000	69270000	20	1	155	0.16	-4.1	
3	69370000	69390000	20	1	161	0.16	-4	
3	77550000	77570000	20	1	257	0.15	-4.17	
3	81600000	81630000	30	2	392	0.14 ± 0.007	-4.49 ± 0.178	*RIMS1*
3	99070000	99100000	30	2	202	0.15 ± 0.003	-4.29 ± 0.09	
3	110160000	110290000	130	12	263	0.14 ± 0.006	-4.45 ± 0.162	*EVA1A*
4	25240000	25260000	20	1	196	0.14	-4.47	
4	26410000	26430000	20	1	253	0.15	-4.2	
4	27860000	27910000	50	4	186	0.16 ± 0.013	-4.13 ± 0.341	
4	32410000	32440000	30	2	250	0.13 ± 0.016	-4.79 ± 0.409	*LRBA*
4	38590000	38640000	50	4	247	0.1 ± 0.034	-5.52 ± 0.874	*TACR3*
4	40630000	40650000	20	1	92	0.16	-4.15	
4	56900000	56920000	20	1	224	0.14	-4.59	
4	61190000	61230000	40	3	224	0.16 ± 0.013	-4 ± 0.332	*RP11-215A19.2*
4	61750000	61770000	20	1	218	0.14	-4.59	
4	72190000	72210000	20	1	314	0.15	-4.33	
4	77080000	77100000	20	1	263	0.16	-4.14	*RAB28*
5	3800000	3840000	40	3	64	0.16 ± 0.003	-4.05 ± 0.082	
5	4240000	4260000	20	1	84	0.14	-4.51	
5	21670000	21890000	220	19	71	0.11 ± 0.033	-5.4 ± 0.866	
5	21980000	22010000	30	2	202	0.13 ± 0.009	-4.69 ± 0.237	
5	30360000	30400000	40	3	105	0.15 ± 0.009	-4.34 ± 0.245	
5	39870000	39890000	20	1	164	0.15	-4.33	
5	40060000	40110000	50	4	187	0.07 ± 0.063	-6.32 ± 1.634	*TSHR, GTF2A1*
5	40880000	40910000	30	2	190	0.13 ± 0.043	-4.82 ± 1.111	
5	41100000	41160000	60	5	115	0.13 ± 0.033	-4.89 ± 0.862	
5	46470000	46490000	20	1	110	0.16	-4.12	
6	8780000	8800000	20	1	96	0.15	-4.26	*BKJ*

6	13310000	13330000	20	1	67	0.15	-4.28	*KCNMA1*
6	17860000	17950000	90	8	214	0.15 ± 0.008	-4.18 ± 0.197	*ARHGAP22*
6	18020000	18040000	20	1	257	0.16	-4.14	*WDFY4*
7	7890000	7930000	40	3	121	0.14 ± 0.015	-4.55 ± 0.388	
7	8070000	8100000	30	2	109	0.14 ± 0.024	-4.6 ± 0.629	
7	14810000	14830000	20	1	58	0.16	-4.17	*TTN*
7	36140000	36210000	70	6	72	0.13 ± 0.018	-4.86 ± 0.462	*BAZ2B, MARCH7, NAA20*
8	10630000	10650000	20	1	93	0.15	-4.3	
8	11330000	11360000	30	2	136	0.14 ± 0.019	-4.54 ± 0.484	*SNX7*
8	11420000	11450000	30	2	137	0.16 ± 0.004	-4.1 ± 0.102	
8	12050000	12070000	20	1	70	0.12	-5.06	
8	12090000	12140000	50	4	167	0.15 ± 0.012	-4.21 ± 0.305	
8	16000000	16030000	30	2	274	0.16 ± 0.005	-4.12 ± 0.135	
8	16050000	16070000	20	1	208	0.15	-4.34	
8	28540000	28580000	40	2	86	0.16 ± 0.002	-4.09 ± 0.016	*TYW3*
9	10960000	10980000	20	1	138	0.13	-4.84	*PLOD2*
10	5350000	5390000	40	3	187	0.14 ± 0.016	-4.47 ± 0.402	*APBA2*
11	10400000	10420000	20	1	255	0.15	-4.2	*KIAA0355*
11	12350000	12390000	40	3	199	0.13 ± 0.032	-4.82 ± 0.83	*CDH8*
11	12810000	12830000	20	1	185	0.15	-4.26	
11	14890000	14910000	20	1	222	0.15	-4.32	
11	18720000	18750000	30	2	95	0.14 ± 0.008	-4.61 ± 0.215	
12	16600000	16620000	20	1	138	0.15	-4.38	*PDZRN3*
18	10910000	10940000	30	2	52	0.14 ± 0.022	-4.59 ± 0.565	*SLC16A5, ARMC7, HN1*
18	11080000	11100000	20	1	51	0.15	-4.35	*TMEM94*
18	11120000	11140000	20	1	58	0.16	-4.1	*CASKIN2, MCHR1*
22	230000	300000	70	6	133	0.13 ± 0.016	-4.9 ± 0.403	*ANTXR1, GKN2, BMP10, ARHGAP25*
22	560000	580000	20	1	56	0.15	-4.24	*PPP2R2A*
22	630000	650000	20	1	68	0.16	-4.09	*EBF2*
22	1040000	1070000	30	2	87	0.13 ± 0.009	-4.75 ± 0.227	*ADAM28*
22	1110000	1140000	30	2	58	0.15 ± 0.006	-4.2 ± 0.158	

22	1220000	1250000	30	2	135	0.15 ± 0.009	-4.19 ± 0.246	*LOXL2, R3HCC1*
22	1590000	1610000	20	1	113	0.16	-4.07	*SLC18A1, MAK16, TTI2, RNF122*
23	5420000	5490000	70	6	88	0.14 ± 0.028	-4.66 ± 0.723	*PABPC4, HEYL, NT5C1A, HPCAL4, TRIT1, MYCL*
26	160000	190000	30	2	58	0.15 ± 0.012	-4.33 ± 0.306	*MAPK13, BRPF3*
26	5230000	5250000	20	1	62	0.15	-4.23	*ATP2B4*
26	5290000	5310000	20	1	62	0.15	-4.4	
27	4970000	5040000	70	4	58	0.14 ± 0.01	-4.67 ± 0.262	*MLX, PSMC3IP, FAM134C, TUBG1, PLEKHH3, CNTNAP1, EZH1, RAMP2, WNK4, CCR10*
27	5070000	5110000	40	3	66	0.14 ± 0.015	-4.51 ± 0.379	*AARSD1, RUNDC1, RPL27, IFI35*

Table S5. Candidate selective sweep regions detected in red junglefowl population using Hp. **Chr**: Chromosome. **Ln (kb)**: Fragment size for the sweep regions in kilobase pair. **nWd**: Number of window analysed for the sweep region. **SNP**: Average number of SNPs present in the sweep region. **u*ZH*p:** Mean number of Z-transformed pool heterozygosity.

Chr	**Start**	**Stop**	**Ln (kb)**	**nWd**	**uSNP**	***H*p**	**u*ZH*p**	**Gene**
1	4950000	4970000	20	1	168	0.19	-4.34	
1	16750000	16770000	20	1	181	0.19	-4.59	
1	17290000	17310000	20	1	59	0.19	-4.45	*FAM19A5*
1	17560000	17610000	50	4	147	0.19 ± 0.008	-4.48 ± 0.27	
1	18060000	18090000	30	2	156	0.18 ± 0.002	-4.95 ± 0.065	
1	41210000	41240000	30	2	286	0.18 ± 0.007	-5.01 ± 0.25	
1	41720000	41790000	70	6	104	0.21 ± 0.036	-3.64 ± 1.257	*LRRIQ1, ALX1*
1	41880000	41910000	30	2	80	0.2 ± 0.002	-4.21 ± 0.062	
1	42210000	42300000	90	8	151	0.19 ± 0.007	-4.43 ± 0.255	*MGAT4C*
1	42380000	42410000	30	2	151	0.2 ± 0.004	-4.12 ± 0.142	
1	42980000	43010000	30	2	111	0.19 ± 0.001	-4.36 ± 0.05	*FAR2*
1	73000000	73020000	20	1	194	0.2	-4.18	*FAR2*
1	73320000	73340000	20	1	242	0.19	-4.65	*CCND2*
1	73630000	73700000	70	6	83	0.2 ± 0.006	-4.14 ± 0.228	*KCNA1*
1	73760000	73800000	40	3	106	0.19 ± 0.009	-4.42 ± 0.309	*KCNA5*
1	73990000	74010000	20	1	68	0.2	-4.14	*NTF3*
1	93150000	93180000	30	2	174	0.18 ± 0.011	-4.86 ± 0.39	
1	96500000	96530000	30	2	154	0.2 ± 0.007	-4.19 ± 0.258	
1	100180000	100210000	30	2	256	0.19 ± 0.009	-0.44 ± 0.333	*NCAM2*
1	100940000	101010000	70	6	264	0.17 ± 0.026	-5.32 ± 0.905	
1	102620000	102640000	20	1	136	0.2	-4.04	
1	116290000	116310000	20	1	126	0.2	-4.05	
1	125890000	125910000	20	1	142	0.2	-4.22	*PUDP*
1	126080000	126100000	20	1	140	0.19	-4.31	
1	133830000	133870000	40	3	151	0.2 ± 0.002	-4.22 ± 0.085	
1	140470000	140530000	60	5	145	0.19 ± 0.014	-4.45 ± 0.496	

1	140990000	141050000	60	5	163	0.19 ± 0.012	-4.59 ± 0.43	
1	141550000	141600000	50	4	174	0.2 ± 0.005	-4.17 ± 0.19	
1	141730000	141760000	30	2	240	0.18 ± 0.037	-5.01 ± 1.307	
1	147430000	147450000	20	1	150	0.2	-4.17	
1	147500000	147530000	30	2	108	0.2 ± 0.005	-4.25 ± 0.161	
1	149980000	150000000	20	1	117	0.19	-4.51	
1	150800000	150830000	30	2	230	0.2 ± 0.001	-4.07 ± 0.047	
1	162740000	162770000	30	2	110	0.2 ± 0.005	-4.25 ± 0.16	*PCDH17*
1	163380000	163410000	30	2	208	0.2 ± 0.002	-4.11 ± 0.086	
2	30000	140000	110	10	65	0.2 ± 0.01	-4.03 ± 0.358	*SMARCD3, CHPF2, ABCF2*
2	11110000	11130000	20	1	154	0.2	-4.07	
2	11630000	11650000	20	1	214	0.2	-4.01	
2	12120000	12140000	20	1	189	0.2	-4.08	
2	25810000	25830000	20	1	126	0.2	-4.16	
2	27290000	27310000	20	1	139	0.2	-4.04	
2	35840000	35920000	80	7	109	0.17 ± 0.021	-5.28 ± 0.728	
2	36750000	36770000	20	1	201	0.2	-4	
2	36860000	36890000	30	2	154	0.2 ± 0	-4.12 ± 0.012	*ZNF385D*
2	50110000	50130000	20	1	139	0.19	-4.44	*POU6F2*
2	54830000	54850000	20	1	210	0.18	-4.98	*ADCY1*
2	70370000	70400000	30	2	148	0.19 ± 0.001	-4.47 ± 0.02	
2	70640000	70670000	30	2	154	0.19 ± 0.002	-4.43 ± 0.075	
2	70750000	70770000	20	1	132	0.2	-4.13	
2	70850000	70870000	20	1	126	0.19	-4.37	
2	71030000	71050000	20	1	179	0.2	-4.06	
2	71180000	71200000	20	1	192	0.2	-4.28	
2	72710000	72740000	30	2	163	0.19 ± 0.013	-4.38 ± 0.453	
2	73980000	74010000	30	2	203	0.19 ± 0.003	-4.54 ± 0.091	
2	82420000	82440000	20	1	131	0.2	-4.08	
2	82470000	82490000	20	1	117	0.2	-4.12	
2	82510000	82530000	20	1	151	0.2	-4	

2	82600000	82650000	50	4	187	0.19 ± 0.002	-4.32 ± 0.065	
2	86210000	86280000	70	6	177	0.18 ± 0.029	-4.68 ± 1.012	
2	86350000	86380000	30	2	178	0.18 ± 0.023	-4.85 ± 0.83	
2	86470000	86500000	30	2	119	0.2 ± 0.005	-4.25 ± 0.166	*5S_rRNA*
2	92490000	92510000	20	1	207	0.2	-4.07	*NETO1*
2	92550000	92570000	20	1	203	0.17	-5.03	*MIR1803*
2	93440000	93470000	30	2	156	0.2 ± 0.004	-4.16 ± 0.146	
2	93900000	93920000	20	1	117	0.2	-4.27	*DOK6*
2	94880000	94910000	30	2	203	0.19 ± 0	-4.41 ± 0.007	
2	94940000	95010000	70	6	169	0.18 ± 0.027	-4.85 ± 0.957	
2	95010000	95060000	50	4	153	0.2 ± 0.004	-4.16 ± 0.136	
2	95090000	95120000	30	2	122	0.2 ± 0.001	-4.21 ± 0.047	
2	95750000	95790000	40	3	149	0.18 ± 0.017	-4.89 ± 0.593	
2	105180000	105230000	50	4	171	0.19 ± 0.006	-4.47 ± 0.213	
2	108550000	108570000	20	1	234	0.2	-4.19	
2	119050000	119070000	20	1	96	0.18	-4.88	*ZFHX4*
2	119430000	119450000	20	1	207	0.19	-4.35	
2	132480000	132500000	20	1	110	0.2	-4.03	
2	132540000	132560000	20	1	88	0.2	-4.08	
2	132600000	132620000	20	1	101	0.2	-4.2	
2	132650000	132670000	20	1	129	0.2	-4.17	
2	133630000	133670000	40	3	155	0.18 ± 0.018	-4.91 ± 0.63	
2	133820000	133840000	20	1	111	0.19	-4.45	
3	1680000	1700000	20	1	120	0.2	-4.12	
3	52600000	52620000	20	1	142	0.19	-4.43	
3	52820000	52840000	20	1	160	0.2	-4.03	
3	59920000	59950000	30	2	126	0.19 ± 0.003	-4.31 ± 0.09	*NKAIN2*
3	69730000	69750000	20	1	124	0.19	-4.48	*GRIK2*
3	69850000	69880000	30	2	135	0.2 ± 0.001	-4.12 ± 0.046	*GRIK2*
3	71740000	71770000	30	2	218	0.2 ± 0	-4.22 ± 0.016	
3	82760000	82780000	20	1	198	0.19	-4.33	*BAI3*

3	83840000	83860000	20	1	192	0.2	-4.04	
3	84040000	84070000	30	2	180	0.18 ± 0.01	-4.76 ± 0.347	
3	84590000	84660000	70	6	156	0.19 ± 0.006	-4.36 ± 0.198	
3	89190000	89210000	20	1	167	0.2	-4.28	*CSMD1*
3	110220000	110280000	60	5	110	0.2 ± 0.004	-4.23 ± 0.135	*EVA1A*
4	18220000	18270000	50	4	76	0.21 ± 0.006	-3.88 ± 0.208	*AFF2*
4	18460000	18510000	50	4	54	0.21 ± 0.01	-3.76 ± 0.369	
4	19730000	19750000	20	1	177	0.19	-4.41	
4	19990000	20020000	30	2	236	0.19 ± 0.018	-4.45 ± 0.635	*NPY2R*
4	20200000	20240000	40	3	196	0.18 ± 0.008	-4.80 ± 0.269	*GUCY1A3*
4	20440000	20460000	20	1	180	0.19	-4.38	
4	25730000	25750000	20	1	147	0.2	-4.2	
4	27210000	27270000	60	5	189	0.18 ± 0.008	-4.78 ± 0.272	
4	27700000	27730000	30	2	100	0.2 ± 0.001	-4.11 ± 0.027	
4	27750000	27800000	50	4	154	0.2 ± 0.005	-4.11 ± 0.16	
4	27800000	27820000	20	1	152	0.2	-4.03	
4	27950000	27990000	40	3	175	0.19 ± 0.012	-4.34 ± 0.408	
4	40500000	40530000	30	2	194	0.2 ± 0	-4.1 ± 0.003	
4	41190000	41210000	20	1	112	0.2	-4.14	
4	41440000	41460000	20	1	166	0.2	-4.06	
4	47080000	47130000	50	4	352	0.16 ± 0.009	-5.5 ± 0.309	
4	63130000	63150000	20	1	176	0.2	-4.07	
4	63630000	63650000	20	1	255	0.18	-4.78	*SGCZ*
4	71460000	71510000	50	4	143	0.2 ± 0.01	-4.29 ± 0.364	
4	71530000	71550000	20	1	103	0.2	-4.08	
4	71570000	71590000	20	1	100	0.19	-4.32	
4	71610000	71630000	20	1	211	0.19	-4.37	
4	72120000	72150000	30	2	168	0.19 ± 0.008	-4.42 ± 0.278	
4	74850000	74870000	20	1	141	0.19	-4.35	
5	11530000	11550000	20	1	315	0.2	-4.25	
5	19590000	19610000	20	1	112	0.2	-4.01	

5	20330000	20350000	20	1	266	0.2	-4.01	
5	31590000	31610000	20	1	214	0.2	-4.11	
5	31820000	31870000	50	4	107	0.17 ± 0.031	-5.2 ± 1.108	
5	40650000	40670000	20	1	154	0.19	-4.49	
5	40950000	40970000	20	1	116	0.2	-4.13	
5	41060000	41080000	20	1	116	0.2	-4.03	
5	51720000	51740000	20	1	88	0.15	-5.93	*AHNAK2*
6	9820000	9860000	40	3	137	0.19 ± 0.011	-4.39 ± 0.383	
6	12630000	12650000	20	1	224	0.2	-4.03	
6	13300000	13330000	30	2	64	0.16 ± 0.009	-5.38 ± 0.306	*KCNMA1*
7	30000	50000	20	1	80	0.19	-4.55	*NAB1*
7	470000	500000	30	2	176	0.2 ± 0.002	-4.22 ± 0.069	*COL5A2*
7	7710000	7730000	20	1	77	0.18	-4.7	*TMEFF2*
7	8080000	8100000	20	1	101	0.2	-4.02	
7	8160000	8180000	20	1	107	0.2	-4	
7	14780000	14830000	50	4	70	0.2 ± 0.009	-4.02 ± 0.306	*TTN*
7	18510000	18550000	40	3	198	0.19 ± 0.012	-4.4 ± 0.421	*CERS6*
7	19320000	19370000	50	3	84	0.16 ± 0.028	-5.59 ± 1.001	*SCN2A*
8	330000	350000	20	1	79	0.2	-4.2	
8	9490000	9530000	40	3	132	0.2 ± 0.002	-4.07 ± 0.069	
8	12170000	12190000	20	1	144	0.19	-4.58	
8	12750000	12770000	20	1	147	0.2	-4.04	*BCAR3*
8	28400000	28420000	20	1	170	0.18	-4.71	
8	28540000	28570000	30	2	74	0.19 ± 0.007	-4.61 ± 0.254	*TYW3*
9	6460000	6490000	30	2	222	0.2 ± 0.002	-4.08 ± 0.062	
9	23220000	23290000	70	6	102	0.16 ± 0.022	-5.58 ± 0.767	*PFN2, RNF13, COMMD2, WWTR1*
10	7630000	7650000	20	1	105	0.19	-4.37	*UNC13C*
10	15610000	15630000	20	1	153	0.2	-4.04	
10	19590000	19640000	50	4	115	0.19 ± 0.014	-4.59 ± 0.486	*FES, FURIN, MFAP1, HYPK, SERINC4, PDIA3, CKMT1A*
10	19750000	19770000	20	1	82	0.2	-4.1	*TP53BP1*

11	2540000	2560000	20	1	65	0.2	-4.15	*NECAB2*
11	2580000	2600000	20	1	54	0.19	-4.34	*OSGIN1*
11	3250000	3270000	20	1	66	0.2	-4.18	
12	16560000	16580000	20	1	176	0.2	-4.01	
13	17540000	17560000	20	1	52	0.19	-4.64	*CDC23*
15	320000	350000	30	2	88	0.2 ± 0.007	-4.23 ± 0.258	*AL158801.1*
17	190000	230000	40	3	85	0.17 ± 0.005	-5.14 ± 0.175	
17	10020000	10040000	20	1	101	0.2	-4	
17	10350000	10370000	20	1	55	0.2	-4.16	*RALGPS1*
18	30000	60000	30	2	58	0.19 ± 0.006	-4.38 ± 0.21	*ZNF302*
18	110000	130000	20	1	52	0.19	-4.61	*GSG1L2*
18	150000	190000	40	3	86	0.2 ± 0.009	-4.2 ± 0.33	*GLP2R, GAS7*
18	10910000	10930000	20	1	56	0.19	-4.51	*SLC16A5, ARMC7*
19	9430000	9460000	30	2	88	0.19 ± 0.012	-4.32 ± 0.408	*PAFAH1B1*
22	230000	310000	80	7	119	0.17 ± 0.009	-5.22 ± 0.322	*ANTXR1, GKN2, BMP10, ARHGAP25*
22	340000	370000	30	2	60	0.18 ± 0.003	-4.84 ± 0.107	*TMEM230*
22	440000	500000	60	5	87	0.16 ± 0.008	-5.52 ± 0.289	*PROM2, DPYSL2*
22	570000	720000	150	10	56	0.17 ± 0.018	-5.03 ± 0.641	*PPP2R2A, EBF2*
22	740000	770000	30	2	76	0.18 ± 0.017	-4.86 ± 0.612	
22	790000	830000	40	3	76	0.19 ± 0.005	-4.58 ± 0.173	*CDCA2*
22	870000	960000	90	7	74	0.19 ± 0.009	-4.63 ± 0.314	*KCTD9, GNRH1, DOCK5, NEFL, NEFM*
22	990000	1040000	50	4	76	0.18 ± 0.013	-4.75 ± 0.453	
22	1060000	1090000	30	2	54	0.16 ± 0.001	-5.58 ± 0.051	
22	1090000	1180000	90	8	63	0.17 ± 0.018	-5.27 ± 0.64	*STC1, SLC25A37*
22	1200000	1250000	50	4	133	0.18 ± 0.007	-4.76 ± 0.238	*ENTPD4, LOXL2, R3HCC1*
22	1290000	1330000	40	3	115	0.19 ± 0.008	-4.65 ± 0.271	*RHOBTB2*
22	1380000	1410000	30	2	146	0.2 ± 0.001	-4.06 ± 0.026	*DOK2, GFRA2*
22	1420000	1480000	60	5	174	0.19 ± 0.01	-4.4 ± 0.353	
22	1510000	1610000	100	9	132	0.19 ± 0.011	-4.37 ± 0.402	*LZTS1, ATP6V1B2, SLC18A1, MAK16, TTI2, RNF122*
22	1710000	1740000	30	2	136	0.2 ± 0.006	-4.27 ± 0.22	

22	1760000	1780000	20	1	111	0.2	-4.12	
22	2080000	2100000	20	1	87	0.19	-4.61	
22	2120000	2190000	70	6	113	0.19 ± 0.011	-4.34 ± 0.403	*ERLIN2, PROSC*
22	2270000	2330000	60	5	90	0.17 ± 0.017	-5.03 ± 0.612	*LSM1, BAG4, DDHD2, PPAPDC1B, WHSC1L1, LETM2*
22	2350000	2560000	210	20	95	0.19 ± 0.019	-4.63 ± 0.653	*TACC1, PLEKHA2, TM2D2, ADAM9, INDOL1, ZMAT4, SFRP1*
22	2570000	2600000	30	2	60	0.17 ± 0.002	-5.09 ± 0.087	*GOLGA7, GINS4, AGPAT6*
22	2620000	2660000	40	3	65	0.19 ± 0.009	-4.65 ± 0.31	*GKN2*
23	100000	120000	20	1	88	0.2	-4.16	*RPS6KA1*
23	5450000	5510000	60	5	104	0.17 ± 0.034	-5.26 ± 1.217	*NT5C1A, HPCAL4, TRIT1, MYCL, MFSD2A, CAP1, NIPAL3*
23	5700000	5720000	20	1	58	0.19	-4.46	*RCAN3*
26	100000	120000	20	1	66	0.2	-4.11	*SRPK1, SLC26A8*
26	5190000	5240000	50	4	104	0.2 ± 0.024	-3.98 ± 0.861	*PRELP, OPTC*

Table S6. Gene ontology analysis for domestic chicken and red junglefowl populations.

A. Ethiopian domestic chicken population

GO term	Description	*P* value	Genes
GO:0019229	Regulation of vasoconstriction	0.013932	*HRH1, AGTR1*

B. Saudi Arabian domestic chicken population

GO term	Description	*P* value	Genes
GO:0008015	Blood circulation	0.004766	*PRKG1, PDE5A, SEMA3A, PDE3A, HDAC4, CTNNA3, CACNA1G, GCLC*
GO:0008016	Regulation of heart contraction	0.007859	*PDE5A, SEMA3A, HDAC4, CTNNA3, CACNA1G*
GO:0090257	Regulation of muscle system process	0.009369	*PRKG1, PDE5A, MTPN, HDAC4, CTNNA3*
GO:0051094	Positive regulation of developmental process	0.010359	*TSHR, PDE5A, RHOB, SIRT1, PROX1, FLRT2, SEMA3A, RIMS1, PDE3A, IMPACT, MAP2K1, TMEM30A, HIF1AN*
GO:0006886	Intracellular protein transport	0.011917	*RHOB, SEC24C, RABGAP1L, MYO6, RIMS1*
GO:0060284	Regulation of cell development	0.016126	*PDE5A, PROX1, FLRT2, SEMA3A, RIMS1, PDE3A, HDAC4, IMPACT, MYCL, MAP2K1, TMEM30A*
GO:0070301	Cellular response to hydrogen peroxide	0.017447	*RHOB, SIRT1, IMPACT*
GO:0008206	Bile acid metabolic process	0.018957	*AKR1D1, SIRT1, PROX1*
GO:0045892	Negative regulation of transcription, DNA-templated	0.019806	*SIRT1, TFDP2, PHF14, FRK, PROX1, DACH1, HDAC4, IMPACT, AICDA, HIF1AN, GCLC*
GO:1902679	Negative regulation of RNA biosynthetic process	0.022875	*SIRT1, TFDP2, PHF14, FRK, PROX1, DACH1, HDAC4, IMPACT, AICDA, HIF1AN, GCLC*
GO:0043502	Regulation of muscle adaptation	0.02443	*PDE5A, MTPN, HDAC4*
GO:0005912	Adherens junction	0.025717	*RHOB, TNS3, ARHGAP22, TLN2, FLRT2, RSU1, CTNNA3, RPL4*
GO:0042391	Regulation of membrane potential	0.026217	*RIMS1, NETO1, KCNMA1, CTNNA3, CACNA1G, GCLC*

GO:2000243	Positive regulation of reproductive process	0.028393	*PDE5A, SEMA3A, PDE3A*
GO:0045595	Regulation of cell differentiation	0.030314	*SIRT1, GLG1, PROX1, SEMA3A, RIMS1, MYCL, MAP2K1, TMEM30A, HIF1AN, PDE5A, FLRT2, HDAC4, PDE3A, IMPACT, COL5A2*
GO:0055117	Regulation of cardiac muscle contraction	0.030465	*PDE5A, HDAC4, CTNNA3*
GO:0044712	Single-organism catabolic process	0.030569	*RP11-240B13.2, PDE5A, GK5, AKR1D1, ABHD5, LPIN1, PDE3A, HDAC4, USP13*
GO:0009117	Nucleotide metabolic process	0.039809	*TSHR, PDE5A, PRPS2, RP11-411B6.6, GRM8, MDH1B, PDE3A, HDAC4*
GO:0019934	Cgmp-mediated signaling	0.040082	*PRKG1, PDE3A*
GO:0034983	Peptidyl-lysine deacetylation	0.040082	*SIRT1, HDAC4*
GO:0019693	Ribose phosphate metabolic process	0.040998	*TSHR, PDE5A, PRPS2, RP11-411B6.6, GRM8, PDE3A, HDAC4*
GO:0005875	Microtubule associated complex	0.045734	*DYNC1I1, SNX4, KIF26B, BORCS5*
GO:0044281	Small molecule metabolic process	0.045786	*GK5, PRPS2, SIRT1, LARS2, PROX1, GRM8, LPIN1, ASNSD1, MDH1B, GCLC, PDE5A, TSHR, AKR1D1, RP11-411B6.6, HDAC4, PDE3A*

C. Sri Lankan domestic chicken population

GO term	**Description**	***P* value**	**Genes**
GO:0016500	protein-hormone receptor activity	0.04484	*TSHR, MCHR1*

D. Red Junglefowl population

GO term	**Description**	***P* value**	**Genes**
GO:0050789	Regulation of biological process	0.009225514	*ADCY1, AL158801.1, **EBF2**, KCNA5, **PPP2R2A**, CCND2, NAB1, NEFM, FURIN, **BMP10**, GNRH1, **ARHGAP25**, **MYCL**, NPY2R, RHOBTB2, PDIA3, ZNF302, BCAR3, NKX2-6, ADAM9, ZFHX4, **KCNMA1**, RCAN3, PAFAH1B1*
GO:0051246	Regulation of protein metabolic process	0.015403594	*AL158801.1, CCND2, ADAM9, FURIN, **BMP10***

GO:0006810	Transport	0.020966967	*SCN2A, ATP6V1B2, KCNA5, GRIK2, MFSD2A, SLC25A37, NEFM, KCNA1, FURIN, KCTD9, SLC16A5, KCNMA1, SLC26A8, PAFAH1B1*
GO:0015672	Monovalent inorganic cation transport	0.022221846	*ATP6V1B2, KCNA5, KCNA1, KCTD9,* ***KCNMA1***
GO:0022836	Gated channel activity	0.024342944	*KCNA5, GRIK2, KCNA1, KCTD9,* ***KCNMA1***
GO:0008088	Axon cargo transport	0.039408326	*NEFM, PAFAH1B1*
GO:0051592	Response to calcium ion	0.045827198	*ADAM9,* ***KCNMA1***

Note: Genes bolded in red junglefowl population overlapped with other genes selected in domestic chicken populations

Table S7. Kyoto Encyclopaedia of Genes and Genomes pathway (KEGG) analysis for domestic chicken and red junglefowl populations

A. Ethiopian domestic chicken population

ID	Description	*P* value	Genes
gga04080	Neuroactive ligand-receptor interaction	0.004583	*AGTR1, TACR3, HRH1, TSHR*
gga04020	Calcium signaling pathway	0.008034	*AGTR1, TACR3, HRH1*
gga04261	Adrenergic signaling in cardiomyocytes	0.041098	*AGTR1, ATP1B3*

B. Saudi Arabian domestic chicken population

ID	Description	*P* value	Genes
gga04620	Toll-like receptor signaling pathway	0.038572949	*TLR7, MAP2K1, MAP3K7*
gga04270	Vascular smooth muscle contraction	0.047517517	*KCNMA1, PRKG1, MAP2K1*

C. Sri Lankan domestic chicken population

ID	Description	*P* value	Genes
gga04261	Adrenergic signaling in cardiomyocytes	0.02710155	*ATP2B4, MAPK13,* ***PPP2R2A***
gga04020	Calcium signaling pathway	0.049582166	*ATP2B4, TACR3, EGFR*

D. Red Junglefowl population

ID	Description	*P* value	Genes
gga04914	Progesterone-mediated oocyte maturation	0.033069783	*RPS6KA1, ADCY1, CDC23*
gga00230	Purine metabolism	0.039356647	*ADCY1, ENTPD4, GUCY1A3, NT5C1A*
gga04270	Vascular smooth muscle contraction	0.049550781	***KCNMA1****, GUCY1A3, ADCY1*

Chapter 3

Whole-genome sequences reveals candidate signature of positive selection for environmental pressures in Ceylon (*Gallus lafayettii*) and green (*G. varius*) junglefowls

Table S1. Summary of reads, mappings, and SNPs statistics for individual genome

Sample	Reads	PPR (%)	MR (%)	MD (X)	5_b (%)	10_b (%)	20_b (%)	Hom	Het	Total SNPs
Ceylon Junglefowlk1[a]	334796624	90.03	99.24	30.64	97.9	95.9	84.2	8983216	2267144	11250360
Ceylon junglefowlk2[a]	346513098	93.19	99.38	30.64	97.9	95.9	84.2	8981536	2259292	11240828
Ceylon junglefowlk3[b]	158521405	94.27	99.32	14.9	95.4	82.7	13.6	9344441	1538733	10883174
Ceylon junglefowls1[c]	271286217	91.55	98.51	30.91	97.5	95.8	85.8	8699001	3003932	11702933
Ceylon junglefowls2[d]	280378335	89.93	98.64	31.96	97.5	95.9	86.9	8746361	2927259	11673620
Ceylon junglefowls3[d]	278379517	90.37	98.47	31.63	97.5	96	87.7	8613921	3192114	11806035
Ceylon junglefowls4[d]	255124566	91.85	98.59	29.03	97.3	95.5	83.6	8610768	3191108	11801876
Ceylon Junglefowls5[d]	232018442	88.27	98.35	26.35	97.2	95	77.7	8595155	3204385	11799540
Green junglefowlk1[a]	348251083	85.22	99	30.99	97.8	95.7	84.1	11273114	1553619	12826733
Green junglefowlk2[a]	341241445	89.81	99.04	30.28	97.3	95.8	88	11275036	1535593	12810629
Green junglefowlk3[b]	157834818	94.38	99.25	14.71	95.2	82.9	13.1	11092574	1671345	12763919
Green junglefowlJava18*	73900175	95.58	97.95	6.07	60.2	12.3	0.5	10281276	994490	11275766
Green junglefowlJava19*	65657014	95.84	97.83	5.98	63	11.2	0.4	10029122	1709571	11738693
Green junglefowlMadura702*	76152759	94.88	97.78	6.69	67.1	15.4	0.6	10858469	819757	11678226
Green junglefowlMadura703*	91040628	94.27	97.98	8.03	78.4	25.3	0.8	10971131	1021074	11992205
Green junglefowlMadura704*	65467669	94.64	97.94	5.8	57	9.1	0.5	10487770	853498	11341268
Green junglefowlMadura706*	87366743	93.7	97.77	7.63	74.5	21.4	0.7	10394379	814793	11209172
Green junglefowlMadura710*	58763807	94.54	97.93	5.19	49.5	5.8	0.4	10039321	885439	10924760
Green junglefowlMadura714*	61909786	94.07	97.92	5.45	53.5	6.5	0.5	8674990	367773	9042763
Green junglefowlMadura715*	83568378	94.2	97.87	7.31	71.8	19.2	0.7	10537850	759658	11297508

[a] Blood sample obtained from CC®P, [b] sequences obtained from CC®P, The CC®P is the Cosmopolitan Chicken Research Project (CC®P) (http://www.ccrp.be/), [c] blood samples obtained from Gannoruwa (Sri Lanka) and [d] Wayamba (Sri Lanka), * sequences obtained from "Ulfah, M., Kawahara-Miki, R., Farajalllah, A., Muladno, M., Dorshorst, B., Martin, A. *et al.* Genetic features of red and green junglefowls and relationship with Indonesian native chickens Sumatera and Kedu Hitam. *BMC genomics* **17**, 1 (2016)". **PPR:** Percentage of both mates of a read pair that were mapped to the same chromosome. **MR (%):** Percentage of reads mapped to the *Galgal* 4.0 reference genome. **MD (X):** The mean depth or the genome sequence coverage. **5_b:** Percentage of the genome with bases covered by at least 5 reads. **10_b:** Percentage of the genome with bases

covered by at least 10 reads. **20_b:** Percentage of the genome with bases covered by at least 20 reads. **Hom**: Homozygous SNPs compare to the reference. **Het:** Heterozygous SNPs compare to the reference. The sum of **Hom** and **Het** equal bi-allelic SNPs for each individual autosome.

Table S2. Candidate sweep regions in Ceylon junglefowl using Tajima's D statistics. The following have been shortened: **Chr** (chromosome), **Ln** (sweep length in kilobase pair), Σ_{SNPs} (sum of SNPs within sweep), **μD** (mean Tajima's D), **μZ(D)** (mean Z-transformed Tajima's D). The total size of selective sweep is 1.8 megabase pair and the median was 20 kb. The two regions in bold were assumed to be a single candidate sweep region because the 20-kb window which separated them falls just a little below the genome-wide threshold (see the result section)

Chr	Start	Stop	Ln (kb)	Σ_{SNPs}	μD	μZ(D)	Gene
1	17380000	17400000	20	69	-2.132	-4.604	
1	101360000	101380000	20	178	-1.838	-3.976	
1	159100000	159120000	20	96	-1.944	-4.203	
2	51020000	51060000	40	213	-1.981	-4.282	
2	51080000	51100000	20	77	-1.819	-3.934	
2	51720000	51740000	20	114	-1.869	-4.041	*C7orf25*
2	73500000	73520000	20	94	-1.897	-4.102	
2	86200000	86220000	20	191	-1.898	-4.104	
2	86360000	86440000	80	884	-1.967	-4.252	
2	87040000	87060000	20	193	-1.87	-4.045	
2	142820000	142840000	20	312	-1.815	-3.927	
2	142880000	142920000	40	546	-1.959	-4.235	
2	143420000	143440000	20	347	-1.808	-3.912	
3	35580000	35600000	20	96	-1.969	-4.255	*FMN2*
4	40000	240000	200	2038	-2.155	-4.654	*ZC4H2, ZC3H12B, LAS1L, MSN, FDX1L*
4	61740000	61760000	20	65	-1.902	-4.112	
6	13220000	13260000	40	137	-2.086	-4.506	*KCNMA1*
7	6500000	6520000	20	264	-1.983	-4.285	
7	6680000	6700000	20	262	-1.986	-4.292	*COL18A1, SLC19A1*

7	**6800000**	**7220000**	**420**	**4187**	**-2.079**	**-4.491**	***COL6A2, FTCD, MCM3AP, YBEY, POFUT2, LSS, S100B, TUBA1C, PCNT, C21orf58, KMO, ITGB2***
7	**7240000**	**7600000**	**360**	**3310**	**-2.144**	**-4.631**	***FAM207A, ADARB1, GLS, STAT1, STAT4, MYO1B***
7	7680000	7700000	20	196	-1.965	-4.247	*SDPR*
7	7740000	7760000	20	148	-2.108	-4.554	*TMEFF2*
7	7860000	7920000	60	593	-1.9	-4.108	
8	9440000	9460000	20	35	-1.953	-4.222	
9	5900000	5920000	20	342	-1.864	-4.031	*TRIM42*
11	3000000	3040000	40	92	-2.104	-4.545	*SLC12A4*
14	440000	460000	20	84	-1.862	-4.027	
24	0	20000	20	257	-1.852	-4.006	*C2CD2L, HMBS, VPS11, SIK2*
24	40000	100000	60	748	-1.878	-4.061	
24	6200000	6240000	40	433	-1.916	-4.143	*DIXDC1, CRYAB*
24	6280000	6300000	20	95	-2.309	-4.984	*VWA5A*

4821

Table S3. Candidate sweep regions for Ceylon junglefowl using Hp statistics. The following have been shortened: **Chr** (chromosome), **Ln** (sweep length in kilobase pair), **Σ_{SNPs}** (sum of SNPs within sweep), **μHp** (mean Hp), **μZ(Hp)** (mean Z-transformed Hp). The total size of selective sweep is 1.86 megabase pair and the median was 20 kb.

Chr	Start	Stop	Ln (kb)	Σ_{SNPs}	μHp	μZ(Hp)	Gene
1	43000000	43020000	20	213	0.048	-4.008	
1	72580000	72600000	20	216	0.048	-4.002	*PTHLH*
1	73660000	73680000	20	124	0.046	-4.065	*KCNA1*
1	74680000	74700000	20	199	0.05	-3.954	*PRMT8*
1	74880000	74900000	20	125	0.05	-3.948	
1	159060000	159080000	20	176	0.039	-4.227	
2	52040000	52060000	20	165	0.044	-4.11	
2	52180000	52200000	20	173	0.048	-3.994	*HECW1*
2	52240000	52260000	20	181	0.043	-4.129	
2	52280000	52300000	20	152	0.05	-3.964	*VOPP1*
2	52420000	52440000	20	155	0.044	-4.105	*EGFR*
2	52460000	52480000	20	154	0.037	-4.272	*EGFR*
2	52500000	52520000	20	129	0.048	-4.002	*EGFR*
2	52560000	52580000	20	176	0.042	-4.152	
2	52600000	52620000	20	138	0.032	-4.405	
2	52640000	52660000	20	159	0.043	-4.136	
2	52680000	52720000	40	396	0.039	-4.223	*SEC61G*
2	52840000	52860000	20	153	0.03	-4.452	*TPK1*
2	53120000	53160000	40	340	0.046	-4.049	
2	53240000	53260000	20	96	0.041	-4.191	
2	53280000	53300000	20	170	0.048	-4.01	
4	17920000	17960000	40	368	0.038	-4.254	
4	18140000	18180000	40	302	0.047	-4.039	*AFF2*
5	2600000	2620000	20	150	0.031	-4.446	
5	2660000	2700000	40	374	0.047	-4.031	
5	2720000	2820000	100	732	0.038	-4.264	*ANO5, SLC17A6*

5	2940000	3060000	120	1073	0.046	-4.048	*GAS2, SVIP, FANCF*
5	3120000	3220000	100	671	0.039	-4.24	
5	3240000	3260000	20	147	0.034	-4.359	
5	3420000	3440000	20	169	0.036	-4.321	*SLC5A12*
5	3480000	3500000	20	131	0.037	-4.295	
6	13240000	13260000	20	219	0.044	-4.119	*KCNMA1*
6	13280000	13300000	20	238	0.048	-4.007	*KCNMA1*
7	35960000	35980000	20	179	0.042	-4.166	*TANC1*
7	36000000	36200000	200	1655	0.035	-4.338	*TANC1, WDSUB1, BAZ2B, Mar-07*
8	40000	60000	20	267	0.046	-4.047	*AMY2A*
8	200000	340000	140	2042	0.037	-4.295	
8	9440000	9460000	20	152	0.038	-4.261	
9	20000	60000	40	323	0.045	-4.091	*ARMC8*
11	2800000	2820000	20	145	0.046	-4.048	*NFATC3*
11	2900000	2920000	20	165	0.041	-4.189	*DUS2, DDX28, GALR1L*
11	2980000	3060000	80	712	0.037	-4.285	*SLC12A4, SLC6A2*
11	19240000	19300000	60	678	0.044	-4.105	*PHLPP2, MRVLDC3, TAT*
11	19340000	19360000	20	391	0.026	-4.561	*ADAT1, GABARAPL2, RP11-77K12.7*
17	10420000	10460000	40	272	0.044	-4.097	*GARNL3*
20	20000	80000	60	586	0.028	-4.503	*IGSF1, L3MBTL1*
26	160000	180000	20	190	0.05	-3.951	*MAPK13*
27	5040000	5060000	20	178	0.047	-4.026	*WNK4, BECN1*
27	5100000	5140000	40	303	0.044	-4.108	*RPL27, IFI35, VAT1, RND2*
28	4680000	4700000	20	116	0.037	-4.289	*DPP9, MYDGF*

Table S4. Candidate sweep regions for green junglefowl using Tajima's D statistics. The following have been shortened: **Chr** (chromosome), **Ln** (sweep length in kilobase pair), Σ_{SNPs} (sum of SNPs within sweep), **μD** (mean Tajima's D), **μZ(D)** (mean Z-transformed Tajima's D). The total size of selective sweep is 1.74 megabase pair and the median was 20 kb.

Chr	Start	Stop	Ln (kb)	Σ_{SNPs}	μD	μZ(D)	Gene
1	8900000	8920000	20	35	-2.191	-4.242	
1	9040000	9100000	60	124	-2.099	-4.063	
1	9180000	9200000	20	32	-1.803	-3.492	
1	72440000	72480000	40	76	-2.11	-4.086	*MRPS35, MANSC4*
1	72520000	72540000	20	32	-2.117	-4.099	
1	72580000	72600000	20	43	-1.826	-3.536	*PTHLH*
1	72620000	72640000	20	31	-1.989	-3.851	
1	72800000	72820000	20	59	-2.089	-4.044	
1	72860000	72880000	20	37	-1.846	-3.575	
1	72940000	72960000	20	45	-2.242	-4.339	*FAR2*
1	72980000	73020000	40	89	-2.234	-4.323	*FAR2*
1	73140000	73220000	80	178	-2.096	-4.058	*RAD51AP1, C12orf4*
1	73240000	73280000	40	92	-1.984	-3.841	*FGF23, TIGAR*
1	73300000	73320000	20	43	-2.046	-3.96	*CCND2*
1	73380000	73420000	40	102	-1.871	-3.623	
1	73480000	73540000	60	78	-2.18	-4.22	*RP11-234B24.6*
1	73660000	73700000	40	65	-2.112	-4.088	*KCNA1*
1	73760000	73800000	40	76	-1.881	-3.642	*KCNA5*
1	73860000	73880000	20	37	-1.908	-3.694	
1	73940000	74000000	60	150	-2.009	-3.891	
1	74020000	74060000	40	74	-1.997	-3.866	*NTF3*
1	74100000	74120000	20	30	-1.771	-3.43	
1	74180000	74240000	60	78	-1.948	-3.771	*ANO2*
1	74340000	74400000	60	126	-1.977	-3.828	*VWF*
1	74400000	74420000	20	33	-2.085	-4.036	
1	74540000	74560000	20	38	-2.019	-3.91	*NOBOX*

1	74680000	74700000	20	42	-1.772	-3.432	*PRMT8*
1	74700000	74720000	20	56	-2.073	-4.014	
1	99600000	99620000	20	73	-1.883	-3.647	
1	115680000	115700000	20	132	-1.88	-3.64	*IL1RAPL1*
1	140580000	140620000	40	260	-1.989	-3.851	
1	158400000	158420000	20	42	-1.856	-3.594	
1	160860000	160880000	20	71	-1.822	-3.528	
2	25460000	25480000	20	64	-1.783	-3.454	
2	143240000	143260000	20	67	-1.829	-3.542	
3	140000	160000	20	84	-1.849	-3.581	*CFAP36*
3	57980000	58000000	20	96	-1.8	-3.486	
3	71520000	71540000	20	40	-1.864	-3.611	
3	101160000	101180000	20	327	-1.766	-3.42	
3	101180000	101200000	20	358	-1.822	-3.528	
4	40620000	40640000	20	35	-1.798	-3.482	
4	41320000	41340000	20	34	-1.933	-3.742	
4	41700000	41720000	20	81	-1.773	-3.433	*NEIL3*
4	59140000	59160000	20	90	-1.873	-3.628	
5	2360000	2380000	20	56	-1.803	-3.493	*NELL1*
5	3140000	3160000	20	58	-1.813	-3.511	
5	3560000	3580000	20	36	-1.82	-3.524	
5	3660000	3680000	20	49	-1.764	-3.416	*LGR4*
5	3760000	3780000	20	38	-1.763	-3.414	
6	9840000	9860000	20	409	-1.874	-3.63	
7	7300000	7320000	20	40	-1.82	-3.524	*ADARB1*
8	9540000	9560000	20	41	-1.939	-3.755	
8	10000000	10020000	20	52	-1.932	-3.74	*AMY1A*
8	10080000	10100000	20	40	-1.865	-3.612	
8	10120000	10140000	20	37	-2.061	-3.991	
8	10660000	10680000	20	34	-1.789	-3.465	*DPH5, MIR1610*
9	300000	320000	20	40	-1.833	-3.551	*STAG1*

11	4960000	4980000	20	110	-1.766	-3.42	
11	19200000	19220000	20	65	-2.063	-3.994	*AP1G1*
12	19600000	19620000	20	54	-1.842	-3.568	*TMCC1*
14	3460000	3480000	20	275	-1.943	-3.763	*CARD11*
15	240000	260000	20	44	-1.805	-3.495	*HIC2*
15	480000	500000	20	51	-1.889	-3.658	
17	10220000	10240000	20	37	-2.044	-3.958	
20	14000000	14020000	20	35	-2.025	-3.921	
20	14040000	14060000	20	40	-1.925	-3.727	*CEBPB*

4832

Table S5. Candidate sweep regions for green junglefowl using Hp statistics. The following have been shortened: **Chr** (chromosome), **Ln** (sweep length in kb), Σ_{SNPs} (sum of SNPs within sweep), **μHp** (mean Hp), **μZ(Hp)** (mean Z-transformed Hp). The total size of selective sweep is 1.80 mega base and the median was 20 kb.

Chr	Start	Stop	Ln (kb)	Σ_{SNPs}	μHp	μZ(Hp)	Gene
1	8880000	8940000	60	435	0.02	-2.933	
1	8960000	8980000	20	210	0.021	-2.898	
1	9000000	9220000	220	2673	0.019	-2.954	
1	72300000	72400000	100	1023	0.019	-2.947	
1	72420000	72560000	140	1376	0.02	-2.916	*MRPS35, MANSC4, KLHL42*
1	72580000	72600000	20	219	0.026	-2.746	*PTHLH*
1	72620000	72640000	20	175	0.024	-2.807	
1	72660000	72680000	20	191	0.025	-2.781	
1	72700000	72760000	60	698	0.02	-2.933	
1	72920000	72940000	20	169	0.016	-3.017	*FAR2*
1	72960000	73040000	80	882	0.018	-2.979	*FAR2*
1	73060000	73260000	200	2654	0.017	-2.996	*RAD51AP1, C12orf4, FGF6, FGF23*
1	73300000	73320000	20	269	0.021	-2.901	*CCND2*
1	73340000	73360000	20	275	0.014	-3.082	*CCND2*
1	73380000	73400000	20	293	0.023	-2.829	
1	73440000	73460000	20	263	0.021	-2.893	
1	73480000	73500000	20	239	0.019	-2.937	
1	73520000	73540000	20	195	0.02	-2.922	*RP11-234B24.6*
1	73660000	73680000	20	158	0.023	-2.836	*KCNA1*
1	73760000	73780000	20	184	0.026	-2.745	*KCNA5*
1	73860000	73880000	20	194	0.024	-2.813	
1	73940000	73960000	20	230	0.024	-2.824	
1	74040000	74060000	20	210	0.022	-2.862	*NTF3*
1	74080000	74160000	80	692	0.022	-2.872	*NTF3*
1	74180000	74200000	20	222	0.017	-3.001	*ANO2*
1	74260000	74280000	20	176	0.024	-2.816	*ANO2*

1	74340000	74400000	60	553	0.023	-2.839	*VWF*
1	74440000	74480000	40	394	0.022	-2.871	
1	74520000	74600000	80	728	0.02	-2.907	*CD9, NOBOX, ARHGEF19*
1	74680000	74700000	20	223	0.025	-2.788	*PRMT8*
1	74720000	74740000	20	185	0.021	-2.897	
1	74800000	74840000	40	299	0.023	-2.85	*PARP11, TSPAN9*
1	74900000	74920000	20	170	0.017	-2.992	
1	158420000	158440000	20	221	0.025	-2.779	
1	162380000	162400000	20	206	0.026	-2.761	
4	40020000	40060000	40	417	0.021	-2.904	
4	40620000	40660000	40	418	0.024	-2.812	
4	41320000	41340000	20	230	0.019	-2.937	
5	3560000	3580000	20	200	0.026	-2.757	
8	10120000	10140000	20	189	0.026	-2.761	
9	540000	560000	20	250	0.027	-2.735	*GIGYF2*
20	14000000	14020000	20	173	0.026	-2.762	

4836
4837
4838

Table S6. Overlapped sweep regions between Tajima's D and Hp methods and within Ceylon junglefowl

Chr	Start	Stop	Ln (kb)	Gene
6	13220000	13260000	40	*KCNMA1*
8	9440000	9460000	20	
11	3000000	3040000	40	*SLC12A4*

Table S7. Overlapped sweep regions between Tajima's D and Hp methods and within green junglefowl

Chr	Start	Stop	Ln (kb)	Genes
1	8900000	8920000	20	
1	9040000	9100000	60	
1	9180000	9200000	20	
1	72440000	72480000	40	*MRPS35, MANSC4*
1	72520000	72540000	20	
1	72580000	72600000	20	*PTHLH*
1	72620000	72640000	20	
1	72980000	73020000	40	*FAR2*
1	73140000	73220000	80	*RAD51AP1, C12orf4*
1	73240000	73280000	40	*FGF23, TIGAR*
1	73300000	73320000	20	*CCND2*
1	73380000	73420000	40	
1	73480000	73540000	60	*RP11-234B24.6*
1	73660000	73700000	40	*KCNA1*
1	73760000	73800000	40	*KCNA5*
1	73860000	73880000	20	
1	73940000	74000000	60	
1	74020000	74060000	40	*NTF3*
1	74100000	74120000	20	
1	74180000	74240000	60	*ANO2*
1	74340000	74400000	60	*VWF*
1	74540000	74560000	20	*NOBOX*
1	74680000	74700000	20	*PRMT8*
4	40620000	40640000	20	
4	41320000	41340000	20	
5	3560000	3580000	20	
8	10120000	10140000	20	
20	14000000	14020000	20	

Table S8. Overlap sweep regions between Ceylon and green junglefowls

Chr	Start	Stop	Ln (kb)	Methods	Genes
1	72580000	72600000	20	Hp	*PTHLH*
1	73660000	73680000	20	Hp	*KCNA1*
1	74680000	74700000	20	Hp	*PRMT8*
7	7240000	7600000	360	Tajima's D	*FAM207A, ADARB1, GLS, STAT1, STAT4, MYO1B*

4900 **Table S9.** Functional analysis for Ceylon junglefowl using the GO and KEGG pathway

Term ID	Term description	P value	Overrepresented genes
GO:0019239	Deaminase activity	0.004430385	*ADAT1, ADARB1, FTCD*
GO:0006536	Glutamate metabolic process	0.004520639	*GLS, FTCD, TAT*
GO:1901606	Alpha-amino acid catabolic process	0.010440707	*GLS, KMO, TAT*
GO:0072341	Modified amino acid binding	0.012353541	*SDPR, FTCD, SLC19A1*
GO:0022857	Transmembrane transporter activity	0.025013245	*KCNMA1, SLC17A6, SLC6A2, SLC12A4, KCNA1, ANO5, SLC19A1, SEC61G, SLC5A12*
GO:0050730	Regulation of peptidyl-tyrosine phosphorylation	0.027495015	*EGFR, ITGB2*
GO:0031988	Membrane-bounded vesicle	0.029542119	*EGFR, COL18A1, KCNMA1, MYDGF, BECN1, CRYAB, MYO1B, FTCD, RPL27, KMO, ITGB2, VAT1, SLC17A6, COL6A2, VOPP1, SLC5A12*
GO:0008509	Anion transmembrane transporter activity	0.033882493	*SLC17A6, SLC12A4, ANO5, SLC5A12*
GO:0006970	Response to osmotic stress	0.038284358	*KCNMA1, MAPK13*
GO:0009628	Response to abiotic stimulus	0.046633135	*EGFR, KCNMA1, CRYAB, MAPK13, KCNA1, TANC1, KMO*
GO:0015291	Secondary active transmembrane transporter activity	0.048155511	*SLC6A2, SLC12A4, SLC19A1, SLC5A12*
KEGG:gga04320	Dorso-ventral axis formation	0.008107	*FMN2, EGFR*
KEGG:gga04140	Regulation of autophagy	0.014452	*GABARAPL2, BECN1*
KEGG:gga04068	FoxO signaling pathway	0.028836	*MAPK13, GABARAPL2, EGFR*

4901

Table S10. Functional analysis for green junglefowl using the GO and KEGG pathway

Term ID	**Term description**	***P* value**	**Overrepresented genes**
GO:0042325	Regulation of phosphorylation	0.028036324	*CARD11, ADARB1, NTF3, CCND2, FGF23, TIGAR*
GO:1901214	Regulation of neuron death	0.026236716	*CEBPB, NTF3, TIGAR*
GO:0071495	Cellular response to endogenous stimulus	0.037178941	*FGF6, CD9, CEBPB, FGF23, TIGAR*
GO:0007169	Transmembrane receptor protein tyrosine kinase signaling pathway	0.035395338	*FGF6, NTF3, FGF23, GIGYF2*
GO:0051174	Regulation of phosphorus metabolic process	0.013314753	*PTHLH, CARD11, ADARB1, NTF3, CCND2, FGF23, TIGAR*
GO:0001649	Osteoblast differentiation	0.002536029	*PTHLH, CEBPB, FGF23, LGR4*
GO:0032502	Developmental process	0.048166723	*PTHLH, FGF6, CARD11, CD9, CEBPB, NTF3, AP1G1, KCNA1, FGF23, GIGYF2, LGR4, TIGAR*
GO:0051260	Protein homooligomerization	0.024923382	*VWF, KCNA1, KCNA5*
KEGG:gga04010	MAPK signaling pathway	0.01507	*NTF3, FGF23, FGF6*

Chapter 4

The autosomal genome ancestry of domestic chicken: Introgression from grey and Ceylon junglefowls

Table S1. Samples information

Sample Name	Sampling location	Country	Population	Category	Scientific name
HA2A25B	Horro district	Ethiopia	Horro	Domestic	*Gallus domesticus*
HA1B25B	Horro district	Ethiopia	Horro	Domestic	*Gallus domesticus*
HB1A16A	Horro district	Ethiopia	Horro	Domestic	*Gallus domesticus*
HB1B21B	Horro district	Ethiopia	Horro	Domestic	*Gallus domesticus*
HA1A22A	Horro district	Ethiopia	Horro	Domestic	*Gallus domesticus*
HA2A10B	Horro district	Ethiopia	Horro	Domestic	*Gallus domesticus*
JA2A10B	Jarso district	Ethiopia	Jarso	Domestic	*Gallus domesticus*
JB1B16A	Jarso district	Ethiopia	Jarso	Domestic	*Gallus domesticus*
JB1A25B	Jarso district	Ethiopia	Jarso	Domestic	*Gallus domesticus*
JB2A04B	Jarso district	Ethiopia	Jarso	Domestic	*Gallus domesticus*
JA1A17A	Jarso district	Ethiopia	Jarso	Domestic	*Gallus domesticus*
Saudi Arabia1	Al-Hafof city	Saudi Arabia	Saudi Arabia	Domestic	*Gallus domesticus*
Saudi Arabia2	Al-Hafof city	Saudi Arabia	Saudi Arabia	Domestic	*Gallus domesticus*
Saudi Arabia3	Al-Hafof city	Saudi Arabia	Saudi Arabia	Domestic	*Gallus domesticus*
Saudi Arabia4	Al-Hafof city	Saudi Arabia	Saudi Arabia	Domestic	*Gallus domesticus*
Saudi Arabia5	Al-Hafof city	Saudi Arabia	Saudi Arabia	Domestic	*Gallus domesticus*
Sri Lanka1	Puttlam district	Sri Lanka	Sri Lanka	Domestic	*Gallus domesticus*
Sri Lanka2	Puttlam district	Sri Lanka	Sri Lanka	Domestic	*Gallus domesticus*
Sri Lanka3	Puttlam district	Sri Lanka	Sri Lanka	Domestic	*Gallus domesticus*
Sri Lanka4	Puttlam district	Sri Lanka	Sri Lanka	Domestic	*Gallus domesticus*
Sri Lanka5	Puttlam district	Sri Lanka	Sri Lanka	Domestic	*Gallus domesticus*
Sri Lanka9	Puttlam district	Sri Lanka	Sri Lanka	Domestic	*Gallus domesticus*
Sri Lanka10	Puttlam district	Sri Lanka	Sri Lanka	Domestic	*Gallus domesticus*
Sri Lanka12	Puttlam district	Sri Lanka	Sri Lanka	Domestic	*Gallus domesticus*
Sri Lanka13	Puttlam district	Sri Lanka	Sri Lanka	Domestic	*Gallus domesticus*
Sri Lanka15	Puttlam district	Sri Lanka	Sri Lanka	Domestic	*Gallus domesticus*

Sri Lanka16	Puttlam district	Sri Lanka	Sri Lanka	Domestic	*Gallus domesticus*
Mechelse koekoek	Private collection[1]	Belgium	Fancy	Domestic	*Gallus domesticus*
Poulet de Bresse	Private collection[1]	France	Fancy	Domestic	*Gallus domesticus*
Mechelse styrian	Private collection[1]	Belgium	Fancy	Domestic	*Gallus domesticus*
Red junglefowl1	Yunnan Province[2]	China	Red junglefowl	Wild	*Gallus gallus*
Red junglefowl2	Yunnan Province[2]	China	Red junglefowl	Wild	*Gallus gallus*
Red junglefowl3	Yunnan Province[2]	China	Red junglefowl	Wild	*Gallus gallus*
Red junglefowl4	Hainan Province[2]	China	Red junglefowl	Wild	*Gallus gallus*
Red junglefowl5	Yunnan Province, Xishuanghanna[2]	China	Red junglefowl	Wild	*Gallus gallus*
Red junglefowl-koen	Private collection[3]	Belgium	Red junglefowl	Wild	*Gallus gallus*
Grey junglefowl1k	Private collection[3]	Belgium	Grey junglefowl	Wild	*Gallus sonneratii*
Grey junglefowl2k	Private collection[3]	Belgium	Grey junglefowl	Wild	*Gallus sonneratii*
Grey junglefowl3k	Private collection[3]	Belgium	Grey junglefowl	Wild	*Gallus sonneratii*
Ceylon Junglefowlk1	Private collection[3]	Belgium	Ceylon Junglefowl	Wild	*Gallus lafayetii*
Ceylon junglefowlk2	Private collection[3]	Belgium	Ceylon Junglefowl	Wild	*Gallus lafayetii*
Ceylon junglefowlk3	Private collection[3]	Belgium	Ceylon Junglefowl	Wild	*Gallus lafayetii*
Ceylon junglefowlsr1	Uva province	Sri Lanka	Ceylon Junglefowl	Wild	*Gallus lafayetii*
Ceylon junglefowlsr2	Uva province	Sri Lanka	Ceylon Junglefowl	Wild	*Gallus lafayetii*
Ceylon junglefowlsr3	Uva province	Sri Lanka	Ceylon Junglefowl	Wild	*Gallus lafayetii*
Ceylon junglefowl1sr4	Uva province	Sri Lanka	Ceylon Junglefowl	Wild	*Gallus lafayetii*
Ceylon Junglefowl1sr5	Uva province	Sri Lanka	Ceylon Junglefowl	Wild	*Gallus lafayetii*
Green junglefowl1k	Private collection[3]	Belgium	Green junglefowl	Wild	*Gallus varius*
Green junglefowl2k	Private collection[3]	Belgium	Green junglefowl	Wild	*Gallus varius*
Green junglefowl3k	Private collection[3]	Belgium	Green junglefowl	Wild	*Gallus varius*

[1]Bird samples obtained from the Cosmopolitan Chicken Research Project in Belgium. They include the Mechelse koekoek and Poulet de Bresse (parental lines) and the Mechelse styrian, the 16th generation crossbred (http://www.ccrp.be/#/0), [2]Wang, M.-S., Li, Y., Peng, M.-S., Zhong, L., Wang, Z.-J., Li, Q.-Y. *et al.* Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. *Molecular biology and evolution* **32**, 1880-1889 (2015). [3]Koen Vanmechelen Private Bird collection

Table S2. Summary sequencing statistics and numbers of SNPs

Samples	Reads	PPR (%)	MR (%)	MD (X)	5_b (%)	10_b (%)	20_b (%)	Hom	Het	Total SNPs
SNPs information for domestic chicken from Ethiopia, Saudi Arabia, Sri Lanka, and red and Ceylon junglefowls, are found in the supplementary Table S1 of chapters 2 and 3										
Mechelse koekoek	101373357	90.01	96.01	11.17	90.3	56.9	4	3164689	2465899	5630588
Poulet de Bresse	171141396	91.78	96.91	19.27	97.7	92.6	42.5	2915896	3172160	6088056
Mechelse styrian	158798357	91.22	96.16	17.66	97.3	88.5	31.6	2345156	4357926	6703082
Grey junglefowl1	343715230	91.05	99.28	31.28	98.2	96.3	85.3	8366471	2387736	10754207
Grey junglefowl2	343383598	87.11	99.12	31.57	97.9	96.5	87.9	8186940	2634640	10821580
Grey junglefowl3	158709787	94.69	99.35	14.93	95.6	83.3	13.7	8570915	1539929	10110844
Green junglefowl1	348251083	85.22	99	30.99	97.8	95.7	84.1	11273114	1553619	12826733
Green junglefowl2	341241445	89.81	99.04	30.28	97.3	95.8	88	11275036	1535593	12810629
Green junglefowl3	157834818	94.38	99.25	14.71	95.2	82.9	13.1	11092574	1671345	12763919

PPR: Percentage of both mates of a read pair that were properly mapped to the same chromosome. **MR:** Total reads mapped to the *Galgal* 4.0 reference genome. **MD:** The mean sequence depth or the genome sequence coverage. **5_b:** Percentage of the genome with bases covered by at least 5 reads. **10_b:** Percentage of the genome with bases covered by at least 10 reads. **20_b:** Percentage of the genome with bases covered by at least 20 reads. **Hom**: Homozygous SNPs compare to the reference. **Het:** Heterozygous SNPs compare to the reference. The **Total SNPs** include the bi-allelic **Hom** + **Het.**

Table S3. Summary information for the mitochondrial DNA genomes downloaded from NCBI

Sample Name	**Accession number**	**Country**	**Category**	**Scientific Name**
Silky	AB086102.1	Japan: Hiroshima	Domestic chicken	*Gallus domesticus*
Chigulu	GU261684.1	China: Yunnan	Domestic chicken	*Gallus domesticus*
Tulufan	GU261683.1	China: Xinjiang	Domestic chicken	*Gallus domesticus*
Xianju	GU261677.1	China: Zhejiang	Domestic chicken	*Gallus domesticus*
YP20488	GU261686.1	China: Henan	Domestic chicken	*Gallus domesticus*
YP63	GU261680.1	Southern India	Domestic chicken	*Gallus domesticus*
Yin21b	GU261698.1	Northeast India	Domestic chicken	*Gallus domesticus*
Yin46	HQ857212	Northeast India	Domestic chicken	*Gallus domesticus*
Yin12	GU261685	Northeast India	Domestic chicken	*Gallus domesticus*
Yin125	HQ857211	-	Domestic chicken	*Gallus domesticus*
WhitePlymouth Rock	AP003318.1	-	Domestic chicken	*Gallus domesticus*
White Leghorn	AP003317.1	-	Domestic chicken	*Gallus domesticus*
Sonneratii1	AP003320.1	Lao Zoological Park,Vientiane, Lao	Wild	*Gallus sonneratii*
Sonneratii2	AP006741.1	Delhi National Park, Delhi, India	Wild	*Gallus sonneratii*
Sonneratii3	AP006746.1	Japan: Tokyo, Tama Zoological Park	Wild	*Gallus sonneratii*
Sonneratii4	NC_007240.1		Wild	*Gallus sonneratii*
Lafayetii1	AP003325.1	Tama Zoological Park, Tokyo, Japan	Wild	*Gallus lafayetii*
Lafayetii2	NC_007239.1	Tama Zoological Park, Tokyo, Japan	Wild	*Gallus lafayetii*
Varius1	AP003324.1	Bali, Indonesia	Wild	*Gallus varius*
Varius2	NC_007238.1	Bali, Indonesia	Wild	*Gallus varius*

Table S4. Population and variants SNPs statistics and annotations

Variant consequences	**Ethiopia[a]**	**Saudi Arabia[a]**	**Sri Lanka[a]**	**Fancy birds[a]**	**Red[b]**	**Grey[b]**	**Ceylon[b]**	**Green[b]**
Samples number	11	5	11	3	6	3	8	3
Total number of bi-allelic SNPs	13,075,004	10,231,837	14,455,255	9,746,503	15,317,274	11,959,750	17,316,240	13,715,157
Average Properly paired reads (%)	93.50	92.67	92.62	91.00	89.57	90.95	91.98	89.80
Average mapped reads	388,850,785	107,489,443	260,240,115	138,629,051	230,353,835	279,751,704	266,460,953	279,783,600
Average mapping rate (%)	98.75	98.56	98.75	96.36	98.08	99.25	98.81	99.10
Novel Variants	1,765,053	1,029,993	2,325,291	956,881	4,451,313	6,764,127	12,019,614	9,010,868
Existing Variants	11,309,951	9,201,844	12,129,964	8,789,622	10,865,961	5,195,623	5,296,626	4,704,289
Overlapped genes	39,892	39,868	39,901	39,853	39,922	39,830	39,930	39,888
Overlapped transcript	47,659	47,647	47,659	47,660	47,677	47,594	47,665	47,657
Splice donor variant	643	444	693	472	730	556	791	654
Splice acceptor variant	570	496	603	401	653	531	899	580
Stop gained	1154	886	1302	757	1281	873	1571	1008
Stop lost	307	265	345	262	371	331	419	354
Start lost	365	303	416	280	428	356	513	392
Missense variant	74,488	57,637	83,438	54,350	87,272	61,538	99,949	73,064
Splice region variant	26,015	20,522	28,811	19,369	30,350	22,684	33,848	25,705
Stop retained variant	79	59	80	55	92	79	126	108
Synonymous variant	123,041	98,486	138,077	92,719	145,705	105,585	160,411	120,323
Coding sequence variant	3	1	4	1	1	1	11	1
Mature miRNA variant	170	133	197	119	214	151	242	199
5 prime UTR variant	51,655	40,869	57,787	40,179	60,366	46,030	68,764	52,323
3 prime UTR variant	246,270	193,040	273,686	181,557	289,436	218,014	323,714	251,044
Non-coding transcript exon variant	133,818	10,6657	148,262	100,596	158,914	121,734	177,987	138,486
Intron variant	7,058,272	554,4587	7,791,825	5,274,339	8,244,547	6,347,064	9,167,140	7,249,461
Upstream gene variant	758,660	604,830	838,568	574,103	885,235	674,736	981,457	766,315
Downstream gene variant	568,176	451,103	626,089	426,544	664,720	508,043	735,887	579,829

Intergenic variant	4,031,318	3,111,519	4,465,072	2,980,400	4,746,959	3,851,442	5,562,507	4,455,310

[a] Domestic chicken population. [b] Junglefowl population

Table S5. Within population nucleotide diversity (π) calculated at 20 kb windows.

Population name	**Number of samples**	**Genome mean nucleotide diversity**
Ethiopia domestic chicken	11	0.0041 ± 0.0018
Saudi Arabia domestic chicken	5	0.0039 ± 0.0018
Sri Lanka domestic chicken	11	0.0046 ± 0.0018
Fancy birds	3	0.0040 ± 0.0018
Red junglefowl	6	0.0052 ± 0.0019
Grey junglefowl	3	0.0024 ± 0.0020
Ceylon junglefowl	8	0.0034 ± 0.0013
Green junglefowl	3	0.0014 ± 0.0008

Table S6. Pairwise F_{ST} values between populations. I measured the pairwise genetic distance between population at 20 kb windows. The green junglefowl among all the species remain the most distinct populations while the red junglefowl remain the closest species to the domestic chicken. The domestic chicken populations on this table includes **Ethiopia, Saudi Arabia, Sri Lanka** and **Fancy birds**. The junglefowl populations on this table includes **Red, Grey, Ceylon** and **Green junglefowl**.

	Ethiopia	**Saudi Arabia**	**Sri Lanka**	**Fancy birds**	**Red**	**Grey**	**Ceylon**	**Green**
Ethiopia	0							
Saudi Arabia	0.081	0						
Sri Lanka	0.092	0.082	0					
Fancy birds	0.122	0.097	0.07	0				
Red	0.096	0.095	0.062	0.063	0			
Grey	0.651	0.666	0.621	0.676	0.596	0		
Ceylon	0.658	0.662	0.633	0.662	0.615	0.625	0	
Green	0.739	0.762	0.714	0.779	0.699	0.856	0.776	0

Table S7. Pairwise nucleotide divergence (d_{XY}) values between populations at 20 kb windows

Pop	Domestic Chicken	Red Junglefowl	Grey Junglefowl	Ceylon Junglefowl	Green Junglefowl
Domestic Chicken	0				
Red Junglefowl	0.110 ± 0.025	0			
Grey Junglefowl	0.226 ± 0.044	0.227 ± 0.042	0		
Ceylon Junglefowl	0.244 ± 0.030	0.243 ± 0.030	0.185 ± 0.039	0	
Green Junglefowl	0.288 ± 0.035	0.287 ± 0.035	0.283 ± 0.033	0.284 ± 0.033	0

Table S8.1. Akaike information criterion for model predicted for mitochondrial SNPs**.** There are 24 models in the 100% confidence interval: [GTR+I+G GTR+G HKY+G HKY+I+G SYM+I+G SYM+G K80+I+G K80+G GTR+I HKY+I SYM+I K80+I GTR HKY SYM K80 F81+I+G F81+G JC+I+G JC+G F81+I JC+I F81 JC]

Model	-InL	K	AIC	Delta	weight	cumWeight
GTR+I+G	35828.00707	150	71956.01414	0	0.999476	0.999476
HKY+I+G	35839.55977	146	71971.11954	15.1054	5.24E-04	1
GTR+G	35846.61177	149	71991.22354	35.2094	2.26E-08	1
HKY+G	35859	145	72008	51.98586	5.14E-12	1
GTR+I	35875.21996	149	72048.43992	92.42578	8.51E-21	1
HKY+I	35890.10871	145	72070.21742	114.20328	1.59E-25	1
SYM+I+G	36706.62636	147	73707.25272	1751.23858	0.00E+00	1
SYM+G	36726.85439	146	73745.70878	1789.69464	0.00E+00	1
SYM+I	36760.00406	146	73812.00812	1855.99398	0.00E+00	1
K80+I+G	36765.91735	143	73817.8347	1861.82056	0.00E+00	1
K80+G	36784.30836	142	73852.61672	1896.60258	0.00E+00	1
GTR	36796.55229	148	73889.10458	1933.09044	0.00E+00	1
K80+I	36813.11867	142	73910.23734	1954.2232	0.00E+00	1
HKY	36813.9857	144	73915.9714	1959.95726	0.00E+00	1
F81+I+G	37594.26081	145	75478.52162	3522.50748	0.00E+00	1
F81+G	37602.90156	144	75493.80312	3537.78898	0.00E+00	1
F81+I	37621.52127	144	75531.04254	3575.0284	0.00E+00	1
SYM	37642.01989	145	75574.03978	3618.02564	0.00E+00	1
K80	37711.45265	141	75704.9053	3748.89116	0.00E+00	1
F81	38441.25203	143	77168.50406	5212.48992	0.00E+00	1
JC+I+G	38456.74036	142	77197.48072	5241.46658	0.00E+00	1
JC+G	38466.41713	141	77214.83426	5258.82012	0.00E+00	1
JC+I	38487.33409	141	77256.66818	5300.65404	0.00E+00	1
JC	39290.64518	140	78861.29036	6905.27622	0.00E+00	1

-lnL: negative log likelihood; **K**: number of estimated parameters; **AIC**: Akaike Information Criterion; **delta**: AIC difference; **weight**: AIC weight; **cumWeight** cumulative AIC weight.

Table S8.2. Akaike information criterion for model predicted for autosomal SNPs. There are 24 models in the 100% confidence interval: [GTR+I+G GTR+G HKY+G HKY+I+G SYM+I+G SYM+G K80+I+G K80+G GTR+I HKY+I SYM+I K80+I GTR HKY SYM K80 F81+I+G F81+G JC+I+G JC+G F81+I JC+I F81 JC]

Model	**-lnL**	**K**	**AIC**	**Delta**	**Weight**	**cumWeight**
GTR+I+G	625114.9322	108	1250445.864	0	1	1
GTR+G	625144.7019	107	1250503.404	57.53938	3.20E-13	1
HKY+G	625476.1983	103	1251158.397	712.5321	1.89E-155	1
HKY+I+G	625509.8497	104	1251227.699	781.83504	1.69E-170	1
SYM+I+G	625640.6705	105	1251491.341	1045.47646	0.00E+00	1
SYM+G	625672.0258	104	1251552.052	1106.18716	0.00E+00	1
K80+I+G	625783.5333	101	1251769.067	1323.20214	0.00E+00	1
K80+G	625813.6795	100	1251827.359	1381.49456	0.00E+00	1
GTR+I	635313.3424	107	1270840.685	20394.8203	0.00E+00	1
HKY+I	635639.4619	103	1271484.924	21039.0594	0.00E+00	1
SYM+I	635738.8863	104	1271685.773	21239.9082	0.00E+00	1
K80+I	635887.8273	100	1271975.655	21529.7902	0.00E+00	1
GTR	639098.473	106	1278408.946	27963.0815	0.00E+00	1
HKY	639444.171	102	1279092.342	28646.4776	0.00E+00	1
SYM	639552.2801	103	1279310.56	28864.6958	0.00E+00	1
K80	639723.5678	99	1279645.136	29199.2711	0.00E+00	1
F81+I+G	661098.4909	103	1322402.982	71957.1173	0.00E+00	1
F81+G	661122.4359	102	1322448.872	72003.0074	0.00E+00	1
JC+I+G	661170.7073	100	1322541.415	72095.5502	0.00E+00	1
JC+G	661194.2113	99	1322586.423	72140.5582	0.00E+00	1
F81+I	669729.4639	102	1339662.928	89217.0633	0.00E+00	1
JC+I	669800.8896	99	1339799.779	89353.9147	0.00E+00	1
F81	673085.1266	101	1346372.253	95926.3887	0.00E+00	1
JC	673166.8299	98	1346529.66	96083.7953	0.00E+00	1

-lnL: negative log likelihood; **K**: number of estimated parameters; **AIC**: Akaike Information Criterion; **delta**: AIC difference; **weight**: AIC weight; **cumWeight** cumulative AIC weight.

Table S9. Chromosomal *D* statistics: Average *D* statistic value for each chromosome (20 kb window) between domestic chicken and three junglefowl species (red, grey, Ceylon). The negative D statistic indicates more shared variation between P1 (Red junglefowl) and P3 (either grey or Ceylon junglefowl) than between P2 (domestic chicken) and P3. In other words, an indication of excess of BABA over ABBA.

	D statistic values		
Chr	**Domestic chicken - red**	**Domestic chicken - grey**	**Domestic chicken - Ceylon**
1	0.77 ± 0.168	0.05 ± 0.227	0.02 ± 0.210
2	0.77 ± 0.213	0.04 ± 0.230	0.01 ± 0.208
3	0.81 ± 0.127	0.02 ± 0.207	0.002 ± 0.197
4	0.64 ± 0.267	0.07 ± 0.215	0.01 ± 0.198
5	0.81 ± 0.118	0.02 ± 0.239	0.01 ± 0.213
6	0.72 ± 0.216	0.03 ± 0.179	0.004 ± 0.165
7	0.81 ± 0.104	0.03 ± 0.220	0.02 ± 0.184
8	0.81 ± 0.135	0.01 ± 0.240	0.01 ± 0.225
9	0.79 ± 0.105	0.02 ± 0.163	0.01 ± 0.140
10	0.78 ± 0.107	0.02 ± 0.162	0.002 ± 0.136
11	0.71 ± 0.260	0.04 ± 0.216	0.03 ± 0.198
12	0.76 ± 0.111	0.04 ± 0.159	0.02 ± 0.146
13	0.70 ± 0.207	0.05 ± 0.169	0.01 ± 0.168
14	0.79 ± 0.117	0.01 ± 0.180	-0.0004 ± 0.161
15	0.78 ± 0.146	0.03 ± 0.209	-0.007 ± 0.209
16	0.13 ± 0.270	-0.06 ± 0.143	0.01 ± 0.085
17	0.75 ± 0.131	0.06 ± 0.197	0.01 ± 0.190
18	0.58 ± 0.332	0.04 ± 0.160	0.02 ± 0.167
19	0.45 ± 0.394	0.07 ± 0.226	-0.009 ± 0.202
20	0.72 ± 0.157	0.08 ± 0.205	-0.003 ± 0.205
21	0.58 ± 0.222	0.08 ± 0.168	0.008 ± 0.157
22	0.89 ± 0.132	0.05 ± 0.305	0.03 ± 0.293
23	0.73 ± 0.118	0.03 ± 0.173	0.03 ± 0.130
24	0.71 ± 0.097	0.04 ± 0.140	0.01 ± 0.121
25	0.76 ± 0.093	0.01 ± 0.157	0.01 ± 0.148
26	0.67 ±0.220	0.10 ± 0.199	0.03 ± 0.167
27	0.65 ± 0.195	0.04 ± 0.181	-0.009 ± 0.191
28	0.78 ± 0.100	0.02 ± 0.188	0.03 ± 0.185

Table S10A. Characteristics of the introgressed regions from domestic chicken into grey junglefowl (n = 317). Other bidirectional regions which also shows introgression from domestic chicken into grey junglefowl are located on Table 1A (n = 32), Table 1B (n = 2), Table 1C (n = 4) and Table 1D (n = 15). The **Ggal** is the reference genome for 4 or 5 version and **Ln4** and **Ln5** are the length in kilobase for the **Ggal4** and **5** coordinates respectively. **Ggal4_μ(D)** and **Ggal4_ μ*(Zfd)*** are the mean of D and the *Zfd* statistics based on *Galgal* 4.0 reference genome respectively

Chr	Ggal4_start	Ggal4_stop	Ln4	Ggal4_μ(D)	Ggal4_ μ*(Zfd)*	Ggal4_Genes	Ggal5_start	Ggal5_stop	Ln5
1	14885001	14905001	20	0.29	4.23	*ENSGALT00000039327*	15086012	15106012	20
1	15705001	15740001	35	0.37 ± 0.035	4.68 ± 0.593	*GRAMD4*	15906899	15942087	35
1	15810001	15840001	30	0.29 ± 0.038	4.01 ± 0.305	*GRAMD4*	16012141	16042686	31
1	16105001	16135001	30	0.37 ± 0.03	4.81 ± 0.265	-	16307735	16337736	30
1	16195001	16215001	20	0.35	4.44	-	16397740	16417750	20
1	16325001	16360001	35	0.50 ± 0.033	3.85 ± 0.351	*ENSGALT00000045258*	16527685	16562728	35
1	16430001	16460001	30	0.40 ± 0.030	4.56 ± 0.262	*ENSGALT00000045258*	16632195	16662201	30
1	16520001	16560001	40	0.75 ± 0.077	8.64 ± 1.110	-	16722200	16762203	40
1	17175001	17205001	30	0.46 ± 0.023	4.35 ± 0.199	*FAM19A5*	17376339	17406336	30
1	18090001	18120001	30	0.66 ± 0.023	6.20 ± 0.202	-	18291775	18321767	30
1	18555001	18635001	80	0.40 ± 0.106	4.19 ± 1.491	*BRD1*	18762639	18842590	80
1	18900001	18920001	20	0.5	4.42	-	19117118	19137120	20
1	19060001	19100001	40	0.55 ± 0.067	5.12 ± 0.650	-	19276925	19316219	39
1	19175001	19195001	20	0.79	6.07	-	19391172	19411172	20
1	19280001	19305001	25	0.28 ± 0.026	4.02 ± 0.020	-	19496145	19519784	24
1	19540001	19565001	25	0.54 ± 0.007	4.51 ± 0.308	-	19754894	19779897	25
1	19595001	19625001	30	0.65 ± 0.101	6.24 ± 0.130	-	19809811	19836579	27
1	19710001	19740001	30	0.45 ± 0.013	4.69 ± 0.467	-	19921566	19951568	30
1	88910001	88930001	20	0.39	15.49	-	No Map	No Map	
1	100465001	100515001	50	0.74 ± 0.011	6.79 ± 0.162	-	101129899	101179901	50
1	138815001	138835001	20	0.43	6.53	-	139480509	139500509	20
1	139040001	139070001	30	0.52 ± 0.078	6.88 ± 0.769	-	139705347	139735348	30
1	139620001	139685001	65	0.44 ± 0.170	3.88 ± 1.360	-	140324648	140389619	65
1	139785001	139810001	25	0.55 ± 0.013	6.30 ± 0.071	-	140489618	140514616	25
1	139865001	139910001	45	0.61 ± 0.068	6.30 ± 0.257	-	140569250	140614251	45

1	139960001	139990001	30	0.78 ± 0.021	4.67 ± 0.457	-	140664253	140694247	30
1	140090001	140115001	25	0.71 ± 0.016	4.33 ± 0.136	-	140794241	140819244	25
1	140235001	140270001	35	0.60 ± 0.160	4.75 ± 0.399	-	140939258	140990041	51
1	140350001	140385001	35	0.46 ± 0.043	5.05 ± 0.645	*EFNB2*	141070041	141105041	35
1	140865001	140895001	30	0.67 ± 0.010	4.52 ± 0.126	-	141594945	141625085	30
1	141060001	141090001	30	0.80 ± 0.077	4.91 ± 0.234	-	141792279	141821997	30
1	141110001	141170001	60	0.76 ± 0.208	4.10 ± 0.626	-	141842002	141901996	60
1	141285001	141310001	25	0.70 ± 0.028	4.41 ± 0.109	-	142033389	142058391	25
1	141355001	141425001	70	0.74 ± 0.036	6.68 ± 0.530	-	142103012	142173036	70
1	141850001	141875001	25	0.55 ± 0.031	6.32 ± 0.063	-	142598084	142623084	25
1	141915001	141945001	30	0.37 ± 0.055	4.76 ± 0.233	*SLC10A2*	142663098	142693090	30
1	142025001	142045001	20	0.38	4.11	*BIVM*	142773101	142793101	20
1	142350001	142375001	25	0.60 ± 0.018	6.22 ± 0.079	-	143098197	143123197	25
1	142550001	142575001	25	0.34 ± 0.014	4.25 ± 0.343	*FGF14*	143298029	143323029	25
1	142625001	142655001	30	0.34 ± 0.075	4.03 ± 0.214	*FGF14*	143373029	143403030	30
1	142885001	142910001	25	0.33 ± 0.008	4.22 ± 0.018	*ITGBL1*	143632913	143657913	25
1	142970001	143005001	35	0.32 ± 0.060	4.77 ± 0.523	*NALCN*	143717911	143752910	35
1	143035001	143065001	30	0.35 ± 0.033	4.78 ± 0.205	-	143782910	143812911	30
1	143255001	143280001	25	0.40 ± 0.031	6.42 ± 0.532	*PCCA*	144003807	144028812	25
1	143320001	143340001	20	0.32	4.01	-	144068812	144088812	20
1	143920001	143945001	25	0.36 ± 0.022	5.11 ± 0.445	-	144670701	144695183	24
1	144070001	144090001	20	0.28	4.96	*SLC15A1, STK24*	144820182	144840182	20
1	144130001	144160001	30	0.45 ± 0.011	4.47 ± 0.273	*FARP1*	144880144	144910147	30
1	144265001	144290001	25	0.38 ± 0.063	4.77 ± 0.885	-	145015148	145040148	25
1	144440001	144475001	35	0.49 ± 0.052	6.99 ± 0.977	-	145199017	145234016	35
1	144555001	144605001	50	0.39 ± 0.038	4.88 ± 0.640	*RAP2A, MBNL2*	145313996	145363632	50
1	145140001	145170001	30	0.33 ± 0.042	4.78 ± 0.526	*UGGT2*	145908741	145938743	30
1	145285001	145310001	25	0.35 ± 0.016	4.18 ± 0.246	-	146053742	146078735	25
1	145455001	145475001	20	0.37	4.35	*ABCC4*	146223730	146243727	20
1	145540001	145560001	20	0.44	6.098	*DCT*	146308726	146328819	20
1	149920001	149955001	35	0.54 ± 0.052	4.62 ± 0.385	-	150693919	150744623	51

1	150000001	150035001	35	0.54 ± 0.024	4.76 ± 0.546	-	150789621	150824620	35
1	150310001	150345001	35	0.70 ± 0.027	4.41 ± 0.294	-	151099200	151134200	35
1	153515001	153545001	30	0.27 ± 0.047	4.84 ± 0.452	*MYCBP2*	154307490	154337489	30
1	154080001	154125001	45	0.33 ± 0.037	3.91 ± 0.402	*LMO7*	154886101	154931100	45
1	154995001	155075001	80	0.39 ± 0.055	3.79 ± 0.385	*KLF12*	155812372	155892960	81
1	156280001	156340001	60	0.57 ± 0.079	4.86 ± 0.475	-	157096605	157156606	60
1	156375001	156400001	25	0.62 ± 0.043	6.08 ± 0.088	-	157191608	157216022	24
1	156840001	156860001	20	0.47	4.06	*KLHL1*	157658687	157678686	20
1	157065001	157090001	25	0.60 ± 0.070	4.32 ± 0.057	*KLHL1*	157897340	157922344	25
1	160175001	160205001	30	0.60 ± 0.046	4.11 ± 0.065	-	161005254	161035253	30
1	160410001	160430001	20	0.65	4.67	-	161239992	161259992	20
1	161530001	161550001	20	0.51	4.03	*TDRD3*	162216893	162236893	20
1	161585001	161660001	75	0.42 ± 0.031	3.79 ± 0.514	*TDRD3*	162271892	162346894	75
1	161690001	161715001	25	0.65 ± 0.015	6.02 ± 0.020	*DIAPH3*	162375915	162400915	25
1	161740001	161760001	20	0.58	6.22	*DIAPH3*	162425916	162445917	20
1	161815001	161840001	25	0.76 ± 0.025	6.06 ± 0.023	*DIAPH3*	162500823	162525827	25
1	161920001	161960001	40	0.67 ± 0.042	5.59 ± 0.423	-	162605834	162645718	40
1	164095001	164135001	40	0.55 ± 0.041	4.23 ± 0.193	-	164783773	164823849	40
1	194975001	194995001	20	0.27	4.3	*CHRDL2*	195903400	195923255	20
2	10525001	10560001	35	0.54 ± 0.056	7.12 ± 0.591	*ADARB2*	10588186	10623186	35
2	10665001	10710001	45	0.37 ± 0.023	4.81 ± 0.596	-	10728213	10773034	45
2	10890001	10910001	20	0.62	6.07	-	10952889	10972892	20
2	11115001	11275001	160	0.54 ± 0.141	7.23 ± 0.731	-	11177890	11337885	160
2	11195001	11310001	115	0.53 ± 0.173	7.29 ± 1.247	-	11257887	11372886	115
2	11435001	11460001	25	0.34 ± 0.013	4.81 ± 0.366	-	11497885	11522885	25
2	11735001	11760001	25	0.25 ± 0.022	4.18 ± 0.119	*KLF6*	11798732	11823813	25
2	11785001	11850001	65	0.33 ± 0.038	4.86 ± 0.605	-	11848813	11913814	65
2	11890001	11960001	70	0.42 ± 0.051	4.58 ± 0.362	-	11964105	12034217	70
2	12060001	12145001	85	0.58 ± 0.061	7.24 ± 0.489	*gga-mir-6628*	12130838	12215911	85
2	12170001	12220001	50	0.42 ± 0.040	4.90 ± 0.494	-	12240914	12290912	50
2	12280001	12350001	70	0.38 ± 0.129	4.21 ± 1.638	-	12350912	12420928	70

2	12930001	12960001	30	0.41 ± 0.008	4.36 ± 0.127	*CREM*	13000371	13030373	30
2	13075001	13095001	20	0.35	4.2	*PARD3*	13145102	13165102	20
2	13200001	13275001	75	0.30 ± 0.103	3.78 ± 0.800	*PARD3*	13270100	13345100	75
2	13655001	13675001	20	0.49	6.69	-	13725219	13745219	20
2	13855001	13925001	70	0.47 ± 0.063	5.30 ± 0.776	*NRP1*	13925164	13995163	70
2	14520001	14540001	20	0.32	4.07	*ZNF438*	14605103	14625102	20
2	14630001	14660001	30	0.33 ± 0.018	4.80 ± 0.372	*SVIL*	14715096	14745096	30
2	15255001	15290001	35	0.55 ± 0.011	7.15 ± 0.359	*WAC*	15340103	15375105	35
2	15445001	15495001	50	0.46 ± 0.030	4.81 ± 0.324	*MPP7*	15529901	15579903	50
2	15580001	15665001	85	0.43 ± 0.128	4.07 ± 1.307	*ARMC4, MKX*	15664905	15749863	85
2	15765001	15800001	35	0.45 ± 0.029	4.30 ± 0.150	*PTCHD3*	NoMap	NoMap	
2	16100001	16130001	30	0.41 ± 0.043	4.43 ± 0.469	*ENSGALT00000031243*	16220407	16251476	31
2	16225001	16255001	30	0.51 ± 0.036	4.28 ± 0.253	*MYO3A*	16346294	16376296	30
2	16360001	16425001	65	0.45 ± 0.060	5.11 ± 0.665	*GPR158*	16481053	16546053	65
2	16620001	16645001	25	0.32 ± 0.018	4.19 ± 0.109	*ENKUR, PRTFDC1*	16741069	16766070	25
2	17750001	17780001	30	0.59 ± 0.026	6.50 ± 0.445	*DNAJC1*	17884053	17914052	30
2	17895001	17920001	25	0.44 ± 0.034	4.61 ± 0.295	*MLLT10*	18029053	18053887	25
2	18135001	18160001	25	0.48 ± 0.053	4.80 ± 0.112	*NEBL*	18268404	18293404	25
2	18175001	18195001	20	0.47	4.61	*NEBL*	18308418	18328419	20
2	18450001	18485001	35	0.49 ± 0.140	5.02 ± 0.815	-	18583547	18618355	35
2	18520001	18550001	30	0.70 ± 0.071	7.68 ± 0.759	*PLXDC2*	18665494	18695493	30
2	18575001	18595001	20	0.42	4.54	-	18720493	18740496	20
2	18685001	18715001	30	0.67 ± 0.043	7.40 ± 0.606	*ENSGALG00000007962*	18830335	18860334	30
2	18735001	18765001	30	0.52 ± 0.019	5.01 ± 0.457	-	18880334	18910337	30
2	18885001	18950001	65	0.57 ± 0.137	6.35 ± 1.491	-	19030339	19095342	65
2	19485001	19520001	35	0.51 ± 0.064	5.13 ± 0.521	-	19631014	19665908	35
2	40185001	40245001	60	0.40 ± 0.106	3.63 ± 0.751	-	39745165	39805168	60
2	40515001	40545001	30	0.52 ± 0.043	6.89 ± 0.535	*GADL1*	40075164	40105164	30
2	40750001	40775001	25	0.44 ± 0.014	4.49 ± 0.224	*STT3B*	40310141	40335141	25
2	40815001	40850001	35	0.43 ± 0.022	5.17 ± 0.553	*OSBPL10*	40375139	40410141	35

2	40995001	41045001	50	0.41 ± 0.032	5.77 ± 0.578	*CMTM7, DYNC1LI1*	40555139	40605139	50
2	41415001	41465001	50	0.30 ± 0.029	4.35 ± 0.425	*ABHD5*	40972323	41022326	50
2	41560001	41580001	20	0.45	6.24	-	41117331	41137331	20
2	41610001	41640001	30	0.32 ± 0.020	4.08 ± 0.113	*SNRK*	41167331	41197331	30
2	41930001	41950001	20	0.42	6.36	*MTURN, gga-mir-1723*	41487495	41507494	20
2	42045001	42070001	25	0.37 ± 0.022	5.35 ± 0.593	*SH3BP5*	41602435	41627434	25
2	42165001	42190001	25	0.45 ± 0.045	4.37 ± 0.472	*COL6A6*	41722437	41747435	25
2	67630001	67655001	25	0.39 ± 0.017	4.33 ± 0.062	*MYLK4*	67806256	67831256	25
2	68220001	68290001	70	0.44 ± 0.028	4.43 ± 0.469	*ZCCHC2*	68396145	68466121	70
3	20460001	20505001	45	0.31 ± 0.089	4.32 ± 0.539	*PTPN14*	21147729	21192728	45
3	20600001	20635001	35	0.51 ± 0.076	7.31 ± 1.227	*PROX1*	21287728	21322729	35
3	20705001	20725001	20	0.4	4.37	-	21392684	21412683	20
3	21115001	21145001	30	0.36 ± 0.034	4.82 ± 0.541	*FLVCR1, TATDN3, NSL1*	21802684	21832684	30
3	21220001	21250001	30	0.53 ± 0.014	6.53 ± 0.297	-	21907672	21937672	30
4	16985001	17035001	50	0.42 ± 0.053	3.79 ± 0.422	*IL1RAPL2*	17093966	17143965	50
4	17200001	17230001	30	0.57 ± 0.028	4.25 ± 0.040	*NRK*	17308966	17338966	30
4	19880001	19970001	90	0.47 ± 0.100	3.98 ± 0.496	*RBM46*	20571153	20661145	90
4	19990001	20010001	20	0.61	5.4	-	20681148	20701149	20
4	20320001	20345001	25	0.43 ± 0.055	4.75 ± 0.320	*GUCY1B3, ASIC5*	21017249	21042249	25
4	20920001	20950001	30	0.59 ± 0.128	4.80 ± 0.738	*GRIA2*	21617229	21647231	30
4	21765001	21795001	30	0.48 ± 0.053	4.56 ± 0.341	-	22462240	22492341	30
4	21800001	21820001	20	0.45	4.58	-	22497341	22517340	20
4	21960001	22050001	90	0.61 ± 0.099	4.39 ± 0.738	-	22657340	22752038	95
4	23220001	23330001	110	0.49 ± 0.064	4.62 ± 0.482	*ENSGALT00000038777, TMEM192, KLHL2, MSMO1*	23924533	24034531	110
4	23475001	23510001	35	0.59 ± 0.044	4.45 ± 0.315	-	24179531	24214531	35
4	23890001	23985001	95	0.36 ± 0.087	3.72 ± 0.897	*SPOCK3*	24616920	24711925	95
4	24000001	24035001	35	0.46 ± 0.083	4.87 ± 0.125	*SPOCK3*	24726924	24761924	35
4	25180001	25220001	40	0.39 ± 0.019	5.12 ± 0.332	-	25921872	25961870	40

4	25520001	25555001	35	0.66 ± 0.043	6.31 ± 0.318	-	26261684	26296780	35
4	25580001	25625001	45	0.40 ± 0.107	4.15 ± 0.532	-	26321212	26365750	45
4	26310001	26340001	30	0.49 ± 0.082	4.92 ± 0.596	-	27050587	27080586	30
4	26595001	26635001	40	0.43 ± 0.067	4.26 ± 0.158	-	27339321	27379311	40
4	26795001	26820001	25	0.42 ± 0.009	4.61 ± 0.482	-	27539202	27564203	25
4	26970001	27000001	30	0.78 ± 0.033	6.74 ± 0.245	-	27714140	27743761	30
4	27070001	27100001	30	0.71 ± 0.024	4.03 ± 0.031	-	27813766	27843767	30
4	27530001	27550001	20	0.37	4.02	-	28278669	28298665	20
4	27975001	28050001	75	0.57 ± 0.130	6.79 ± 0.566	-	28736323	28811311	75
4	28095001	28130001	35	0.46 ± 0.017	7.03 ± 0.347	*PCDH18*	28856313	28891315	35
4	28325001	28350001	25	0.53 ± 0.045	6.35 ± 0.415	*SLC7A11*	29086318	29111318	25
4	28385001	28430001	45	0.43 ± 0.035	6.21 ± 0.162	-	29146318	29191316	45
4	29750001	29770001	20	0.38	4.11	-	30519555	30539555	20
4	31910001	31930001	20	0.48	4.22	-	32718324	32738192	20
4	32515001	32540001	25	0.45 ± 0.001	4.05 ± 0.001	*LRBA*	33322829	33347829	25
4	32800001	32820001	20	0.23	4.01	-	33607841	33627840	20
4	32850001	32875001	25	0.45 ± 0.045	4.88 ± 0.381	*FAM160A1*	33657841	33682841	25
4	33355001	33380001	25	0.54 ± 0.030	7.26 ± 0.276	*ARFIP1, FHDC1*	34162825	34187638	25
4	33490001	33525001	35	0.36 ± 0.036	5.17 ± 0.450	*CDKL2*	34306209	34341209	35
4	33550001	33645001	95	0.49 ± 0.048	7.85 ± 1.156	*ENSGALG00000023351, C4H8ORF40, SLC20A2, SGK196, INTS10, HGSNAT, SMIM19*	34366209	34461235	95
4	33895001	33975001	80	0.38 ± 0.027	6.64 ± 0.778	-	34711308	34791309	80
4	33990001	34035001	45	0.49 ± 0.036	7.28 ± 0.719	*JADE1*	34806311	34851139	45
4	34040001	34105001	65	0.44 ± 0.028	6.81 ± 0.500	*ENSGALT00000043425.1, RCHY1, PARM1, BTC, GSR, PPP2CB*	34856139	34920881	65
4	39240001	39275001	35	0.45 ± 0.069	7.61 ± 1.438	*ING2*	40070724	40105703	35
4	39335001	39360001	25	0.30 ± 0.015	4.12 ± 0.090	-	40165703	40190703	25

4	39385001	39405001	20	0.32	6.22	*WWC2*	40215703	40235703	20
4	39450001	39475001	25	0.24 ± 0.031	5.24 ± 0.223	*DCTD*	40280692	40305691	25
4	39495001	39570001	75	0.35 ± 0.073	4.56 ± 0.977	*TENM3*	40325691	40400692	75
4	39765001	39785001	20	0.26	4.02	*TENM3*	40595749	40615750	20
4	40855001	40880001	25	0.63 ± 0.016	4.31 ± 0.043	-	41696828	41721828	25
4	40925001	40985001	60	0.48 ± 0.070	5.06 ± 0.652	-	41766834	41826836	60
4	41020001	41055001	35	0.85 ± 0.031	7.64 ± 0.881	-	41861837	41896838	35
4	41520001	41545001	25	0.48 ± 0.024	5.06 ± 0.134	-	42361886	42386887	25
4	41615001	41640001	25	0.31 ± 0.076	4.55 ± 0.720	-	42456887	42481888	25
4	41810001	41840001	30	0.39 ± 0.049	5.11 ± 1.038	-	42651887	42681888	30
4	42665001	42695001	30	0.37 ± 0.061	7.92 ± 1.108	*GALNT7*	43506958	43536900	30
4	42710001	42735001	25	0.39 ± 0.031	6.73 ± 0.911	-	43551900	43576900	25
4	42855001	42875001	20	0.22	4.37	-	43695732	43715735	20
4	42980001	43000001	20	0.38	4.98	-	43820740	43840738	20
4	43265001	43290001	25	0.47 ± 0.022	4.57 ± 0.259	-	44105736	44130736	25
4	43300001	43345001	45	0.20 ± 0.126	2.27 ± 1.689	*GLRA3*	44140737	44185738	45
4	43455001	43505001	50	0.29 ± 0.014	4.86 ± 0.379	-	44295735	44388514	93
4	43680001	43705001	25	0.30 ± 0.031	4.92 ± 1.017	-	44563509	44588508	25
4	43735001	43755001	20	0.39	4.038	*WDR17*	44619054	44639056	20
4	43950001	44055001	105	0.34 ± 0.050	7.11 ± 0.849	*VEGFC*	44850607	44955611	105
4	44110001	44140001	30	0.30 ± 0.021	4.32 ± 0.390	*MTHFD2L, EPGN*	45010614	45040614	30
4	44200001	44230001	30	0.25 ± 0.021	5.42 ± 0.043	*USO1*	45099697	45129498	30
4	44265001	44305001	40	0.41 ± 0.054	7.89 ± 1.337	*ENSGALG00000010883*	45164497	45204497	40
4	44380001	44490001	110	0.25 ± 0.078	3.50 ± 1.129	-	45279498	45389497	110
4	44615001	44660001	45	0.46 ± 0.035	7.37 ± 0.494	*ANTXR2*	45514496	45559495	45
4	44785001	44820001	35	0.26 ± 0.032	4.55 ± 0.619	-	45685466	45720465	35
4	45150001	45190001	40	0.30 ± 0.020	4.61 ± 0.289	*KLHL8*	46050388	46090390	40
4	45220001	45255001	35	0.33 ± 0.045	5.46 ± 0.407	*AFF1*	46120391	46155391	35
4	45270001	45330001	60	0.43 ± 0.048	7.04 ± 0.752	*TPN13*	46170391	46230360	60
4	45365001	45390001	25	0.41 ± 0.028	6.33 ± 0.324	*MAPK10*	46265352	46290354	25

4	46220001	46240001	20	0.33	6.2	*CDS1*	47122742	47142742	20
4	46860001	46880001	20	0.46	4.92	-	47767512	47787512	20
4	47375001	47500001	125	0.25 ± 0.078	3.50 ± 1.129	*LPHN3*	48282240	48407244	125
4	47525001	47615001	90	0.36 ± 0.084	4.79 ± 1.702	*LPHN3*	48432245	48522253	90
4	47780001	47805001	25	0.40 ± 0.048	5.07 ± 0.532	*LPHN3*	48687256	48712255	25
4	48010001	48030001	20	0.38	4.24	-	48917212	48937212	20
4	48060001	48085001	25	0.30 ± 0.019	5.59 ± 0.341	-	48967213	48992212	25
4	48190001	48235001	45	0.51 ± 0.065	7.00 ± 0.693	-	49097218	49142222	45
4	48295001	48320001	25	0.44 ± 0.010	6.92 ± 0.157	-	49202225	49227108	25
4	48525001	48555001	30	0.33 ± 0.045	5.04 ± 0.619	*TECRL*	49432114	49462114	30
4	48610001	48640001	30	0.38 ± 0.041	5.15 ± 0.658	*POLR2B, NOA1, REST*	49517114	49547188	30
4	49380001	49410001	30	0.33 ± 0.031	4.66 ± 0.397	*PPEF2, NAAA, SDAD1*	50295754	50325753	30
4	49560001	49590001	30	0.52 ± 0.079	6.87 ± 0.892	*SLC4A4*	50475772	50505772	30
4	49625001	49660001	35	0.45 ± 0.033	6.80 ± 0.551	-	50540679	50575689	35
4	50890001	50920001	30	0.31 ± 0.015	4.38 ± 0.120	-	51816730	51846730	30
4	75755001	75775001	20	0.34	4.11	*LDB2*	76742996	76762841	20
4	75790001	75810001	20	0.4	4.44	*LDB2*	76777844	76797838	20
4	75835001	75920001	85	0.41 ± 0.043	4.73 ± 0.486	-	76822832	76911969	89
4	76400001	76430001	30	0.41 ± 0.026	4.01 ± 0.413	*CPEB2*	77391817	77421817	30
4	76755001	76790001	35	0.46 ± 0.051	4.46 ± 0.250	-	77742859	77777858	35
4	76815001	76860001	45	0.38 ± 0.059	4.00 ± 0.477	-	77802859	77847778	45
4	76905001	77285001	380	0.58 ± 0.167	8.84 ± 3.077	*BOD1L1, RAB28, ENSGALG00000028079*	77892778	78282988	390
4	77355001	77375001	20	0.66	4.44	-	78352988	78372935	20
4	77405001	77555001	150	0.67 ± 0.050		-	78402935	78552940	150
4	77600001	77625001	25	0.46 ± 0.035	5.49 ± 0.887	-	78598180	78623180	25
4	77685001	77705001	20	0.56	7.15	-	78683181	78703182	20
4	77910001	78060001	150	0.42 ± 0.076	5.02 ± 0.664	*SLC2A9*	78908078	79070963	163

4	78095001	78125001	30	0.30 ± 0.036	4.65 ± 0.471	*OTOP1, TMEM128, LYAR, ZBTB49, ZBTB49,*	79105895	79135795	30
4	78200001	78220001	20	0.34	4.35	*STX18*	79210794	79230793	20
4	78255001	78300001	45	0.52 ± 0.056	6.85 ± 0.544	-	79265793	79310788	45
4	78375001	78405001	30	0.33 ± 0.018	4.01 ± 0.177	*MSX1*	79385788	79415653	30
4	78495001	78515001	20	0.58	6.12	*STK32B*	79505655	79525655	20
4	78605001	78710001	105	0.45 ± 0.067	7.09 ± 0.706	*EVC2*	79615649	79721202	106
6	210001	230001	20	0.42	4.12	-	709627	729627	20
6	680001	730001	50	0.33 ± 0.029	4.11 ± 0.414	-	1179512	1229513	50
6	790001	810001	20	0.35	4.45	-	1289270	1309270	20
6	1320001	1400001	80	0.25 ± 0.121	3.52 ± 1.854	-	1819045	1899048	80
6	1470001	1500001	30	0.41 ± 0.027	5.58 ± 0.092	-	1969050	1999052	30
6	1565001	1620001	55	0.32 ± 0.046	4.14 ± 0.456	*ENSGALT00000034358*	2064050	2119052	55
6	1660001	1685001	25	0.35 ± 0.018	4.38 ± 0.105	-	2158969	2183969	25
6	1695001	1725001	30	0.37 ± 0.028	4.26 ± 0.156	-	2193971	2223956	30
6	1785001	1815001	30	0.27 ± 0.017	4.66 ± 0.248	*WAPL*	2283850	2313721	30
6	1900001	1955001	55	0.35 ± 0.039	4.97 ± 0.687	*LDB3*	2398723	2453723	55
6	2035001	2070001	35	0.32 ± 0.024	4.54 ± 0.360	*BMPR1A*	2534120	2569123	35
6	2510001	2530001	20	0.27	4.36	*LRIT1*	3009148	3029147	20
6	3055001	3075001	20	0.48	6.11	*NRG3*	3568845	3588846	20
6	3375001	3415001	40	0.42 ± 0.031	5.25 ± 0.394	-	3897034	3937034	40
6	3430001	3485001	55	0.40 ± 0.034	4.75 ± 0.481	-	3952035	4006883	55
6	4125001	4170001	45	0.32 ± 0.041	5.32 ± 0.794	-	4646814	4691814	45
6	4465001	4525001	60	0.35 ± 0.038	4.89 ± 0.472	-	4986908	5046912	60
6	4930001	4950001	20	0.34	4.08	-	5451521	5471521	20
6	6150001	6175001	25	0.40 ± 0.046	5.40 ± 0.222	*CTNNA3*	6715264	6740264	25
6	6525001	6550001	25	0.57 ± 0.023	4.37 ± 0.478	*CTNNA3*	7090264	7115263	25
6	6660001	6765001	105	0.51 ± 0.074	6.67 ± 0.518	-	7231892	7336894	105
6	6855001	6915001	60	0.59 ± 0.076	7.45 ± 0.490	-	7426895	7486894	60

7	22200001	22245001	45	0.67 ± 0.033	3.96 ± 0.185	-	22825997	22870996	45
8	22575001	22595001	20	0.41	4.28	*FAF1*	23753975	23773976	20
8	22700001	22720001	20	0.32	4.2	-	23879246	23899246	20
8	27525001	27545001	20	0.48	6.56	*SLC35D1, IL23R*	28722309	28742308	20
11	7030001	7050001	20	0.39	4.22	*PHKB*	7667109	7687109	20
11	7385001	7405001	20	0.33	4.04	-	8021802	8041802	20
11	7645001	7680001	35	0.27 ± 0.039	4.76 ± 0.831	-	8281804	8316804	35
11	9175001	9195001	20	0.4	4.32	-	9819648	9848202	29
11	9355001	9395001	40	0.31 ± 0.028	4.86 ± 0.454	*ANKRD27, RGS9BP*	10008037	10047931	40
11	9780001	9800001	20	0.28	4.2	*CEBPG*	10432773	10452903	20
11	10520001	10585001	65	0.36 ± 0.076	4.77 ± 0.970	*ENSGALT00000043640*	11172618	11237637	65
13	45001	70001	25	0.24 ± 0.029	4.60 ± 0.044	-	547133	560468	13
13	150001	170001	20	0.27	4.36	-	670972	690973	20
13	235001	255001	20	0.37 ± 0.124	6.11 ± 0.124	*PCDHA11*	755974	775974	20
13	430001	490001	60	0.36 ± 0.033	4.76 ± 0.677	*TMCO6*	981997	1044000	62
13	610001	645001	35	0.58 ± 0.038	5.46 ± 0.383	*PSD2*	1164030	1199031	35
13	780001	820001	40	0.43 ± 0.032	5.21 ± 0.607	*NRG2*	1334114	1374116	40
13	910001	965001	55	0.48 ± 0.030	4.70 ± 0.272	*PFDN1, HBEGF*	1464118	1519120	55
13	975001	1090001	115	0.41 ± 0.068	4.80 ± 0.575	*SLC4A9, ANKHD1*	1529030	1644035	115
13	2275001	2295001	20	0.34	5.39	-	2905101	2925101	20
17	5410001	5450001	40	0.54 ± 0.041	4.79 ± 0.487	*RP11-101E3.5, SH3GLB2*	5873557	5913557	40
18	95001	150001	55	0.57 ± 0.198	4.80 ± 1.216	*DHRS7C, GSG1L2, GLP2R*	144530	199529	55
18	180001	200001	20	0.61	4.94	*GAS7*	229528	249529	20
18	205001	225001	20	0.26	4.8	-	254529	274529	20
19	425001	480001	55	0.31 ± 0.029	4.88 ± 0.354	*BZRAP1*	443508	498508	55
19	505001	545001	40	0.41 ± 0.025	5.26 ± 0.566	*RNF43*	523512	563511	40
19	1405001	1445001	40	0.38 ± 0.039	5.33 ± 0.728	-	1383248	1423176	40
19	1520001	1595001	75	0.33 ± 0.017	4.41 ± 0.508	-	1498175	1573173	75
19	2965001	2990001	25	0.34 ± 0.016	4.35 ± 0.262	*SPNS2*	2937876	2962876	25

19	3345001	3365001	20	0.61	6.52	*CAMKK1*	3317773	3337773	20
19	4030001	4060001	30	0.47 ± 0.043	4.74 ± 0.584	*ENSGALT00000000279*	4003164	4033164	30
19	4515001	4545001	30	0.42 ± 0.056	4.89 ± 0.689	-	4489124	4519124	30
19	4595001	4625001	30	0.41 ± 0.037	4.96 ± 0.071	-	4569125	4599125	30
19	4695001	4725001	30	0.36 ± 0.028	4.51 ± 0.446	-	4669126	4699126	30
19	4805001	4835001	30	0.50 ± 0.067	6.71 ± 0.481	*ENSGALG00000001908, ENSGALG00000027051, ENSGALG00000002367, NF2L*	4774265	4810514	36
19	5145001	5165001	20	0.4	4.07	*NCBP3, CRK*	5131823	5151823	20
20	11075001	11110001	35	0.37 ± 0.054	5.28 ± 0.788	*SLC35C2, ELMO2, SNRPB, DDX27*	10932198	10967472	35
20	11130001	11150001	20	0.37	4.23	*EDN3*	10987472	11007476	20
20	11420001	11440001	20	0.33	4.4	*STX16*	11277277	11297277	20
20	11785001	11815001	30	0.37 ± 0.014	4.10 ± 0.072	-	11649020	11679019	30
20	11990001	12025001	35	0.42 ± 0.023	6.49 ± 0.289	*RAE1*	11855441	11890442	35
20	12200001	12315001	115	0.35 ± 0.075	4.94 ± 0.699	*TFAP2C, RTFDC1*	12065504	12180100	115
20	12355001	12375001	20	0.45	6.09	*CSTF1, AURKA, FAM210B*	12220100	12240100	20
20	12415001	12440001	25	0.28 ± 0.026	4.26 ± 0.150	*ENSGALT00000033520*	12280182	12305183	25
20	12525001	12560001	35	0.28 ± 0.024	4.40 ± 0.326	*ENSGALT00000035361*	12390183	12425184	35
20	12600001	12635001	35	0.49 ± 0.033	6.37 ± 0.236	*DOK5*	12465184	12500184	35
20	12655001	12690001	35	0.50 ± 0.028	6.74 ± 0.462	*PFDN4*	12520174	12555171	35
20	12720001	12755001	35	0.51 ± 0.029	6.37 ± 0.141	*BCAS1*	12585175	12620180	35
20	12800001	12905001	105	0.46 ± 0.066	4.73 ± 1.208	*ZNF217*	12665180	12779169	114
20	13095001	13115001	20	0.63	6.33	*TSHZ2*	12969198	12989209	20
20	13165001	13210001	45	0.60 ± 0.038	6.94 ± 0.484	-	13039195	13084177	45
20	13245001	13265001	20	0.27	4.16	-	13119185	13139187	20

20	13565001	13595001	30	0.48 ± 0.035	4.37 ± 0.432	-	13437401	13467401	30
20	13610001	13630001	20	0.4	4.11	-	13482400	13502399	20
20	13855001	13890001	35	0.61 ± 0.012	5.79 ± 0.351	*FAM65C, PTPN1*	13727210	13762210	35
21	2695001	2730001	35	0.48 ± 0.025	4.72 ± 0.339	*ENSGALT00000000312, AGRN*	2689010	2724112	35
21	3785001	3805001	20	0.33	4.22	*PEX14*	3781198	3801190	20
21	3810001	3855001	45	0.43 ± 0.089	5.01 ± 0.646	*CASZ1*	3806187	3851200	45
21	4275001	4295001	20	0.39	4.14	*SPEN*	4271609	4291606	20
21	4320001	4355001	35	0.49 ± 0.079	4.94 ± 0.471	*EPHA2, FBXO42*	4316607	4353202	37
26	1870001	1890001	20	0.46	6.04	*NFASC*	1917556	1937556	20
26	2040001	2075001	35	0.35 ± 0.027	4.9 ± 0.366	*ENSGALT00000035231*	2098561	2133500	35
26	2120001	2145001	25	0.34 ± 0.012	4.75 ± 0.347	*MFSD4A, ELK4*	2178411	2203412	25
26	2170001	2190001	20	0.29	4.17	*SLC45A3*	2228411	2248412	20
26	2215001	2245001	30	0.42 ± 0.070	5.16 ± 0.781	*RAB7L1, SLC41A1, MIR1797, gga-mir-6602*	2273412	2303394	30

Table S10B. Regions of ambiguity. Introgressed regions, which passed the genome-wide threshold (*Zfd* > 4.0) for the grey junglefowl - domestic chicken ABBA/BABA introgression analysis but for which the haplotype tree does not allow to identify unambigously the donors/introgressed species. The **Ggal** is the reference genome for 4 or 5 version and **Ln4** and **Ln5** are the length in kilobase for the **Ggal4** and **5** coordinates respectively. **Ggal4_μ(D), Ggal4_ μ*(fd)*** and **Ggal4_ μ*(Zfd)*** are the mean of D, *fd* and *Zfd* statistics based on *Galgal* 4.0 reference genome respectively

Chr	Ggal4_start	Ggal4_stop	Ln4	Ggal4_μ(D)	Ggal4_ μ(*fd*)	Ggal4_ μ (*Zfd*)	Ggal5_start	Ggal5_stop	Ln5	Possible Donor	Possible Receiver
1	77730001	77750001	20	0.38	0.23	9.77	78234538	78254537	20	Ceylon/Grey	Domestic
1	88835001	88860001	25	0.40 ± 0.030	0.20 ± 0.031	8.38 ± 1.313	4975	17909	13	Ceylon/Grey	Domestic
1	100455001	100550001	95	0.64 ± 0.078	0.23 ± 0.072	9.84 ± 3.047	101119900	101215093	95	Ceylon/Grey	Domestic
1	186290001	186320001	30	0.53 ± 0.021	0.19 ± 0.026	7.88 ± 1.084	187111135	187141553	30	Ceylon/Grey	Domestic
2	142360001	142395001	35	0.49 ± 0.010	0.20 ± 0.008	8.34 ± 0.352	No Map	No Map	No Map	Grey	Domestic
3	12170001	12190001	20	0.43	0.154	6.44	12815296	12835295	20	Grey	Domestic
4	27295001	27320001	25	0.53 ± 0.011	0.45 ± 0.002	6.07 ± 0.029	28038844	28063843	25	Domestic	Grey
4	45865001	45915000	50	0.31 ± 0.044	0.29 ± 0.022	4.08 ± 0.324	46765443	46815400	50	Grey/Domestic	Grey/Domestic
4	73505001	73525001	20	0.86	0.15	6.14	74386539	74405828	19	Grey	Domestic
4	85240001	85260001	20	0.2	0.14	6.02	No Map	No Map	No Map	Grey	Domestic
4	85380001	85415001	35	0.17 ± 0.018	0.18 ± 0.012	7.565 ± 0.525	86382559	86415596	33	Grey	Domestic
14	13625001	13690001	65	0.24 ± 0.077	0.24 ± 0.059	10.15 ± 2.510	No Map	No Map	No Map	Grey, Ceylon	Domestic
14	14255001	14315001	60	0.25 ± 0.046	0.28 ± 0.070	11.69 ± 2.961	14276883	14340251	63	Grey, Ceylon	Domestic
17	5420001	5445001	25	0.47 ± 0.030	0.16 ± 0.001	6.76 ± 0.062	5883558	5908557	25	Grey	Domestic
23	4005001	4035001	30	0.35 ± 0.032	0.16 ± 0.015	6.85 ± 0.635	4067896	4097908	30	Grey	Domestic

Table S11A. Population distribution and frequencies of grey junglefowl introgressed haplotypes into domestic chicken and red junglefowl. Analysis was based on ABBA – BABA test grey junglefowl – domestic chicken. The frequencies were calculated by dividing the total number of introgressed haplotypes in any population with the total number of haplotype present in that population. Total number of haplotypes indicates below each population. **Jarso** (Ethiopia), **Horro** (Ethiopia), **SA:** Saudi Arabia chicken**, SL:** Sri Lanka chicken**, Red:** Red junglefowl**, PdB:** Poule de Bresse**, MK:** Mechelse Koekoe**, MS:** Mechelse styrian.

Chr	Donor species	Ggal4_start	Ggal4_stop	Jarso n = 10	Horro n = 12	SA n = 10	SL n = 22	Red n = 12	PdB n = 2	MK n = 2	MS n = 2
1	Grey	15180001	15255001	-	-	-	4.5	-	-	-	-
1	Grey	18335001	18355001	-	-	10	-	-	-	-	
1	Grey	90920001	90940001	-	25	50	18.1	-	-	-	-
1	Grey	99665001	99740001	30	16.7	70	40.1	8.3	-	-	50
1	Grey	99915001	99935001	30	66.7	80	45.5	16.7	-	-	100
1	Grey	100195001	100240001	80	83.3	20	36.4	-	100	100	100
1	Grey	100785001	100885001	60	66.7	30	63.7	25	50	100	50
1	Grey	139220001	139245001	-	-	-	9.1	-	-	-	-
1	Grey	141700001	141755001	10	-	-	4.5	-	-	-	-
1	Grey	145600001	145630001	-	-	-	4.5	-	-	-	-
1	Grey	147640001	147660001	-	-	10	-	-	-	-	-
1	Grey	147760001	147810001	-	-	10	-	-	-	-	-
1	Grey	185245001	185300001	50	33.3	60	9.1	-	-	-	-
1	Grey	185435001	185485001	10	41.7	10	4.5	-	-	-	-
4	Grey	17015001	17035001	-	-	-	4.5	-	-	-	-
4	Grey	33685001	33720001	50	-	-	-	-	-	-	50
4	Grey	37045001	37070001	-	8.3	-	4.5	-	-	50	-
4	Grey	41895001	41930001	-	-	-	31.8	-	-	-	-
4	Grey	41960001	41985001	-	-	-	4.5	8.3	-	-	-
4	Grey	42020001	42040001	-	-	-	4.5	8.3	-	-	-

4	Grey	42085001	42115001	-	-	-	9.1	8.3	-	-	-
4	Grey	45220001	45255001	-	-	-	4.5	-	-	50	-
4	Grey	45865001	45915001	20	33.3	40	-	-	100	50	50
4	Grey	45950001	45970001	20	-	-	-	-	-	-	-
4	Grey	46970001	47005001	-	-	-	9.1	-	-	-	-
6	Grey	1260001	1280001	-	25	-	-	-	-	-	-
6	Grey	1370001	1400001	10	-	-	-	-	-	-	-
6	Grey	1415001	1445001	-	-	-	13.6	-	-	-	-
7	Grey	22200001	22245000	60	83.3	-	18.2	-	-	-	-
13	Grey	2165001	2190001	-	-	-	9.1	-	-	-	
13	Grey	2395001	2415001	10	-	20	-	-	-	-	-
19	Grey	1635001	1680001	-	-	-	4.5	-	-	-	-
20	Grey	13345001	13370001	-	-	-	4.5	-	-	-	-
20	Grey	13505001	13540001	-	-	-	4.5	-	-	-	-
21	Grey	3005001	3030001	-	-	-	9.1	-	-	-	-
21	Grey	3230001	3250001	-	-	-	4.5	-	-	-	-
21	Grey	3335001	3360001	-	-	-	4.5	-	-	-	-
21	Grey	3495001	3520001	-	-	-	9.1	-	-	-	-
21	Grey	3570001	3595001	-	-	30	4.5	-	-	-	-
7	Grey	22200001	22245001	60	83.3	-	18.2	-	-	-	-
24	Grey	6130001	6160000	100	100	90	86	33.3	-	50	50
28	Grey	2110001	2130001	20	41.6	-	9.1	-	-		50
Number of candidate introgressed regions				16	14	14	33	7	3	6	8
Total number of introgressed haplotypes for all regions				62	85	53	113	13	5	8	10
% of introgressed haplotypes				38.8	50.6	37.9	15.5	15.5	83.3	66.7	62.5

Table S11B. Population distribution and frequencies of green junglefowl introgressed haplotypes into domestic. Analysis was based on ABBA – BABA test grey junglefowl – domestic chicken. The frequencies were calculated by dividing the total number of introgressed haplotypes in any population with the total number of haplotype present in that population. Total number of haplotypes indicates below each population. **Jarso** (Ethiopia), **Horro** (Ethiopia), **SA:** Saudi Arabia chicken**, SL:** Sri Lanka chicken**, Red:** Red junglefowl**, PdB:** Poule de Bresse**, MK:** Mechelse Koekoe**, MS:** Mechelse styrian.

Chr	**Donor species**	**Ggal4_start**	**Ggal4_stop**	**Jarso n = 10**	**Horro n = 12**	**SA n = 10**	**SL n = 22**	**Red n = 12**	**PdB n = 2**	**MK n = 2**	**MS n = 2**
2	Green	18310001	18345001	-	-	-	9.1	-	-	-	-
2	Green	18365001	18395001	-	-	-	9.1	-	-	-	-
Number of candidate introgressed regions				0	0	0	2	0	0	0	0
Total number of introgressed haplotypes for all regions				0	0	0	4	0	0	0	0
% of introgressed haplotypes				0	0	0	11	0	0	0	0

Table S12A. Characteristics of the introgressed regions from domestic chicken into Ceylon junglefowl, ABBA – BABA test Ceylon junglefowl - domestic chicken (n = 2). The **Ggal** is the reference genome for 4 or 5 version and **Ln4** and **Ln5** are the length in kilobase for the **Ggal4** and **5** coordinates respectively. **Ggal4_μ(D)** and **Ggal4_ μ(*Zfd*)** are the mean of D and *Zfd* statistics based on *Galgal* 4.0 reference genome respectively

Chr	**Ggal4_start**	**Ggal4_stop**	**Ln4**	**Ggal4_μ(D)**	**Ggal4_μ(*Zfd*)**	**Ggal4_Genes**	**Ggal5_start**	**Ggal5_stop**	**Ln5**	**Ggal5_Genes**
2	76905001	76935001	30	0.67 ± 0.053	6.18 ± 0.038	-	77161051	77186542	25.5	
24	6120001	6190001	70	0.52 ± 0.158	11.97 ± 5.223	*PTS, BCDO2, IL18, AP002884.2, C11orf57, PIH1D2, DLAT*	6097101	6167118	70	*PTS, BCDO2, IL18, C11orf57, DLAT*

Table S12B. Regions of ambiguity Introgressed regions, which passed the genome-wide threshold (*Zfd* > 6.0) for the Ceylon junglefowl domestic chicken ABBA - BABA introgression analysis but for which the haplotype tree does not allow to identify unambiguously the donors/introgressed species. The **Ggal** is the reference genome for 4 or 5 version and **Ln4** and **Ln5** are the length in kilobase for the **Ggal4** and **5** coordinates respectively. **Ggal4_μ(D), Ggal4_ μ(*fd*)** and **Ggal4_ μ(*Zfd*)** are the mean of D, *fd* and *Zfd* statistics based on *Galgal* 4.0 reference genome respectively

Chr	Ggal4_start	Ggal4_stop	Ln4	Ggal4_ μ(D)	Ggal4_ μ(*fd*)	Ggal4_ μ(*Zfd*)	Ggal5_start	Ggal5_stop	Ln5	Possible Donor	Possible Receiver
1	23740001	23760001	20	0.76	0.148	6.22	23970099	23990098	20	Ceylon	Domestic
1	77730001	77750001	20	0.38	0.23	9.77	78234538	78254537	20	Ceylon/Grey	Domestic
1	88835001	88860001	25	0.40 ± 0.030	0.20 ± 0.031	8.38 ± 1.313	4975	17909	13	Ceylon/Grey	Domestic
1	185510001	185530001	20	0.59		6.06	186320605	186340604	20	Ceylon/Grey	Domestic
1	100455001	100550001	95	0.64 ± 0.078	0.23 ± 0.072	9.84 ± 3.047	101119900	101215093	95	Ceylon/Grey	Domestic
1	186290001	186320001	30	0.53 ± 0.021	0.19 ± 0.026	7.88 ± 1.084	187111135	187141553	30	Ceylon/Grey	Domestic
2	81835001	81860001	25	0.55 ± 0.023	0.18 ± 0.033	7.36 ± 1.387	82085320	82100687	15	Ceylon	Domestic
2	142360001	142395001	35	0.49 ± 0.010	0.20 ± 0.008	8.34 ± 0.352	No Map	No Map	No Map	Grey	Domestic
3	12170001	12190001	20	0.43	0.154	6.44	12815296	12835295	20	Grey	Domestic
4	73505001	73525001	20	0.86	0.15	6.14	74386539	74405828	19	Grey	Domestic
4	85380001	85415001	35	0.17 ± 0.018	0.18 ± 0.012	7.565 ± 0.525	86382559	86415596	33	Grey	Domestic
14	13625001	13690001	65	0.25 ± 0.046	0.28 ± 0.070	11.69 ± 2.961	No Map	No Map	No Map	Grey, Ceylon	Domestic
14	14255001	14315001	60	0.47 ± 0.030	0.16 ± 0.001	6.76 ± 0.062	14276883	14340251	63	Grey	Domestic
17	5420001	5445001	25	0.41 ± 0.024	0.16 ± 0.014	6.81 ± 0.628	5883558	5908557	25	Ceylon/Grey	Domestic
19	2680001	2735000	55	0.35 ± 0.032	0.16 ± 0.015	6.85 ± 0.635	2668450	2723449	55	Grey	Domestic
23	4005001	4035001	30	0.45 ± 0.012	0.15 ± 0.002	6.12 ± 0.072	4067896	4097908	30	Grey	Domestic
25	325001	350001	25	0.76	0.148	6.22	1301102	1328686	28	Ceylon	Domestic

Table S13A. Population distribution and frequencies of introgressed haplotypes from the grey junglefowl into domestic chicken. Analysis was based on ABBA – BABA test Ceylon junglefowl – domestic chicken. The region in bold shows two directional introgression, grey junglefowl into domestic chicken and same region also shows in Ceylon junglefowl into domestic chicken. The frequencies were calculated by dividing the total number of introgressed haplotypes in any population with the total number of haplotype present in that population. Total number of haplotypes indicates below each population. **Jarso** (Ethiopia), **Horro** (Ethiopia), **SA:** Saudi Arabia chicken**, SL:** Sri Lanka chicken**, Red:** Red junglefowl**, PdB:** Poule de Bresse**, MK:** Mechelse Koekoe**, MS:** Mechelse styrian.

Chr	Donor species	Ggal4_start	Ggal4_stop	Jarso n = 10	Horro n = 12	SA n = 10	SL n = 22	Red n = 12	PdB n = 2	MK n = 2	MS n = 2
1	Grey	13070001	13095001	50	41.7	30	-	-	-	-	-
1	Grey	14950001	15005001	-	-	-	9.1	-	-	-	-
1	**Grey**	**25800001**	**25835001**	**70**	**50**	**-**	**-**	**-**	**-**	**-**	**-**
1	Grey	52835001	52865001	20	33.3	60	13.6	-	-	-	-
1	Grey	90490001	90530001	-	58.3	40	13.6	-	-	-	-
1	Grey	192750001	192770001	10	33.3	50	4.5	-	100	50	-
2	Grey	79185001	79205001	30	-	-	18.2	-	-	-	-
2	Grey	115180001	115295001	50	75	-	13.1	-	50	-	-
3	Grey	106610001	106645001	30	58.3	30	9.1	-	-	-	-
4	Grey	2195001	2225001	50	66.7	10	-	-	-	-	50
4	Grey	12010001	12060001	60	16.7	20	9.1	-	-	-	50
4	Grey	13130001	13180001	-	16.7	50	31.8	-	50	-	50
4	Grey	13265001	13285001	60	33.3	30	13.6	-	50	-	50
5	Grey	44955001	44975000	50	16.7	10	13.6	-	-	50	-
7	Grey	22795001	22825001	40	50	20	4.5		100	100	-
11	Grey	14570001	14590001	50	8.3	-	45.5	-	50	-	-
Number of candidate introgressed regions				13	14	11	13	0	6	3	4
Total number of introgressed haplotypes for all regions				57	67	35	44	0	8	4	4

% of introgressed haplotypes	43.8	39.9	31.8	15.4	0	66.7	66.7	50

Table S13B. Population distribution and frequencies of introgressed haplotypes from the Ceylon junglefowl into domestic chicken. Analysis was based on ABBA – BABA test Ceylon junglefowl – domestic chicken. The region in bold shows two directional introgression, Ceylon junglefowl into domestic chicken and same region also shows in grey junglefowl into domestic chicken. The frequencies were calculated by dividing the total number of introgressed haplotypes in any population with the total number of haplotype present in that population. Total number of haplotypes indicates below each population. **Jarso** (Ethiopia), **Horro** (Ethiopia), **SA:** Saudi Arabia chicken**, SL:** Sri Lanka chicken**, Red:** Red junglefowl**, PdB:** Poule de Bresse**, MK:** Mechelse Koekoe**, MS:** Mechelse styrian.

Chr	Donor species	Ggal4_start	Ggal4_stop	Jarso n = 10	Horro n = 12	SA n = 10	SL n = 22	Red n = 12	PdB n = 2	MK n = 2	MS n = 2
1	Ceylon	15515001	15580001	50	66.7	20	4.5	-	-	-	-
1	Ceylon	24985001	29480001	-	-	-	4.5	-	-	-	-
1	Ceylon	145735001	147155001	-	-	-	4.5	-	-	-	-
4	Ceylon	23930001	23985000	-	8.3	-	4.5	-	-	-	-
Number of candidate introgressed regions				1	2	1	4	0	0	0	0
Total number of introgressed haplotypes for all regions				5	9	2	4	0	0	0	0
% of introgressed haplotypes				12.5	18.8	5	4.5	0	0	0	0

Table S14. Topology weightings estimation for all relationships between taxa at the chromosome level (see **Figure 5A**). Each value corresponds to the proportion of admixture between taxa. The **Red vs Domestic** has the highest admixture fraction or genome similarity across the autosomes. I expect slight deviation from 1 due to approximation criteria. However, all the fifteen possible topologies are expected to sum to 1. The **Ceylon, Grey** and **Red** are the three junglefowls included in the topology weightings. The **domestic** include the 30 domestic chickens.

Chr	**Ceylon vs Domestic (%)** (topo1 + topo2 + topo3)	**Red vs Domestic (%)** (topo4 + topo5 + topo6)	**Grey vs Domestic (%)** (topo7 + topo8 + topo9)	**Red vs Grey vs Ceylon (%)** (topo10 + topo11 + topo12)	**Red vs Grey vs Ceylon (%)** (topo13 + topo14 + topo15)
1	2.21	74.45	6.14	6.71	10.59
2	2.37	74.97	5.75	6.61	10.30
3	1.92	79.25	3.31	6.68	8.84
4	2.71	63.80	12.92	6.11	14.46
5	2.15	80.05	2.57	6.52	8.71
6	2.31	71.65	7.83	6.51	11.69
7	1.79	80.52	2.25	6.79	8.65
8	1.49	81.06	2.52	6.44	8.49
9	1.70	79.24	3.11	6.84	9.12
10	1.56	79.51	2.73	7.03	9.17
11	2.24	72.51	7.80	6.00	11.45
12	1.87	79.01	4.48	7.26	10.38
13	2.04	72.33	7.17	6.67	11.80
14	1.28	81.12	1.91	7.08	8.61
15	1.48	78.57	4.79	6.00	9.16
16	12.89	33.57	13.38	20.01	20.15
17	1.61	75.83	5.31	7.02	10.24
18	2.55	66.71	9.56	7.09	14.10
19	3.29	55.11	18.78	5.78	17.05

20	1.69	71.89	7.87	6.85	11.70
21	2.79	61.28	13.96	6.76	15.21
22	0.74	89.04	0.90	4.35	4.96
23	2.73	73.70	5.10	7.28	11.20
24	3.74	70.98	4.95	8.13	12.19
25	1.37	79.40	2.01	7.61	9.61
26	1.82	70.71	7.49	7.78	12.21
27	2.69	68.45	7.54	8.46	12.86
28	1.24	80.30	2.00	7.50	8.96

Table S15. Overrepresented gene ontology terms and KEGG pathways for grey junglefowl introgressed genes into domestic chicken

Term ID	**Term description**	***P* value**	**Genes**
GO:0004672	Protein kinase activity	0.044	*FASTKD3, RELN, EPHA6, GUCY2F*

Table S16. GO and KEGG enrichment analyses for the introgression of domestic chicken into grey junglefowl

Term	**Term description**	***P* value**	**Genes**
GO:0001945	Lymph vessel development	0.001	*EFNB2, PTPN14, PROX1, EPHA2*
GO:0007399	Nervous system development	0.007	*SLC45A3, RAP2A, NRP1, FGF14, STK24, EFNB2, NFASC, ARMC4, AURKA, REST, SPEN, PROX1, EPHA2, FARP1, PCDH18, DCT, ANKRD27, LPHN3, SGK196, MSX1, DOK5, TENM3, HBEGF, BMPR1A*
GO:0005057	Receptor signaling protein activity	0.013	*STK24, DOK5, NRK, MAPK10, BMPR1A*
GO:0016772	Transferase activity, transferring phosphorus-containing groups	0.015	*NRP1, MYO3A, STK24, AURKA, MAPK10, CDS1, CDKL2, EPHA2, CAMKK1, POLR2B, SGK196, STK32B, SNRK, NRK, BMPR1A*
GO:0016021	Integral component of membrane	0.021	*HGSNAT, SLC45A3, NRP1, SLC15A1, GALNT7, BTC, MFSD4A, PTCHD3, FAM210B, GRAMD4, STT3B, TMEM128, DCT, TECRL, SLC35D1, STX18, STX16, PCDHA11, TMEM192, NALCN, IL1RAPL2, DNAJC1, FLVCR1, SMIM19, SLC2A9, GPR158, MSMO1, EFNB2, ASIC5, NFASC, CDS1, SLC10A2, RGS9BP, PCDH18, SLC35C2, EVC2, SGK196, GRIA2, TENM3, PLXDC2, CMTM7, ABCC4*
GO:0001078	Transcriptional repressor activity, RNA polymerase II core promoter proximal region sequence-specific binding	0.024	*TSHZ2, TFAP2C, REST, PROX1*
GO:0016301	Kinase activity	0.024	*SGK196, STK32B, NRP1, MYO3A, SNRK, STK24, NRK, AURKA, MAPK10, CDKL2, EPHA2, CAMKK1, BMPR1A*
GO:0016814	Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in cyclic amidines	0.025	*DCTD, MTHFD2L, ADARB2*
GO:0044089	Positive regulation of cellular component biogenesis	0.027	*LPHN3, NRP1, STX18, LDB2, MPP7, FAF1, ARFIP1, PROX1*

GO:0035051	Cardiocyte differentiation	0.034	*EFNB2, REST, PROX1, NEBL*
GO:0048013	Ephrin receptor signaling pathway	0.035	*EFNB2, CRK*
GO:0004674	Protein serine/threonine kinase activity	0.036	*STK32B, SNRK, STK24, NRK, AURKA, MAPK10, CAMKK1, BMPR1A*
GO:0070309	Lens fiber cell morphogenesis	0.036	*PROX1, EPHA2*
GO:0022857	Transmembrane transporter activity	0.037	*SLC45A3, SLC2A9, SLC15A1, GLRA3, ASIC5, MFSD4A, REST, SLC10A2, SLC7A11, GRIA2, SLC35D1, ABCC4, NALCN, SLC4A4*
GO:0015144	Carbohydrate transmembrane transporter activity	0.042	*SLC45A3, SLC2A9, MFSD4A*
GO:0005829	Cytosol	0.044	*RAP2A, NRP1, PRTFDC1, ABHD5, WWC2, REST, ARFIP1, ELMO2, FARP1, ANKRD27, DCT, JADE1, USO1, SNRPB, FAF1, DNAJC1*
GO:0051179	Localization	0.044	*HGSNAT, DYNC1LI1, SDAD1, NRP1, ABHD5, AURKA, REST, LPHN3, WAPL, STX18, RAB28, STX16, PEX14, NALCN, SLC4A4, NCBP3, DNAJC1, FLVCR1, RAP2A, GPR158, STK24, EFNB2, NFASC, PTPN14, ARMC4, LDB2, MPP7, ARFIP1, DHRS7C, PROX1, CDKL2, EPHA2, SLC7A11, CTNNA3, ELMO2, SLC35C2, ANKRD27, SGK196, MSX1, USO1, ABCC4, OSBPL10, HBEGF, FAF1, BMPR1A*
GO:0006493	Protein O-linked glycosylation	0.045	*SLC35C2, SGK196, GALNT7*
GO:0098609	Cell-cell adhesion	0.046	*LPHN3, SH3GLB2, ELMO2*
GO:0051100	Negative regulation of binding	0.047	*WAPL, MSX1, PEX14, AURKA*
GO:0034613	Cellular protein localization	0.047	*RAP2A, GPR158, NRP1, NFASC, PTPN14, AURKA, ARFIP1, CDKL2, EPHA2, WAPL, MSX1, STX18, STX16, USO1, PEX14, FAF1, BMPR1A*
GO:0000981	RNA polymerase II transcription factor activity, sequence-specific DNA binding	0.048	*TSHZ2, MSX1, KLF12, ELK4, CEBPG, TFAP2C, REST, PROX1, BMPR1A*

GO:0070851	Growth factor receptor binding	0.048	*EPGN, BTC, HBEGF*
GO:0048845	Venous blood vessel morphogenesis	0.048	*EFNB2, PROX1*
GO:0060284	Regulation of cell development	0.049	*SLC45A3, DCT, ANKRD27, RAP2A, NRP1, STK24, AURKA, REST, SPEN, PROX1, BMPR1A*
GO:0060326	Cell chemotaxis	0.049	*HBEGF, ELMO2, EPHA2*
GO:0000982	Transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding	0.050	*TSHZ2, MSX1, CEBPG, TFAP2C, REST, PROX1*
KEGG:gga04012	Erbb signaling pathway	0.002	*NRG3, BTC, HBEGF, MAPK10, CRK, NRG2*

Table S17A. Overlap with QTL: Overlap between domestic chicken introgressed regions from grey junglefowl and reported QTLs (https://www.animalgenome.org/cgi-bin/QTLdb/GG/index). ABBA – BABA test grey junglefowl – domestic chicken

Chr	Ggal4_start	Ggal4_start stop	*Galgal* 4.0 QTL overlap
1	15180001	15255000	Body weight, pectoralis major weight, chest width, fear-stand latency, abdominal fat weight, subcutaneous fat thickness, body weight
1	18335001	18355001	
1	90920001	90940000	Subcutaneous neck fat weight, body weight, spleen weight, age at first egg, shank length, chest width, breast muscle percentage, feed efficiency
1	99665001	99740000	Subcutaneous neck fat weight, body weight, spleen weight, age at first egg, shank length, breast muscle percentage, egg production rate
1	99915001	99935000	Subcutaneous neck fat weight, body weight, spleen weight, age at first egg, shank length
1	100195001	100240001	Subcutaneous neck fat weight, body weight, spleen weight, age at first egg, shank length, bone index, breast muscle weight
1	100785001	100885000	Subcutaneous neck fat weight, body weight, spleen weight, age at first egg, shank length, bone index, breast muscle weight,
1	139220001	139245000	Spleen weight, age at first egg, shank length, breast pH, leg bowing, carcass protein content, body weight
1	141700001	141755000	Spleen weight, age at first egg, shank length, body weight
1	145600001	145630000	Spleen weight, age at first egg, shank length, body weight, egg shell stiffness
1	147640001	147660000	Spleen weight, age at first egg, shank length, body weight, egg shell stiffness
1	147760001	147810000	Spleen weight, shank length, body weight, egg shell stiffness
1	185245001	185300000	Body weight, abdominal fat weight, intestine length, insulin-like growth factor level, drip loss, feed intake
1	185435001	185485000	Body weight, abdominal fat weight, intestine length, insulin-like growth factor level, drip loss, feed intake
4	17015001	17035000	Egg weight, tibia bone mineral density, earlobe colour, residual feed intake, Marek's disease-related traits

4	33685001	33720000	Marek’s disease-related traits, drumstick and thigh muscle weight, egg weight and number, tibia width and weight
4	37045001	37070000	Marek’s disease-related traits, drumstick and thigh muscle weight, egg weight and number, tibia width and weight, tibia plateau angle, body weight,
4	41895001	41930000	Egg weight and number, thigh meat-to-bone ratio, fear-headflicks, abdominal fat weight
4	41960001	41985001	Average daily gain, egg weight, age at first egg, drumstick and thigh muscle weight, egg weight and number, tibia width, feed intake, Marek’s disease-related traits, thigh meat to bone ratio, abdominal fat weight, visceral fat weight
4	42020001	42040001	Average daily gain, egg weight, age at first egg, drumstick and thigh muscle weight, egg weight and number, tibia width, feed intake, Marek’s disease-related traits, thigh meat to bone ratio, abdominal fat weight, visceral fat weight
4	42085001	42115000	Average daily gain, egg weight, age at first egg, drumstick and thigh muscle weight, egg weight and number, tibia width, feed intake, Marek’s disease-related traits, thigh meat to bone ratio, abdominal fat weight, visceral fat weight
4	45220001	45255000	Average daily gain, egg weight, age at first egg, drumstick and thigh muscle weight, egg weight and number, tibia width, feed intake, Marek’s disease-related traits, thigh meat to bone ratio, abdominal fat weight, visceral fat weight
4	45950001	45970000	Average daily gain, egg weight, age at first egg, drumstick and thigh muscle weight, egg weight and number, tibia width, feed intake, Marek’s disease-related traits, thigh meat to bone ratio, abdominal fat weight, visceral fat weight
4	46970001	47005000	Skin fat weight
6	1260001	1280000	
6	1370001	1400000	
6	1415001	1445001	
7	22200001	22245000	Abdominal fat weight, intestine length, Marek’s disease-related traits, wing weight, egg number, albumen height, body weight, feed conversion ratio
13	2165001	2190001	Breast muscle weight, age at first egg
13	2395001	2415001	Breast muscle weight, age at first egg
19	1635001	1680001	Feed intake, dry matter intake and shank diameter
20	13345001	13370001	Body weight and breast pH
20	13505001	13540001	Body weight and breast pH

21	3005001	3030001	
21	3230001	3250001	
21	3335001	3360001	
21	3495001	3520001	
21	3570001	3595001	
7	22200001	22245000	Abdominal fat and body weight, feed conversion ratio, intestine length, Marek’s disease-related traits, wing weights, egg number, albumen height
24	6130001	6160000	Egg number
28	2110001	2130000	Egg shell colour, abdominal fat weight, average daily gain, body weight, antibody response to E. coli, abdominal fat percentage

Table S17B. Overlap between domestic chicken introgressed regions from green junglefowl and reported QTLs. https://www.animalgenome.org/cgi-bin/QTLdb/GG/index). ABBA – BABA test grey junglefowl – domestic chicken

Chr	**Ggal4_start**	**Ggal4_start stop**	***Galgal* 4.0 QTL overlap**
2	18310001	18345001	Body weight, insulin level, earlobe colour, drumstick and thigh muscle weight, insulin-like growth factor level, Marek's disease-related traits, feed intake
2	18365001	18395001	Body weight, insulin level, earlobe colour, drumstick and thigh muscle weight, insulin-like growth factor level, Marek's disease-related traits, feed intake

Table S17C. Overlap between domestic chicken introgressed regions from grey junglefowl and reported QTLs (https://www.animalgenome.org/cgi-bin/QTLdb/GG/index). ABBA – BABA test Ceylon junglefowl – domestic chicken

Chr	Ggal4_start	Ggal4_start stop	*Galgal* 4.0 QTL overlap
1	13070001	13095000	Body weight, pectoralis major weight, chest width, fear-stand latency, abdominal fat weight, subcutaneous fat thickness, body weight
1	14950001	15005000	Body weight, pectoralis major weight, chest width, fear-stand latency, abdominal fat weight, subcutaneous fat thickness, body weight
1	25800001	25835000	Shank length, visceral peritoneum pigmentation, chest width, factor fear, abdominal fat weight, subcutaneous neck fat weight, body weight
1	52835001	52865000	Shank length, subcutaneous neck fat weight, body weight, feed intake, spleen weight, age at first egg, drumstick and thigh weight, egg number, haugh units, femur and tibia bones mineral density, insulin-like growth factor level
1	90490001	90530000	Subcutaneous neck fat weight, body weight, spleen weight, age at first egg, drumstick and thigh weight, shank length, chest width, breast muscle percentage, feed efficiency
1	192750001	192770000	Body weight, abdominal fat weight, intestine length, insulin-like growth factor level, fear-call latency
2	79185001	79205000	Body weight, abdominal fat weight, intestine length, insulin-like growth factor level
2	115180001	115295000	Cloacal bacterial burden after challenge with *Salmonella* E, shank diameter and length, shank growth, body weight, egg shell strength, body weight
3	106610001	106645001	
4	2195001	2225000	Shank growth, age at first egg, visceral fat weight, average daily gain
4	12010001	12060000	Shank growth, age at first egg, visceral fat weight, average daily gain
4	13130001	13180000	Shank growth, age at first egg, visceral fat weight, average daily gain
4	13265001	13285000	Shank growth, age at first egg, visceral fat weight, average daily gain
5	44955001	44975000	Shank diameter, egg number, carcase, egg, and body weight, growth, non-esterified fatty acid level, abdominal fat weight, head width, receiving feather pecking, feed intake
7	22795001	22825000	Abdominal fat and body weight, feed conversion ratio, intestine length, Marek's disease-related traits, wing weights, egg number, albumen height, heart weight

11	14570001	14590000	Carcass weight, body weight and temperature, breast colour, carcass and spleen weight, tibia width, dry matter intake

Table S17D. Overlap between domestic chicken introgressed regions from Ceylon junglefowl and reported QTLs (https://www.animalgenome.org/cgi-bin/QTLdb/GG/index). ABBA – BABA test Ceylon junglefowl – domestic chicken

Chr	Ggal4_start	Ggal4_start stop	Galgal4 QTL overlap
1	15525001	15555000	Body weight, pectoralis major weight, fear-stand latency, abdominal fat weight, shank length, visceral peritoneum pigmentation, chest width, leg twisting
1	25800001	25835000	Body weight, pectoralis major weight, fear-stand latency, abdominal fat weight, shank length, visceral peritoneum pigmentation, chest width, leg twisting
1	145735001	145810000	Body weight, breast pH, leg bowing, carcass protein content, egg shell stiffness
1	146075001	146095000	Body weight, breast pH, leg bowing, carcass protein content, egg shell stiffness
1	146460001	146485000	Body weight, breast pH, leg bowing, carcass protein content, egg shell stiffness
1	146505001	146525000	Body weight, breast pH, leg bowing, carcass protein content, egg shell stiffness
1	146765001	146805000	Body weight, breast pH, leg bowing, carcass protein content, egg shell stiffness
1	146965001	147005000	Body weight, breast pH, leg bowing, carcass protein content, egg shell stiffness
1	147060001	147080000	Body weight, breast pH, leg bowing, carcass protein content, egg shell stiffness
1	147605001	147635000	Body weight, breast pH, leg bowing, carcass protein content, egg shell stiffness
1	185510001	185530001	Body weight, average daily gain, skin fat weight, intestine length, growth, feed intake, dry matter intake, abdominal fat weight, carcass fat content, shank weight percentage, drip loss, insulin-like growth factor level
3	99500001	99575001	Average daily gain, carcass weight, thigh muscle weight, tibia plateau angle, chest width, body weight, tibia bone mineral density, growth, plasma coloration, tibia area and length,
4	23965001	23985000	Abdominal fat weight, average daily gain, egg weight, age at first egg, drumstick and thigh muscle weight, Marek's disease-related traits, Shank length, egg weight and number

Chapter 5

Introgression of Javan red junglefowl *Gallus g. bankiva* to domestic chicken: Insights from autosomal whole-genome sequences

Table S1. Summary sequencing statistics and numbers of SNPs

Samples	Reads	PPR (%)	MR (%)	MD (X)	5_b (%)	10_b (%)	20_b (%)	Hom	Het	Total SNPs
SNPs information for domestic chicken from Ethiopia, Saudi Arabia, Sri Lanka, Fancy birds, red junglefowl, grey junglefowl, Ceylon junglefowl and green junglefowl are found in the supplementary Table S1 of chapter 2, supplementary Table S1 of chapter 3, and supplementary Table S2 of chapter 4										
KeduHitam761[a]	79614260	97.81	99.06	7.44	78.1	21.9	0.6	2406564	3187633	5594197
KeduHitam763[a]	81528449	97.74	99.02	7.62	79.5	23.9	0.7	2390621	3298540	5689161
KeduHitam766[a]	83209097	97.93	99.12	7.78	80.9	26.0	0.8	2852954	2482037	5334991
BankivaJ650[b]	84012601	96.86	98.36	7.71	80.0	25.2	0.6	4140951	2829451	6970402
BankivaJ759[b]	81158057	96.94	98.48	7.51	78.9	22.4	0.7	4103784	2882617	6986401
BankivaJ760[b]	79044674	96.45	98.29	7.27	76.8	19.3	0.6	4099712	2753108	6852820

[a] Domestic chicken, [b] Javan red junglefowl *Gallus gallus bankiva.* **PPR:** Percentage of both mates of a read pair that were properly mapped to the same chromosome. **MR:** Total reads mapped to *Galgal* 4.0 reference genome. **MD:** The mean sequence depth or the genome sequence coverage. **5_b:** Percentage of the genome with bases covered by at least 5 reads. **10_b:** Percentage of the genome with bases covered by at least 10 reads. **20_b:** Percentage of the genome with bases covered by at least 20 reads. **Hom**: Homozygous SNPs compare to the reference. **Het:** Heterozygous SNPs compare to the reference. The **Total SNPs** include the bi-allelic **Hom** + **Het.**

Table S2. Characteristics of the Introgressed regions from domestic chicken into Javan red junglefowl (n = 39). **Tables 2** and **3** also include bidirectional introgression where some haplotypes contained domestic origin

Chr	Start	Stop	kb[d]	D (mean)	*fd* (mean)	*Zfd* (mean)	Overlapping Genes
1	31510000	31530000	20	0.48	0.33	6.38	
1	58890000	58920000	30	0.68 ± 0.058	0.41 ± 0.066	8.17 ± 1.476	
1	62605000	62625000	20	0.33	0.35	6.88	
1 [a]	100590000	100615000	25	0.61 ± 0.065	0.4 ± 0.063	7.89 ± 1.418	
1	102680000	102700000	20	0.25	0.33	6.3	
1	102870000	102890000	20	0.42	0.35	6.71	
1	122100000	122120000	20	0.47	0.39	7.65	
1	125700000	125725000	25	0.47 ± 0.047	0.34 ± 0.034	6.61 ± 0.761	
1 [b]	141305000	141330000	25	0.8 ± 0.004	0.34 ± 0.005	6.68 ± 0.111	
1 [b]	149755000	149775000	20	0.59	0.32	6.24	
1 [b, a]	152125000	152150000	25	0.35 ± 0.022	0.33 ± 0.01	6.37 ± 0.231	
1 [b]	156380000	156410000	30	0.49 ± 0.07	0.39 ± 0.054	7.62 ± 1.202	
1 [b]	160905000	160930000	25	0.46 ± 0.014	0.41 ± 0.065	8.19 ± 1.453	
1 [b]	161815000	161840000	25	0.46 ± 0.036	0.4 ± 0.015	7.91 ± 0.341	*DIAPH3*
2	54905000	54935000	30	0.8 ± 0.054	0.47 ± 0.007	9.46 ± 0.163	
2 [a]	59610000	59630000	20	0.34	0.33	6.3	
2 [a]	95680000	95710000	30	0.58 ± 0.052	0.34 ± 0.008	6.54 ± 0.186	
2	121930000	121965000	35	0.69 ± 0.013	0.46 ± 0.053	9.24 ± 1.181	
3	32160000	32180000	20	0.31	0.34	6.65	
3	89105000	89125000	20	0.53	0.33	6.45	*CSMD1, gga-mir-6634*
3	97745000	97765000	20	0.72	0.32	6.12	
4 [b]	20160000	20185000	25	0.48 ± 0.065	0.36 ± 0.03	7.08 ± 0.667	
4 [b, a]	39245000	39270000	25	0.36 ± 0.023	0.36 ± 0.016	7.06 ± 0.36	*ING2*
5	30970000	30990000	20	0.58	0.45	8.98	
5	31115000	31155000	40	0.55 ± 0.077	0.37 ± 0.038	7.17 ± 0.86	

5	59380000	59400000	20	0.5	0.46	9.2	*LRFN5*
7	6085000	6115000	30	0.54 ± 0.037	0.38 ± 0.016	7.46 ± 0.351	
7	19020000	19050000	30	0.67 ± 0.071	0.38 ± 0.011	7.38 ± 0.255	
8	695000	725000	30	0.72 ± 0.015	0.38 ± 0.02	7.57 ± 0.446	
8	10810000	10845000	35	0.63 ± 0.03	0.38 ± 0.026	7.53 ± 0.578	*VCAM1, RTCA, DBT*
14	10450000	10480000	30	0.72 ± 0.07	0.34 ± 0.008	6.62 ± 0.178	
23 [c]	5020000	5050000	30	0.34 ± 0.027	0.42 ± 0.063	8.28 ± 1.402	*PTP4A2, KHDRBS1*
23 [c]	5225000	5310000	85	0.32 ± 0.13	0.35 ± 0.087	6.77 ± 1.956	*KIAA1522, YARS, FNDC5, HPCA, RNF19B, AK2, AZIN2, NDUFS5*
23 [c]	5330000	5365000	35	0.31 ± 0.07	0.34 ± 0.033	6.48 ± 0.744	*KIAA1522, YARS, FNDC5, HPCA, RNF19B, AK2, AZIN2, NDUFS5*
23 [c]	5510000	5555000	45	0.41 ± 0.086	0.44 ± 0.095	8.77 ± 2.128	*MFSD2A, CAP1, PPT1, RP11-49K24.9, PITHD1, LYPLA2, GALE, HMGCL, FUCA1, PNRC2*
23 [c]	5625000	5645000	20	0.3	0.33	6.41	*MYOM3, IL22RA1, IFNLR1, GRHL3*
23 [c]	5700000	5720000	20	0.37	0.38	7.56	*NIPAL3, RCAN3, CNR2, MIR1780*
27 [a]	160000	200000	40	0.65 ± 0.101	0.47 ± 0.084	9.54 ± 1.878	
27 [b]	5110000	5135000	25	0.67 ± 0.005	0.42 ± 0.041	8.45 ± 0.908	*IFI35, VAT1, RND2*

[a] These regions also show haplotypes with ambiguous relationships. [b] These regions also show introgression of domestic into grey junglefowl. [c] These regions also show introgression of domestic into Ceylon junglefowl. [d] Length in kilobase.

Table S3. Regions of ambiguity: Introgressed regions, which passed the genome-wide threshold (*Zfd* ≥ 6.0) for the Javan red junglefowl - domestic chicken ABBA/BABA introgression analysis but for which the haplotype tree suggests a likely shared ancestry or ambiguous relationships. Other regions showing ambiguity are found at **Table S2**

Chr	Start	Stop	Kb[b]	D (mean)	fd (mean)	Zfd (mean)	Overlapping Genes
2	83180000	83205000	25	0.43 ± 0.074	0.32 ± 0.002	6.04 ± 0.046	*FHOD3*
3	12160000	12185000	25	0.43 ± 0.03	0.33 ± 0.008	6.38 ± 0.181	*SPTLC3*
3 [a]	52495000	52520000	25	0.64 ± 0.011	0.38 ± 0.027	7.4 ± 0.613	
5	58800000	58820000	20	0.66	0.46	9.23	
6	11645000	11665000	20	0.35	0.35	6.86	
11	1310000	1330000	20	0.6	0.53	10.84	*SLC9A5*
17	5425000	5450000	25	0.48 ± 0.005	0.37 ± 0.026	7.16 ± 0.586	*RP11-101E3.5, SH3GLB2*
19	590000	620000	30	0.51 ± 0.014	0.41 ± 0.005	8.09 ± 0.107	*HSF5, gga-mir-1698-2*
20	2820000	2840000	20	0.3	0.48	9.61	

[a] This region also shows introgression of domestic into grey junglefowl. [b] Length in kilobase

Table S4. Population distribution and frequencies of Javan red junglefowl introgressed haplotypes into domestic chicken and red junglefowl. The frequencies were calculated by dividing the total number of introgressed haplotypes in any population with the total number of haplotype present in that population. Total number of haplotypes indicates below each population. **Jarso** (Ethiopia), **Horro** (Ethiopia), **SA:** Saudi Arabian chicken**, SL:** Sri Lankan chicken**, KH:** Kedu*Hitam*, **PdB:** Poule de Bresse**, MK:** Mechelse Koekoe**, MS:** Mechelse styrian, **Red:** Red junglefowl

Chr	**Donor species**	**Start**	**Stop**	**Jarso n = 10**	**Horro n = 12**	**SA n = 10**	**SL n = 22**	**KH N = 6**	**PdB n = 2**	**MK n = 2**	**MS n = 2**	**Red n = 12**
1	Javan	16520000	16540000	10	8.3	10	4.5	50	-	-	-	66.7
1	Javan	23745000	23775000	80	66.7	60	63.4	33.3	100	50	50	41.7
1	Javan	32250000	32275000	80	75	70	50	-	50	50	50	-
1	Javan	176660000	176680000	-	-	-	-	16.7	-	-	-	-
2	Javan	70260000	70295000	-	-	-	13.6	-	-	-	-	-
2	Javan	70675000	70720000	90	91.7	70	68.2	33.3	-	50	50	16.7
2	Javan	71755000	71790000	-	-	-		33.3	-	-	-	41.7
3	Javan	63290000	63310000	-	-	-	-	33.3				16.7
4	Javan	27440000	27470000	-	-	-	-	-	-	-	-	16.7
4	Javan	67000000	67030000	60	50	60	40.1	-	50	50	-	-
5	Javan	40880000	40915000	80	66.7	90	100	66.7	-	50	50	33.3
6	Javan	17915000	17945000	-	-	-	-	-	50	50	-	-
14	Javan	13655000	13770000	40	51.7	70	81.8	50	100	100	100	50
23	Javan	5415000	5485000	-	-	-	45.5	-	-	-	-	25
Number of candidate introgressed regions				7	7	7	9	8	5	7	5	9
Total number of introgressed haplotypes for all regions				44	35	43	103	19	7	8	6	37
% of introgressed haplotypes				62.8	41.7	61.4	52.0	39.6	70	57.1	60	34.3

Table S5. Population distribution and frequencies of grey junglefowl *Gallus sonneratii* introgressed haplotypes into domestic chicken and red junglefowl. The ABBA-BABA test was between domestic chicken and Javan red junglefowl. The frequencies were calculated by dividing the total number of introgressed haplotypes in any population with the total number of haplotype present in that population. Total number of haplotypes indicates below each population. **Jarso** (Ethiopia), **Horro** (Ethiopia), **SA:** Saudi Arabia chicken**, SL:** Sri Lanka chicken**, KH:** KeduHitam, **PdB:** Poule de Bresse**, MK:** Mechelse Koekoe**, MS:** Mechelse styrian, **Red:** Red junglefowl

Chr	**Donor species**	**Start**	**Stop**	**Jarso n = 10**	**Horro n = 12**	**SA n = 10**	**SL n = 22**	**KH N = 6**	**PdB n = 2**	**MK n = 2**	**MS n = 2**	**Red n = 12**
2	Grey junglefowl	80375000	80395000	10	-	-	-	-	-	-	-	-
21	Grey junglefowl	1485000	1505000	-	-	-	-	-	-	-	-	16.7
Number of candidate introgressed regions				1	-	-	-	-	-	-	-	1
Total number of introgressed haplotypes for all regions				1	-	-	-	-	-	-	-	2
% of introgressed haplotypes				10	-	-	-	-	-	-	-	16.7

Supplementary Figures

Chapter 2

Signature of positive selection in red junglefowl and domestic chicken

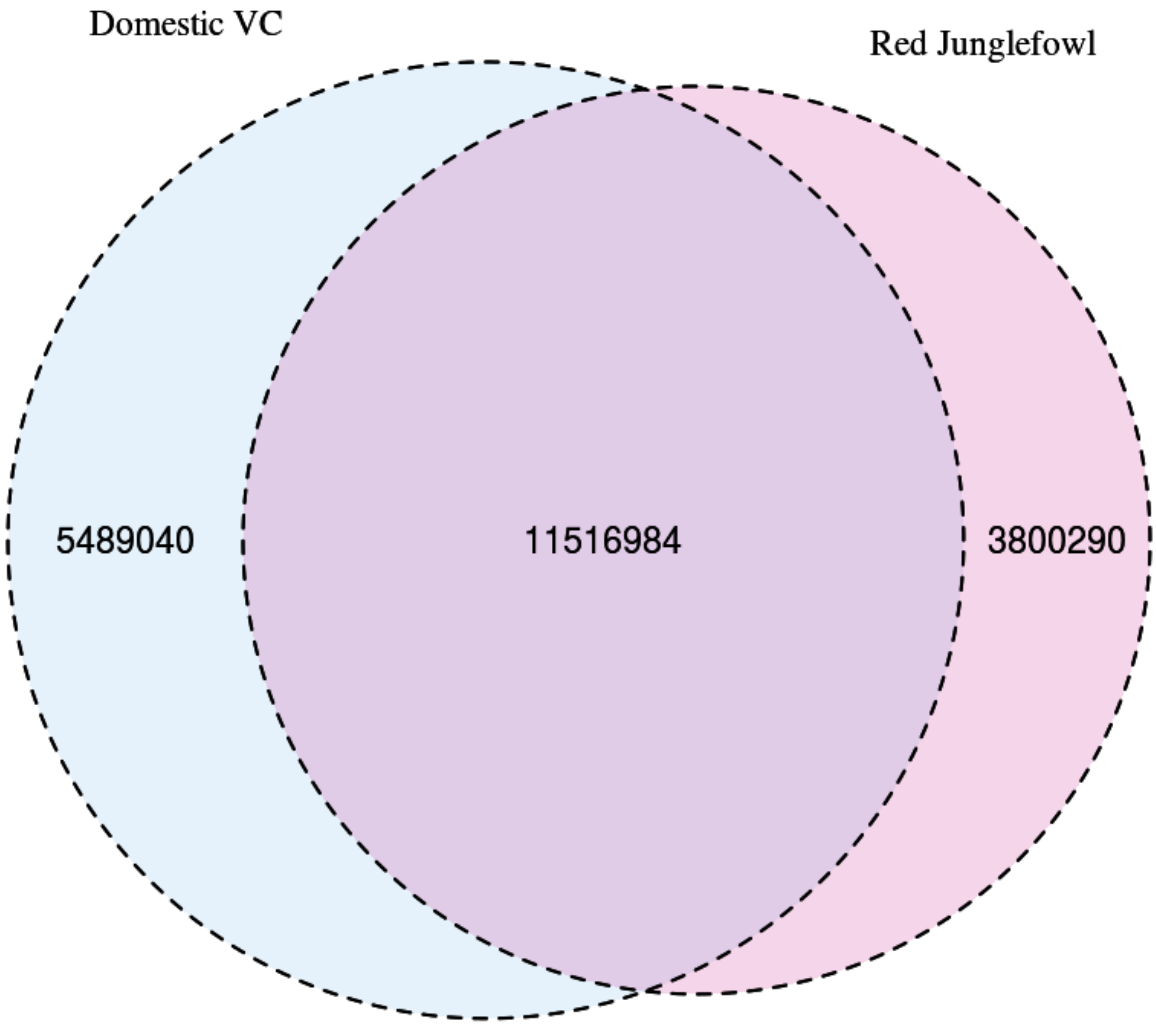


Figure S1. Venn diagram showing the shared and unique SNPs between 27 domestic chickens and six red junglefowls.

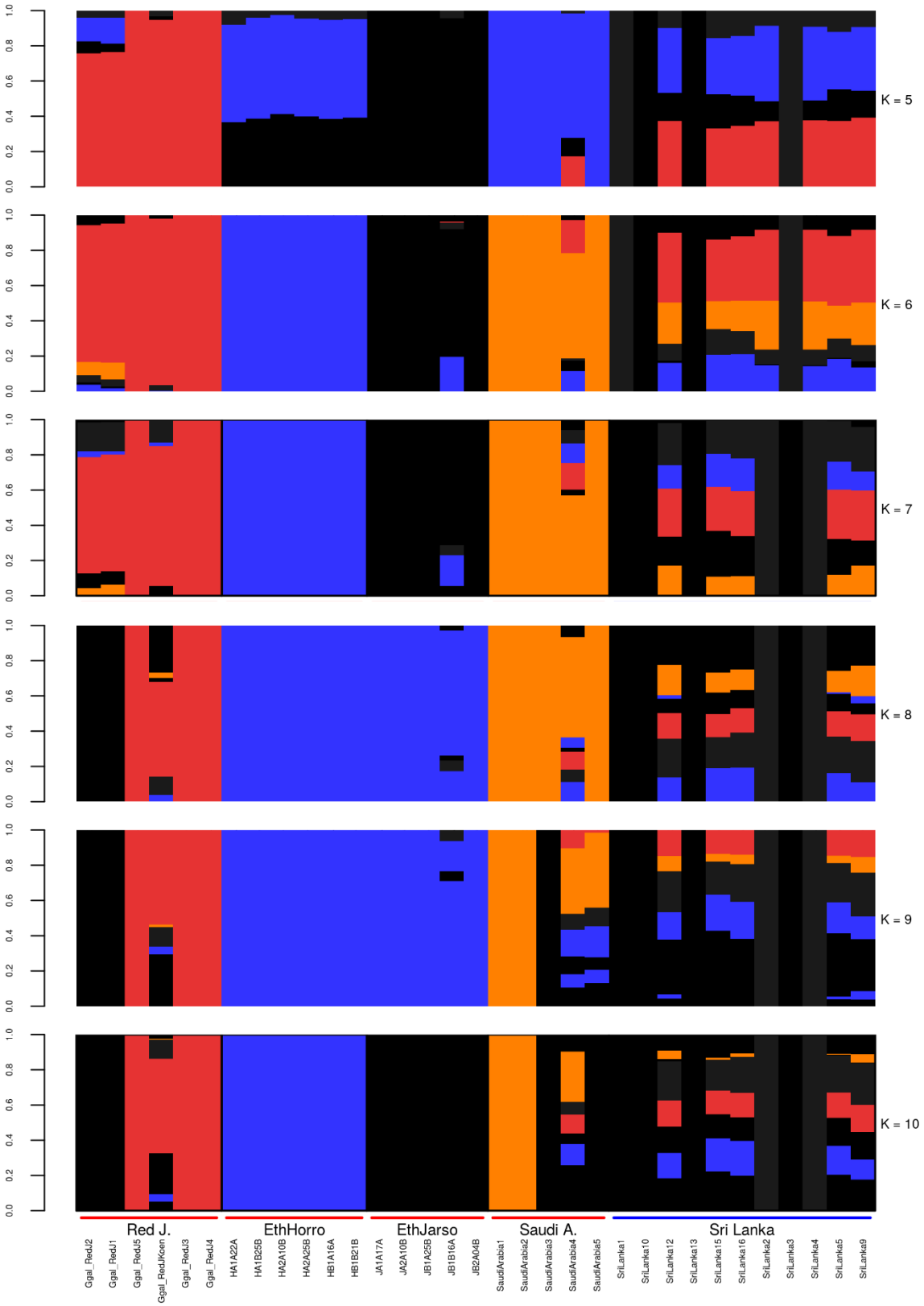
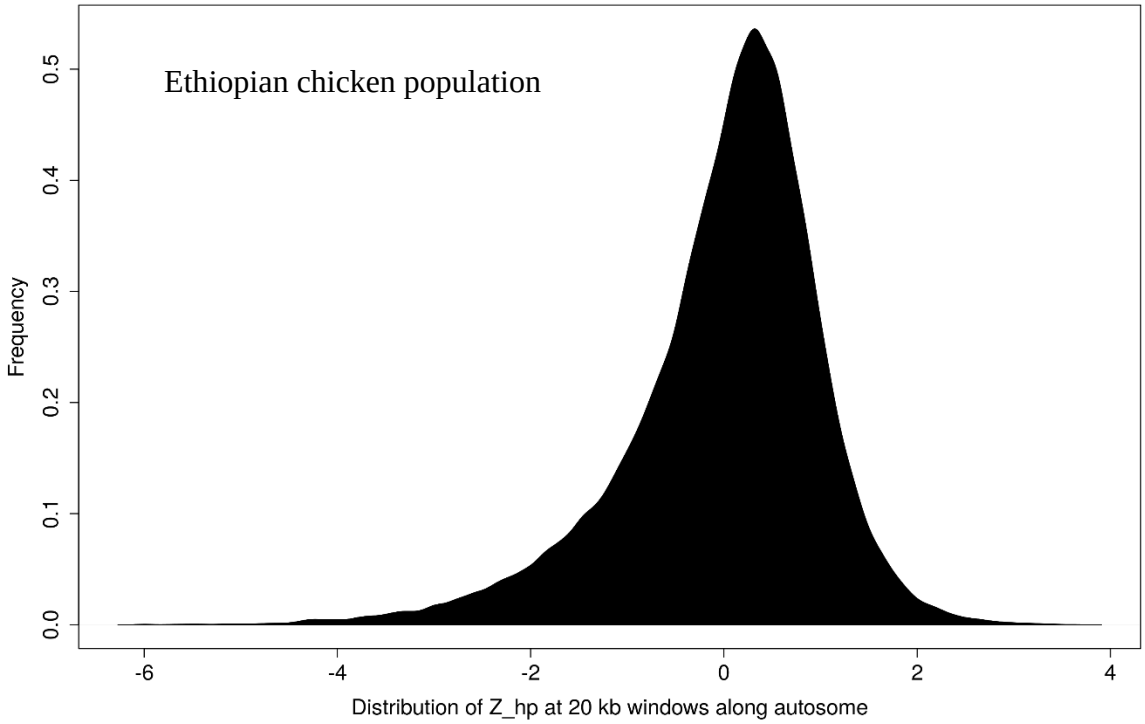


Figure S2. Autosomal admixture plots for the remaining of the 10 runs cluster. See the plot for 2, 3, and 4 clusters in **Figure 1**



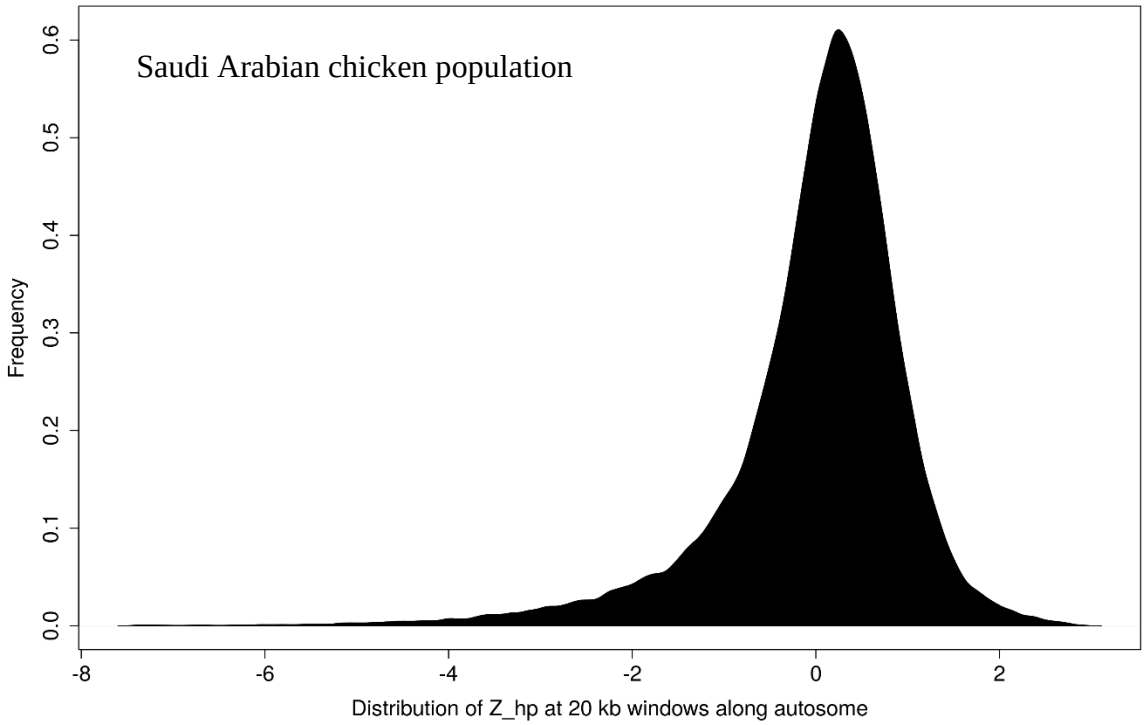
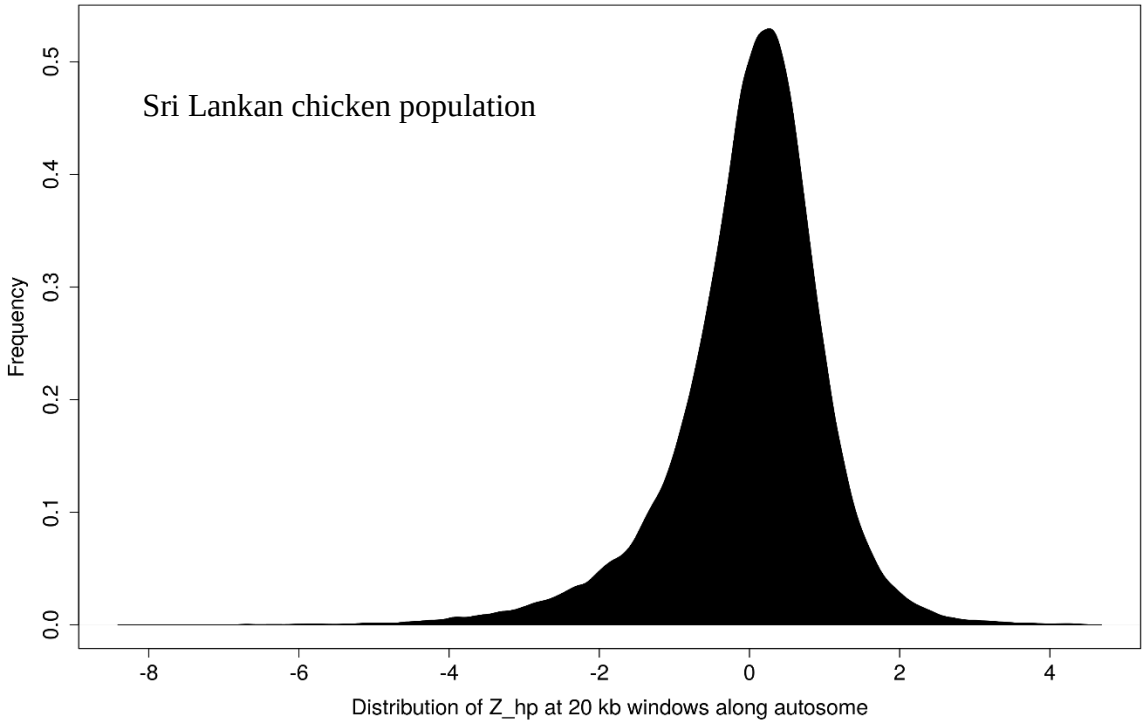


Figure S3A. The distribution of Z transformed pool heterozygosity (*ZH*p) values for the 20-kb windows at 10-kb step size along the autosomes.



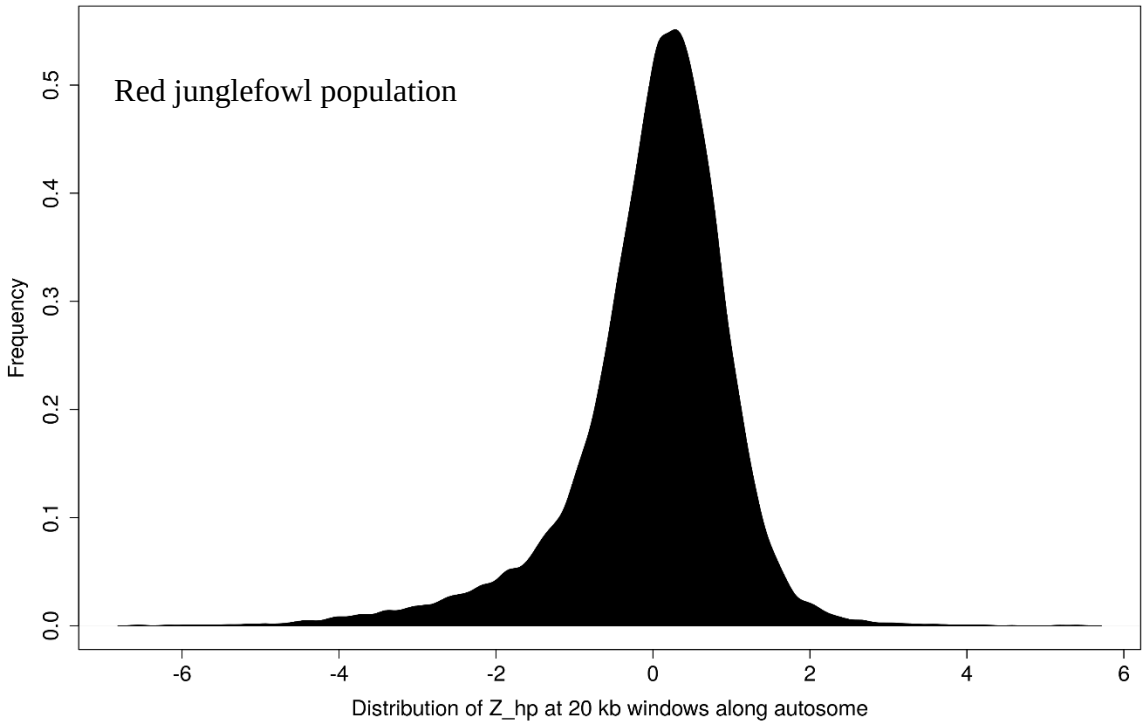


Figure S3B. The distribution of Z-transformed pool heterozygosity (*ZH*p) values for the 20-kb windows at 10-kb step size along the autosomes

Chapter 4

The autosomal genome ancestry of domestic chicken: Introgression from grey and Ceylon junglefowls

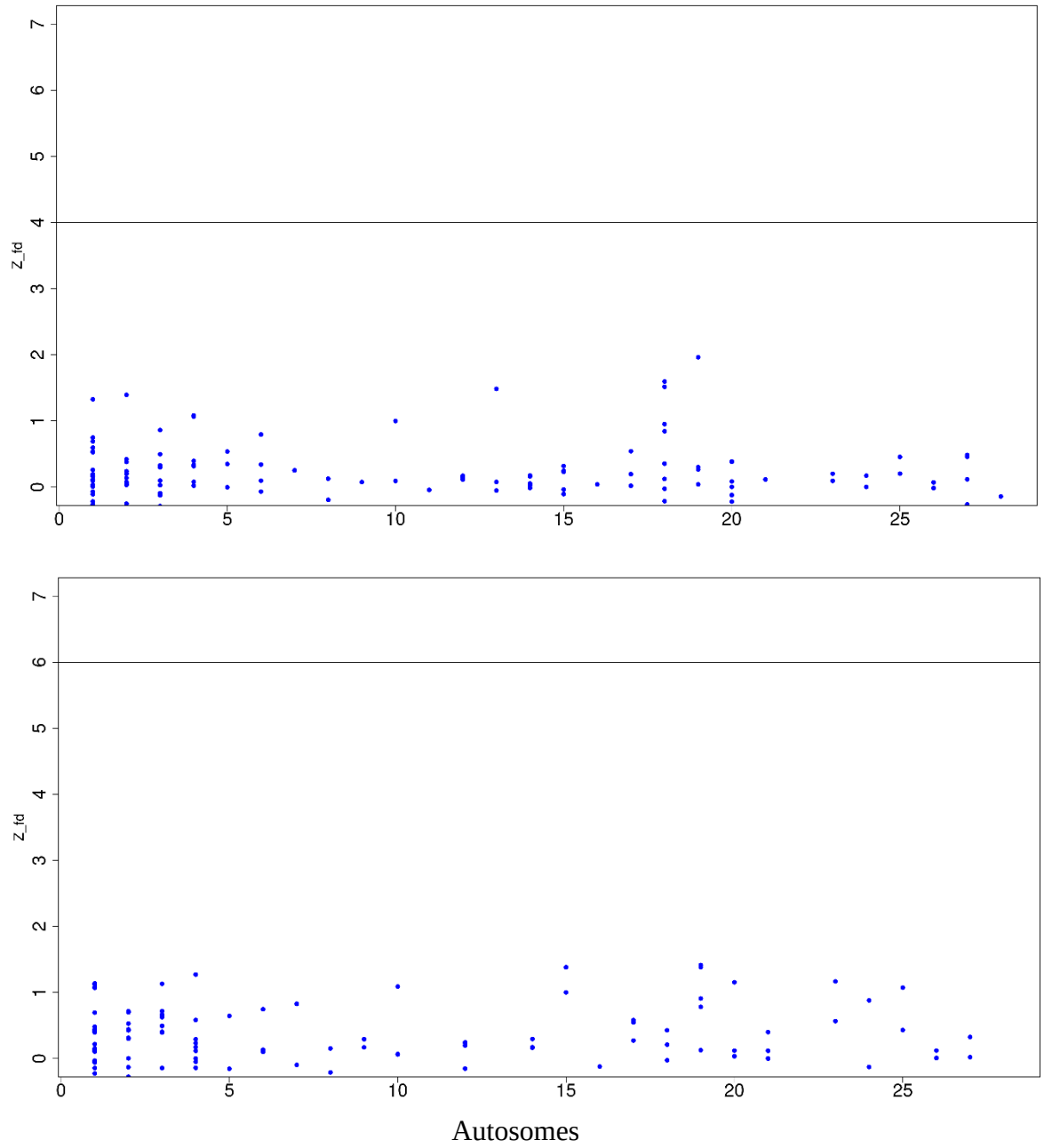


Figure S1. Chromosomal distribution of 20-kb windows with less than 100 SNPs. The Y-axis correspond to the values obtained from *Zfd* score (horizontal line threshold value of significance used in this study). X–axis *Galgal* 4.0 chromosomes 1 – 28. **A.** Grey junglefowl and domestic chicken ABBA/BABA plot (n = 195) **b.** Ceylon junglefowl and domestic chicken ABBA/BAB**A** = 185).

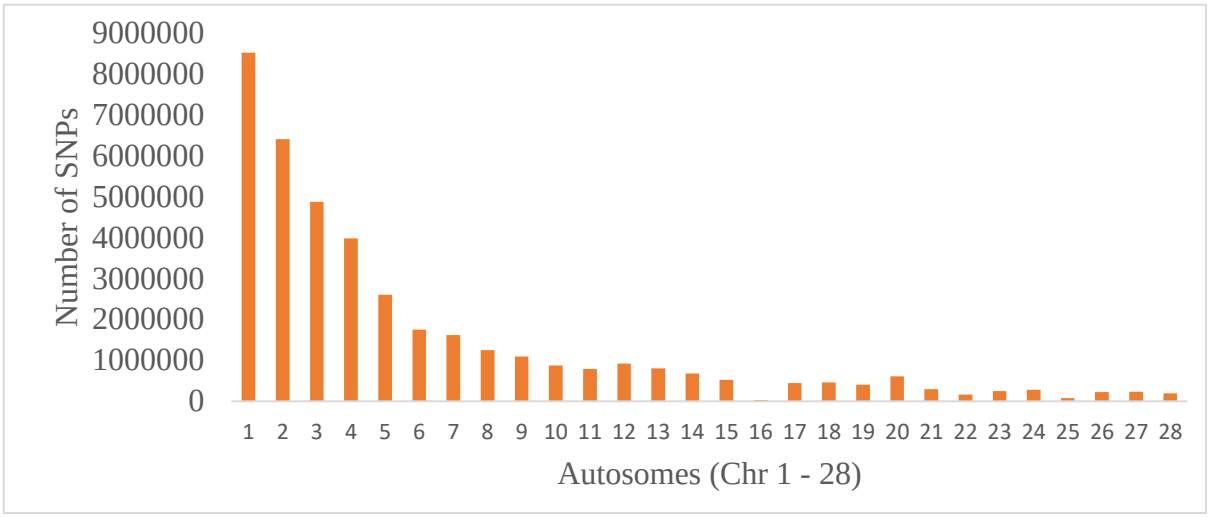


Figure S2. Histogram plot of the number of SNPs per chromosome for the entire dataset (50 genomes).

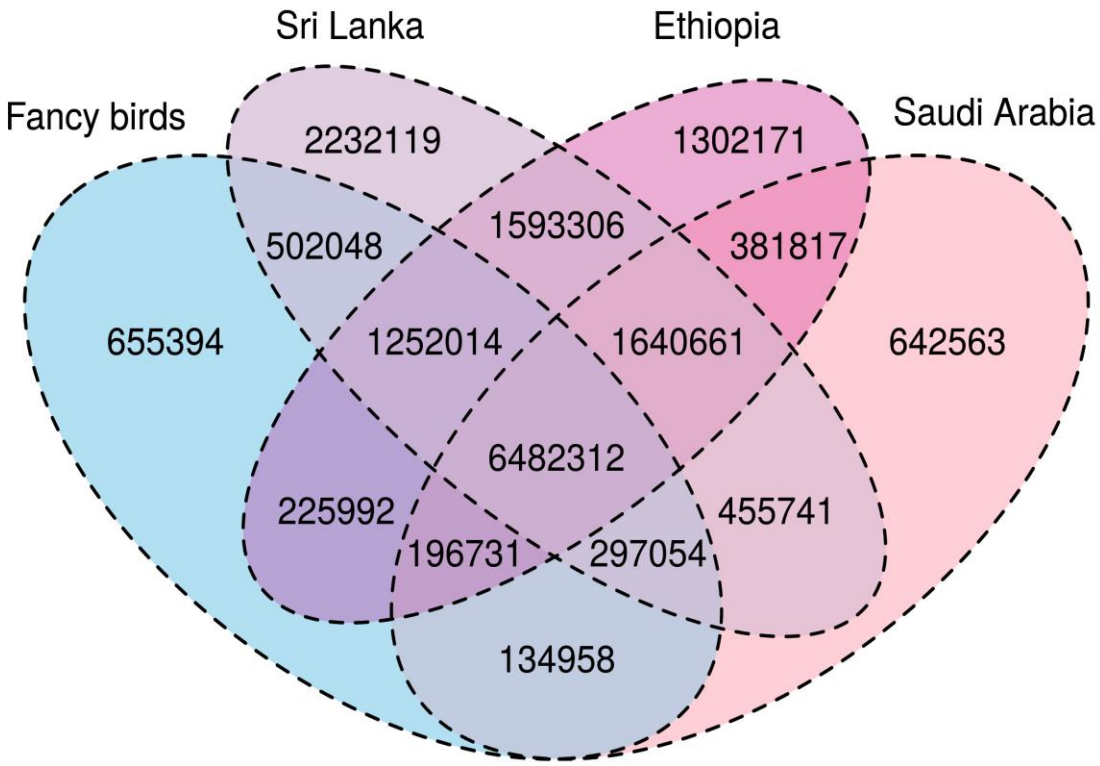


Figure S3.1. Venn diagram showing the shared and unique SNPs in each domestic population.

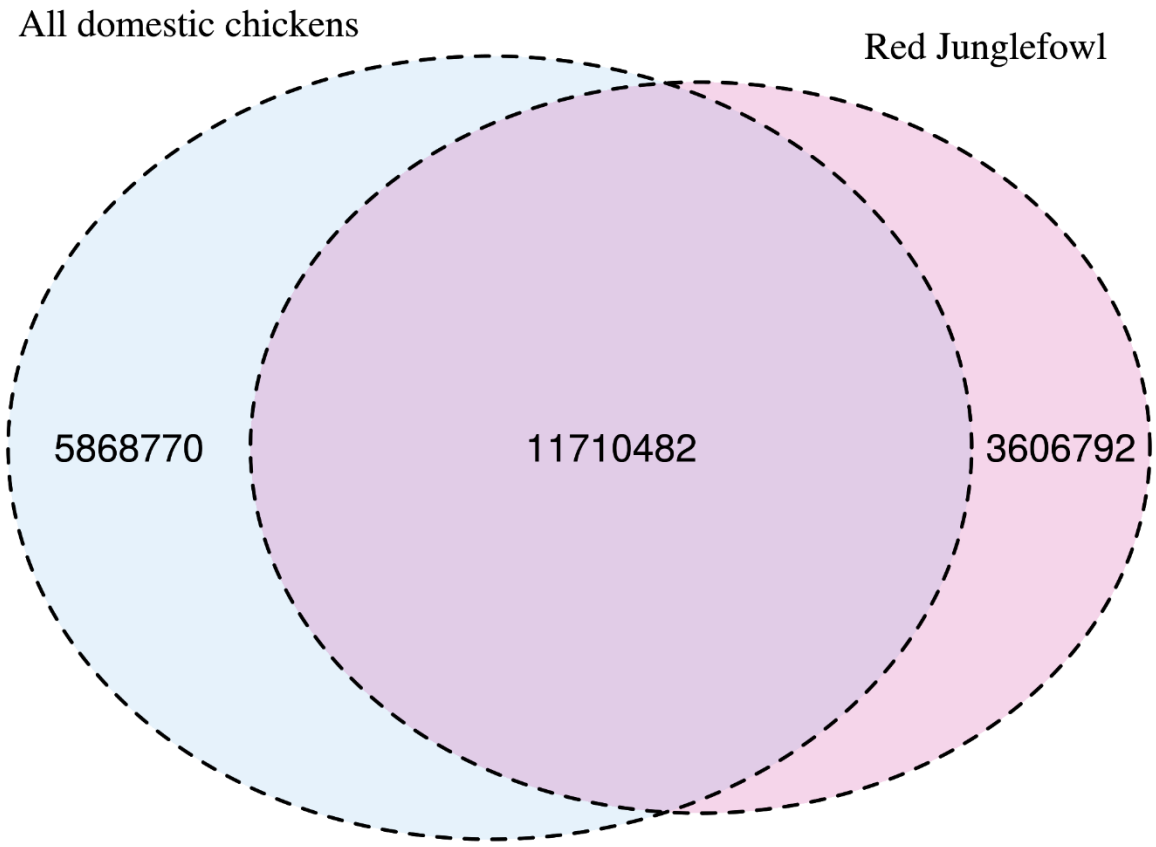


Figure S3.2. Venn diagram showing the number of shared and unique SNPs between domestic chicken and red junglefowl. The "All domestic chickens" category includes 30 birds (Ethiopia, Saudi Arabia, Sri Lanka and fancy).

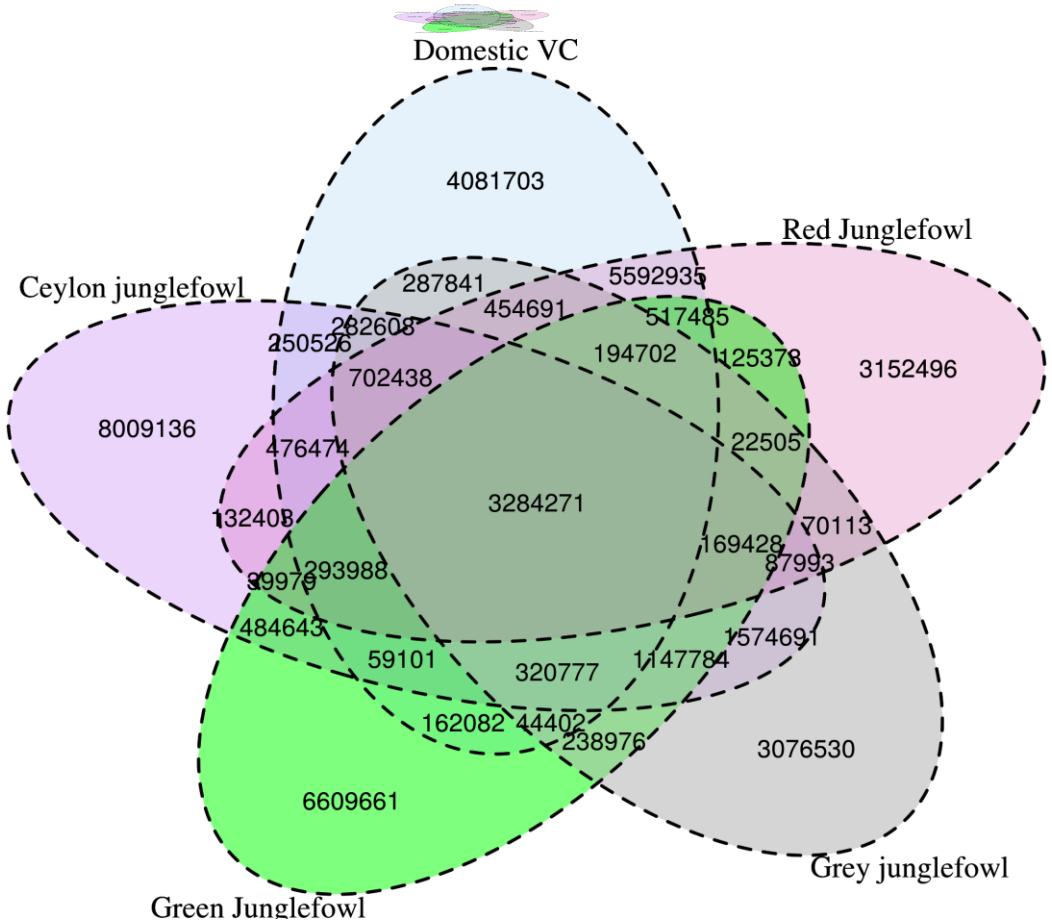


Figure S3.3. Venn diagram showing the number of shared and unique SNPs between domestic chicken and junglefowl populations. The "domestic" category includes 30 domestic chicken (Ethiopia, Saudi Arabia, Sri Lanka and fancy).

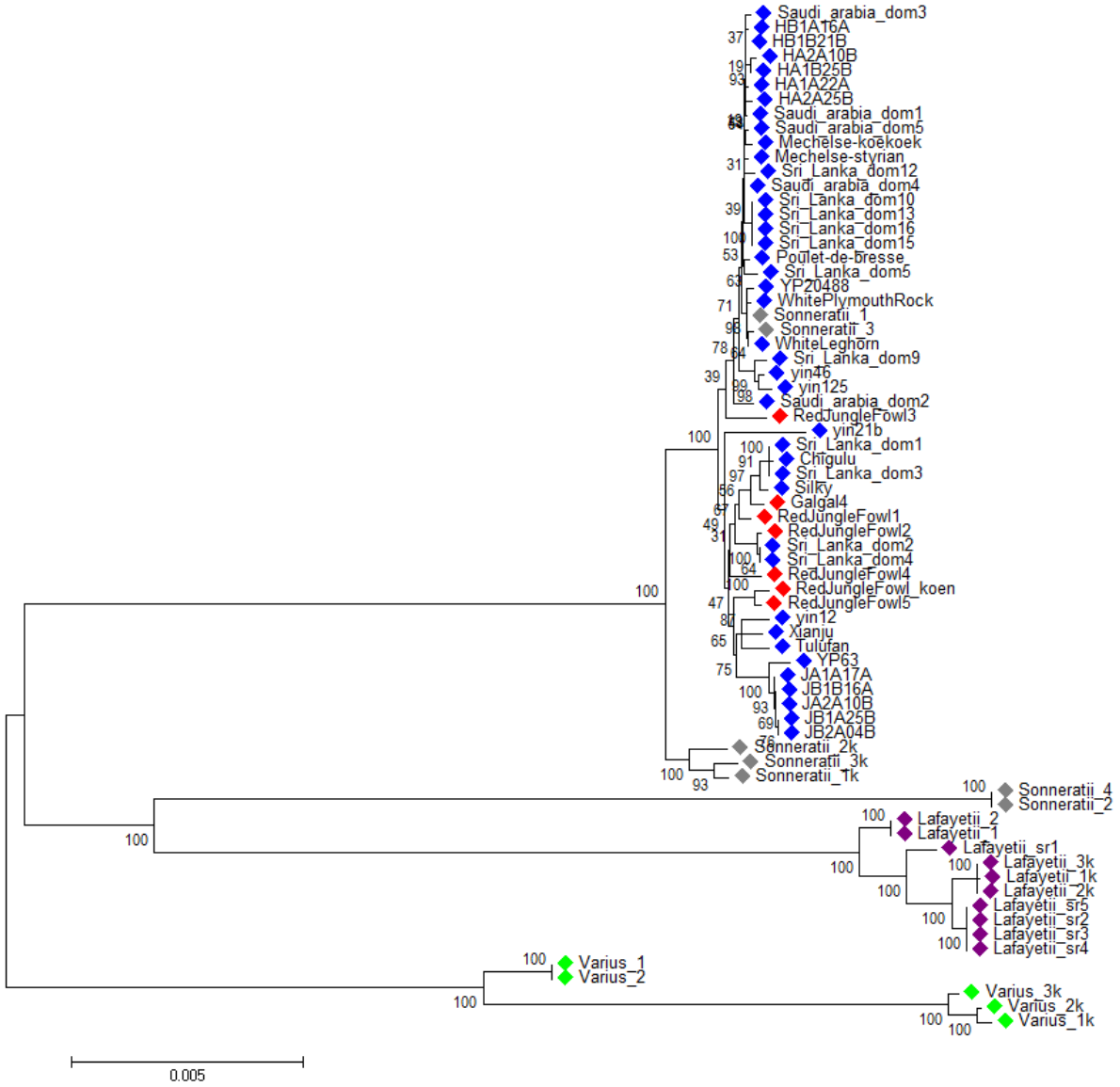


Figure S4. Full mitochondrial DNA neighbour-joining tree. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. Besides the 50 studied birds, the tree also includes the full mtDNA of 11 birds of reference (**Table S3**). Numbers above the nodes are bootstraps values. The tree was constructed based on the predicted GTR+I+G model by the Akaike information criterion.

Figure S5. Haplotypes relationships analysis for candidate regions of introgression from grey junglefowl into domestic chicken. ABBA – BABA test grey junglefowl - domestic chicken.

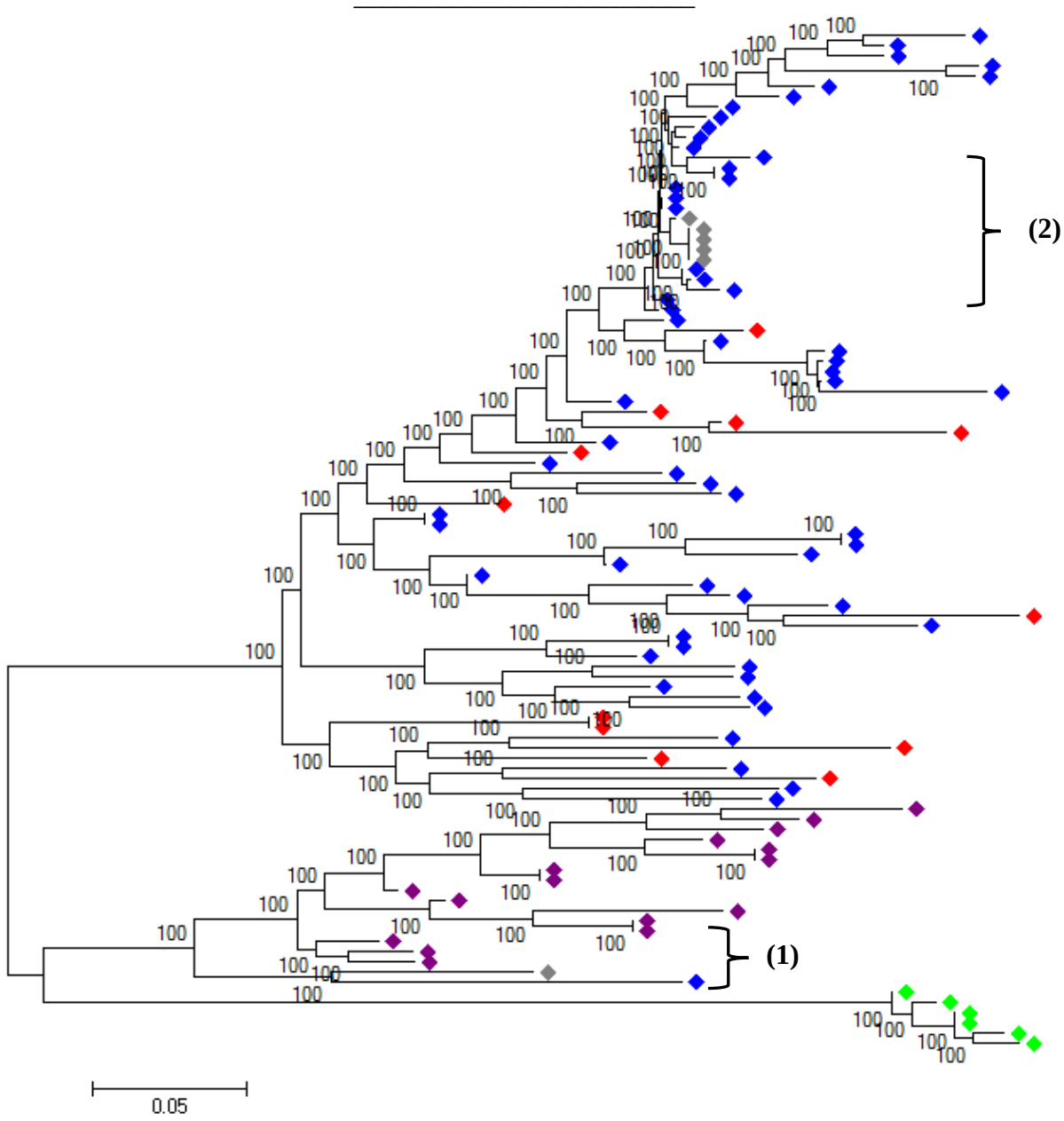


Figure S5.1. Haplotype based maximum likelihood tree for two patterns of introgression based on the 75-kb fragment size that extends from 15180001 to 15255001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Domestic haplotypes of grey junglefowl introgressed origin (Sri Lanka4 sample). (**2**). Grey junglefowl haplotypes of domestic - red junglefowl origin.

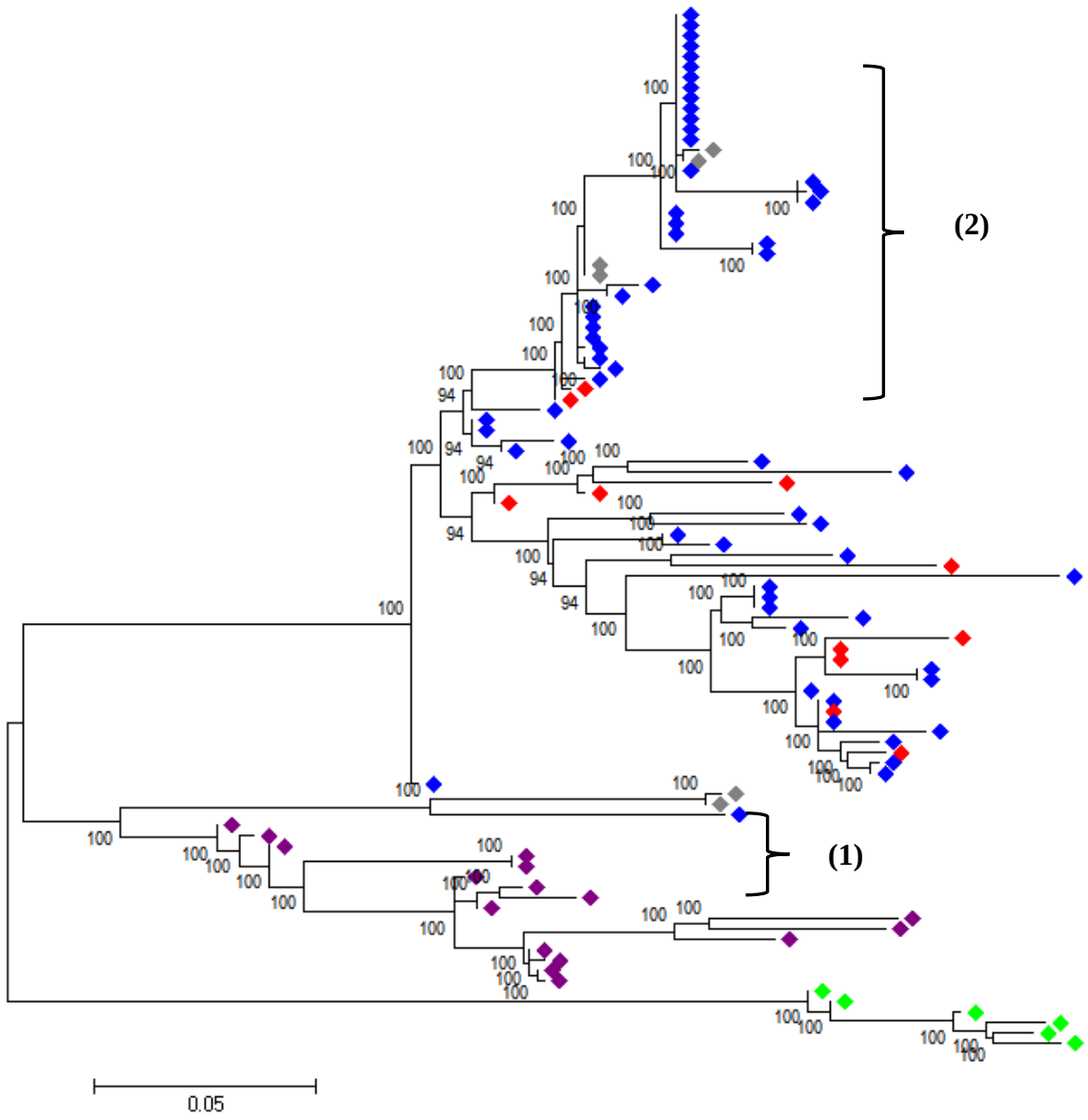


Figure S5.2. Haplotype based maximum likelihood tree for two patterns of introgression based on the 20-kb fragment size that extends from 18335001 to 18355001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Domestic haplotypes of grey junglefowl introgressed origin (Saudi Arabia1 sample). (**2**). Grey junglefowl haplotypes of domestic - red junglefowl origin.

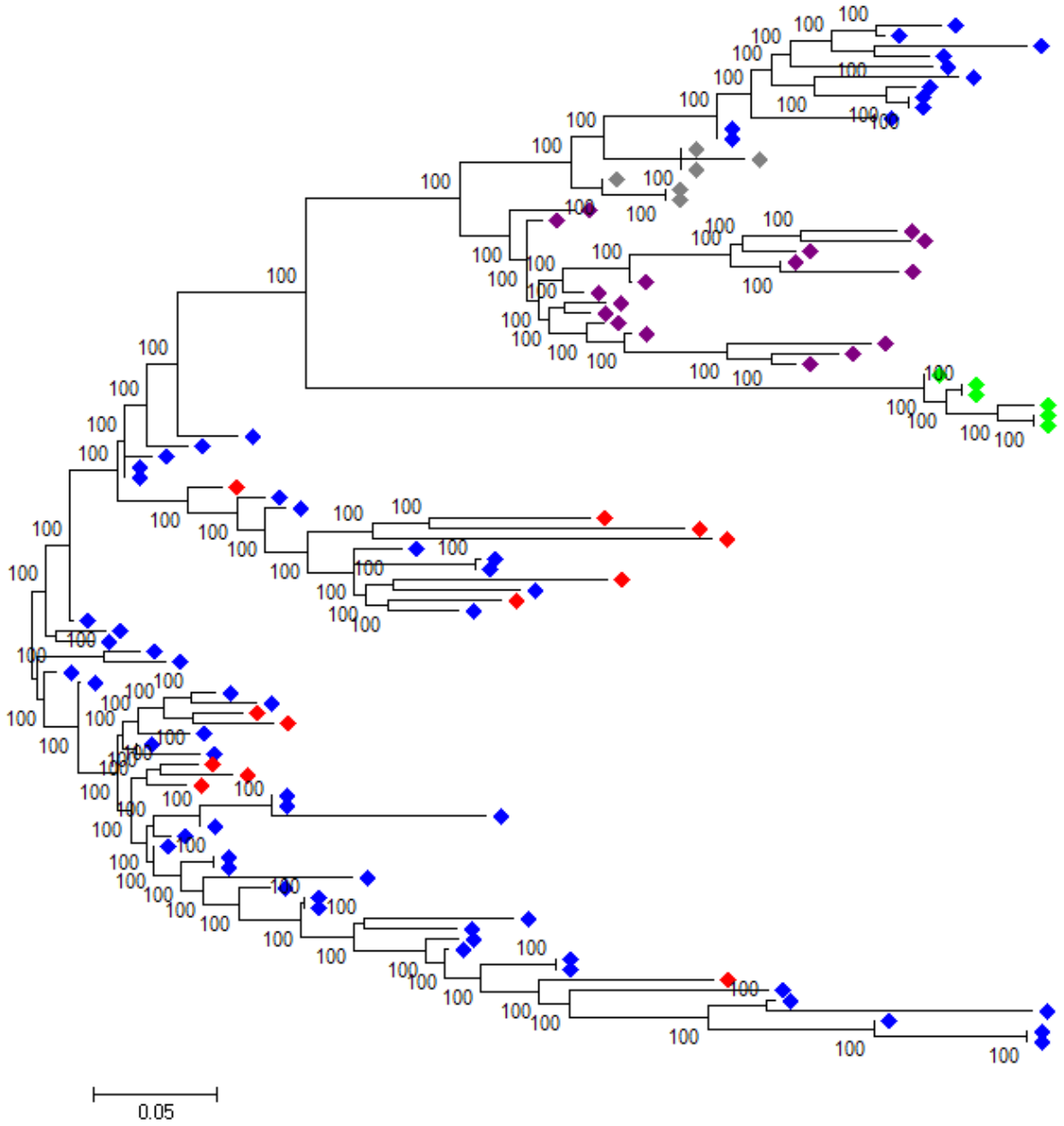


Figure S5.3. Haplotype based maximum likelihood tree for a region of introgression from grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 90920001 to 90940001 bp on chromosome 1.

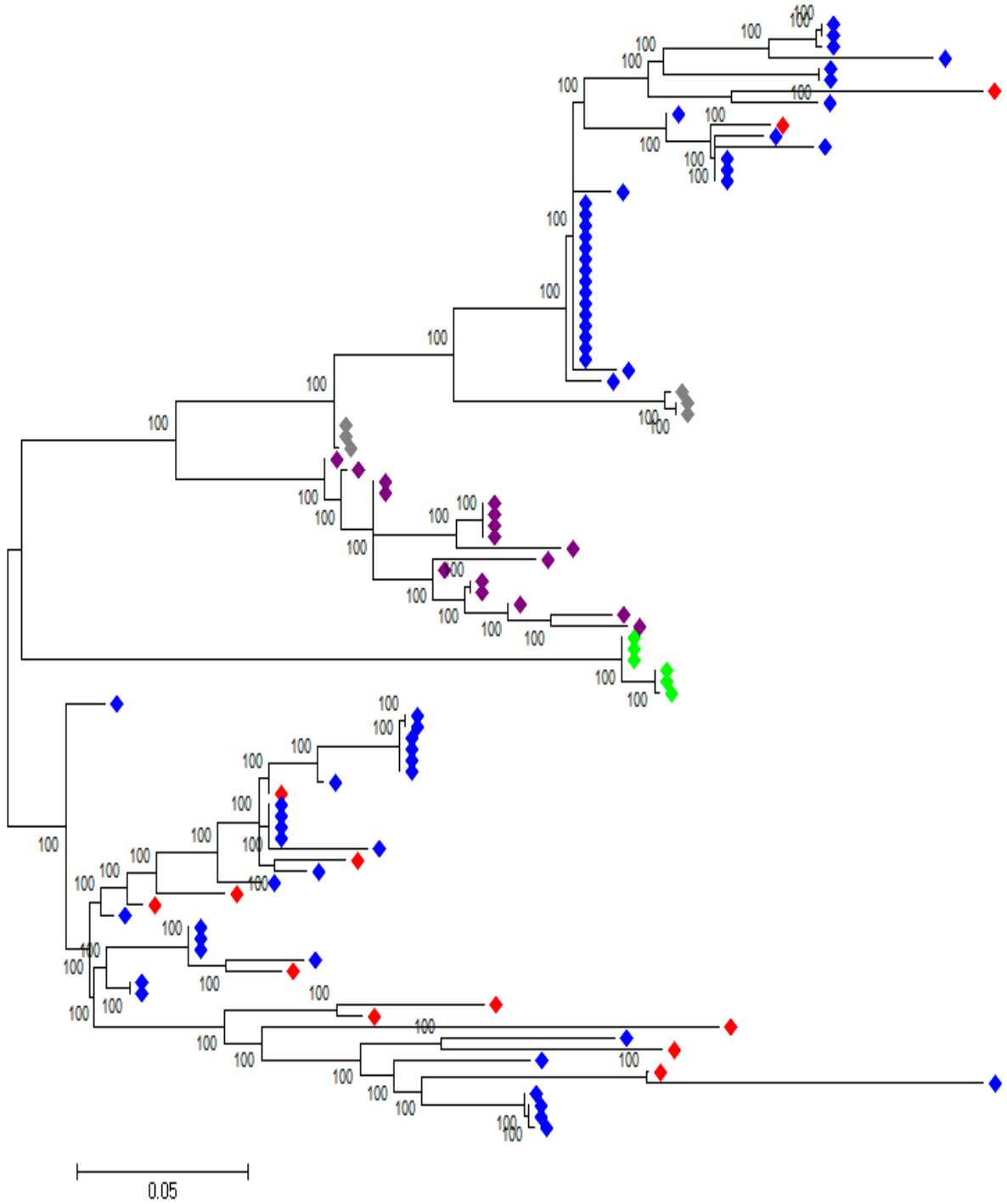


Figure S5.4. Haplotype based maximum likelihood tree for a region of introgression from grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 99915001 to 99935001 bp on chromosome 1.

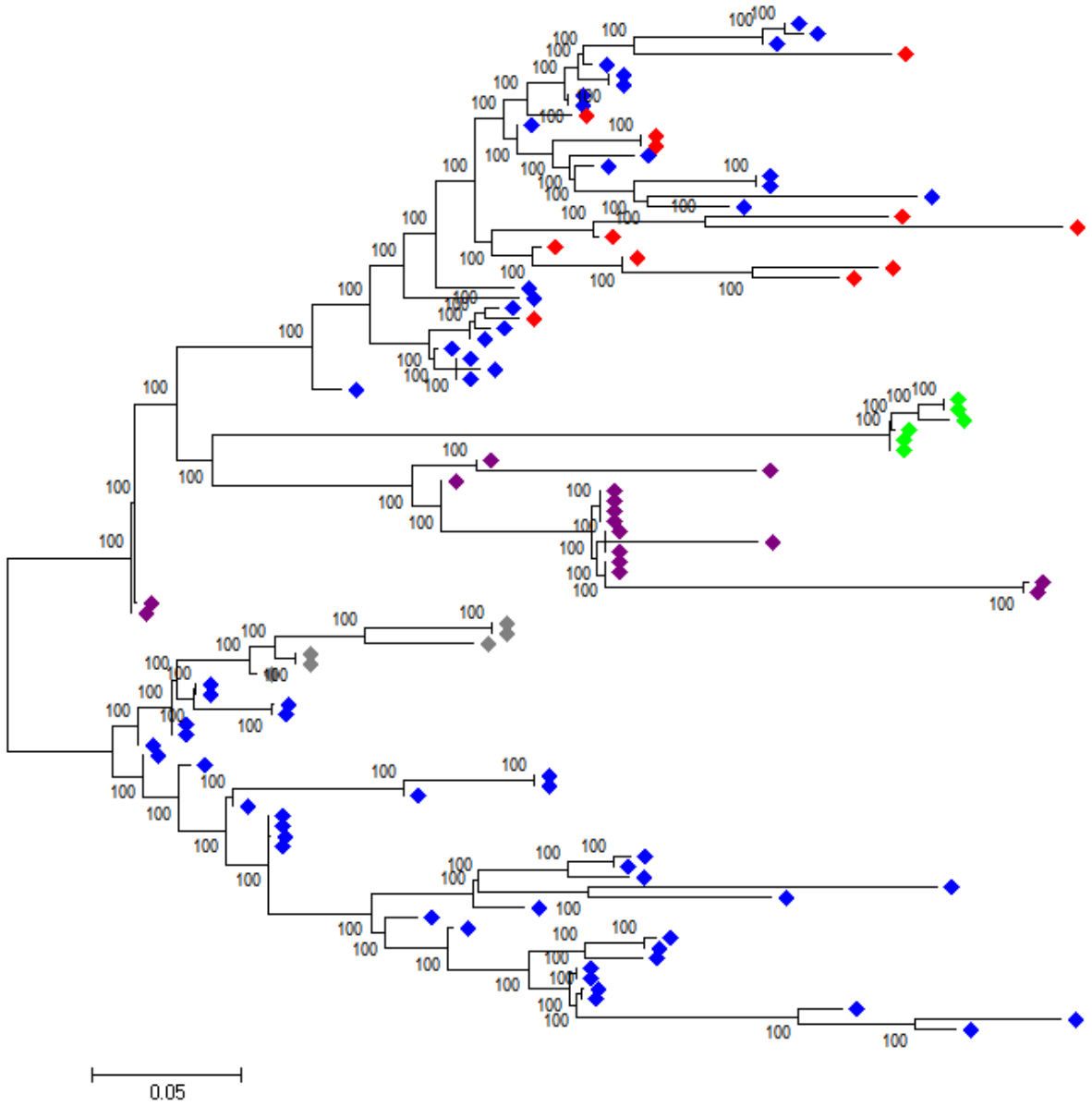


Figure S5.5. Haplotype based maximum likelihood tree for a region of introgression from grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 45-kb fragment size that extends from 100195001 to 100240001 bp on chromosome 1.

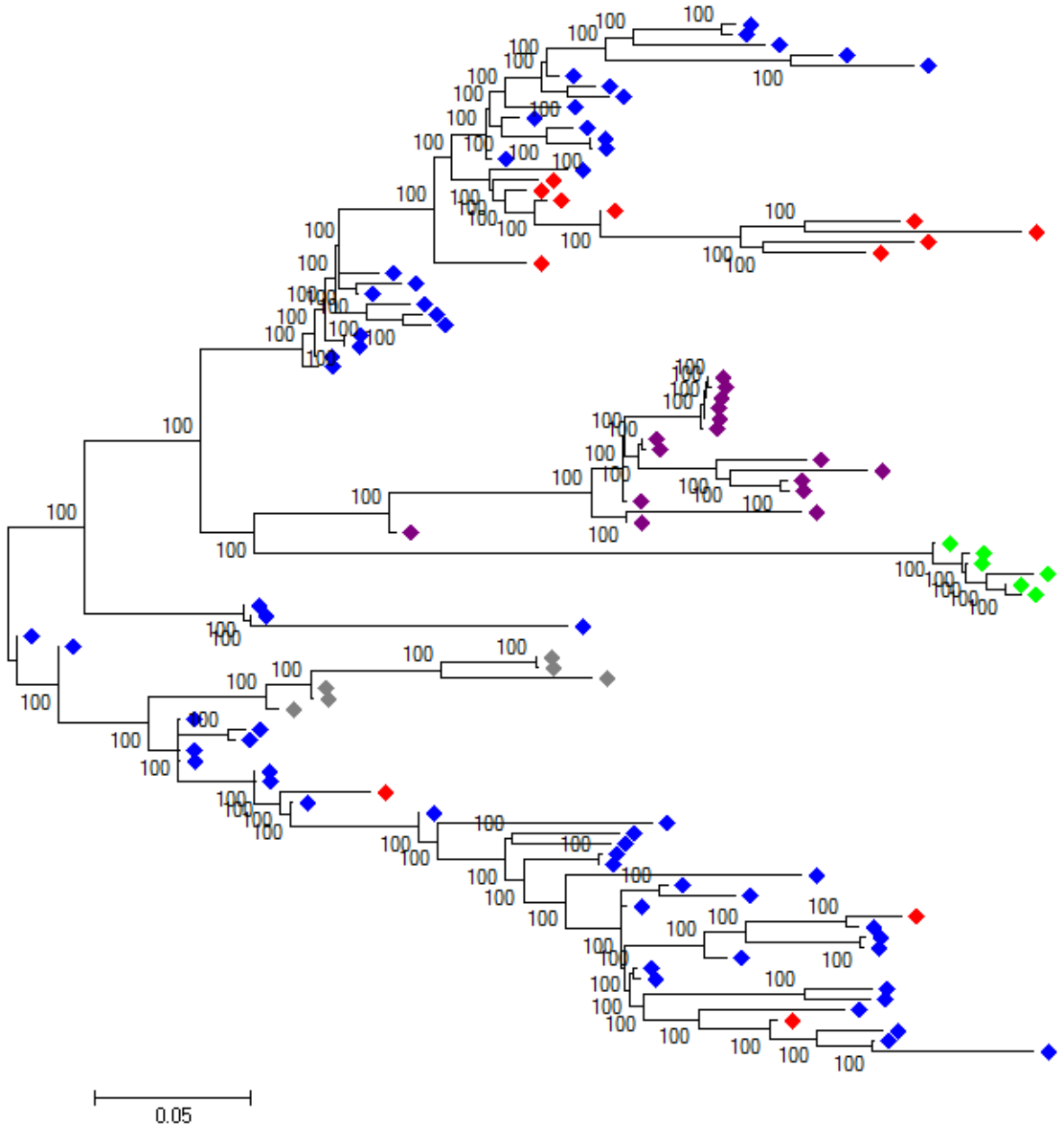


Figure S5.6. Haplotype based maximum likelihood tree for a region of introgression from grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 100-kb fragment size that extends from 100785001 to 100885001 bp on chromosome 1.

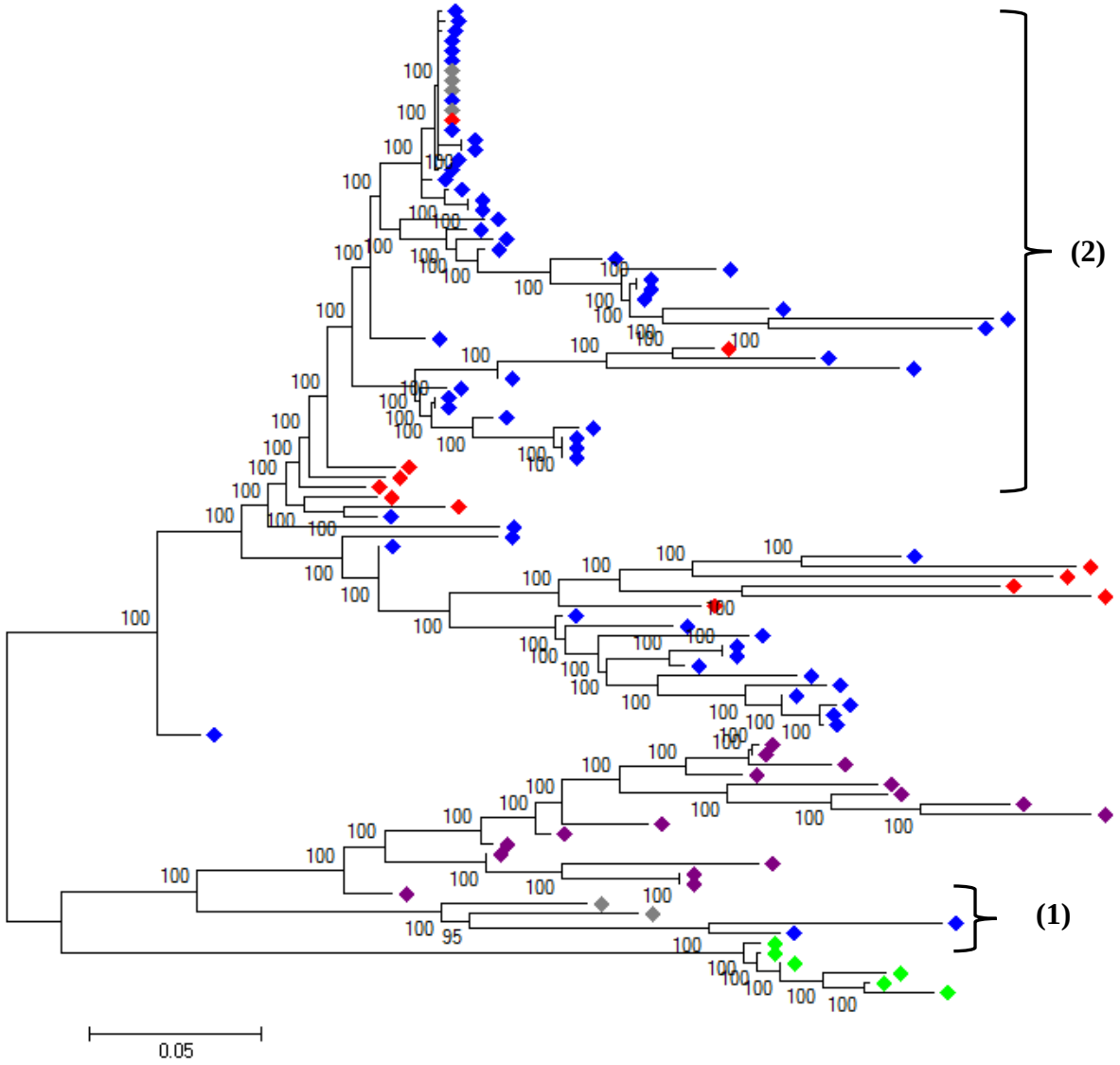


Figure S5.7. Haplotype based maximum likelihood tree for two patterns of introgression based on the 25-kb fragment size that extends from 139220001 to 139245001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. (**1**) Candidate region of introgression of grey junglefowl haplotype to one haplotype from Sri Lanka2 and Sri Lanka4. (**2**) Introgression of domestic chicken into grey junglefowl.

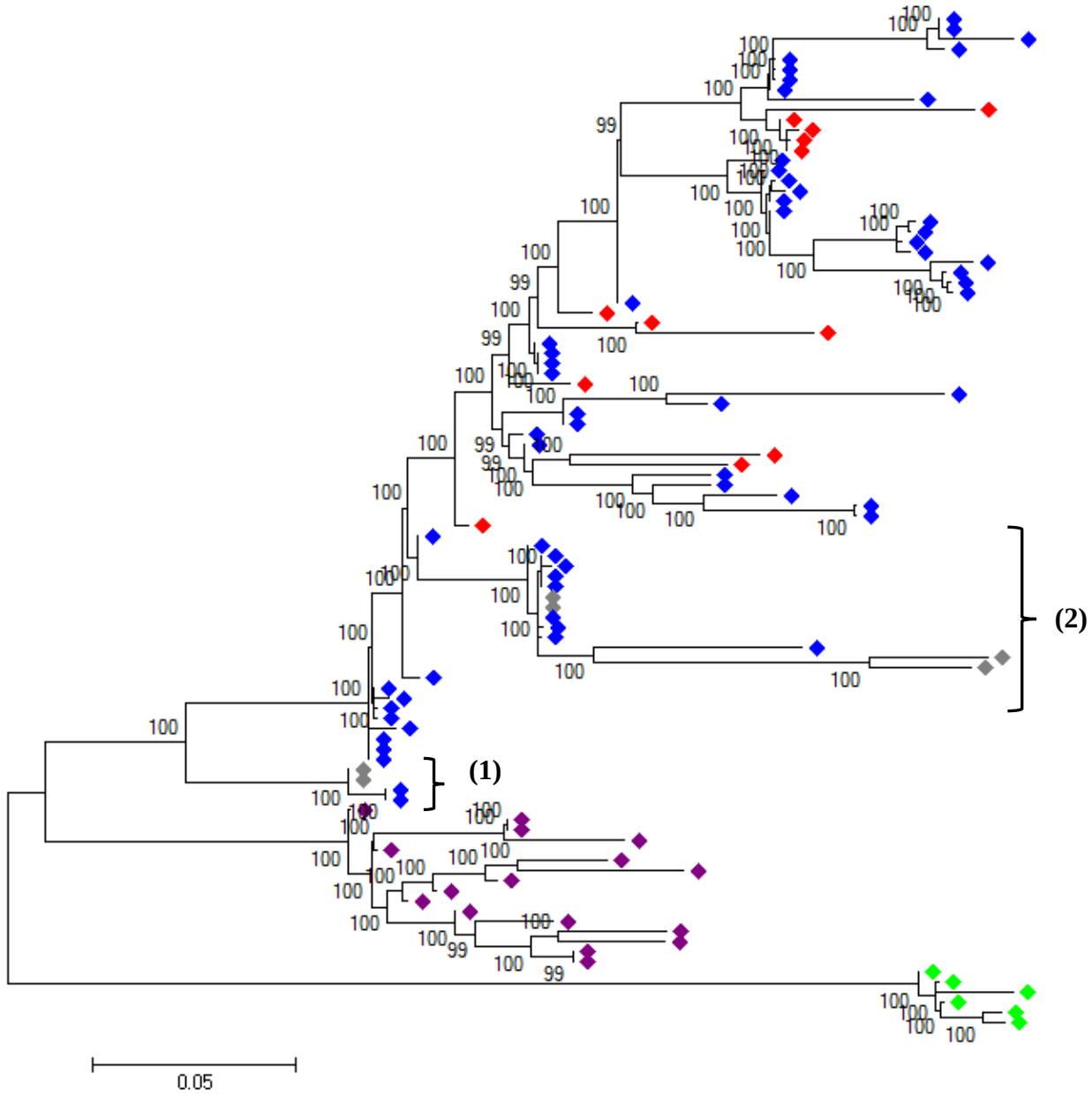


Figure S5.8. Haplotype based maximum likelihood tree for two patterns of introgression based on the 55-kb fragment size that extends from 141700001 to 141755001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. (**1)** Introgression of a grey junglefowl haplotypes to JB2A0B and Sri Lanka4 samples. (**2)** Introgression of a domestic chicken haplotype into grey junglefowl.

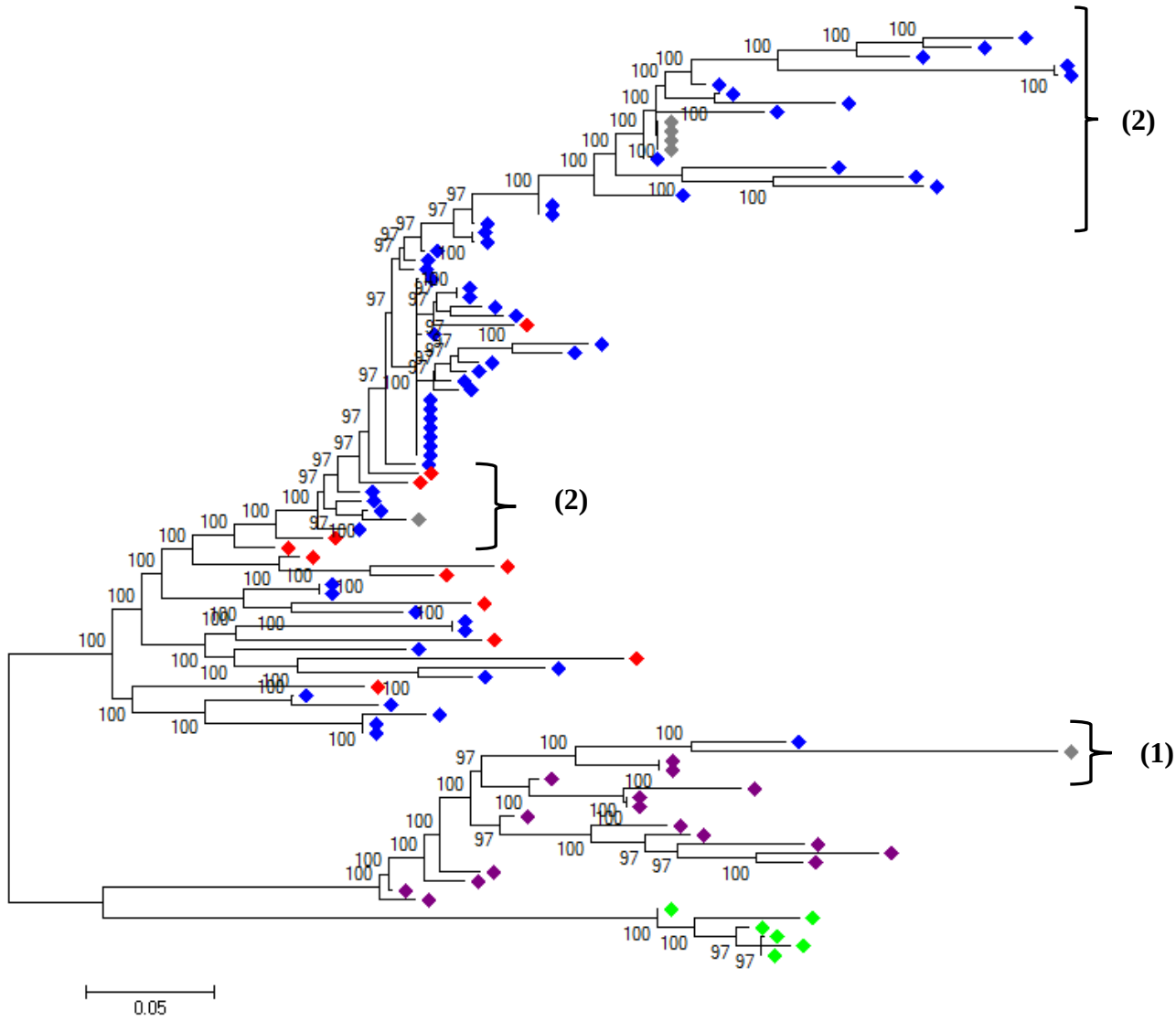


Figure S5.9. Haplotype based maximum likelihood tree for two patterns of introgression based on the 30-kb fragment size that extends from 145600001 to 145630001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl**. (1)** Introgression of grey junglefowl haplotype in Sri Lanka16 sample. (**2)** Introgression of a domestic chicken haplotype into grey junglefowl.

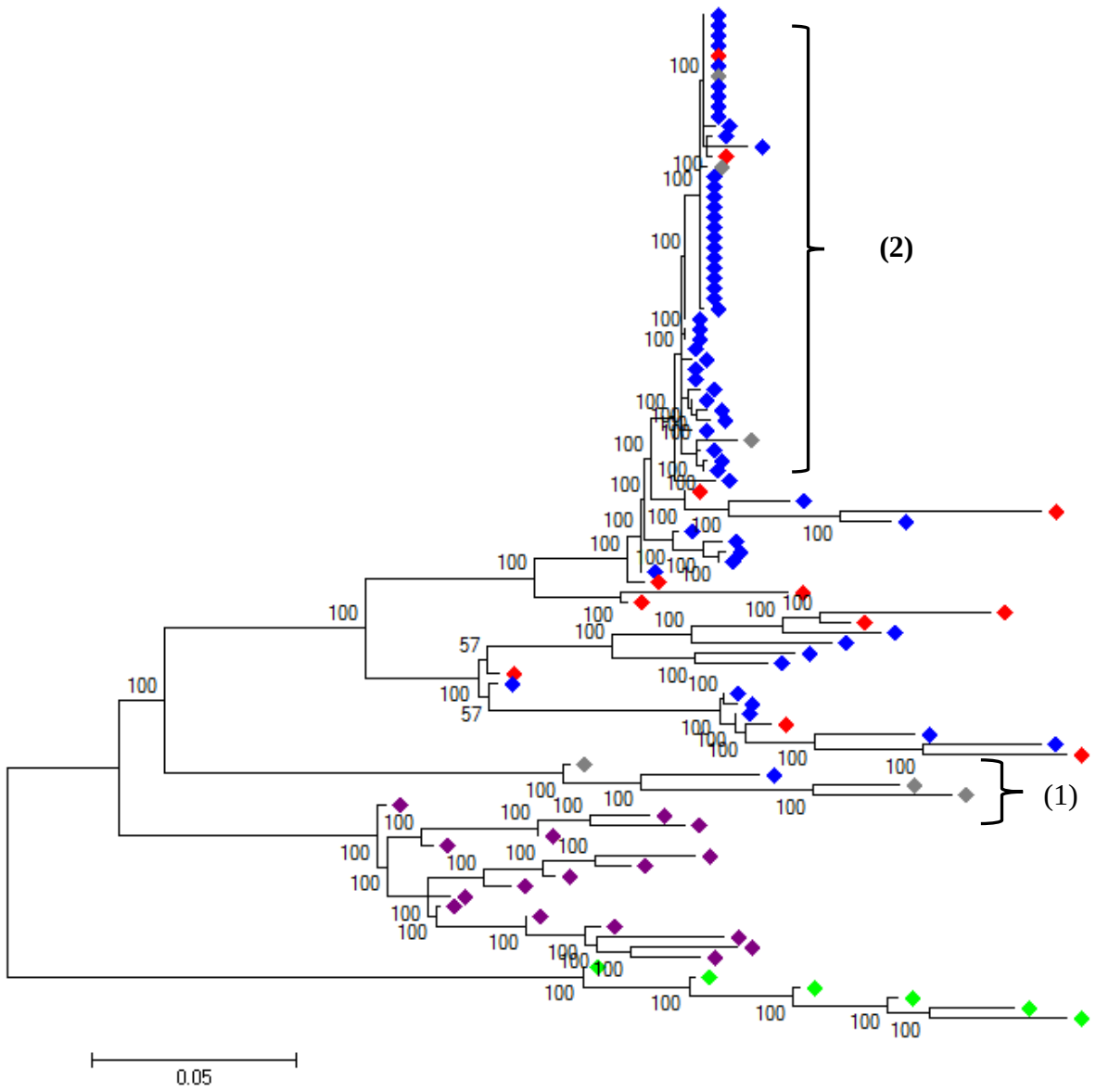


Figure S5.10. Haplotype based maximum likelihood tree for two patterns of introgression based on the 50-kb fragment size that extends from 147760001 to 147810001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl **(1)** Grey junglefowl haplotype introgressed to Sri Lanka1 sample. (**2)** Introgression of domestic chicken haplotypes into grey junglefowl.

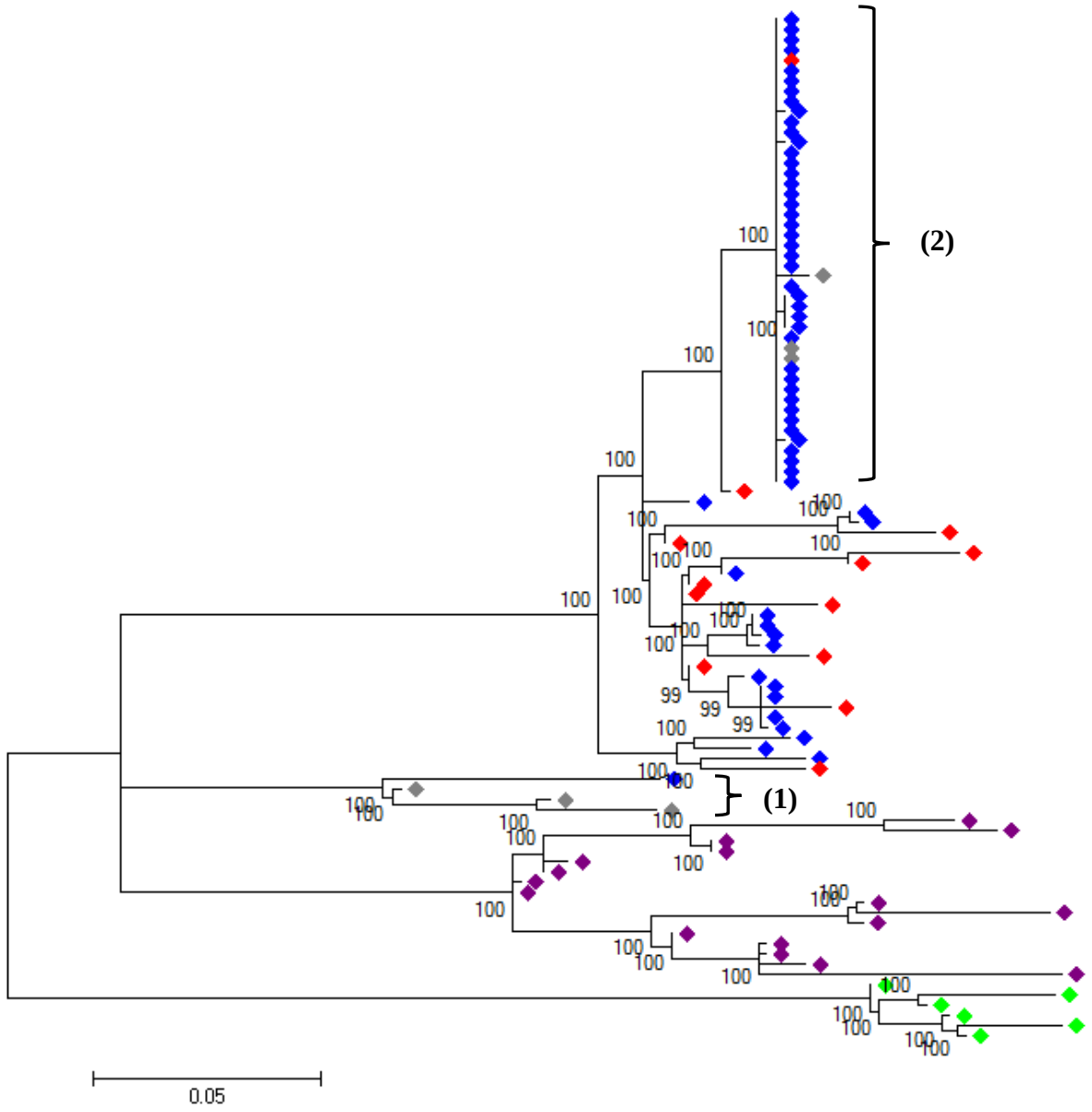


Figure S5.11. Haplotype based maximum likelihood tree for two patterns of introgression based on the 20-kb fragment size that extends from 147640001 to 147660001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of a grey junglefowl haplotypes in a Sri Lanka1 sample. **(2)** Introgression of domestic chicken into grey junglefowl.

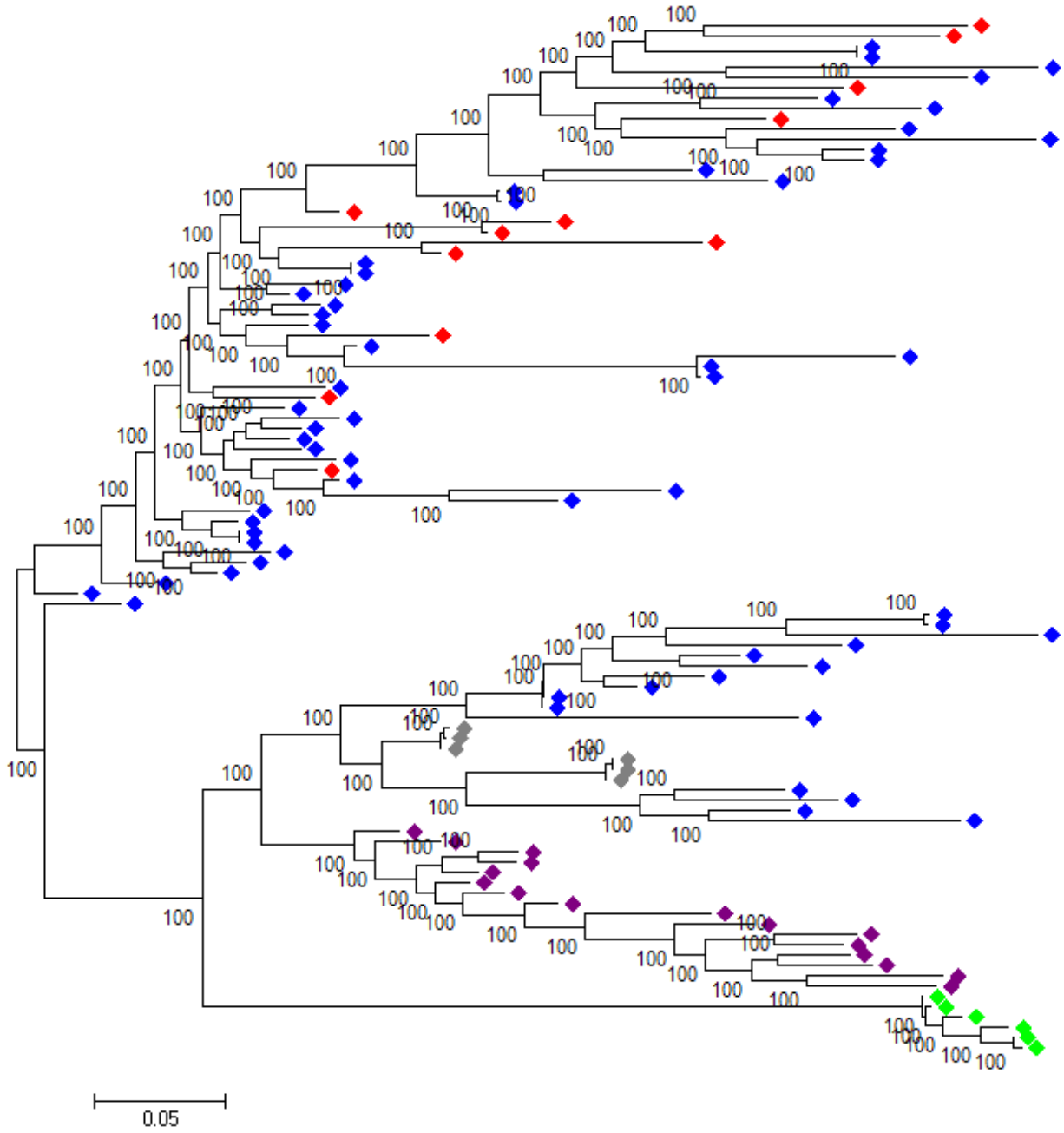


Figure S5.12. Haplotype based maximum likelihood tree for region of introgression from grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 55-kb fragment size that extends from 185245001 to 185300001 bp on chromosome 1.

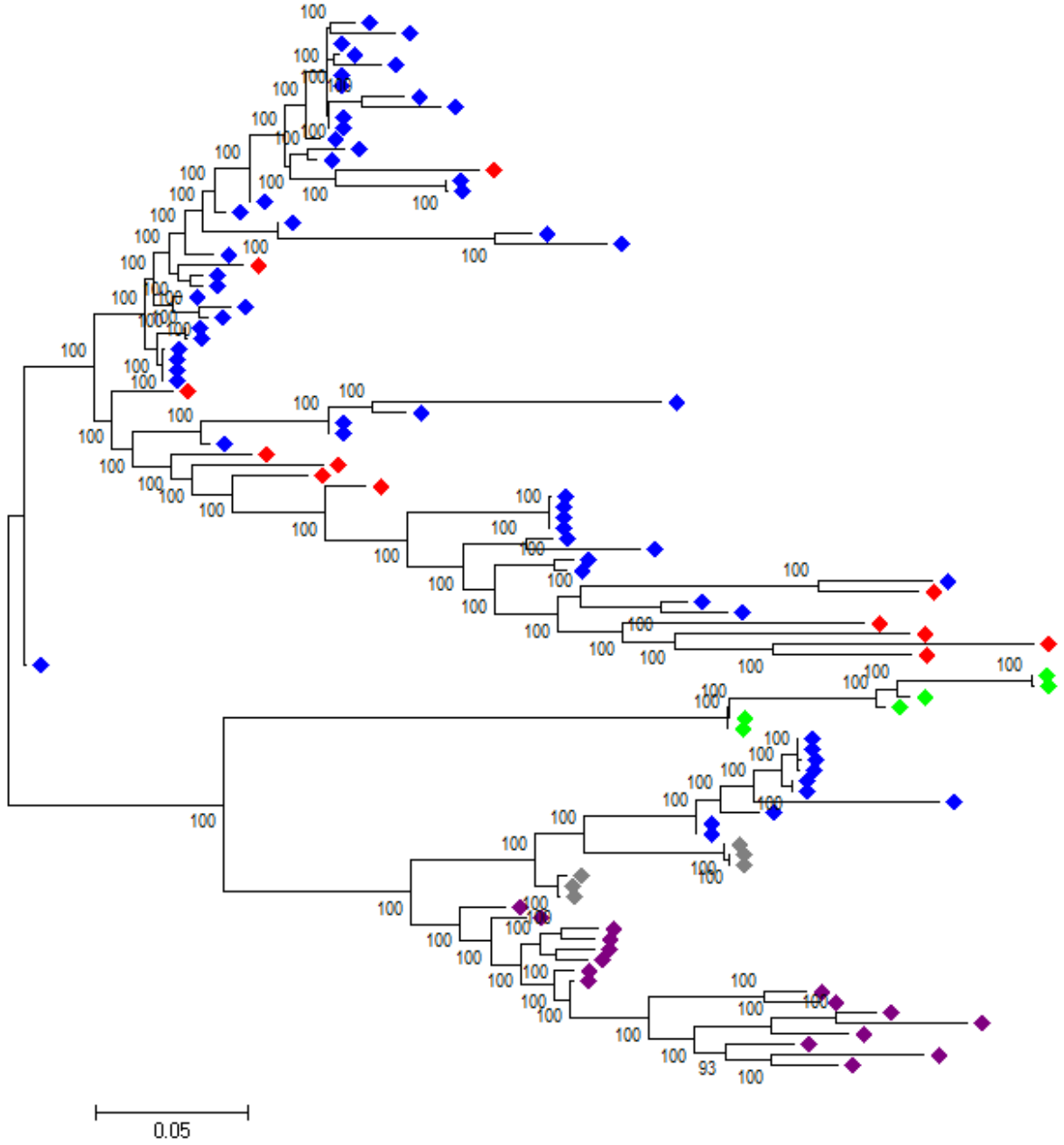


Figure S5.13. Haplotype based maximum likelihood tree for a region of introgression from grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 50-kb fragment size that extends from 185435001 to 185485001 bp on chromosome 1.

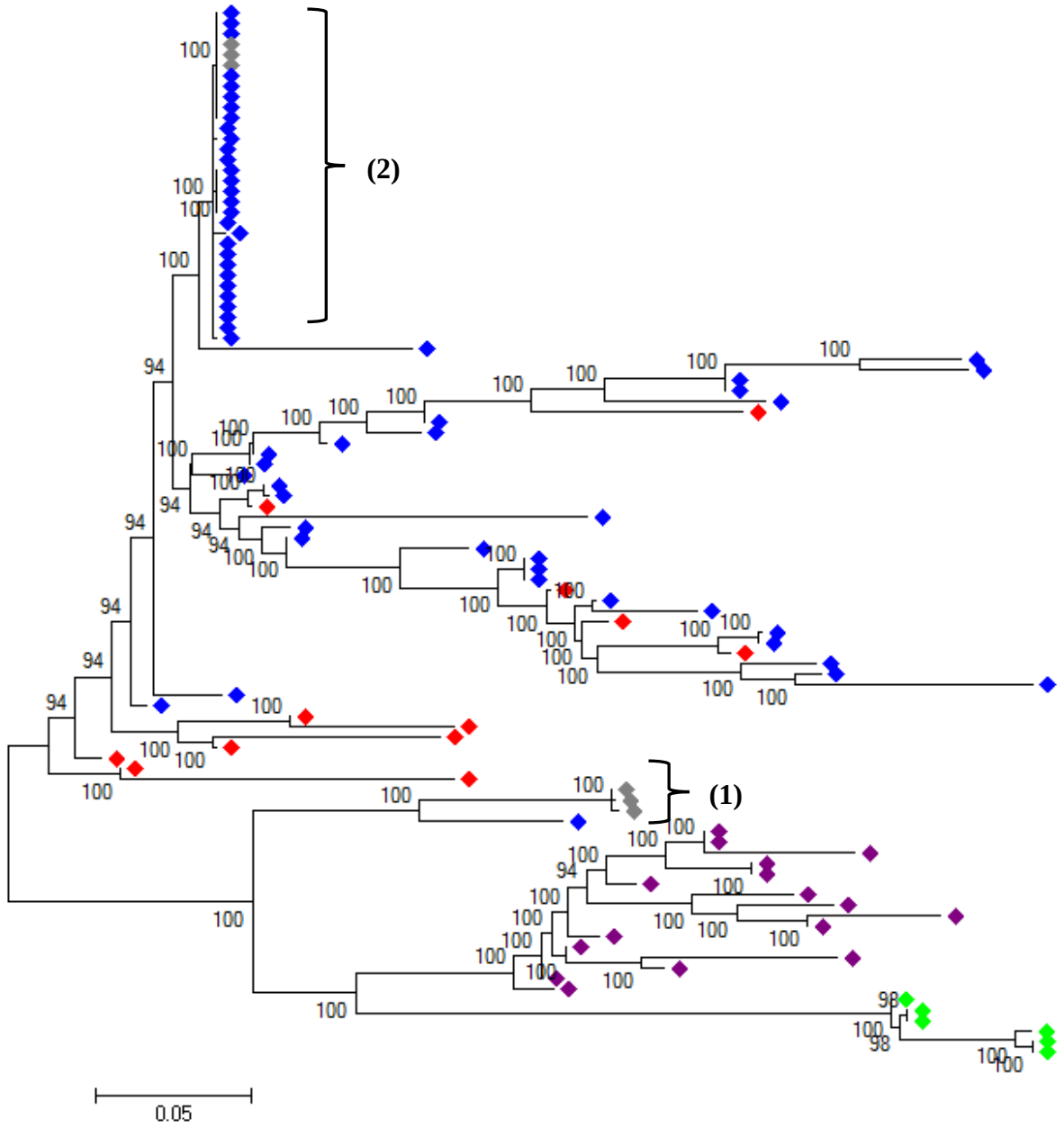


Figure S5.14. Haplotype based maximum likelihood tree for two patterns of introgression based on the 20-kb fragment size that extends from 17015001 to 17035001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of grey junglefowl haplotype into a Sri Lanka9. **(2)** Introgression of a domestic chicken haplotype into grey junglefowl.

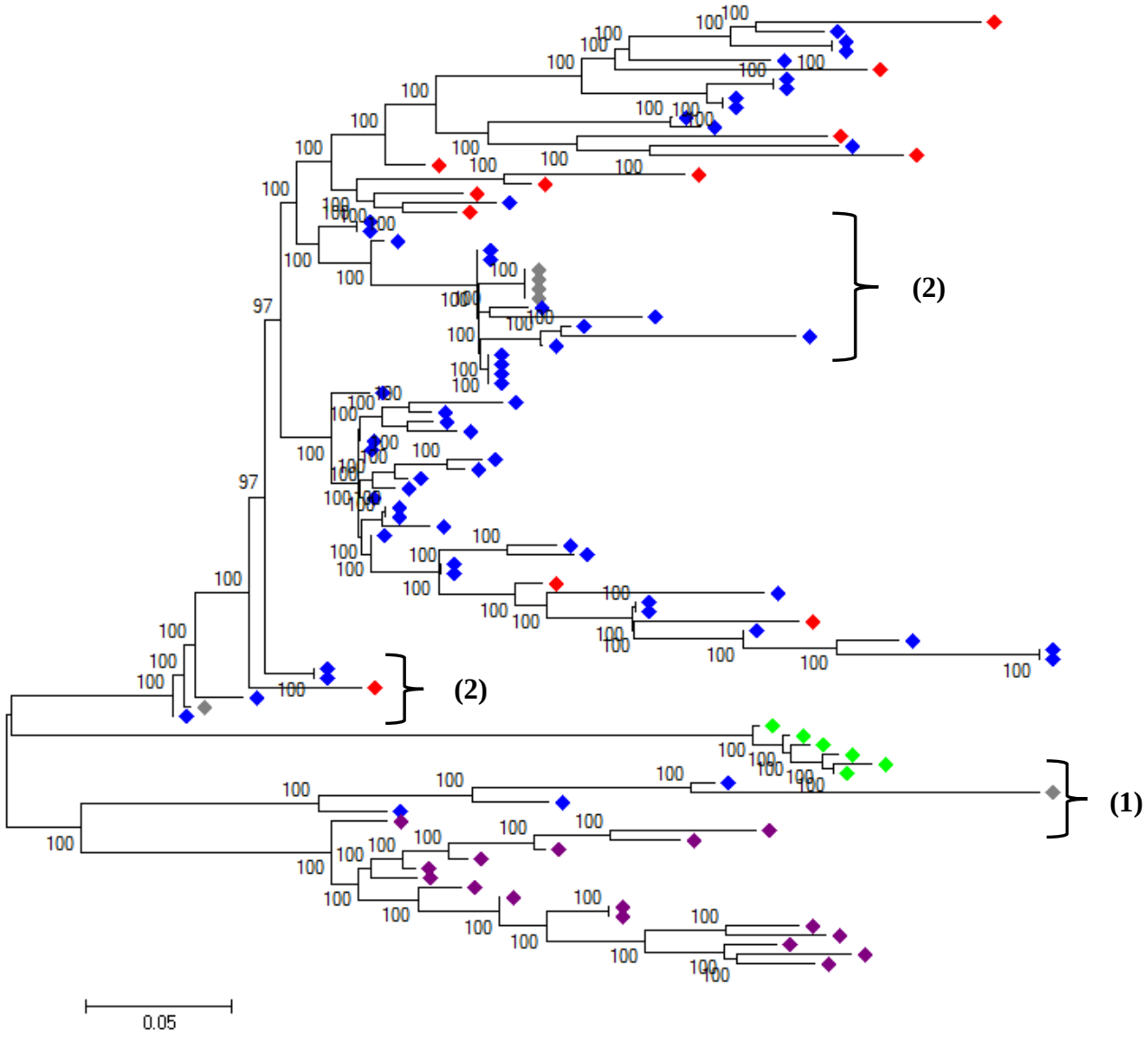


Figure S5.15. Haplotype based maximum likelihood tree for two patterns of introgression based on the 35-kb fragment size that extends from 33685001 to 33720001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl**. (1)** Introgression of grey junglefowl haplotype into Mechelse styrian, Saudi Arabia2 and Saudi Arabia1 samples. **(2)** Introgression of domestic chicken haplotypes into grey junglefowl.

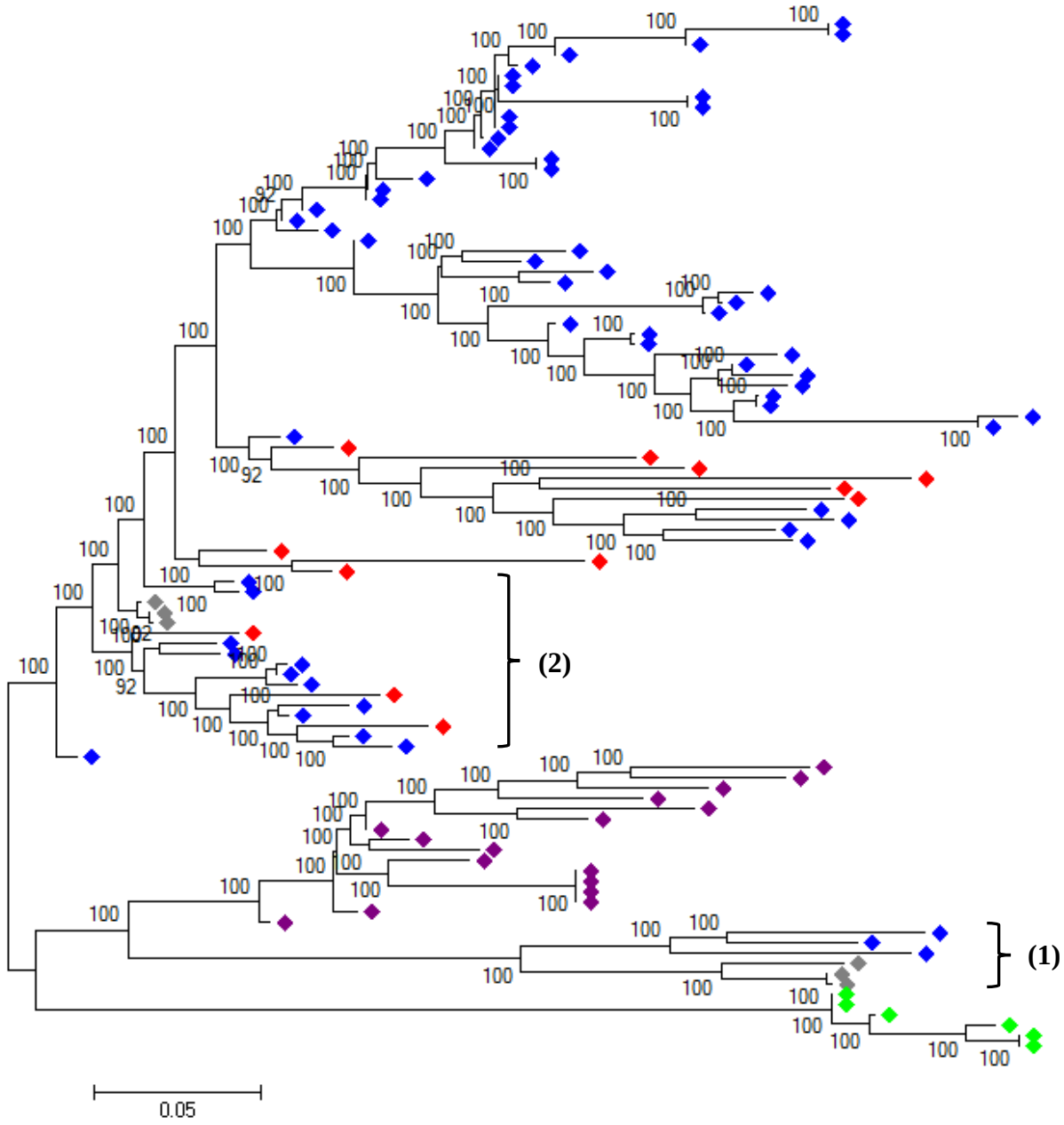


Figure S5.16. Haplotype based maximum likelihood tree for two patterns of introgression based on a 25-kb fragment size that extends from 37045001 to 37070001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl**. (1)** Introgression of a grey junglefowl haplotypes into Mechelse koekoe, HB1A16A and Sri Lanka9 samples. **(2)** Introgression of domestic chicken into grey junglefowl.

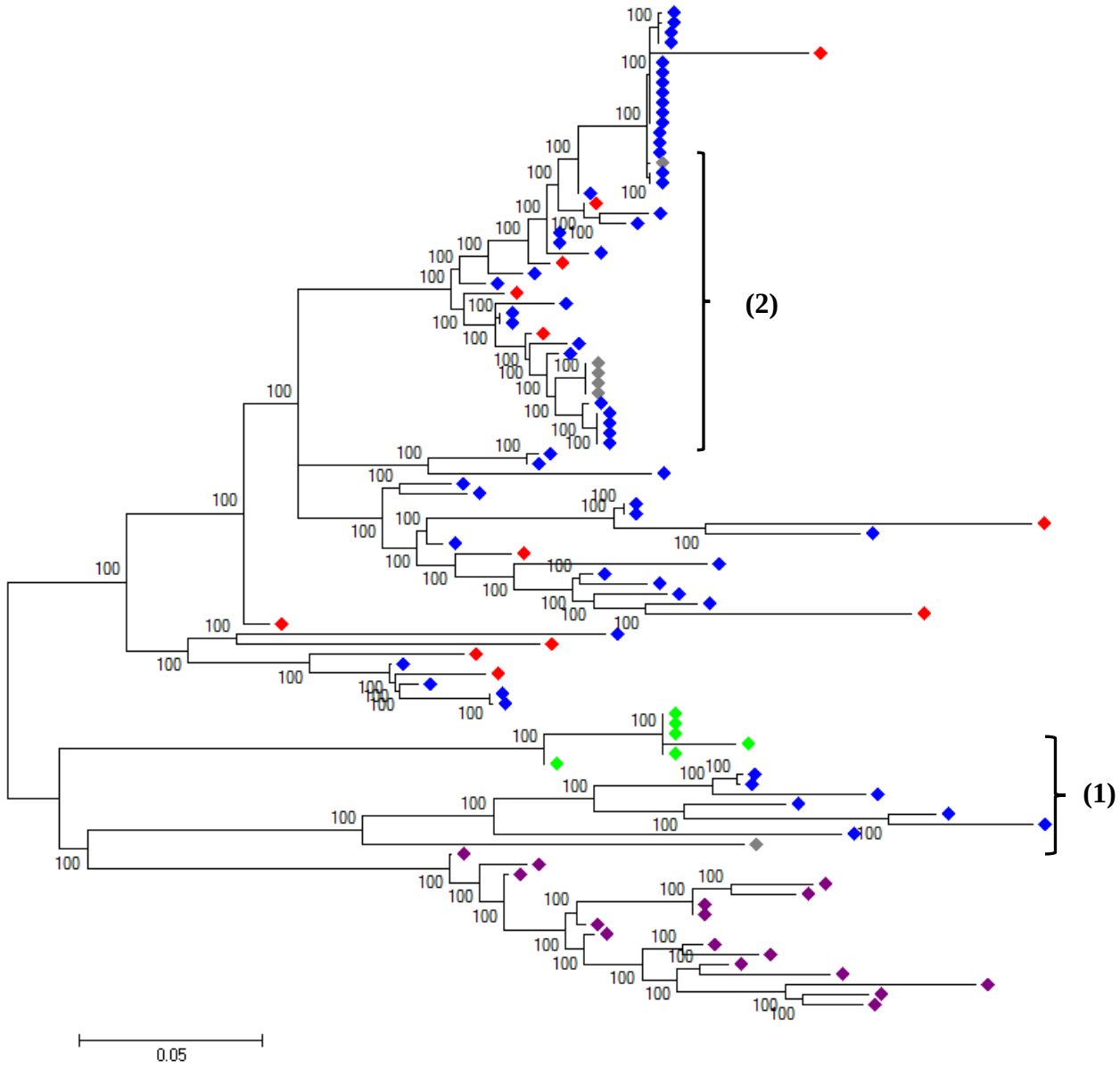


Figure S5.17. Haplotype based maximum likelihood tree for two patterns of introgression based on the 35-kb fragment size that extends from 41895001 to 41930001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of a grey junglefowl haplotype into Sri Lanka samples. **(2)** Introgression of domestic chicken haplotypes into grey junglefowl.

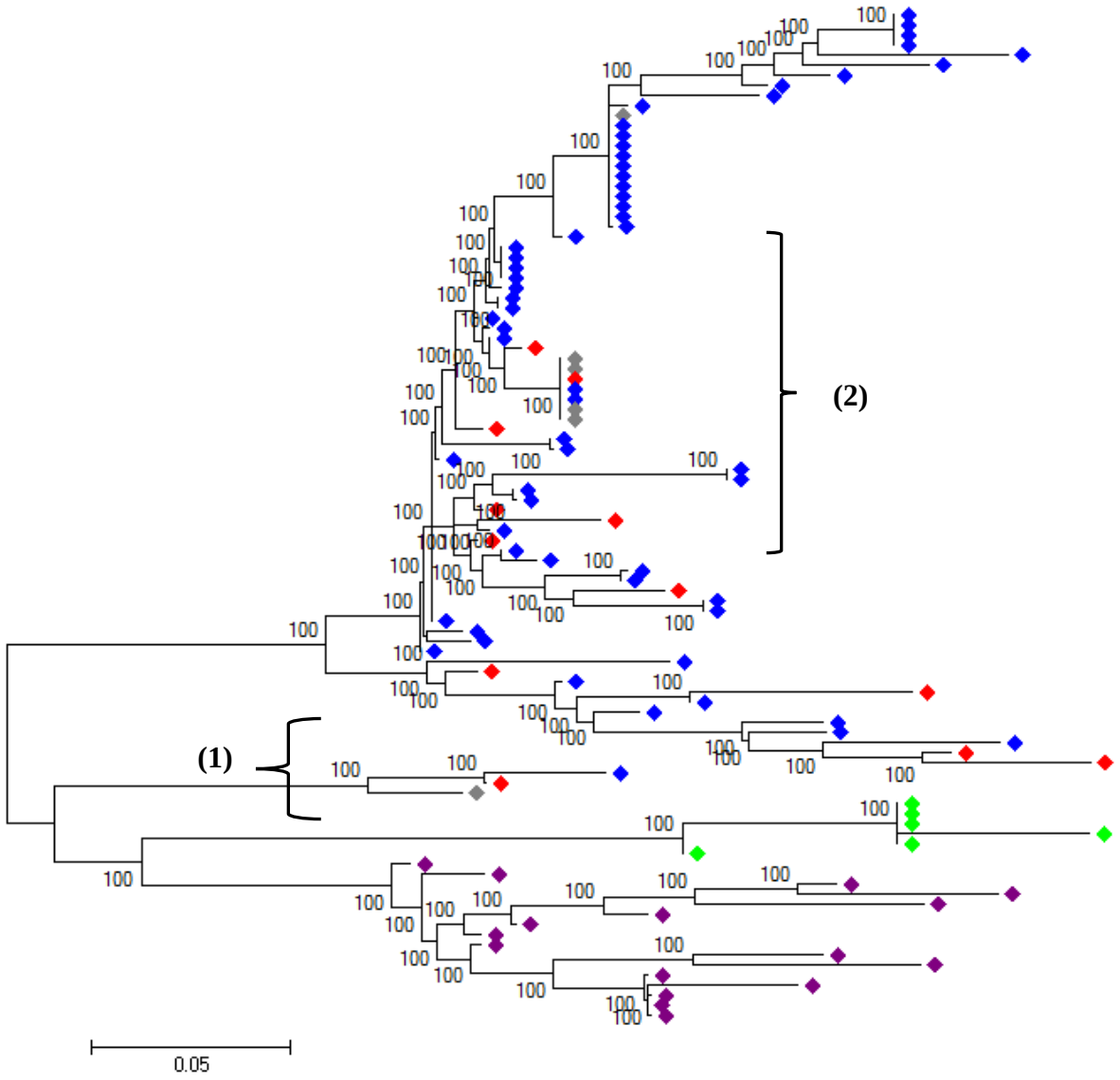


Figure S5.18. Haplotype based maximum likelihood tree for two patterns of introgression based on the 25-kb fragment size that extends from 41960001 to 41985001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of a grey junglefowl haplotype into Sri Lanka12 and red junglefowl2 samples. **(2)** Introgression of domestic chicken haplotypes into grey junglefowl.

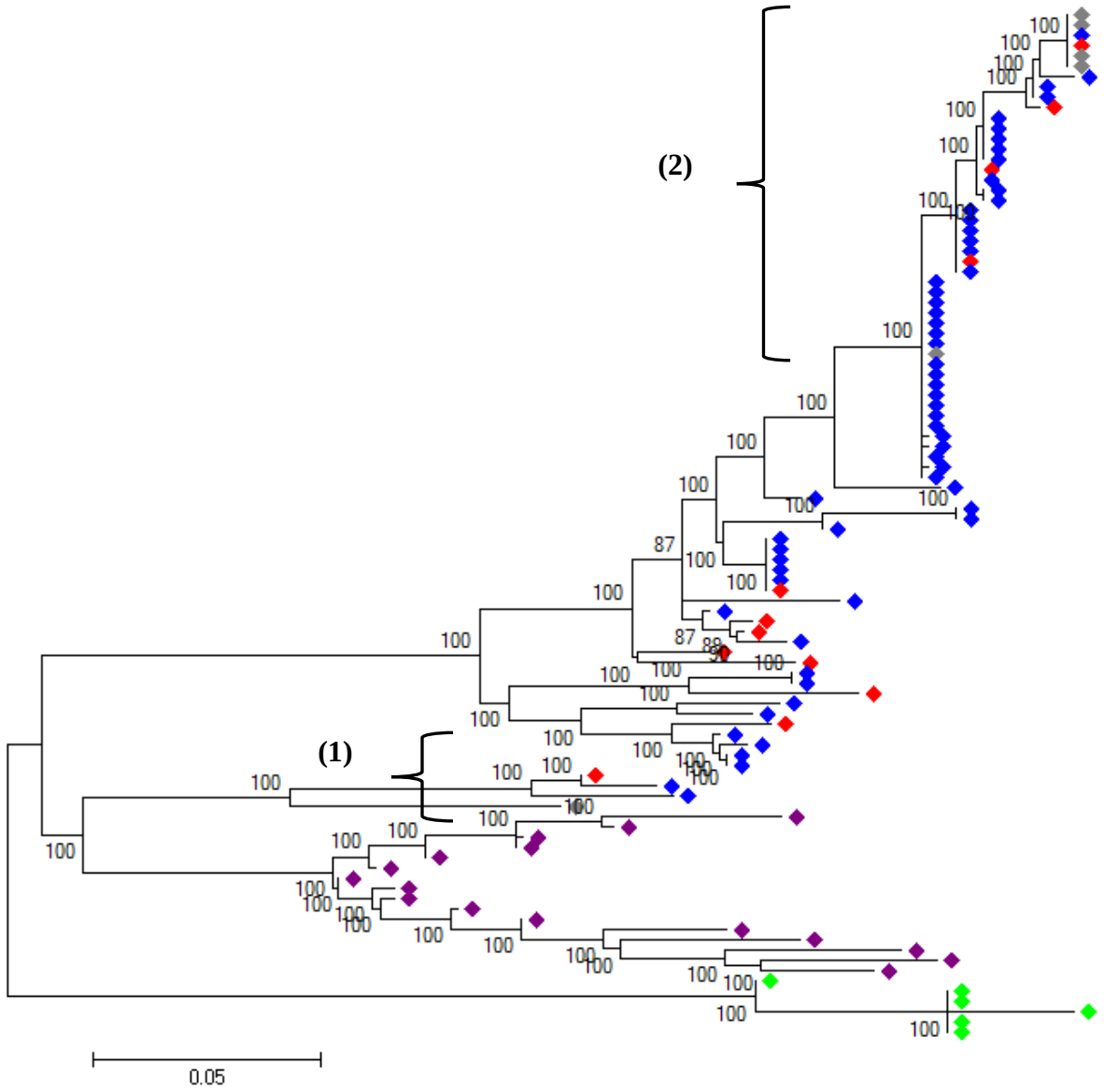


Figure S5.19. Haplotype based maximum likelihood tree for two patterns of introgression based on the 20-kb fragment size that extends from 42020001 to 42040001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of a grey junglefowl haplotype into Sri Lanka12, Sri Lanka9 and red junglefowl2 samples. **(2)** Introgression of domestic chicken haplotypes into grey junglefowl.

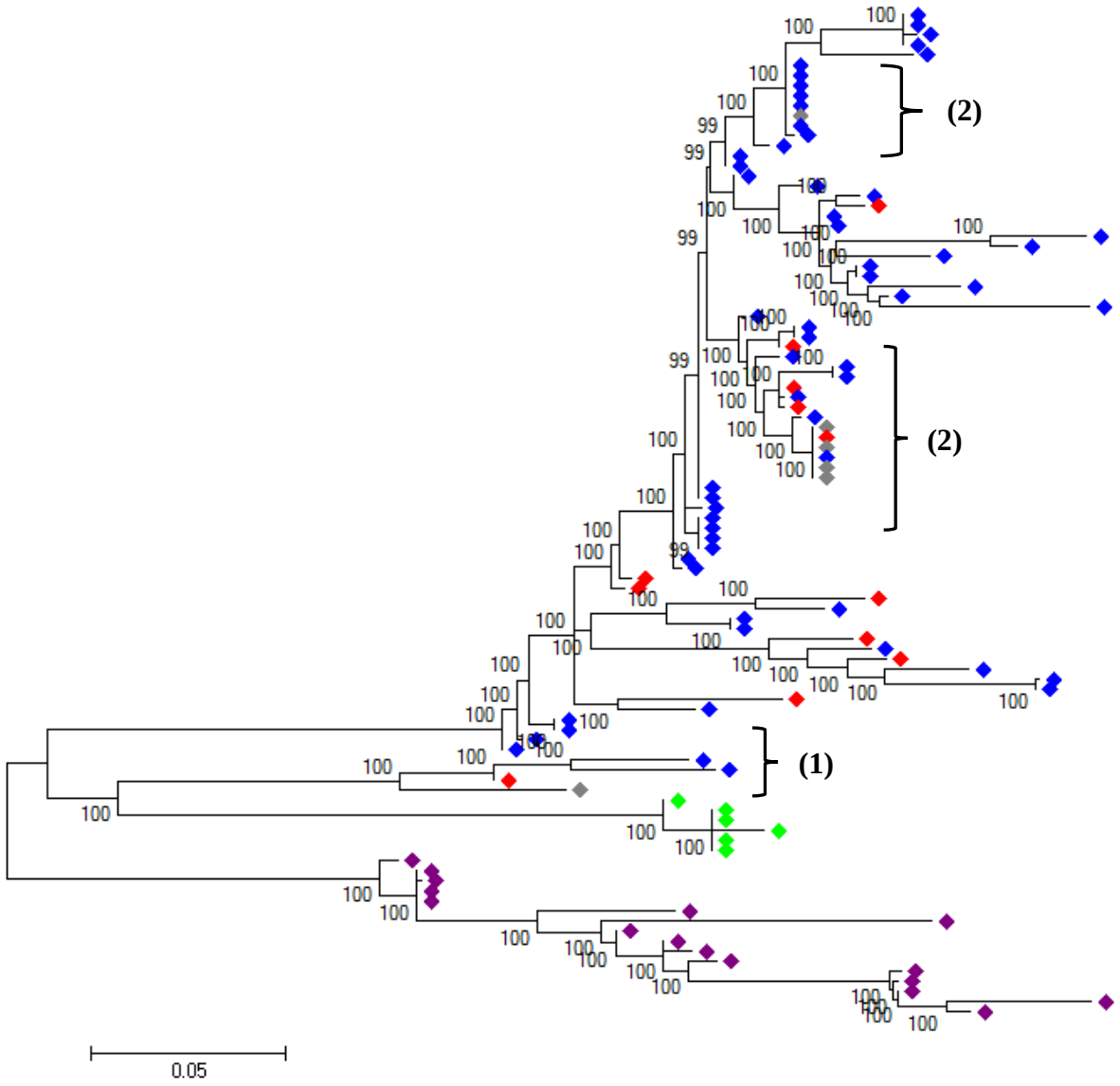


Figure S5.20. Haplotype based maximum likelihood tree for two patterns of introgression for a 30-kb fragment size that extends from 42085001 to 42115001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of a grey junglefowl haplotype into red junglefowl2, Sri Lanka9 and Sri Lanka12 samples. (**2)** Introgression of some domestic chicken haplotypes into grey junglefowl.

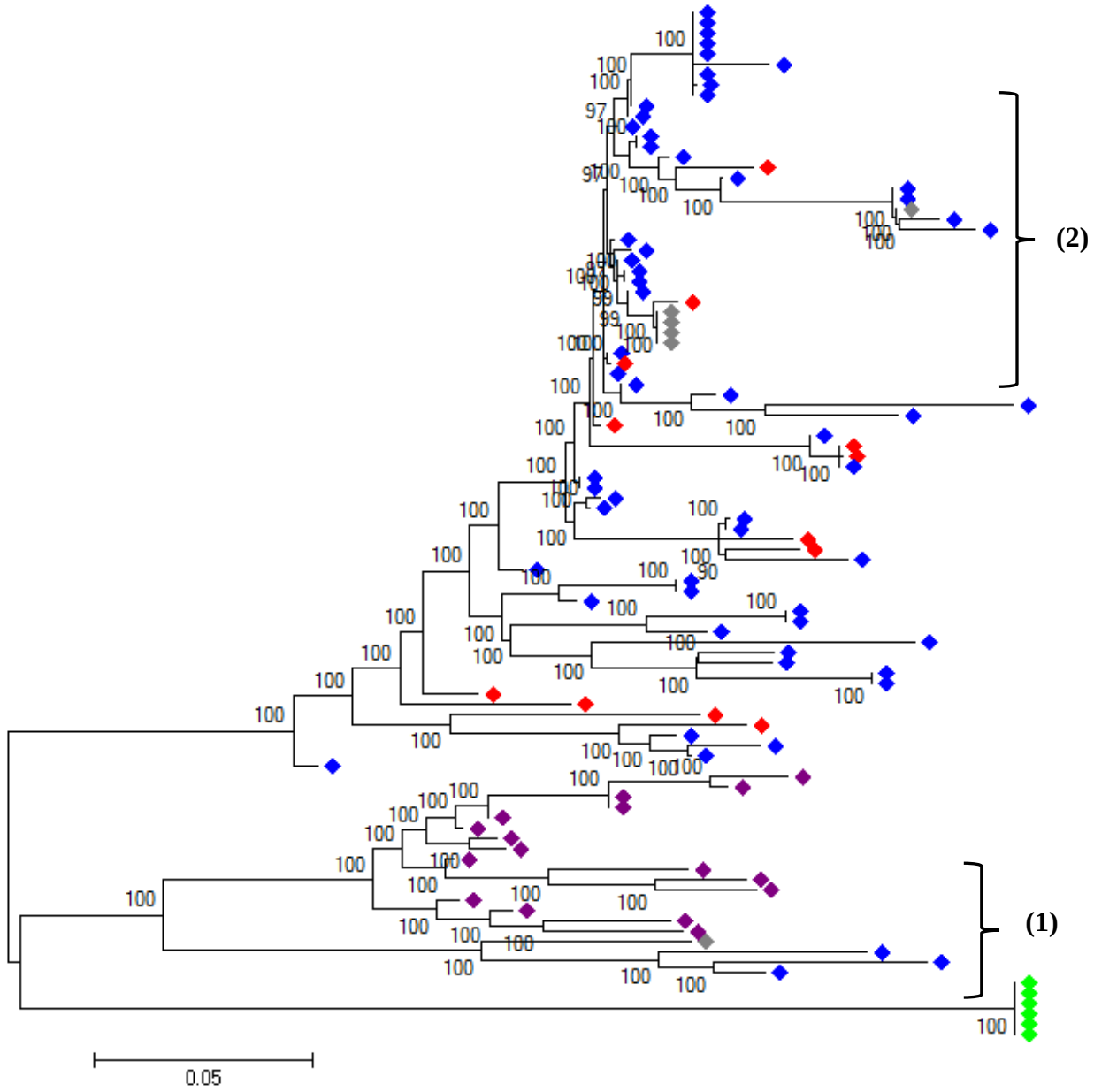


Figure S5.21. Haplotype based maximum likelihood tree for two patterns of introgression for a 35-kb fragment size that extends from 42395001 to 42430001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of a grey junglefowl haplotype into Sri Lanka2, Sri Lanka4 and Sri Lanka12 samples. (**2)** Introgression of some domestic chicken haplotypes into grey junglefowl.

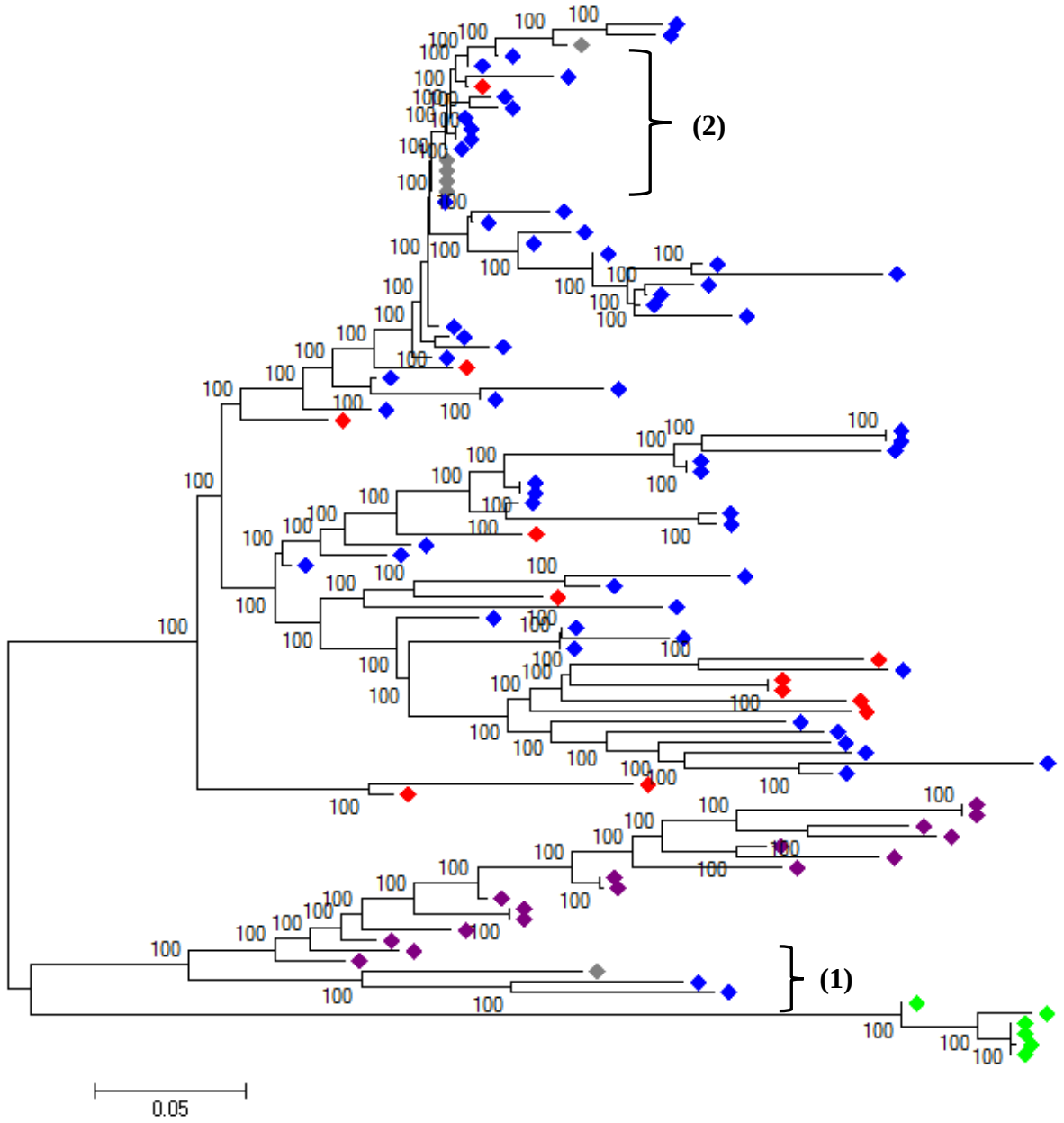


Figure S5.22. Haplotype based maximum likelihood tree for two patterns of introgression based on the 35-kb fragment size that extends from 45220001 to 45255001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of a grey junglefowl haplotype into the Mechelse koekoek and Sri Lanka16 samples. **(2)** Introgression of domestic chicken haplotypes into grey junglefowl.

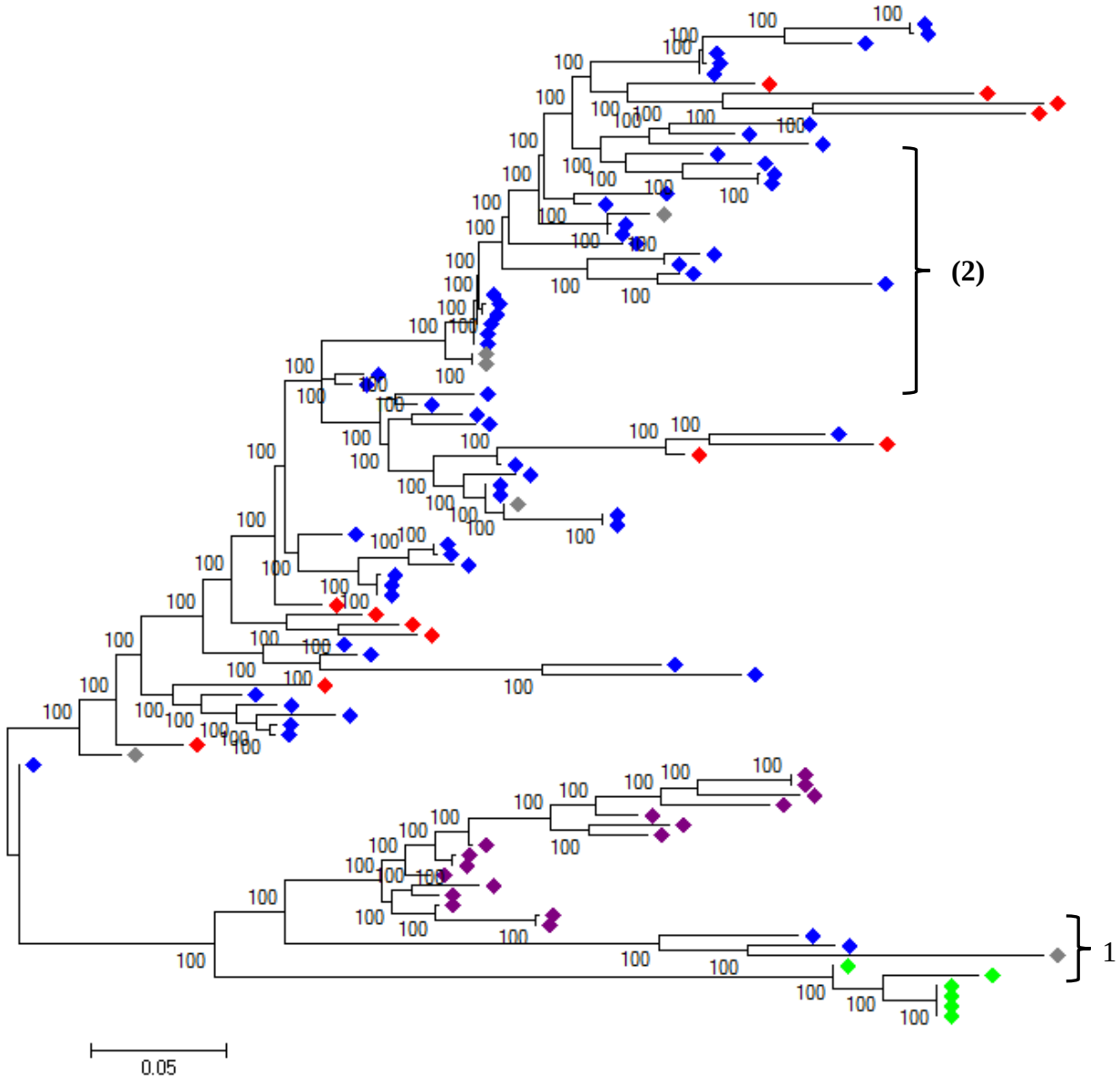


Figure S5.23. Haplotype based maximum likelihood tree for two patterns of introgression based on a 20-kb fragment size that extends from 45950001 to 45970001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl**. (1**) Introgression of grey junglefowl haplotype into JB1A25B and JB1B16A samples. (**2**) Introgression of domestic/red junglefowl haplotypes into grey junglefowl.

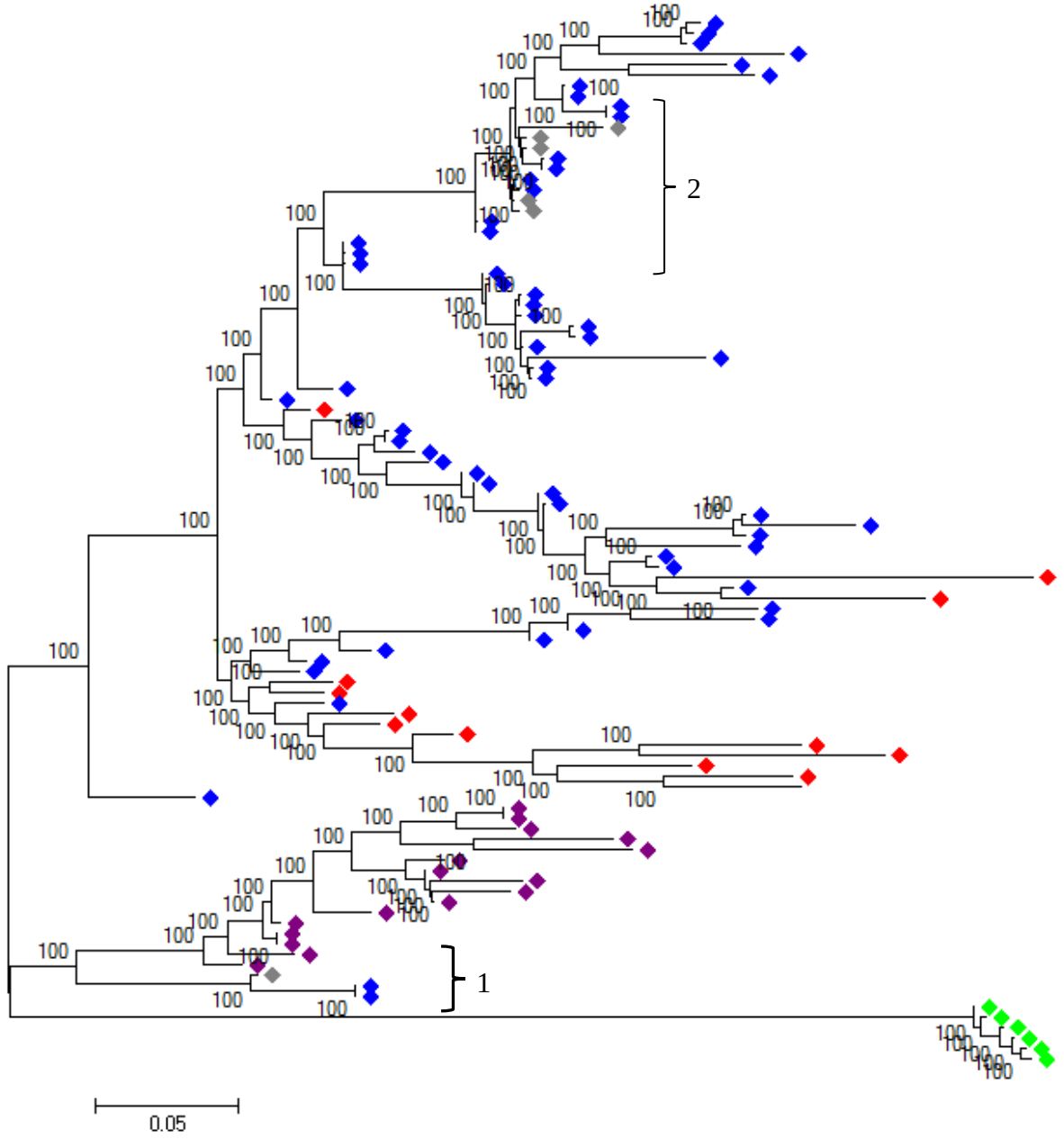


Figure S5.24. Haplotype based maximum likelihood tree for two patterns of introgression based on a 35-kb fragment size that extends from 46970001 to 47005001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl**. (1)** Introgression of one grey junglefowl haplotype into Sri Lanka10 and Sri Lanka13 samples. (**2**) Introgression of domestic chicken haplotypes into grey junglefowl.

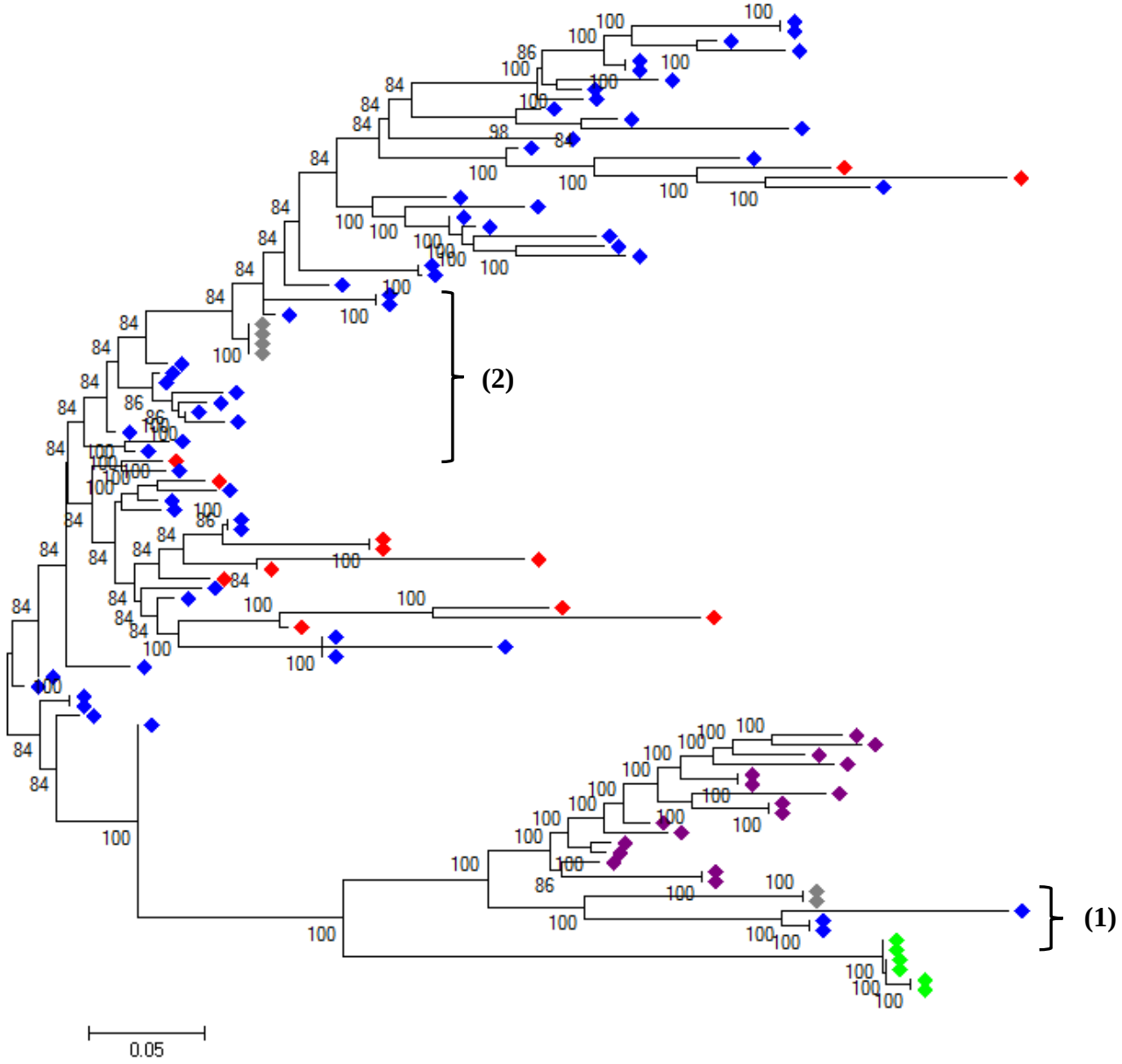


Figure S5.25. Haplotype based maximum likelihood tree for two patterns of introgression based on the 20-kb fragment size that extends from 1260001 to 1280001 bp on chromosome 6. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl **(1)** Introgression of grey junglefowl haplotypes into HB1A16A and HA1A22A. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

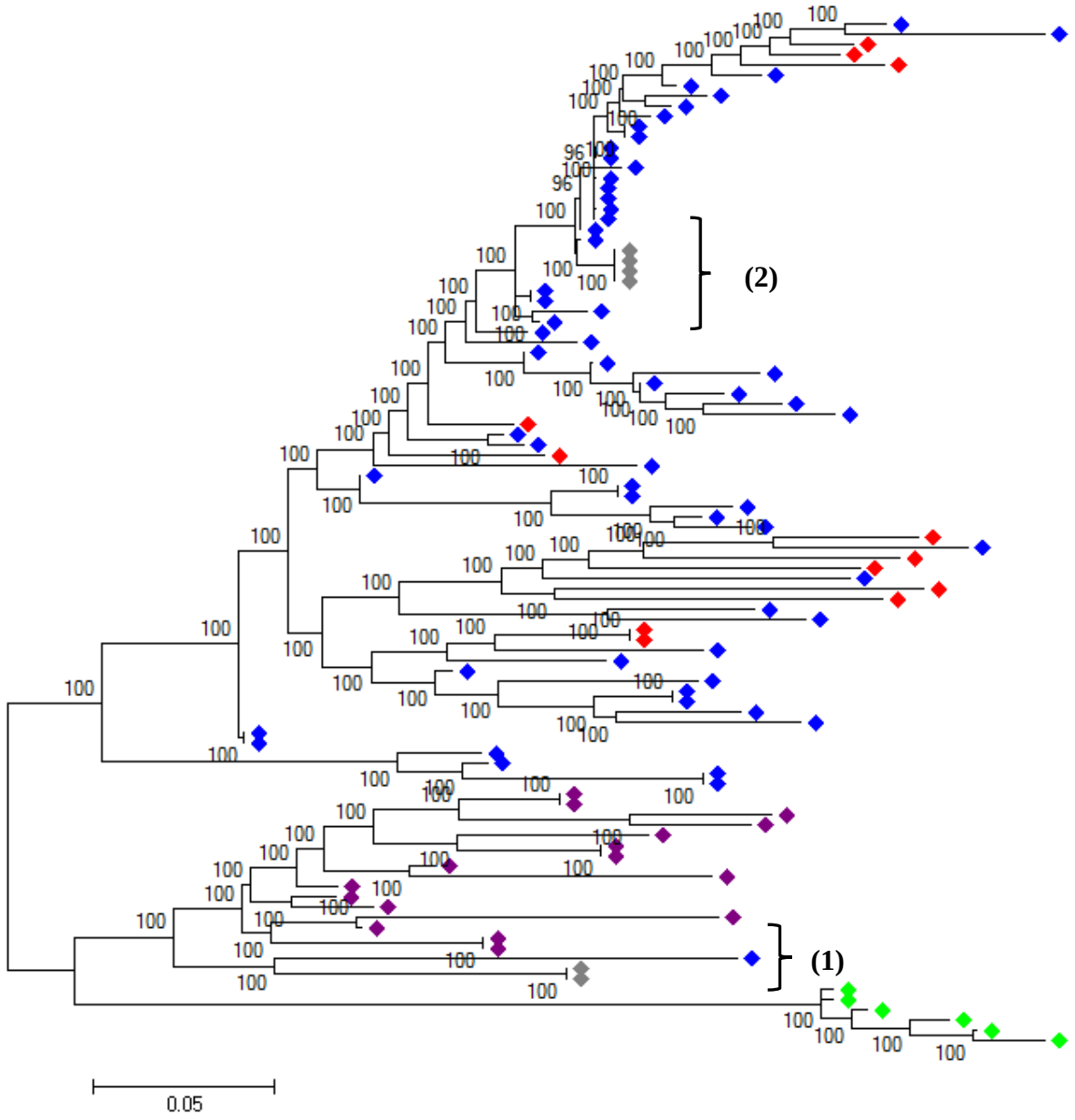


Figure S5.26. Haplotype based maximum likelihood tree for two patterns of introgression based on the 30-kb fragment size that extends from 1370001 to 1400001 bp on chromosome 6. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of one grey junglefowl haplotypes into JB1B16A. **(2)** Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

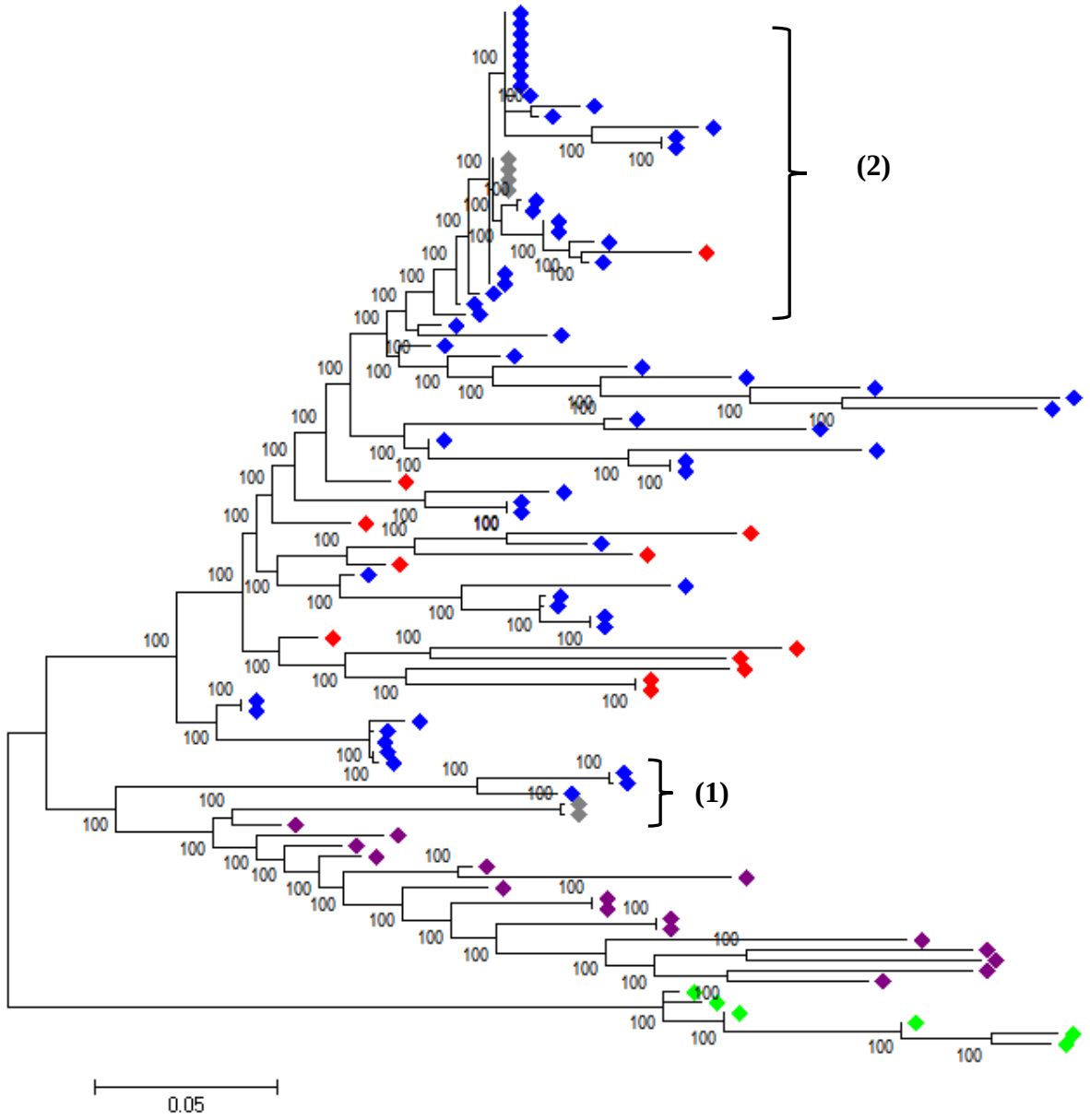


Figure S5.27. Haplotype based maximum likelihood tree for two patterns of introgression based on the 30-kb fragment size that extends from 1415001 to 1445001 bp on chromosome 6. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl **(1)** Introgression of grey junglefowl haplotypes into Sri Lanka1, 2 and 3 samples. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

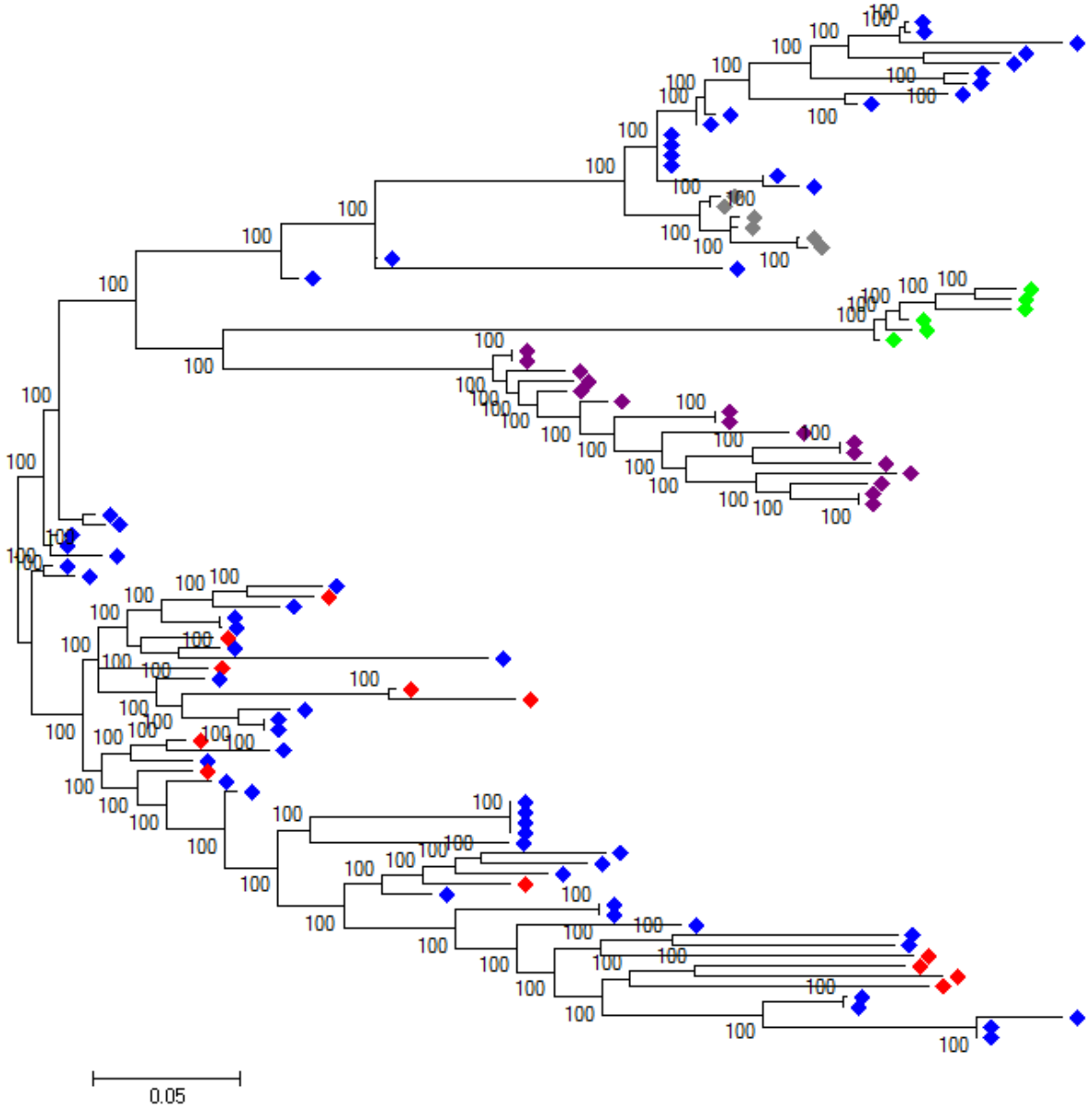


Figure S5.28. Haplotype based maximum likelihood tree of introgression of grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 45-kb fragment size that extends from 22200001 to 22245001 bp on chromosome 7.

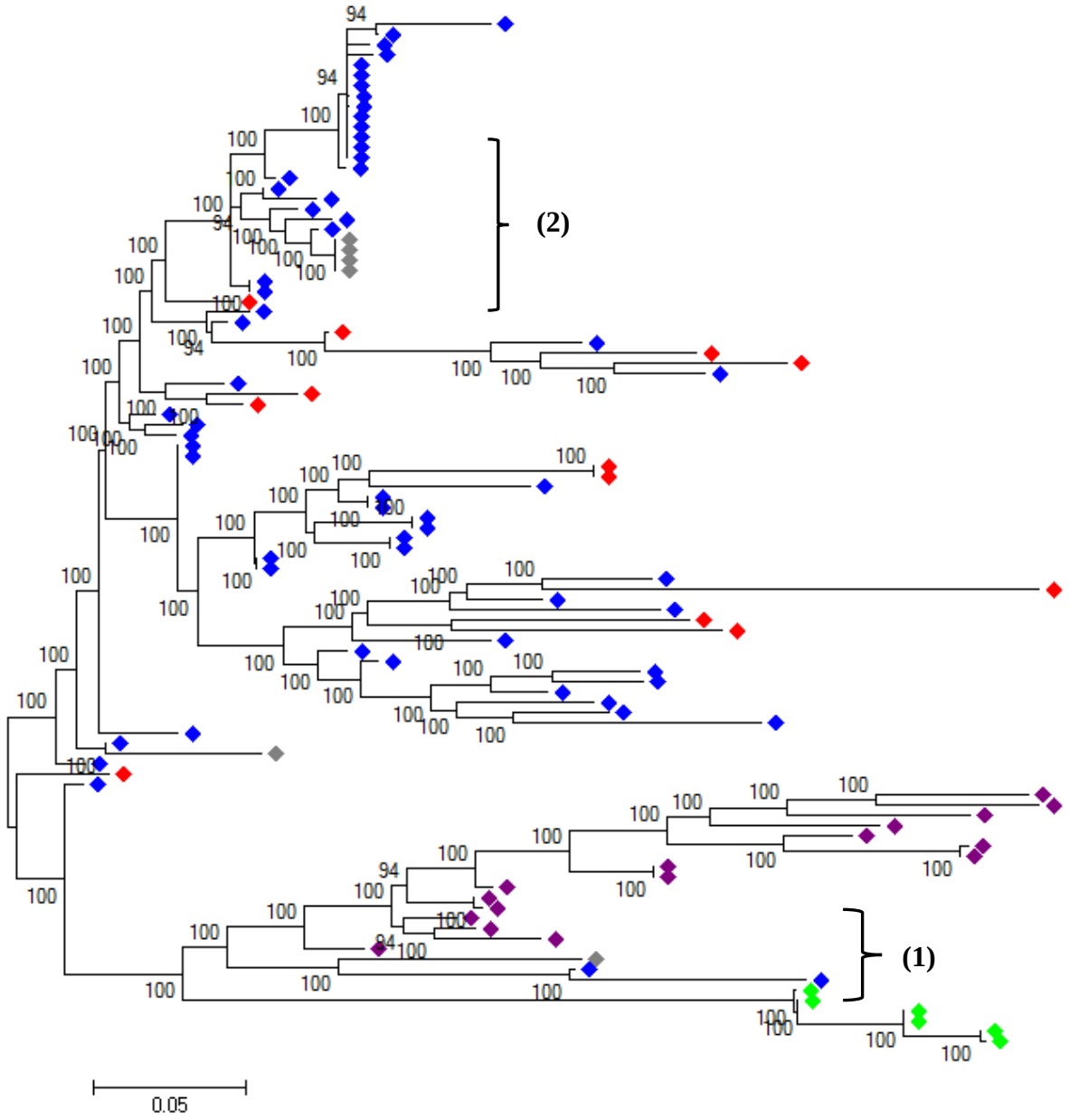


Figure S5.29. Haplotype based maximum likelihood tree for two patterns of introgression based on the 25-kb fragment size that extends from 2165001 to 2190001 bp on chromosome 13. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl **(1)** Introgression of grey junglefowl haplotypes into Sri Lanka2, and 4 samples. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

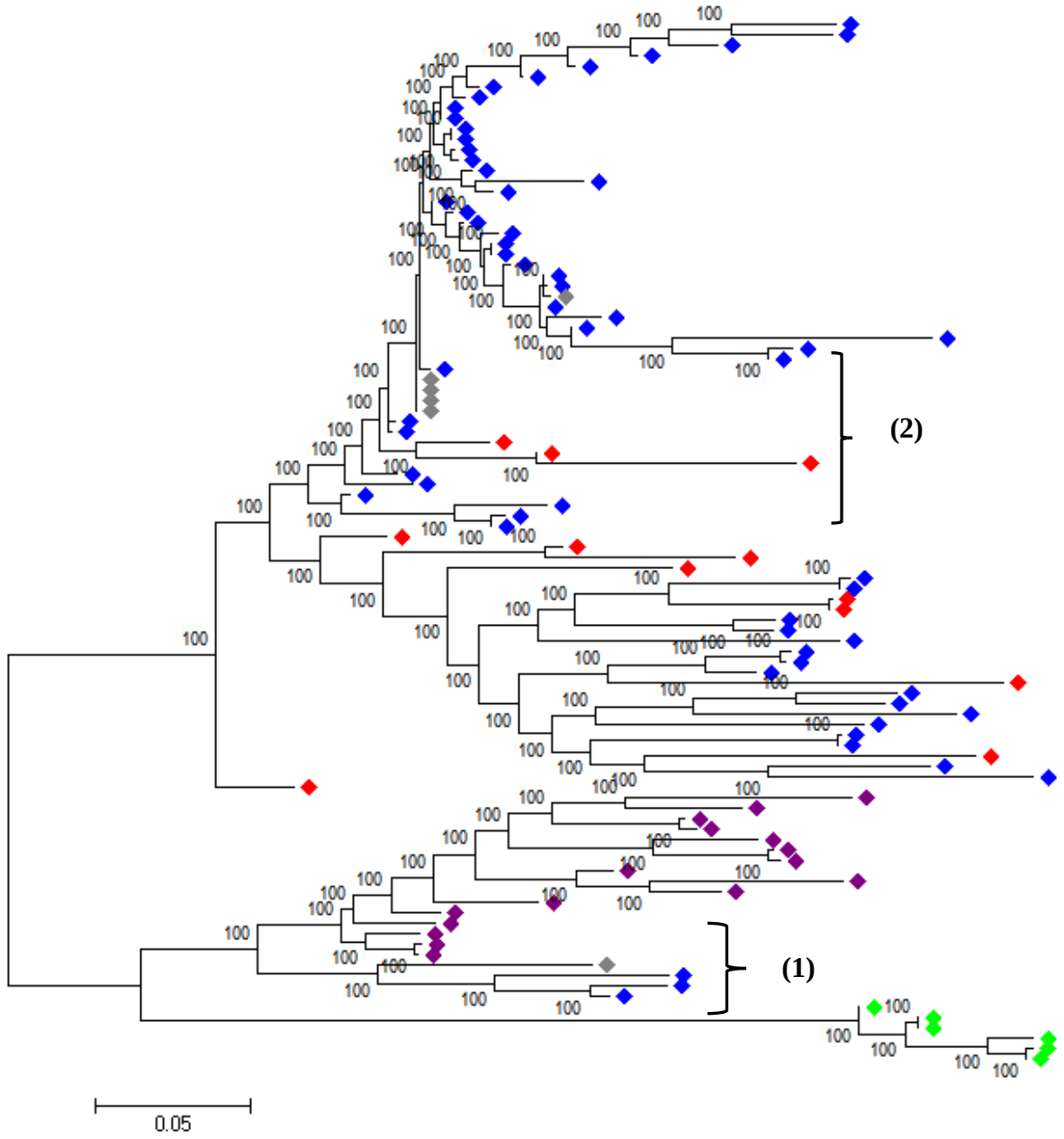


Figure S5.30. Haplotype based maximum likelihood tree for two patterns of introgression based on the 20-kb fragment size that extends from 2395001 to 2415001 bp on chromosome 13. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl **(1)** Introgression of grey junglefowl haplotypes into Saudi Arabia3 and 5, JB1B16A samples. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

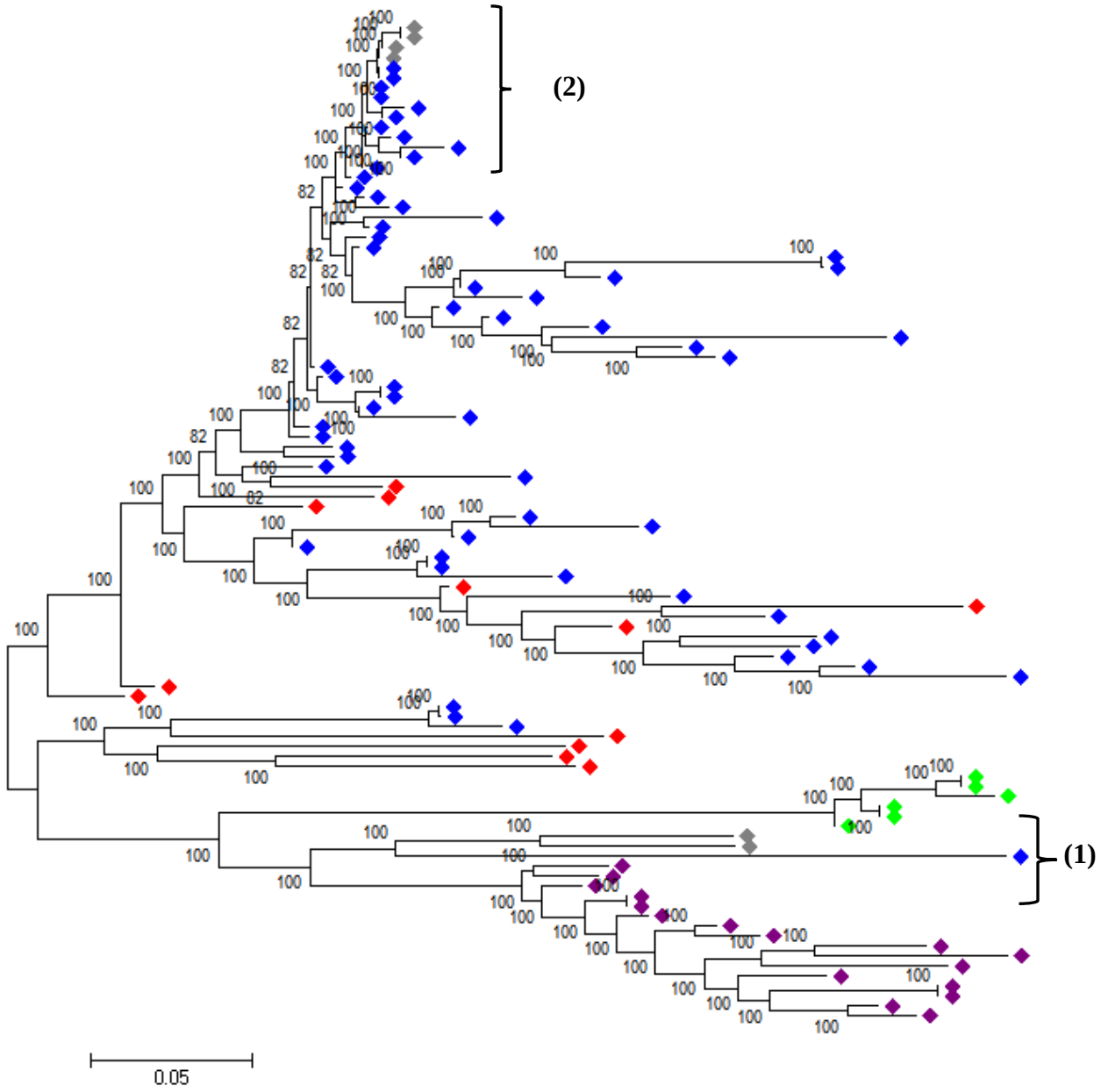


Figure S5.31. Haplotype based maximum likelihood tree for two patterns of introgression based on the 45-kb fragment size that extends from 1635001 to 1680001 bp on chromosome 19. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of grey junglefowl haplotypes into Sri Lanka9 sample. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

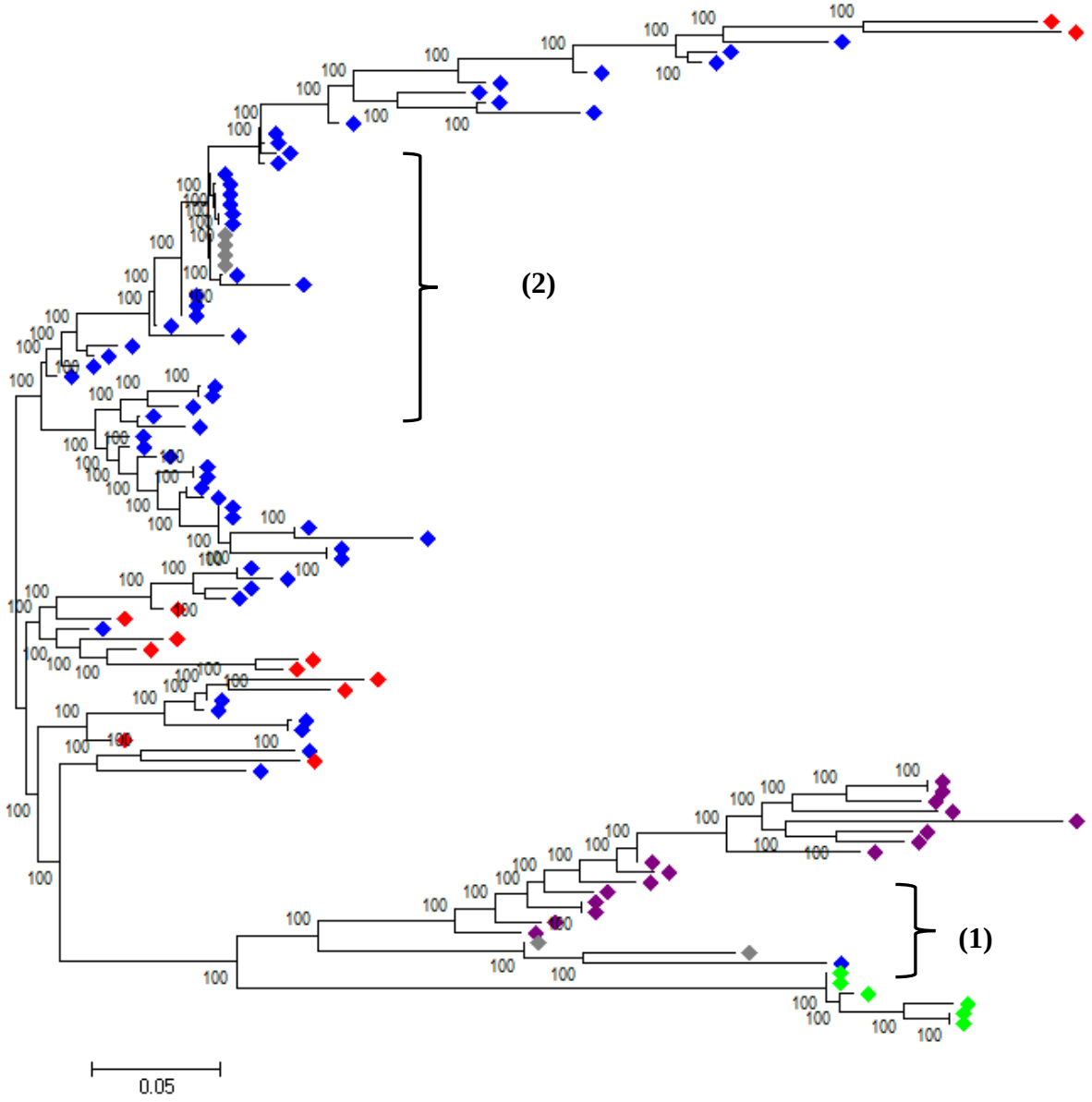


Figure S5.32. Haplotype based maximum likelihood tree for two patterns of introgression based on the 25-kb fragment size that extends from 13345001 – 13370001 bp on chromosome 20. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl **(1)** Introgression of grey junglefowl haplotypes into Sri Lanka4 sample. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

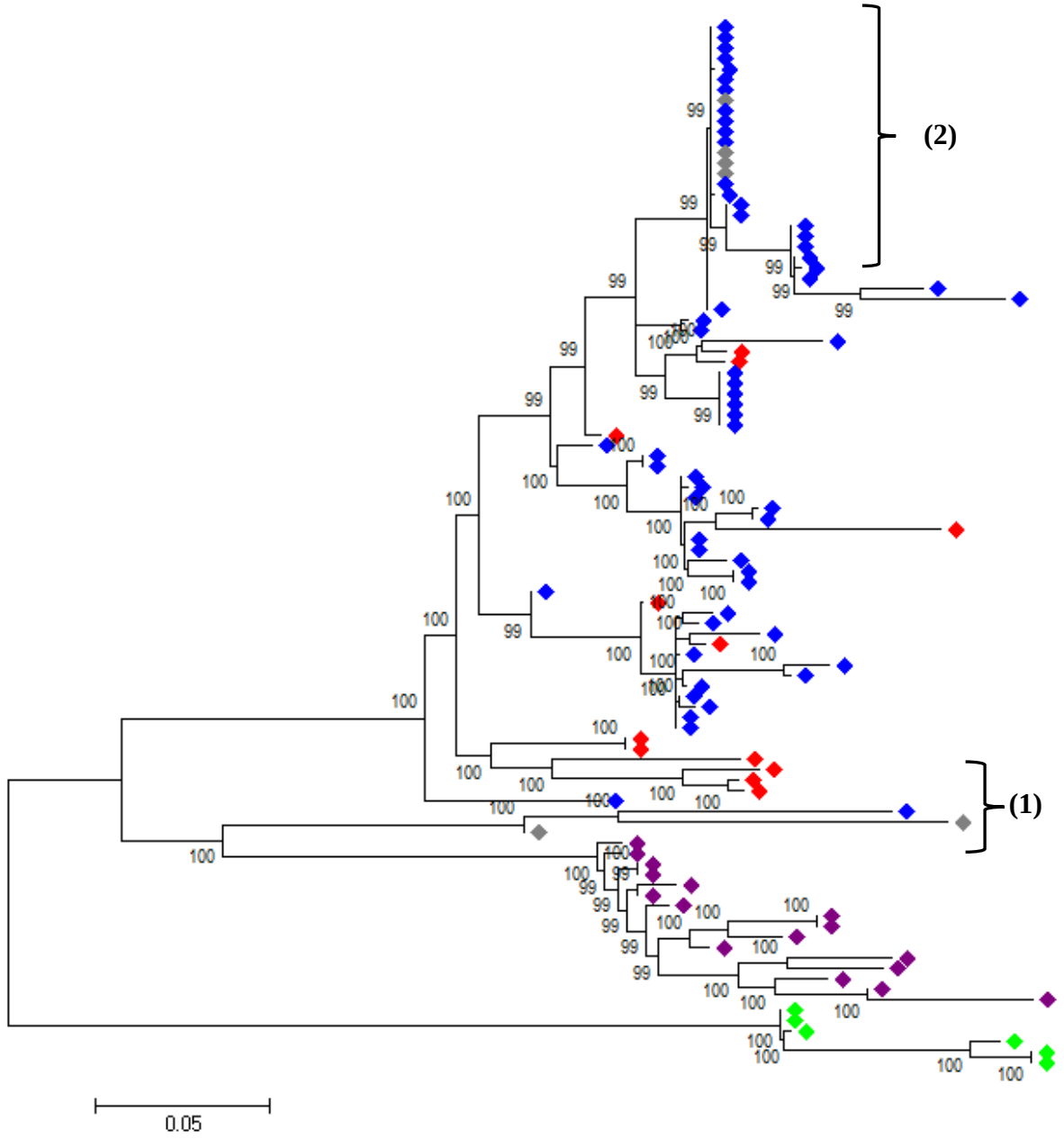


Figure S5.33. Haplotype based maximum likelihood tree for two patterns of introgression based on the 35-kb fragment size that extends from 13505001 to 13540001 bp on chromosome 20. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl **(1)** Introgression of grey junglefowl haplotypes into Sri Lanka5 sample. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

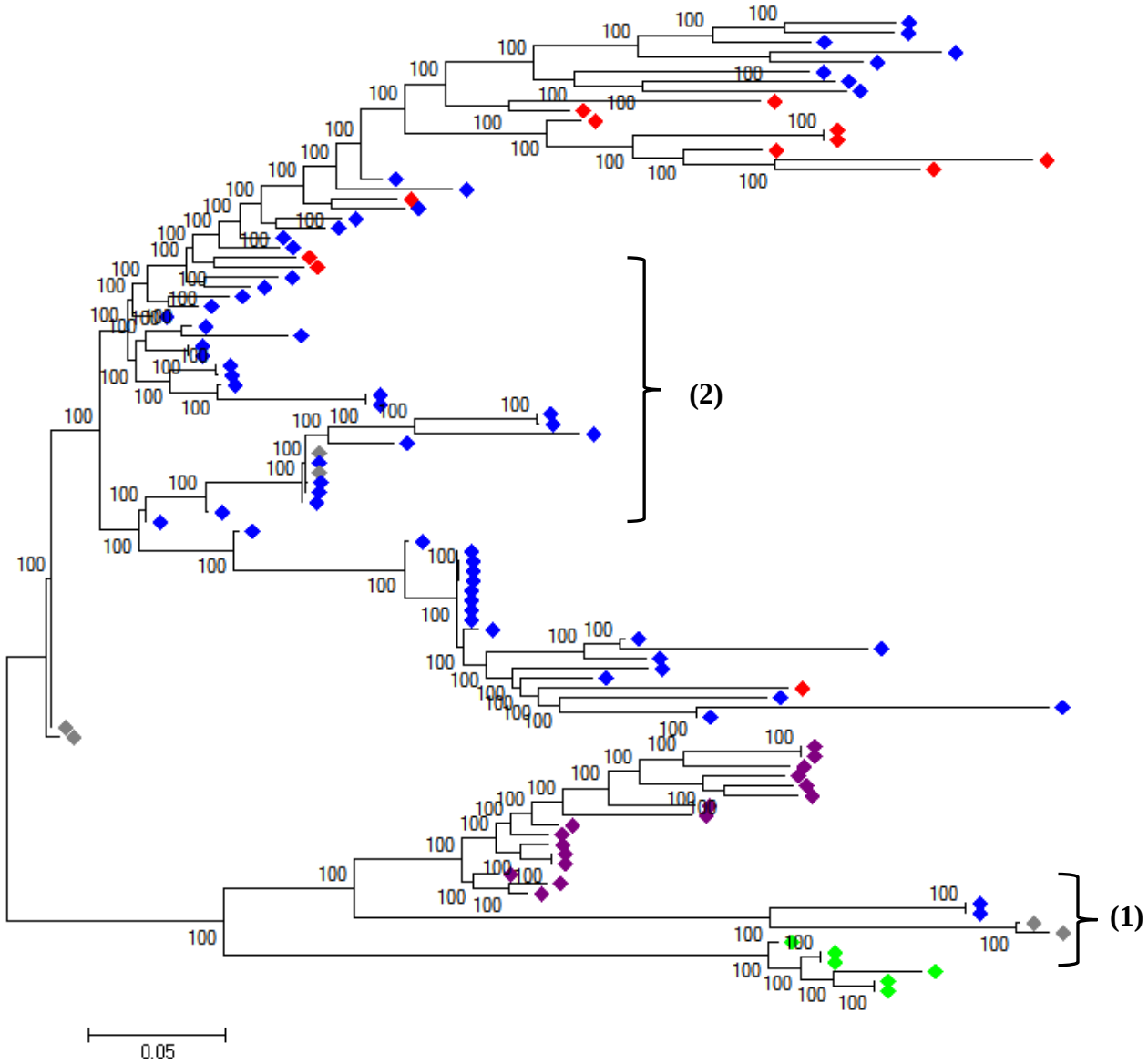


Figure S5.34. Haplotype based maximum likelihood tree for two patterns of introgression based on the 25-kb fragment size that extends from 3005001 to 3030001 bp on chromosome 21. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of grey junglefowl haplotypes into Sri Lanka1 and 2 samples. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

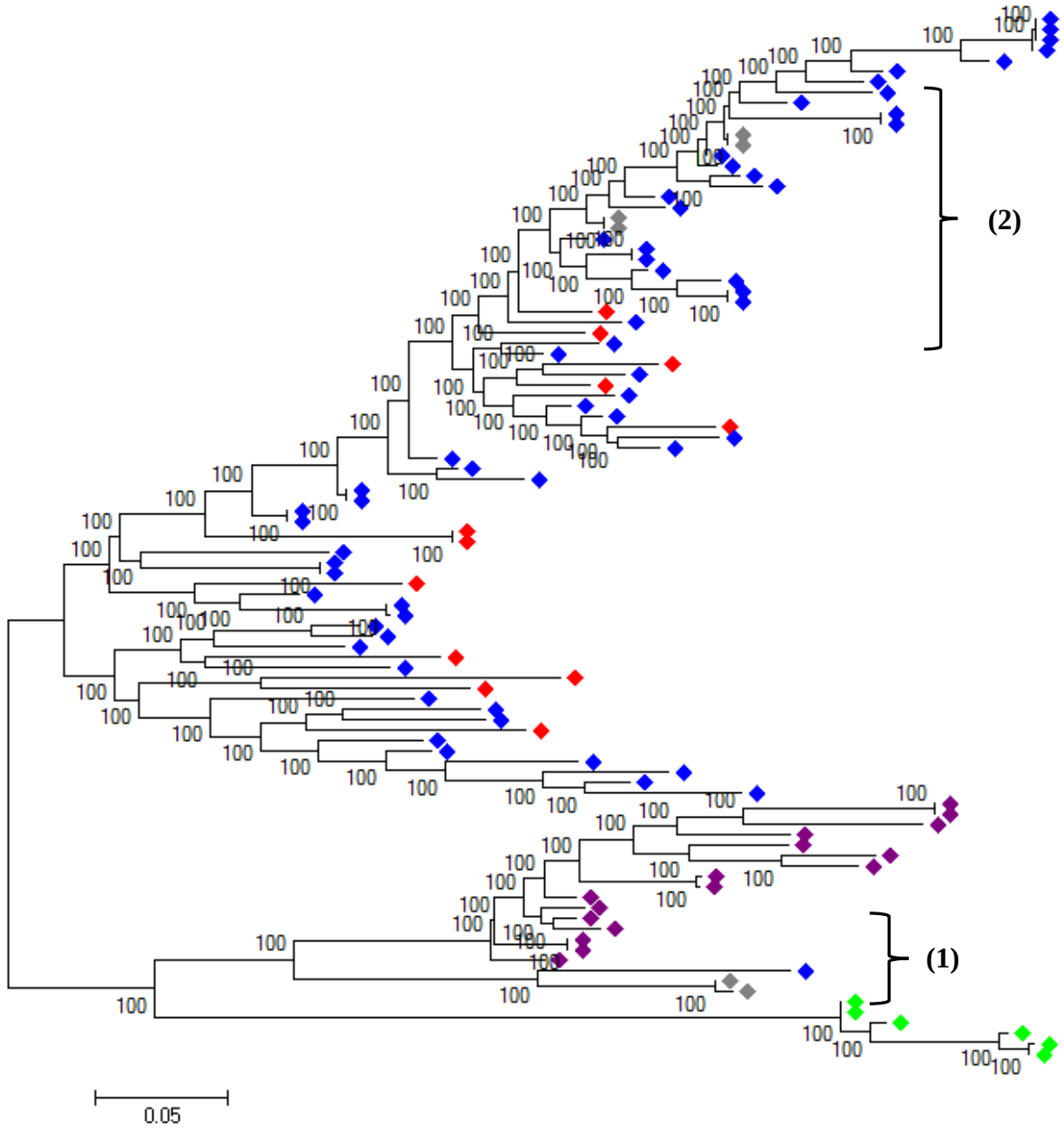


Figure S5.35. Haplotype based maximum likelihood tree for two patterns of introgression based on the 20-kb fragment size that extends from 3230001 to 3250001 bp on chromosome 21. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl **(1)** Introgression of grey junglefowl haplotypes into Sri Lanka5 sample. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

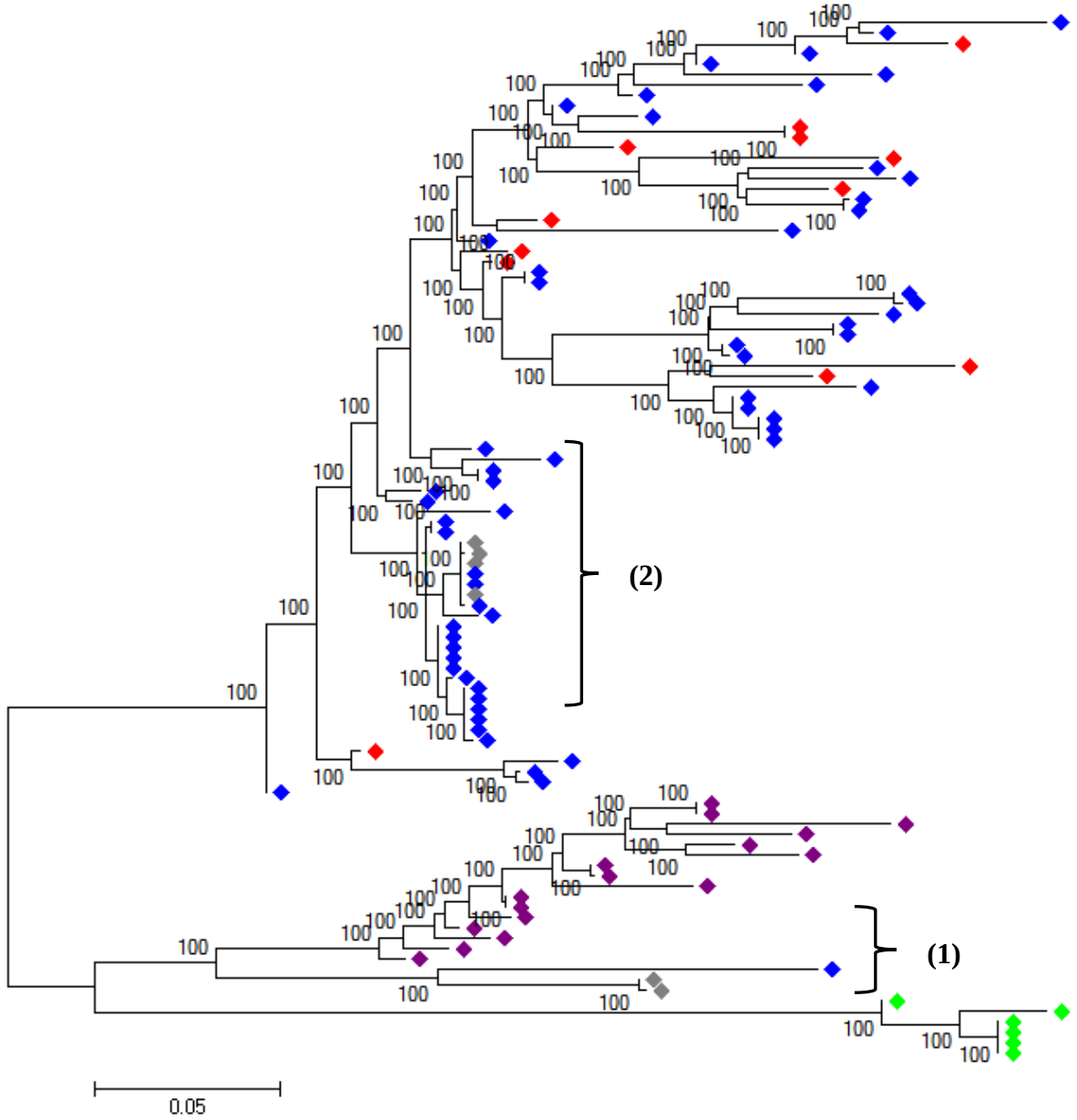


Figure S5.36. Haplotype based maximum likelihood tree for two patterns of introgression based on the 25-kb fragment size that extends from 3335001 to 3360001 bp on chromosome 21. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl **(1)** Introgression of grey junglefowl haplotypes into Sri Lanka6 sample. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

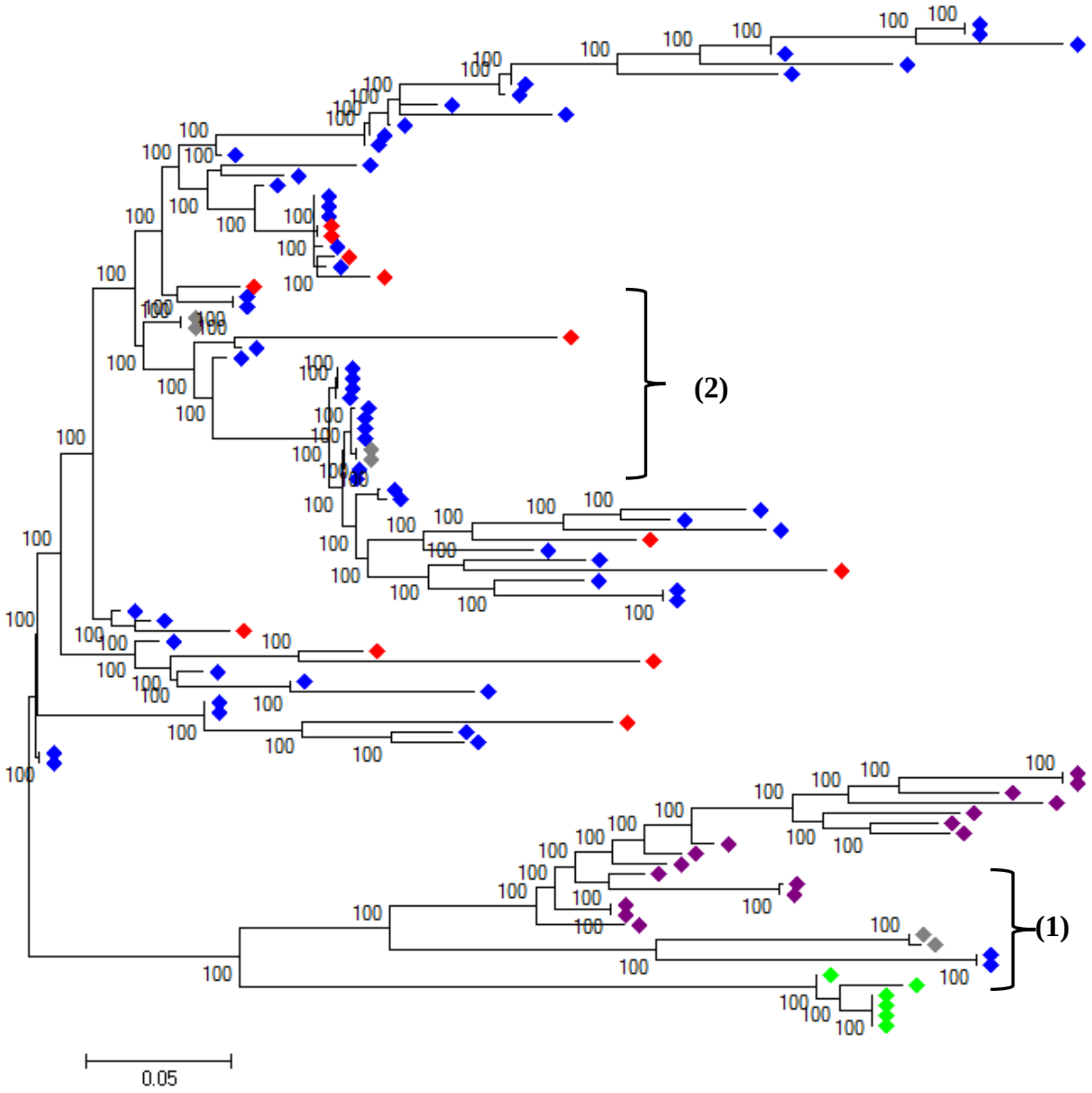


Figure S5.37. Haplotype based maximum likelihood tree for two patterns of introgression based on the 25-kb fragment size that extends from 3495001 – 3520001 bp on chromosome 21. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of grey junglefowl haplotypes into Sri Lanka1 and 3 samples. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

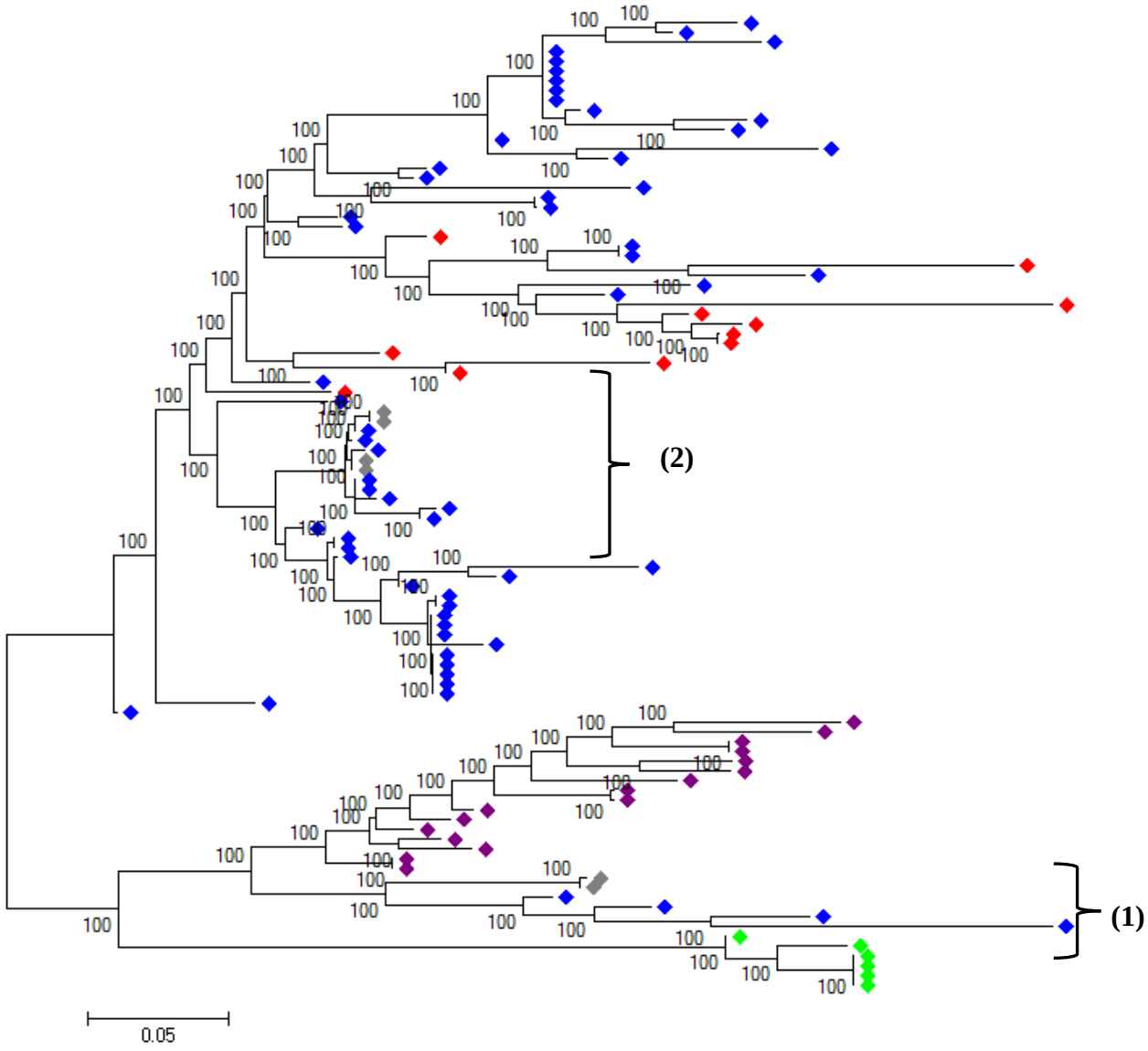


Figure S5.38. Haplotype based maximum likelihood tree for two patterns of introgression based on the 25-kb fragment size that extends from 3570001 to 3595001 bp on chromosome 21. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of grey junglefowl haplotypes into Saudi Arabia1, 2 and 5 and Sri Lanka16 samples. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

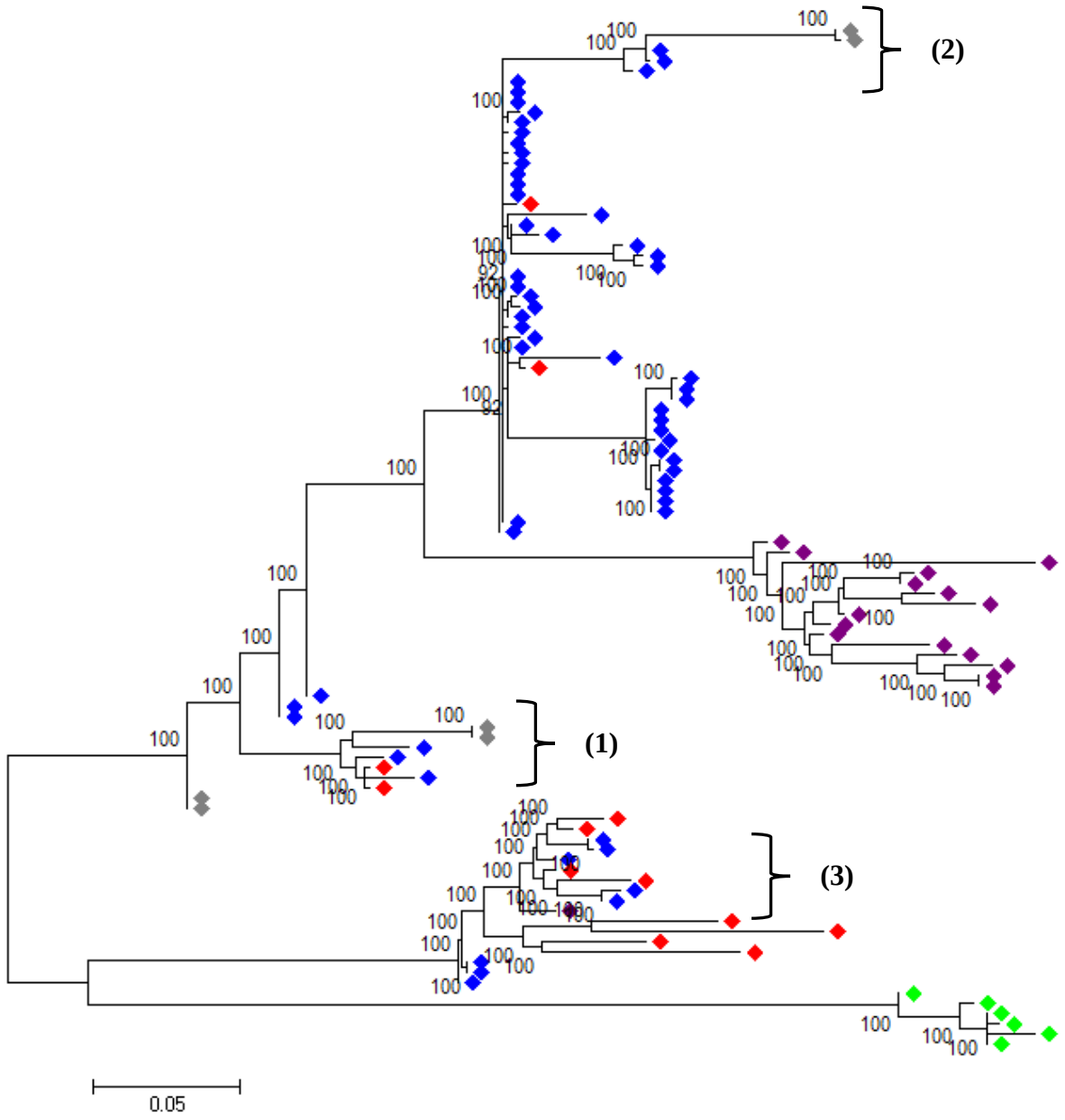


Figure S5.39. Haplotype based maximum likelihood tree of introgression of grey junglefowl into domestic chicken. 30-kb fragment size that extends from 6130001 - 6160000 bp on chromosome 24. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Evidence of introgression of grey junglefowl into domestic chicken. **(2)** Introgression of domestic chicken into grey junglefowl. **(3)** Introgression of domestic chicken in Ceylon junglefowl. The region includes the previously reported grey junglefowl introgressed *BCDO2* region (Eriksson *et al.* 2008).

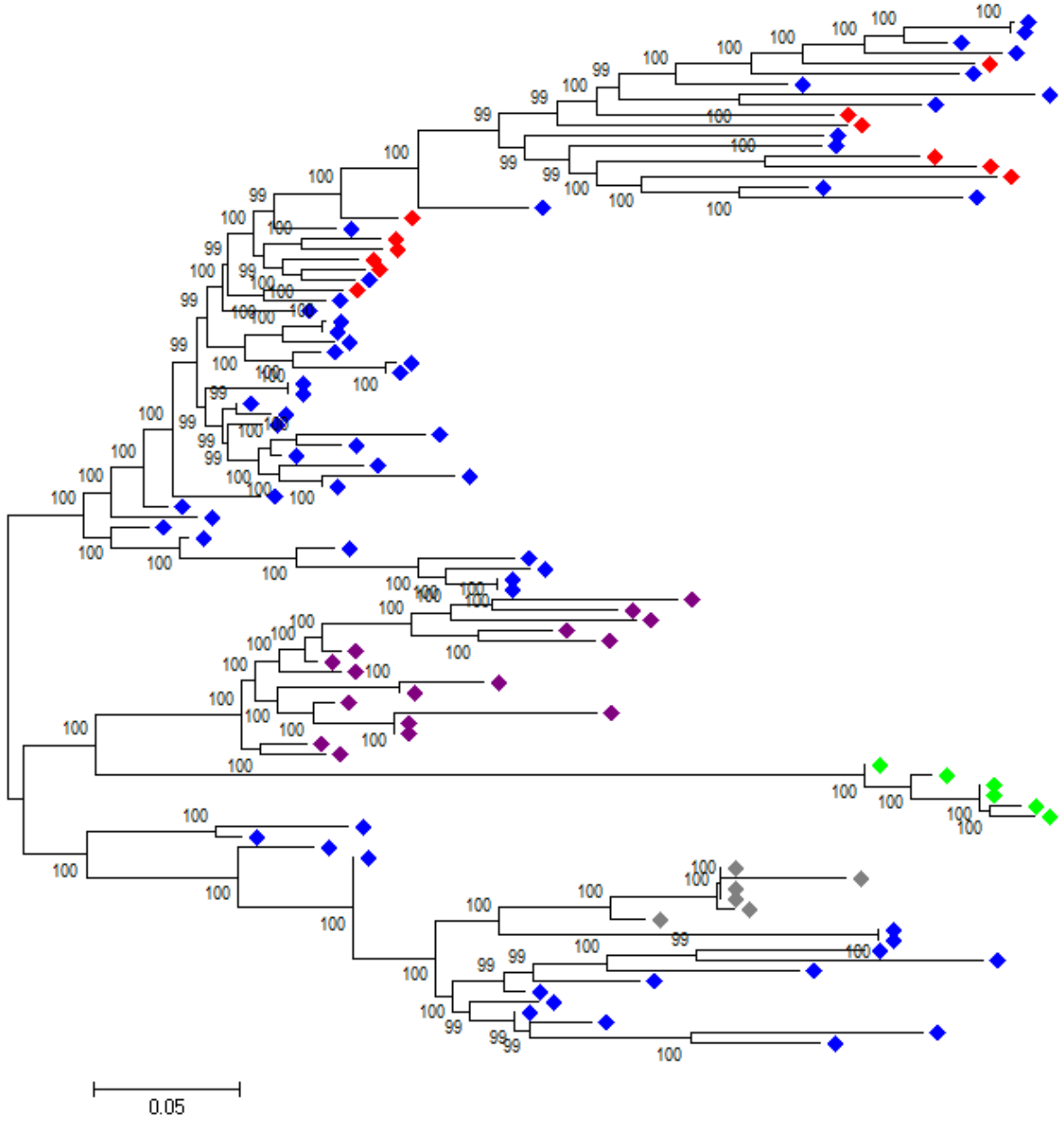


Figure S5.40. Haplotype based maximum likelihood tree of introgression of grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 2110001 to 2130001 bp on chromosome 28.

Figures S6. Haplotype relationships analysis for candidate regions of introgression from green junglefowl into domestic chicken. ABBA – BABA test grey junglefowl - domestic chicken.

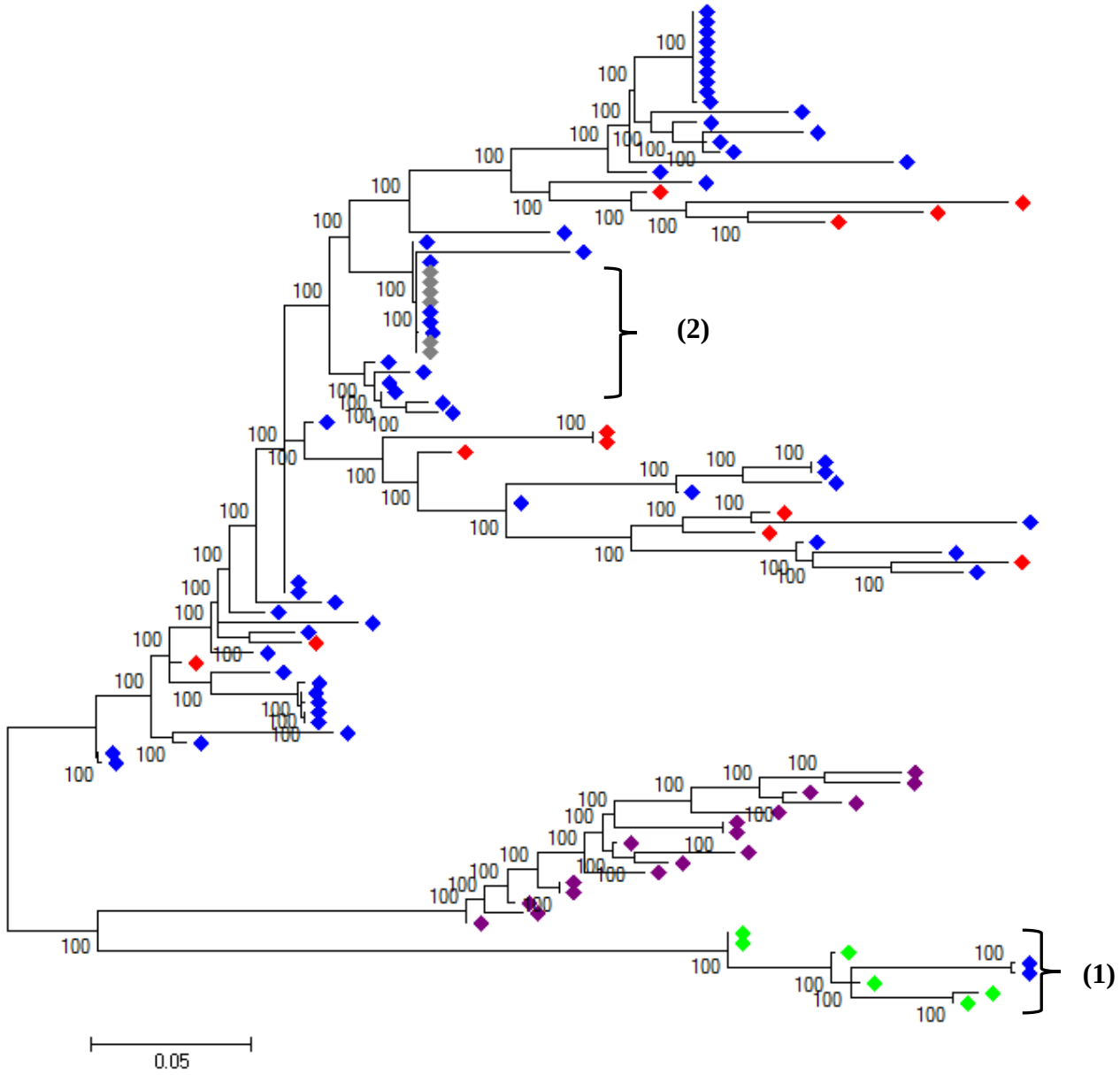


Figure S6.1. Haplotype based maximum likelihood tree for two patterns of introgression based on the 35-kb fragment size that extends from 18310001 to 18345001 bp on chromosome 2. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of green junglefowl haplotypes into Sri Lanka1 and 3. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

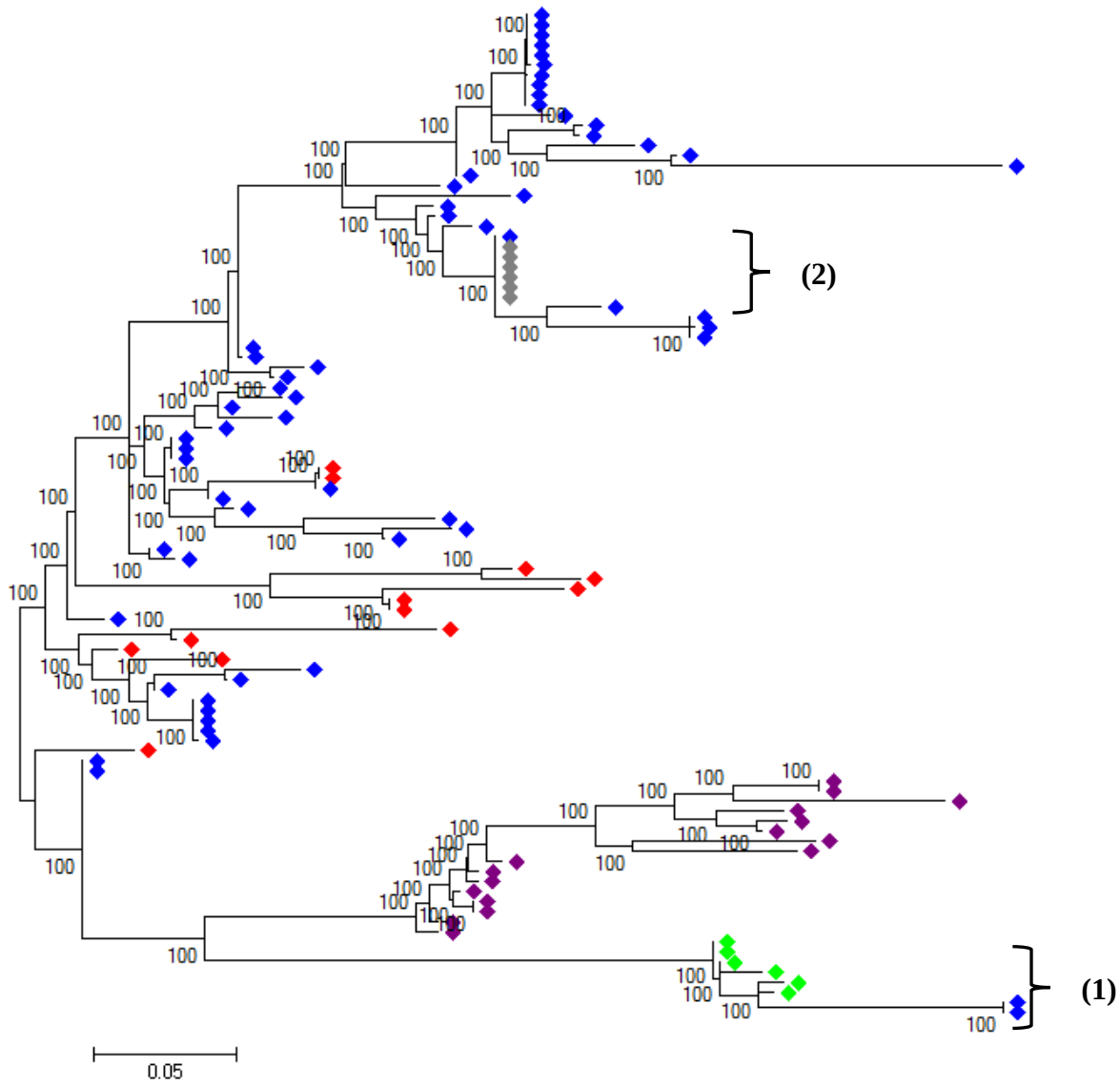


Figure S6.2. Haplotype based maximum likelihood tree for two patterns of introgression based on the 30-kb fragment size that extends from 18365001 to 18395001 bp on chromosome 2. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of a green junglefowl haplotype into Sri Lanka1 and 3. (**2**) Introgression of domestic/red junglefowl chicken haplotypes into grey junglefowl.

Figures S7. Haplotype relationships analysis for candidate regions of introgression from domestic chicken into grey junglefowl. ABBA – BABA test Grey junglefowl - domestic chicken.

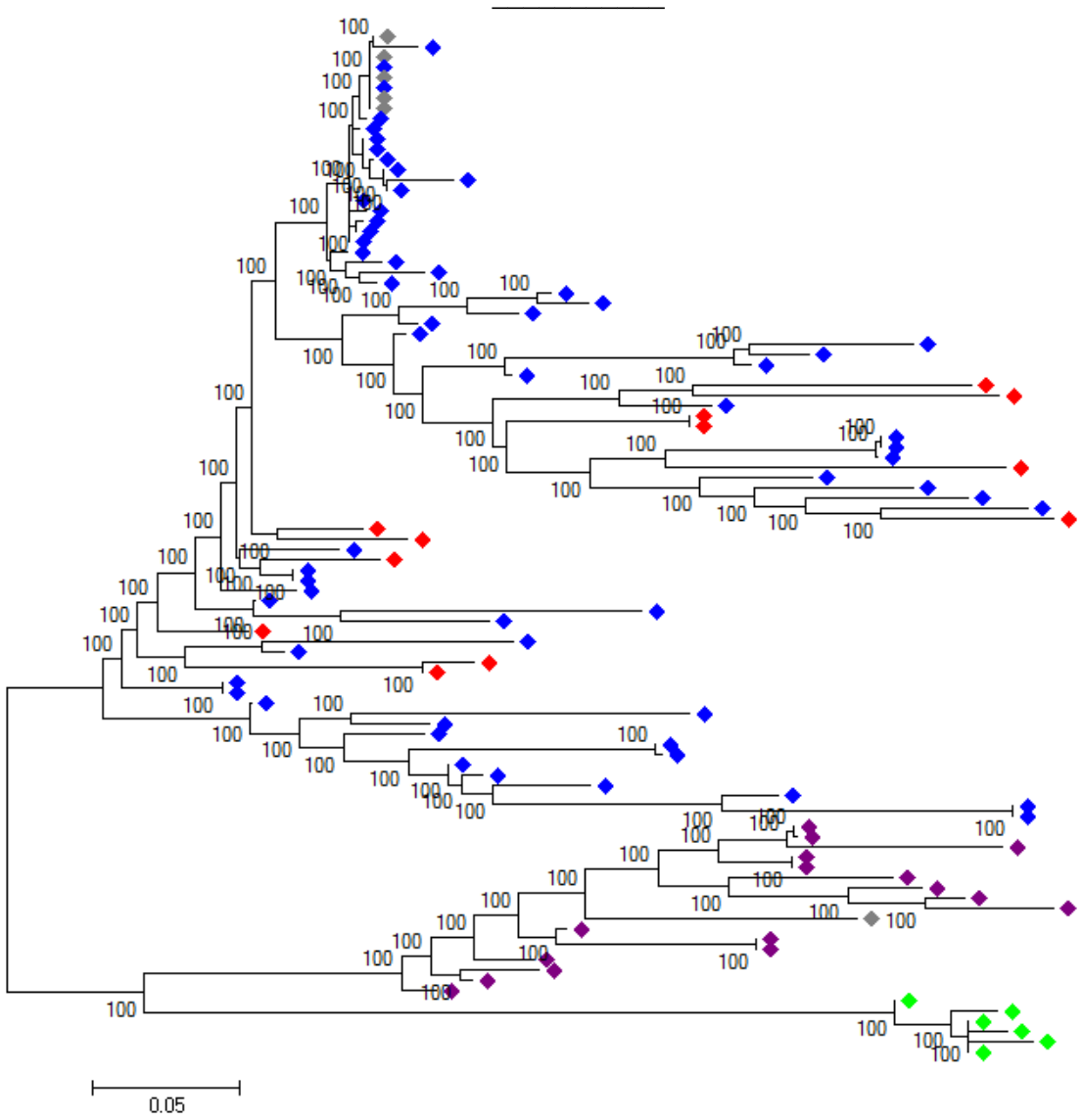


Figure S7.1. Chromosome 1: 14885001 - 14905001

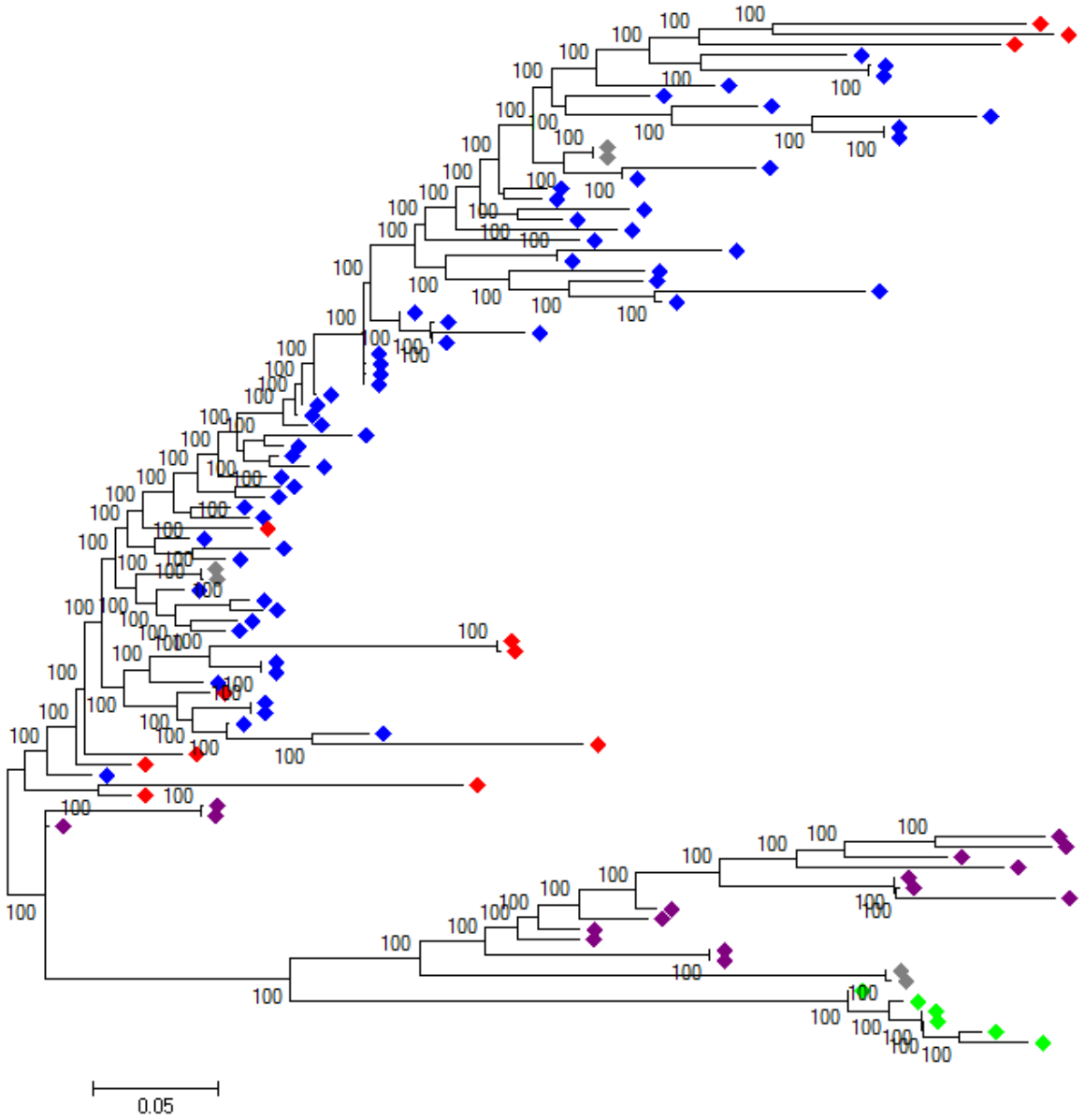


Figure S7.2. Chromosome 1: 15705001 - 15740001

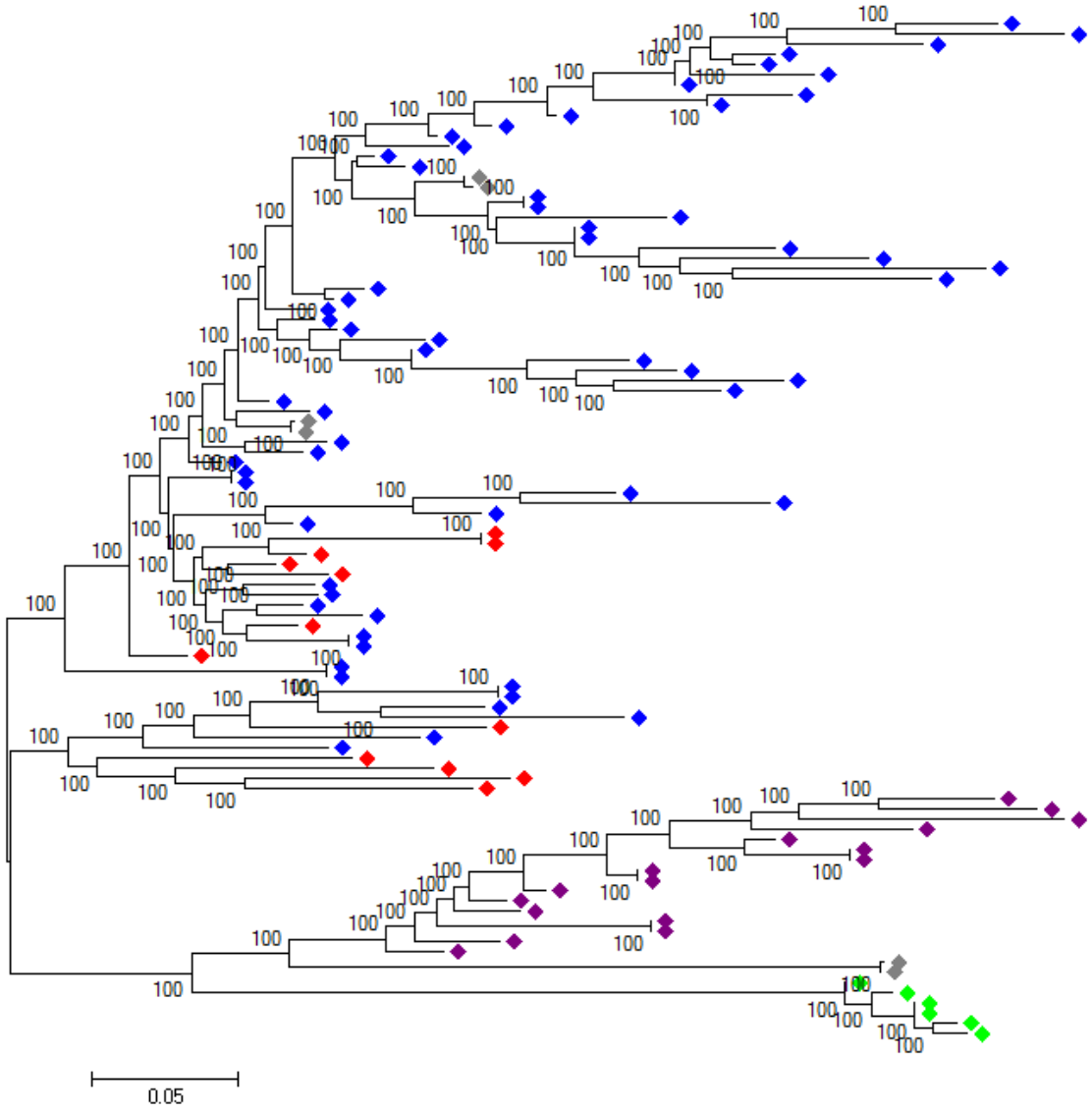


Figure S7.3 Chromosome 1: 15810001 – 15840001.

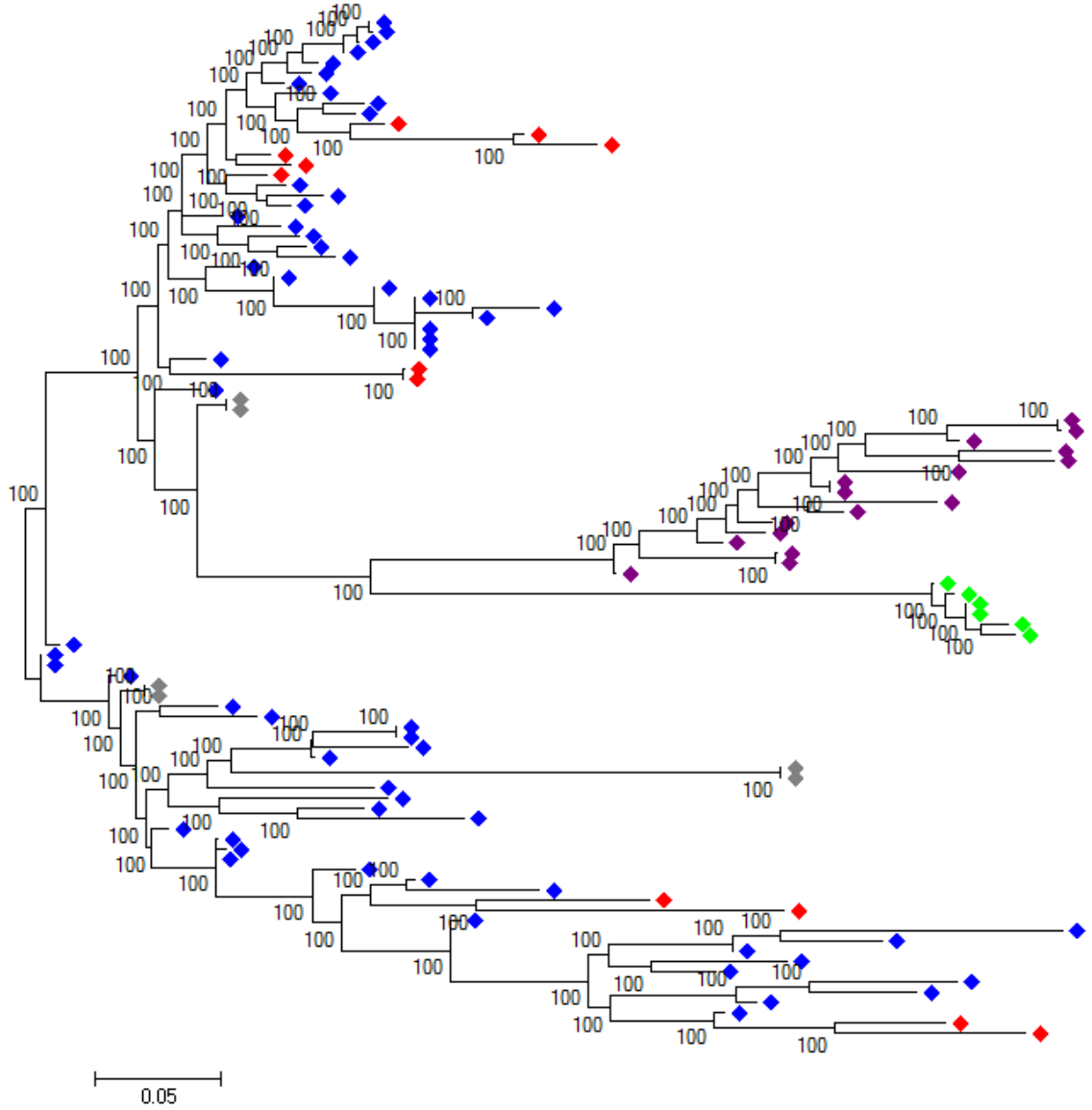


Figure S7.4 Chromosome 1:16105001 - 16135001

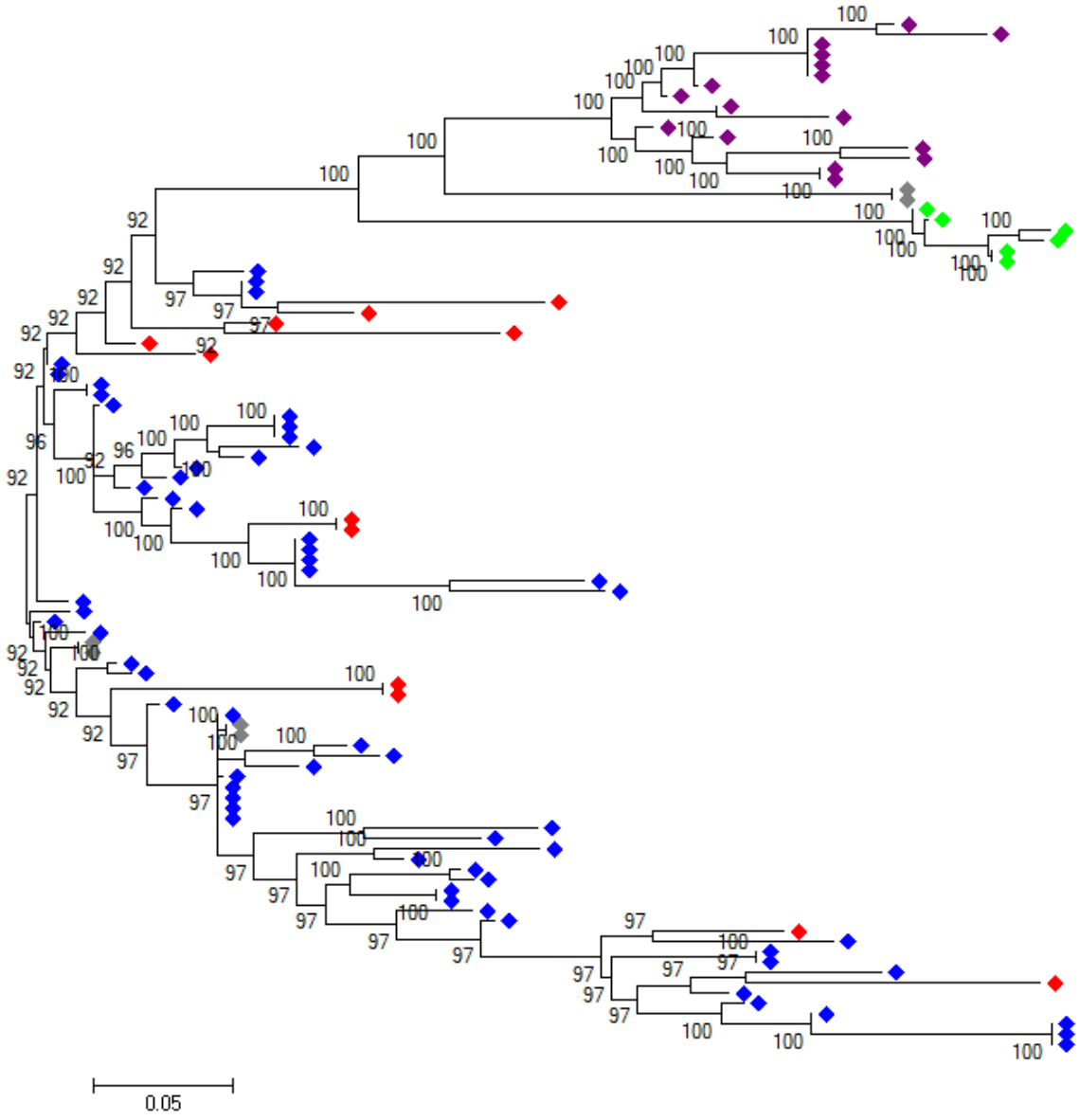


Figure S7.5 Chromosome 1:16195001 - 16215001

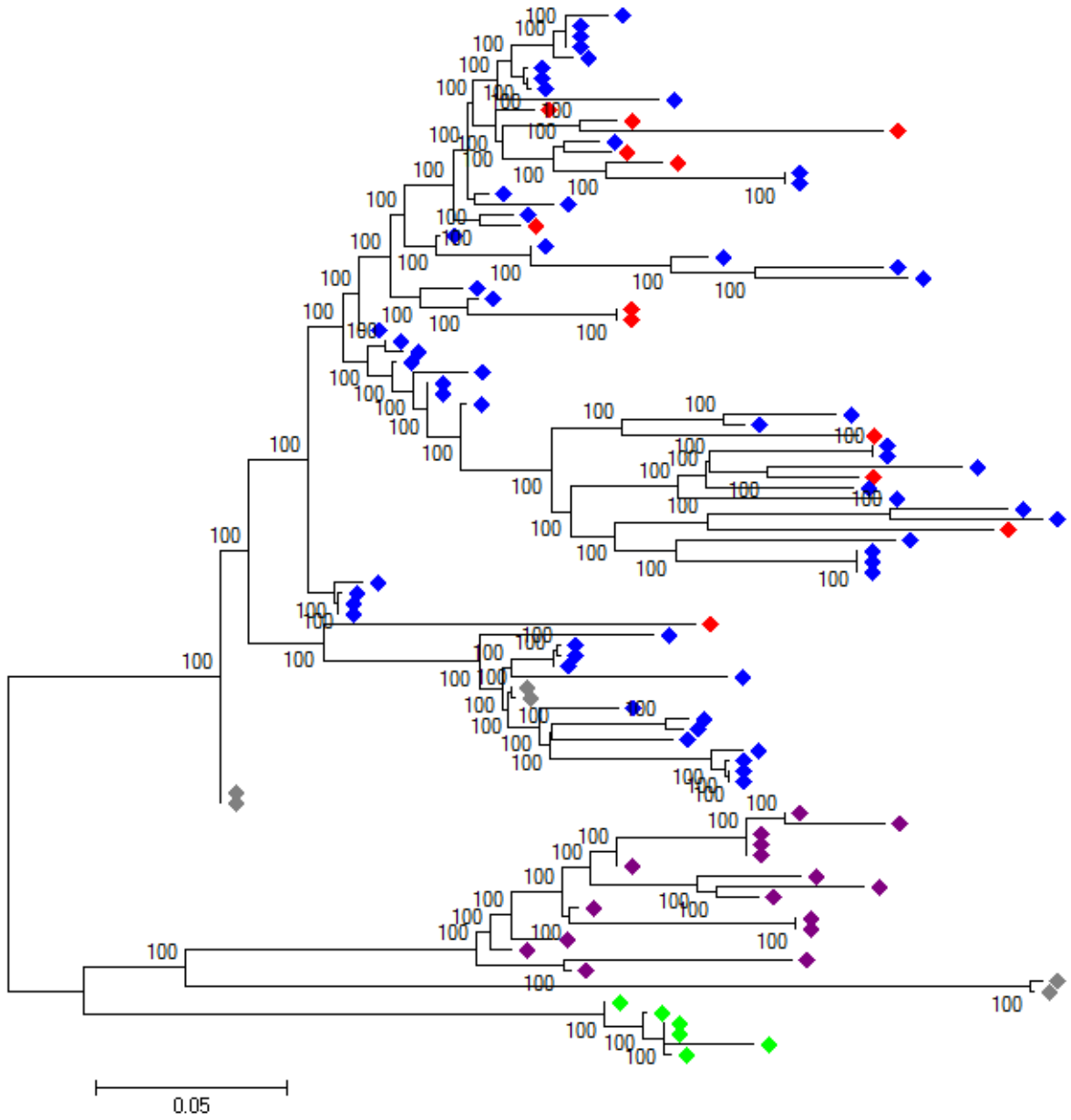


Figure S7.6 Chromosome 1: 16325001 - 16360001

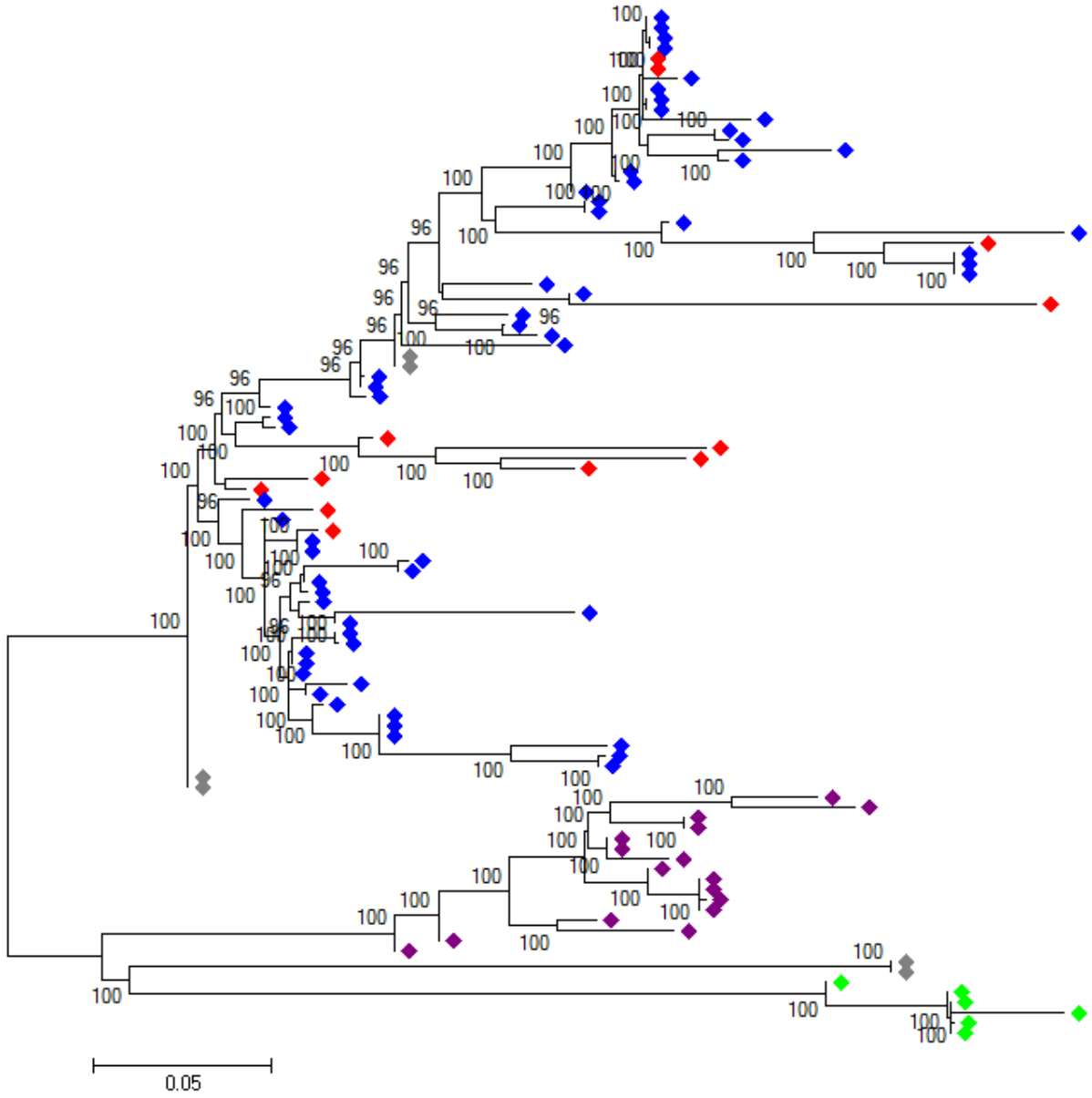


Figure S7.7 Chromosome 1: 16430001 - 16460001

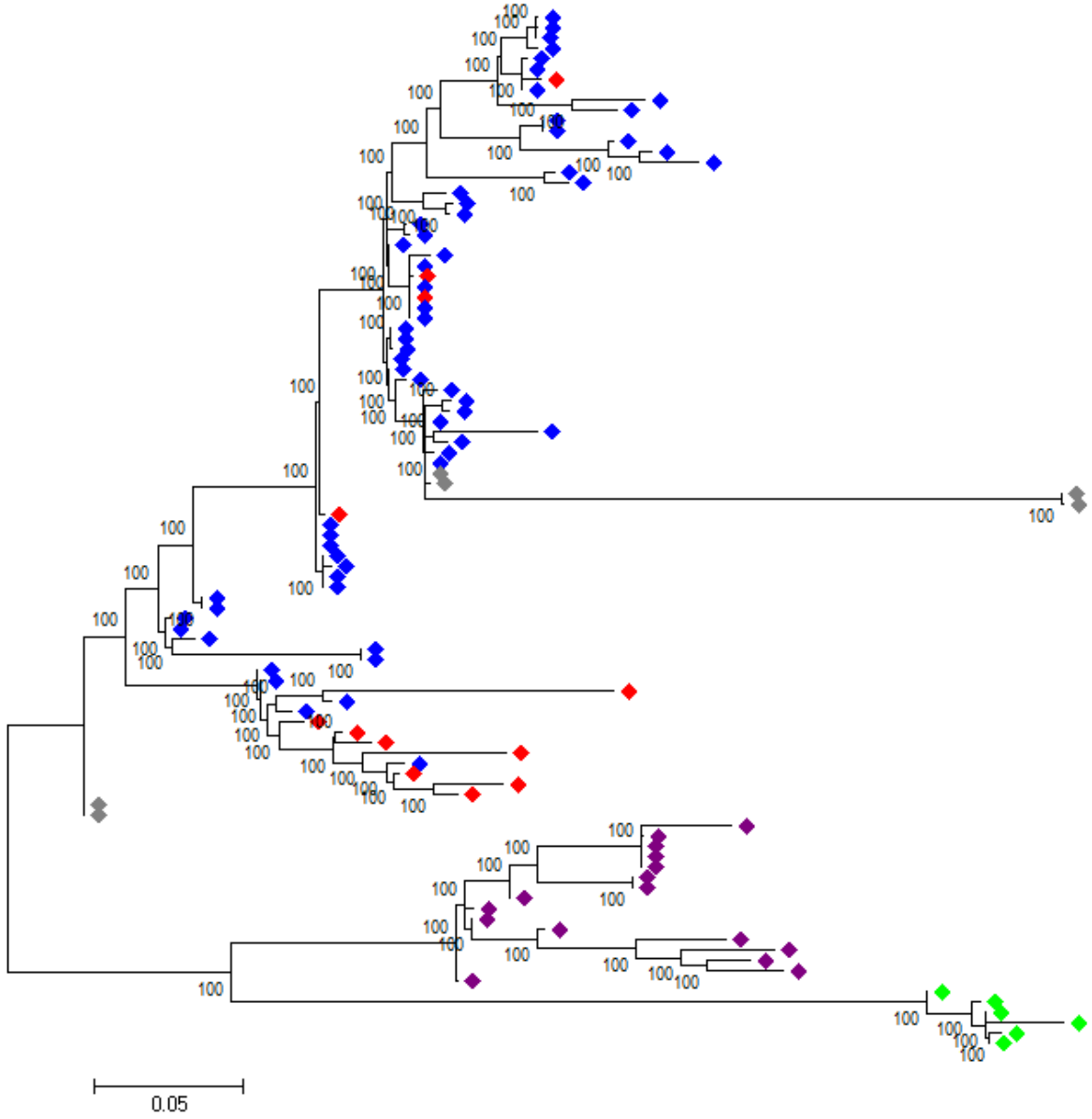


Figure S7.8 Chromosome 1: 16525001 - 16560001

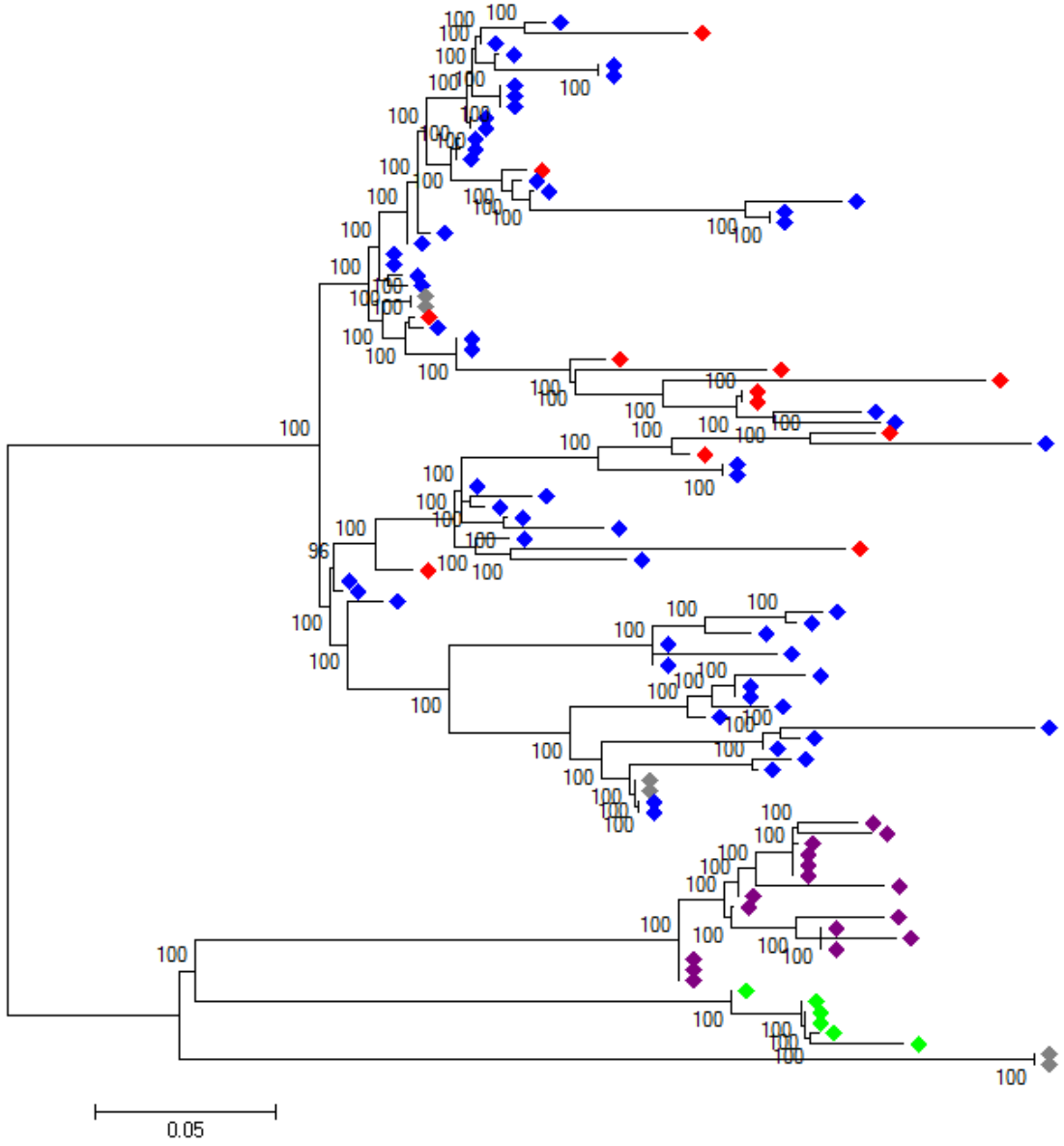


Figure S7.9 Chromosome 1: 17175001 - 17205001

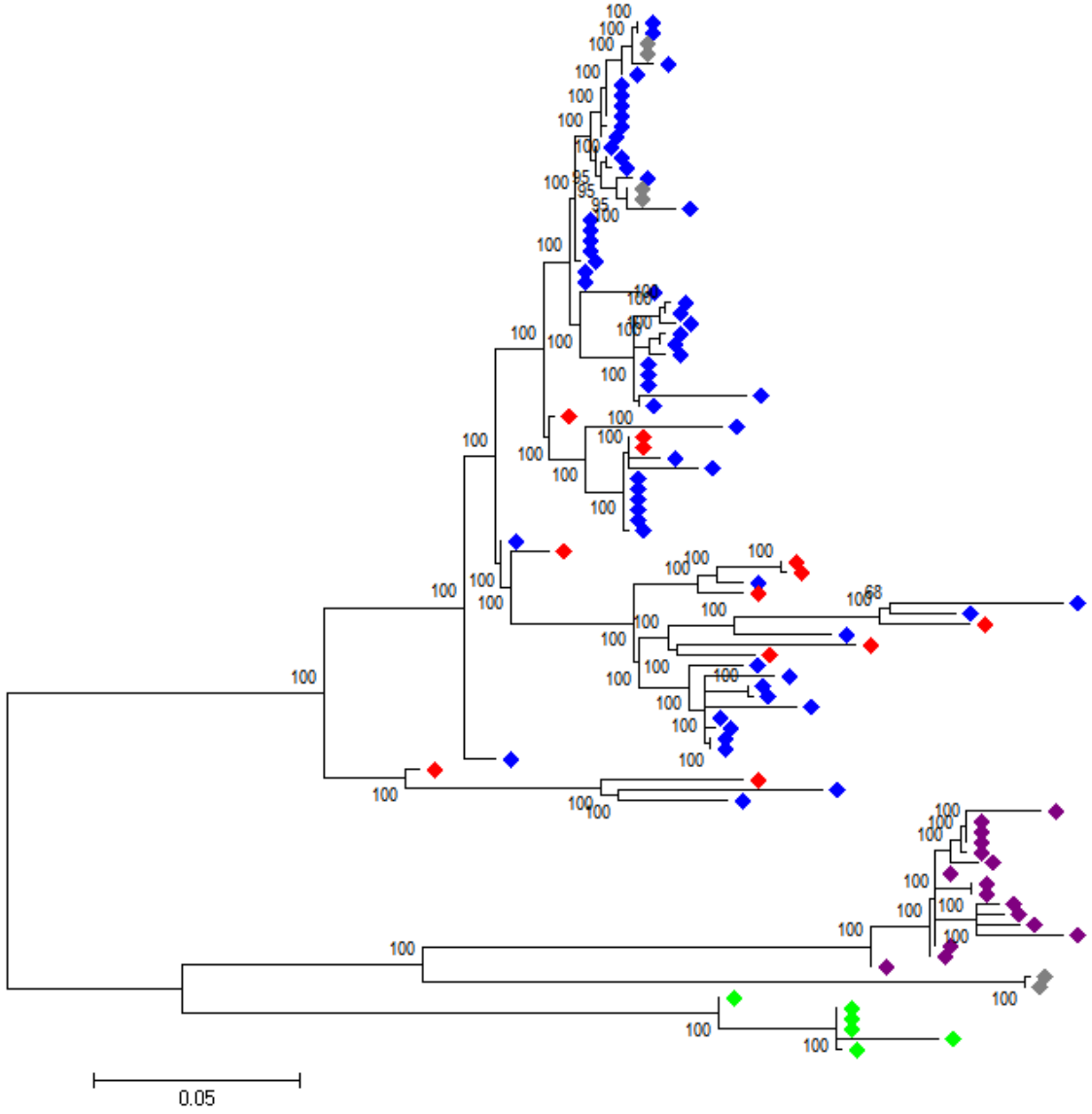


Figure S7.10 Chromosome 1: 18090001 – 18120001

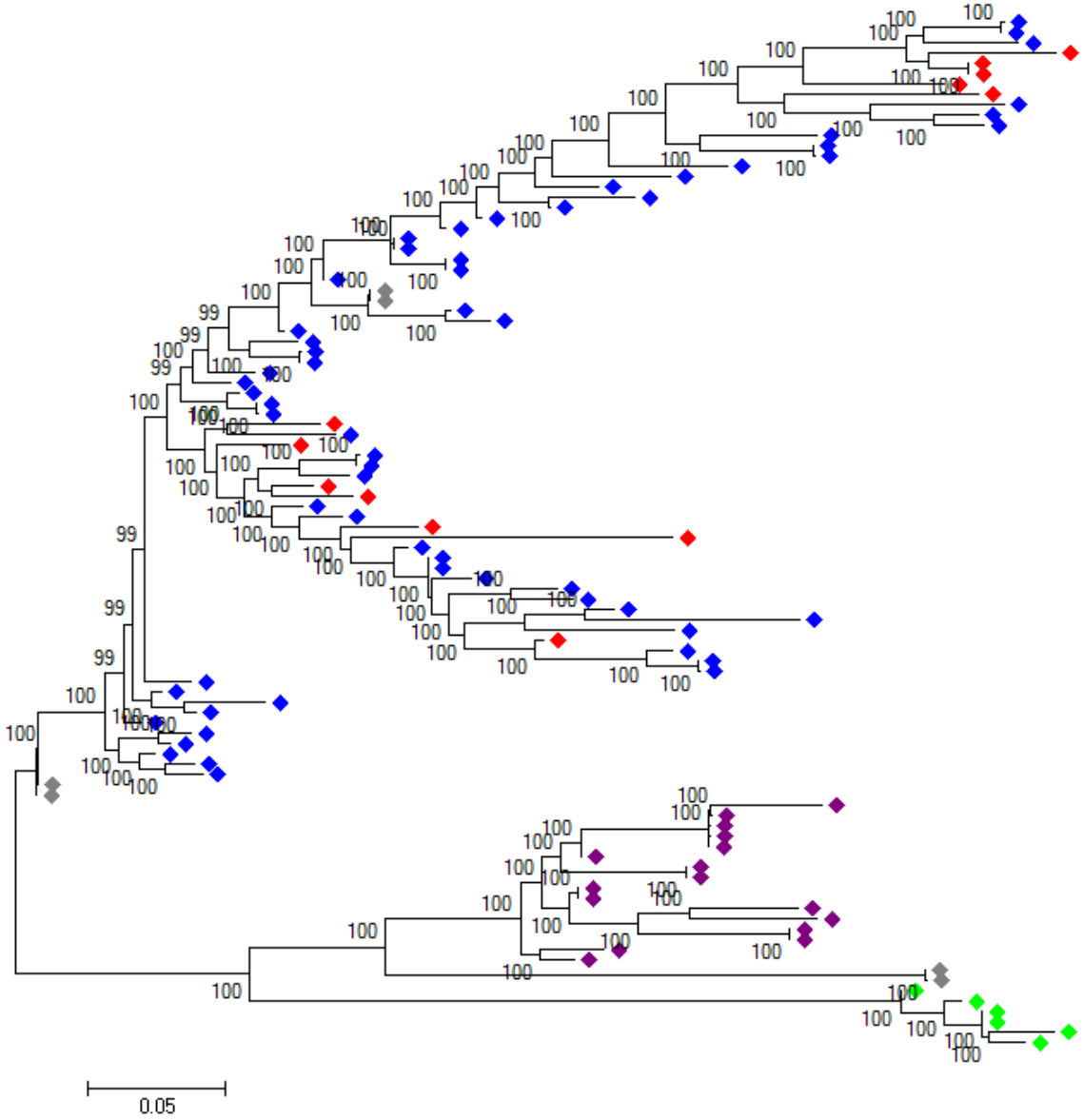


Figure S7.11 Chromosome 1: 18555001 - 18635001

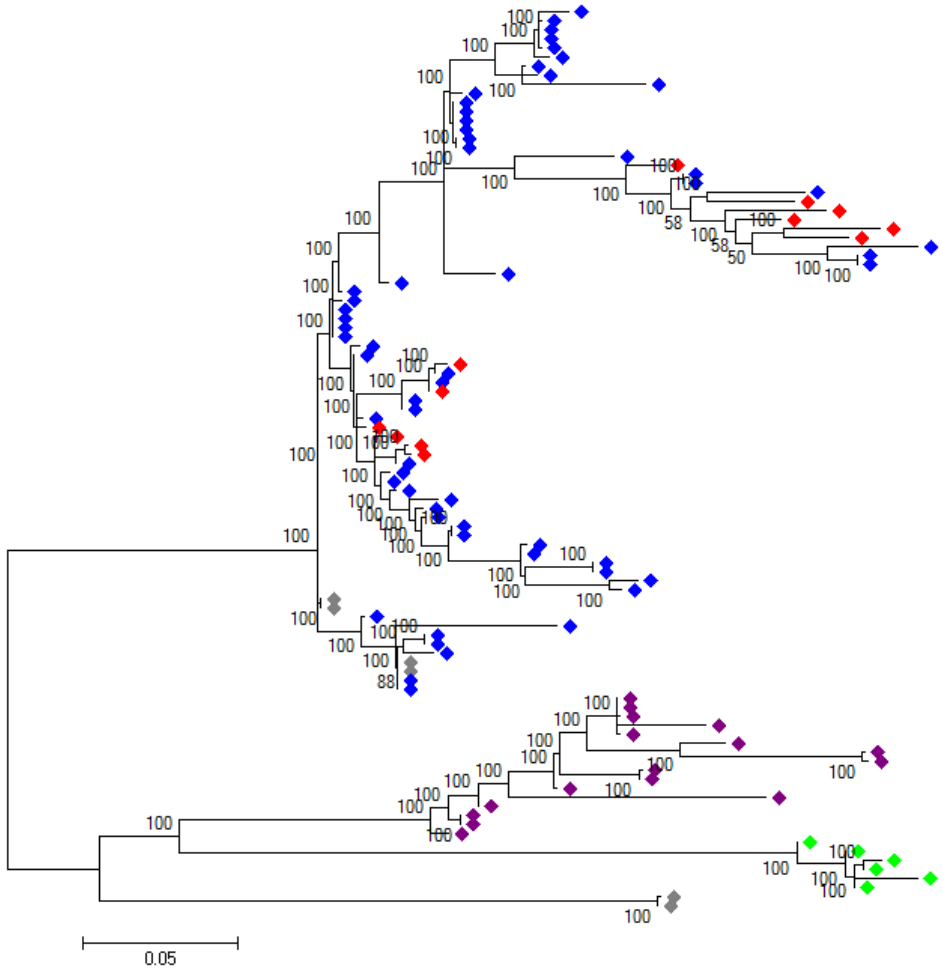


Figure S7.12 Chromosome 1: 19060001 - 19100001

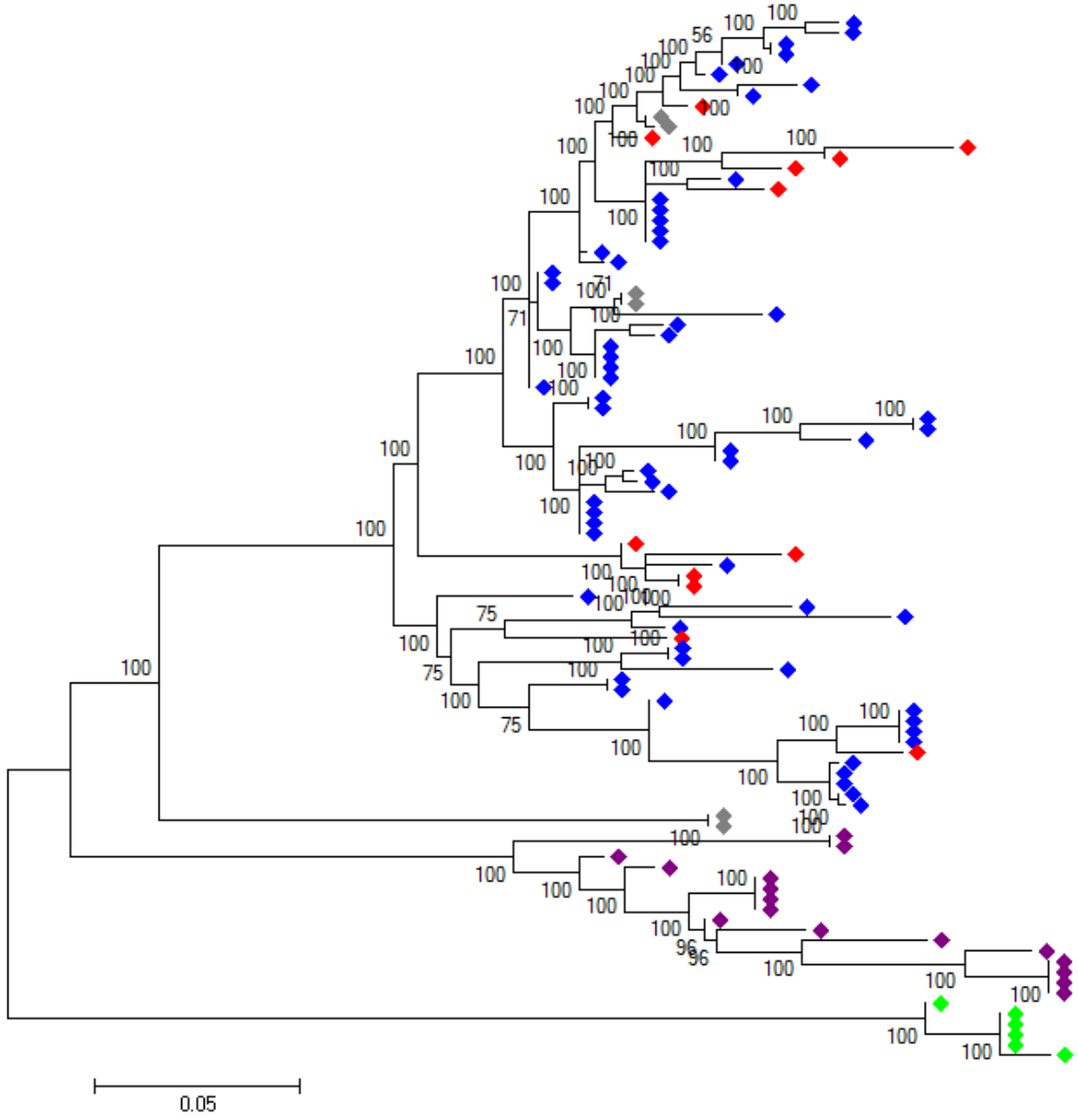


Figure S7.13 Chromosome 1: 18900001 - 18920001

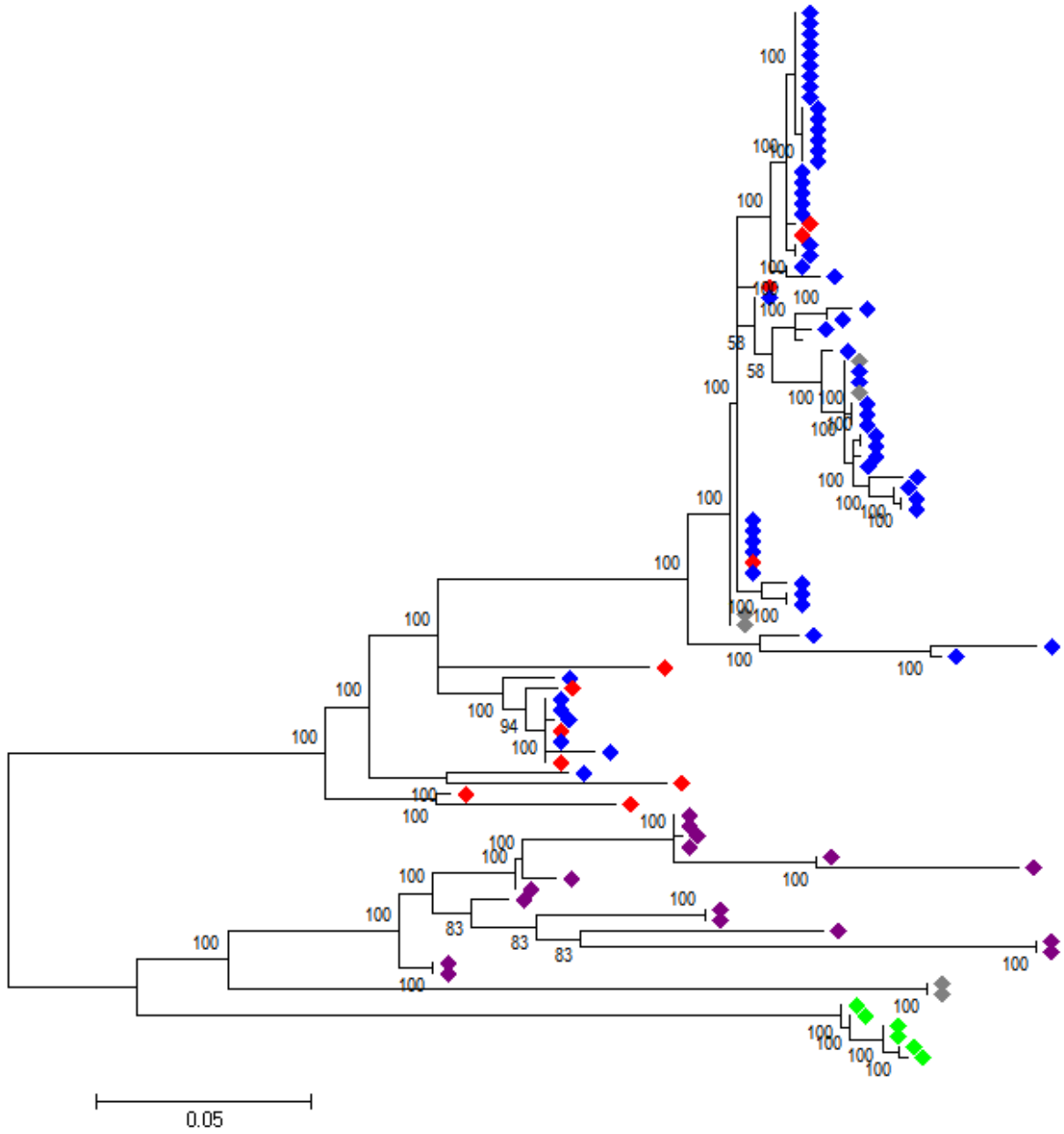


Figure S7.14 Chromosome 1: 19175001 - 19195001

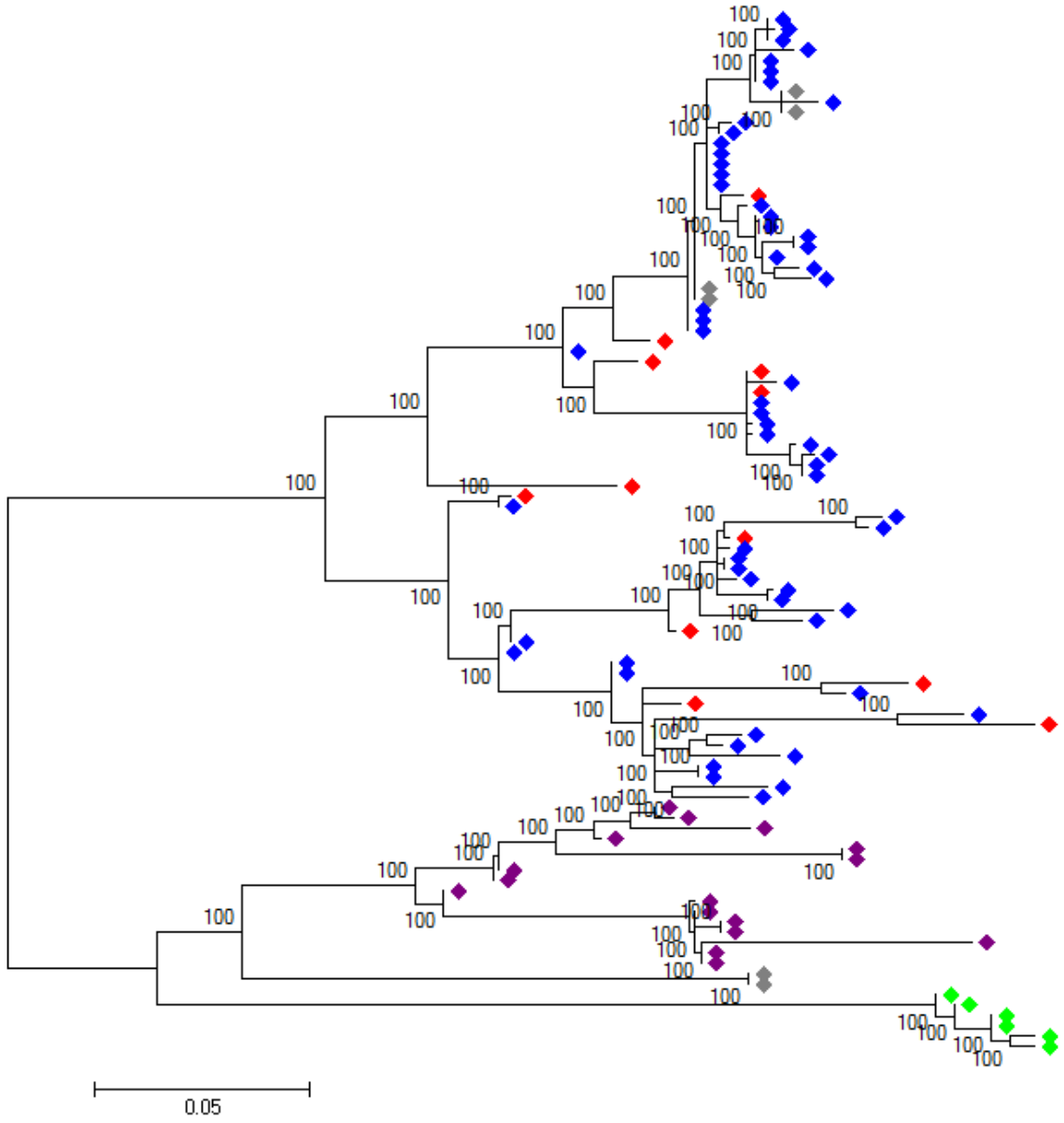


Figure S7.15 Chromosome 1: 19280001 - 19305001

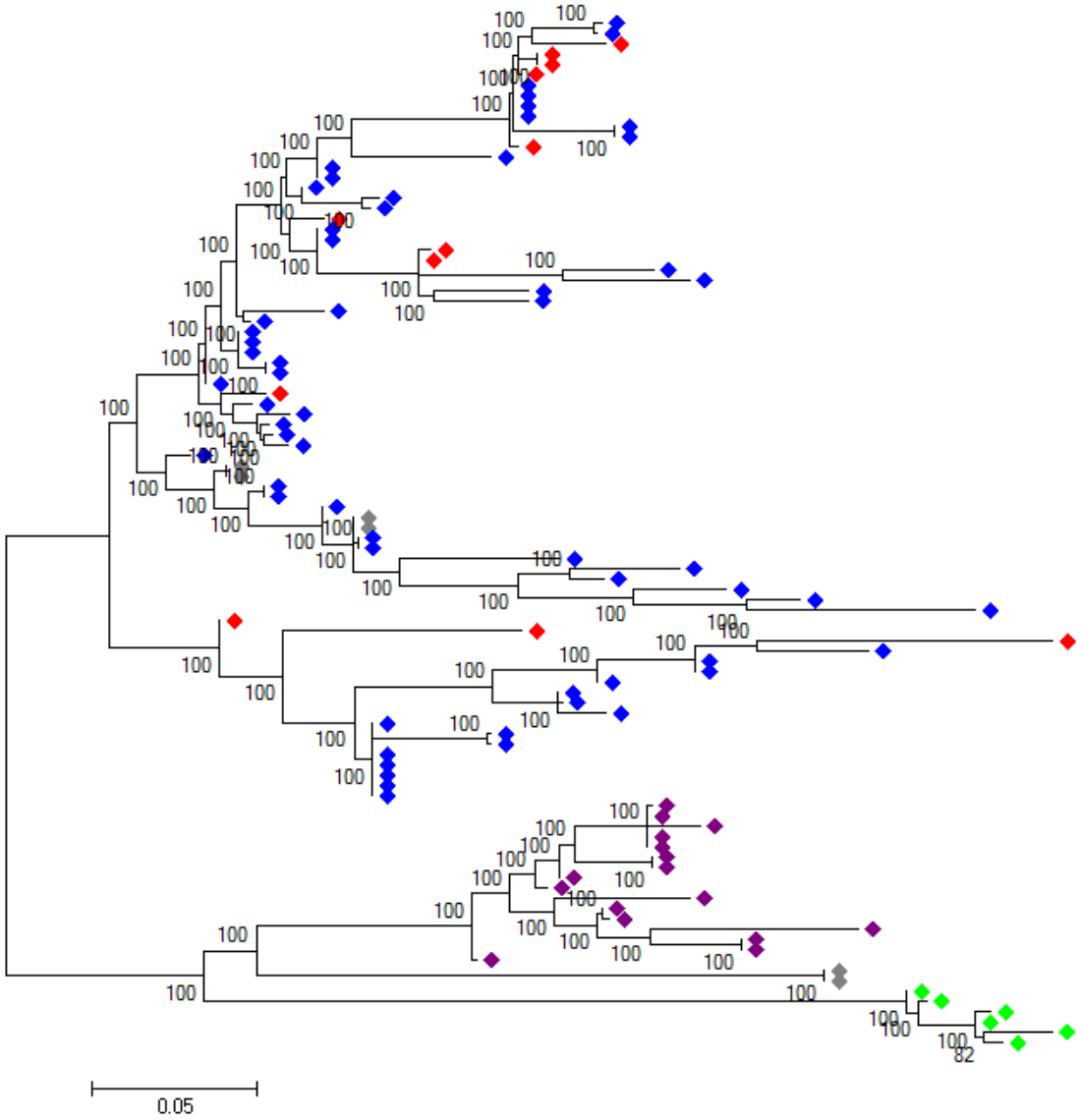


Figure S7.16 Chromosome 1:19540001 - 19565001

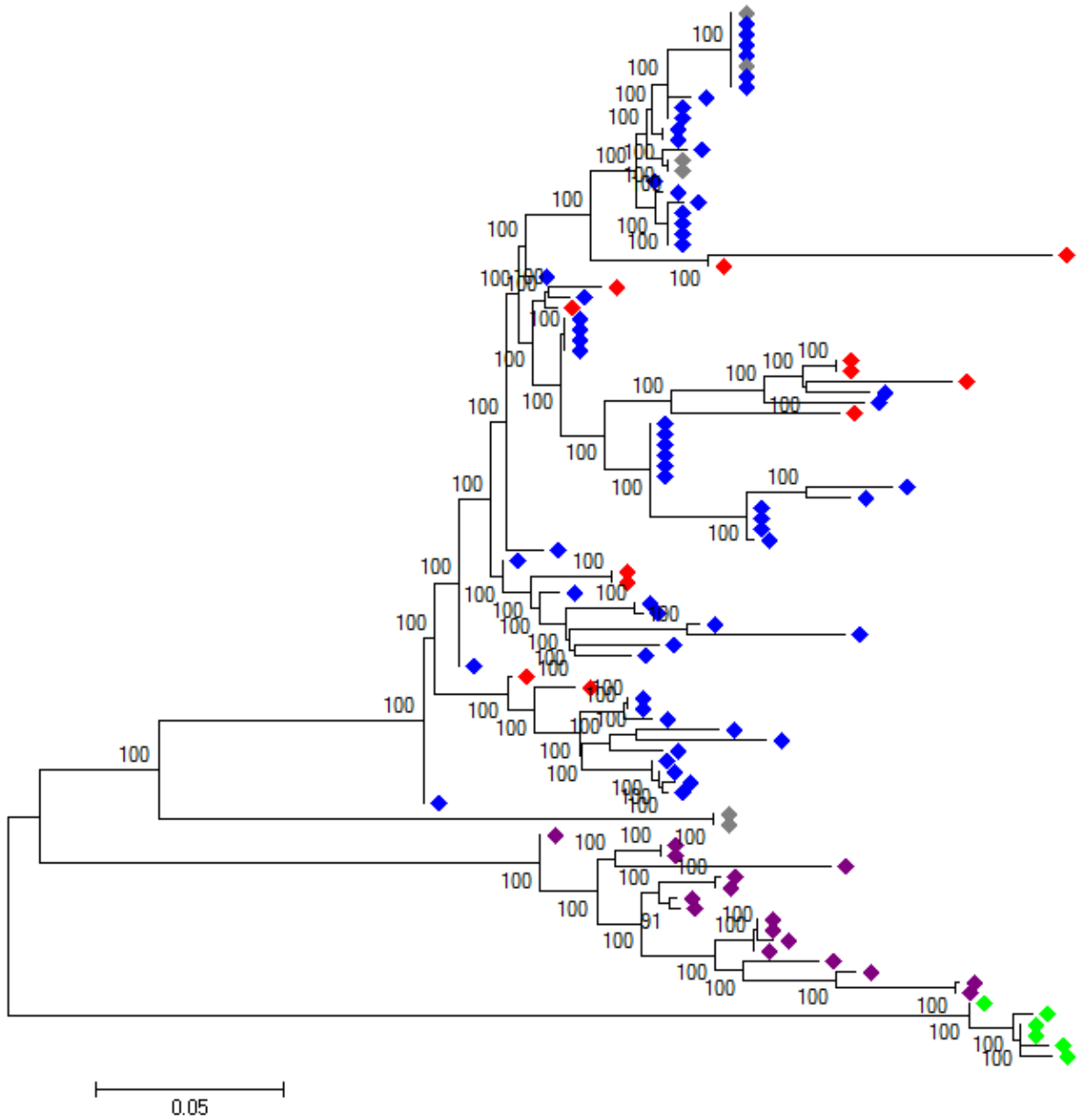


Figure S7.17 Chromosome 1: 19595001 - 19625001

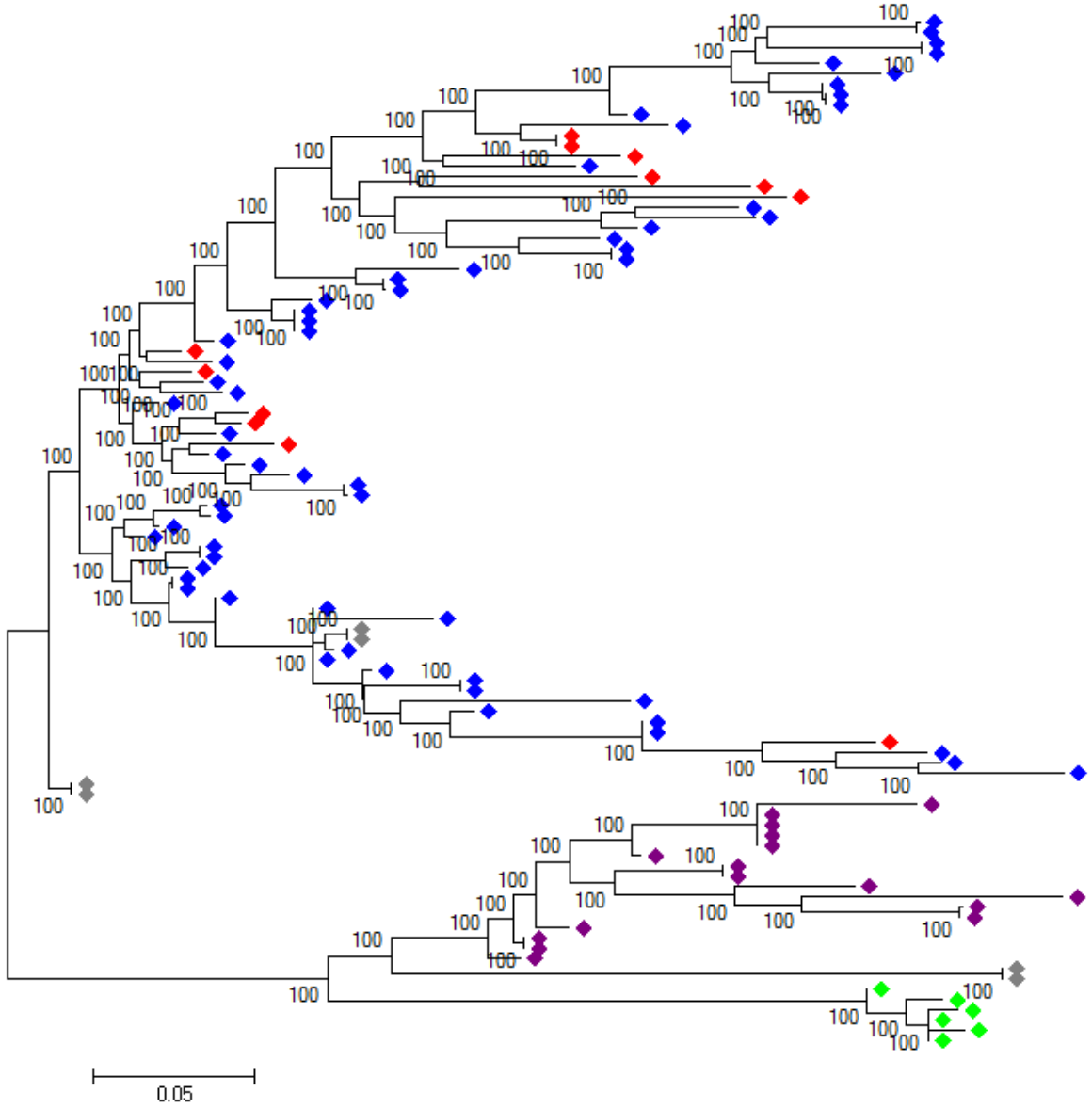


Figure S7.18 Chromosome 1**:**19710001 - 19740001

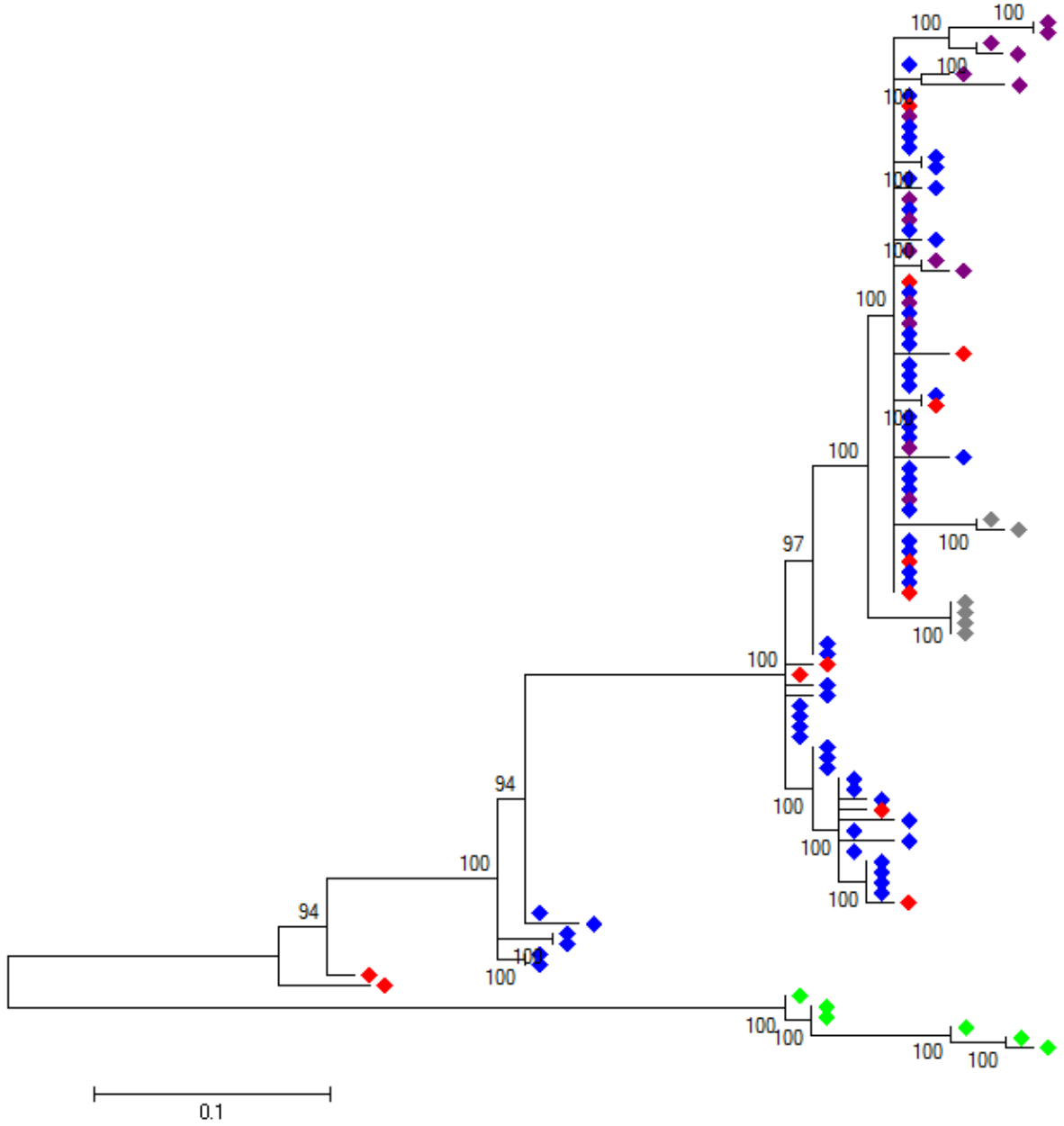


Figure S7.19 Chromosome 1: 88910001 - 88930001

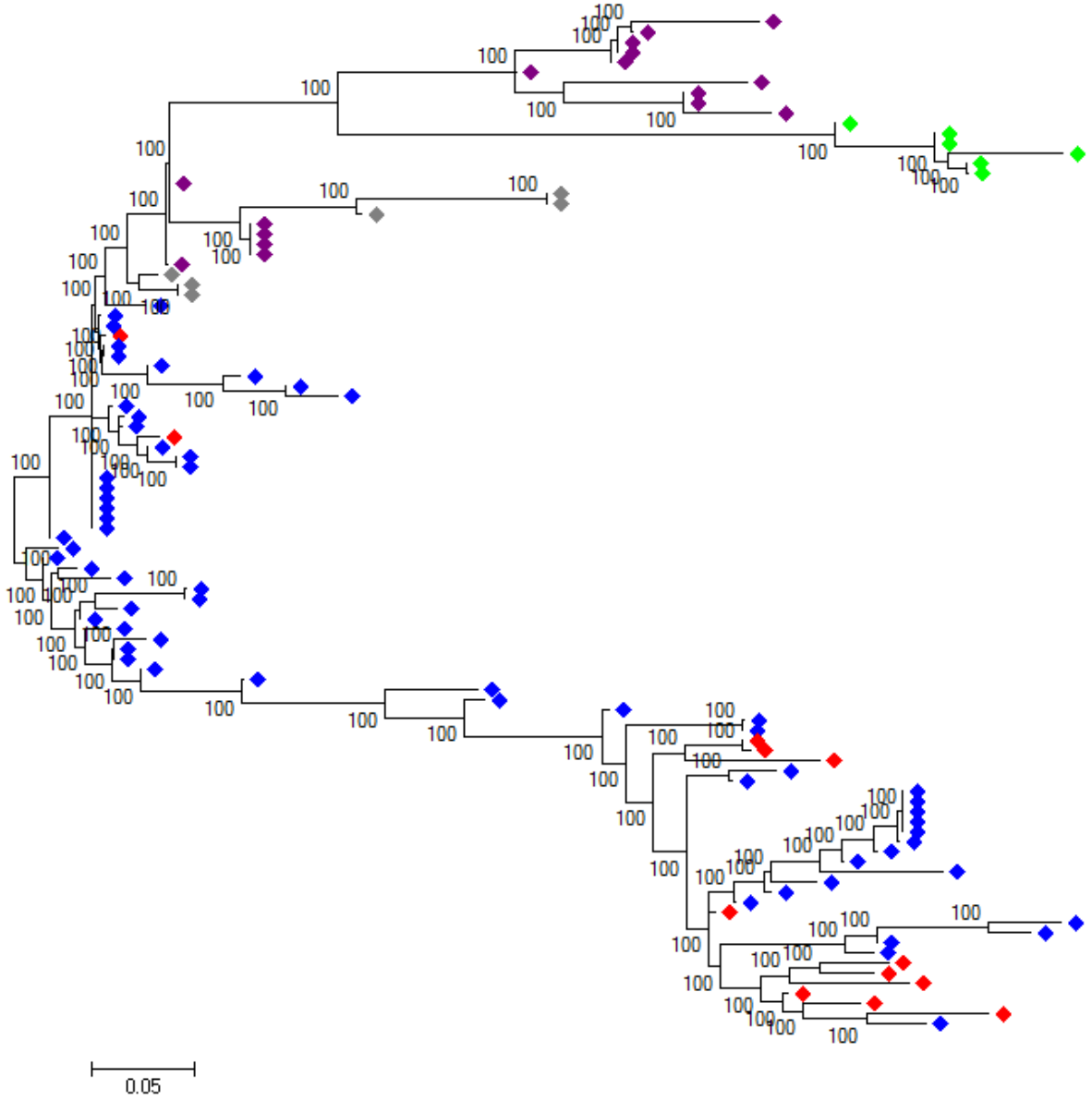


Figure S7.20 Chromosome 1: 100465001 - 100515001

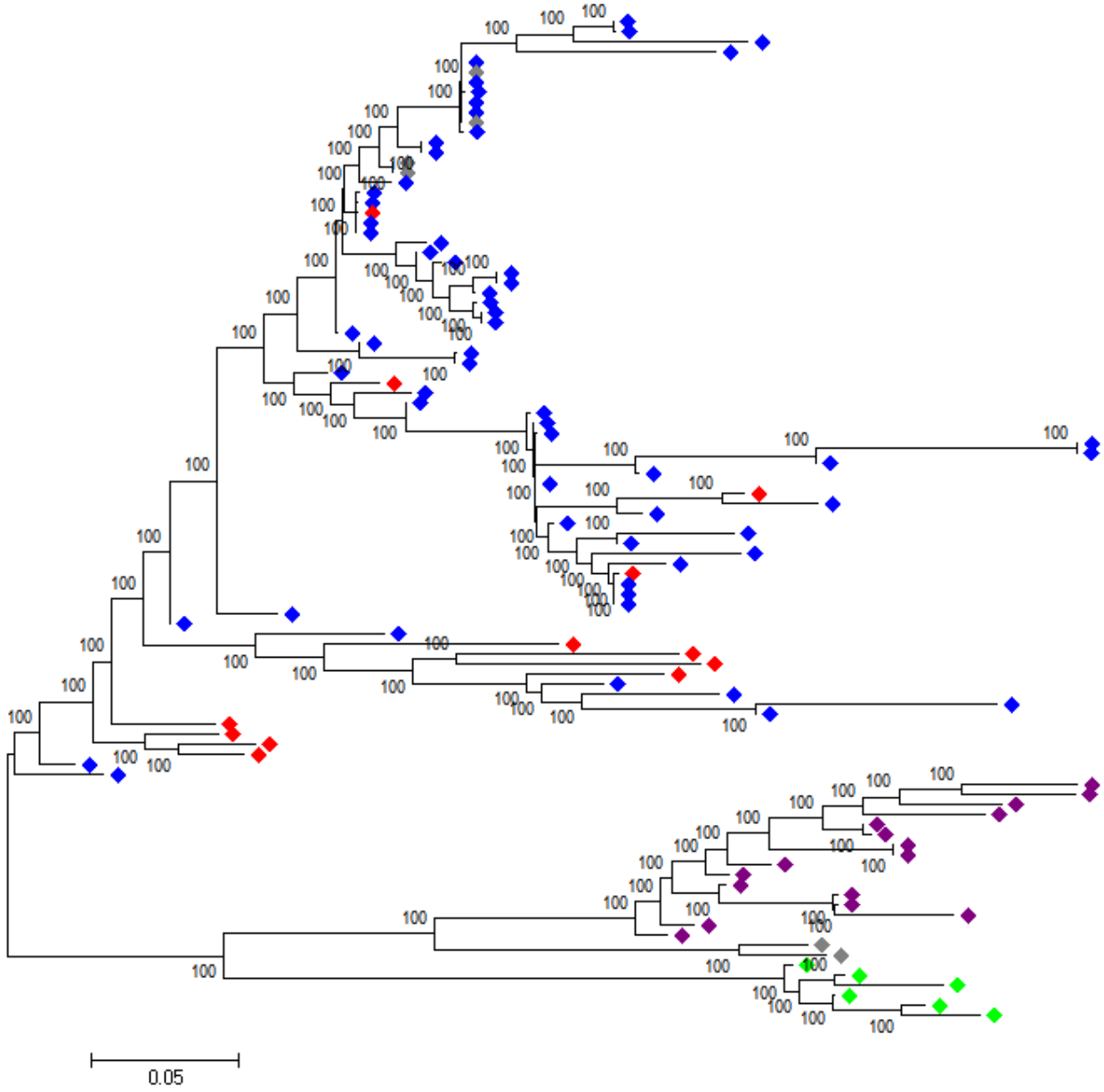


Figure S7.21 Chromosome 1: 138815001 - 138835001

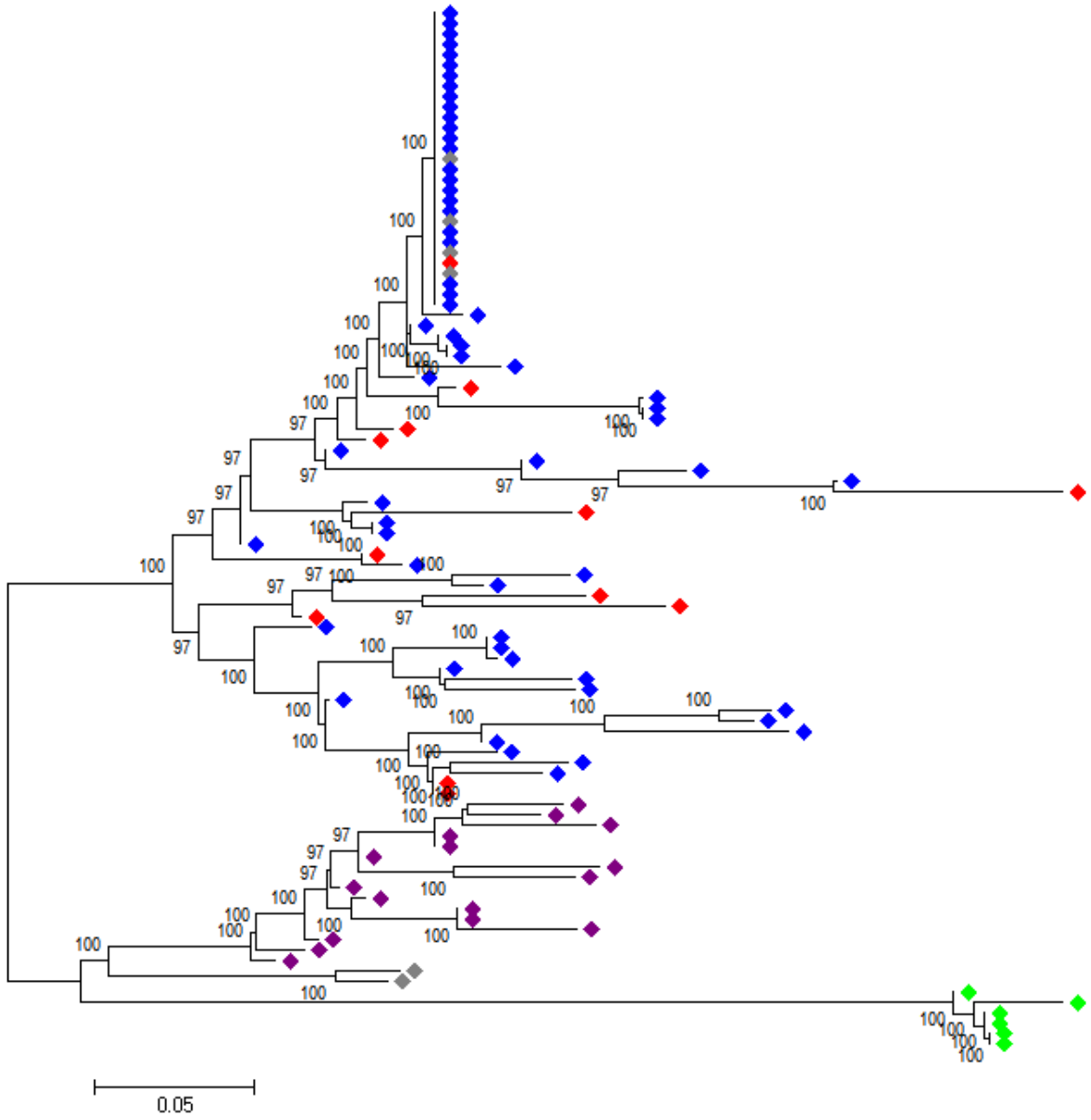


Figure S7.22 Chromosome 1: 139045001 - 139070001

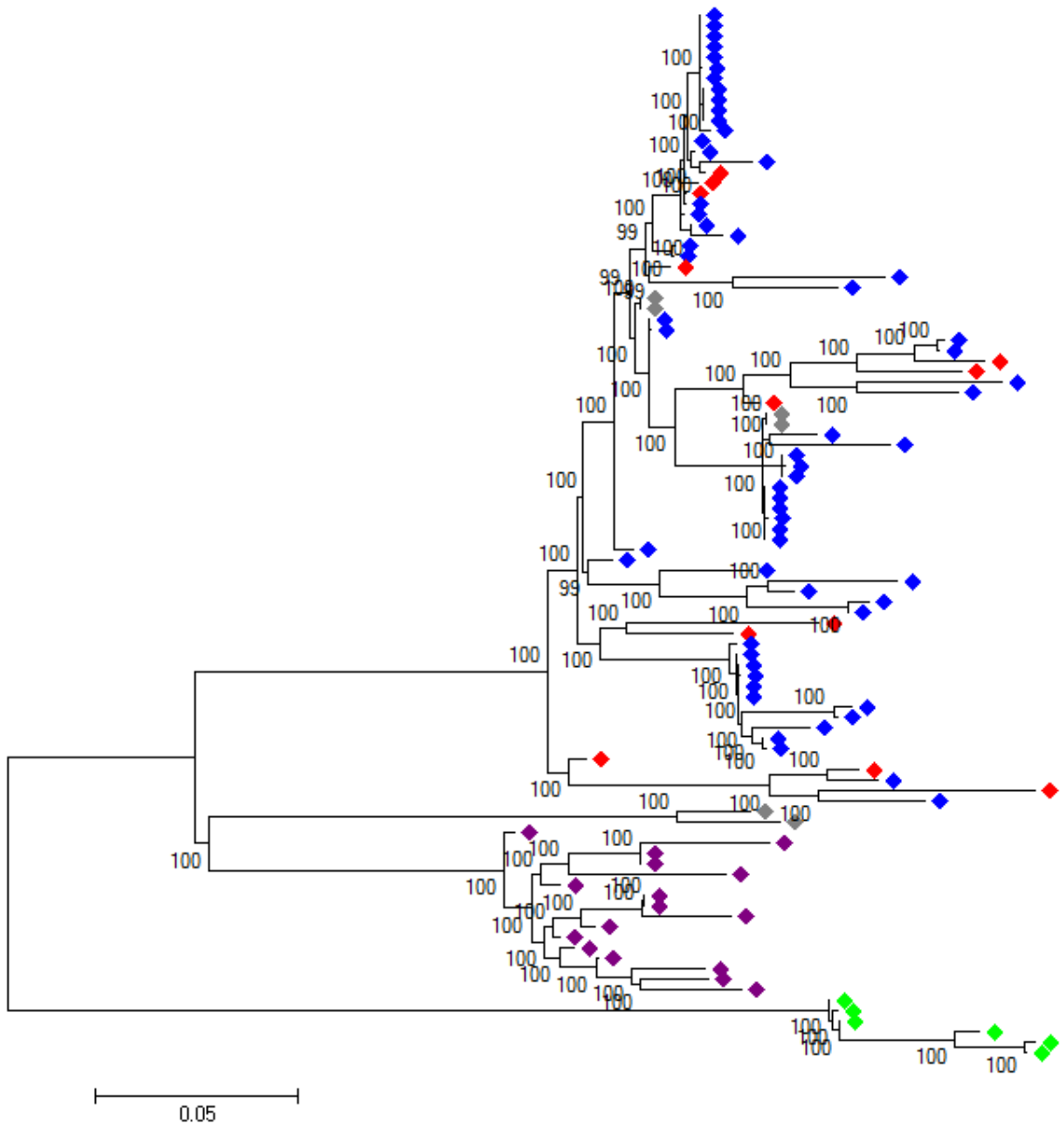


Figure S7.23 Chromosome 1:139620001 - 139685001

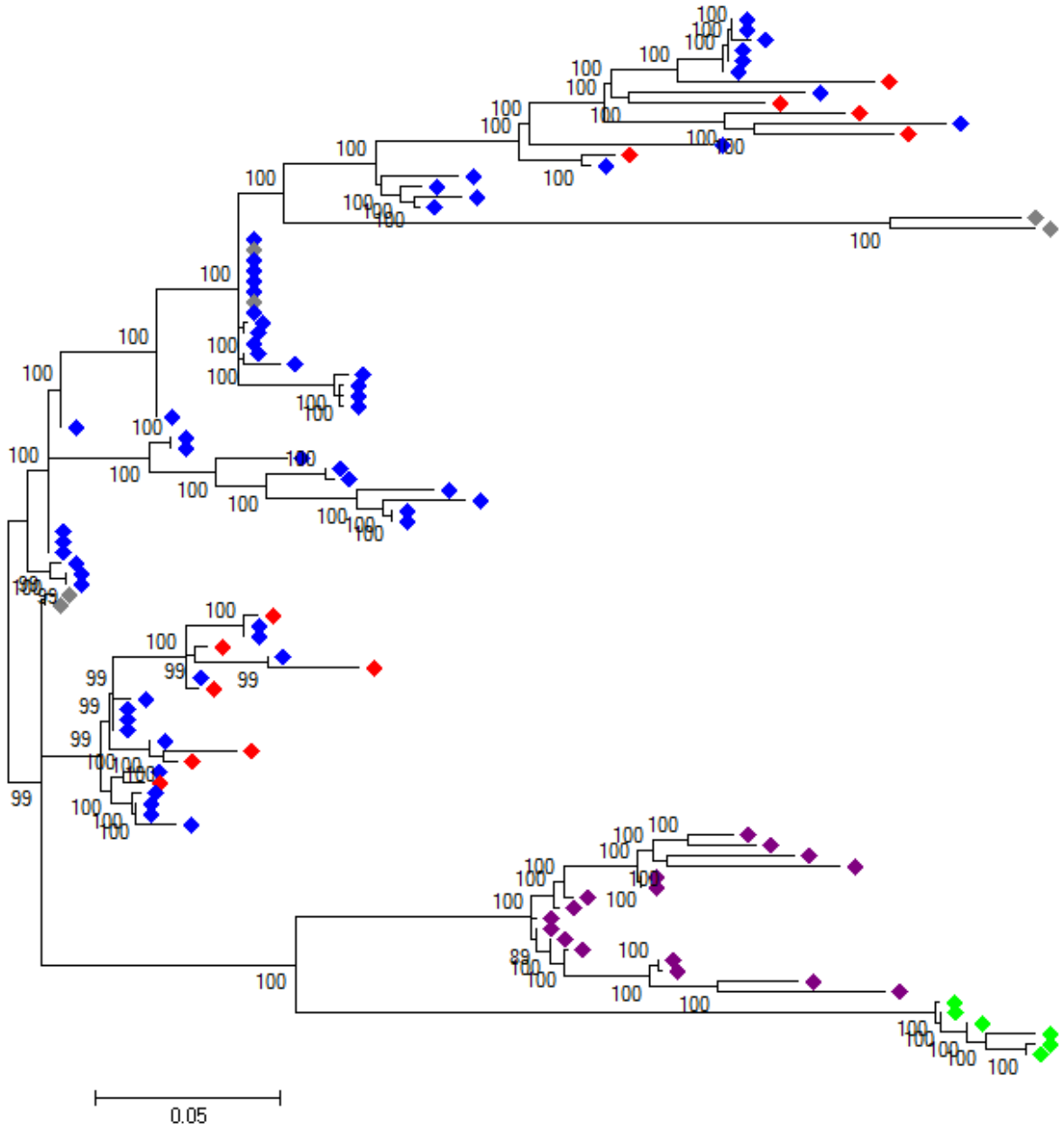


Figure S7.24 Chromosome 1: 139785001 - 139810001

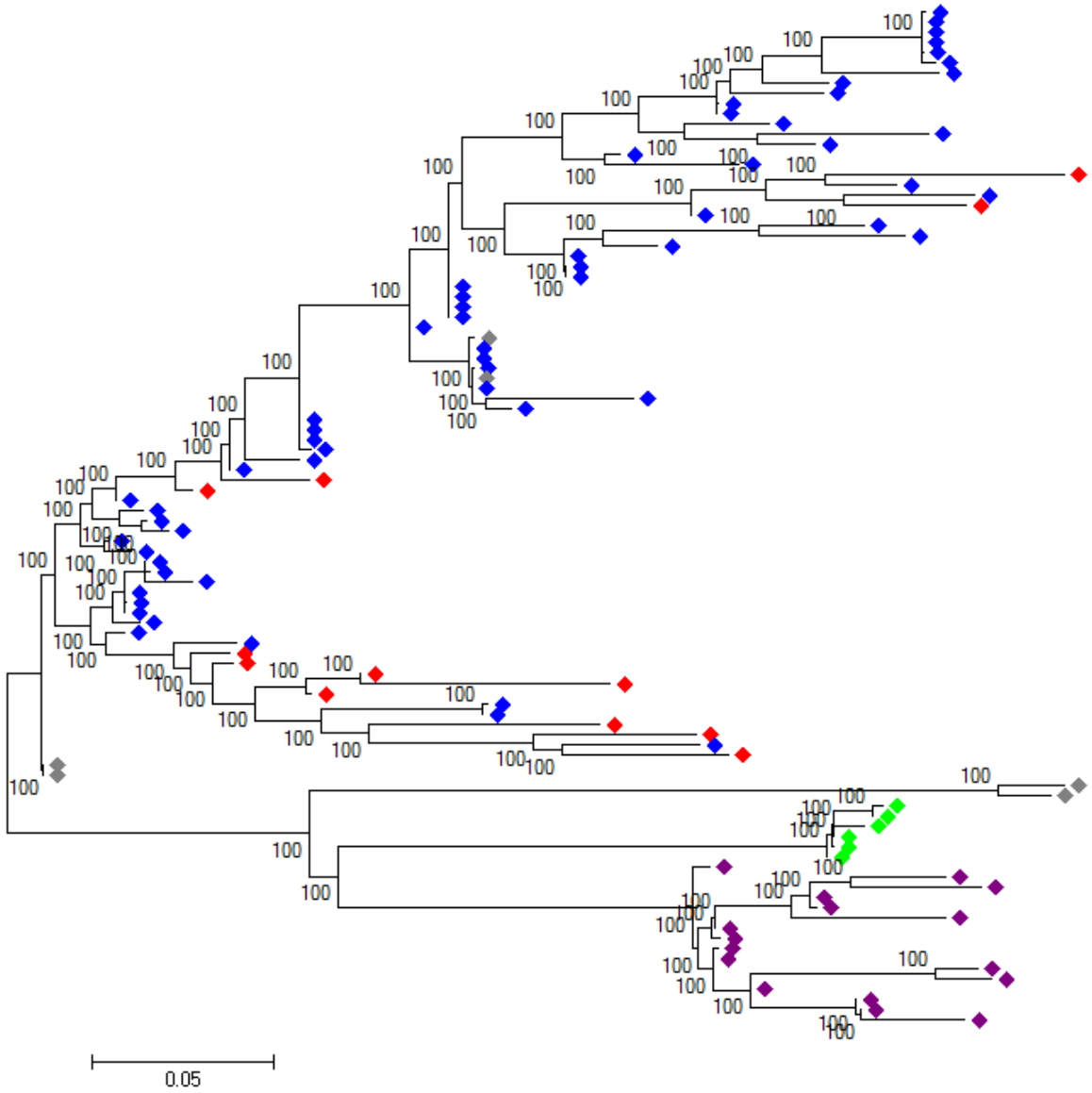


Figure S7.25 Chromosome 1: 139865001 - 139910001

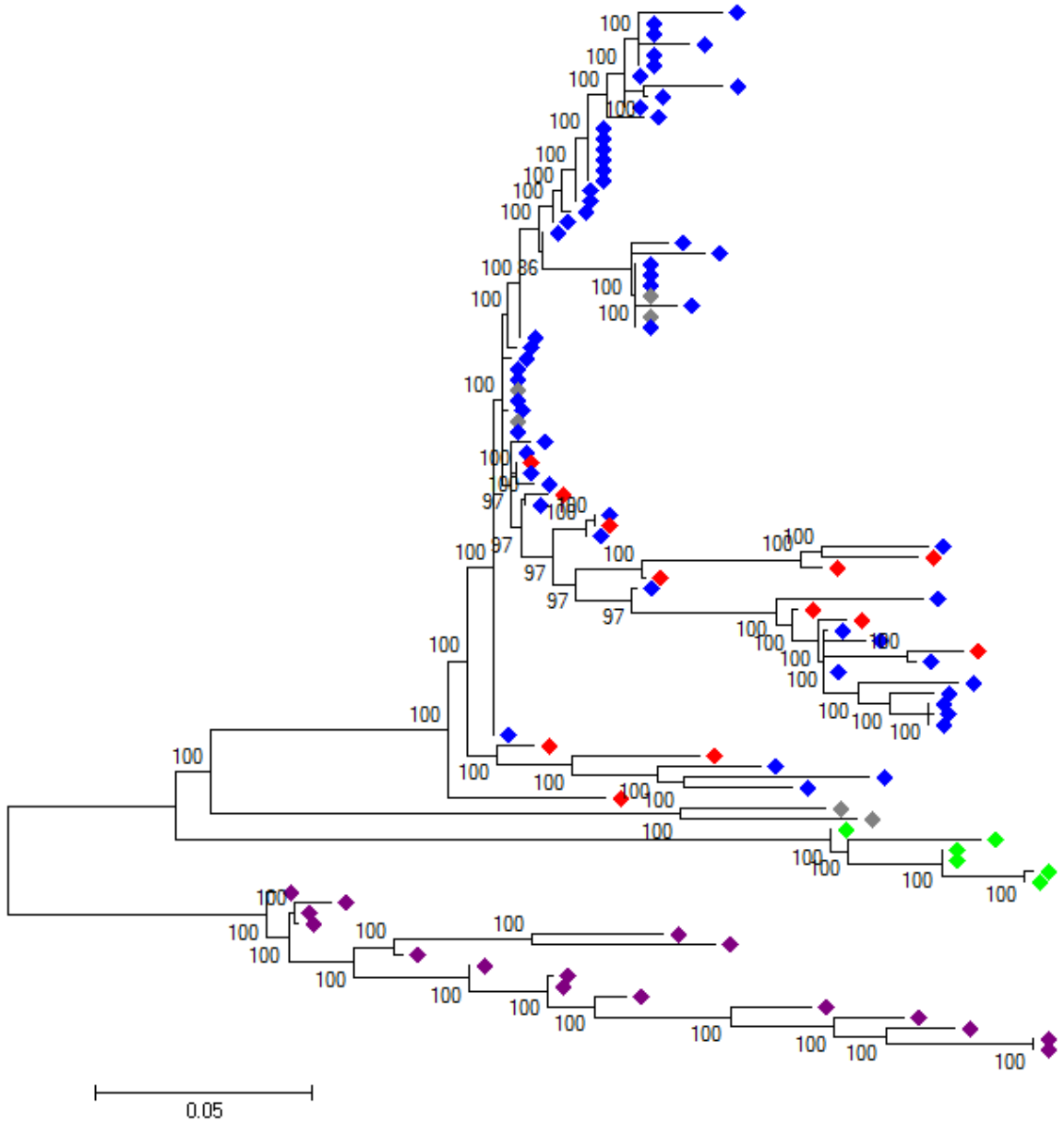


Figure S7.26 Chromosome 1: 139960001 - 139990001

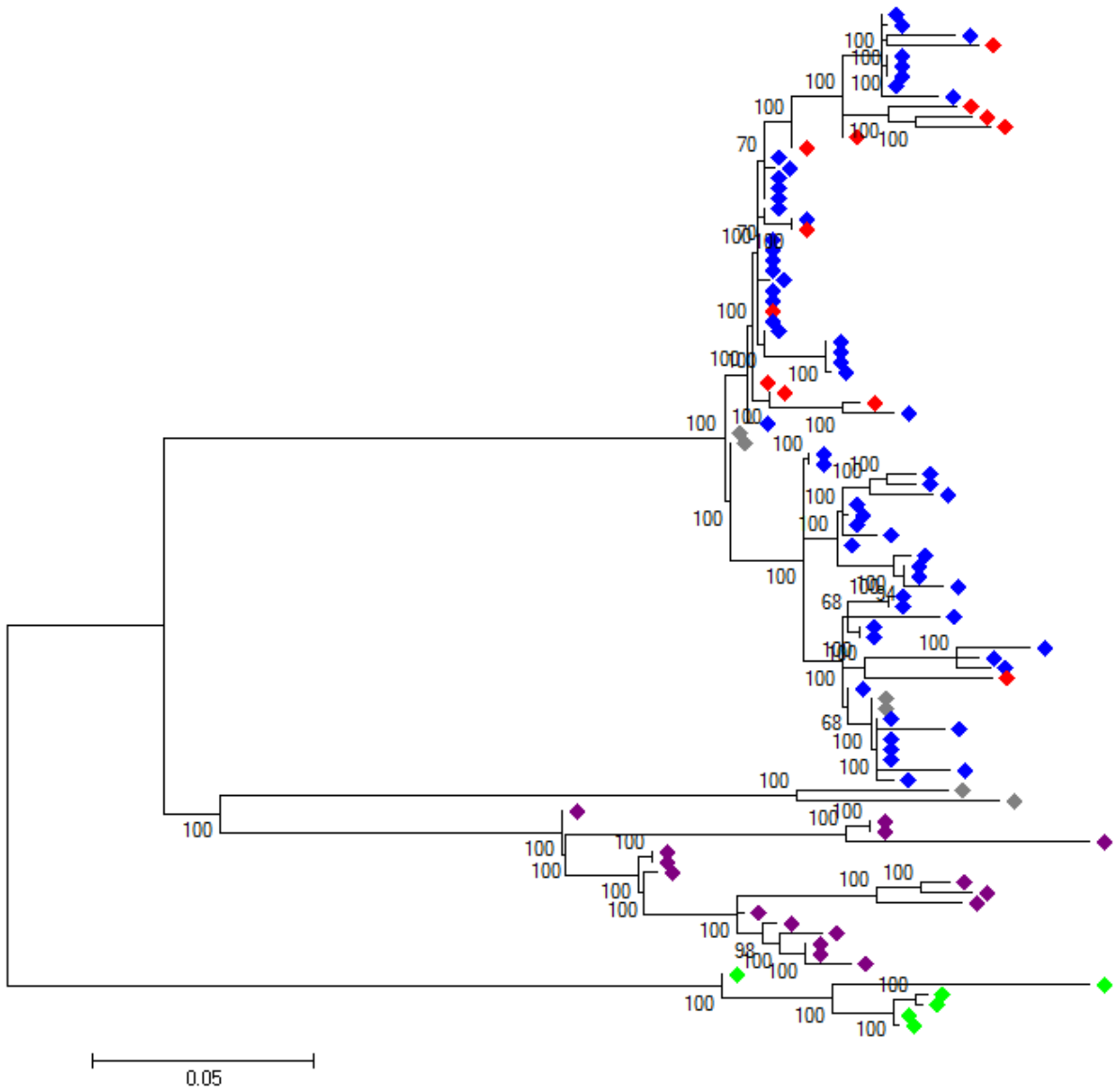


Figure S7.27 Chromosome 1: 140090001 - 140115001

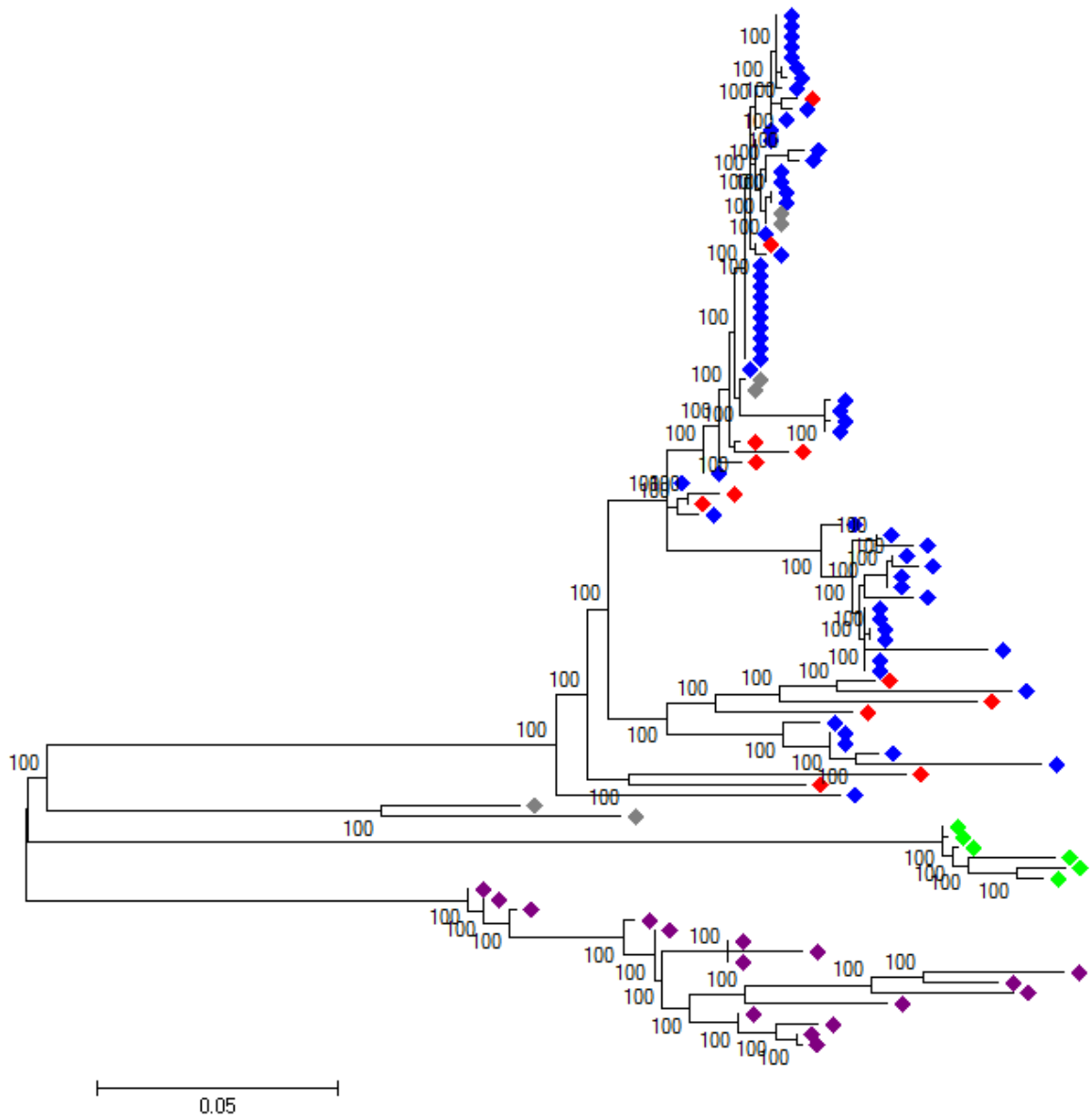


Figure S7.28 Chromosome 1: 140235001 - 140270001

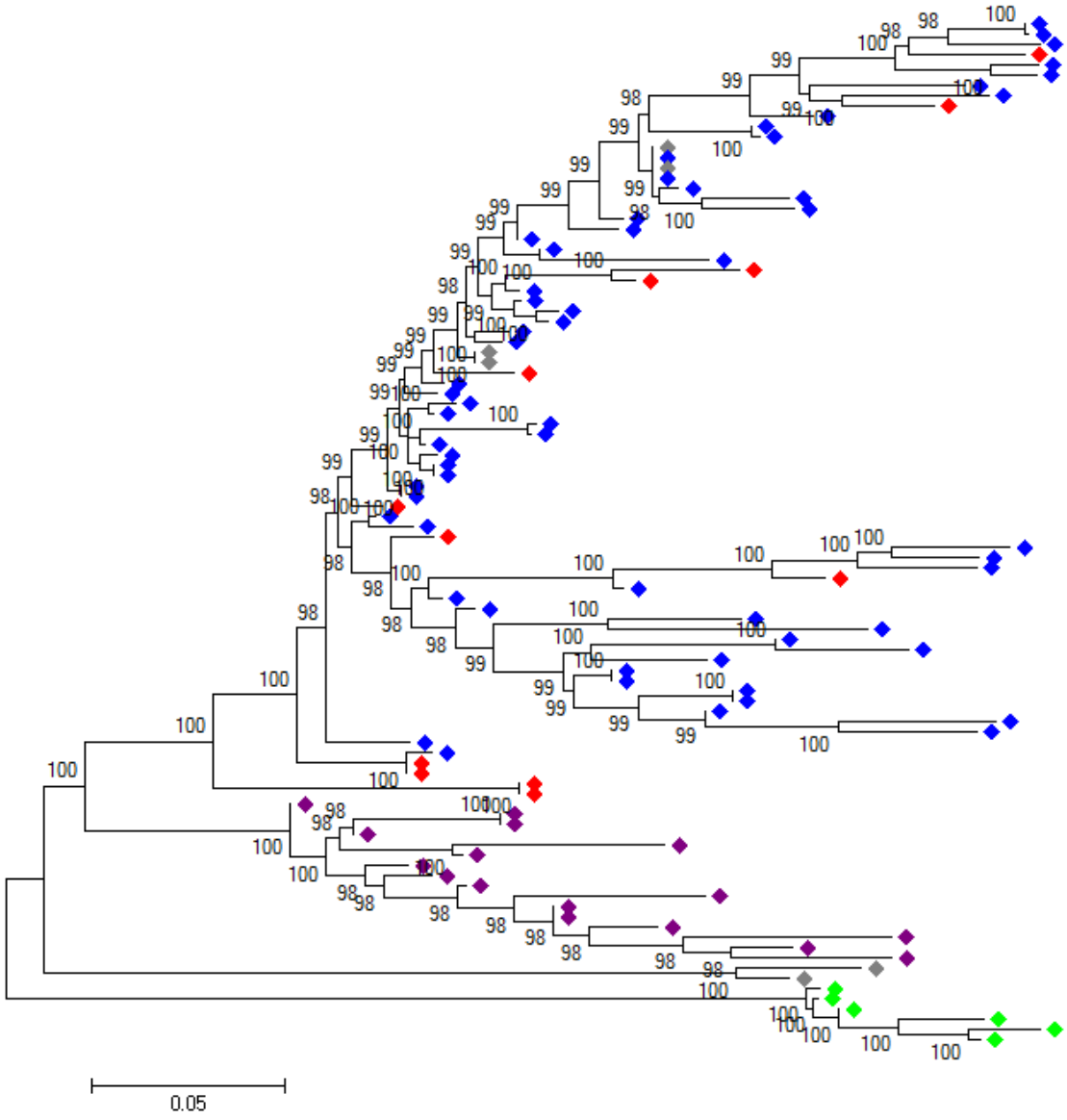


Figure S7.29 Chromosome 1: 140350001 - 140385001

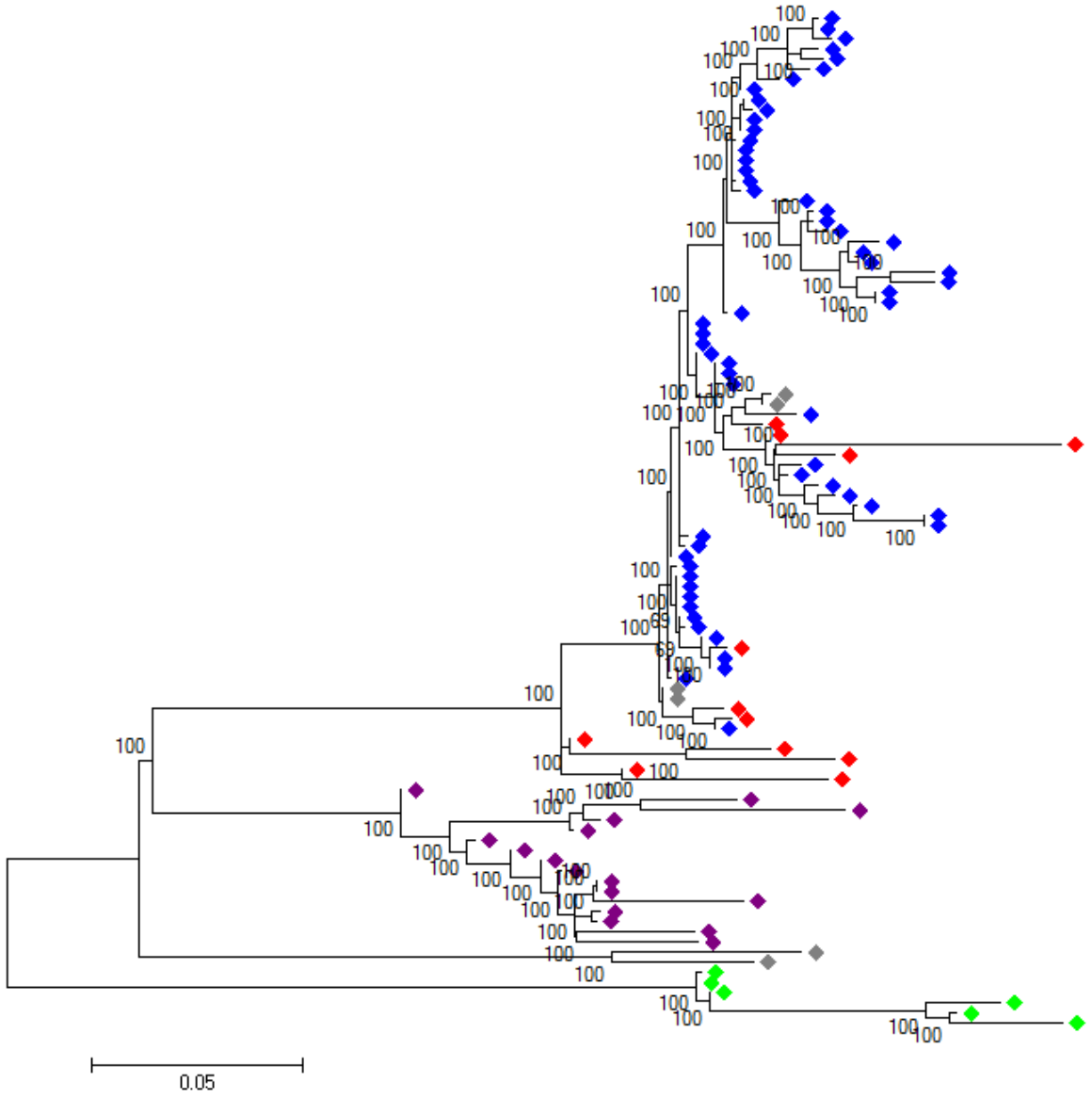


Figure S7.30 Chromosome 1: 140865001 - 140895001

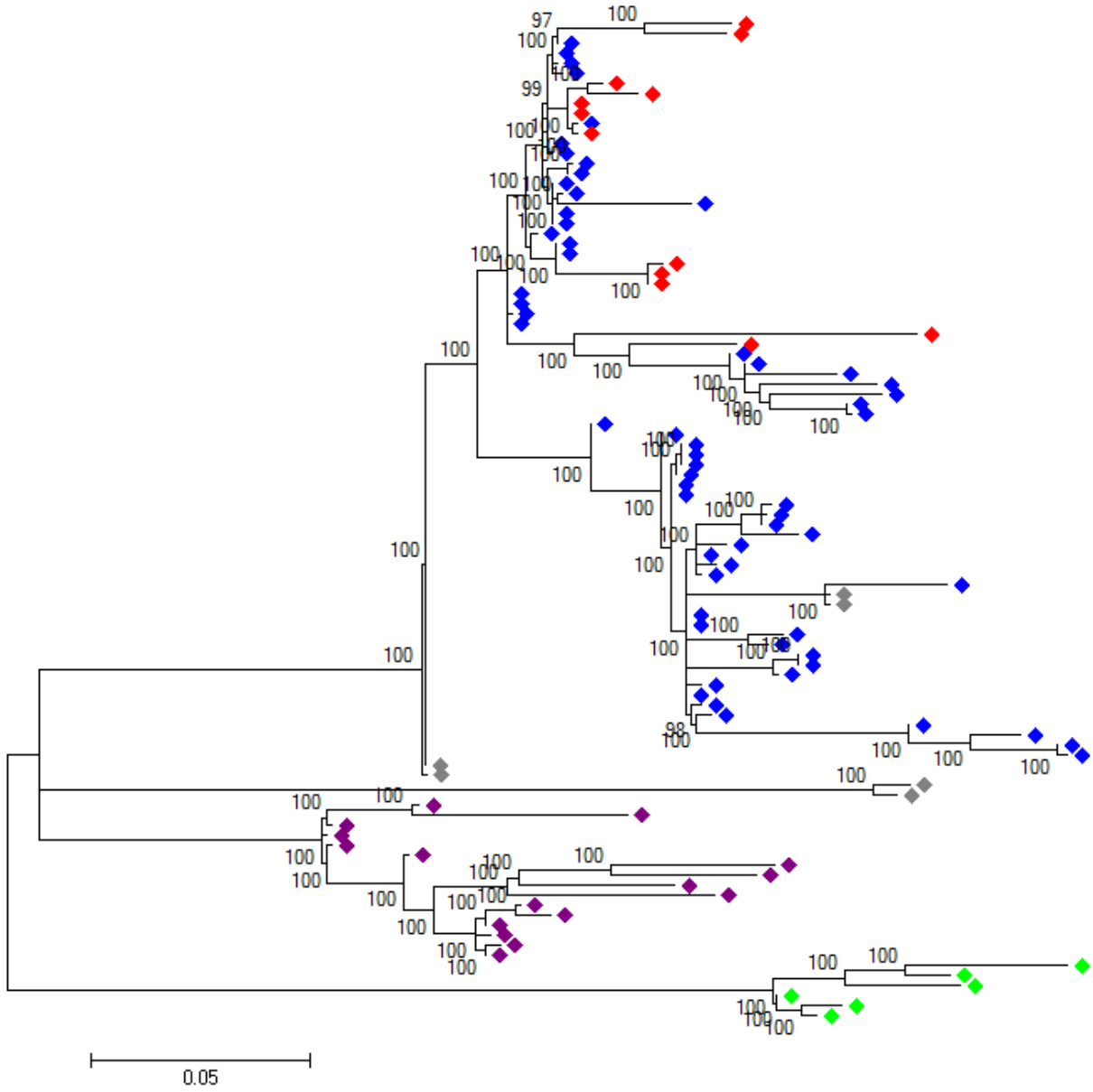


Figure S7.31 Chromosome 1: 141060001 - 141090001

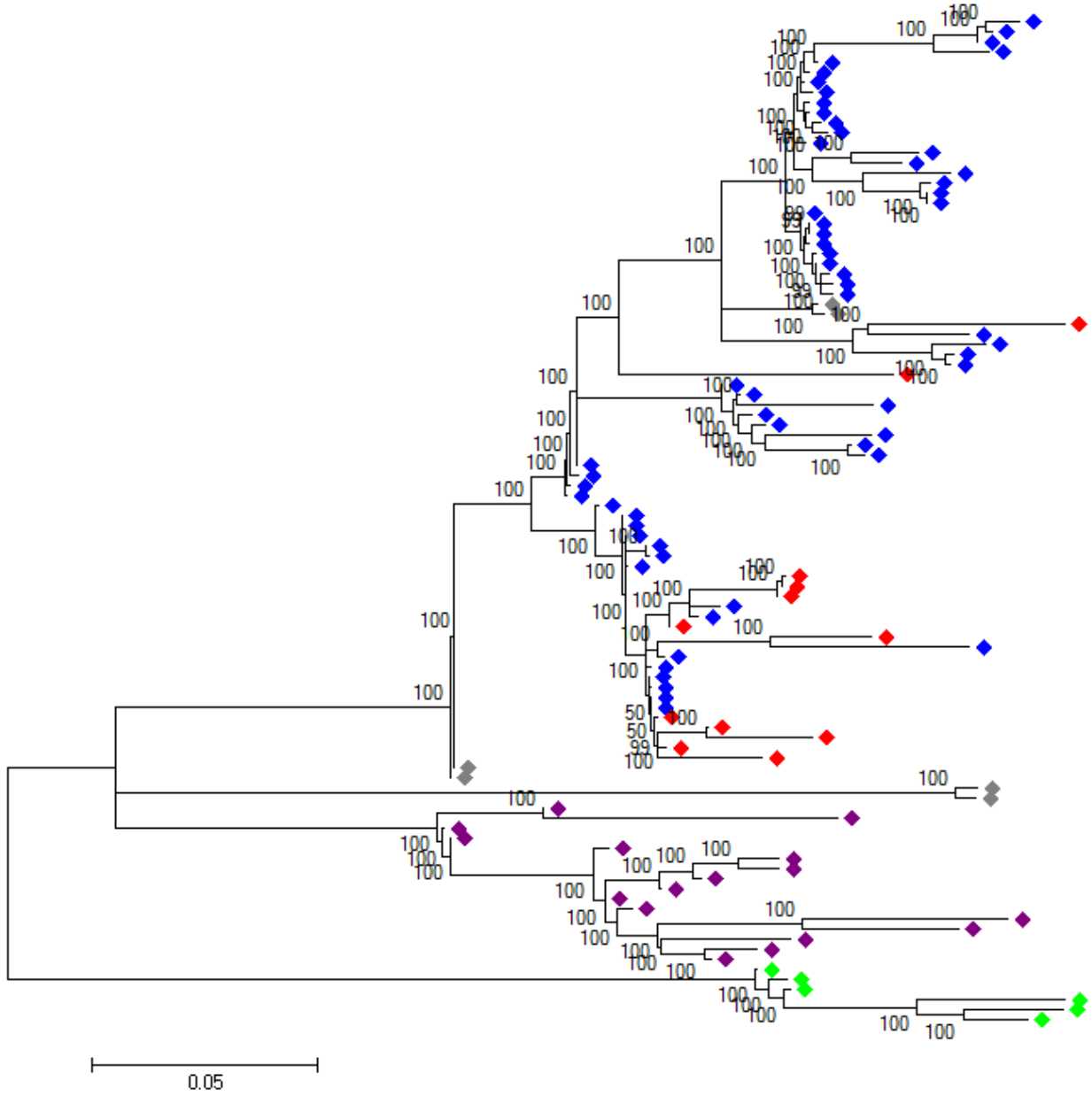


Figure S7.32 Chromosome 1: 141110001 - 141170001

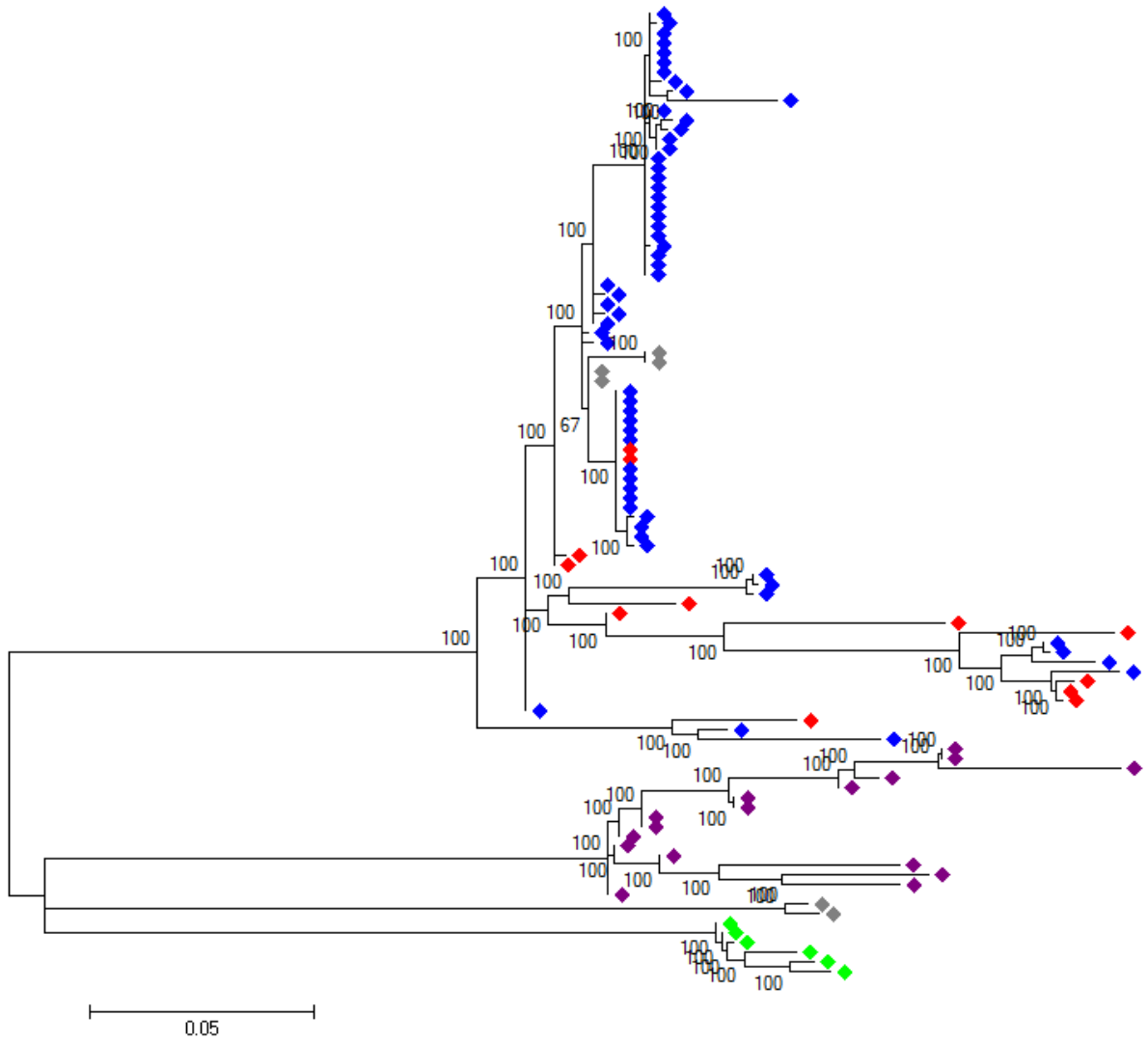


Figure S7.33 Chromosome 1: 141285001 - 141310001

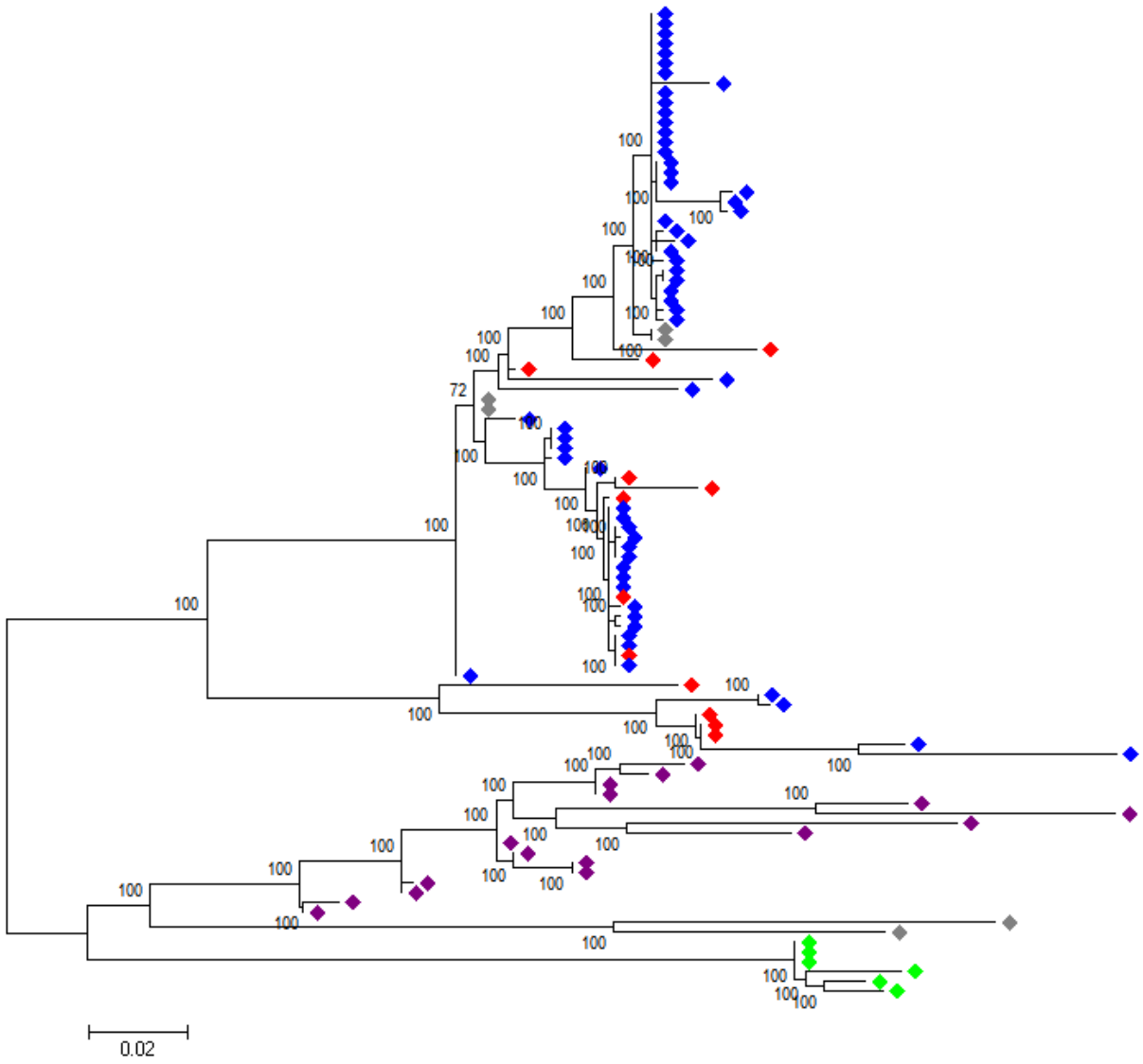


Figure S7.34 Chromosome 1: 141355001 - 141425001

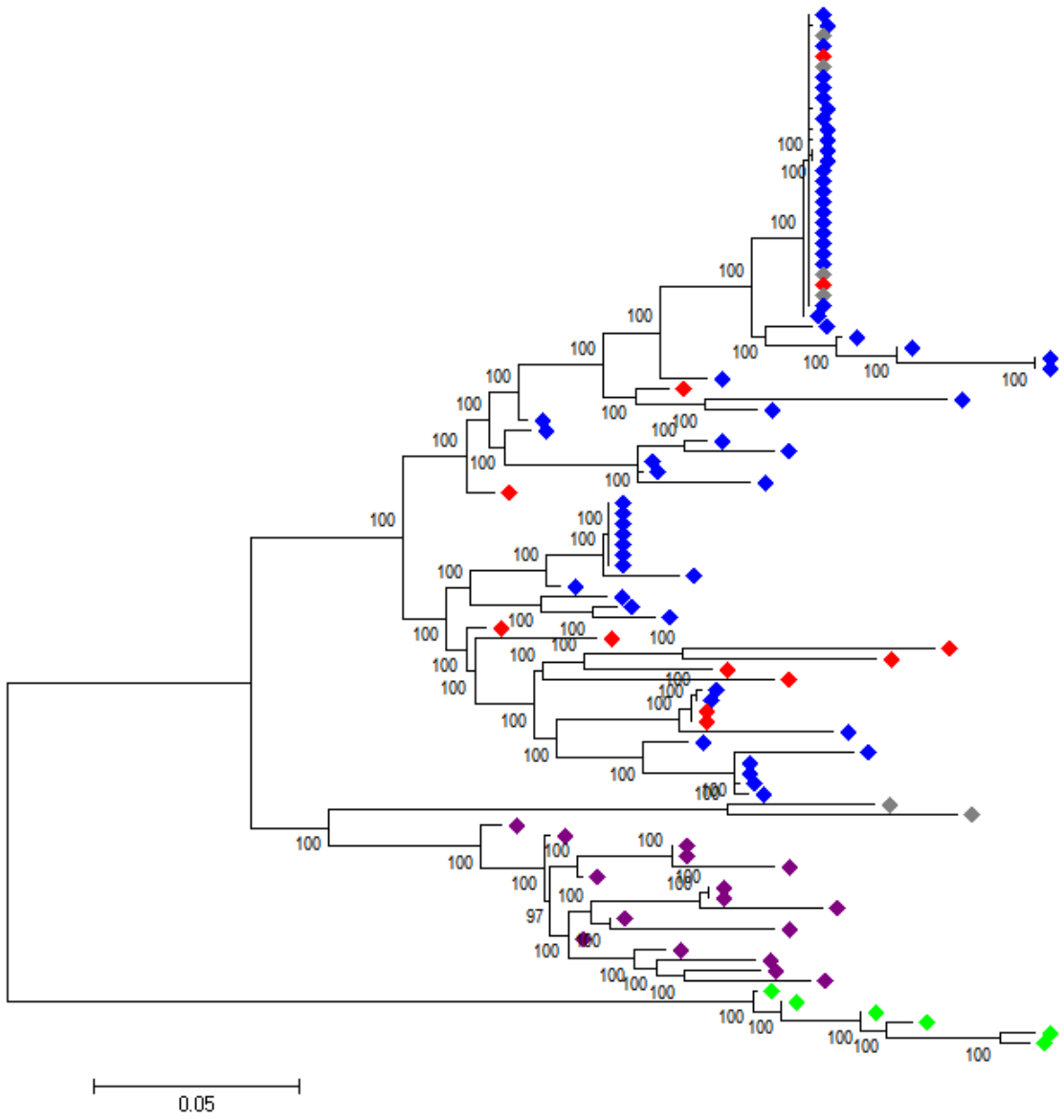


Figure S7.35 Chromosome 1: 141850001 - 141875001

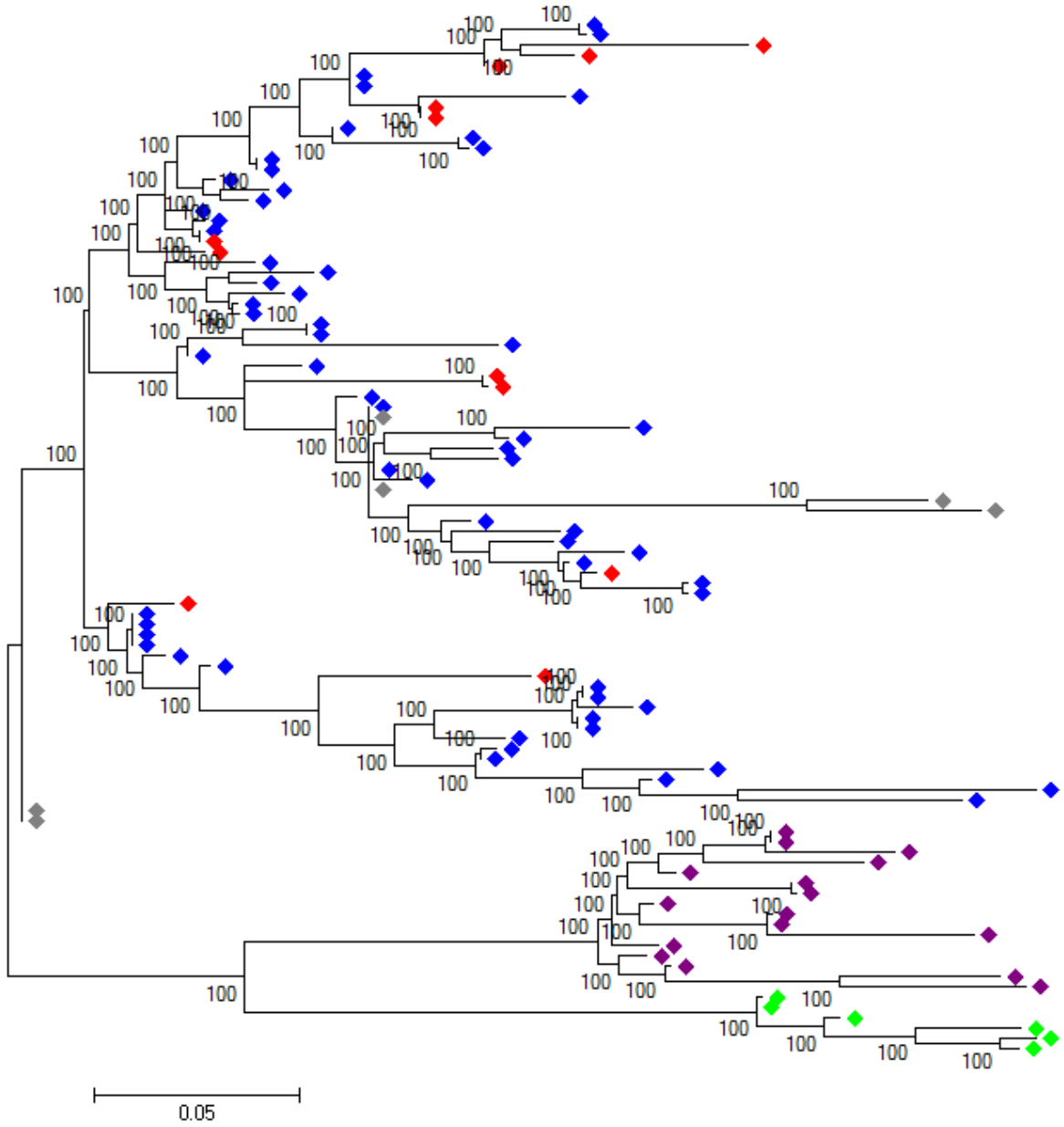


Figure S7.36 Chromosome 1: 141915001 - 141945001

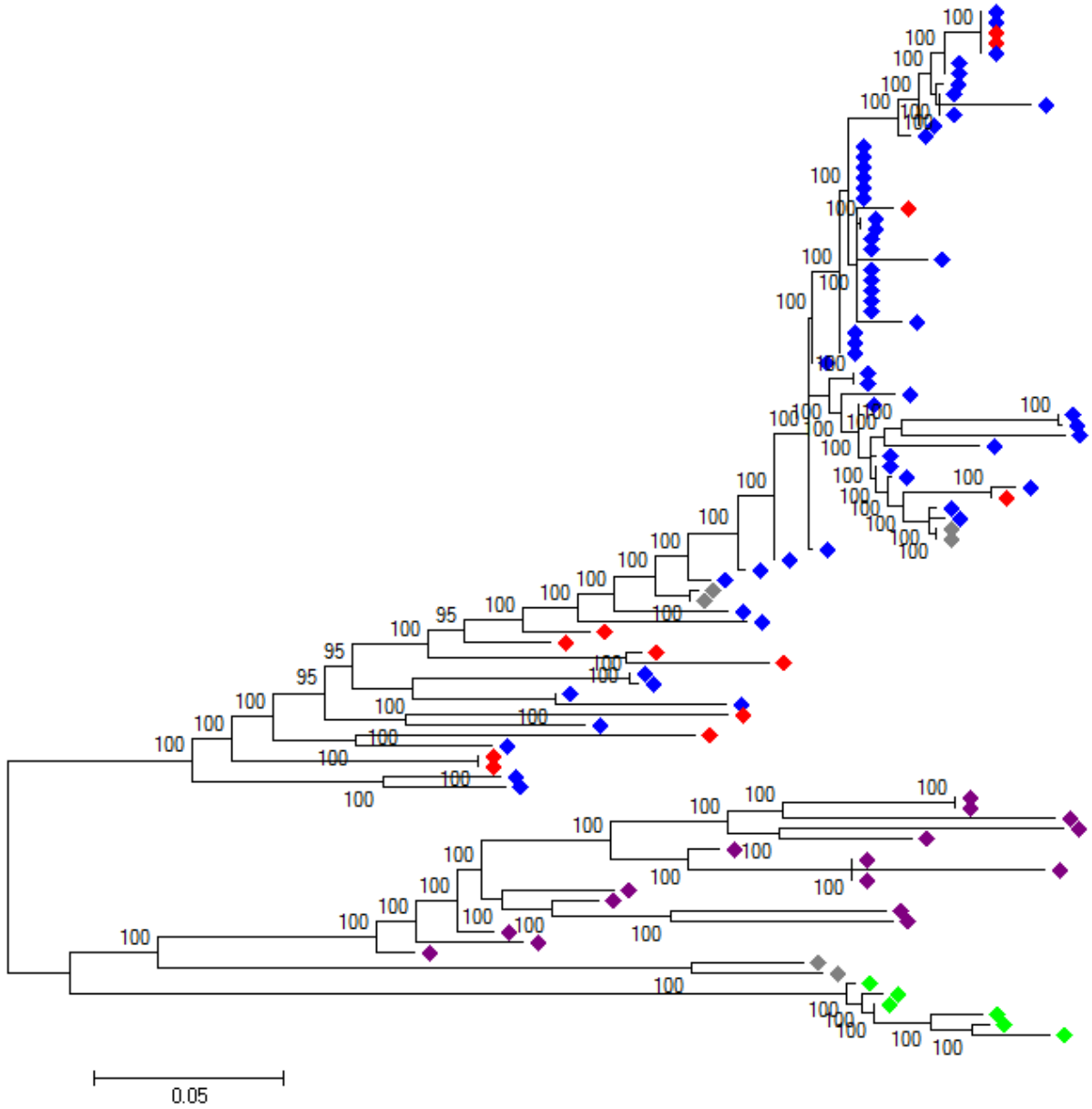


Figure S7.37 Chromosome 1: 142025001 - 142045001

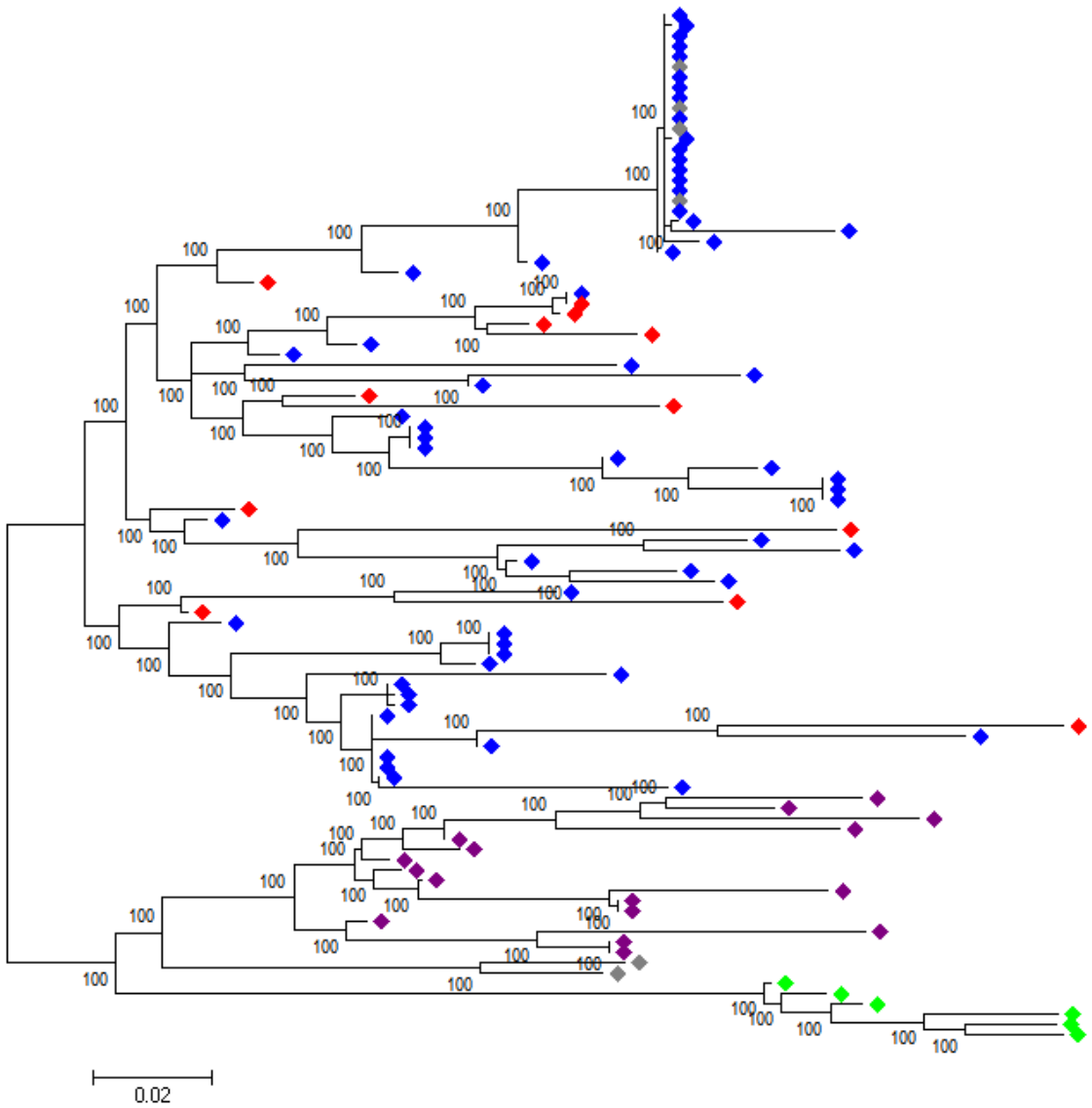


Figure S7.38 Chromosome 1: 142350001 - 142375001

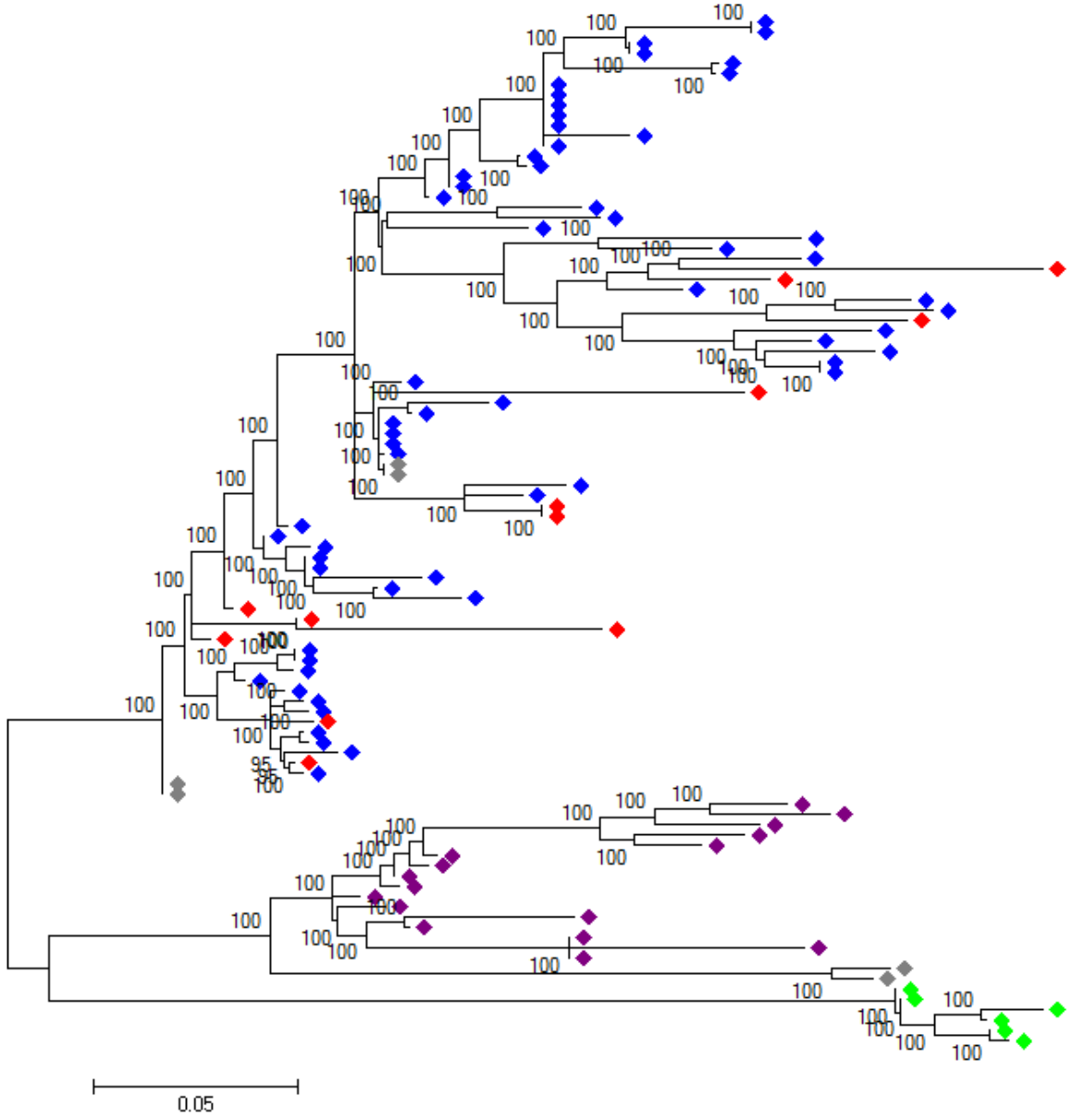


Figure S7.39 Chromosome 1: 142550001 - 142575001

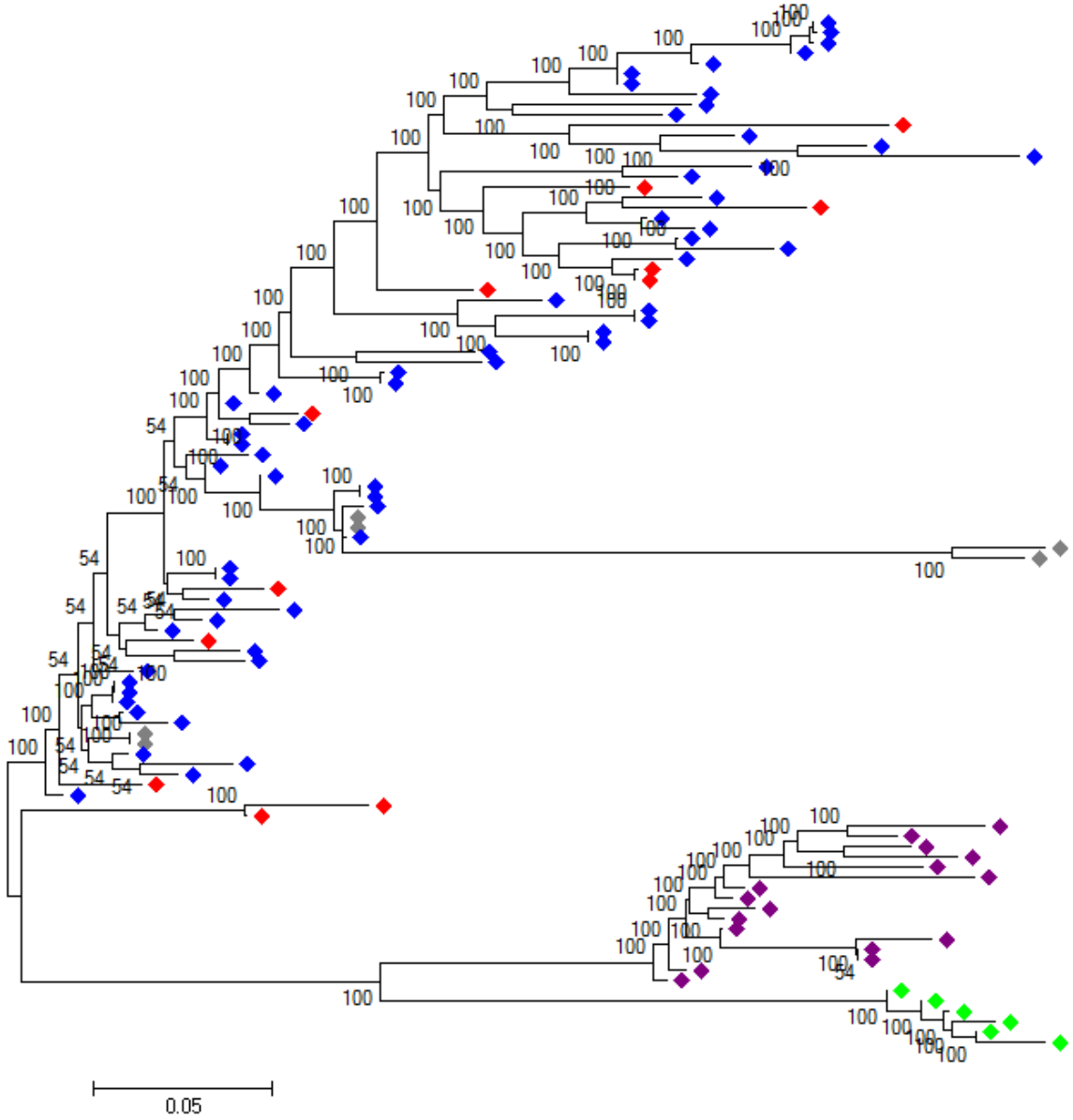


Figure S7.40 Chromosome 1: 142625001 - 142655001

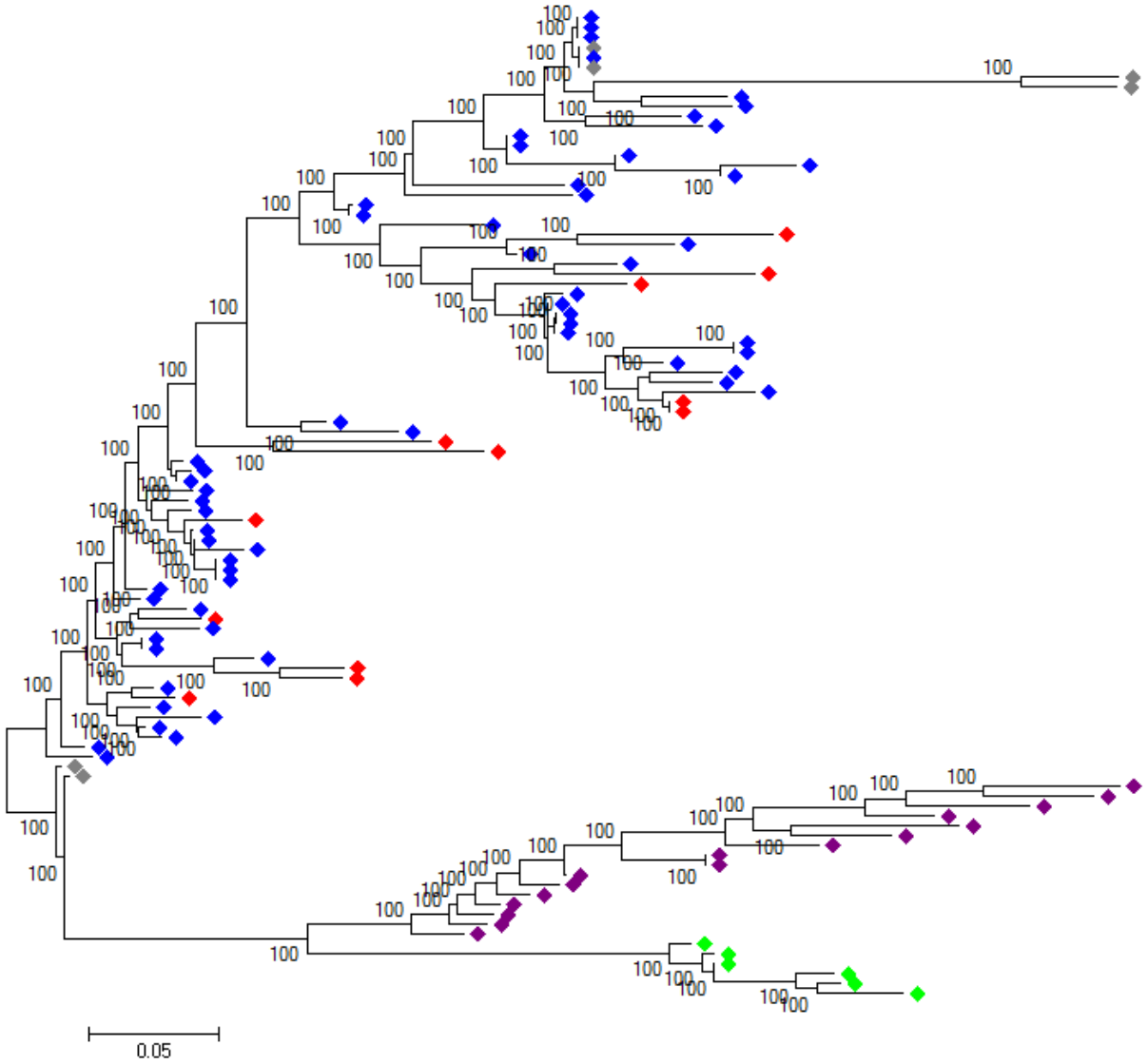


Figure S7.41 Chromosome 1: 142885001 - 142910001

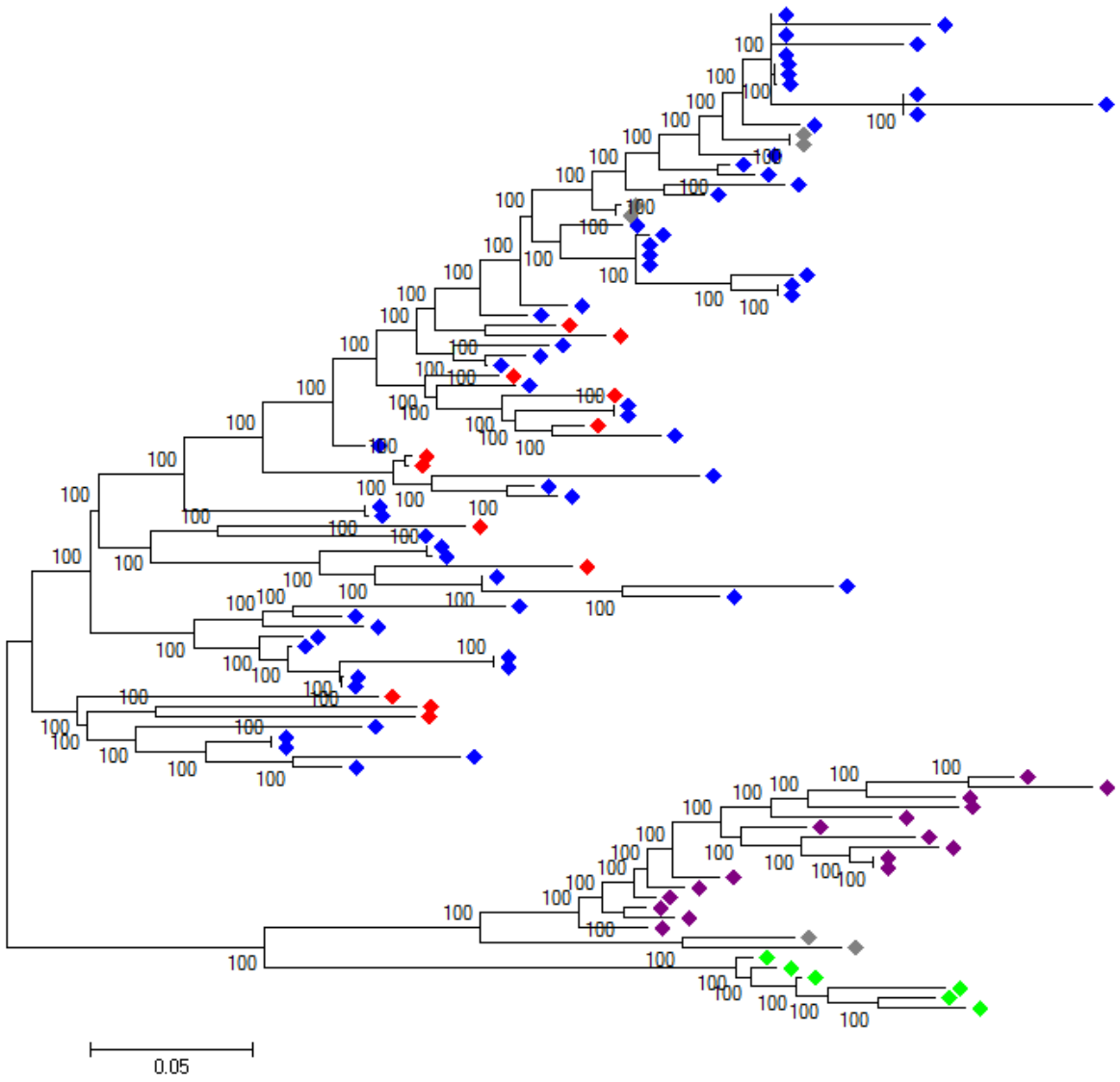


Figure S7.42 Chromosome 1: 142970001 - 143005001

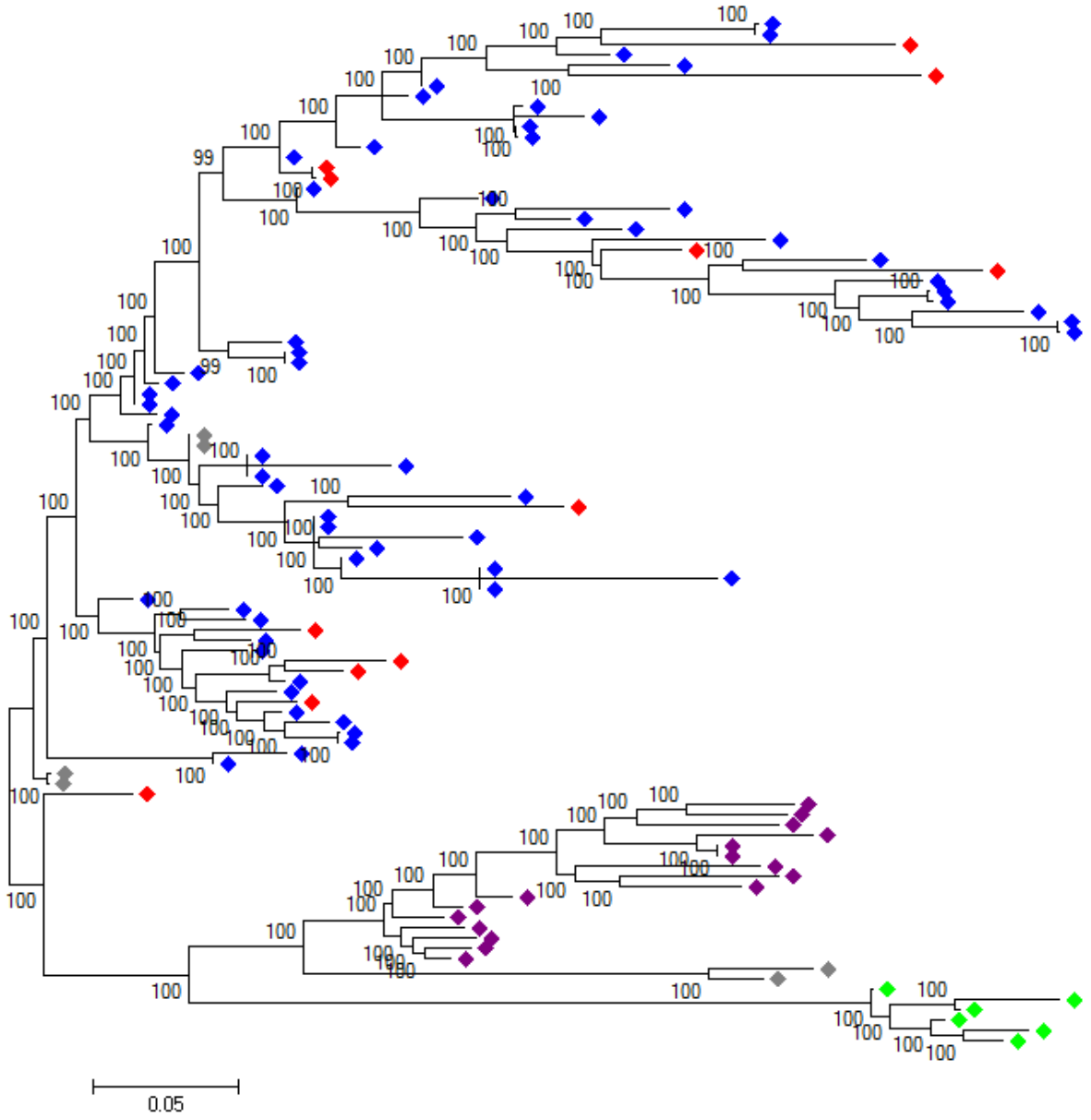


Figure S7.43 Chromosome 1: 143035001 - 143065001

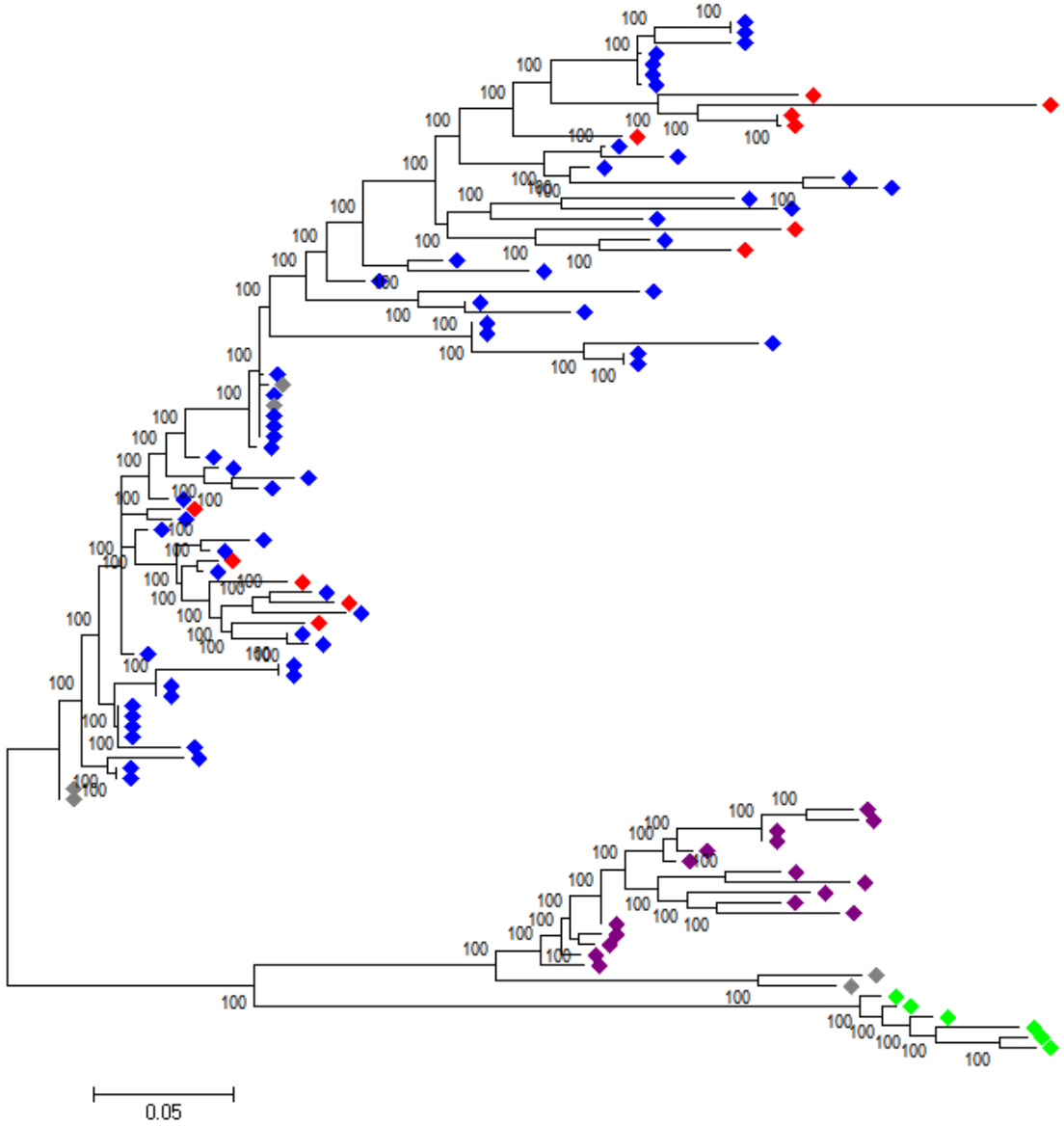


Figure S7.44 Chromosome 1: 143255001 - 143280001

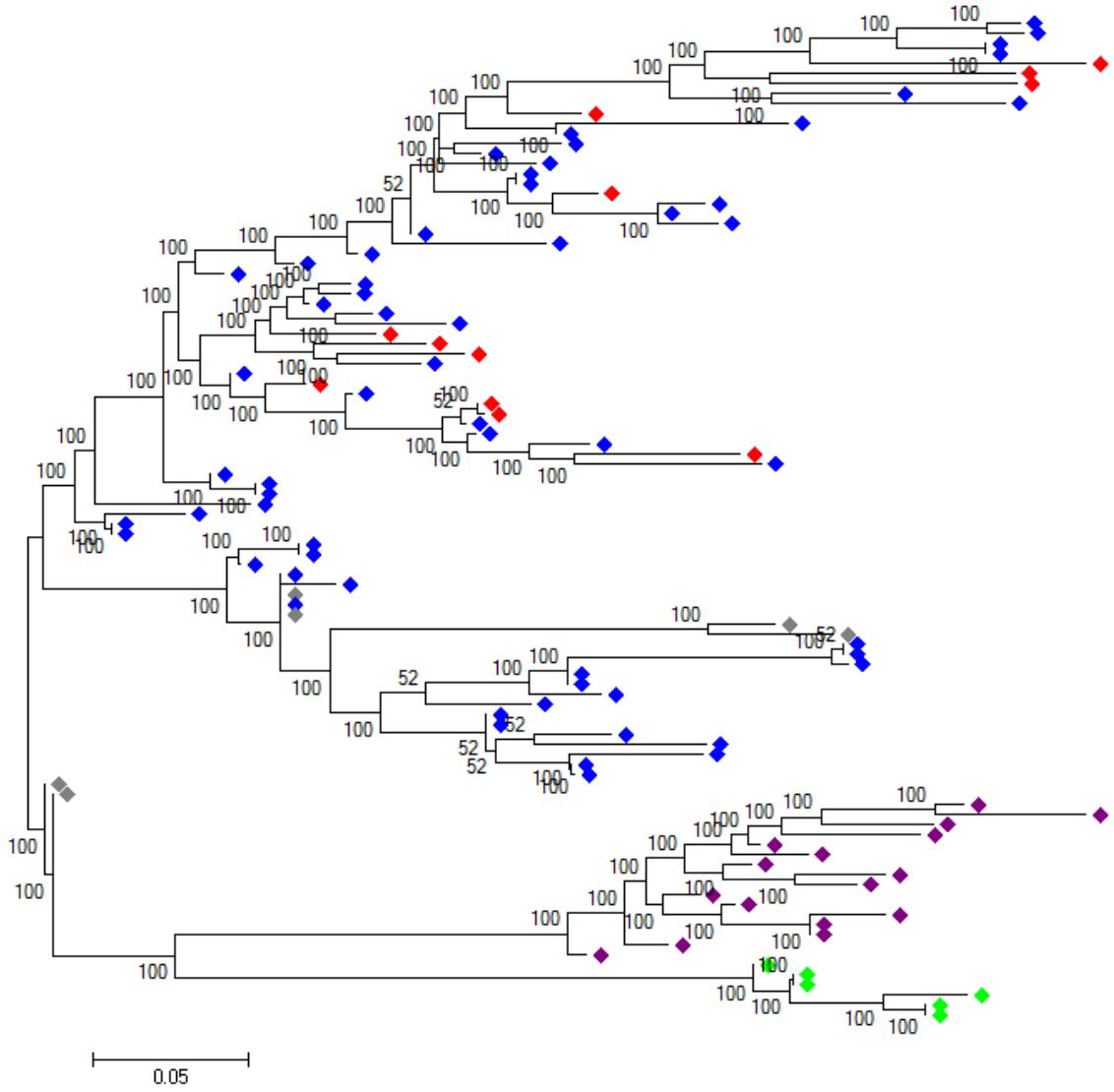


Figure S7.45 Chromosome 1: 143320001 - 143340001

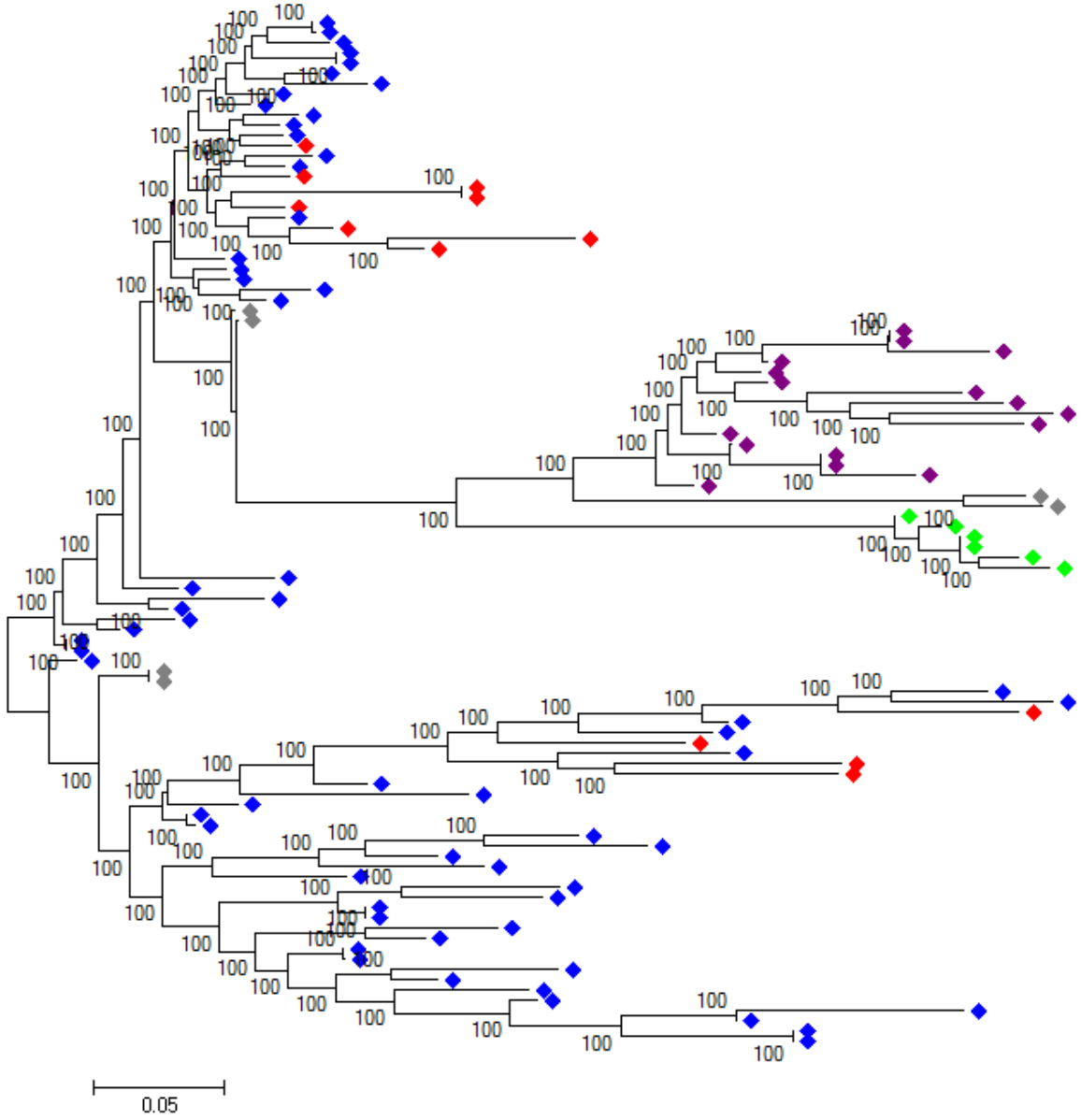


Figure S7.46 Chromosome 1: 143920001 - 143945001

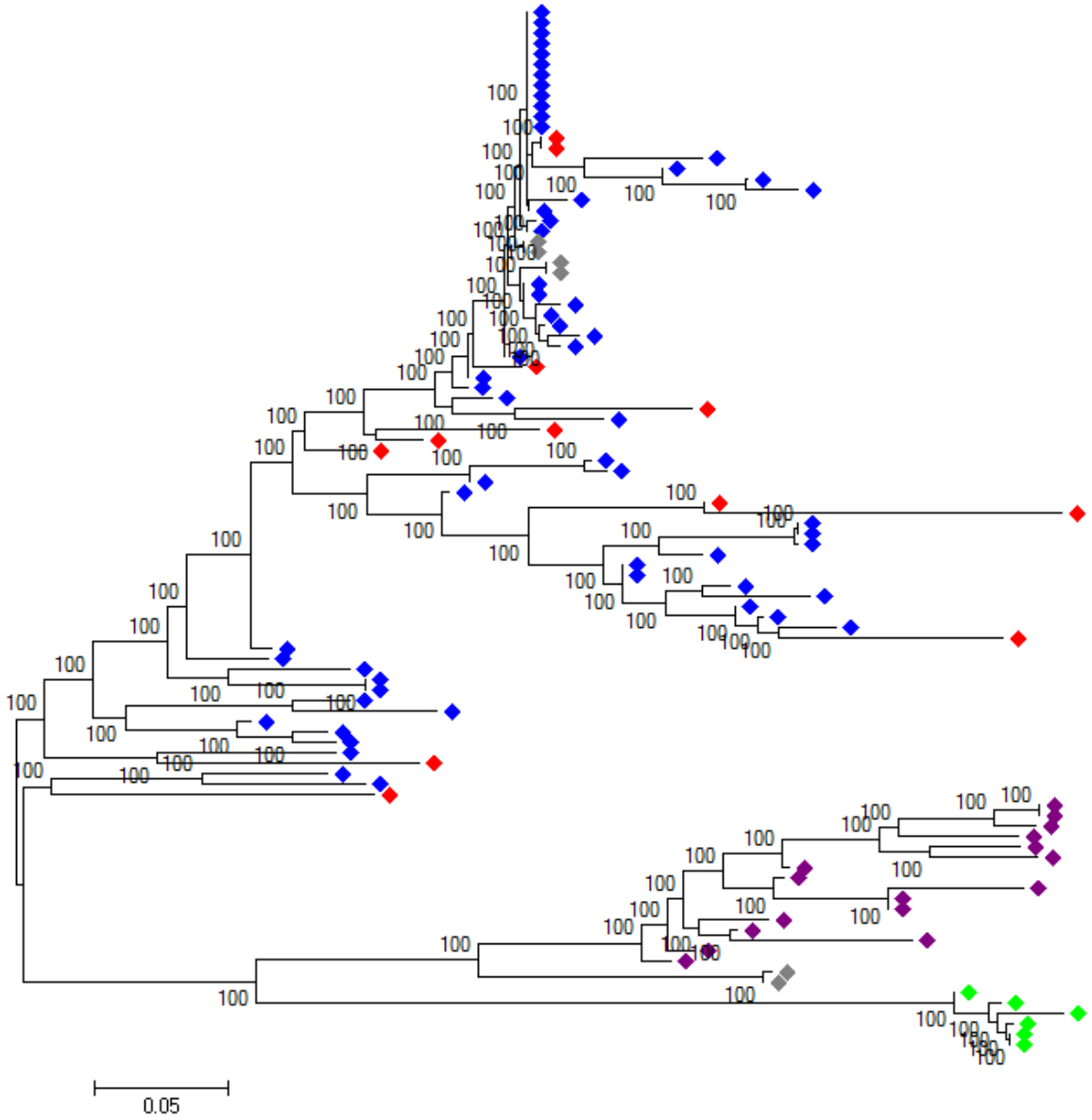


Figure S7.47 Chromosome 1: 144070001 - 144090001

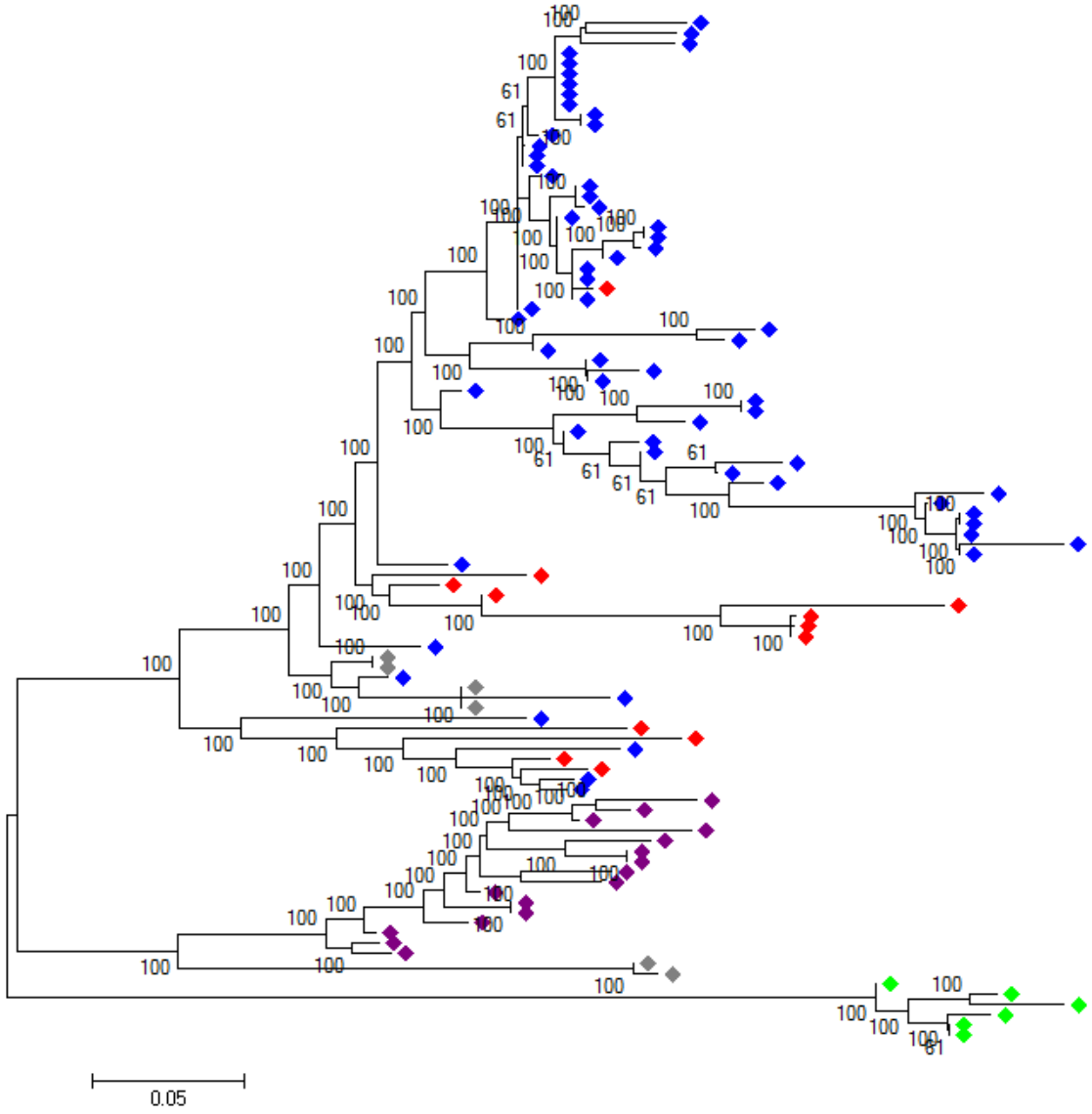


Figure S7.48 Chromosome 1: 144130001 - 144160001

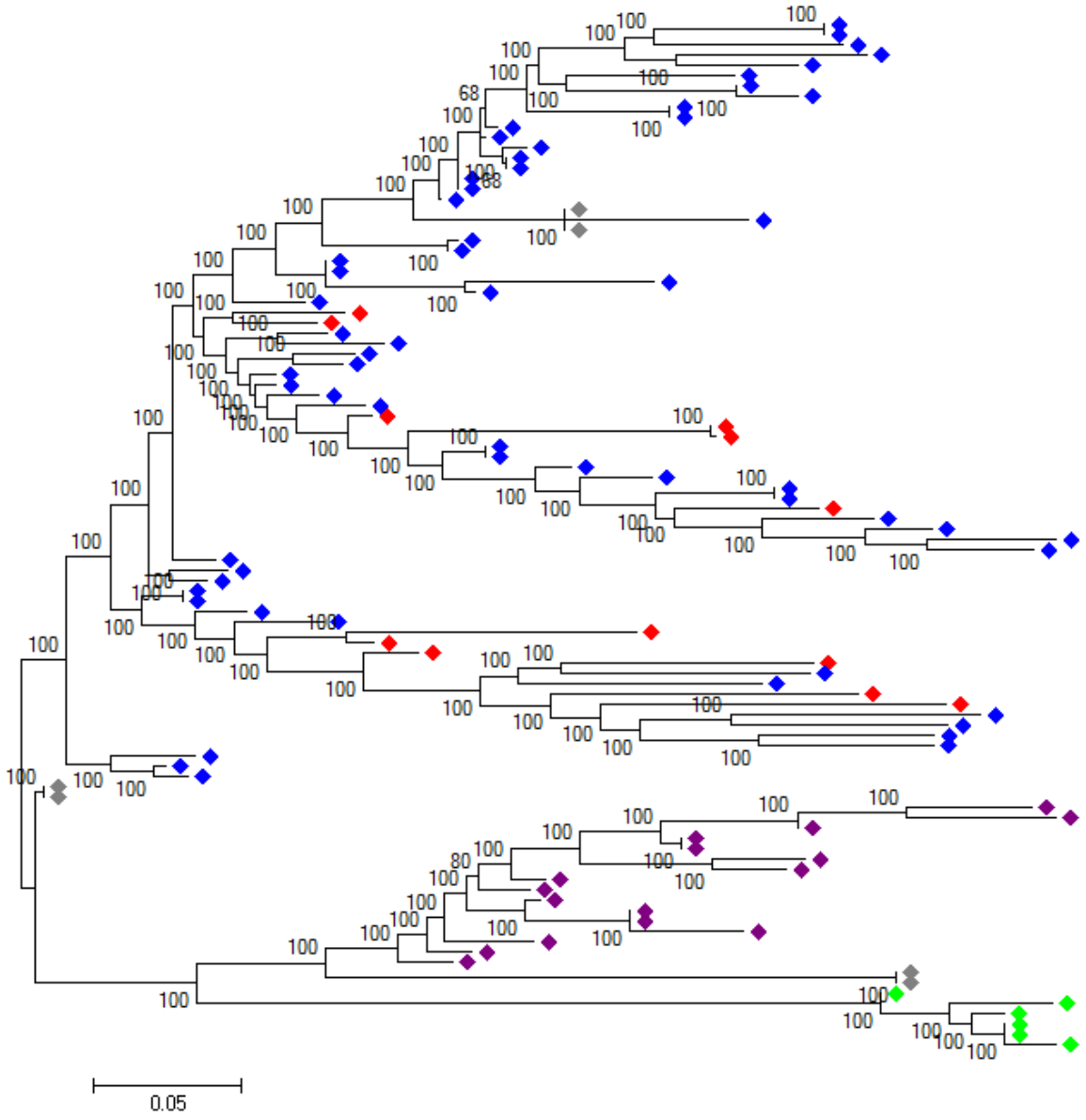


Figure S7.49 Chromosome 1: 144265001 - 144290001

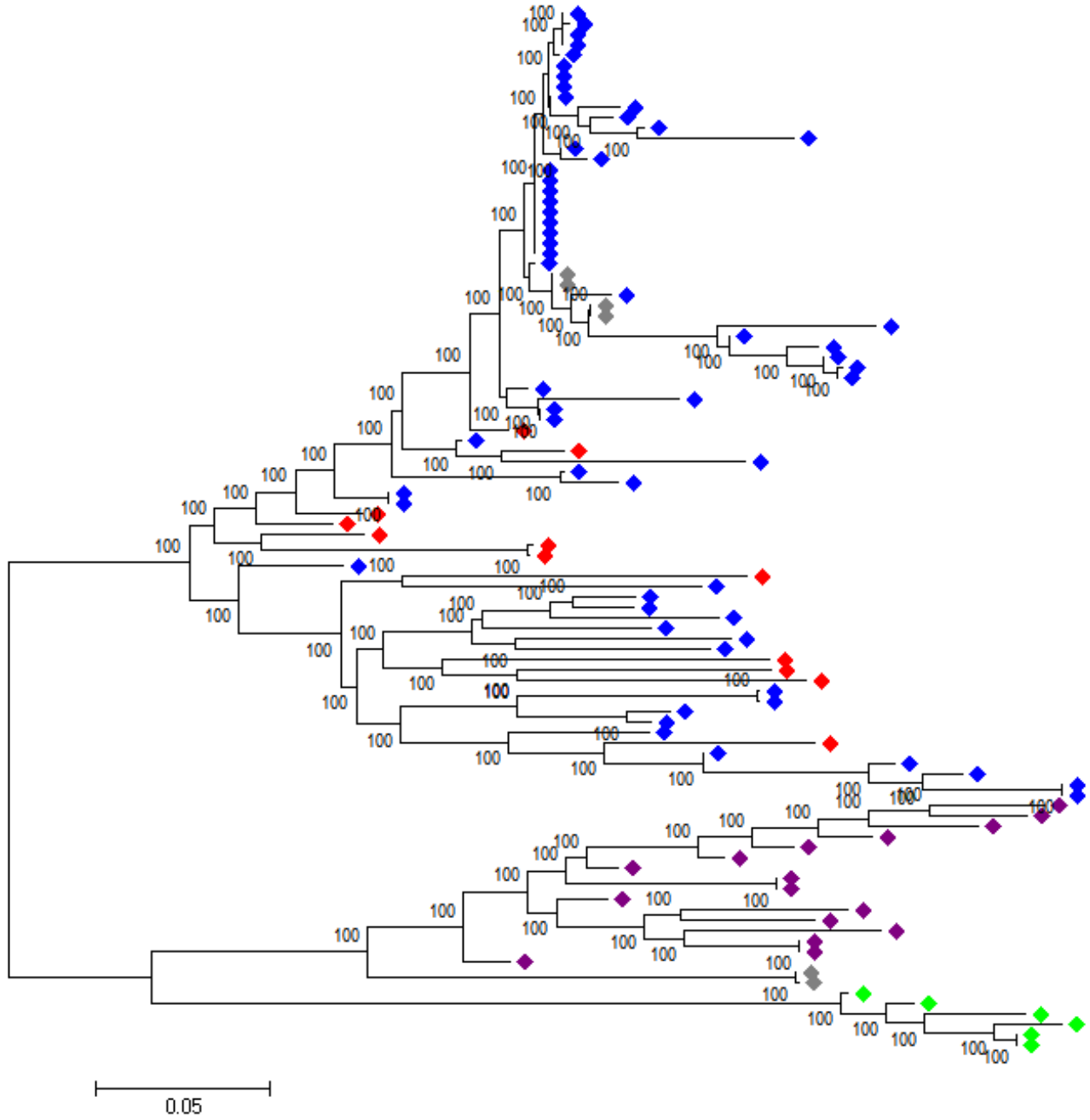


Figure S7.50 Chromosome 1: 144440001 - 144475001

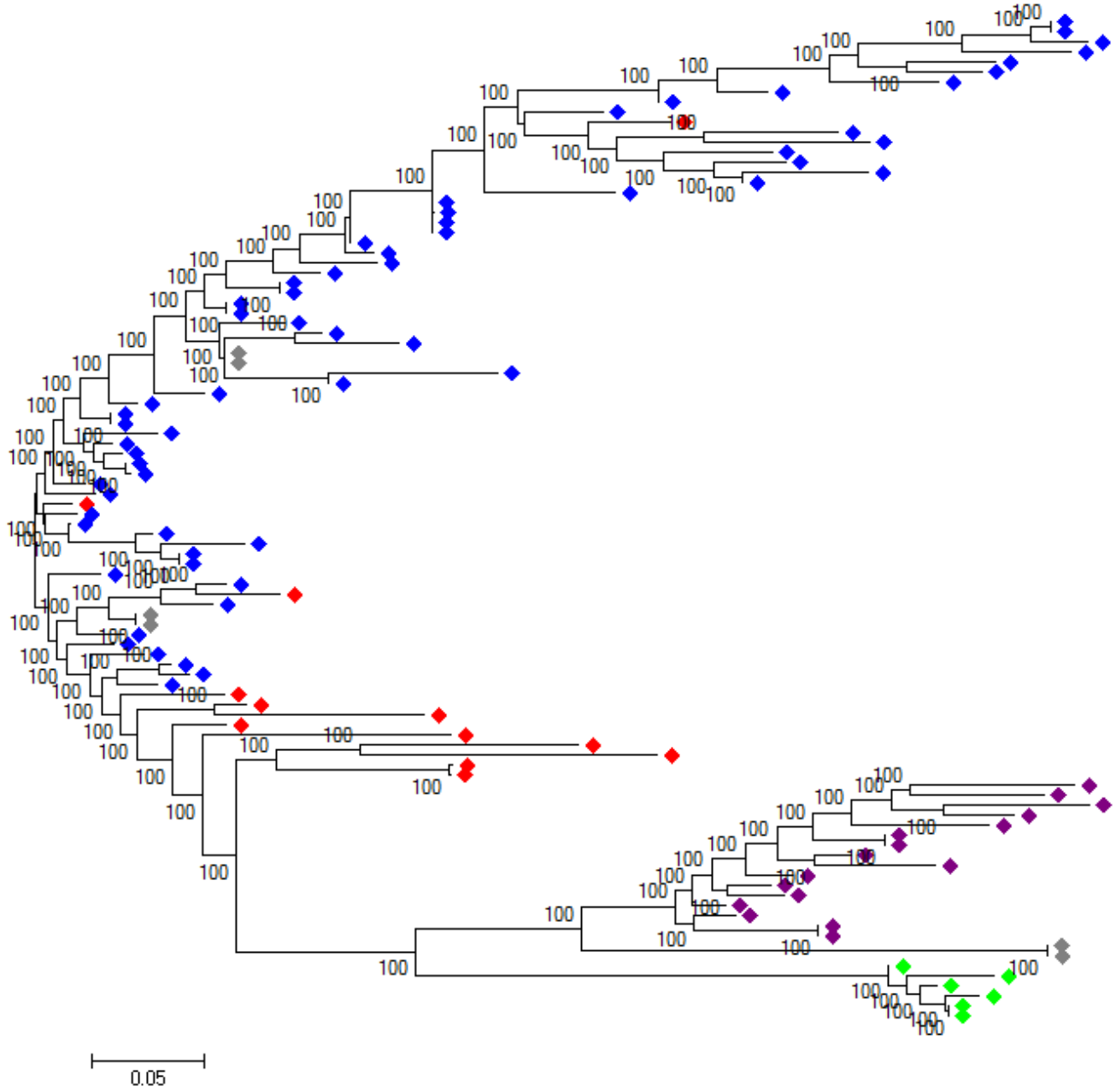


Figure S7.51 Chromosome 1: 144555001 - 144605001

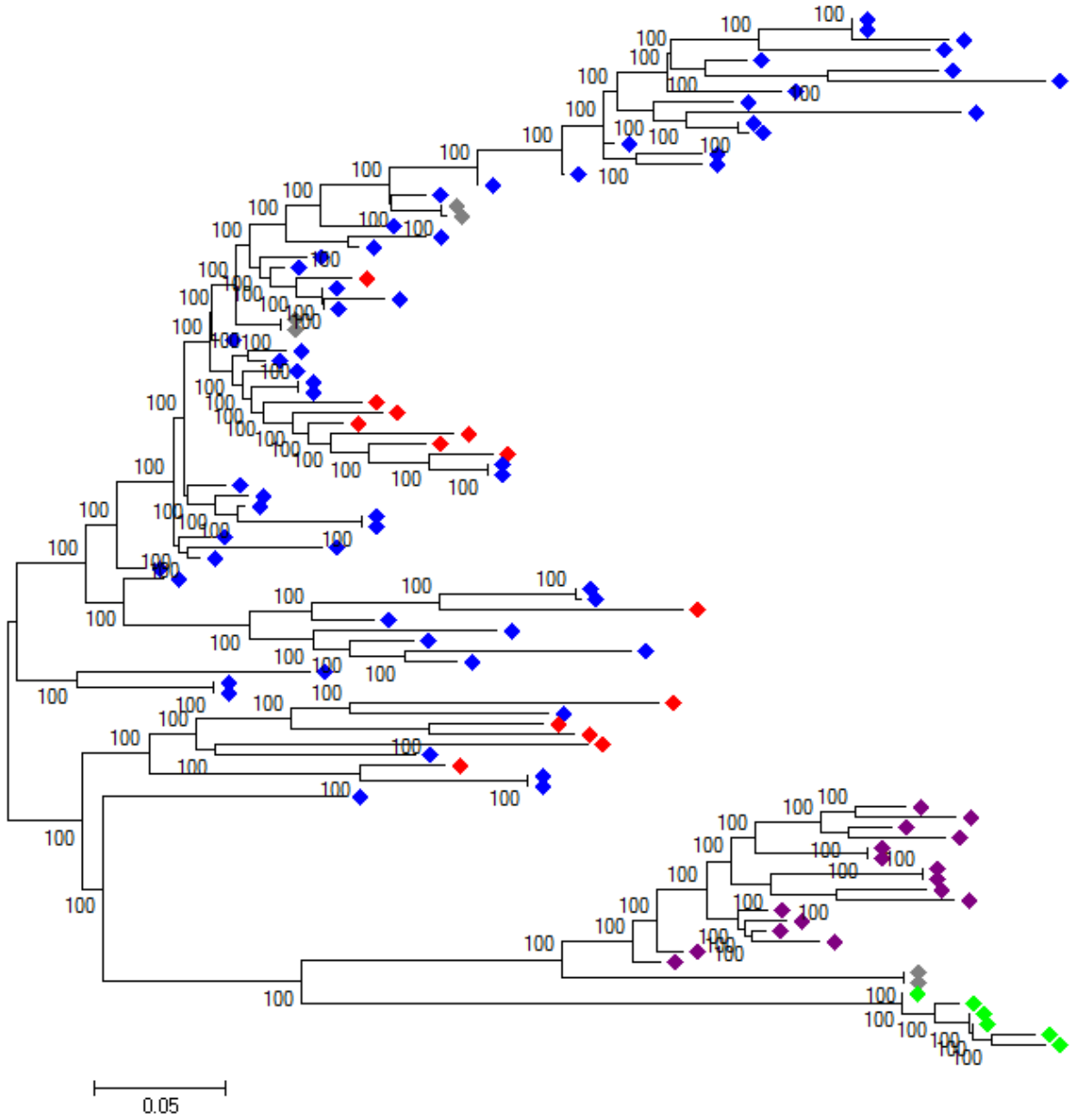


Figure S7.52 Chromosome 1: 145140001 - 145170001

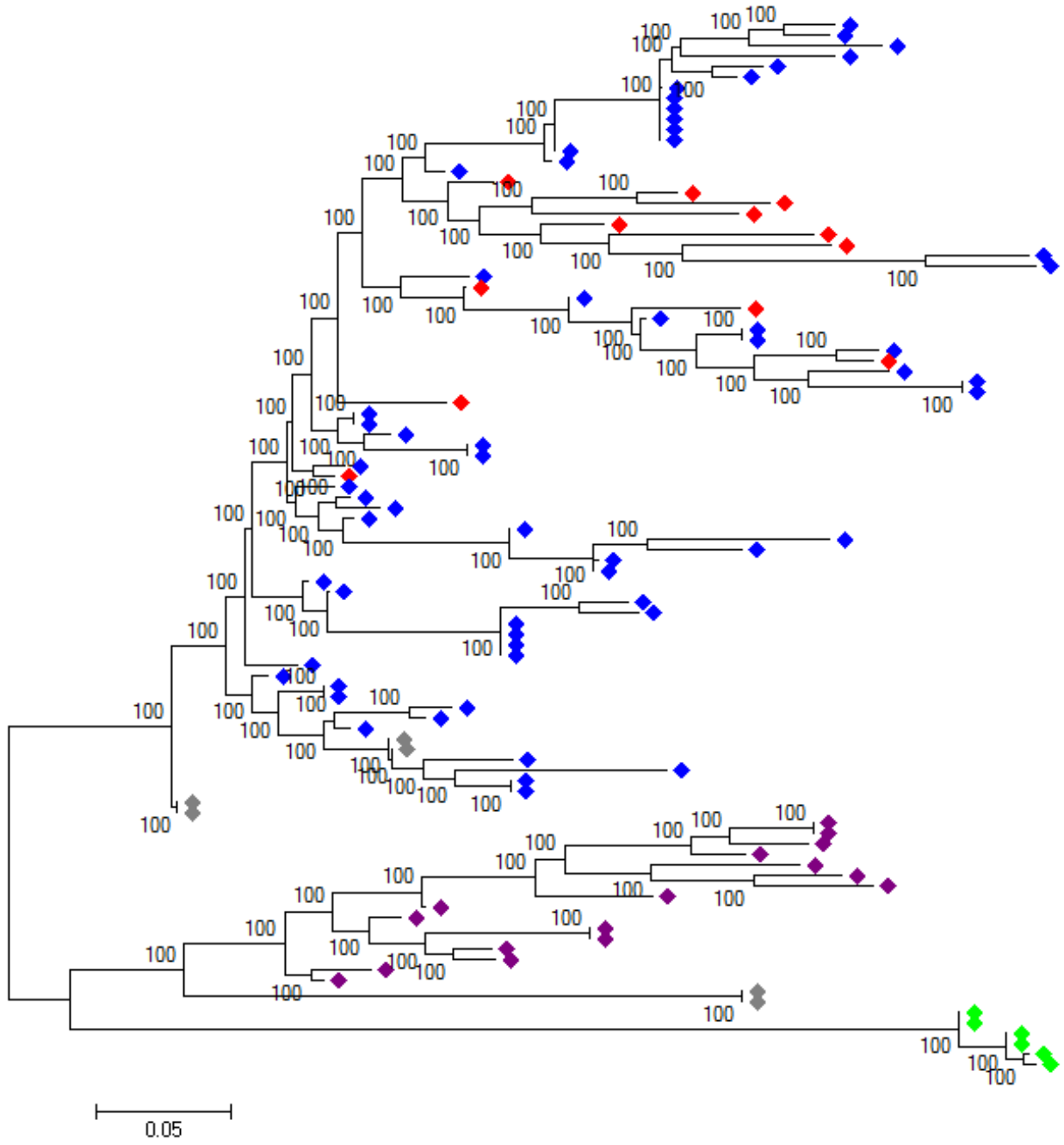


Figure S7.53 Chromosome 1: 145285001 - 145310001

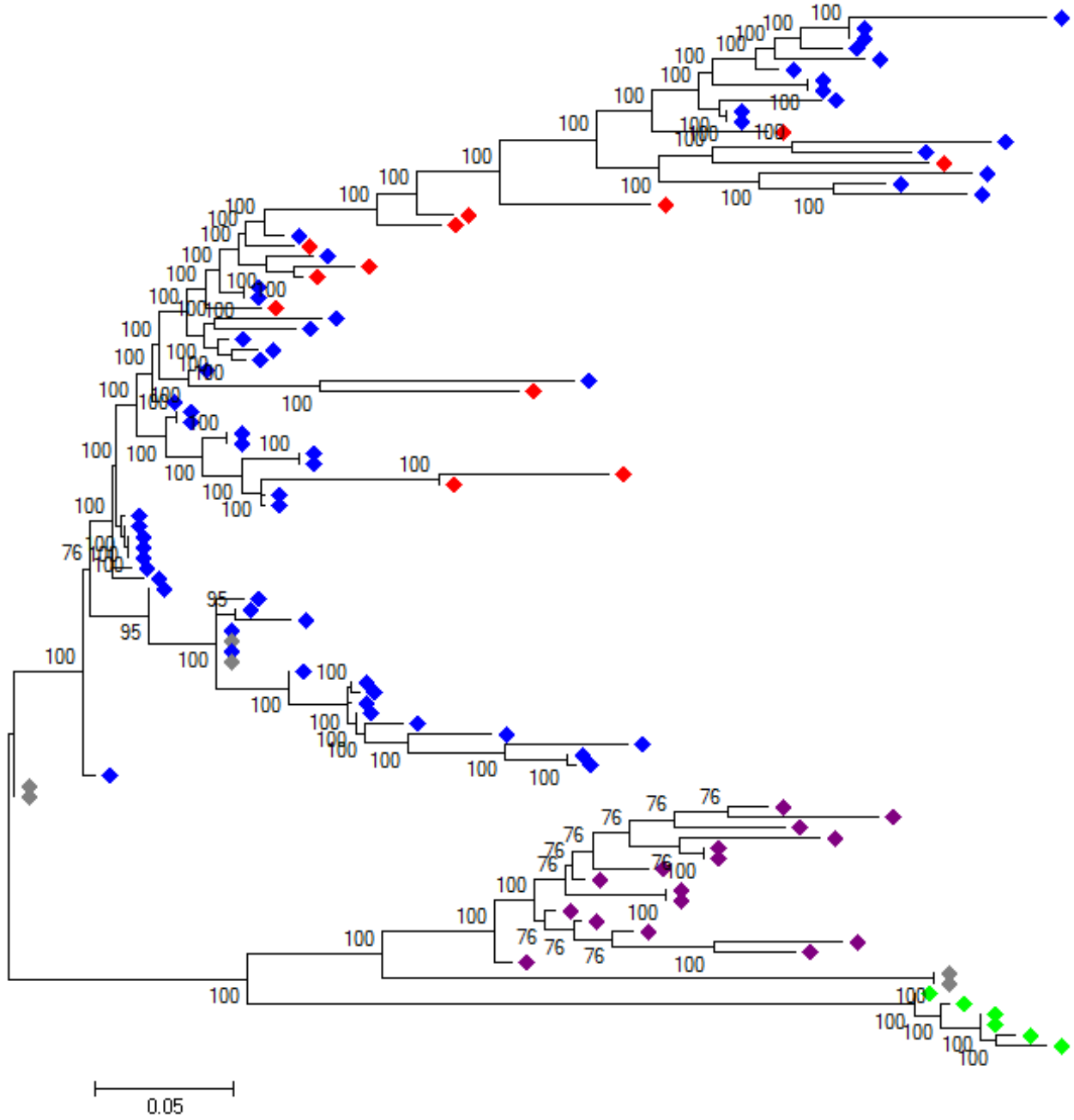


Figure S7.54 Chromosome 1: 145455001 - 145475001

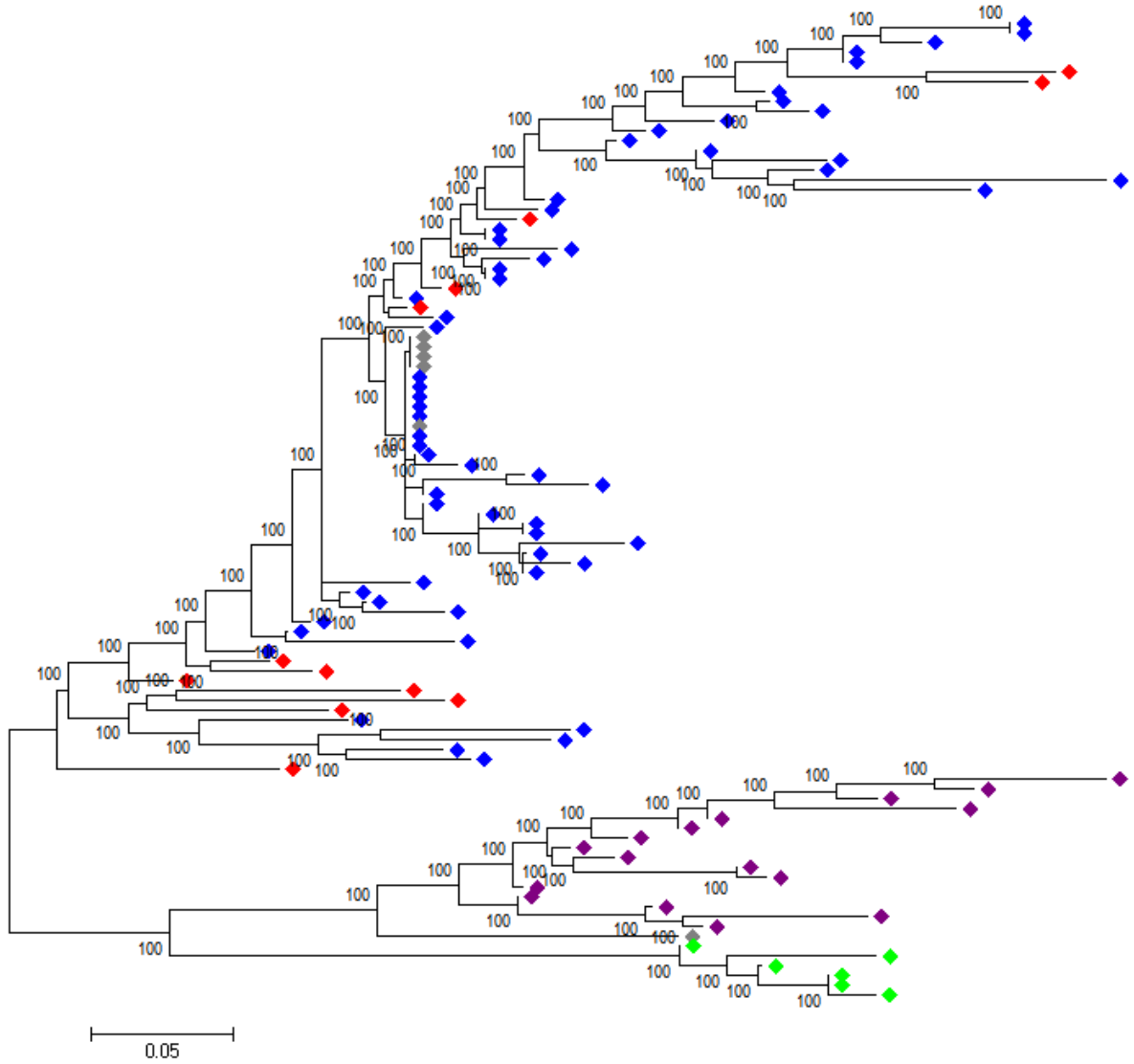


Figure S7.55 Chromosome 1: 145540001 - 145560001

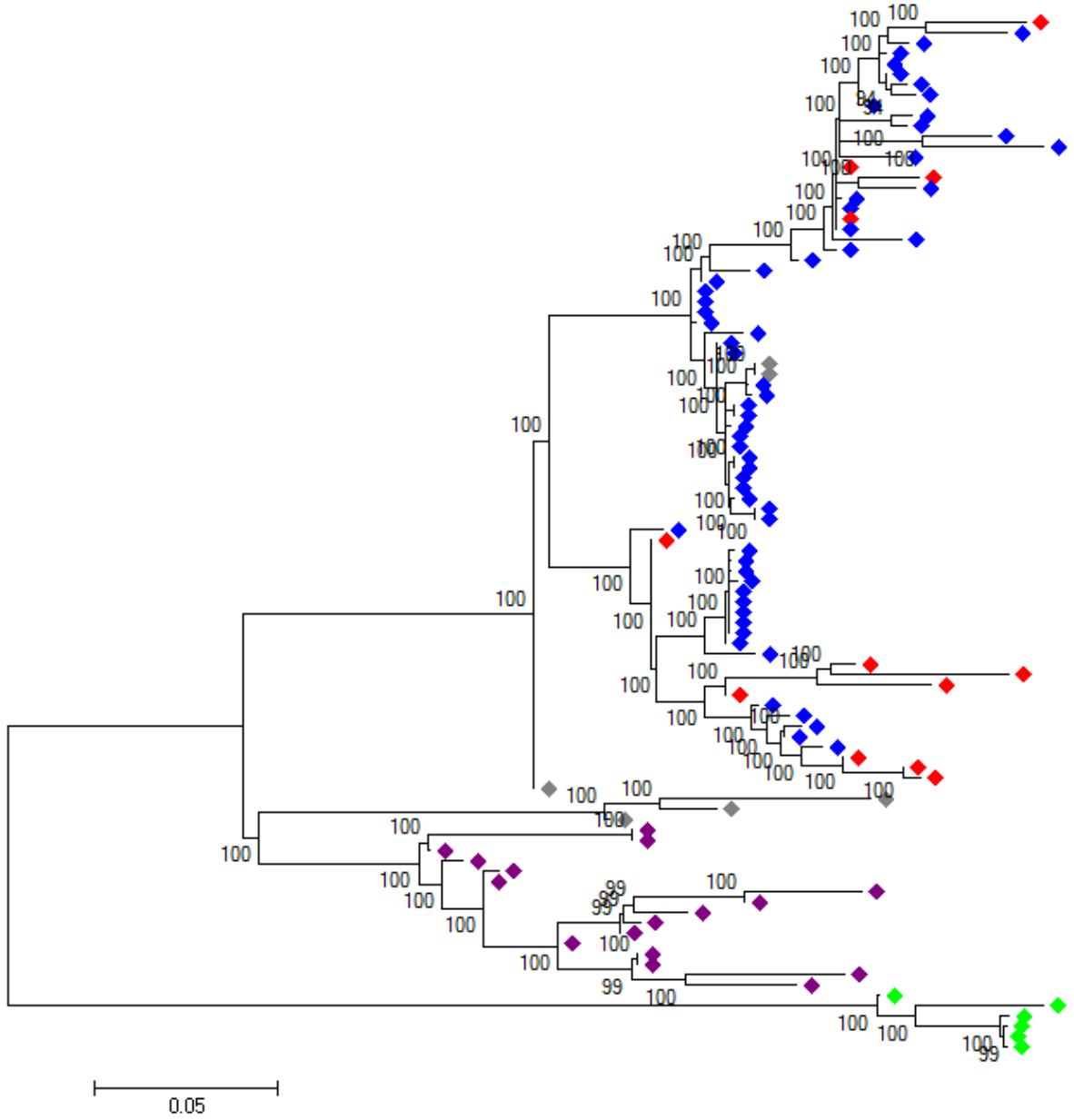


Figure S7.56 Chromosome 1: 149920001 - 149955001

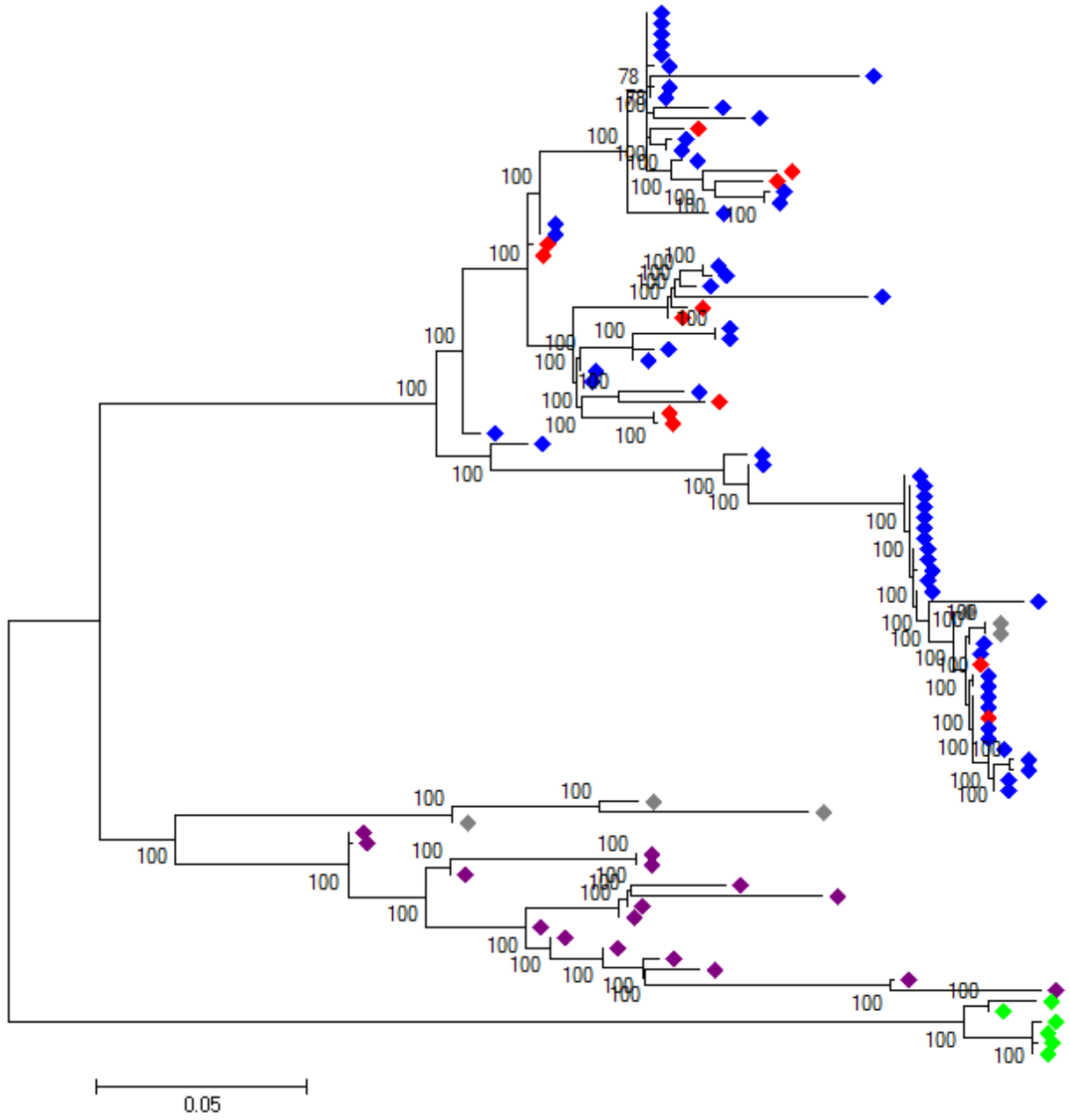


Figure S7.57 Chromosome 1: 150000001 - 150035001

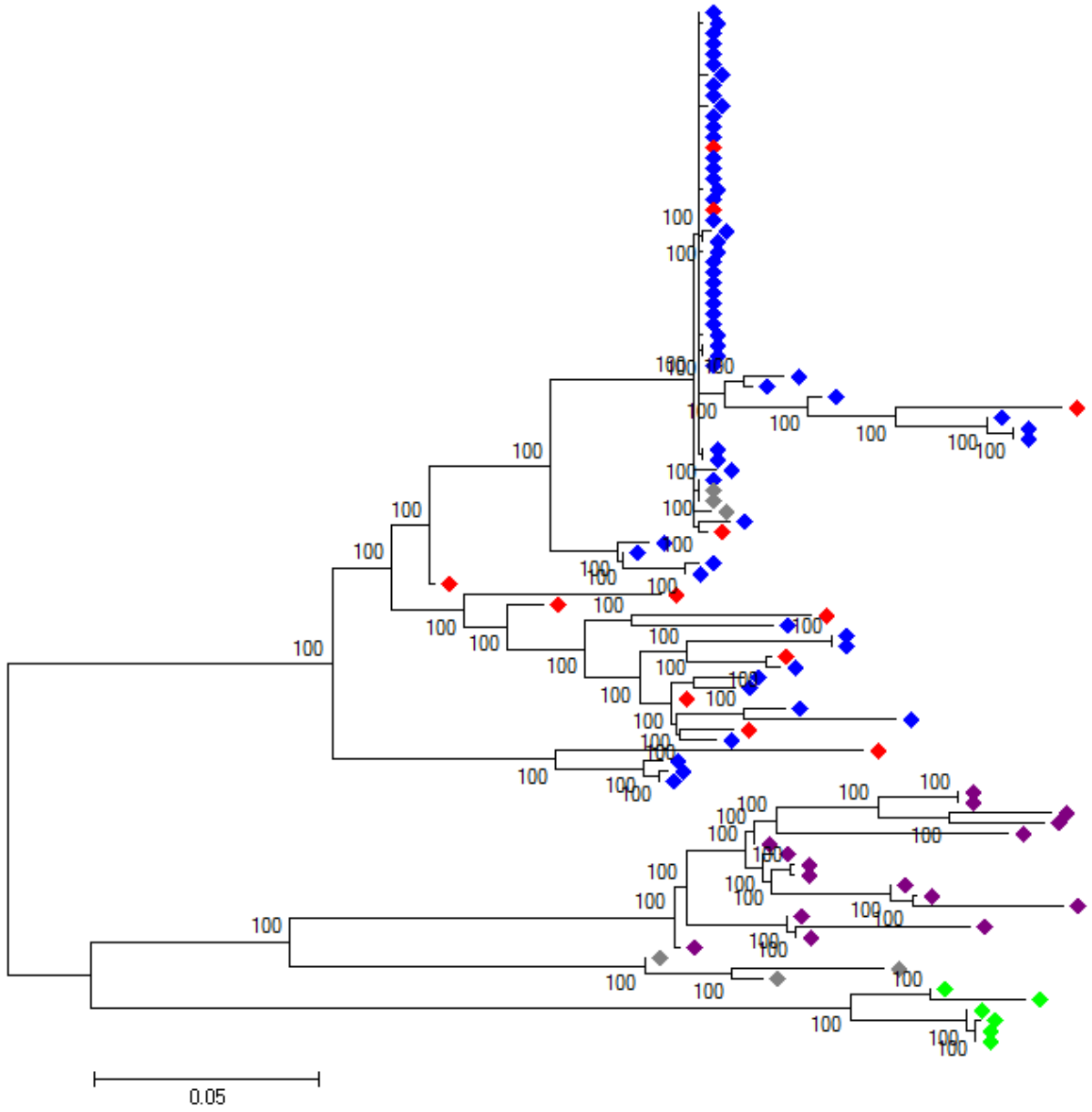


Figure S7.58 Chromosome 1: 150310001 - 150345001

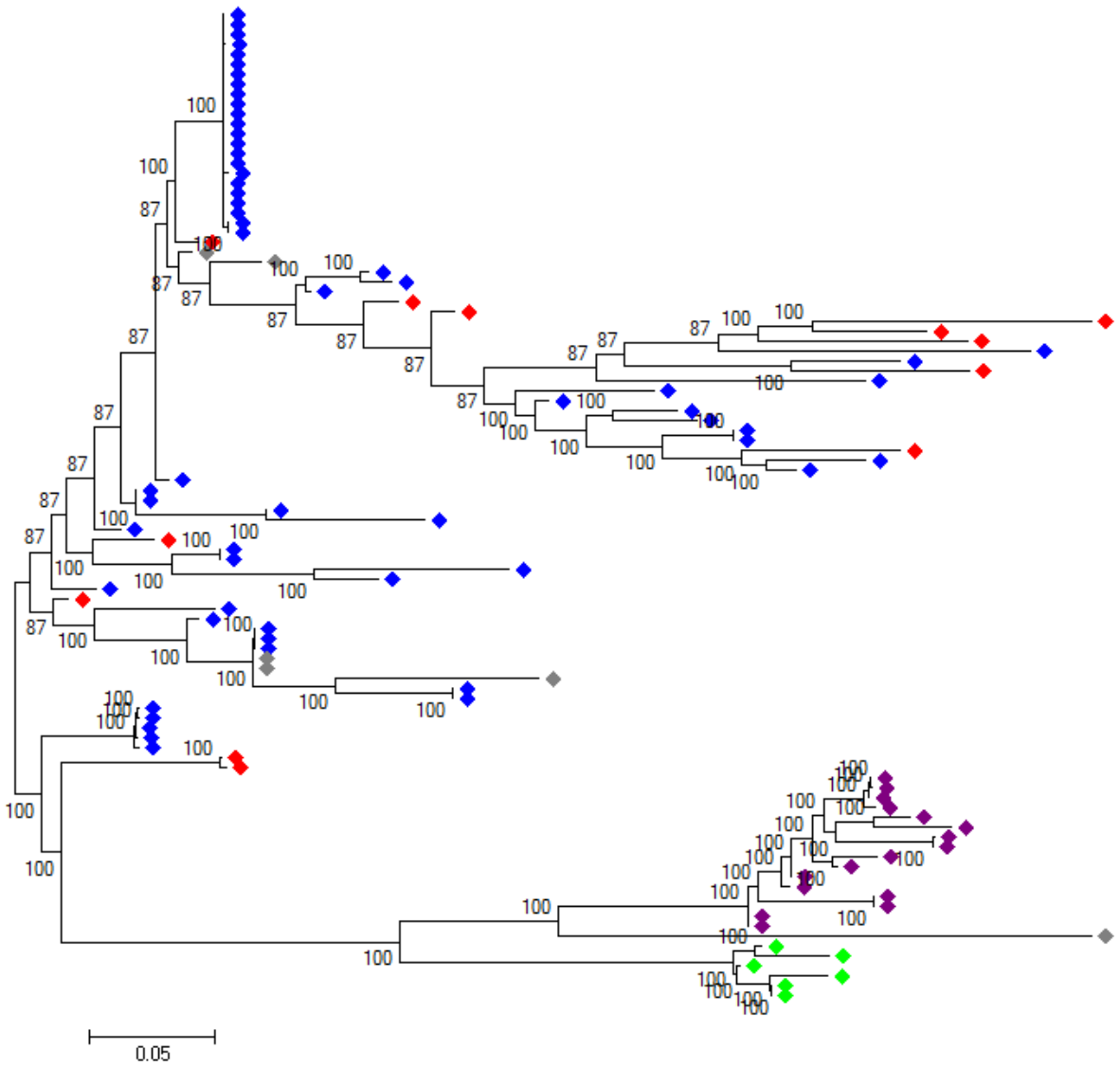


Figure S7.59 Chromosome 1: 153515001 - 153545001

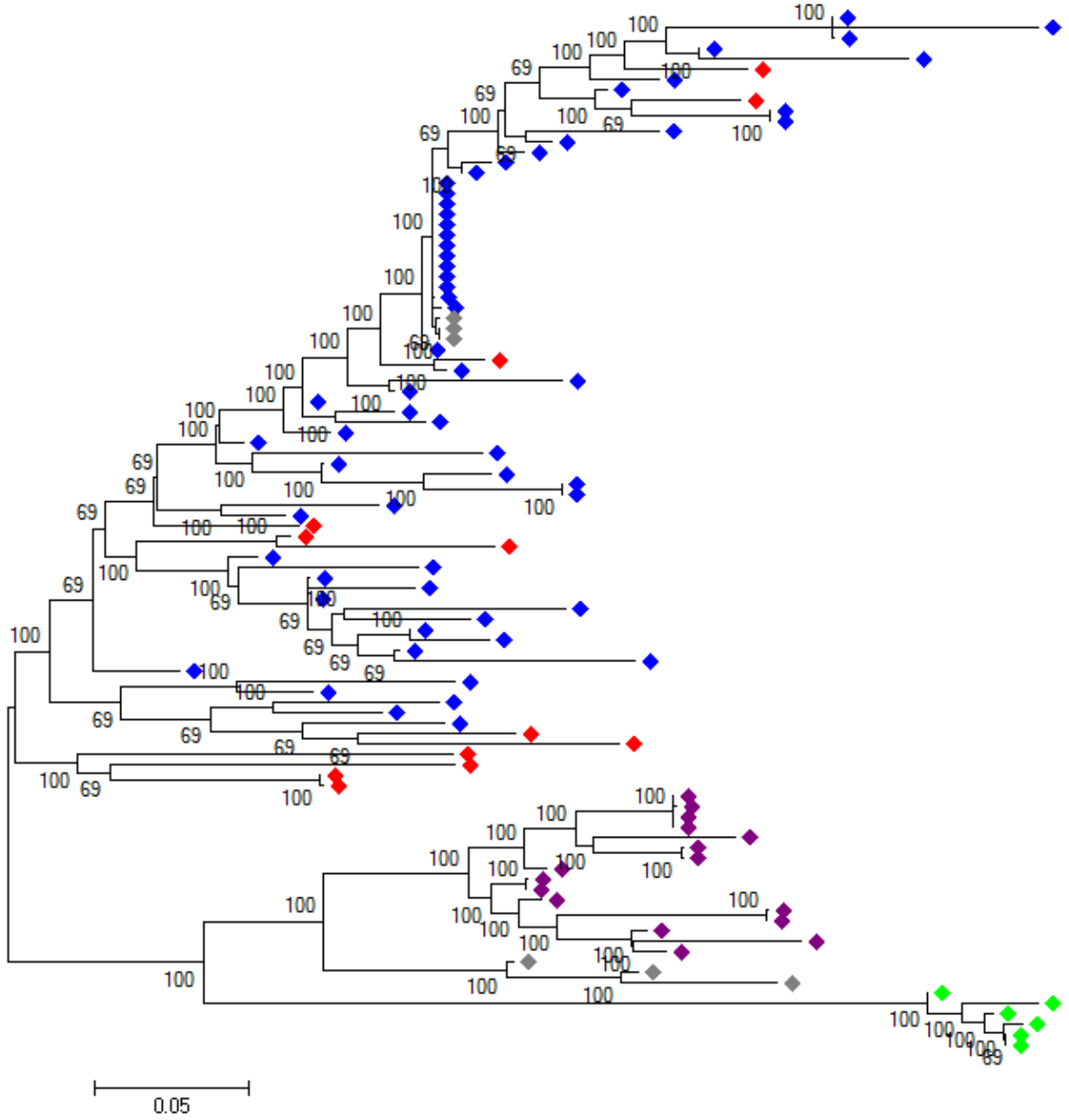


Figure S7.60 Chromosome 1: 154080001 - 154125001

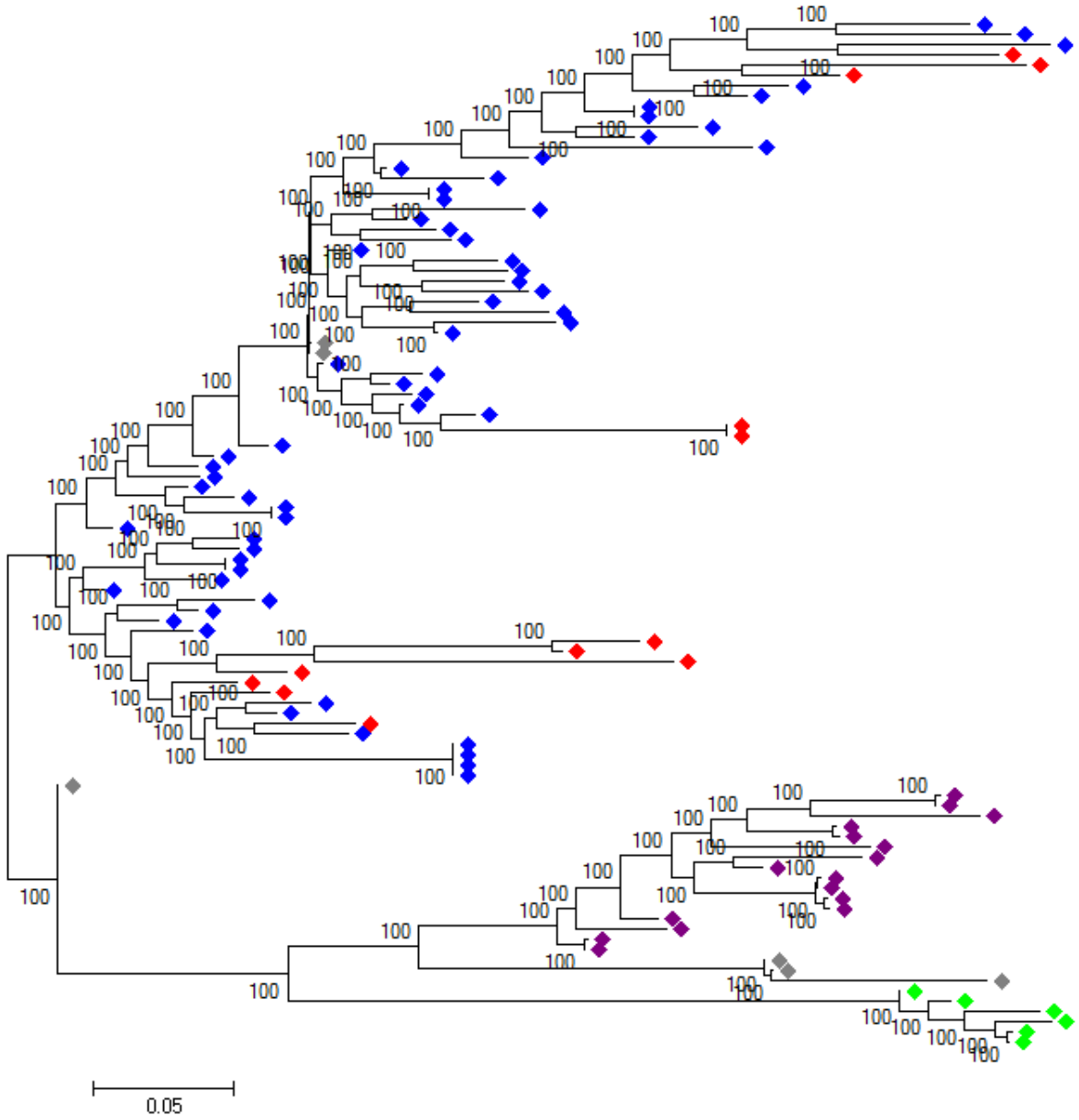


Figure S7.61 Chromosome 1: 154995001 - 155075001

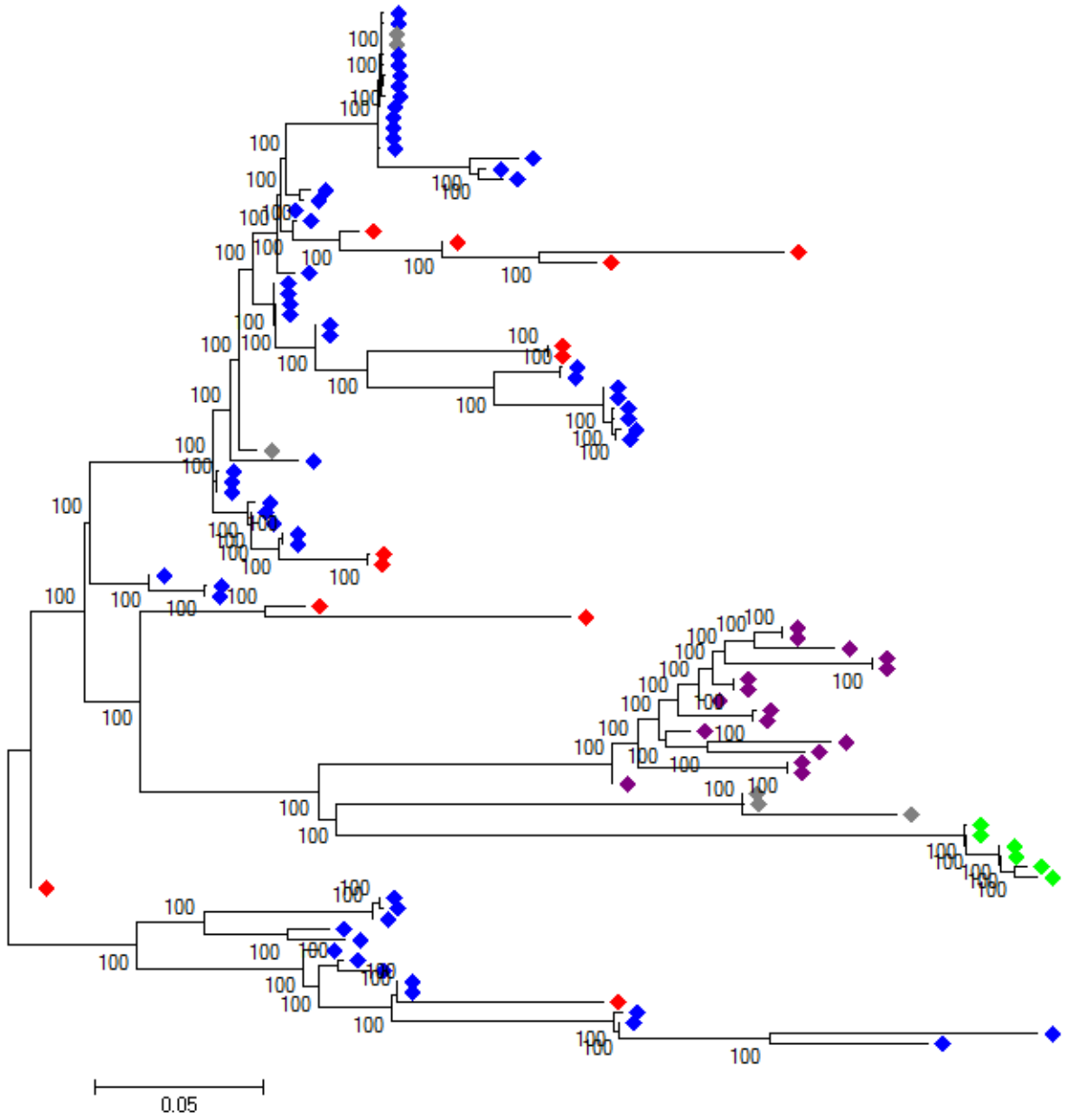


Figure S7.62 Chromosome 1: 156280001 - 156340001

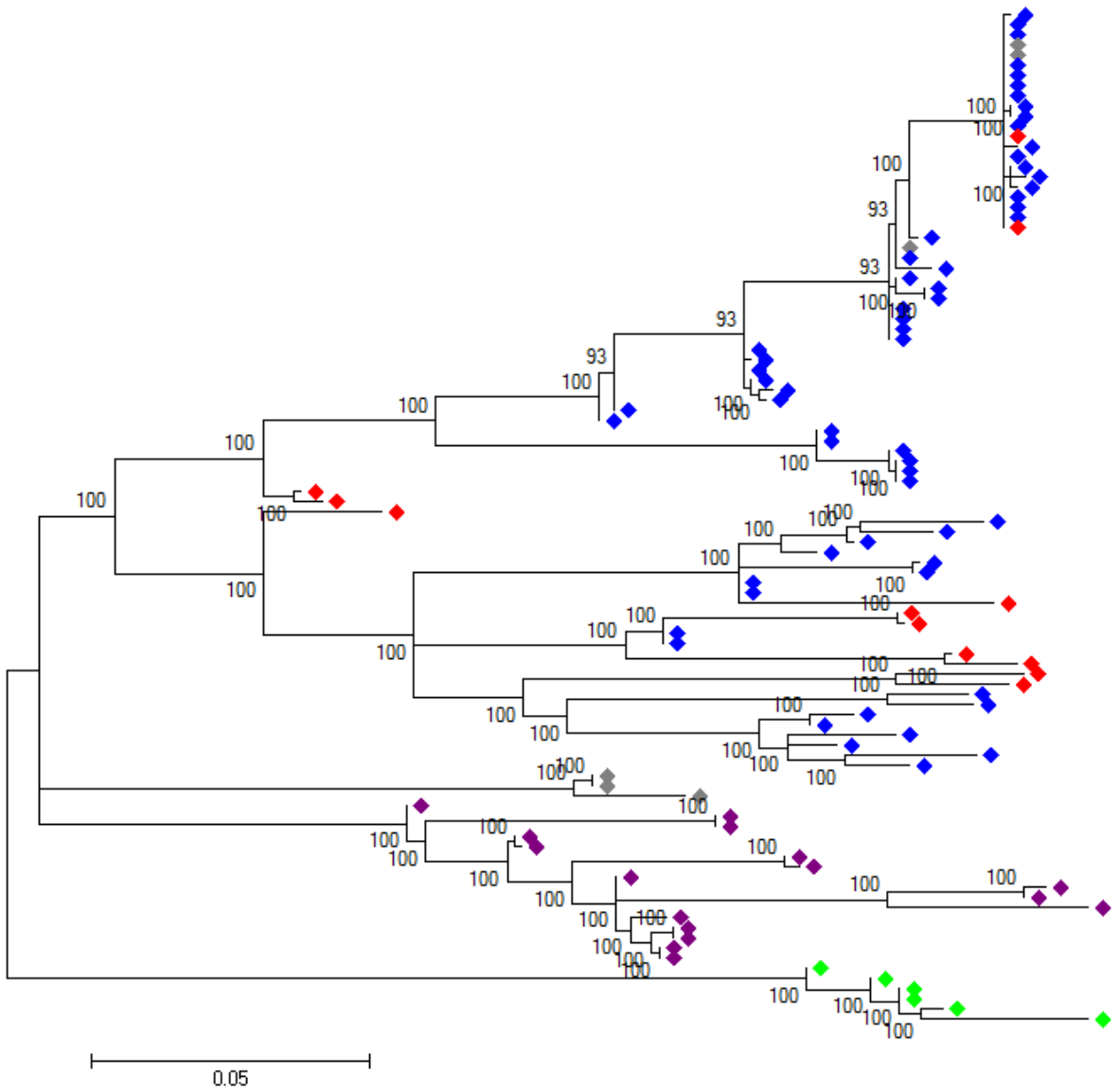


Figure S7.63 Chromosome 1:156375001 - 156400001

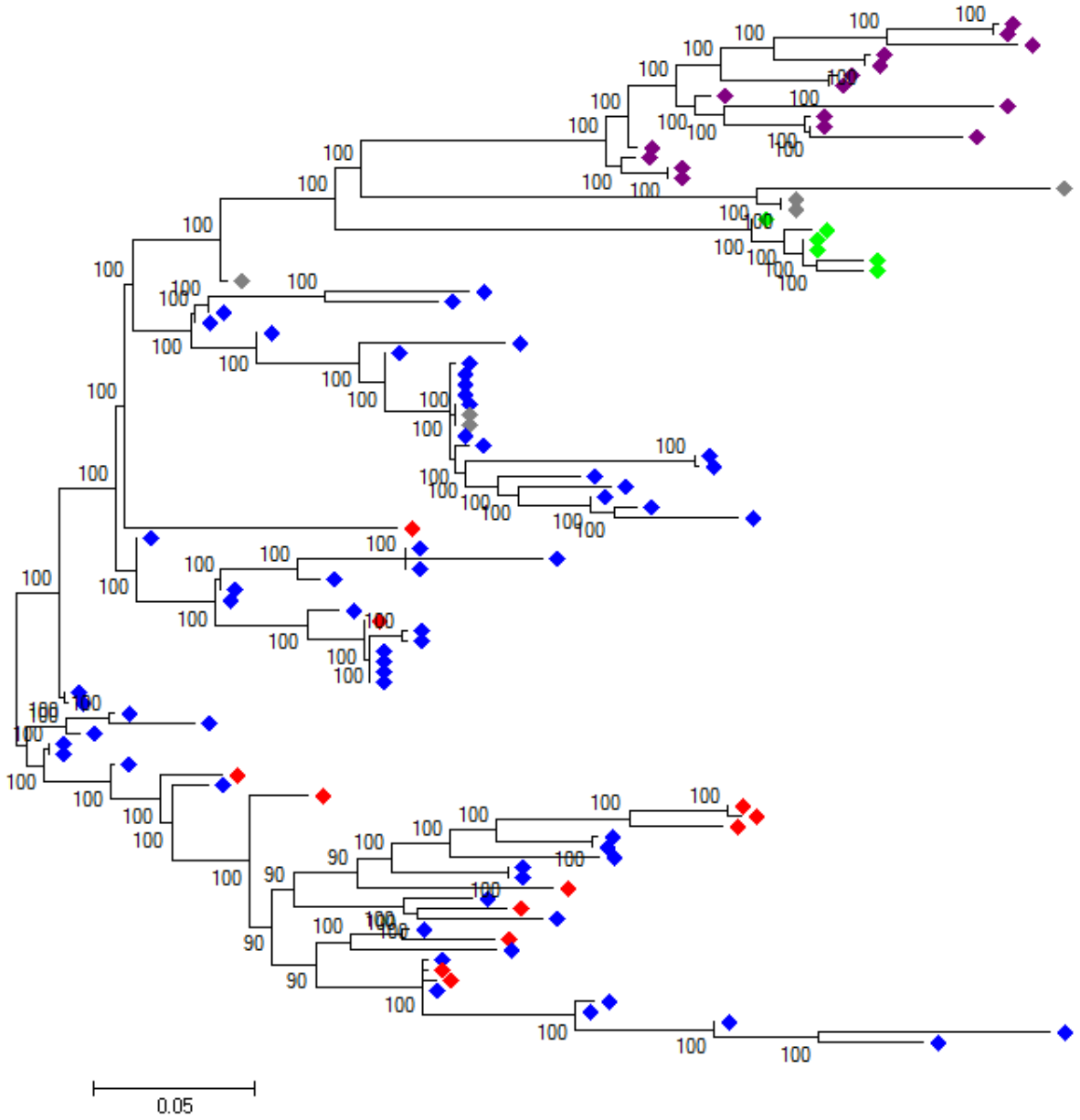


Figure S7.64 Chromosome 1: 156840001 - 156860001

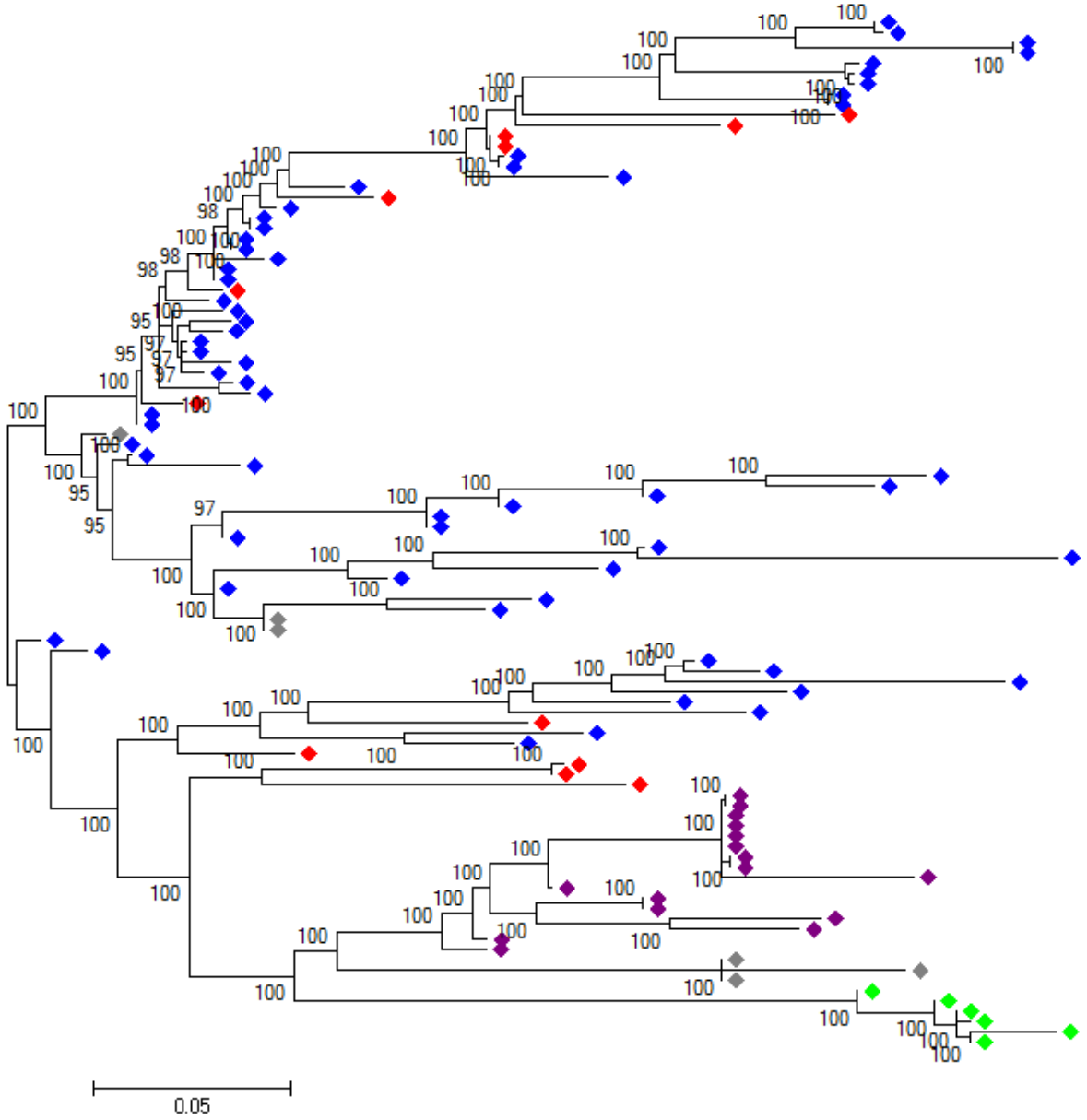


Figure S7.65 Chromosome 1: 157065001 - 157090001

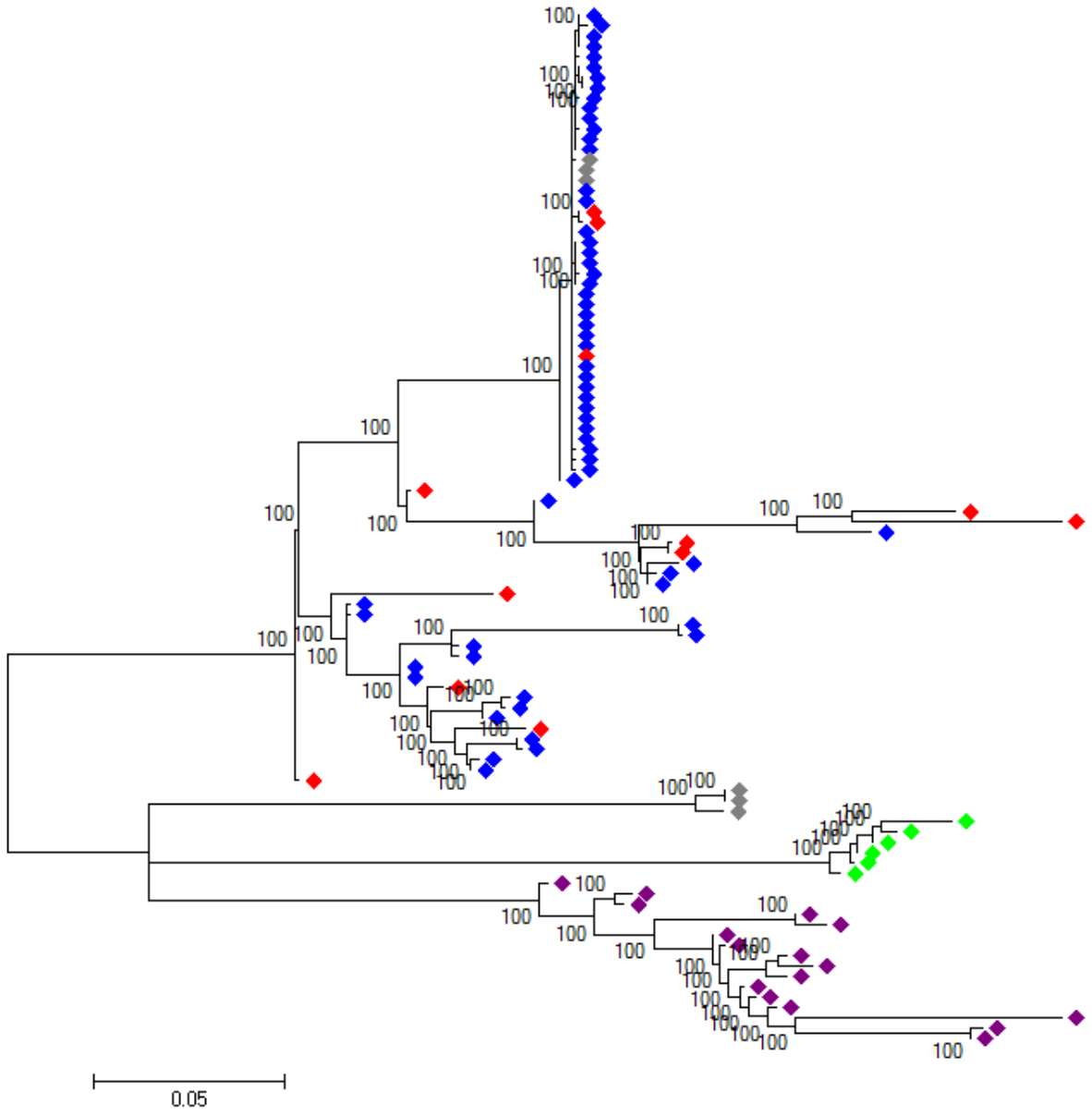


Figure S7.66 Chromosome 1: 160175001 - 160205001

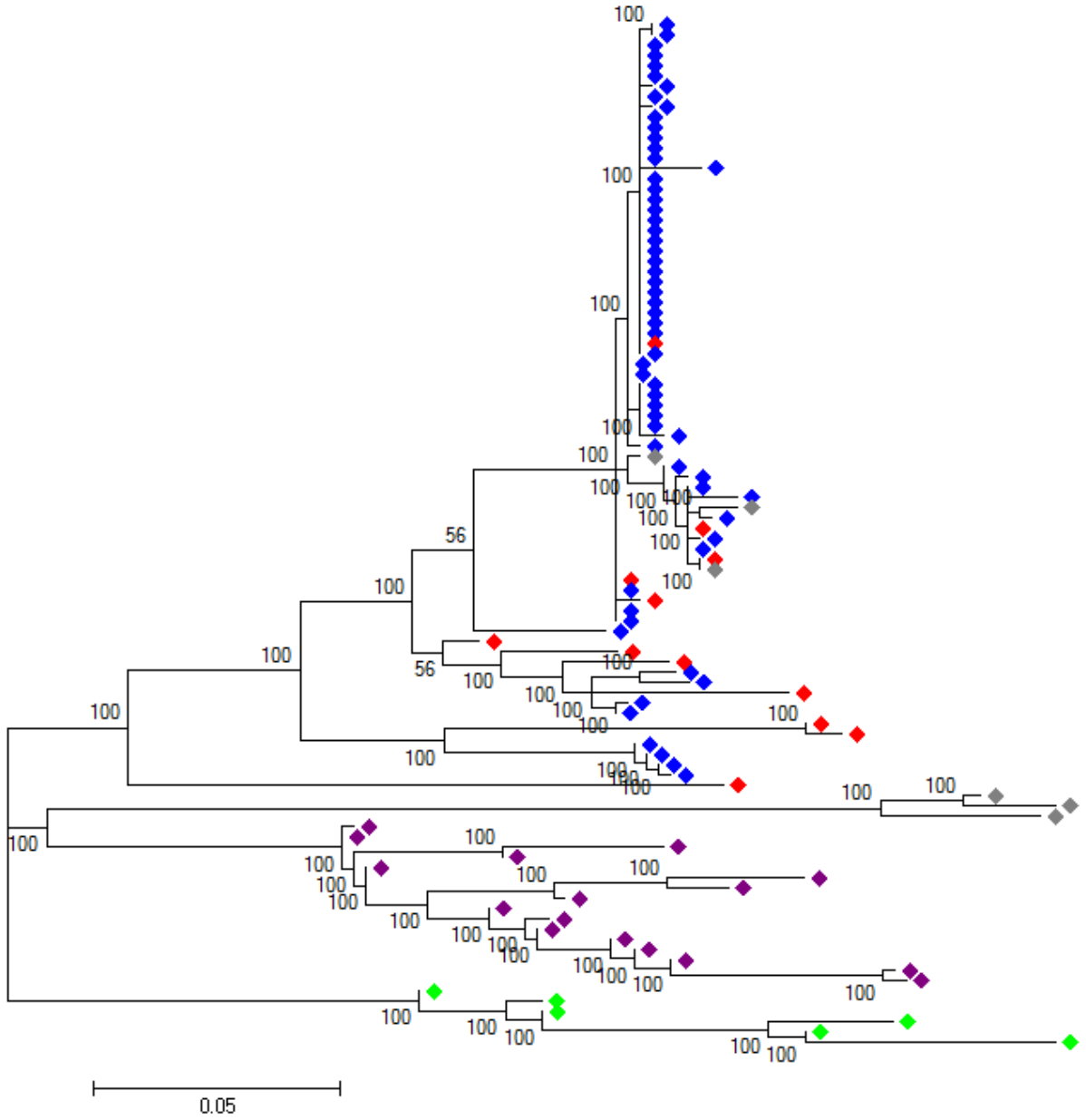


Figure S7.67 Chromosome 1: 160410001 - 160430001

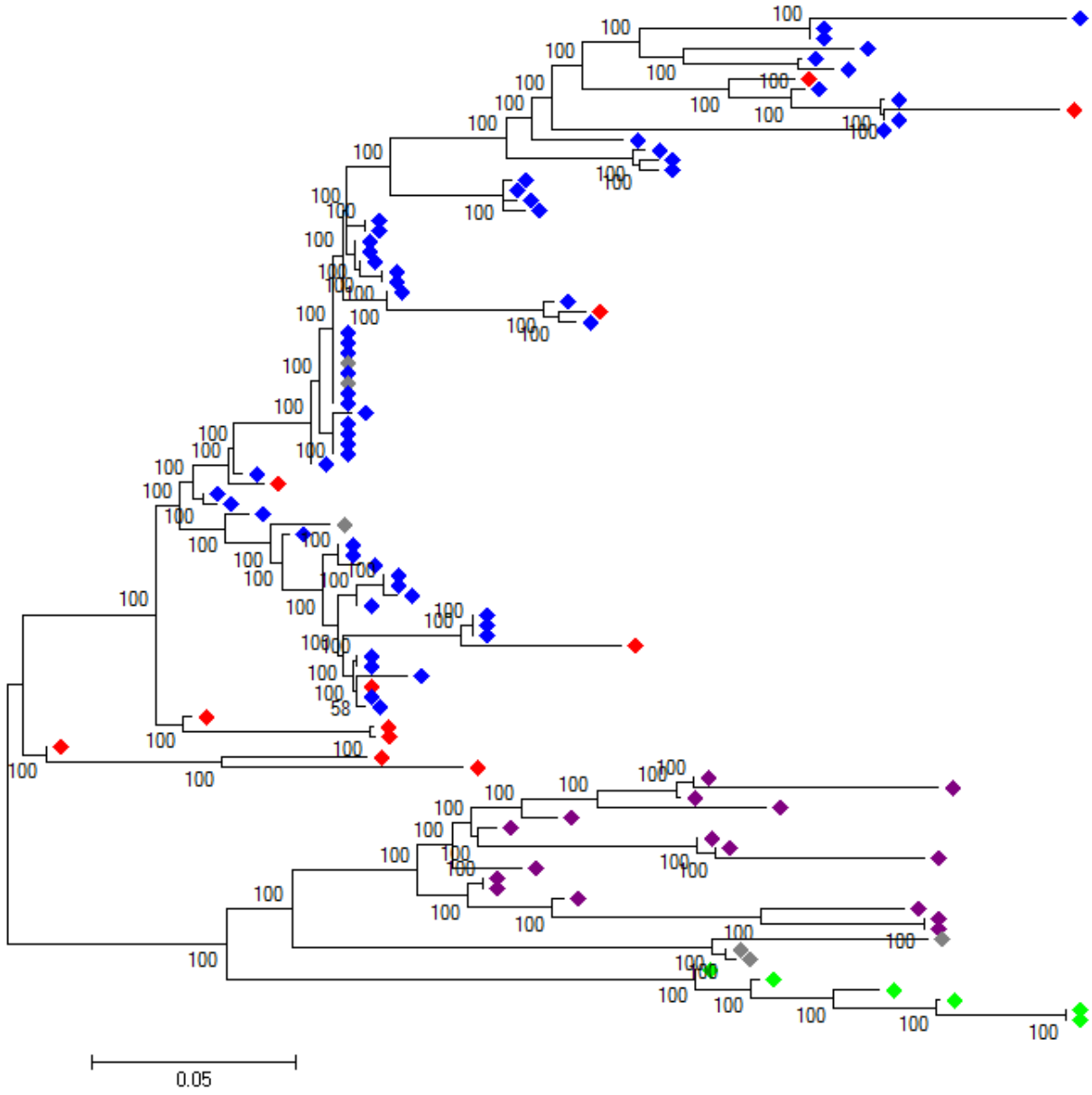


Figure S7.68 Chromosome 1: 161530001 - 161550001

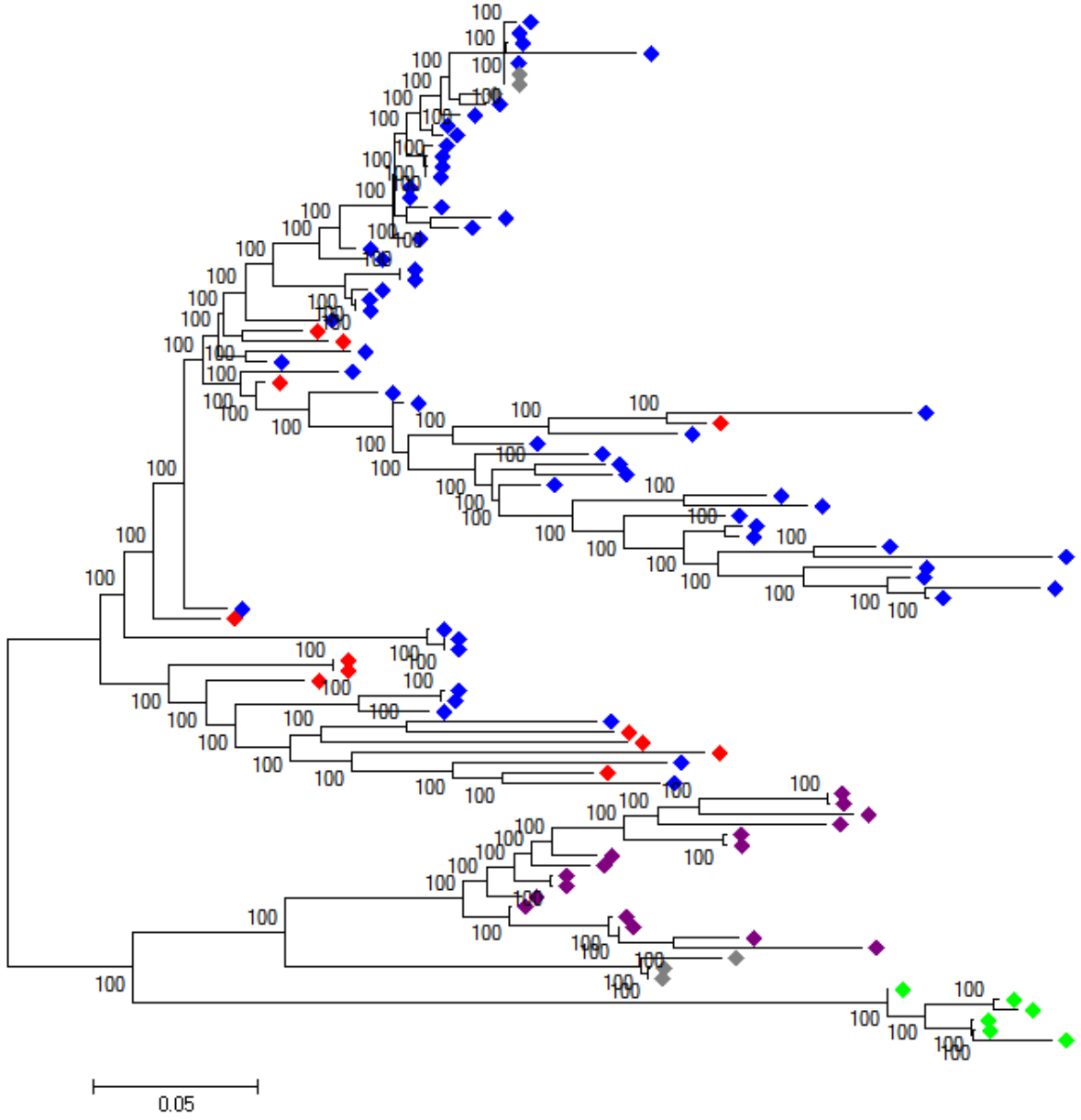


Figure S7.69 Chromosome 1: 161585001 - 161660001

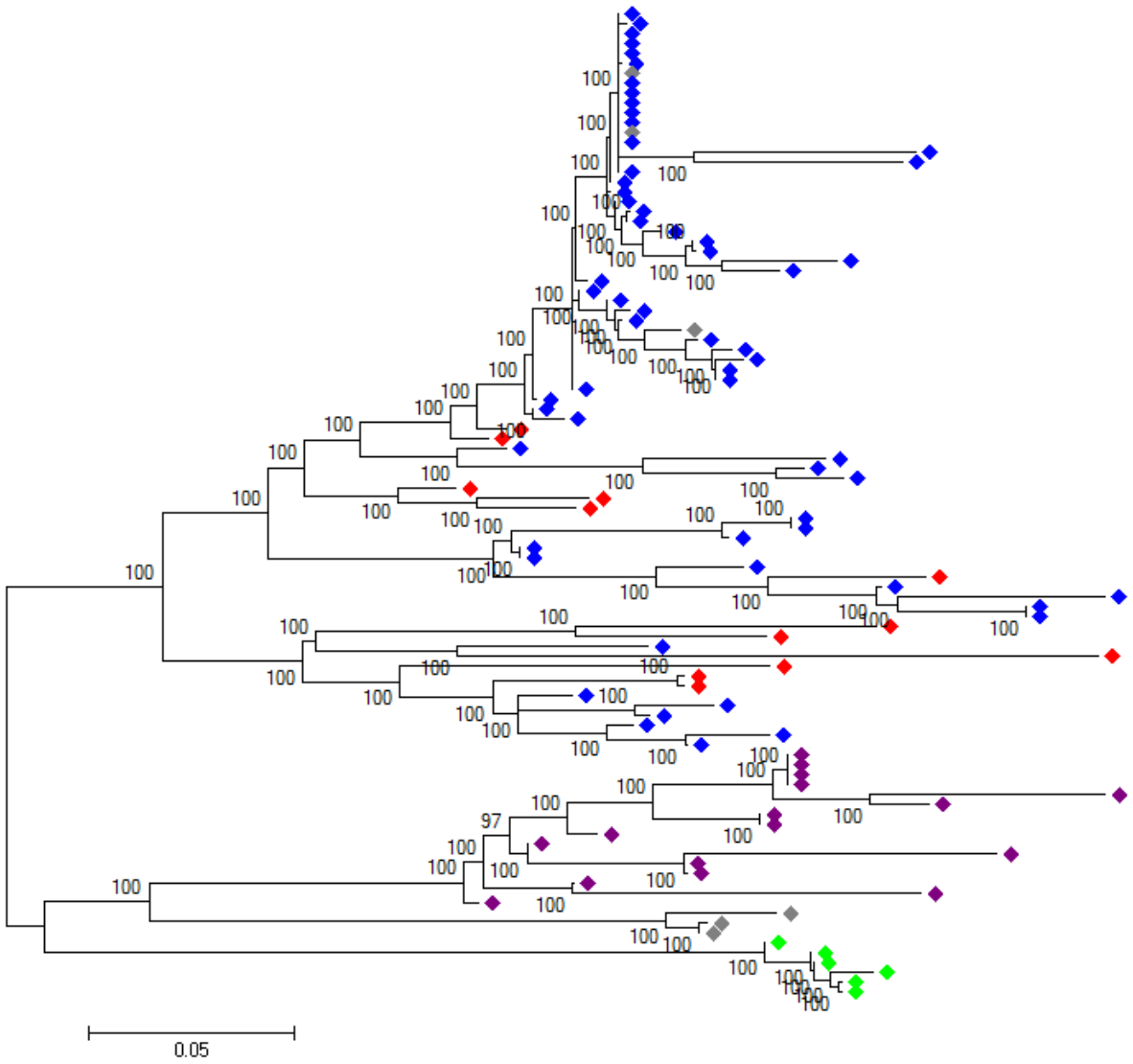


Figure S7.70 Chromosome 1: 161690001 - 161715001

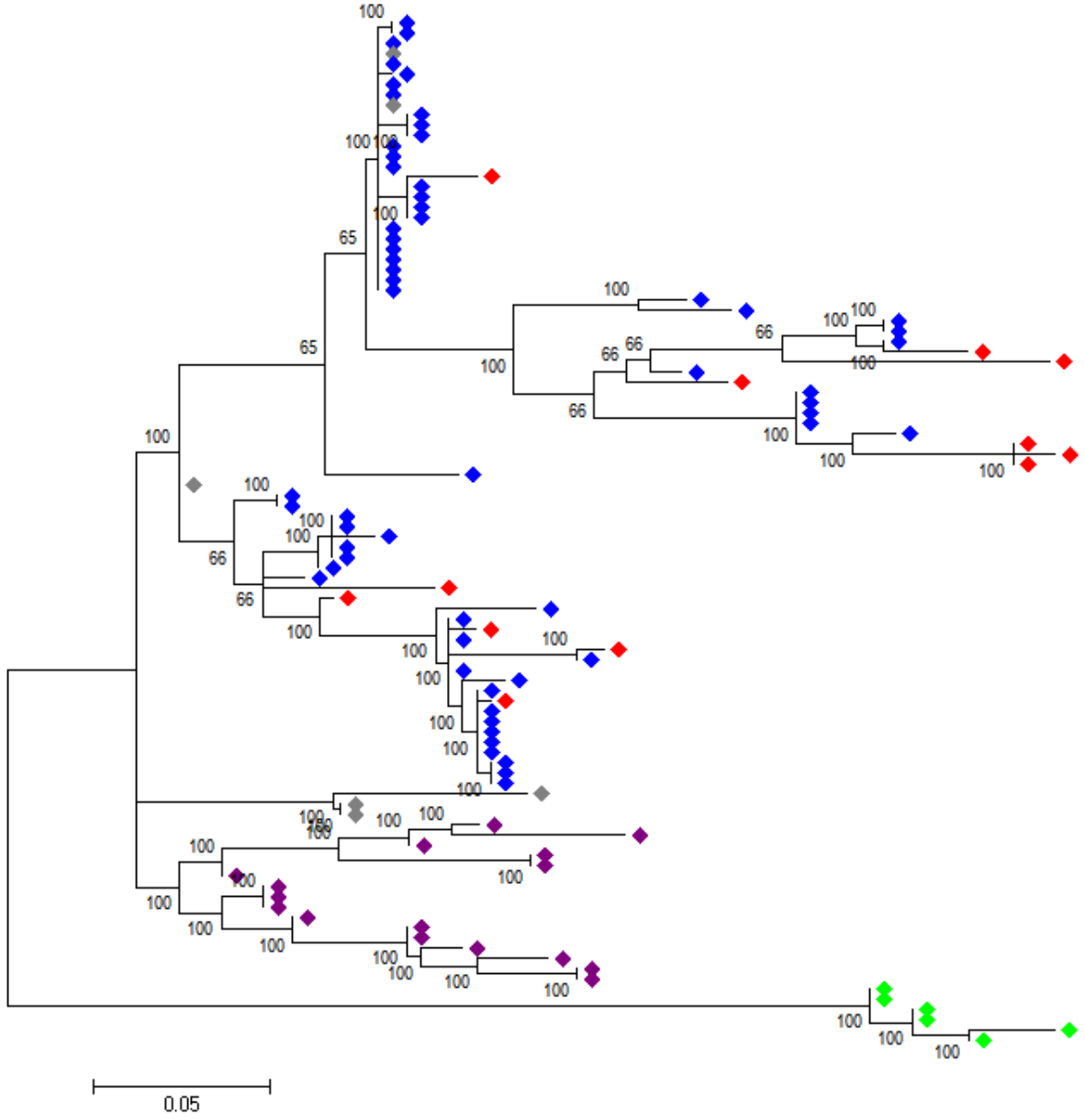


Figure S7.71 Chromosome 1: 161745001 - 161760001

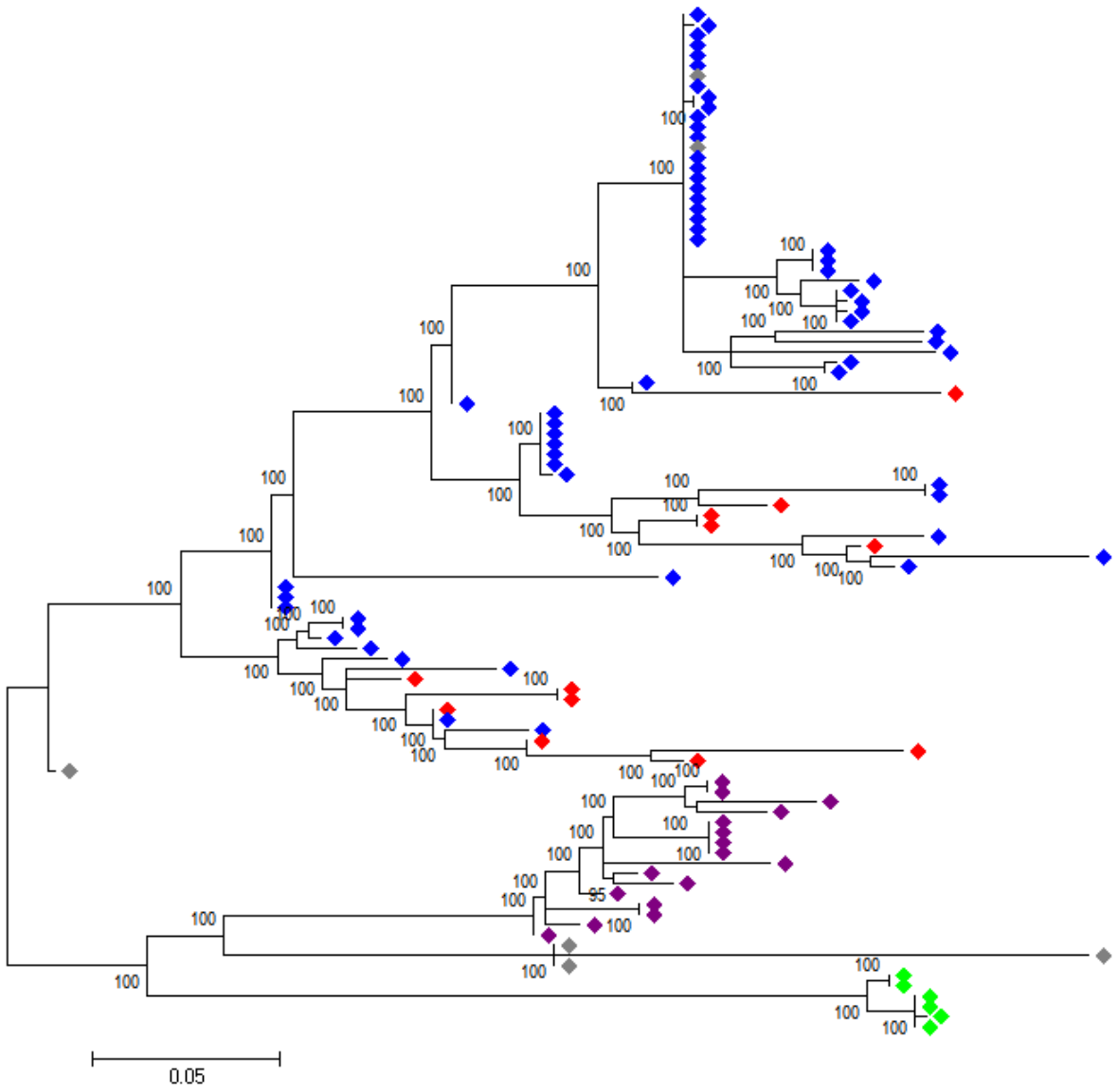


Figure S7.72 Chromosome 1: 161815001 - 161840001

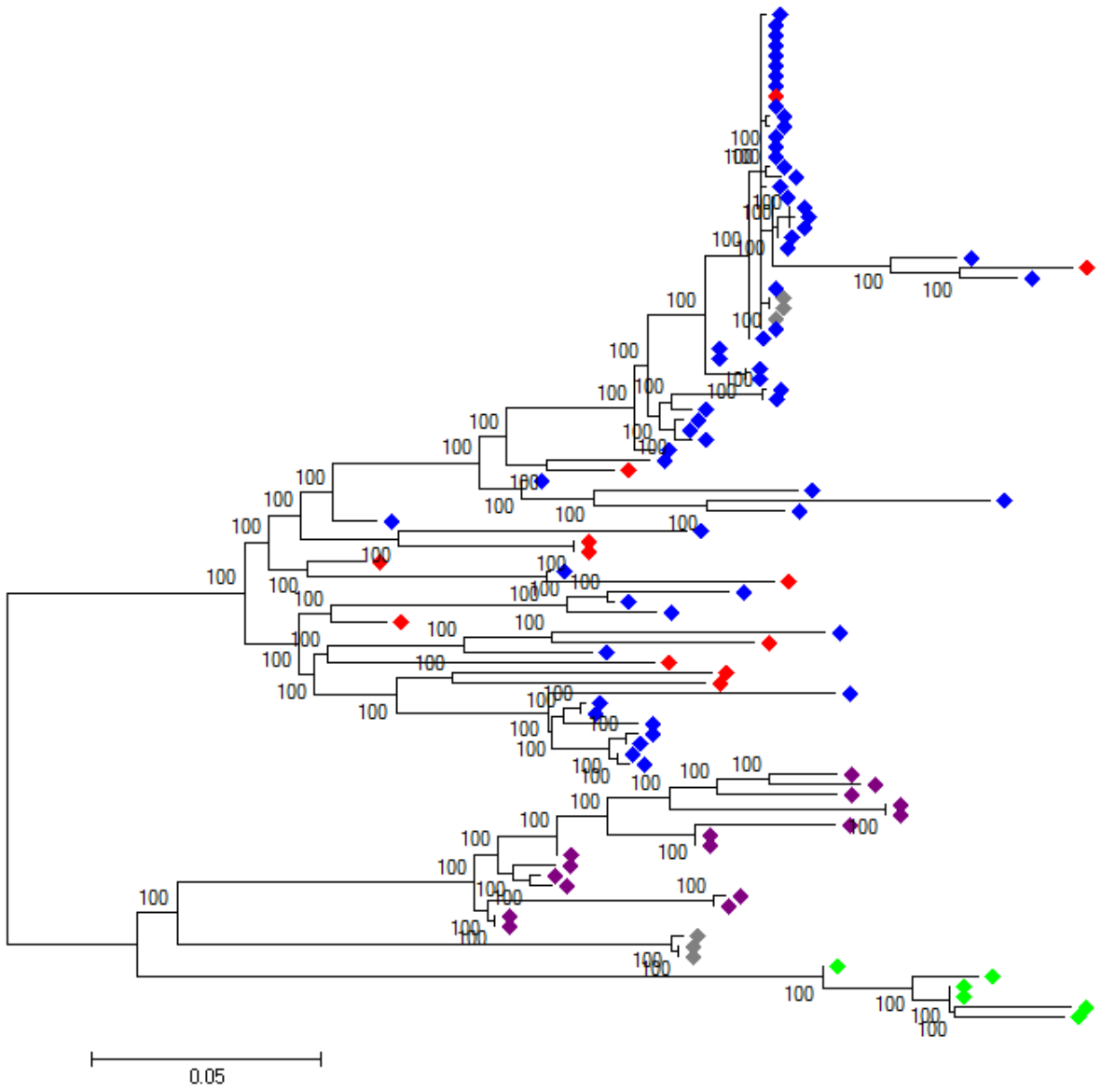


Figure S7.73 Chromosome 1: 161920001 - 161960001

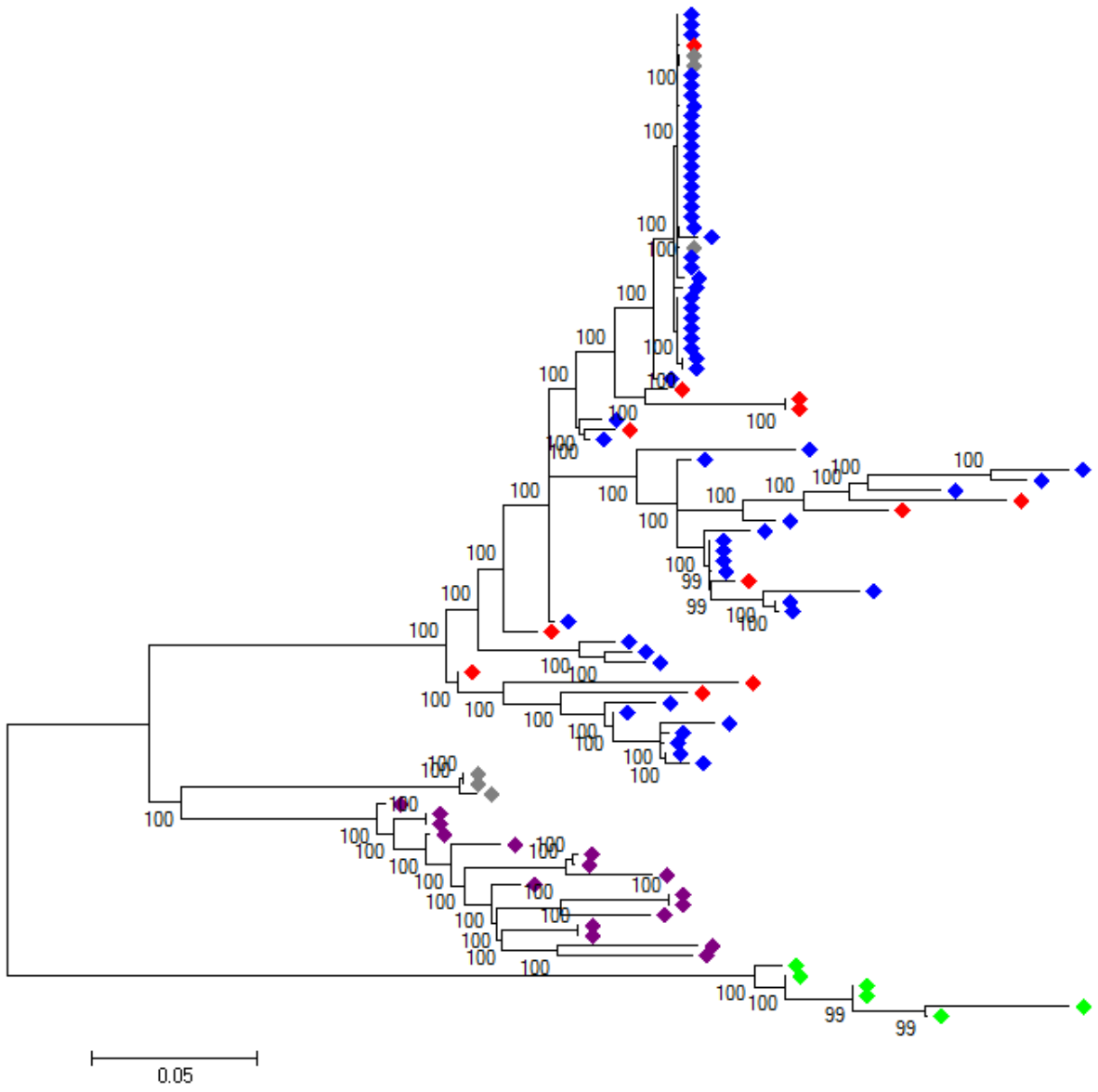


Figure S7.74 Chromosome 1: 164095001 - 164135001

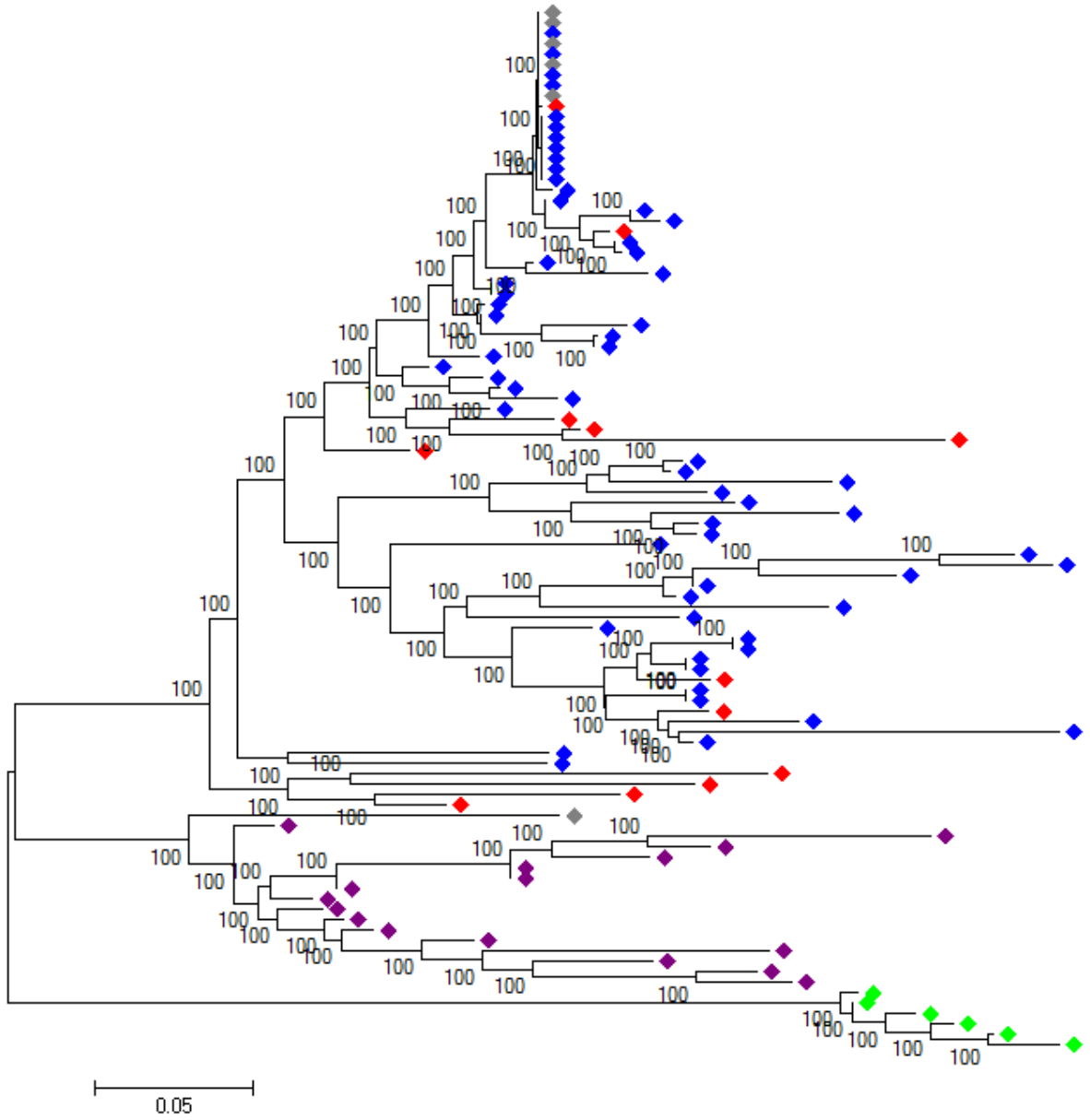


Figure S7.75 Chromosome 1: 194975001 - 194995001

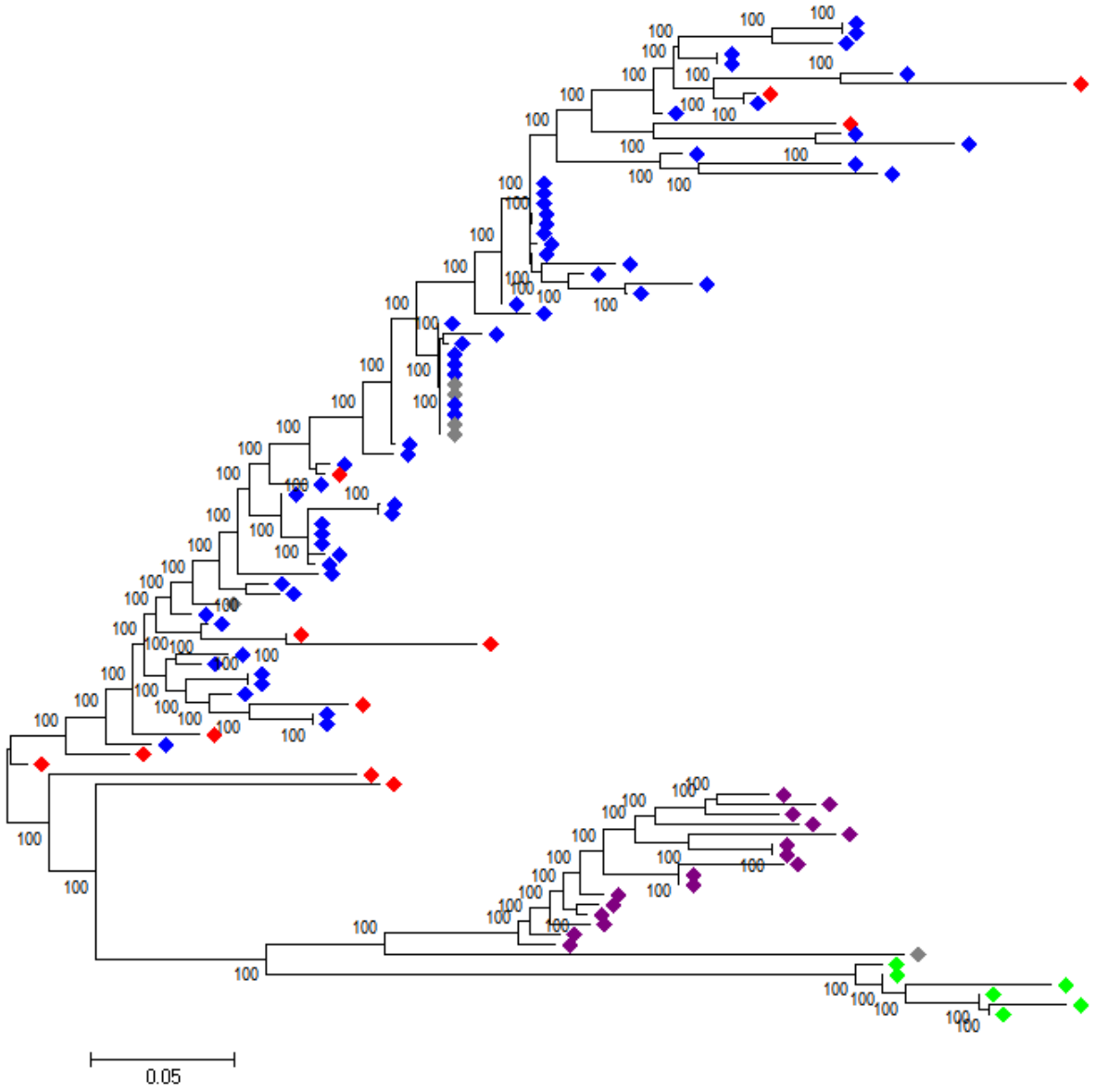


Figure S7.76 Chromosome 2: 10525001 - 10560001

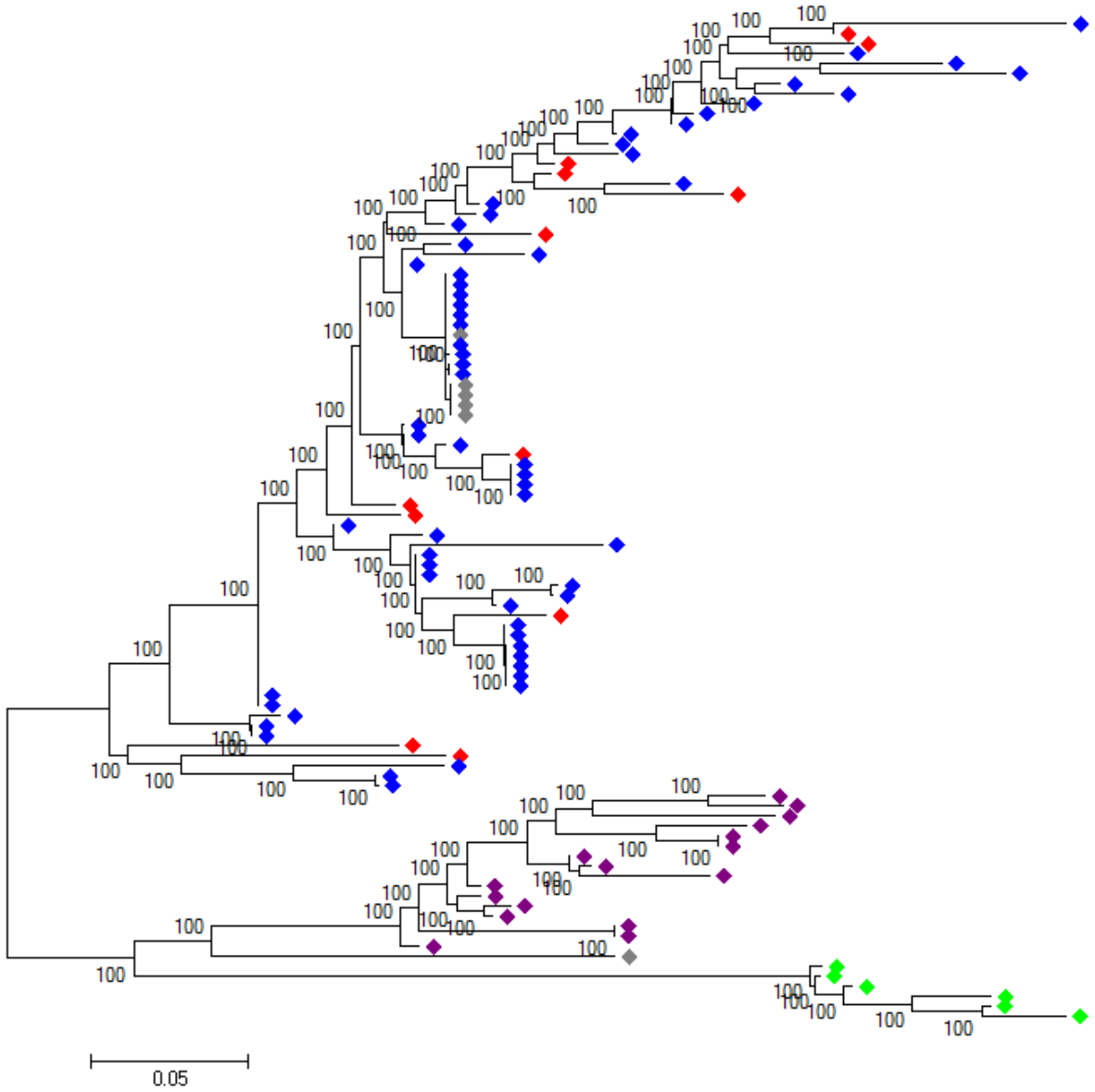


Figure S7.77 Chromosome 2: 10665001 - 10710001

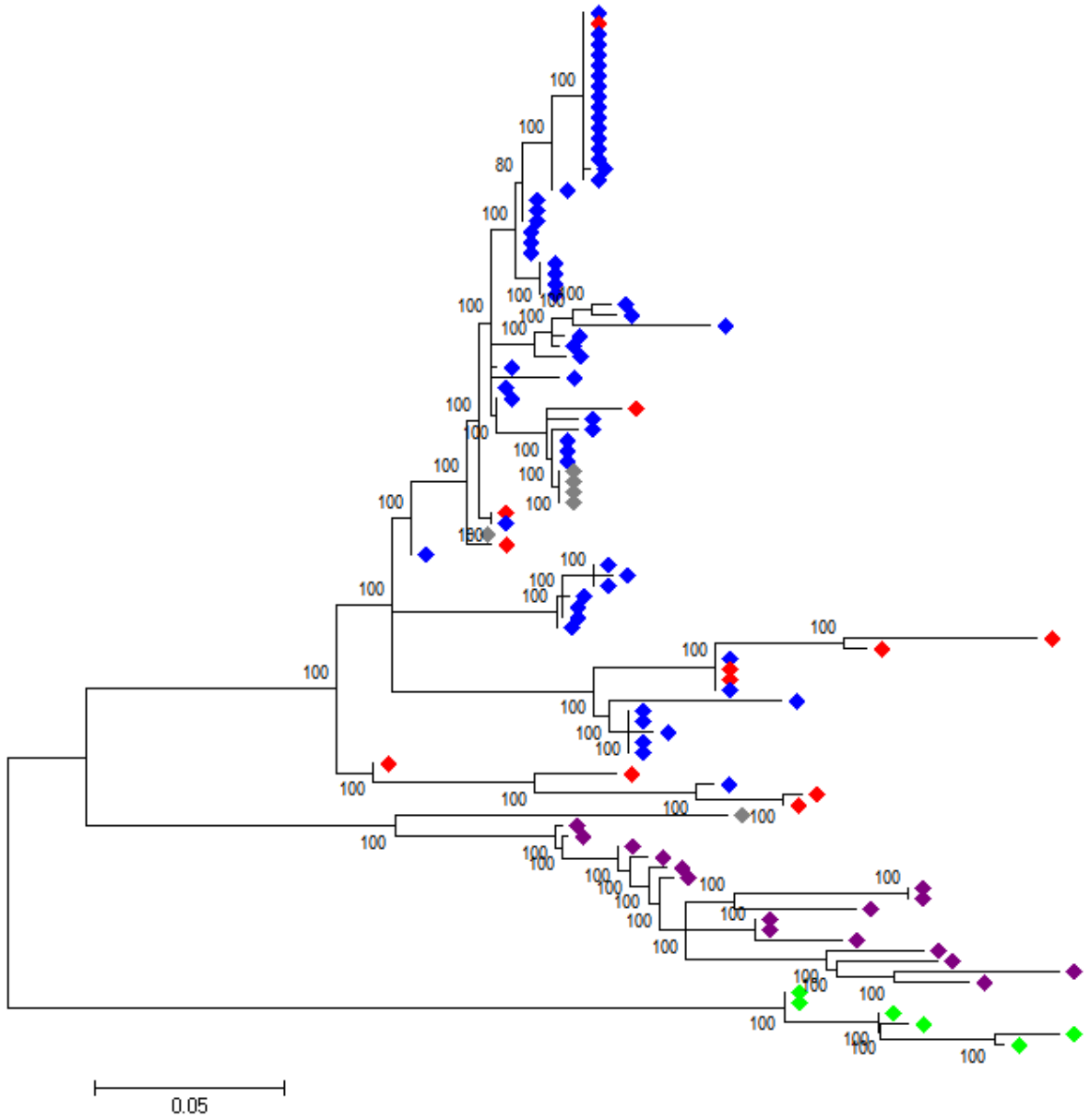


Figure S7.78 Chromosome 2: 10890001 - 10910001

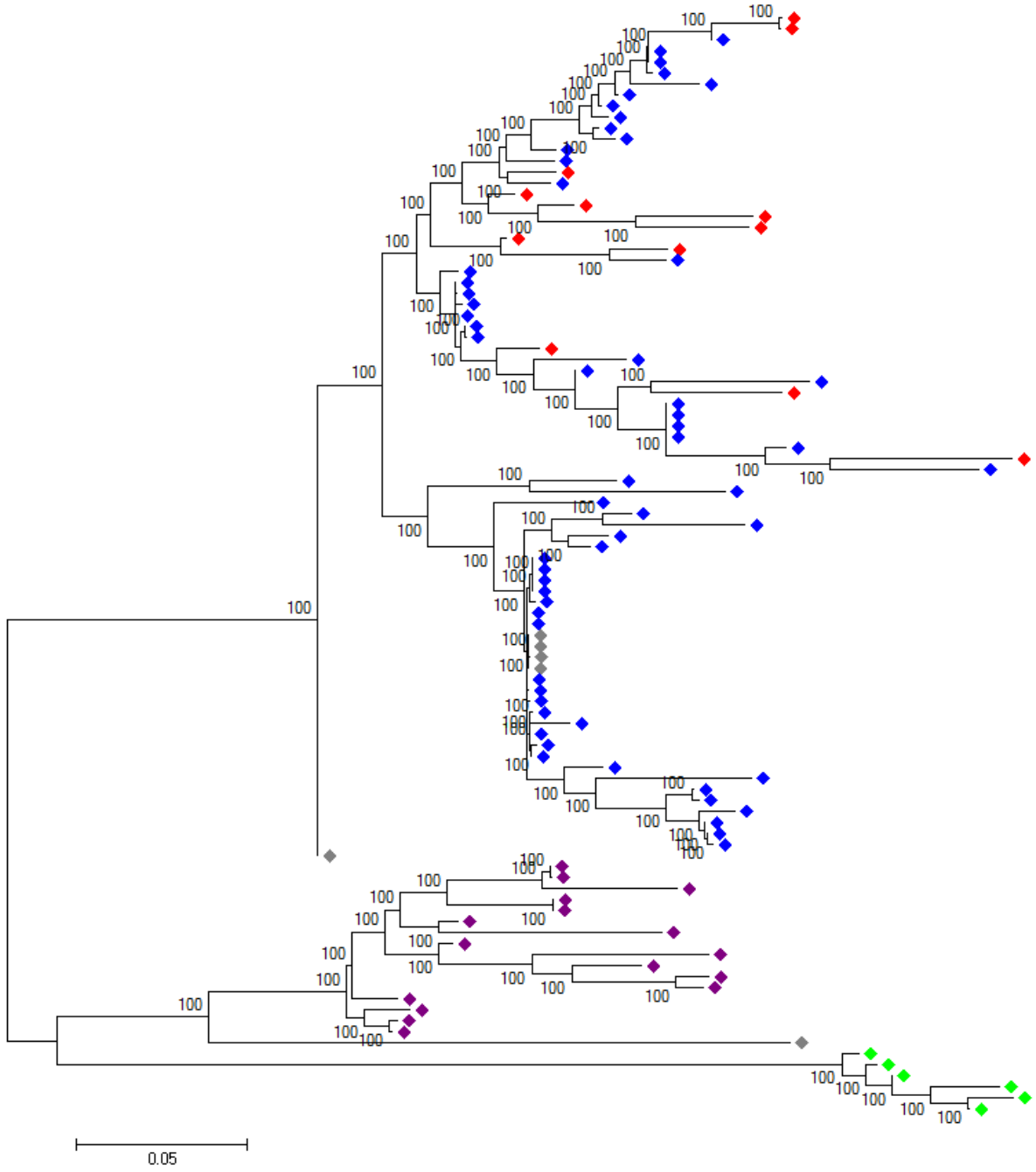


Figure S7.79 Chromosome 2: 11115001 - 11275001

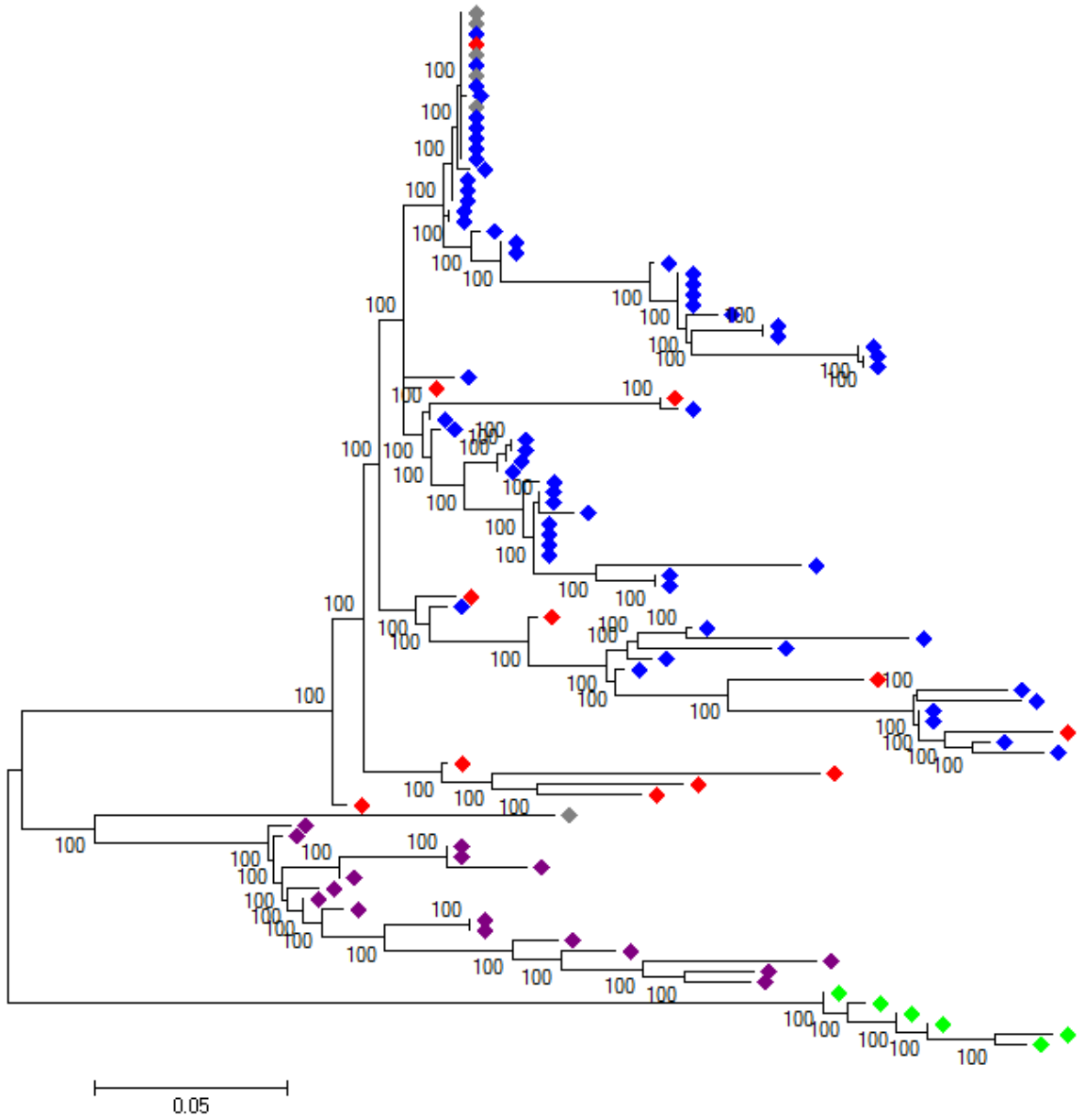


Figure S7.80 Chromosome 2: 11435001 - 11460001

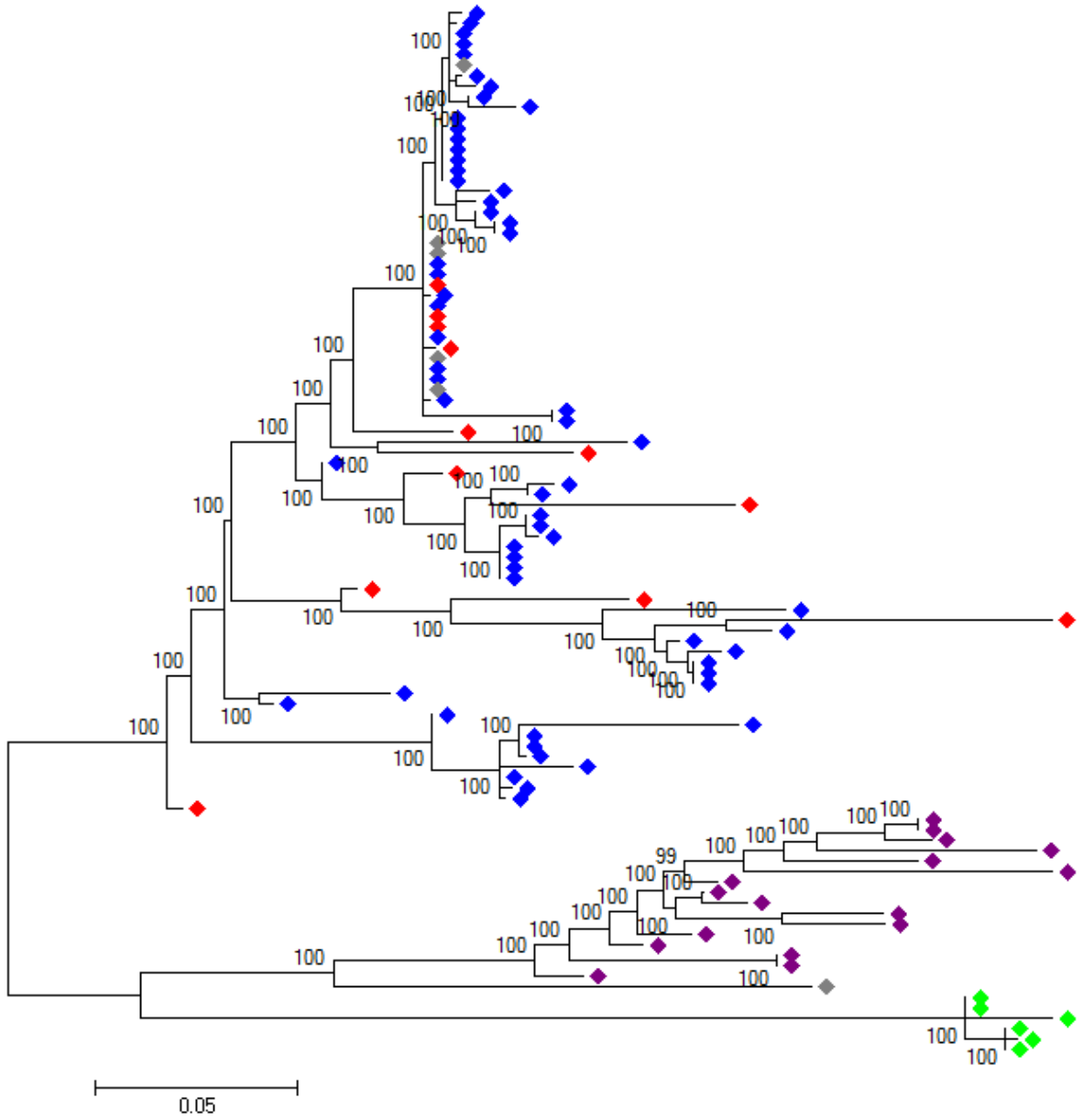


Figure S7.81 Chromosome 2: 11735001 - 11760001

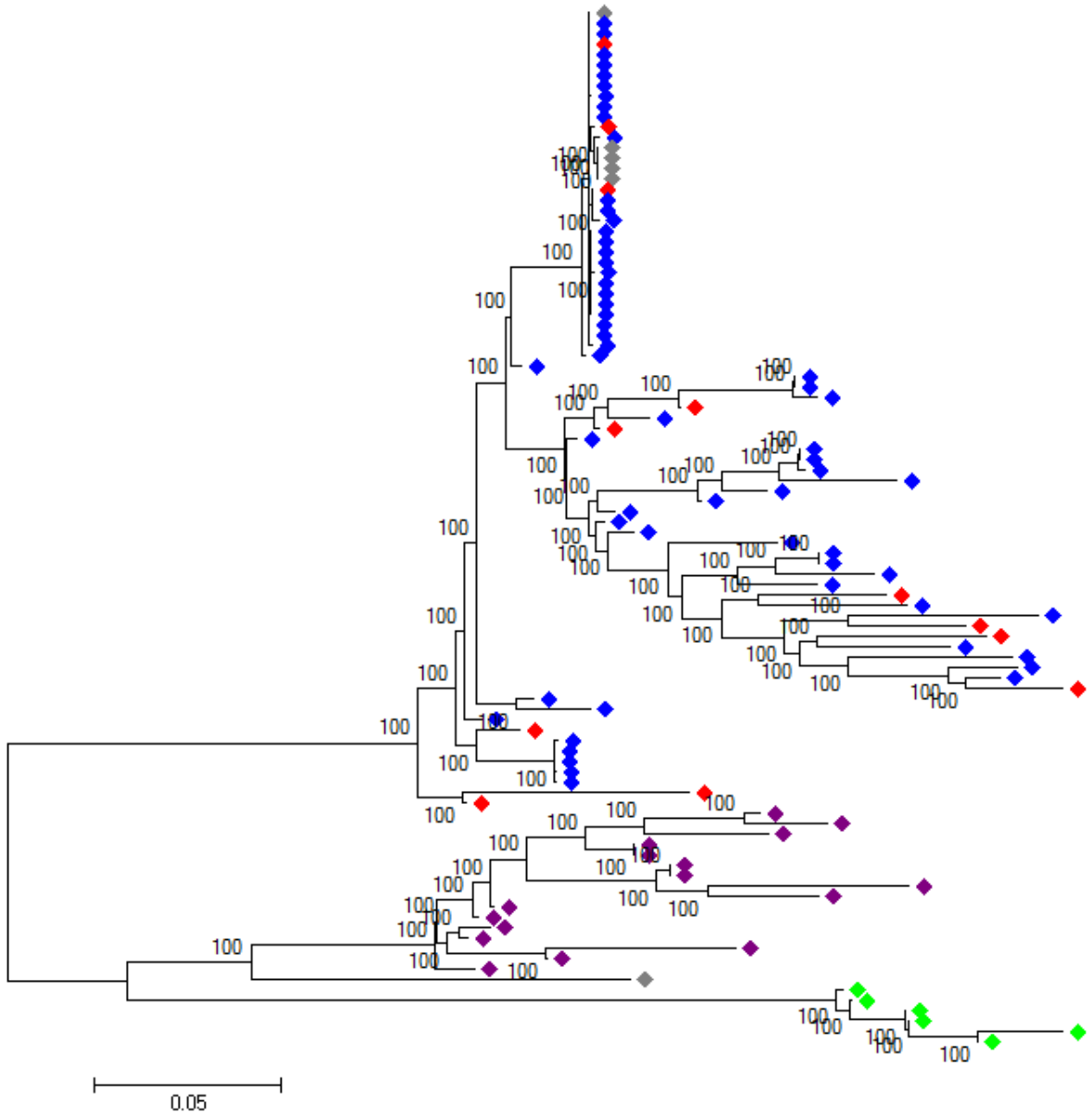


Figure S7.82 Chromosome 2: 11785001 - 11850001

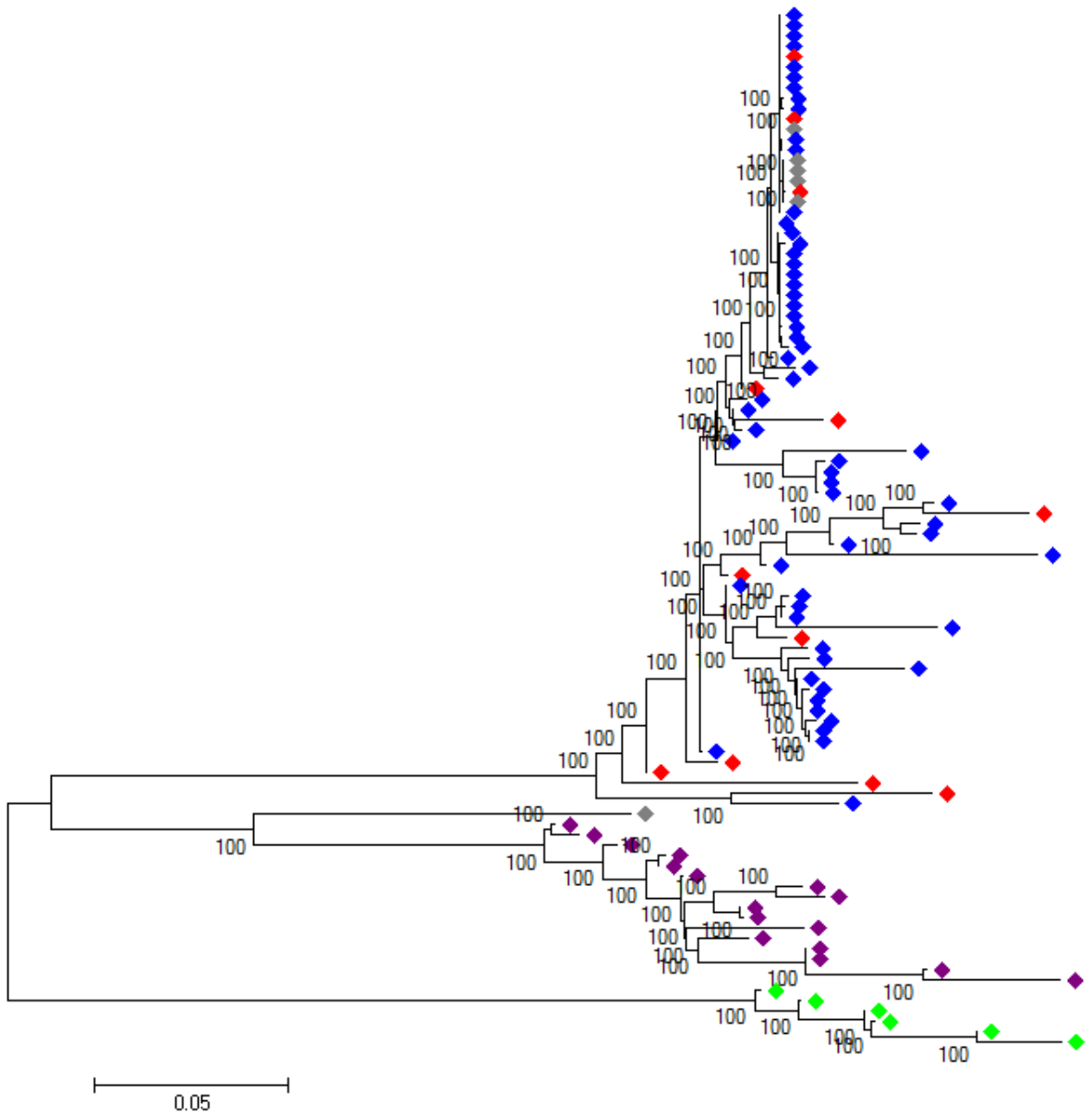


Figure S7.83 Chromosome 2: 11890001 - 11960001

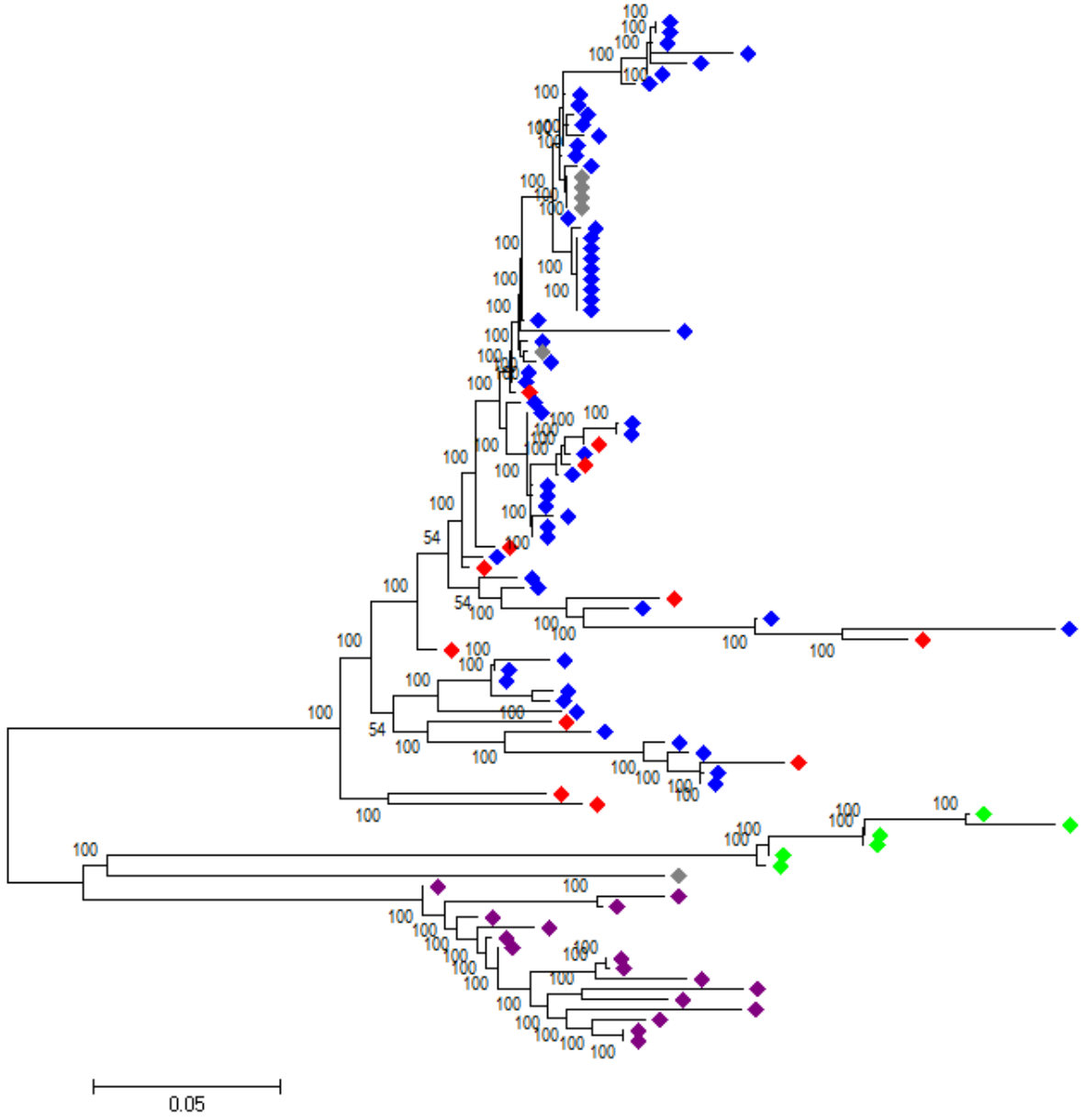


Figure S7.84 Chromosome 2: 12060001 - 12145001

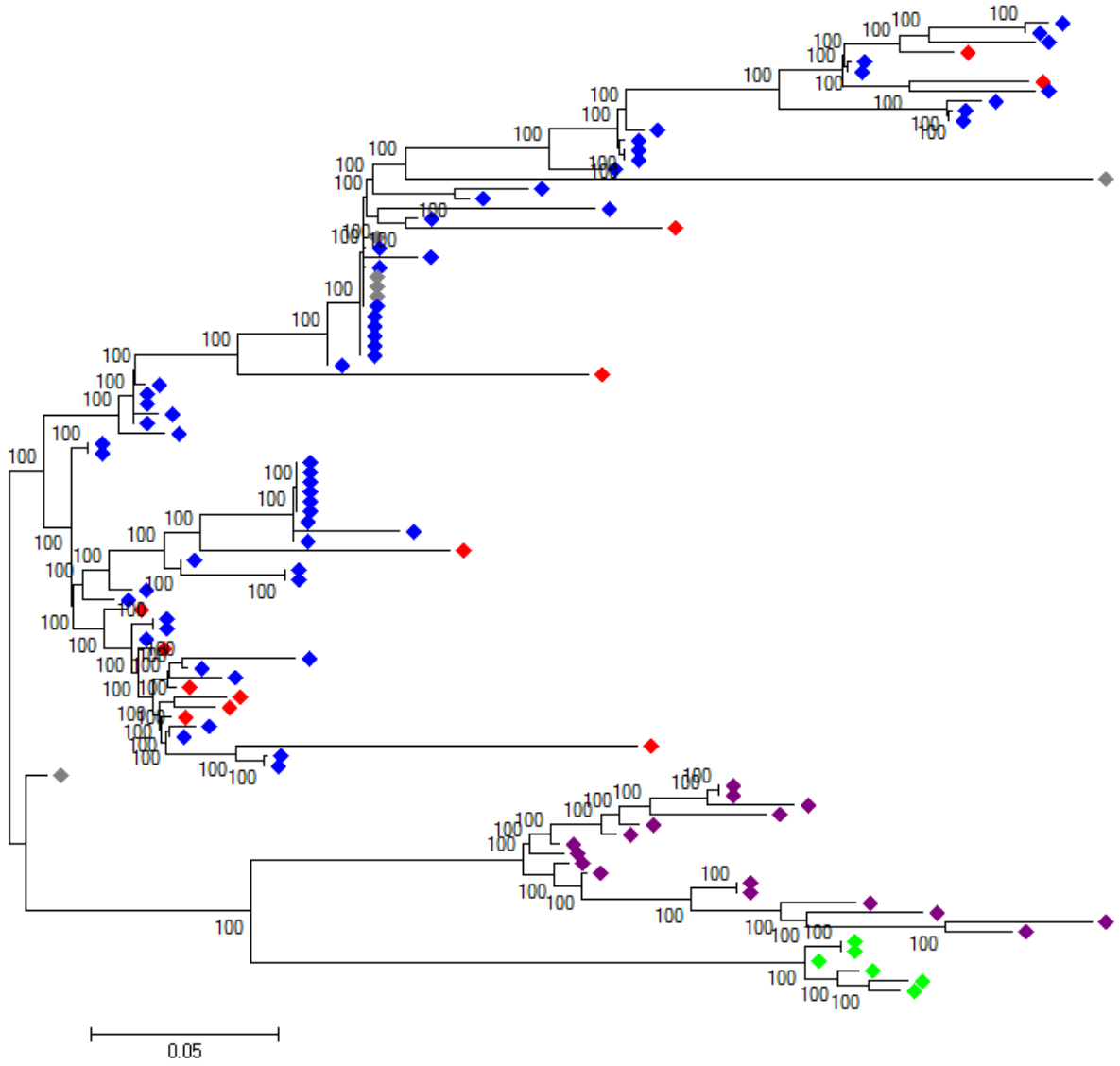


Figure S7.85 Chromosome 2: 12170001 - 12220001

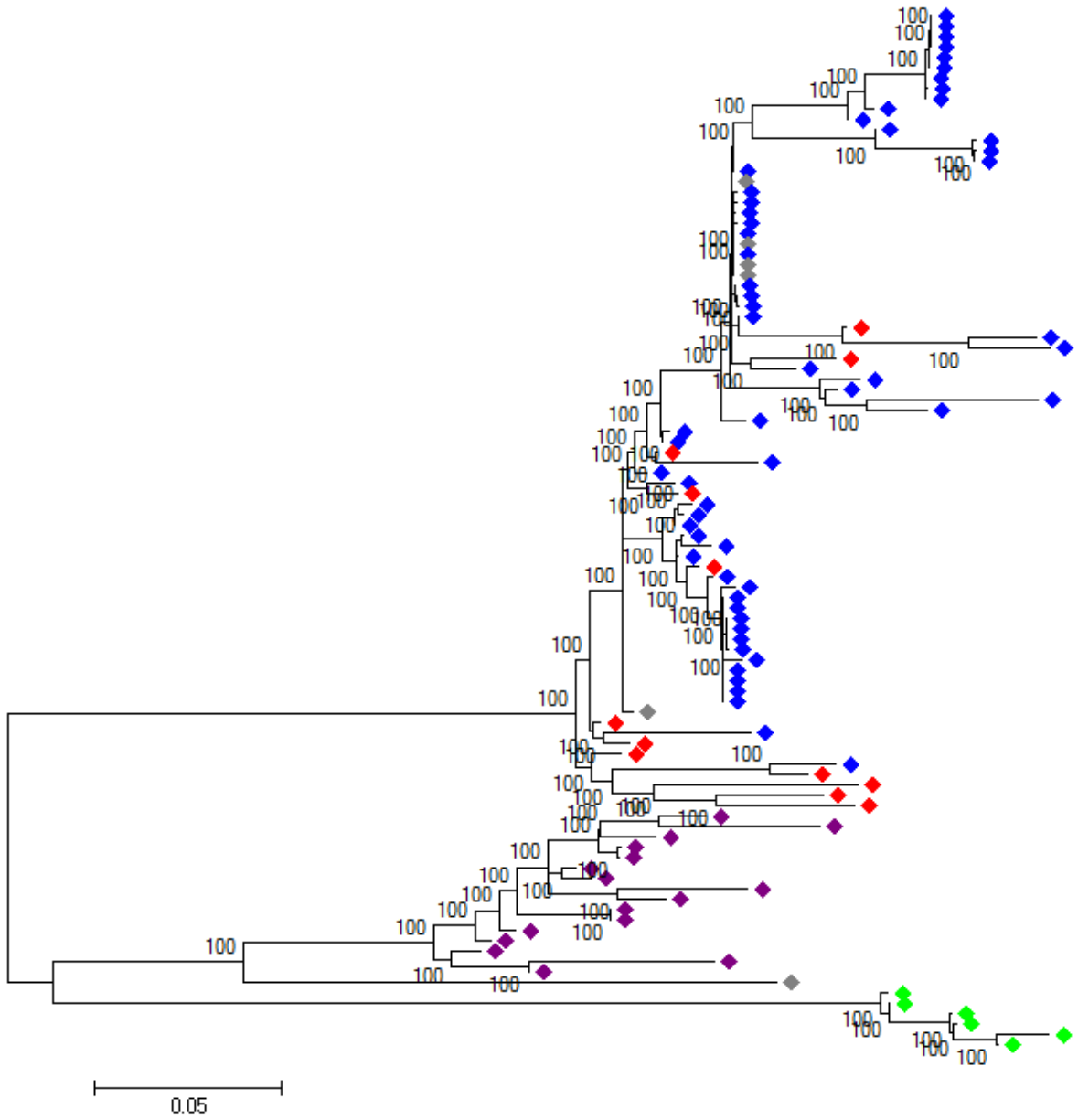


Figure S7.86 Chromosome 2: 12280001 - 12350001

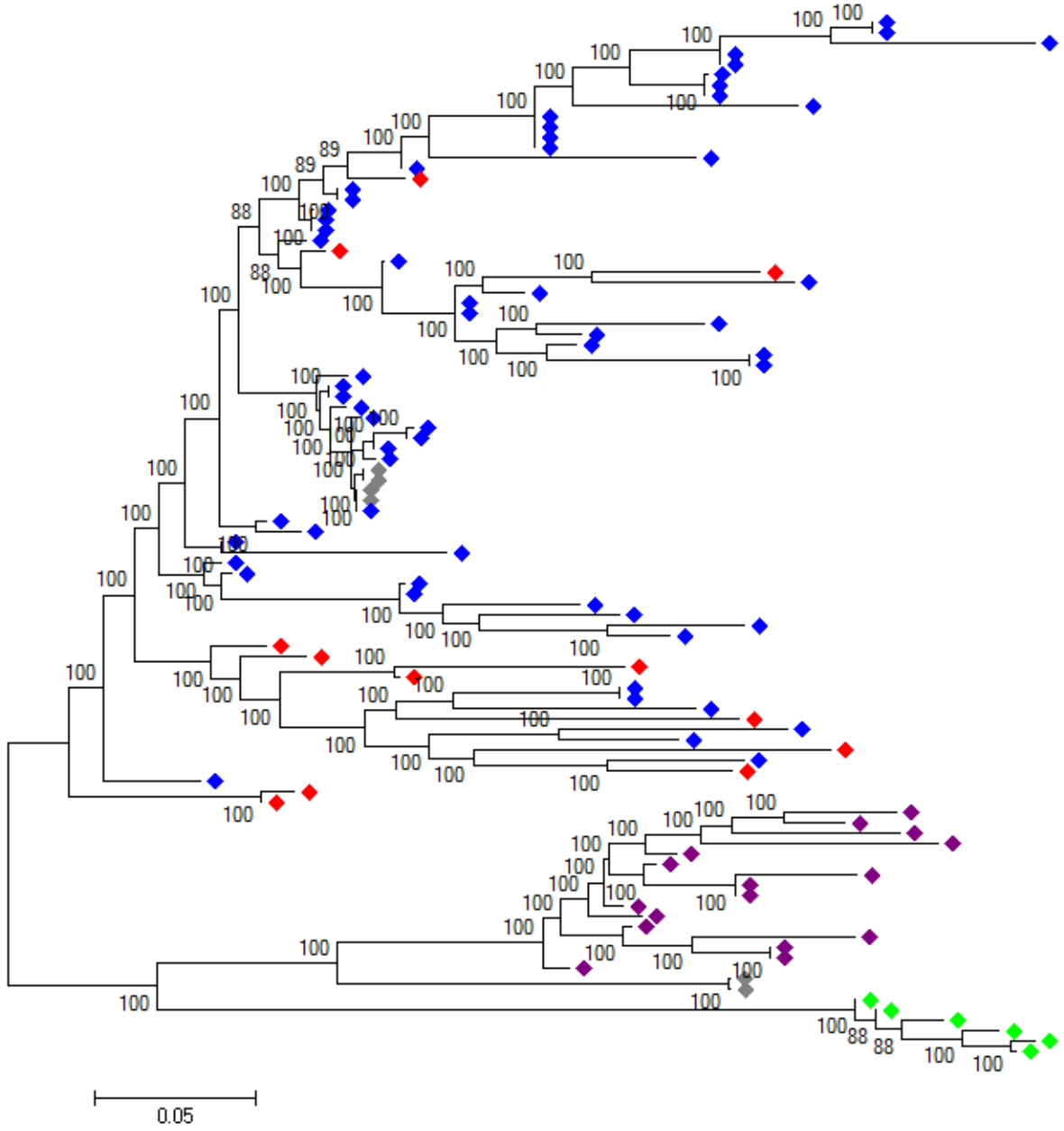


Figure S7.87 Chromosome 2: 12930001 - 12960001

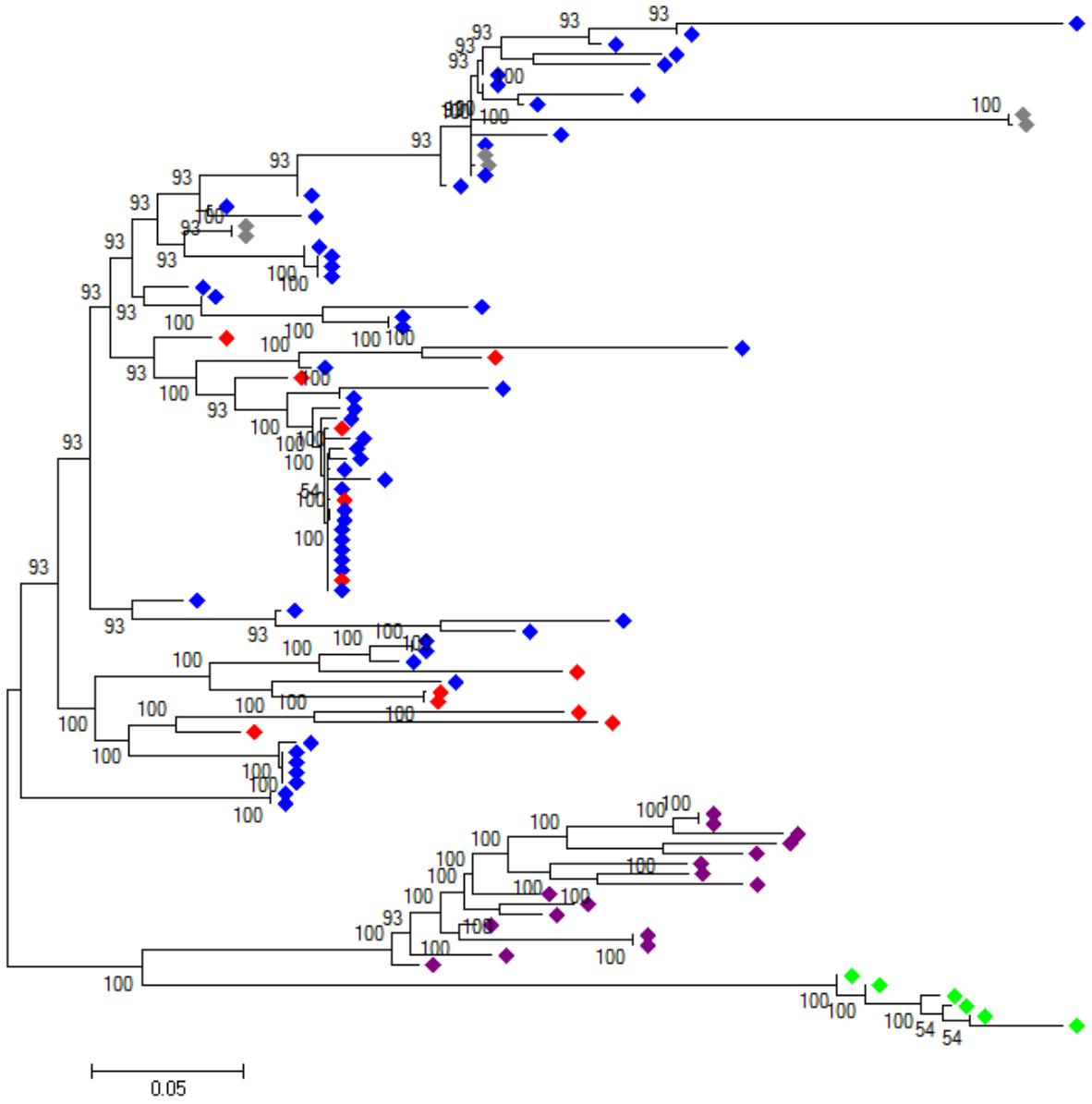


Figure S7.88 Chromosome 2: 13075001 - 13095001

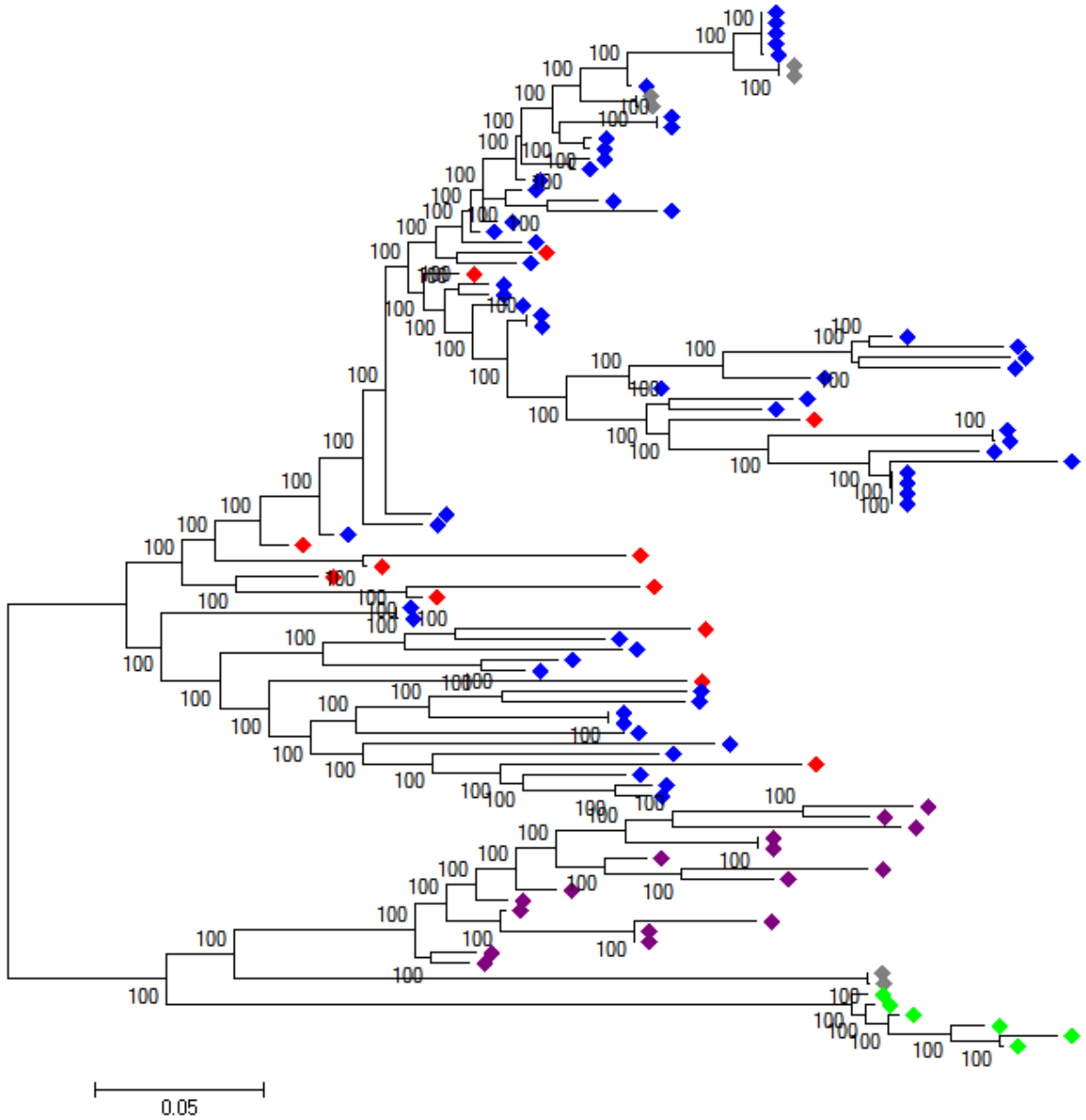


Figure S7.89 Chromosome 2: 13200001 - 13275001

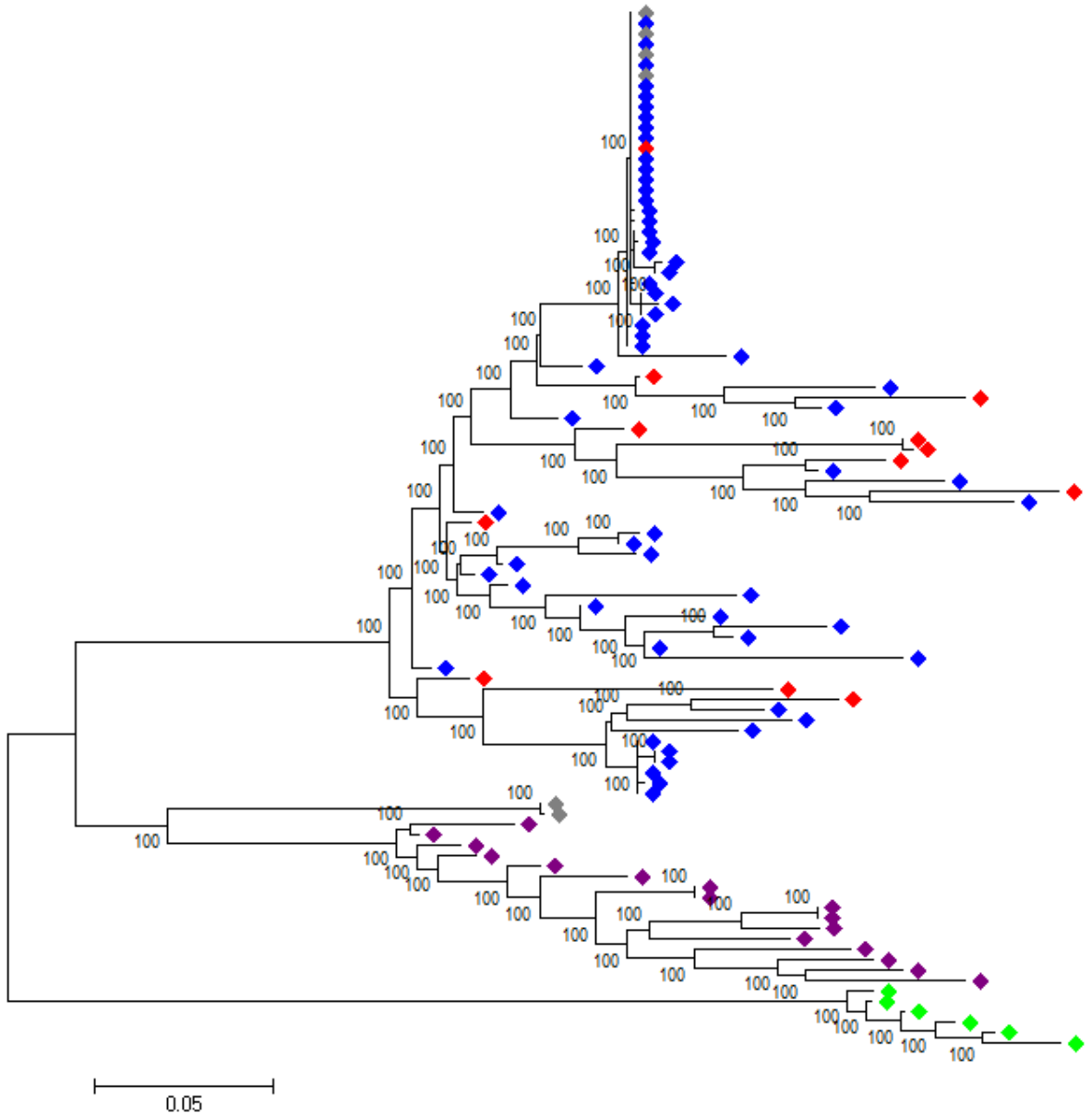


Figure S7.90 Chromosome 2: 13655001 - 13675001

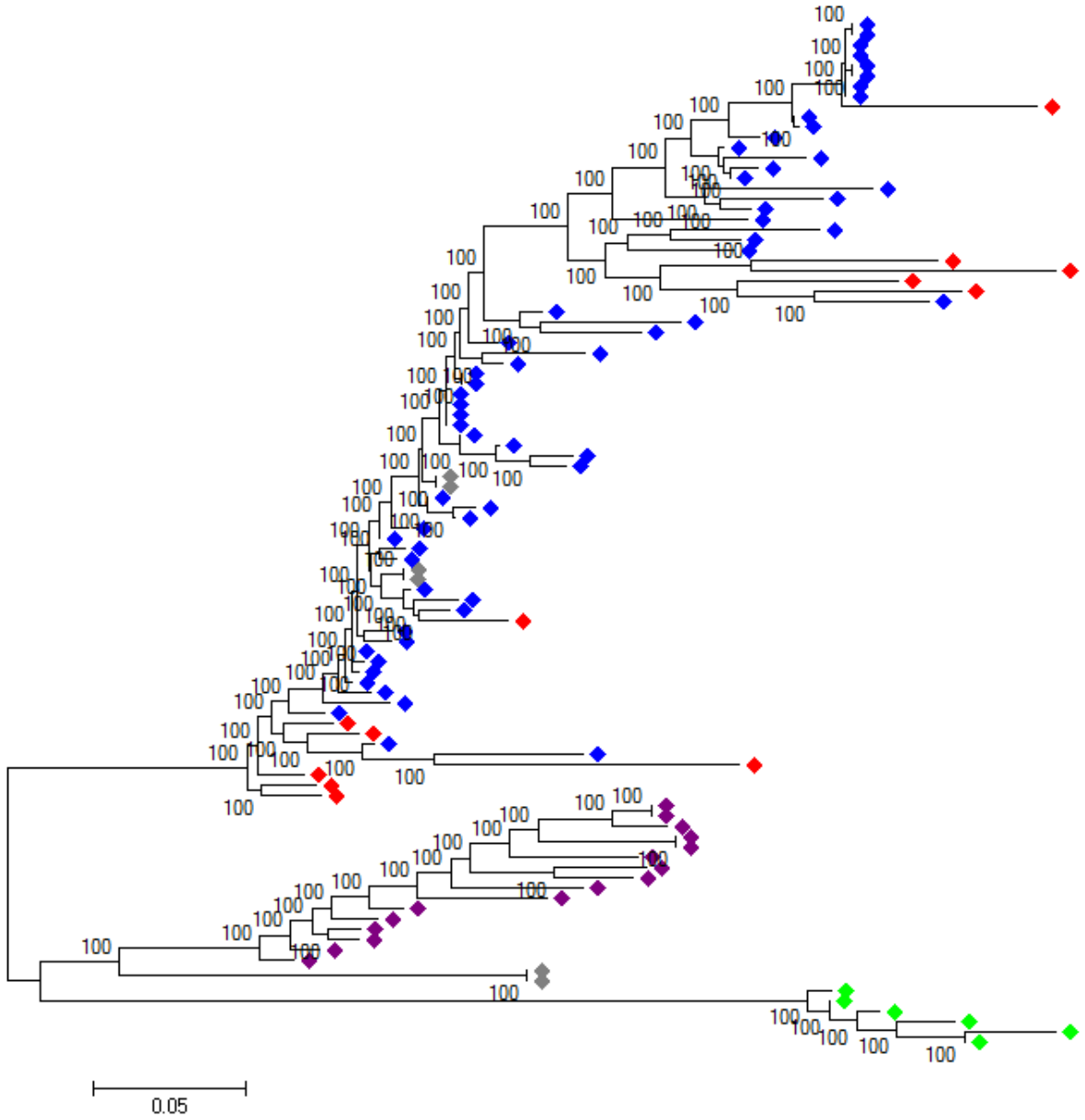


Figure S7.91 Chromosome 2: 13855001 - 13925001

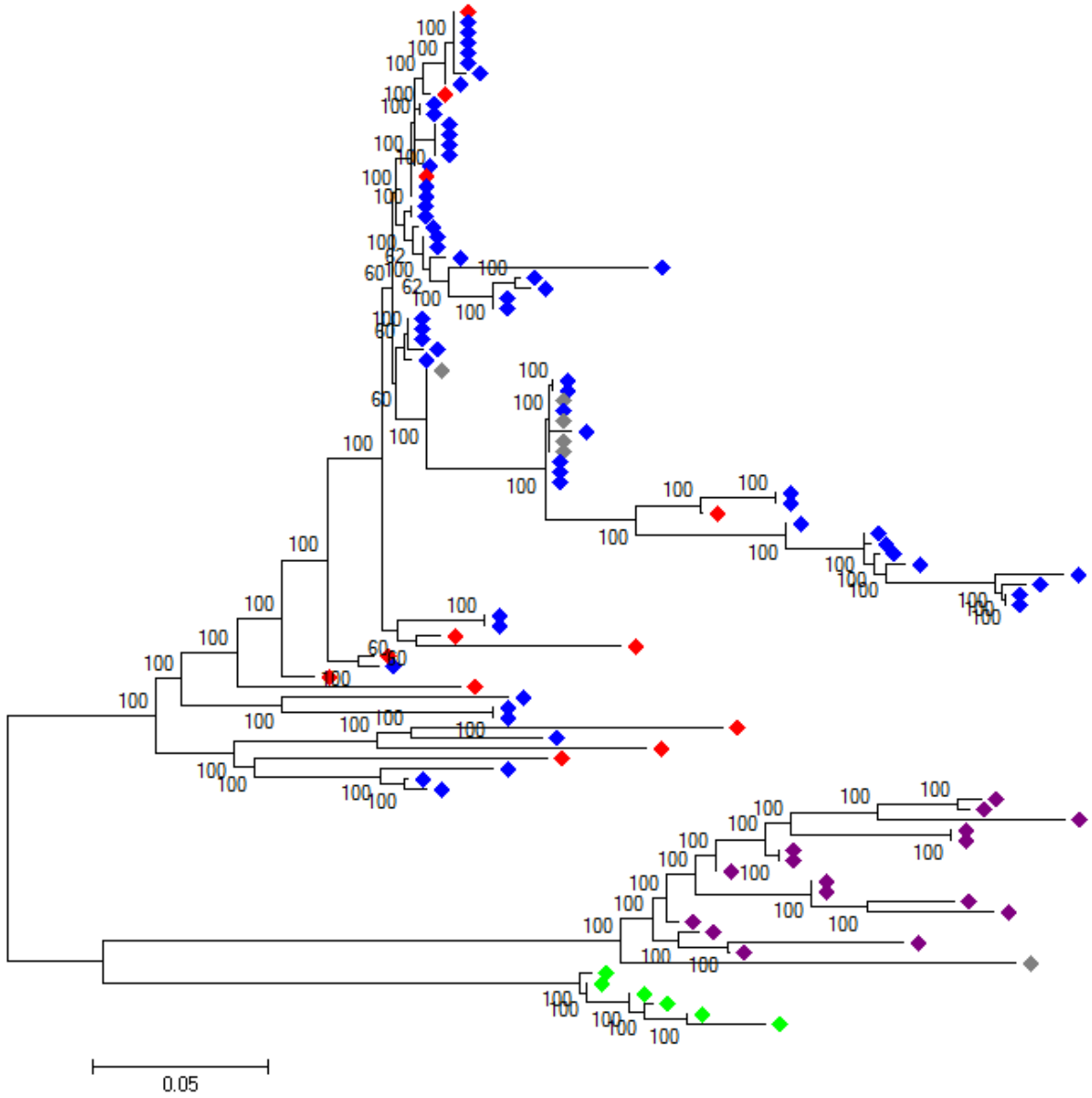


Figure S7.92 Chromosome 2: 14520001 - 14540001

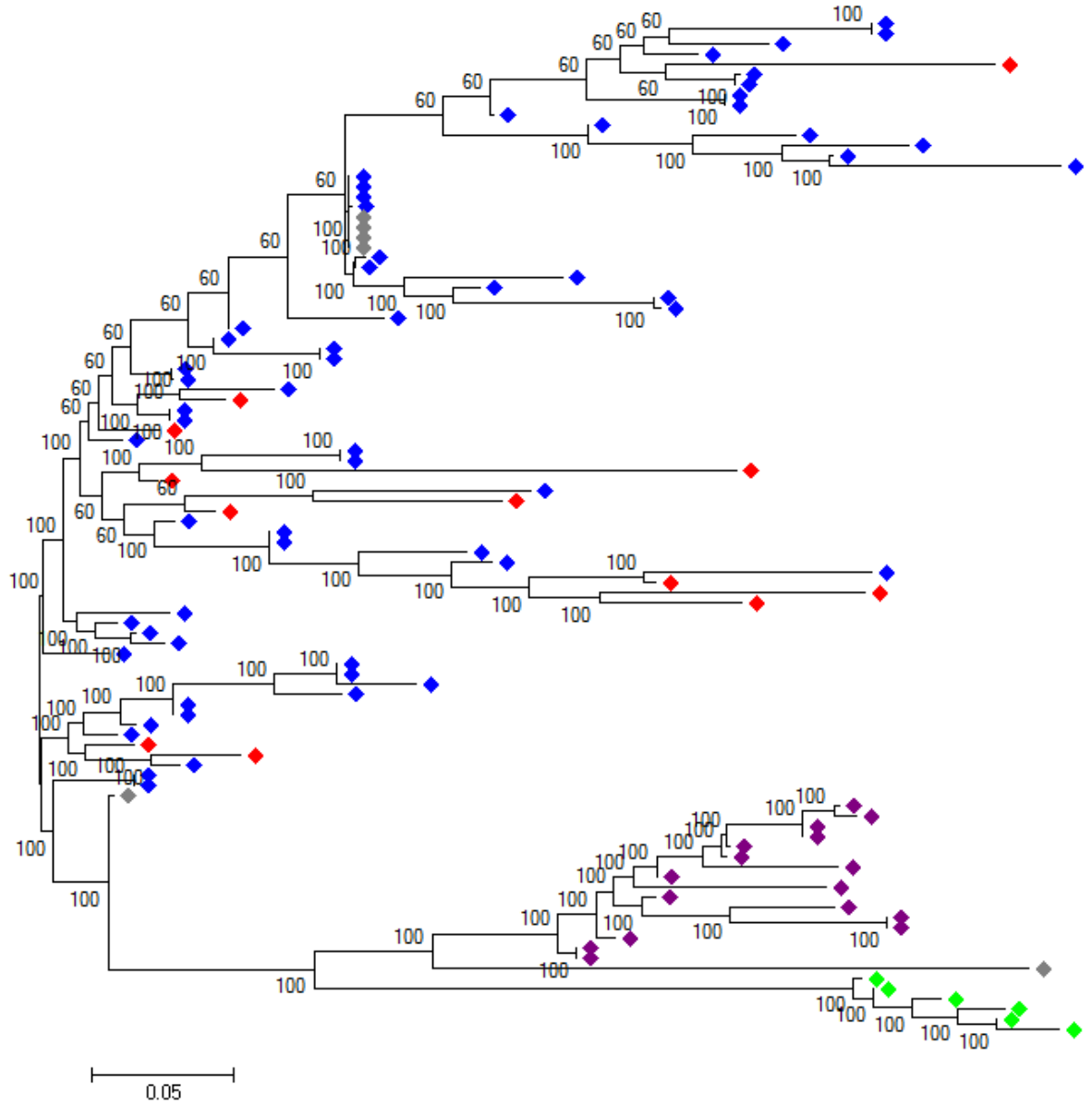


Figure S7.93 Chromosome 2: 14630001 - 14660001

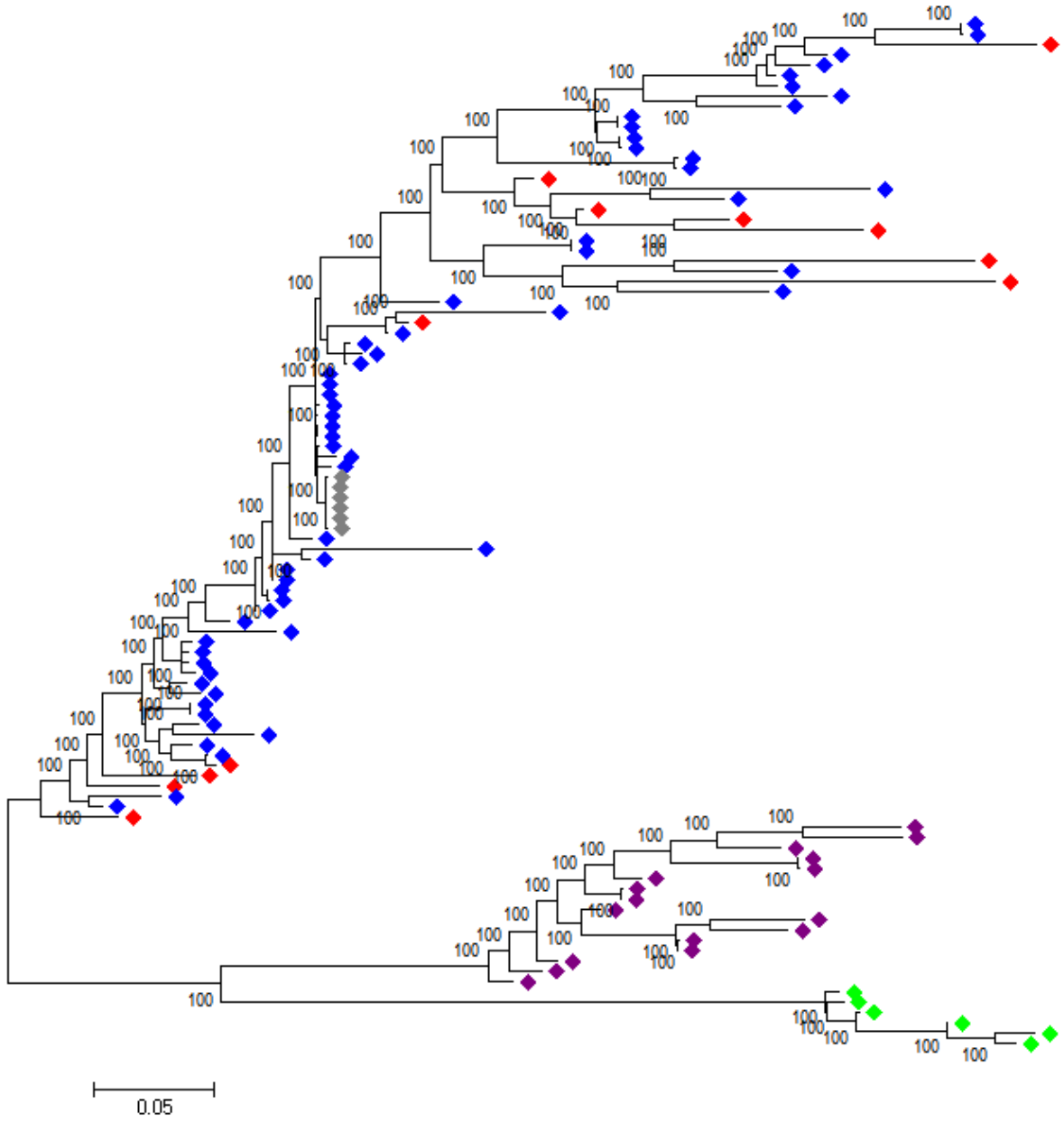


Figure S7.94 Chromosome 2: 15255001 - 15290001

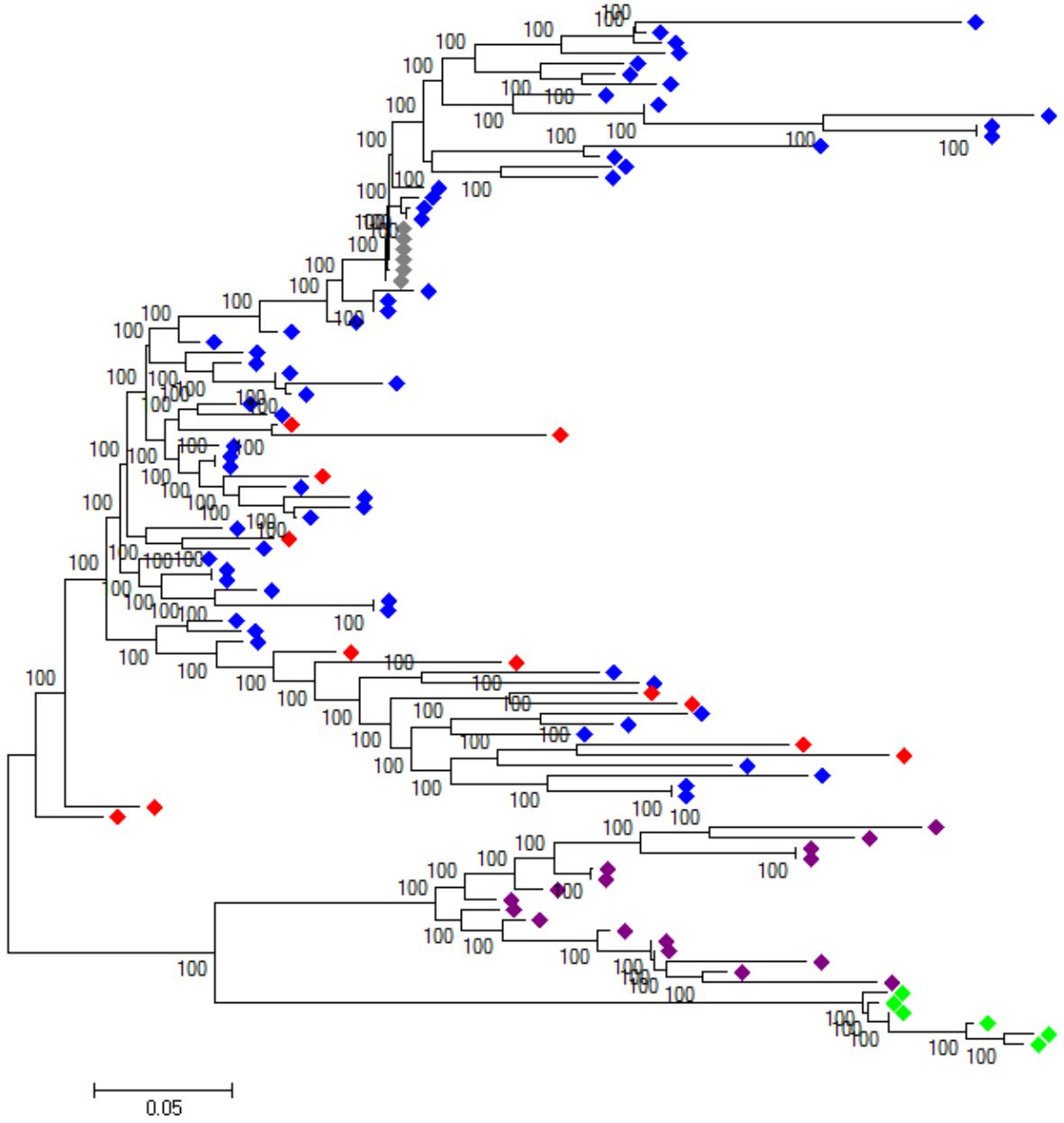


Figure S7.95 Chromosome 2: 15445001 - 15495001

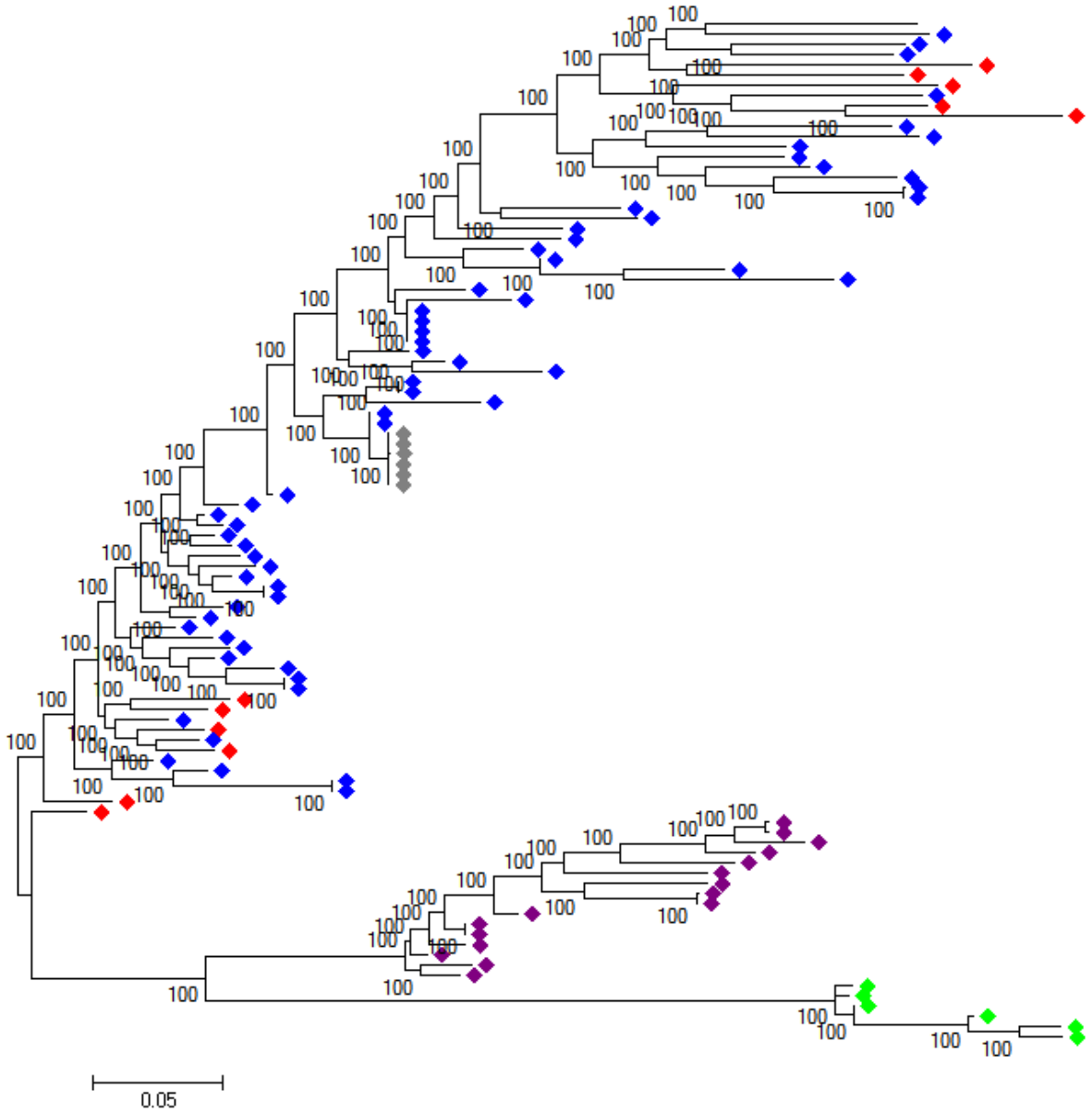


Figure S7.96 Chromosome 2: 15580001 - 15665001

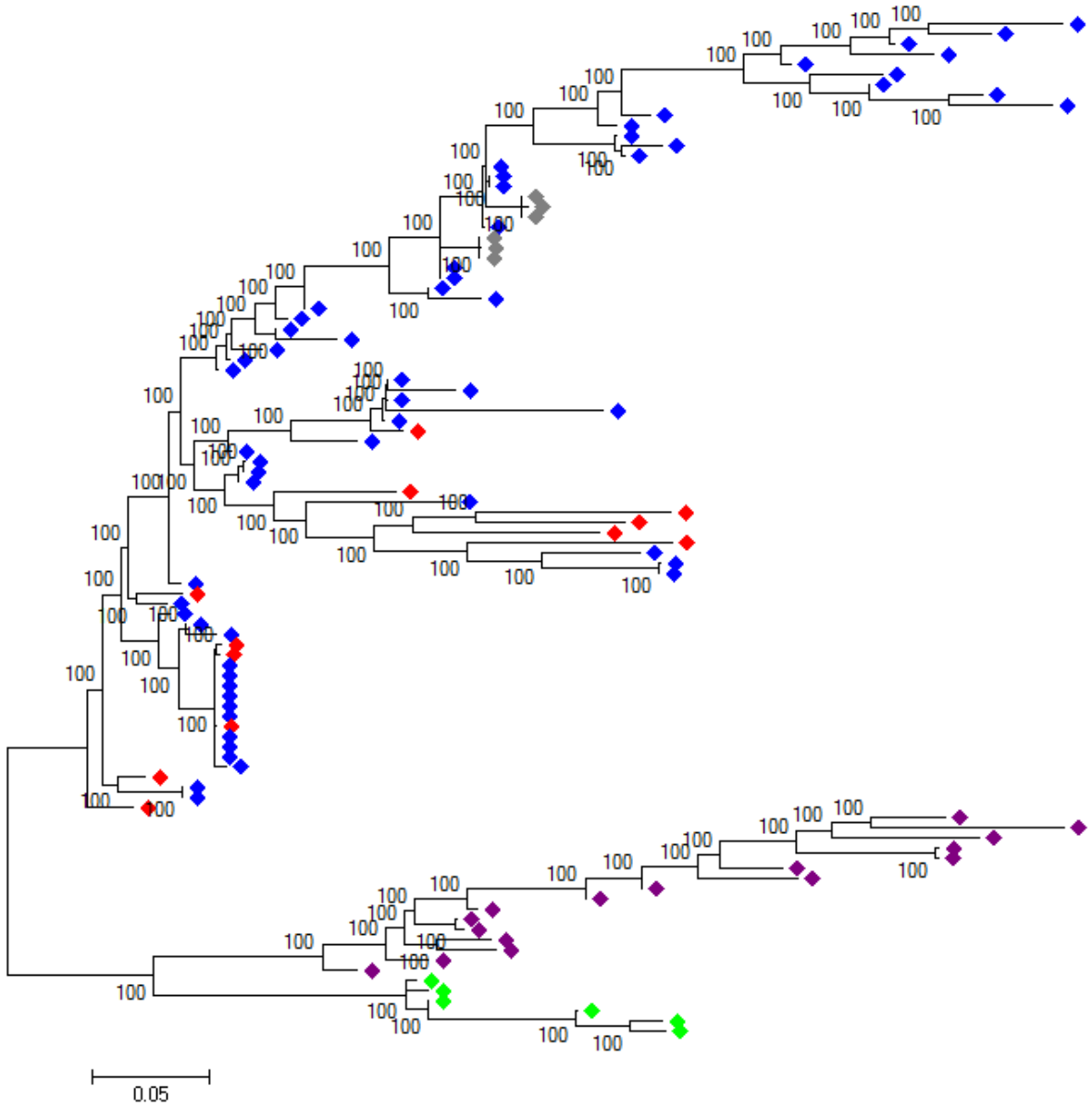


Figure S7.97 Chromosome 2: 15765001 - 15800001

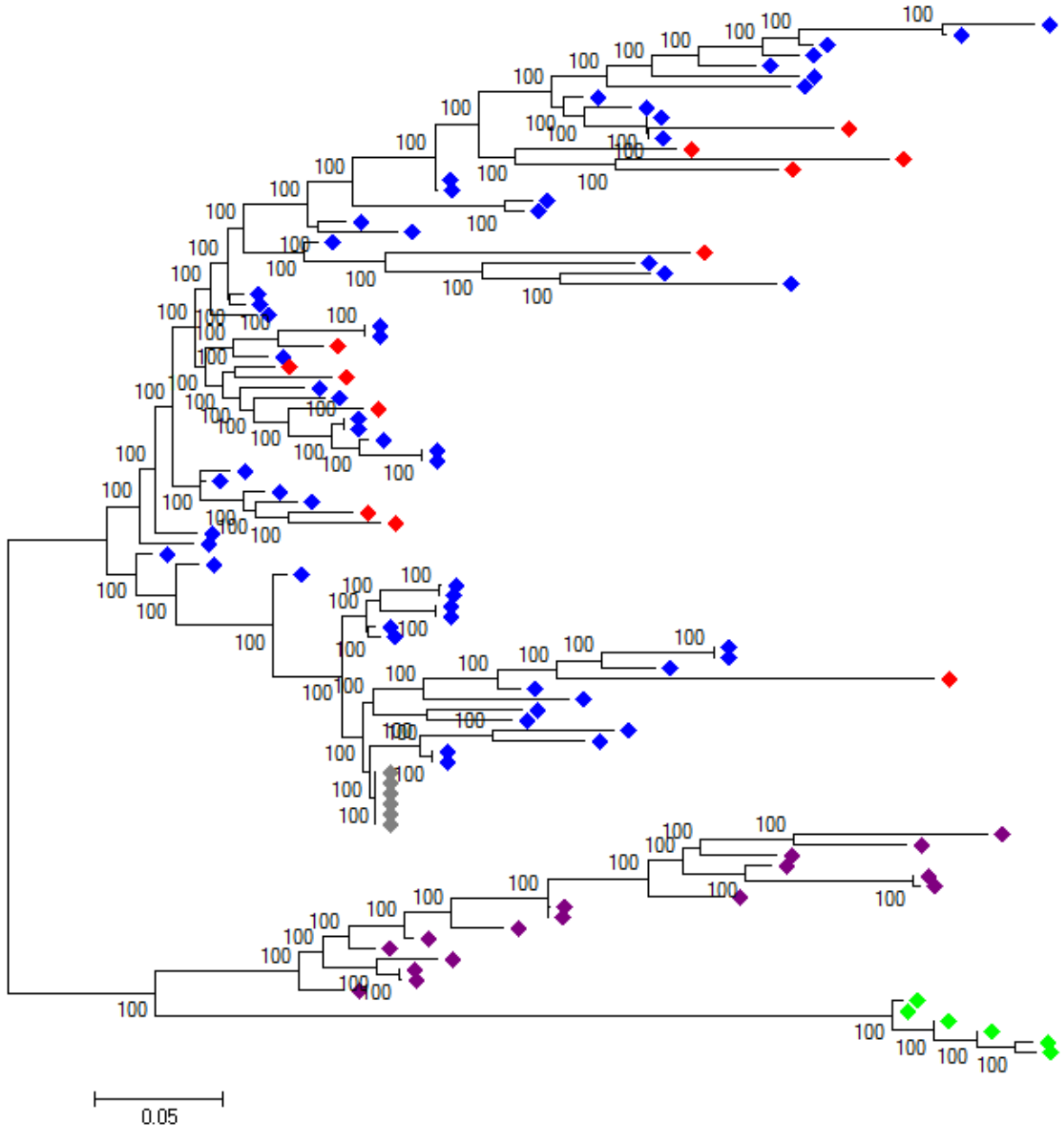


Figure S7.98 Chromosome 2: 16100001 - 16130001

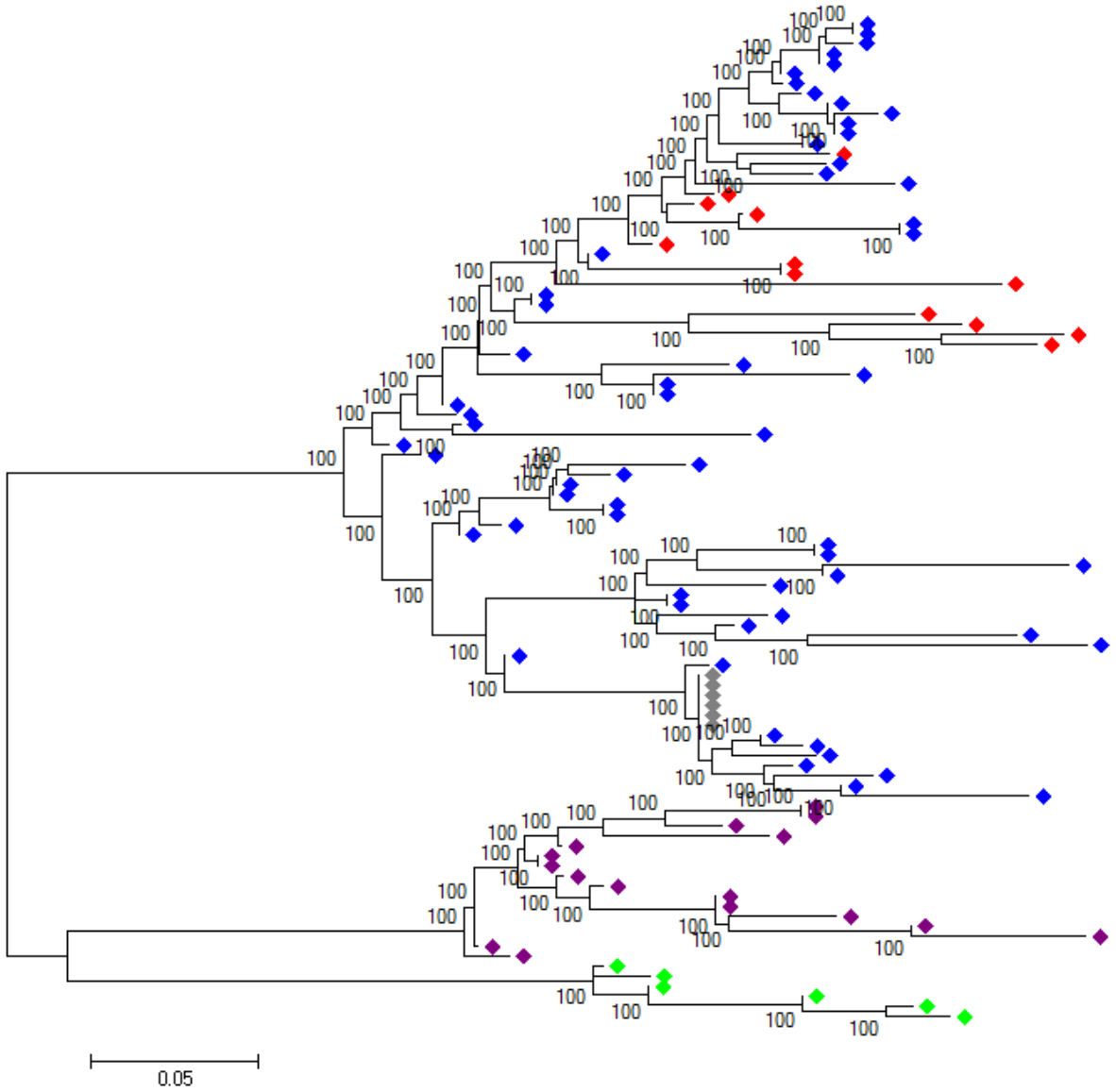


Figure S7.99 Chromosome 2: 16225001 - 16255001

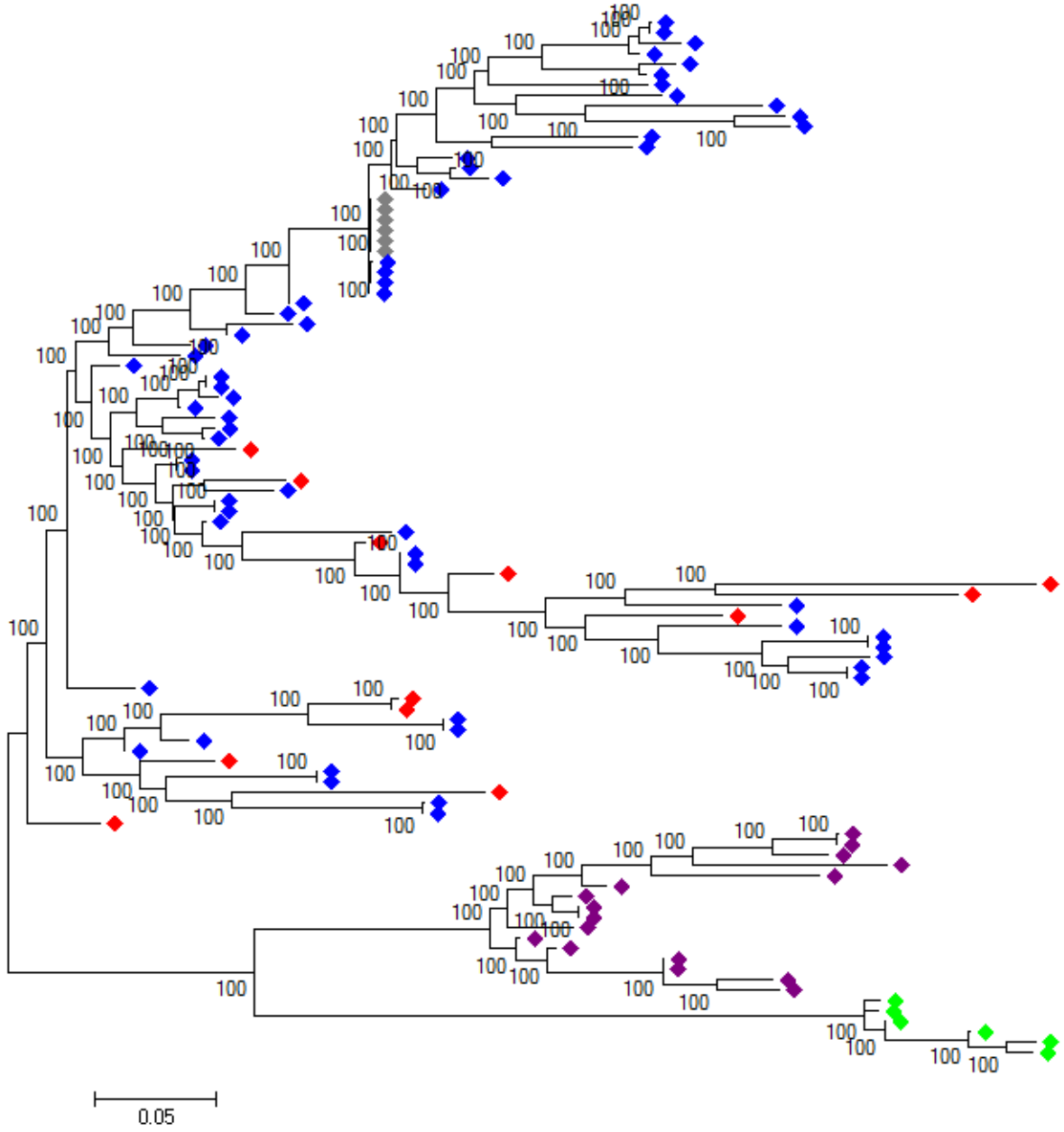


Figure S7.100 Chromosome 2: 16360001 - 16425001

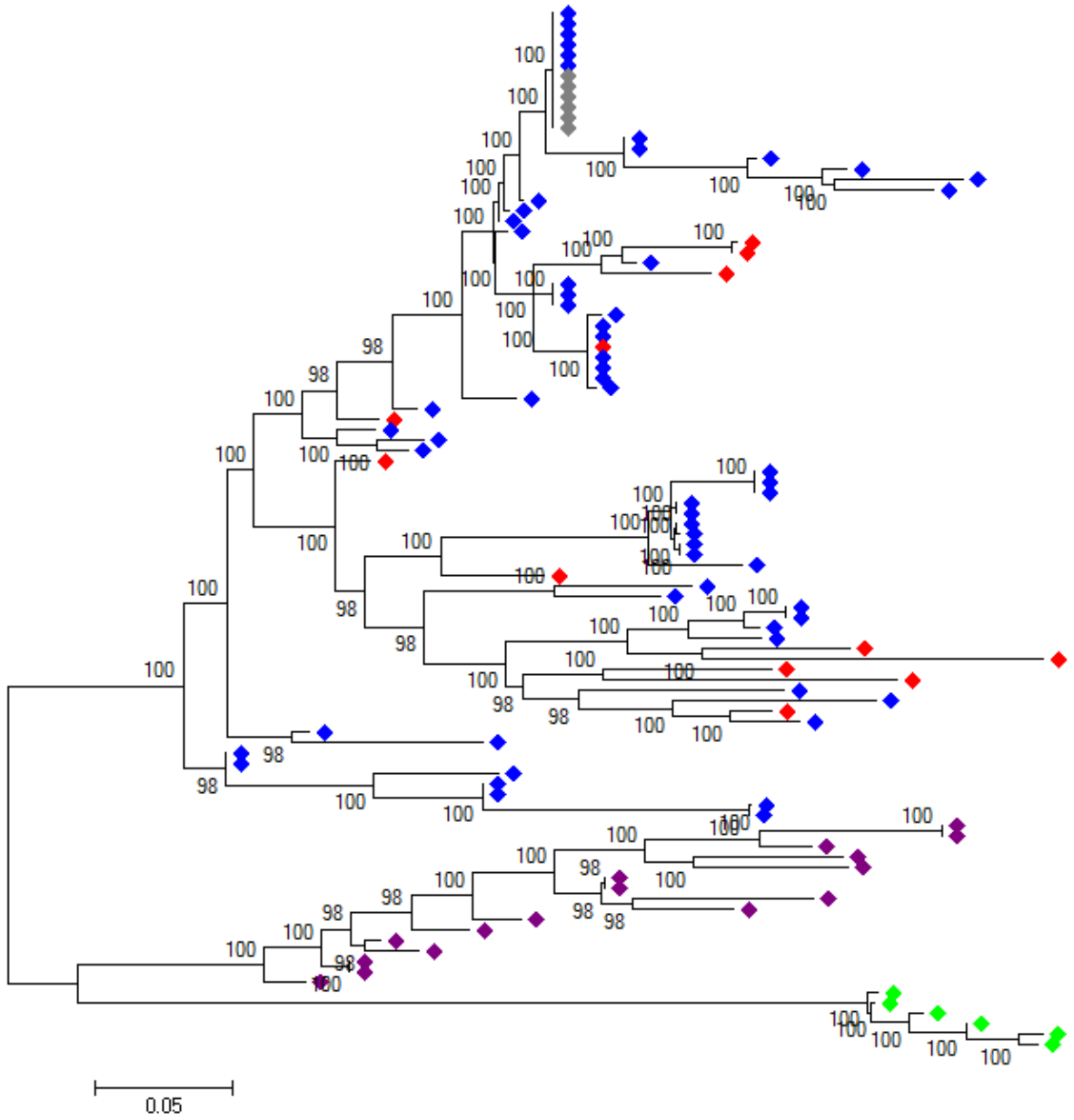


Figure S7.101 Chromosome 2: 16620001 - 16645001

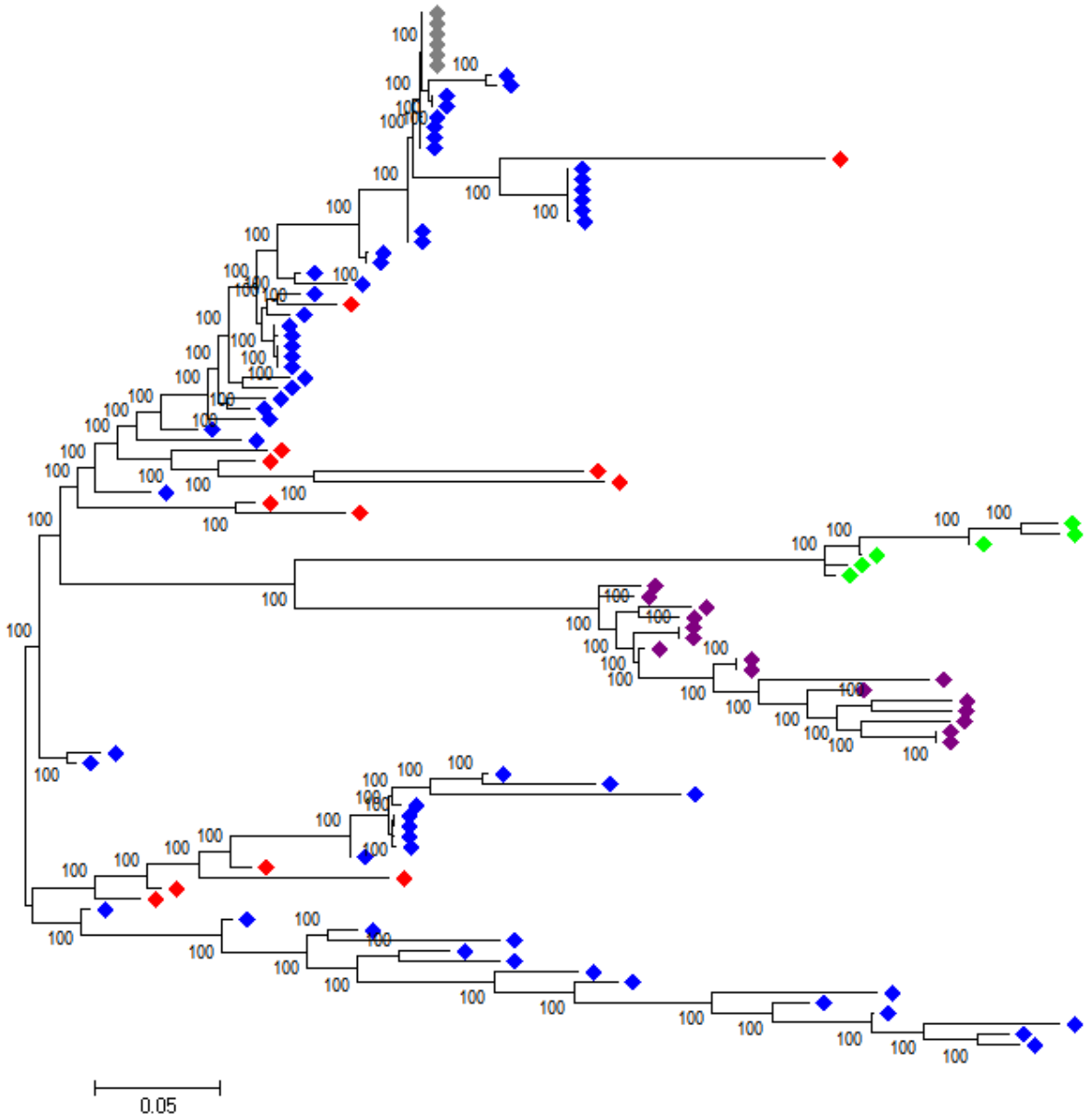


Figure S7.102 Chromosome 2: 17750001 - 17780001

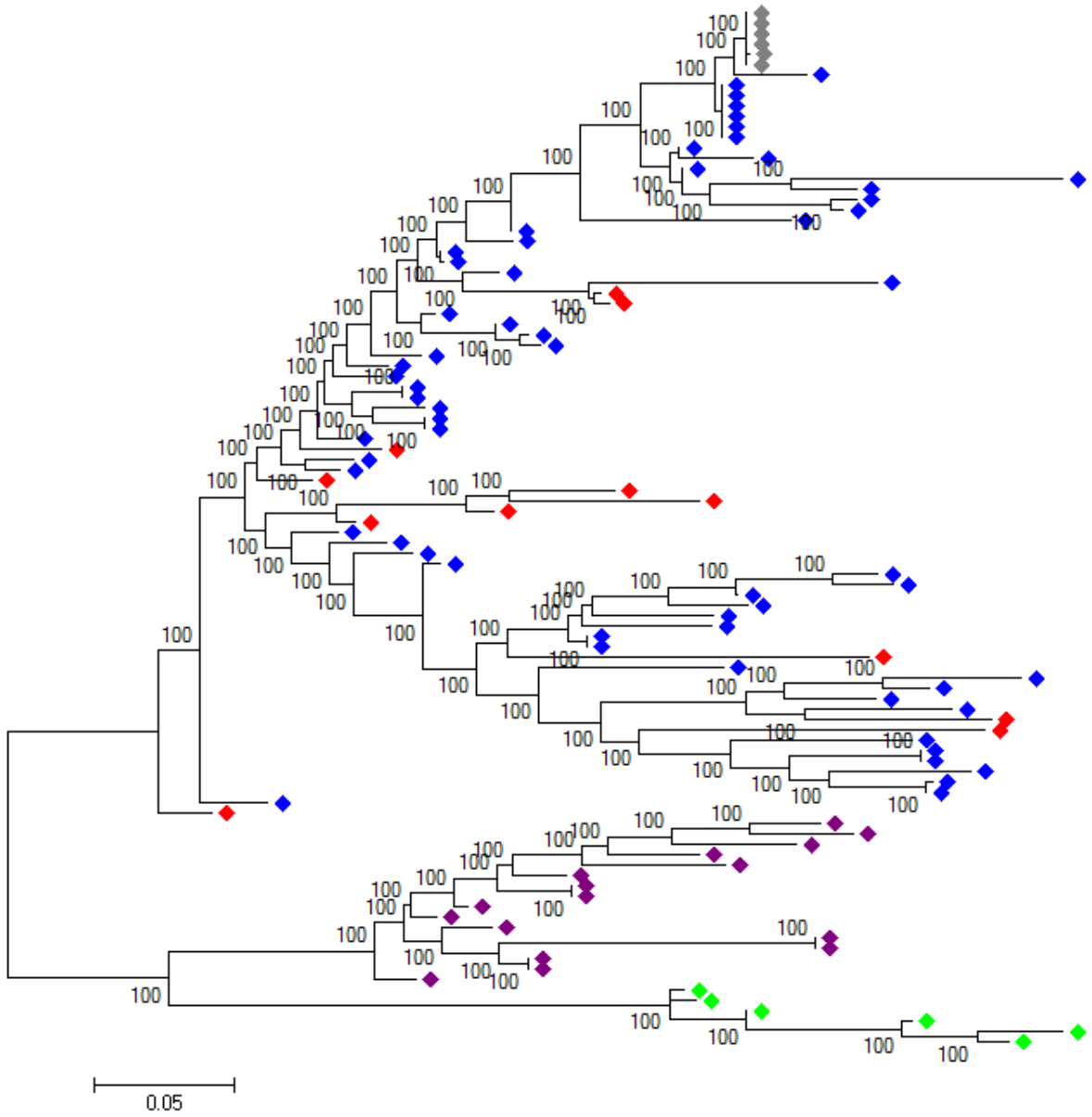


Figure S7.103 Chromosome 2: 17895001 - 17920001

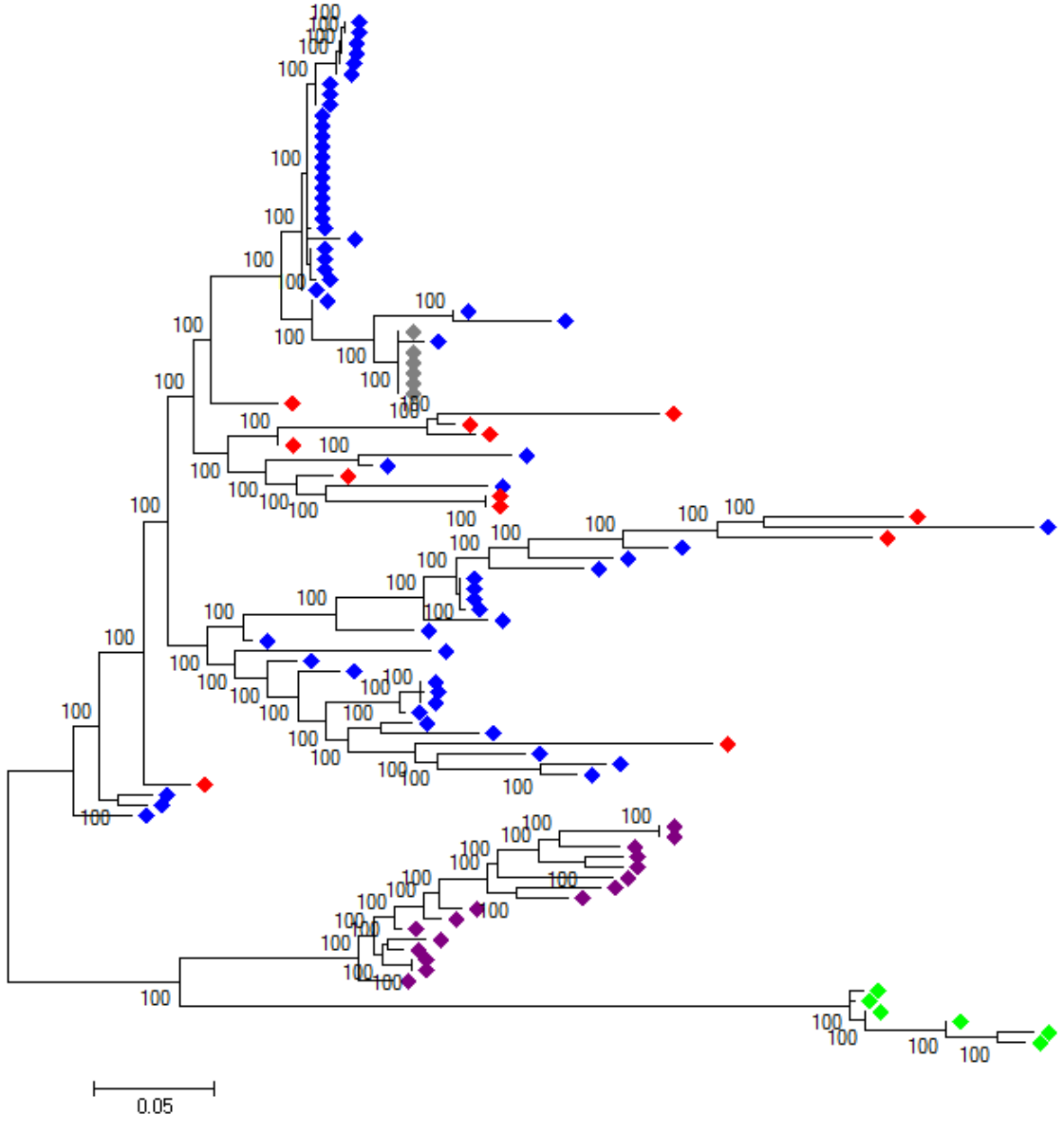


Figure S7.104 Chromosome 2: 18135001 - 18160001

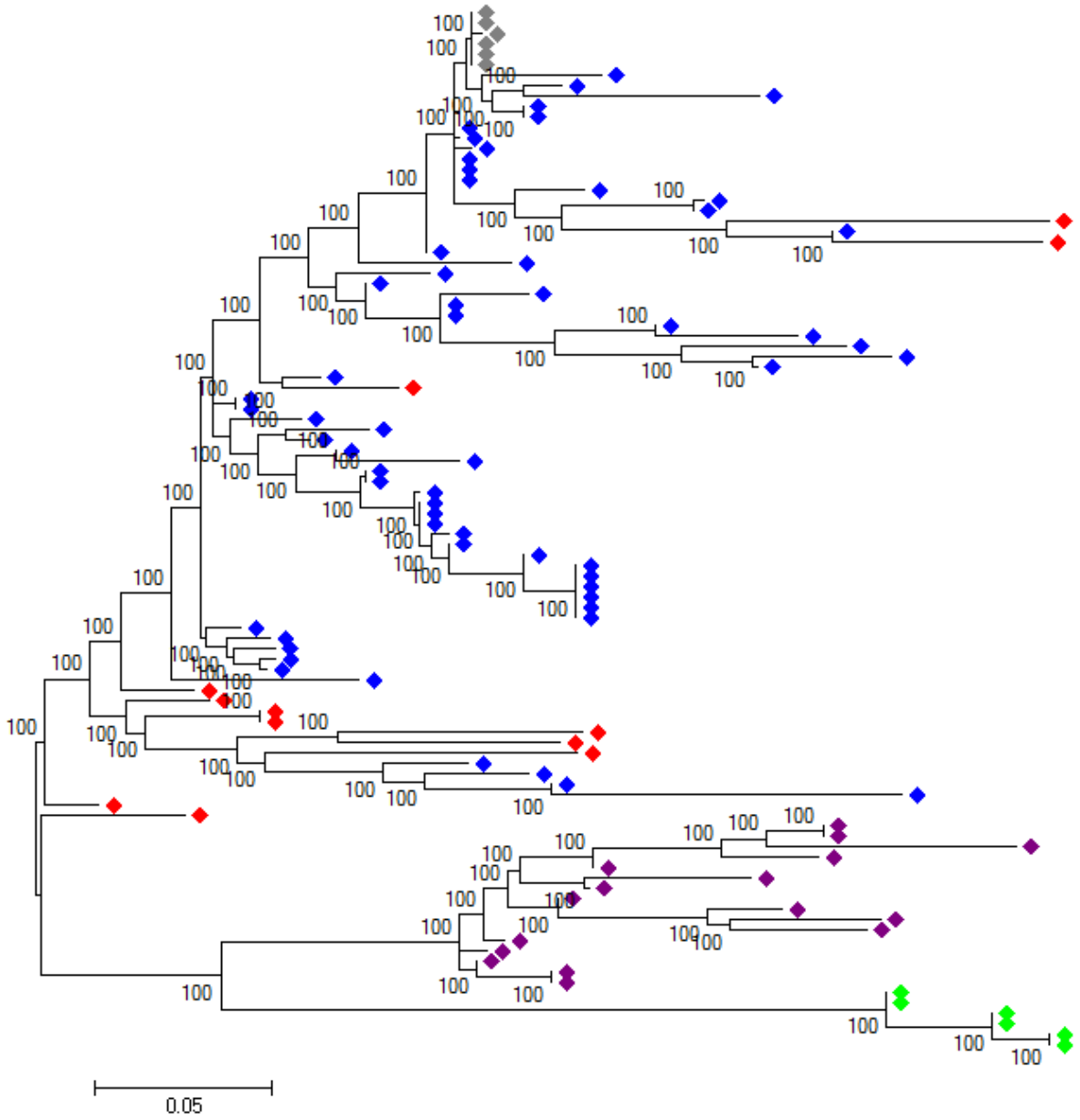


Figure S7.105 Chromosome 2: 18175001 - 18195001

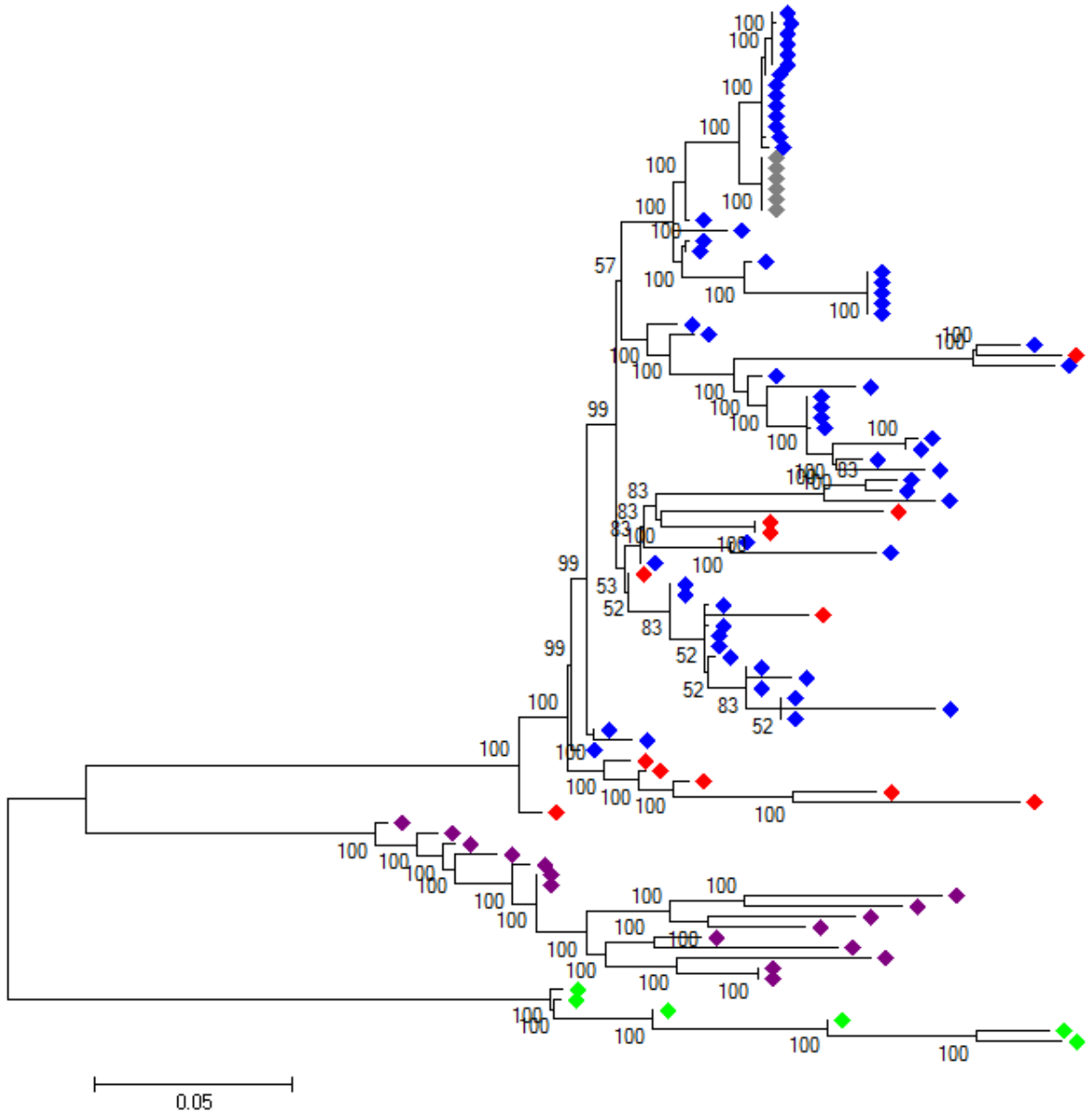


Figure S7.106 Chromosome 2: 18450001 - 18485001

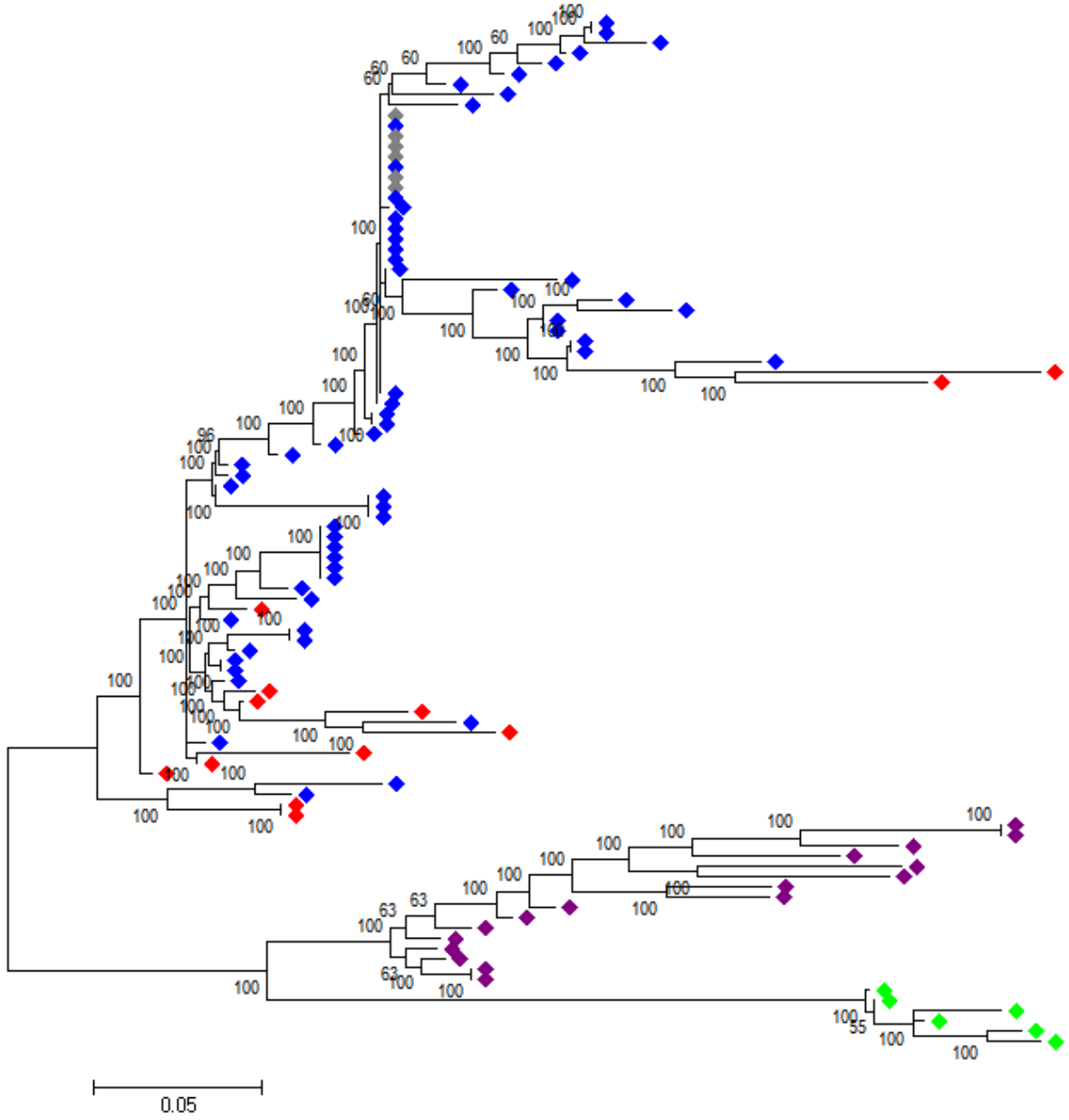


Figure S7.107 Chromosome 2: 18520001 - 18550001

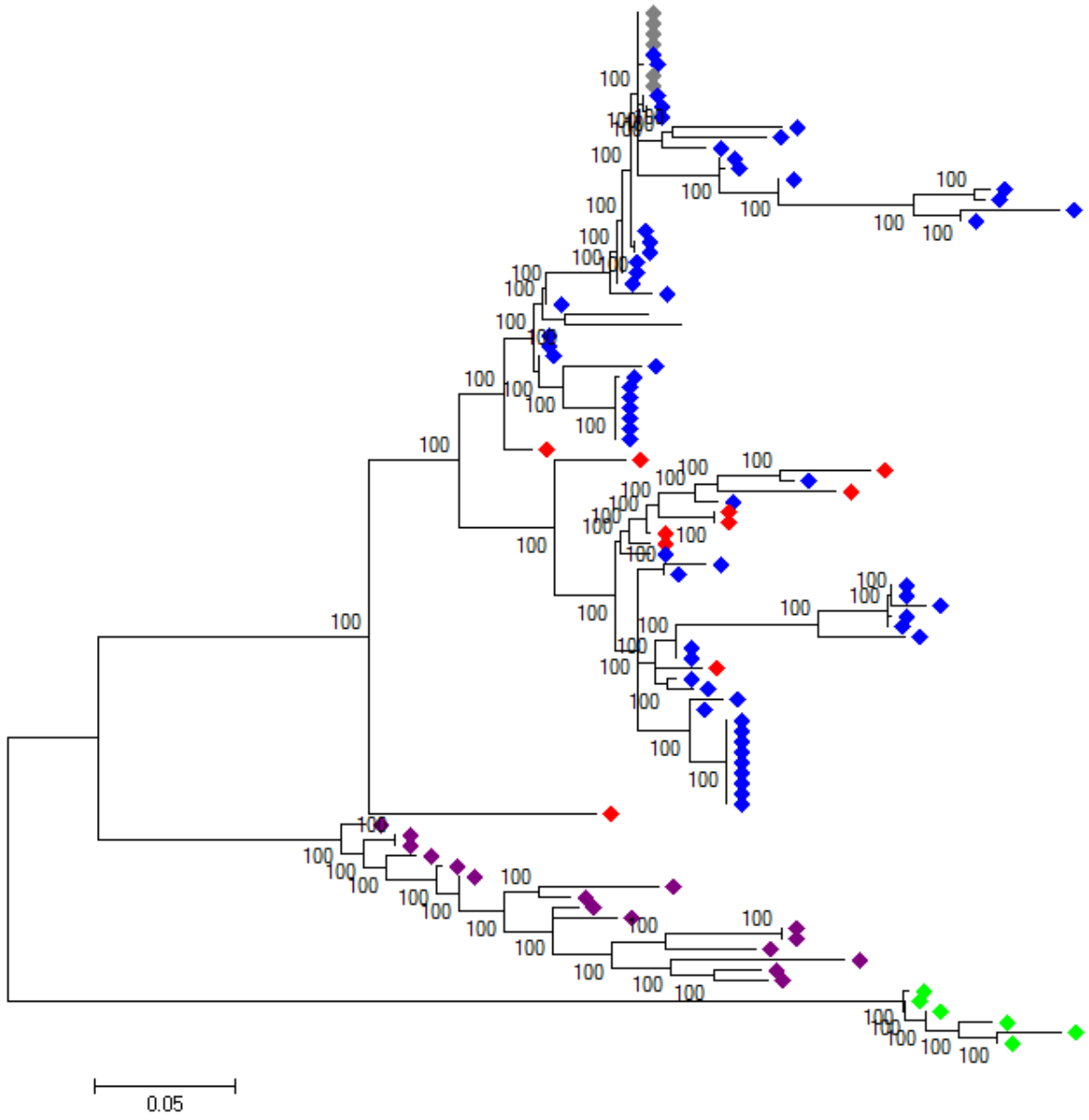


Figure S7.108 Chromosome 2: 18575001 - 18595001

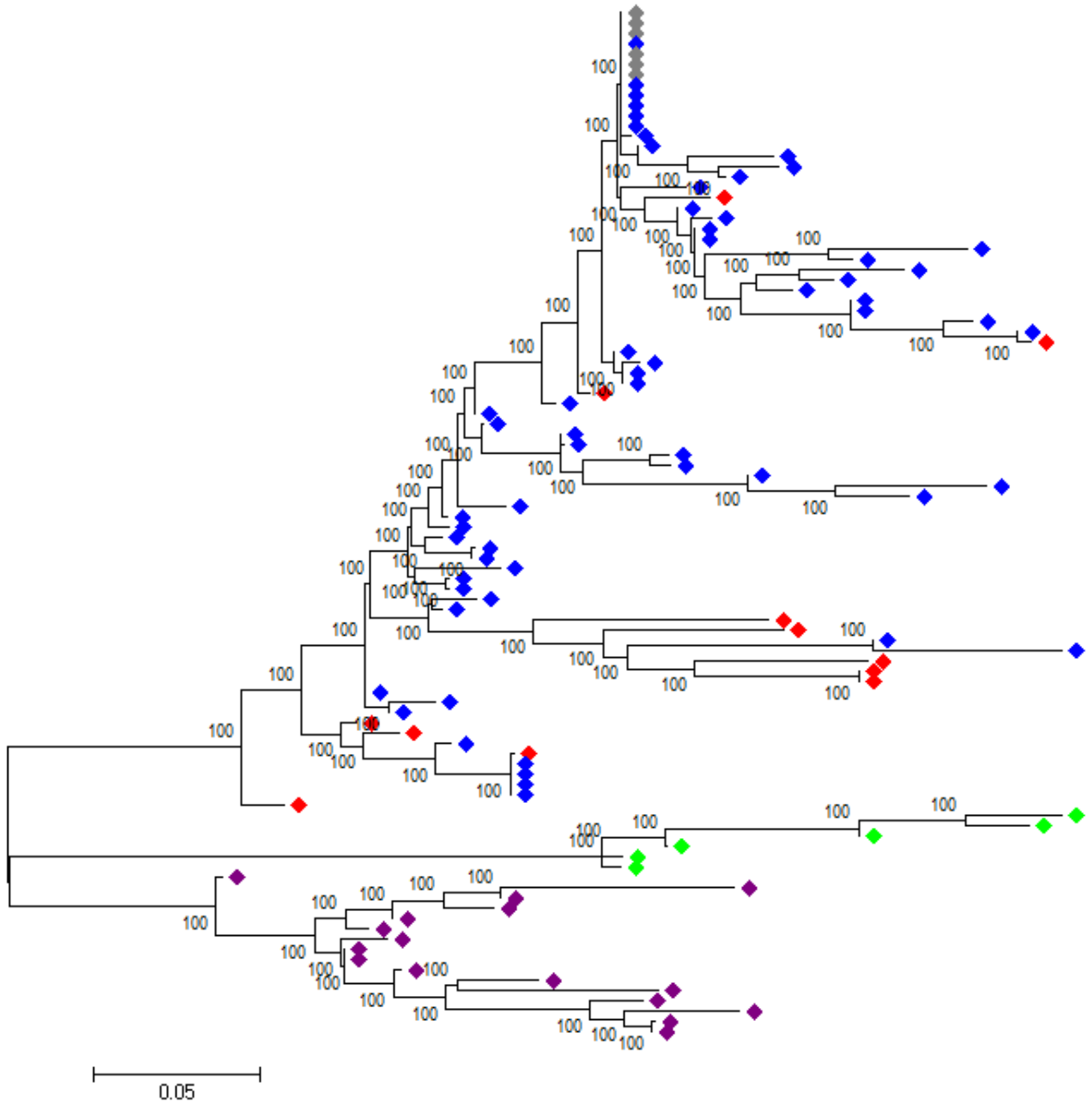


Figure S7.109 Chromosome 2: 18685001 - 18715001

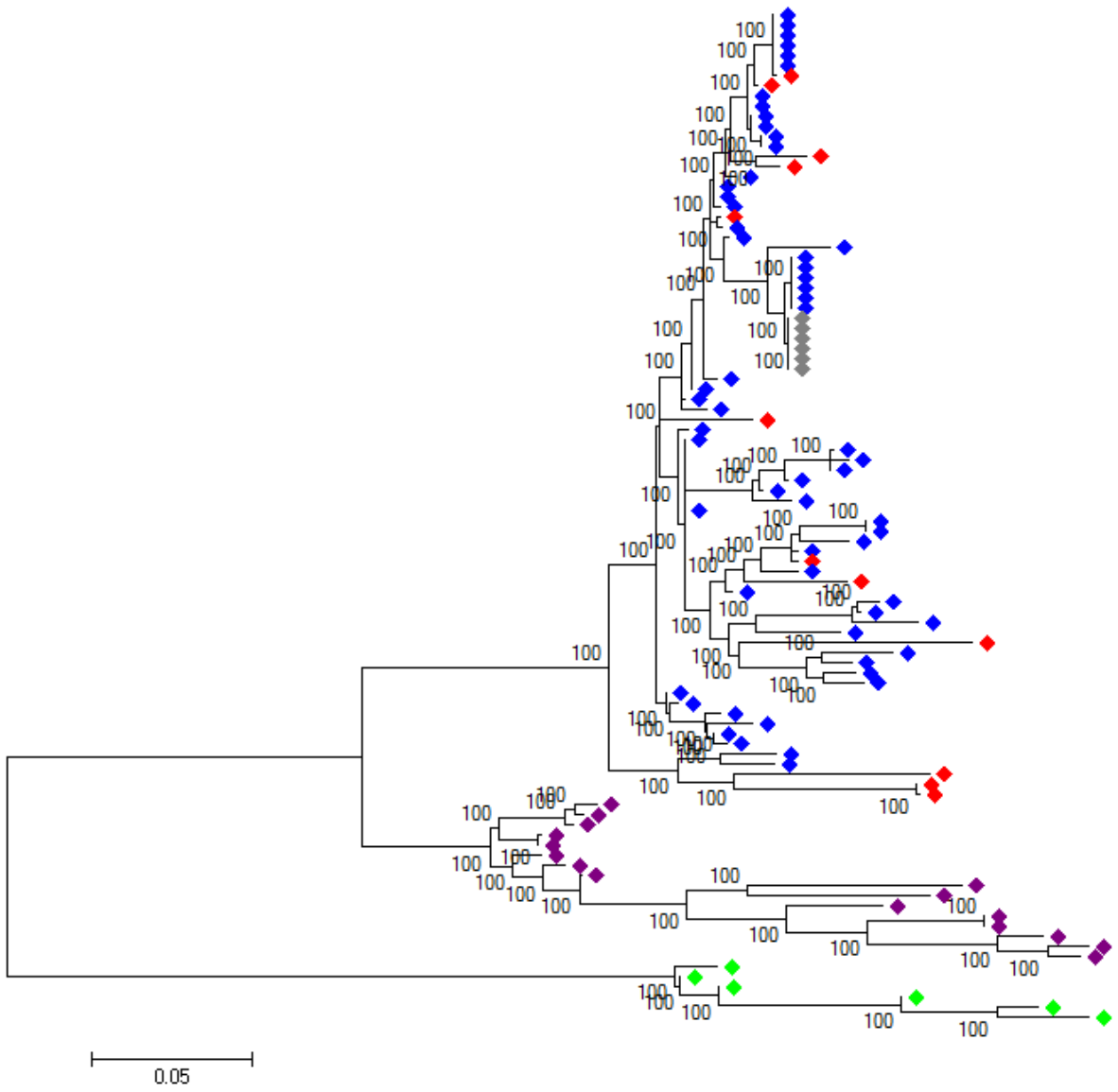


Figure S7.110 Chromosome 2: 18735001 - 18765001

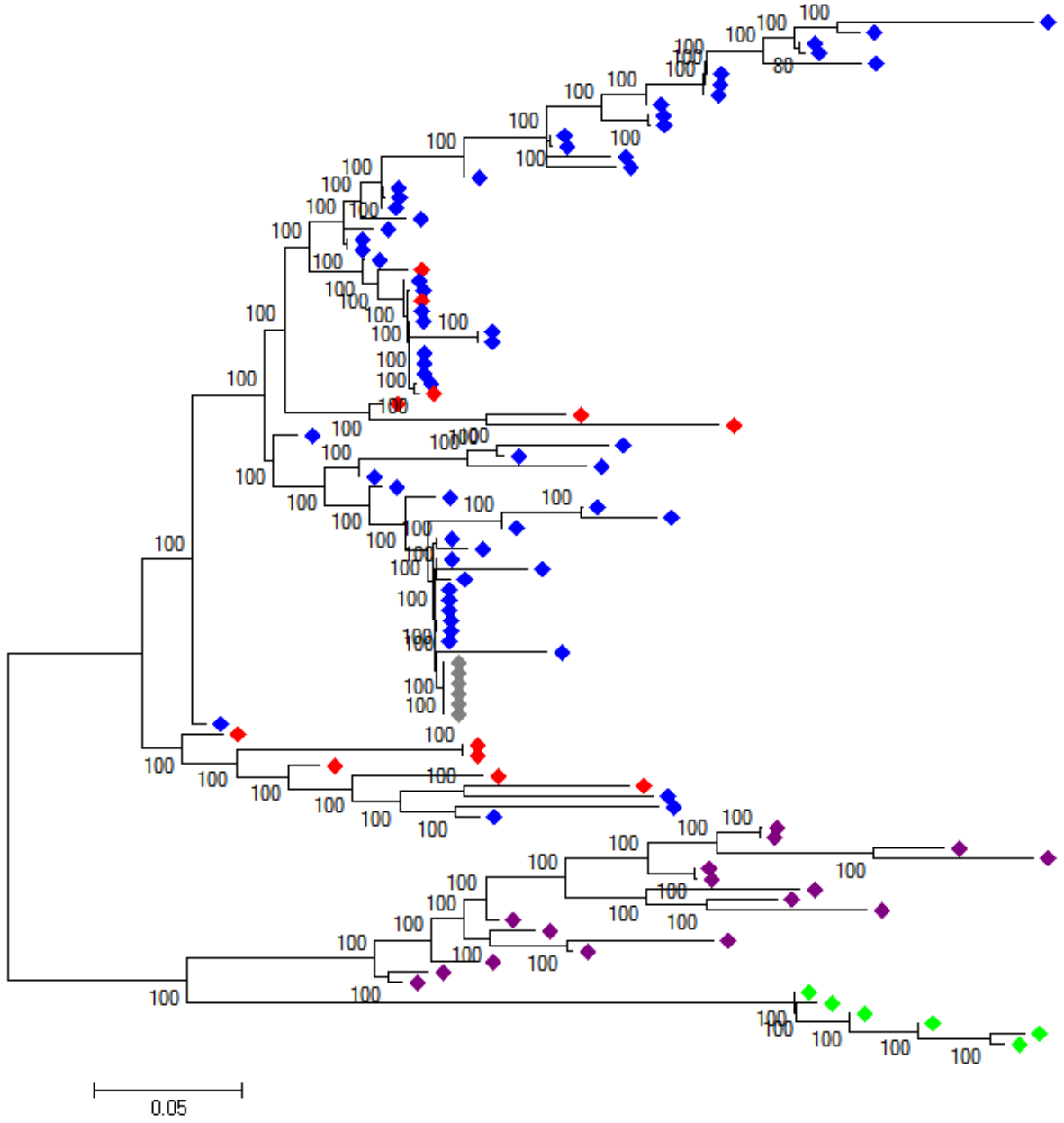


Figure S7.111 Chromosome 2: 18885001 - 18950001

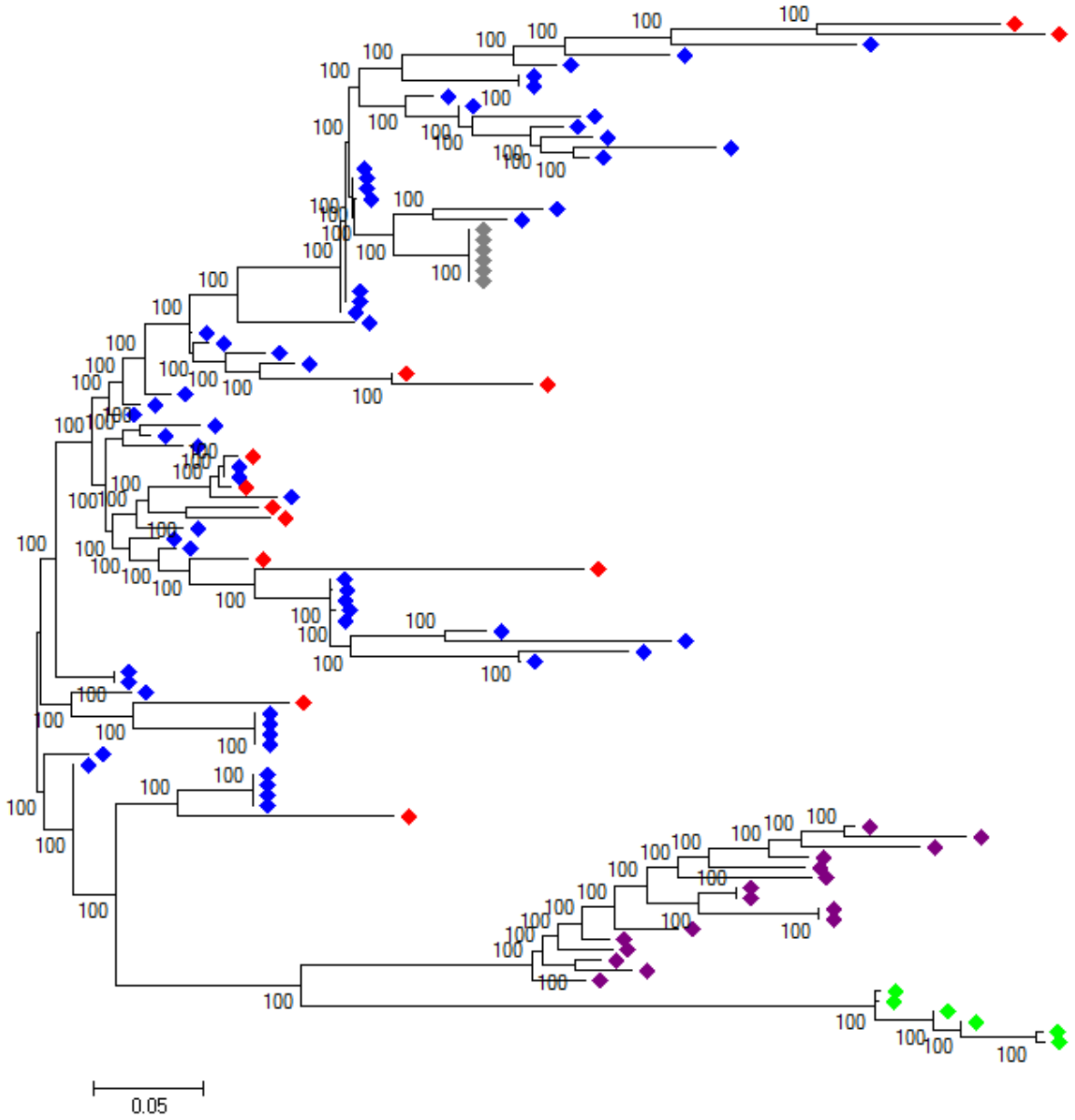


Figure S7.112 Chromosome 2: 19485001 - 19520001

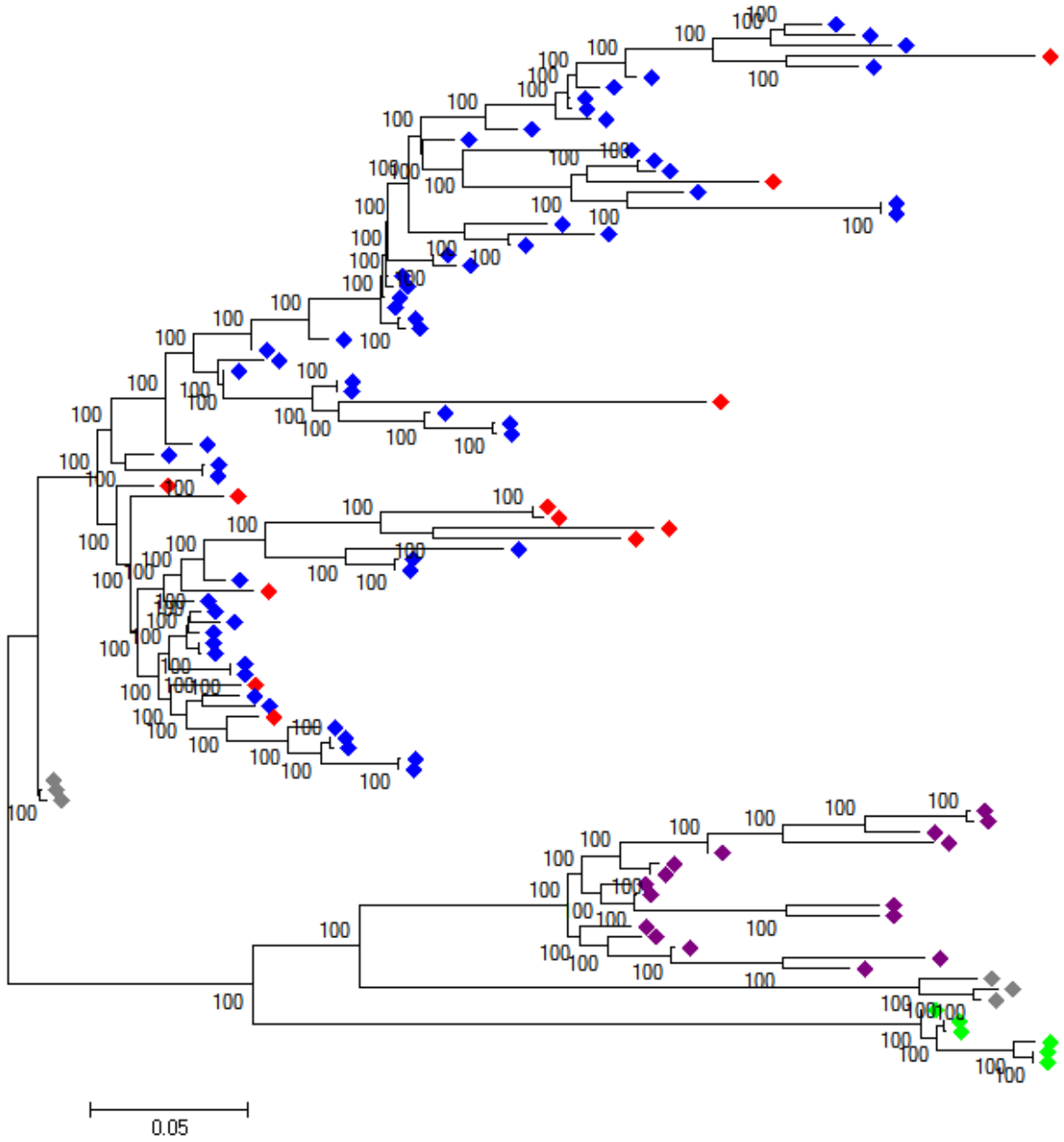


Figure S7.113 Chromosome 2: 40185001 - 40245001

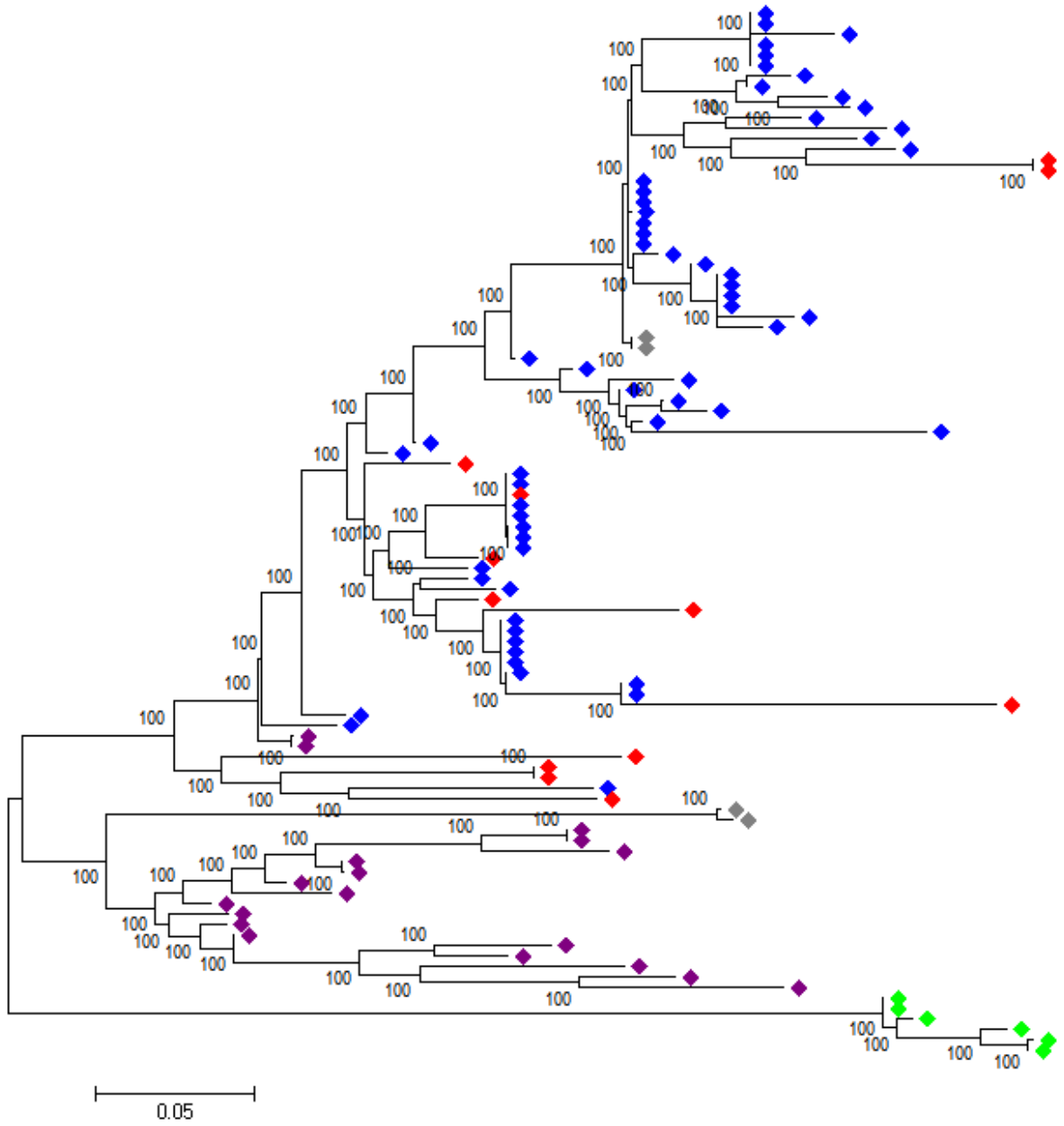


Figure S7.114 Chromosome 2: 40515001 - 40545001

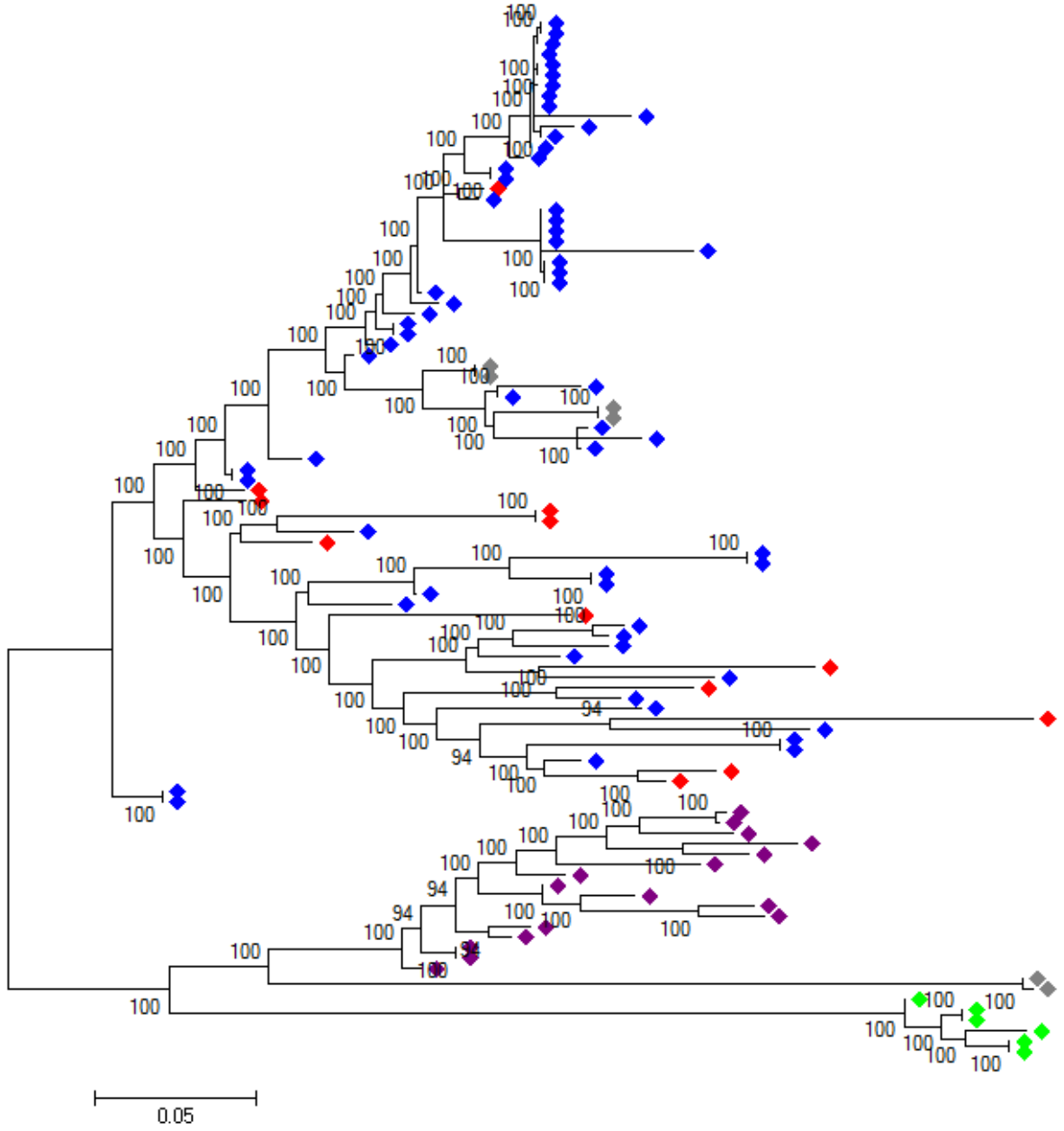


Figure S7.115 Chromosome 2: 40750001 - 40775001

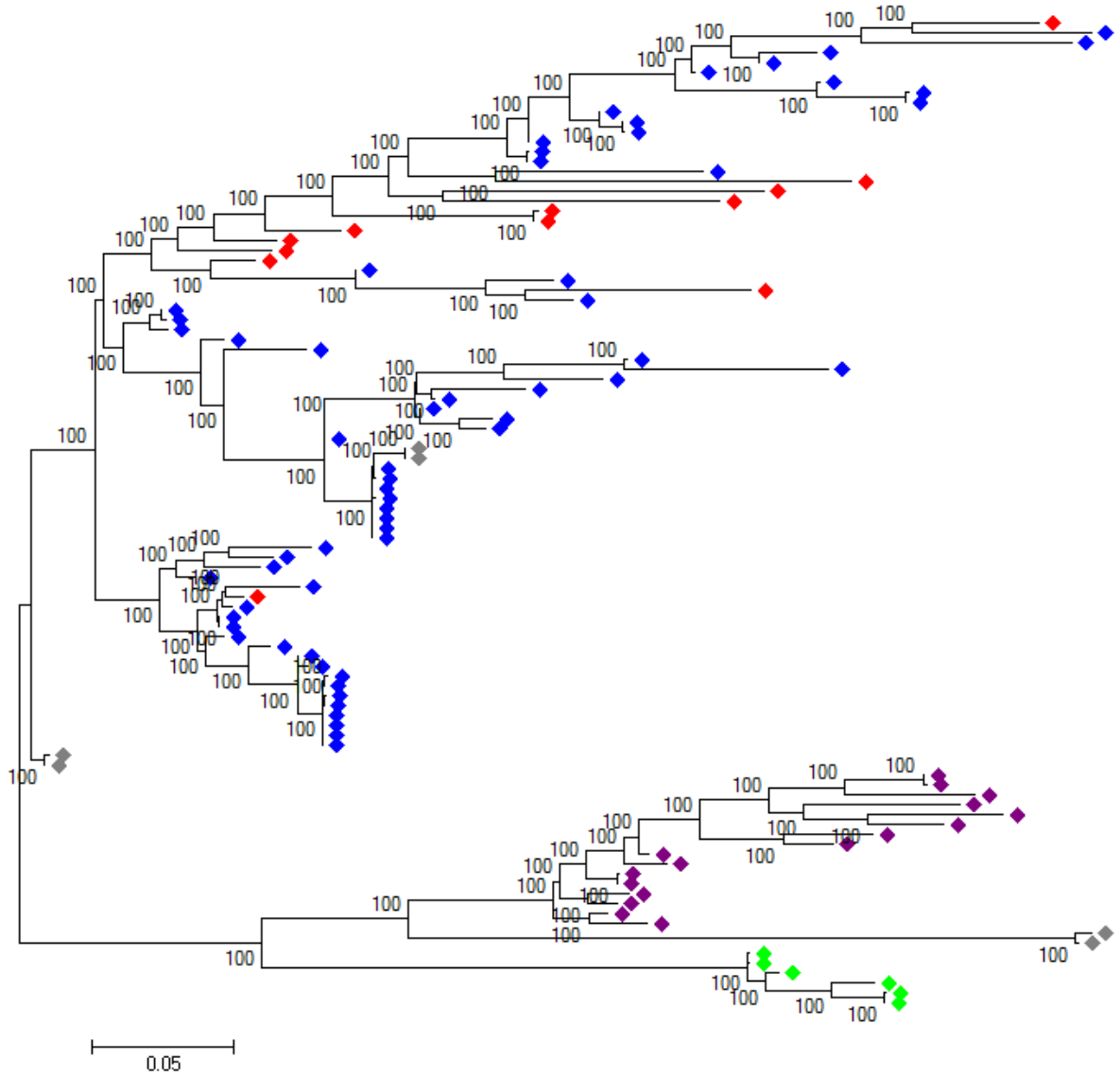


Figure S7.116 Chromosome 2: 40815001 - 40850001

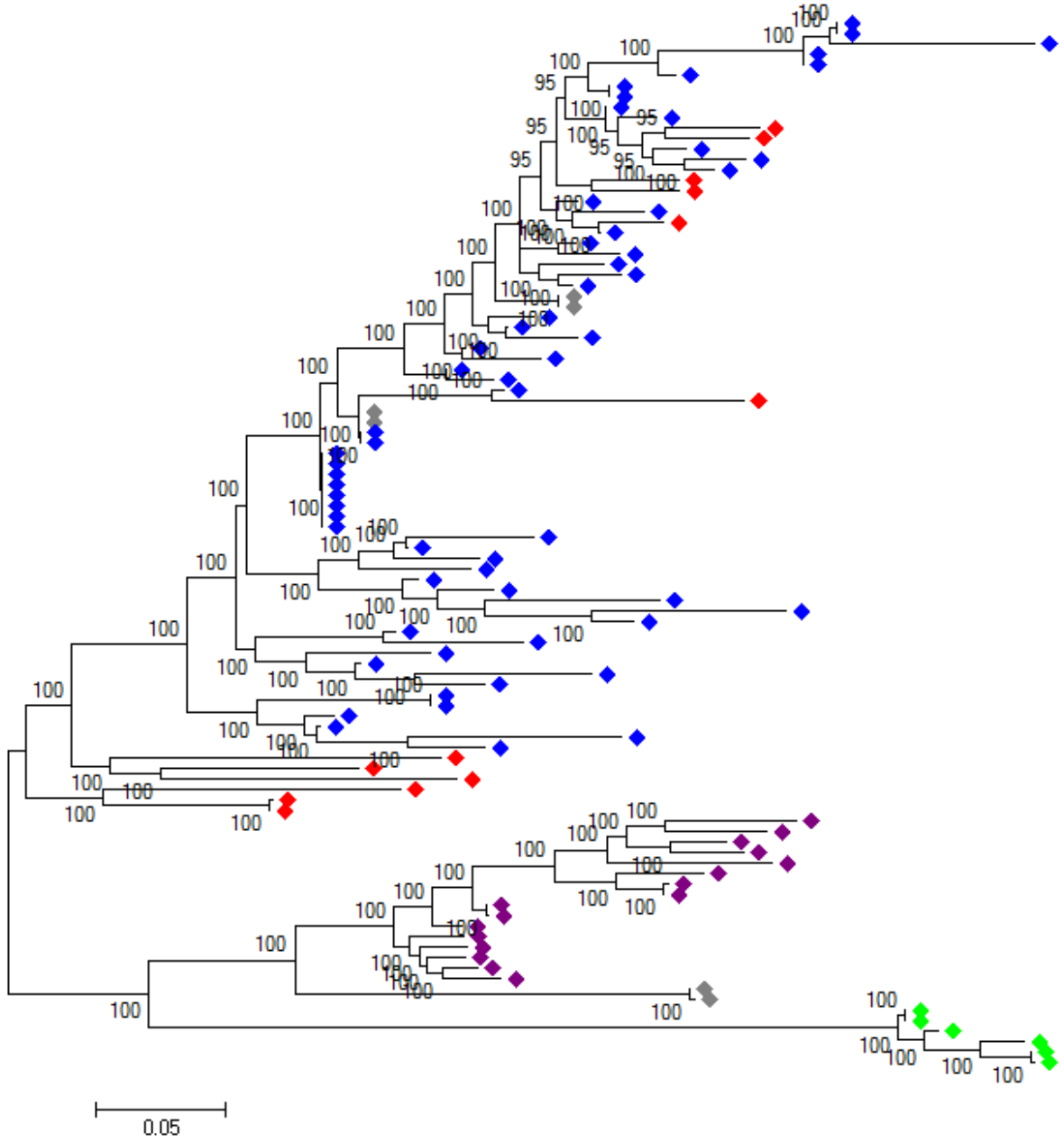


Figure S7.117 Chromosome 2: 40995001 - 41045001

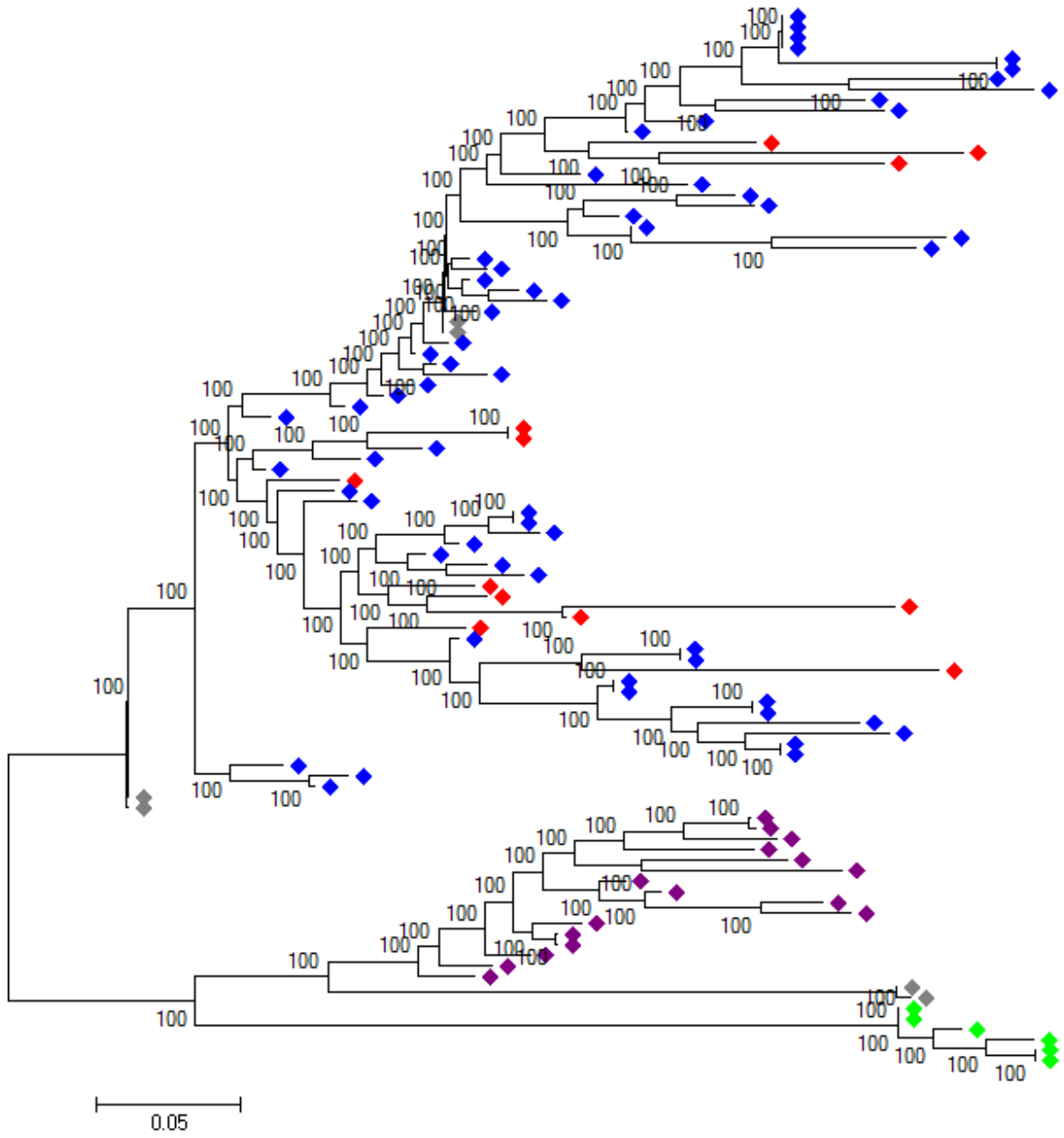


Figure S7.118 Chromosome 2: 41415001 - 41465001

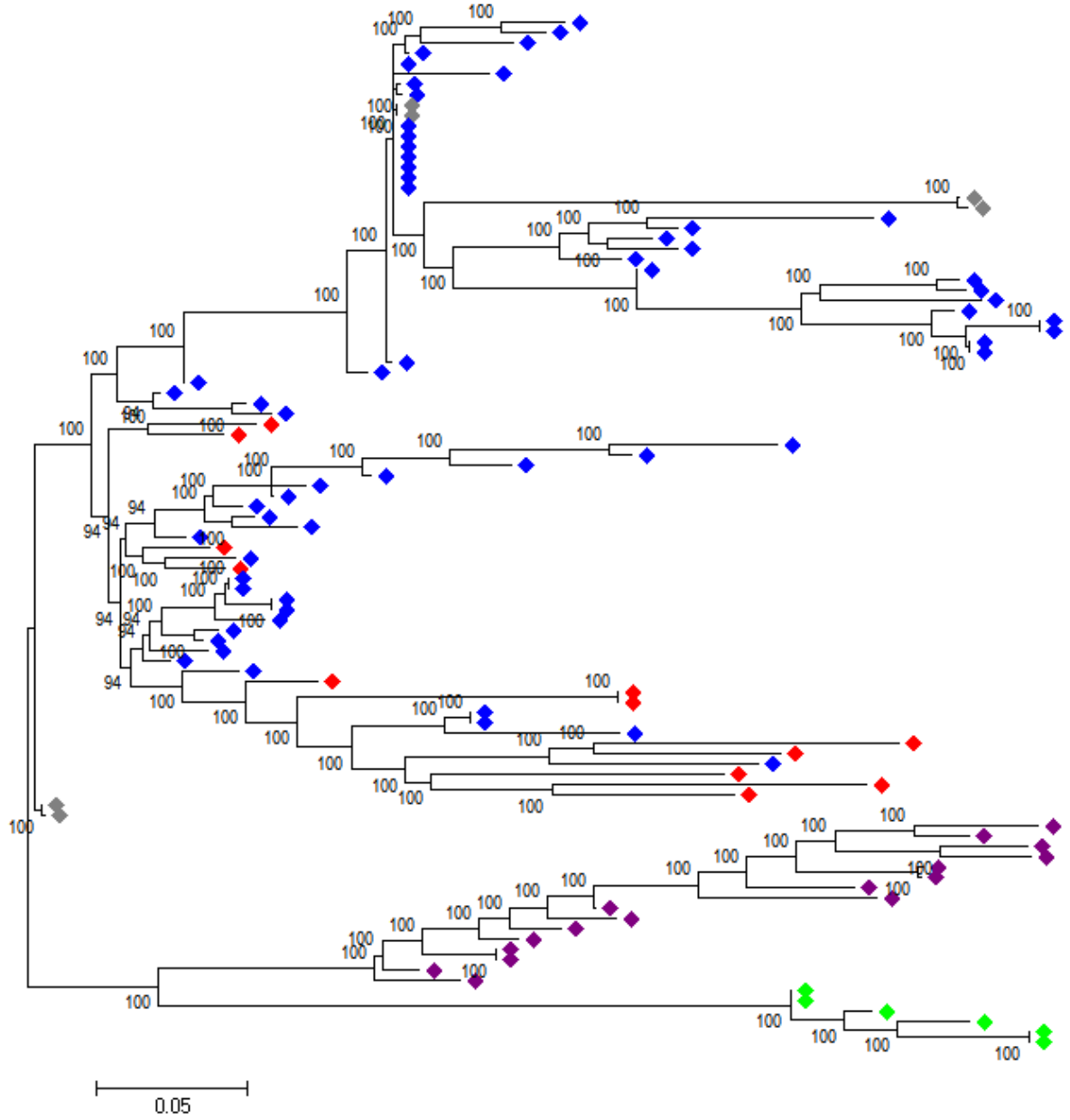


Figure S7.119 Chromosome 2: 41560001 - 41580001

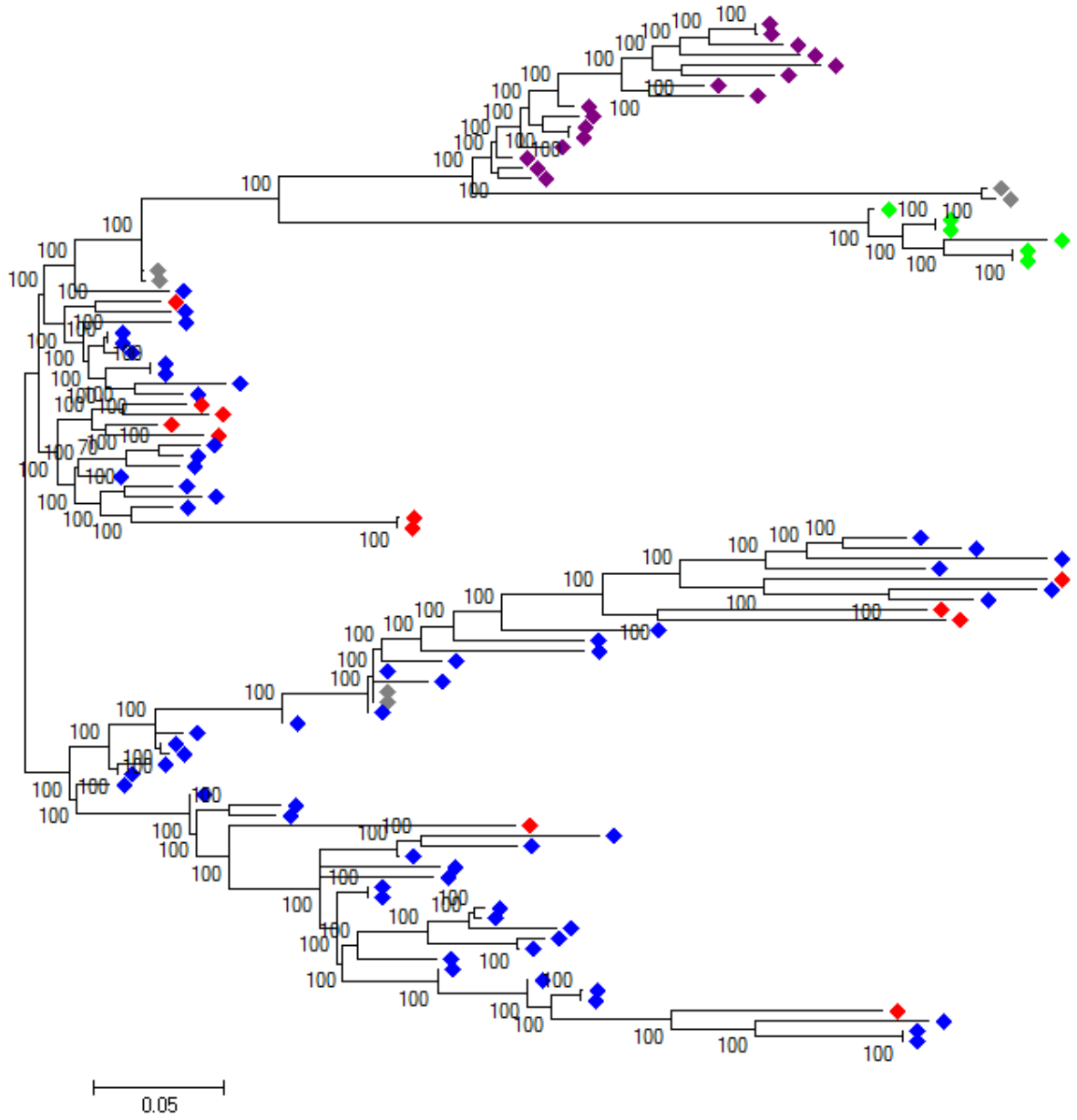


Figure S7.120 Chromosome 2: 41610001 - 41640001

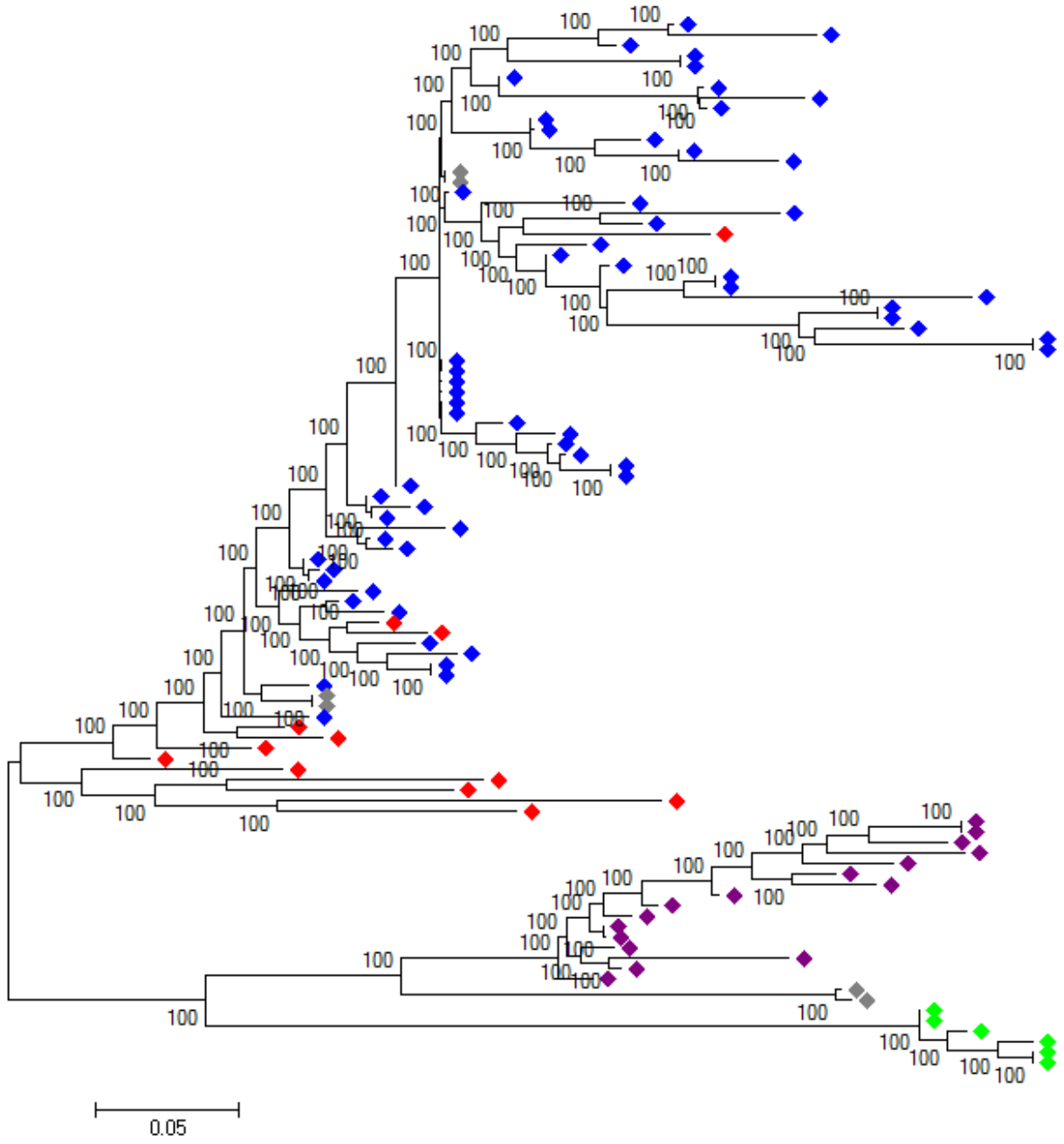


Figure S7.121 Chromosome 2: 41930001 - 41950001

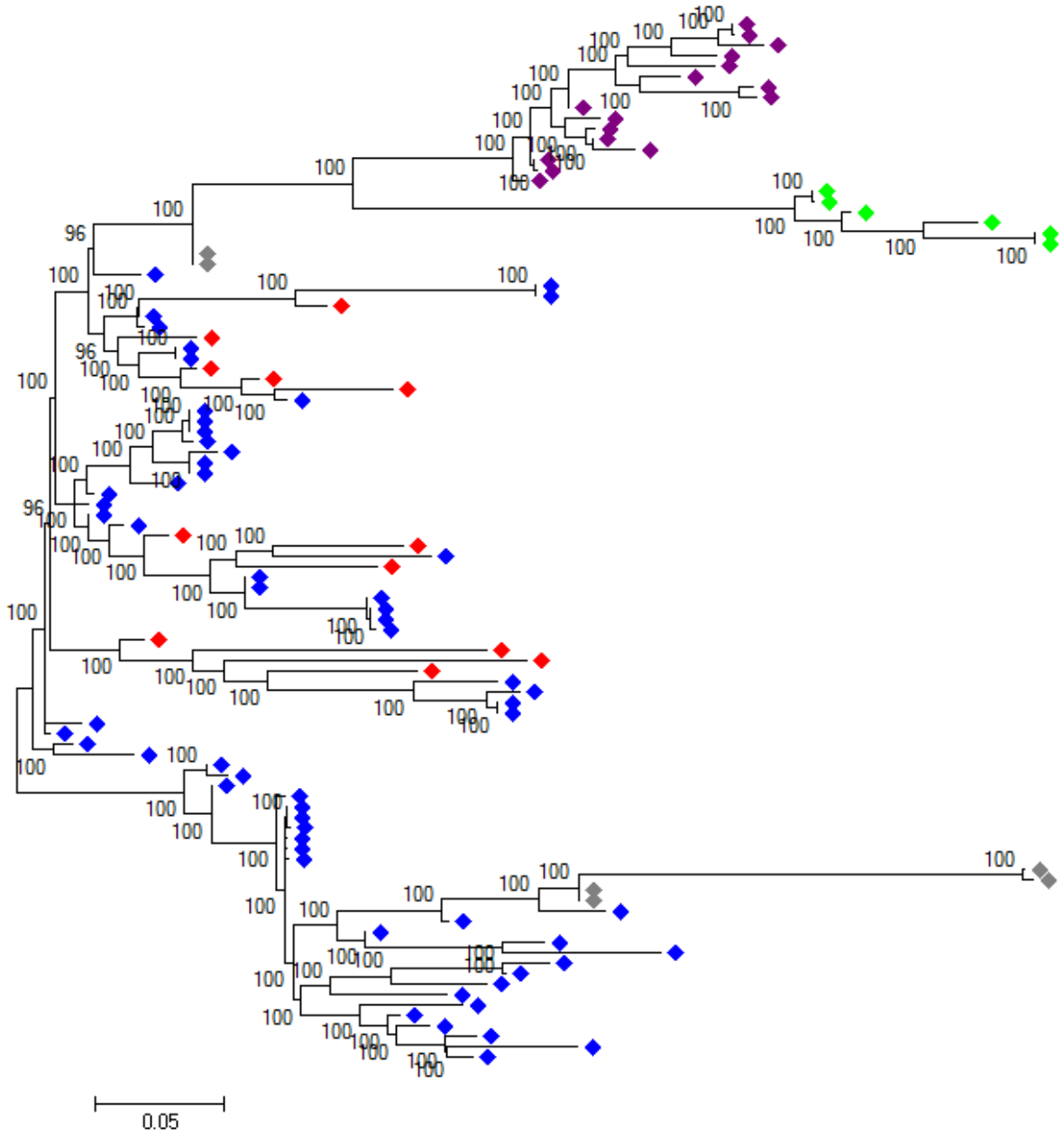


Figure S7.122 Chromosome 2: 42045001 - 42070001

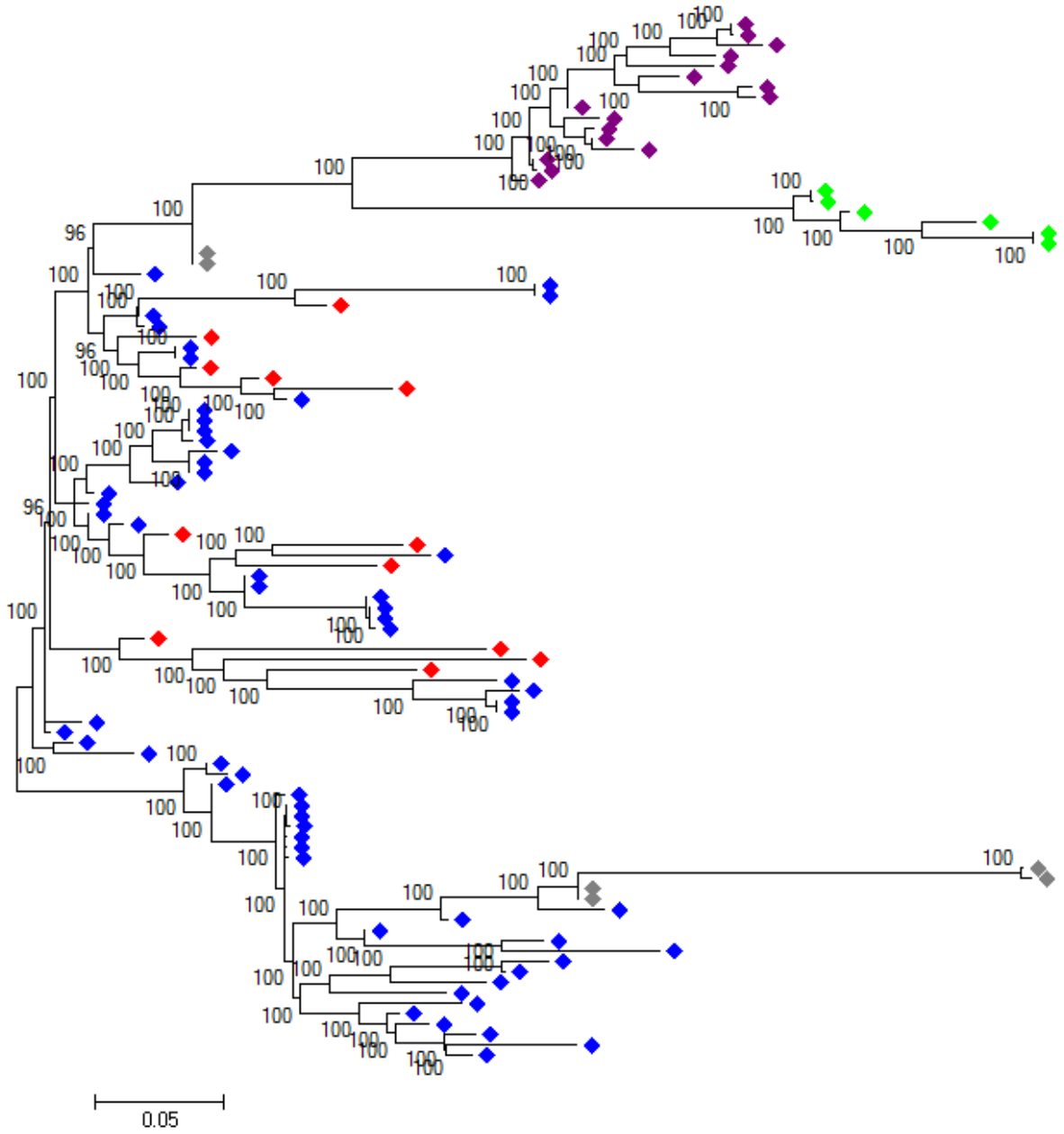


Figure S7.123 Chromosome 2: 42165001 - 42190001

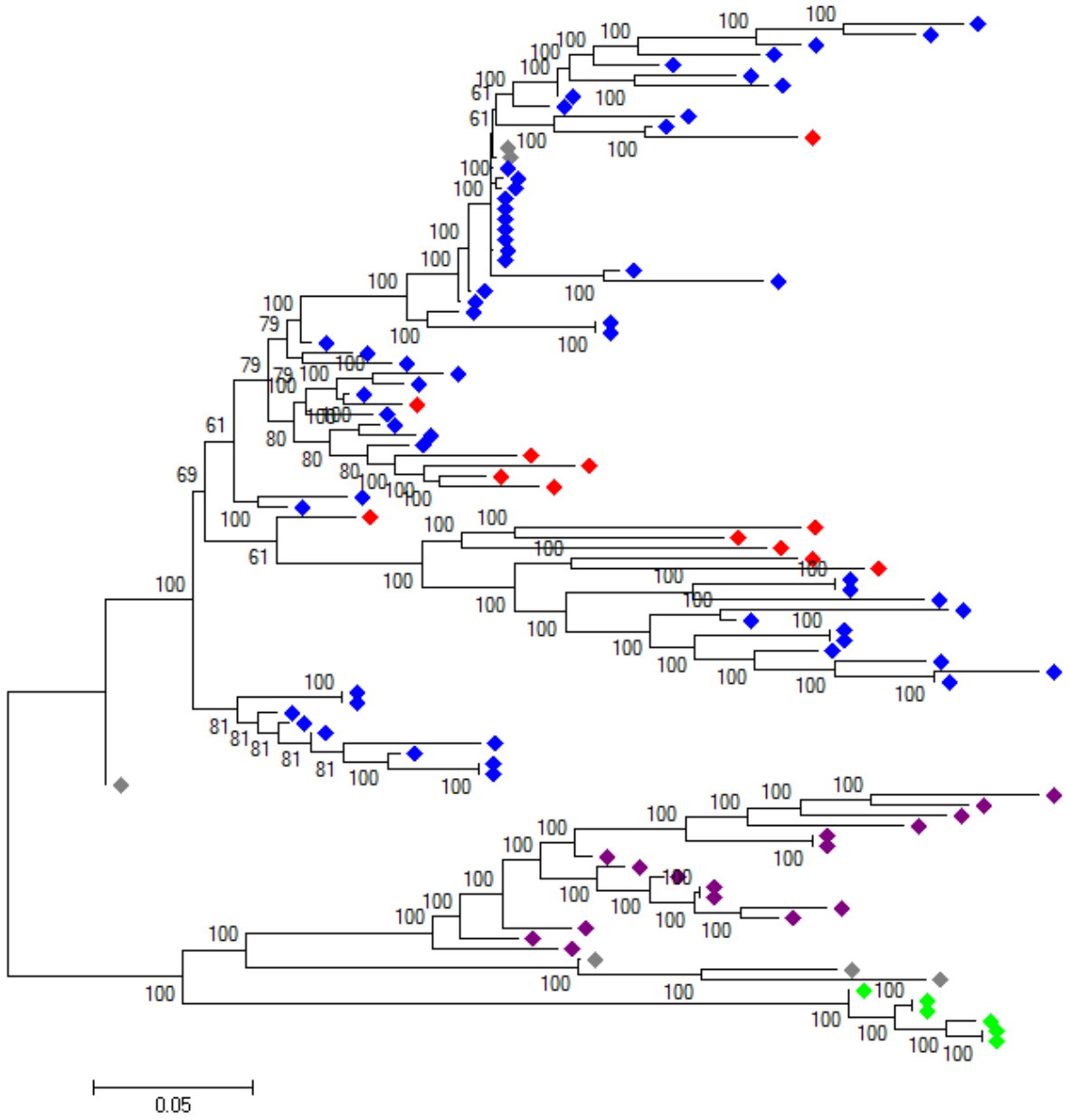


Figure S7.124 Chromosome 2: 67630001 - 67655001

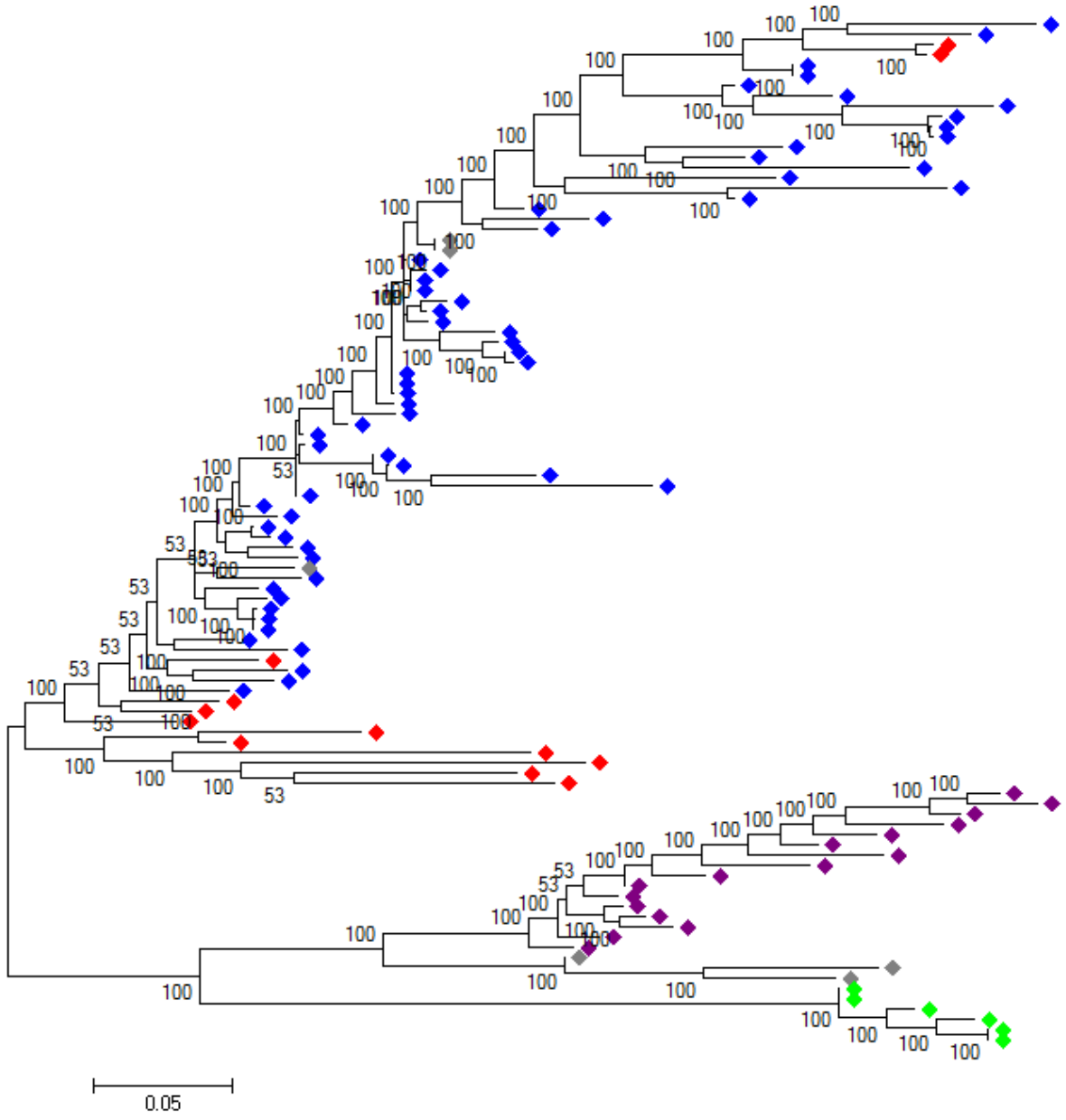


Figure S7.125 Chromosome 2: 68220001 - 68290001

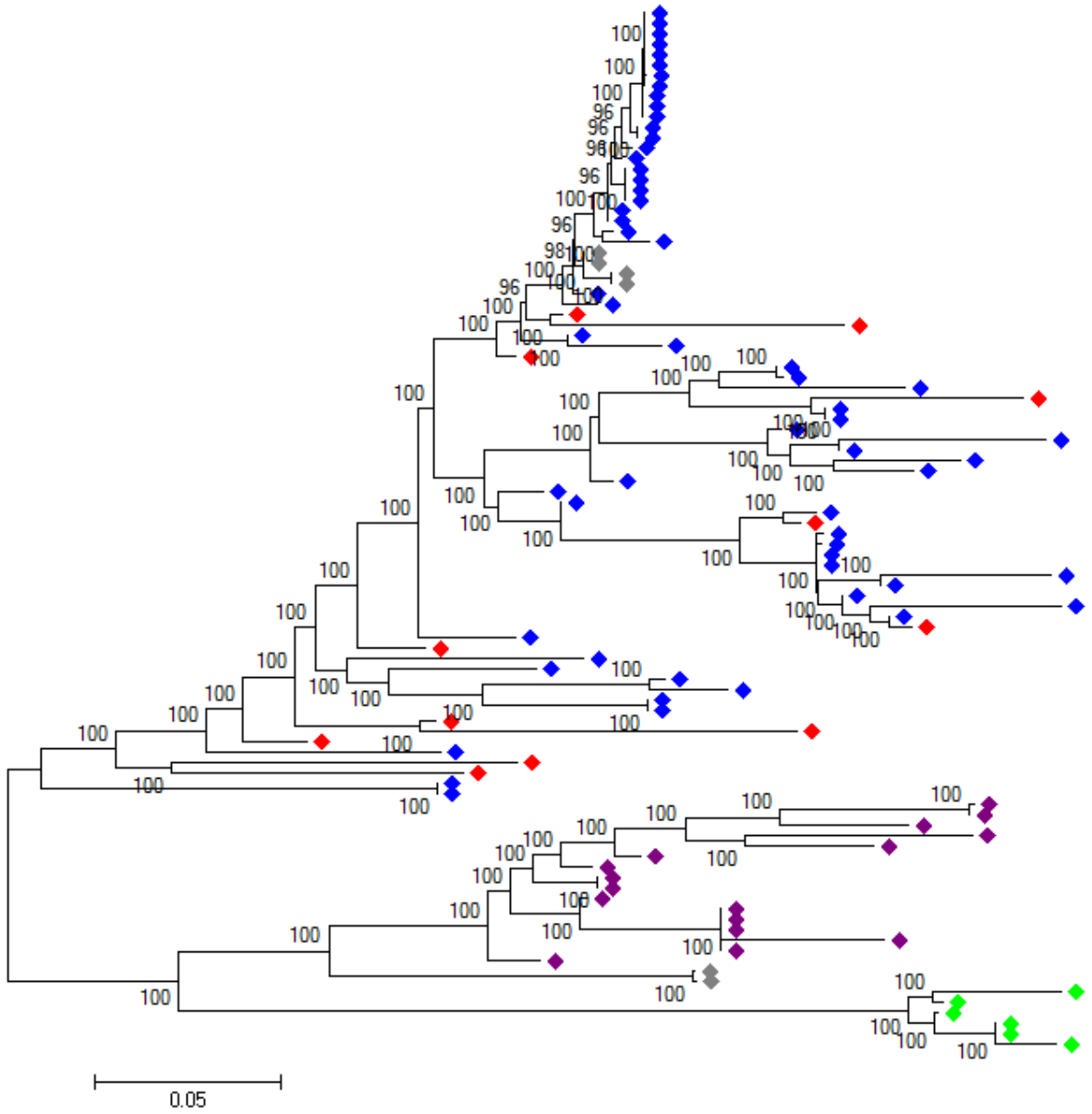


Figure S7.126 Chromosome 3: 20460001 - 20505001

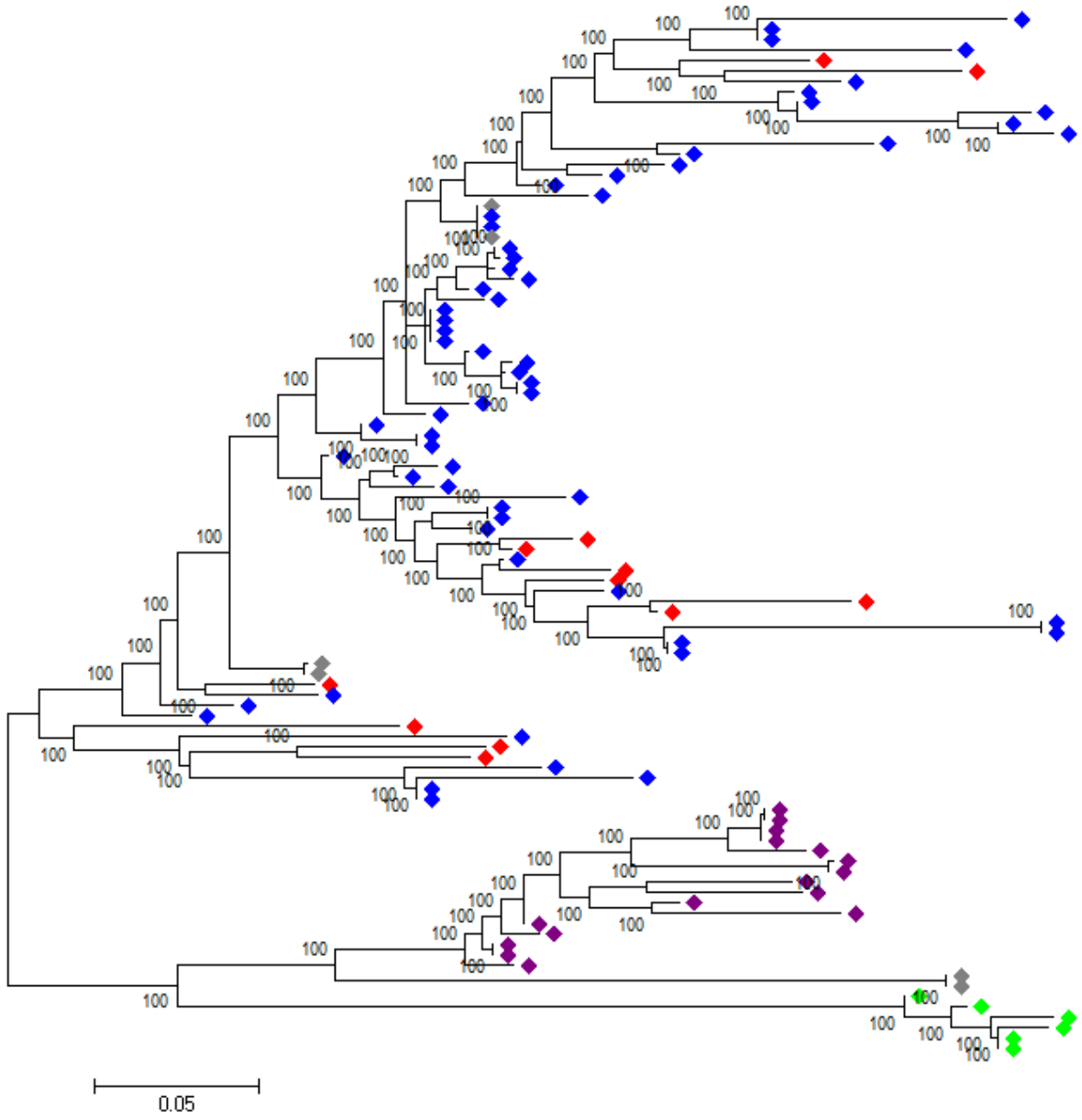


Figure S7.127 Chromosome 3: 20600001 - 20635001

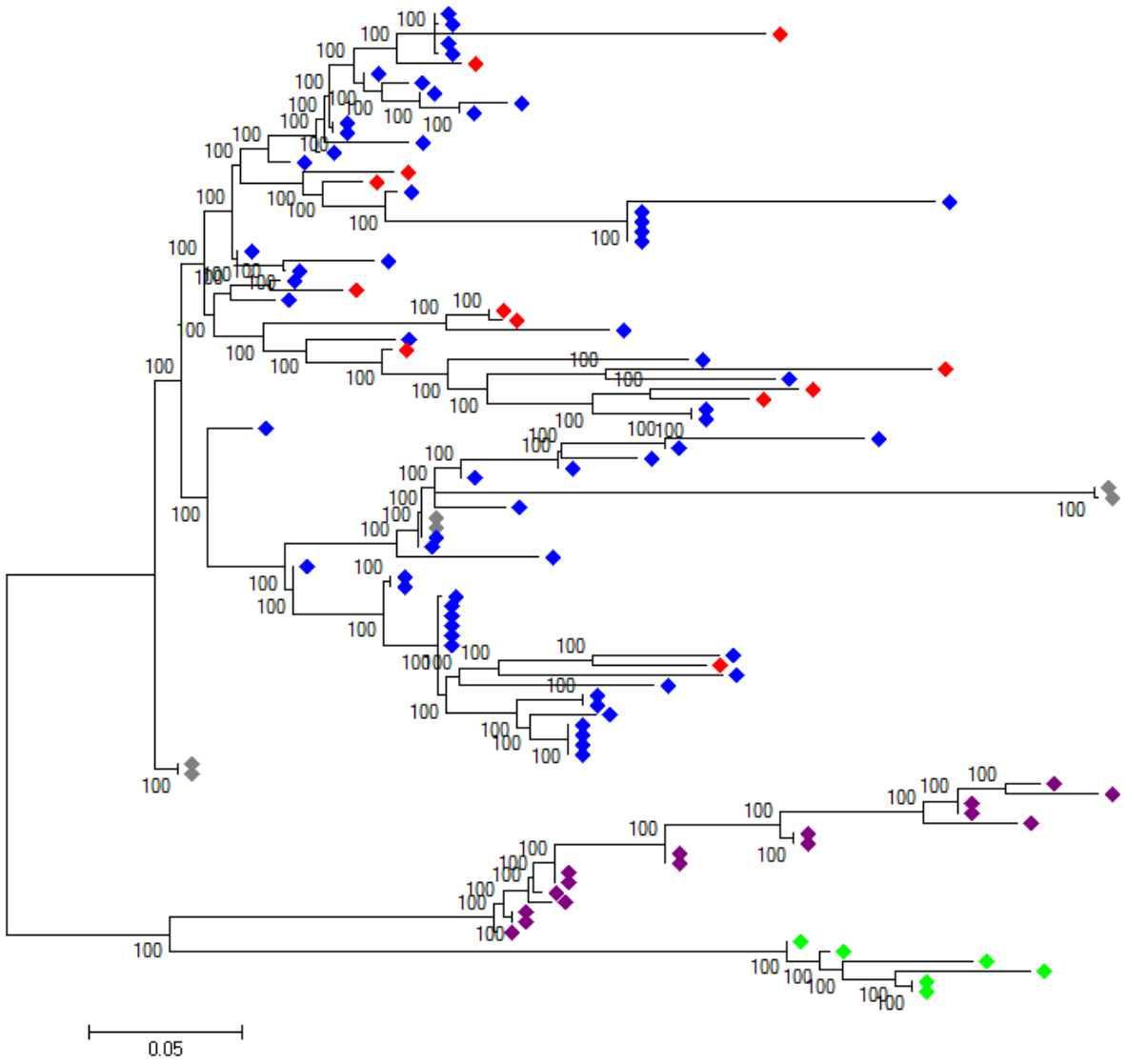


Figure S7.128 Chromosome 3: 20705001 - 20725001

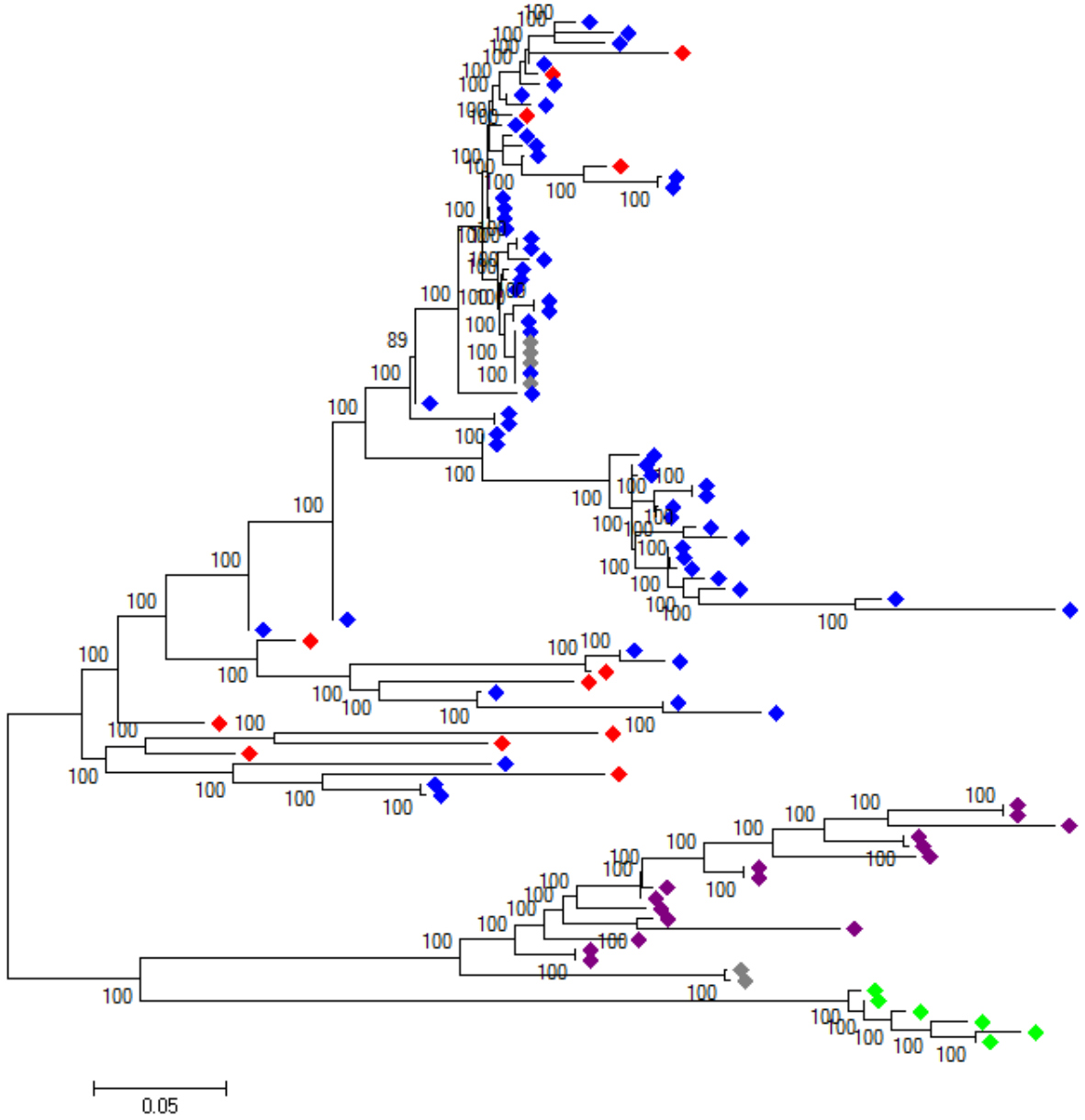


Figure S7.129 Chromosome 3: 21115001 - 21145001

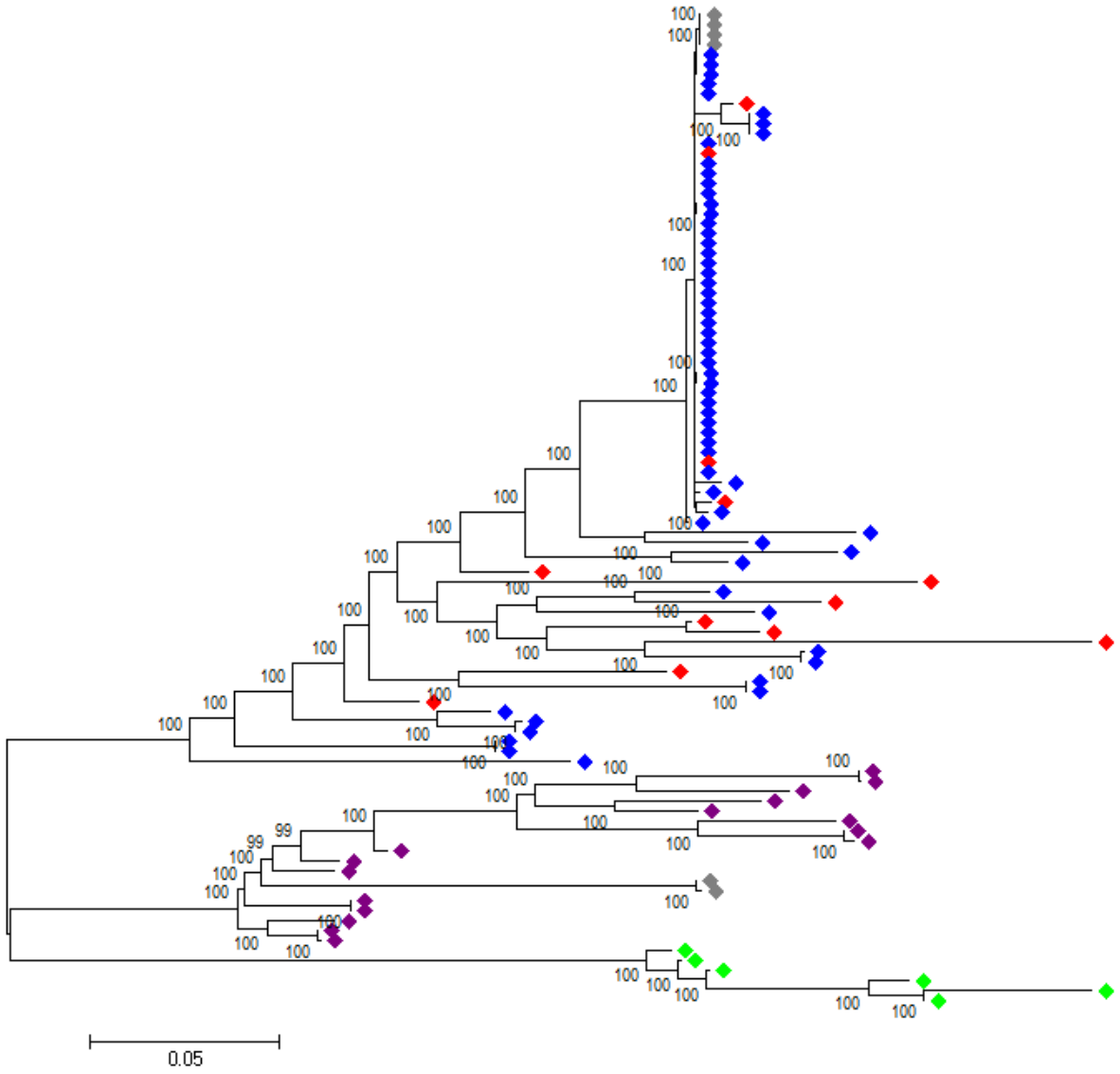


Figure S7.130 Chromosome 3: 21220001 - 21250001

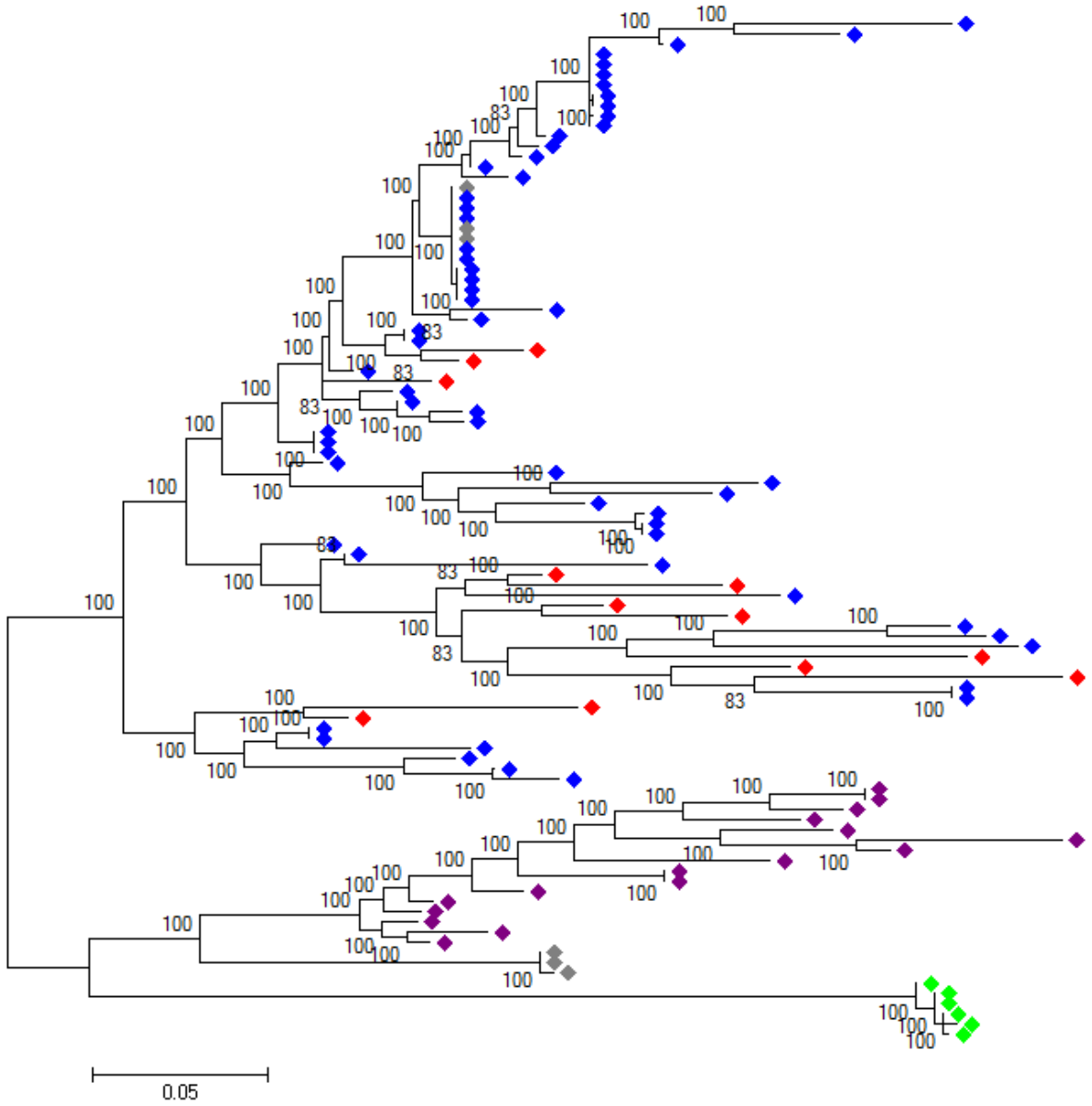


Figure S7.131 Chromosome 4: 16985001 - 17035001

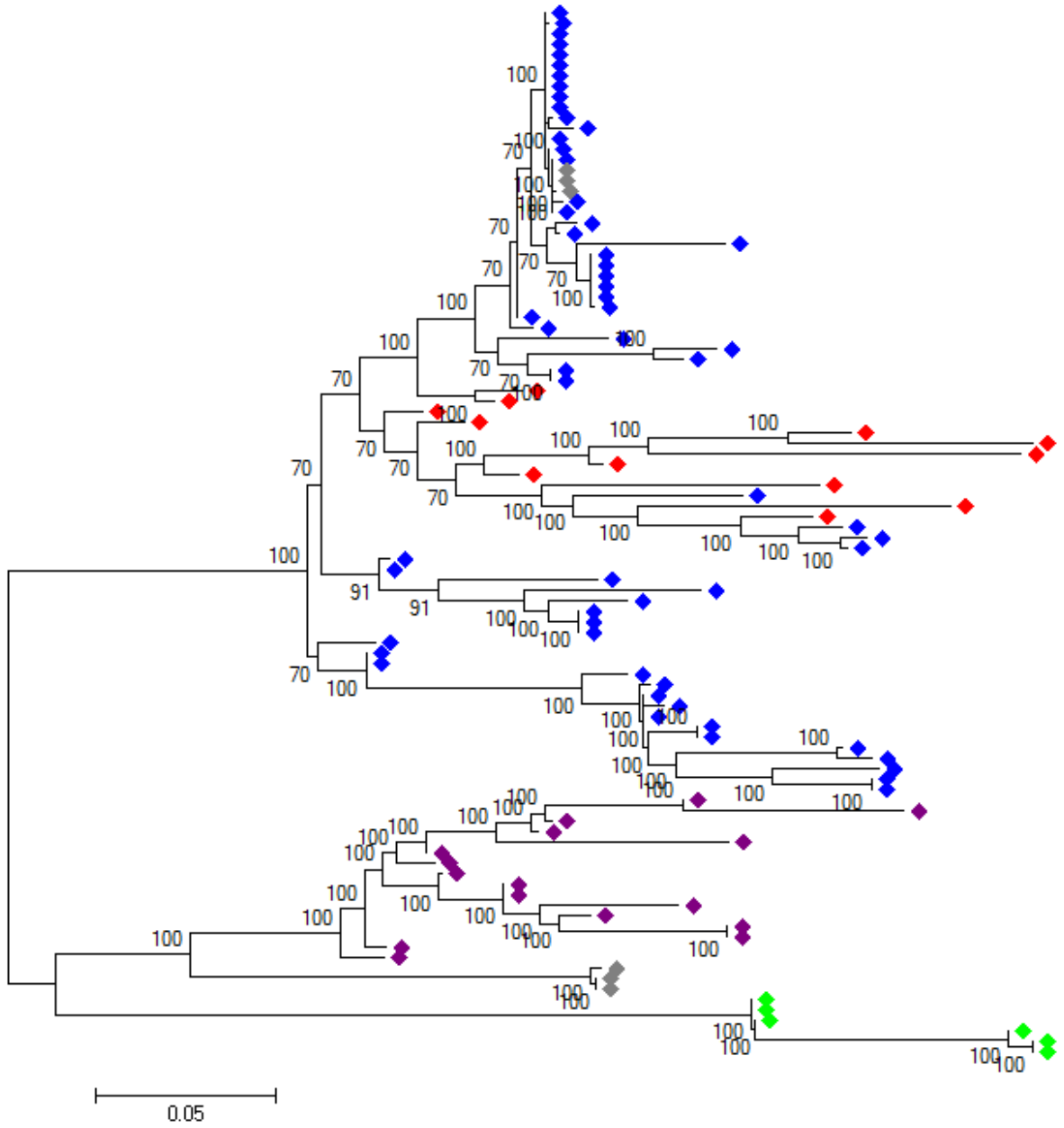


Figure S7.132 Chromosome 4: 17200001 - 17230001

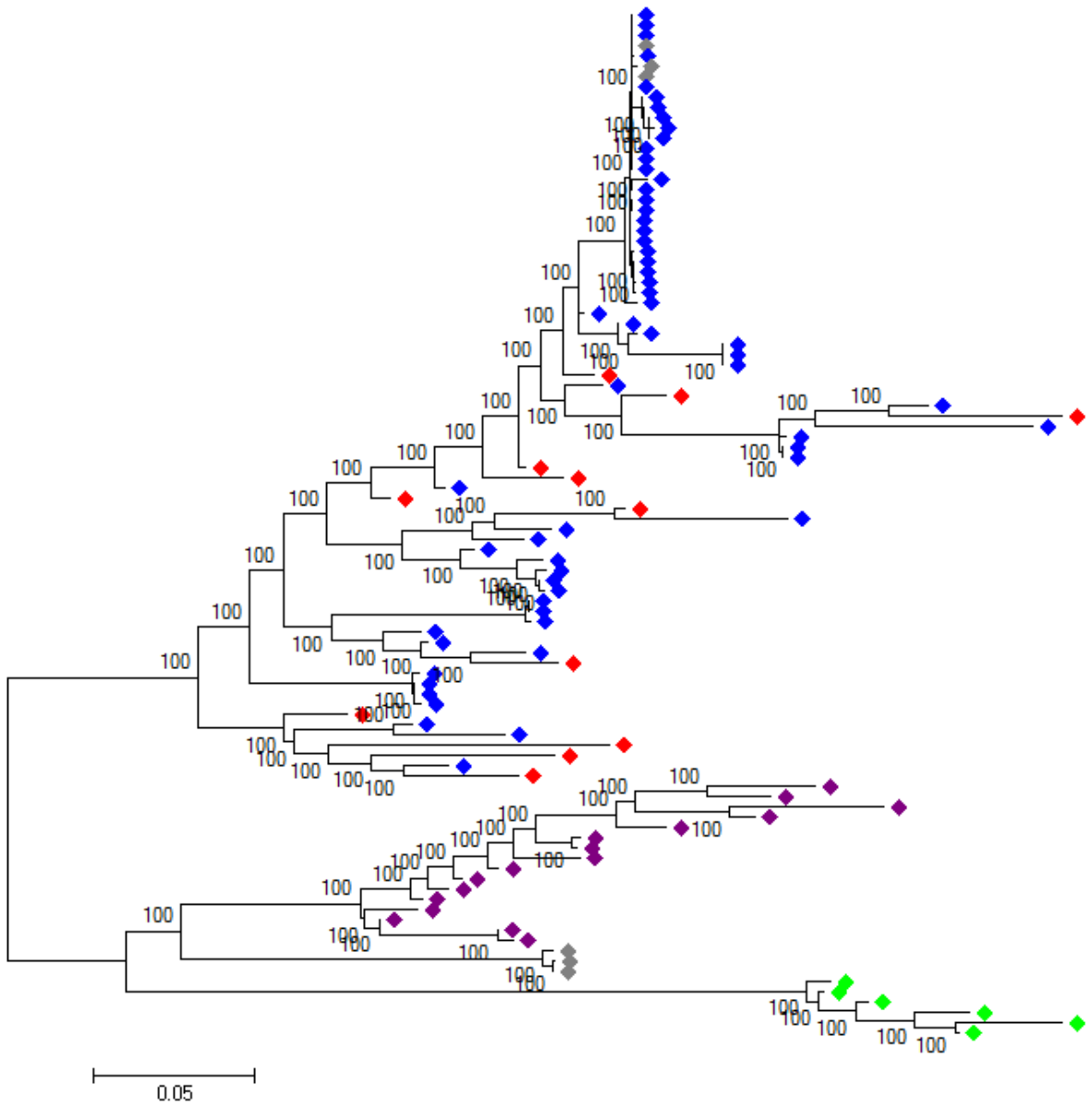


Figure S7.133 Chromosome 4: 19880001 - 19970001

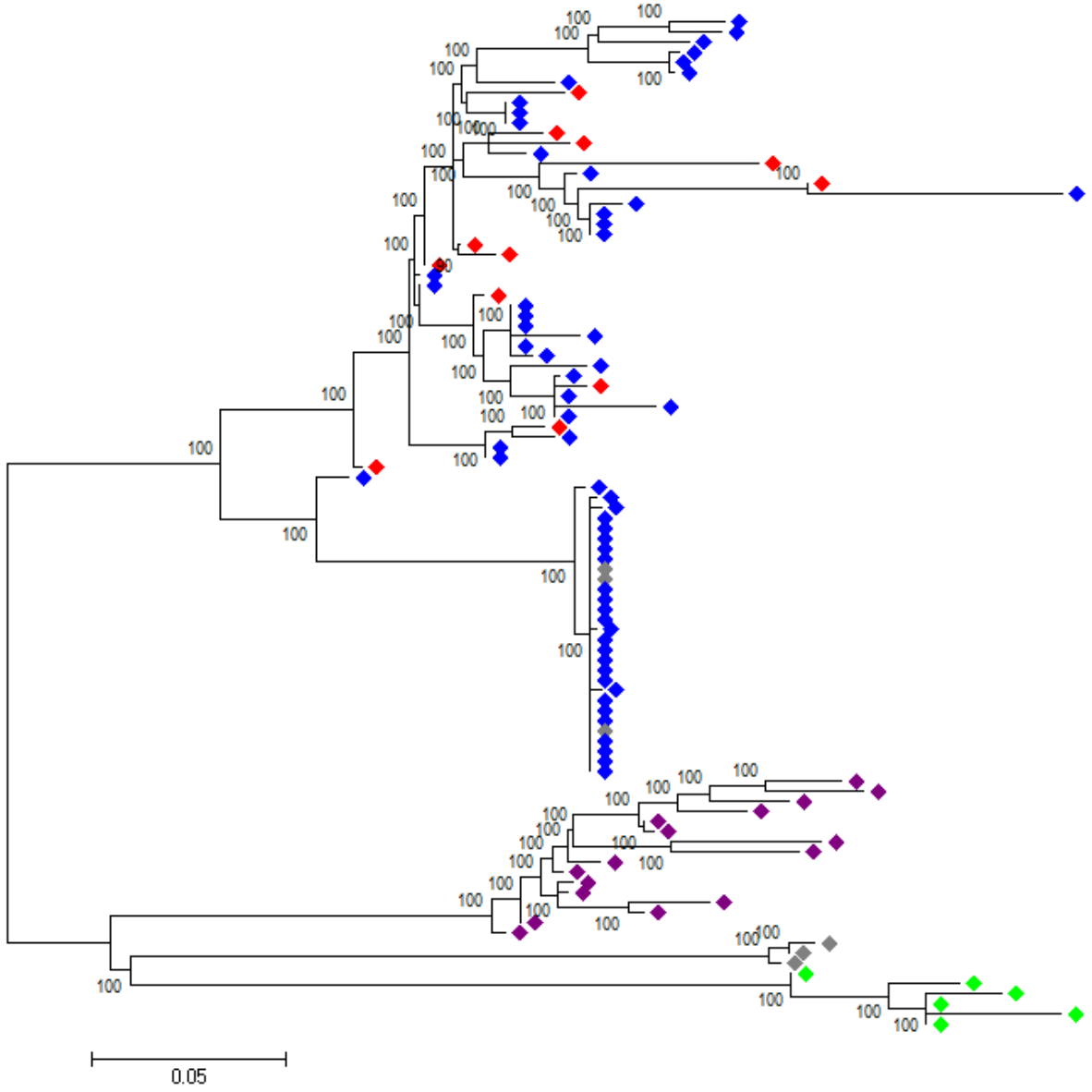


Figure S7.134 Chromosome 4: 19990001 - 20010001

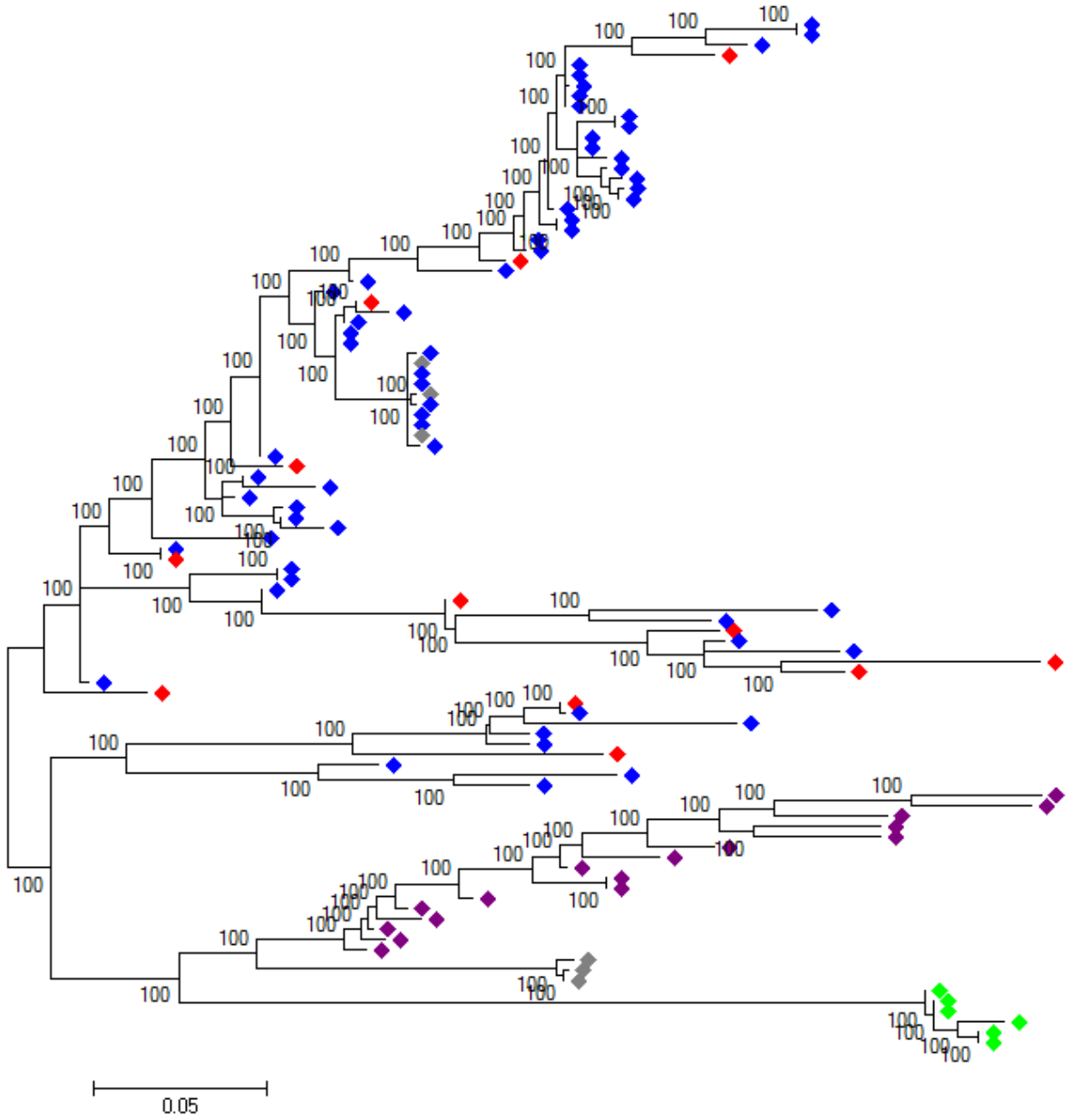


Figure S7.135 Chromosome 4: 20320001 - 20345001

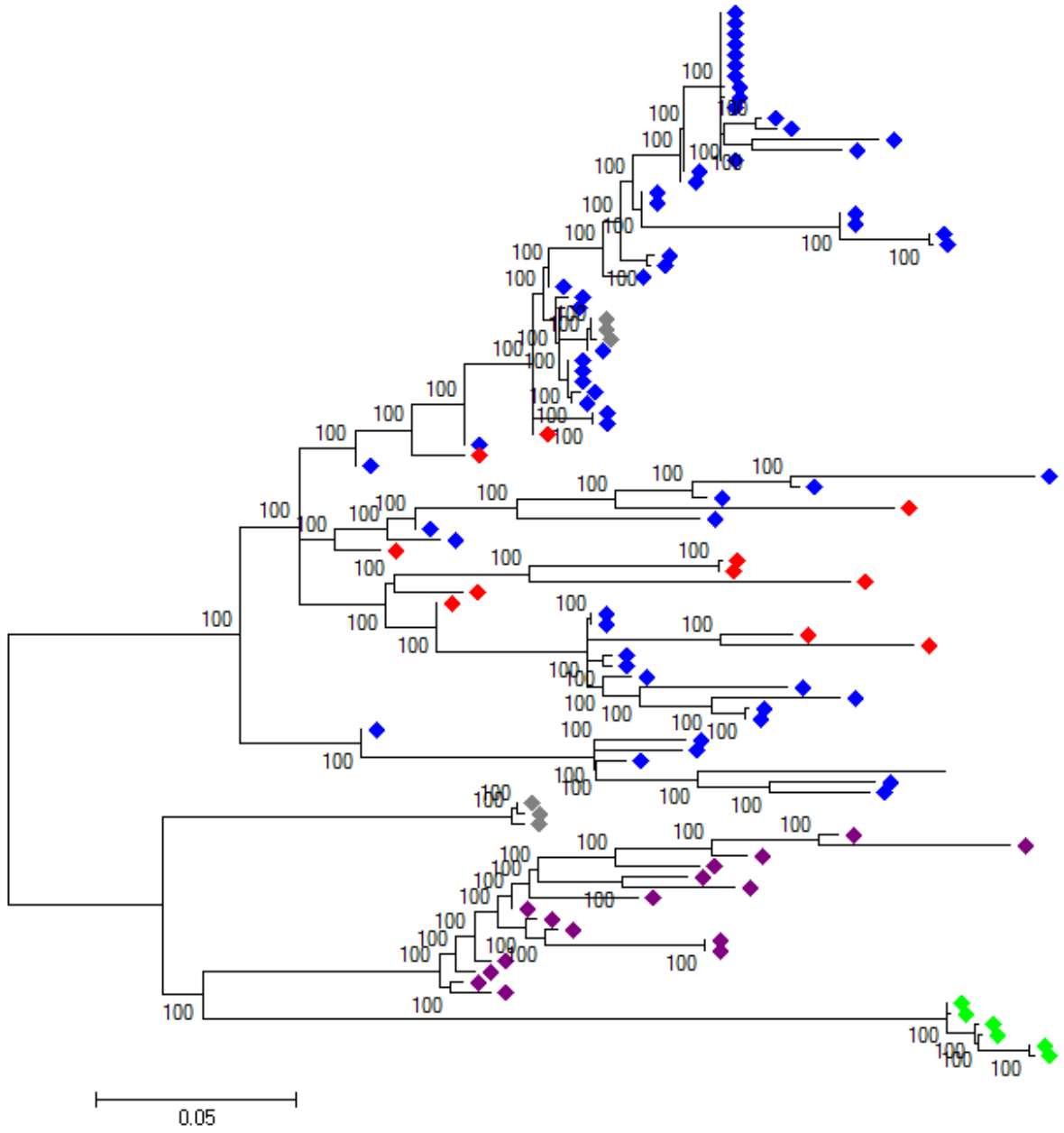


Figure S7.136 Chromosome 4: 20920001 - 20950001

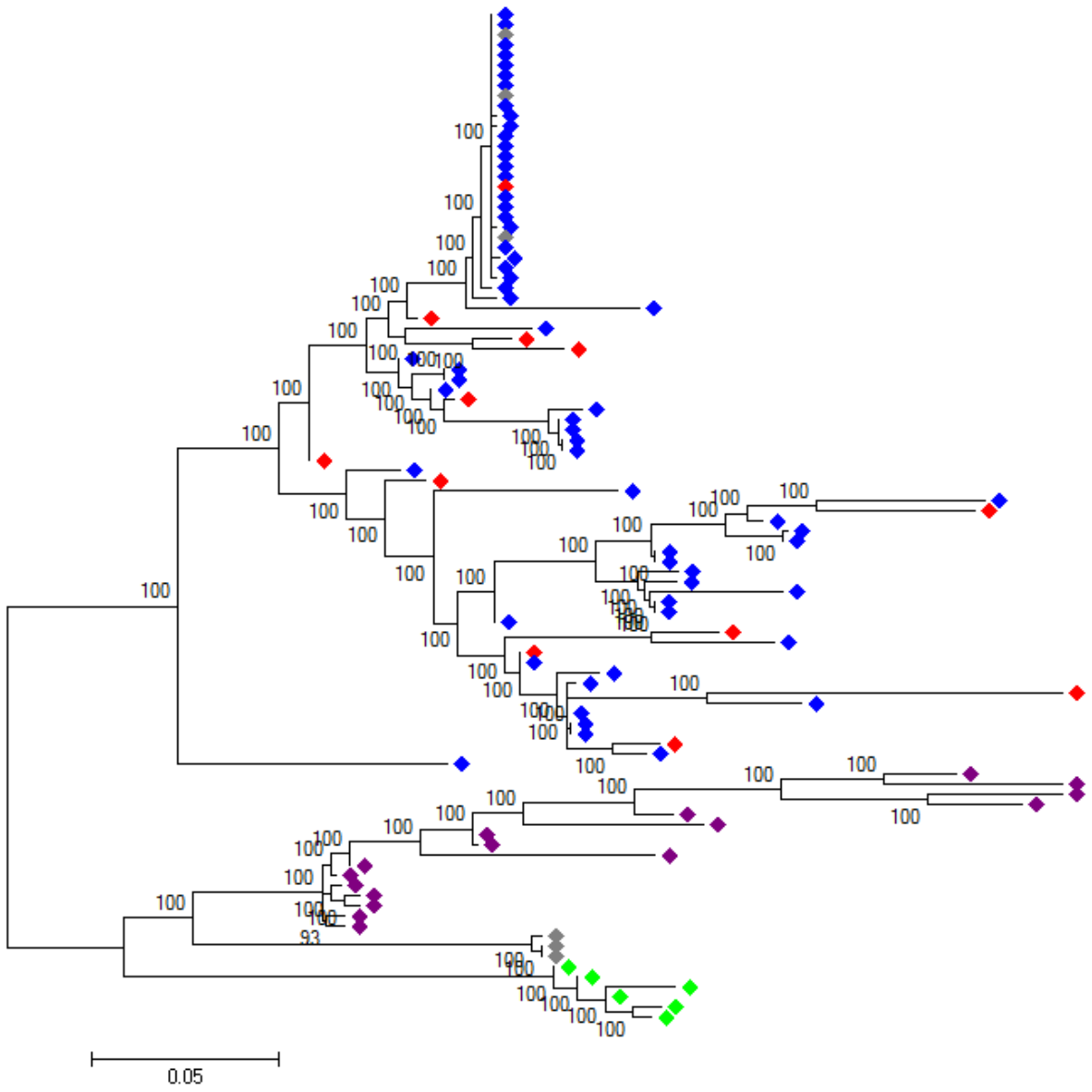


Figure S7.137 Chromosome 4: 21765001 - 21795001

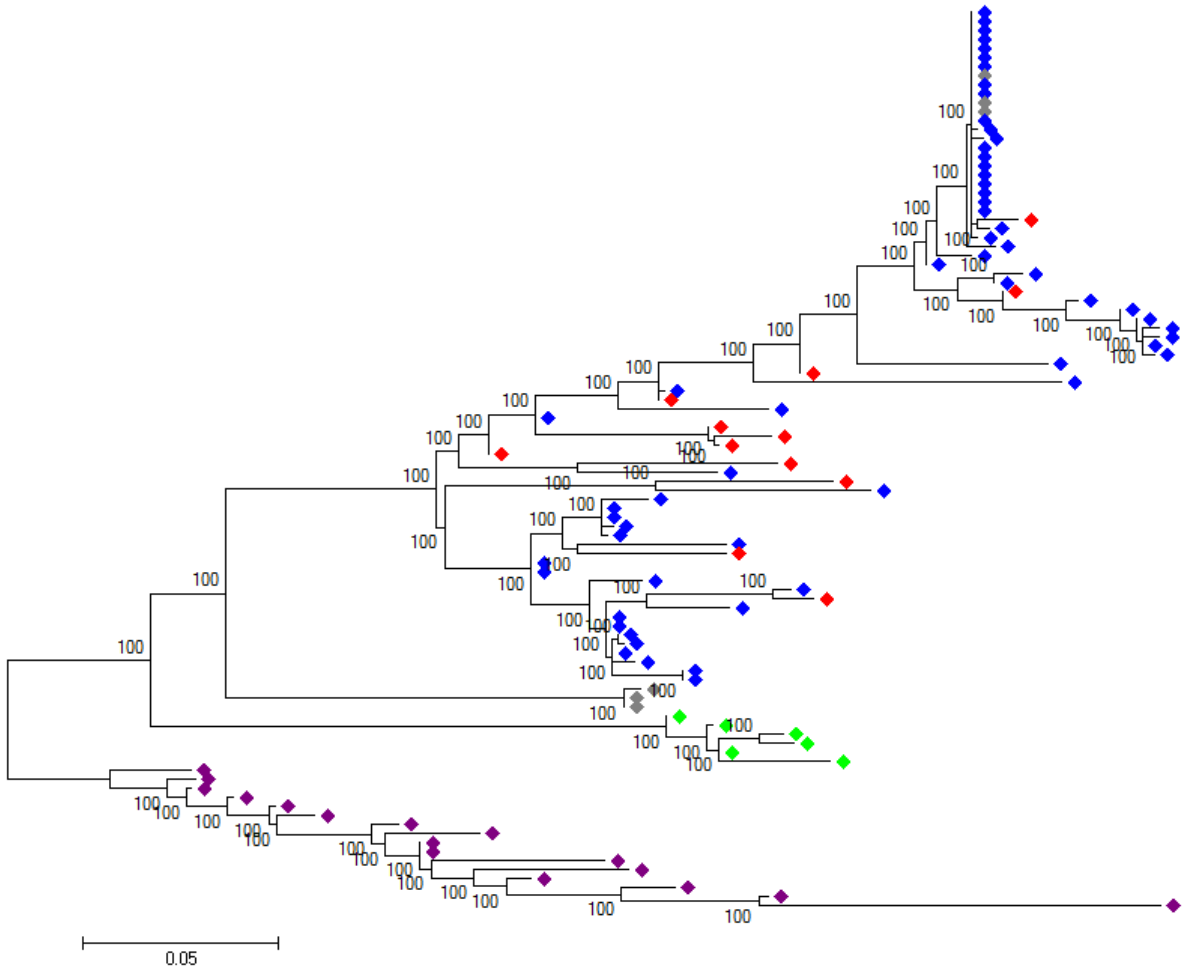


Figure S7.138 Chromosome 4: 21800001 - 21820001

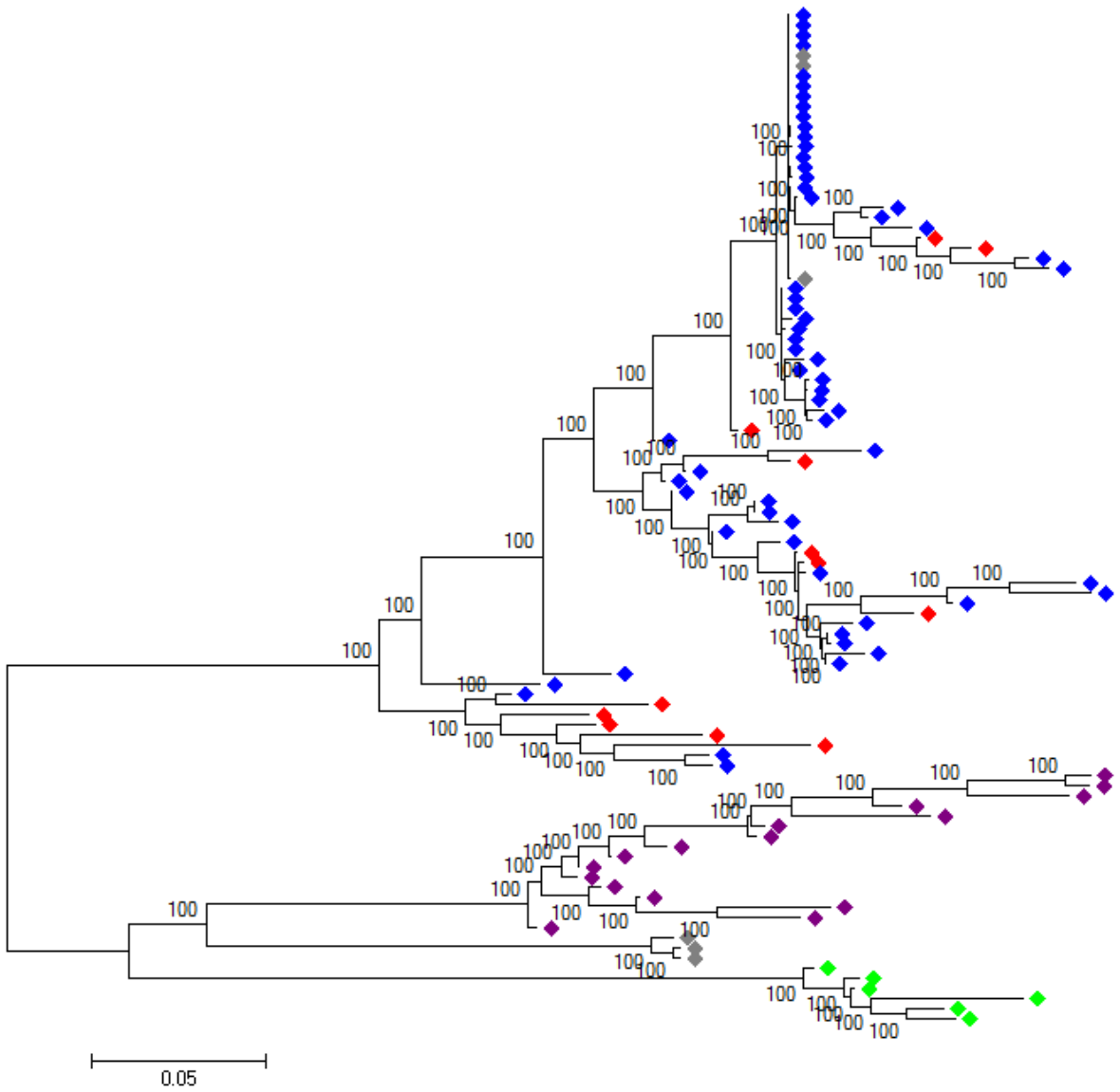


Figure S7.139 Chromosome 4: 21960001 - 22050001

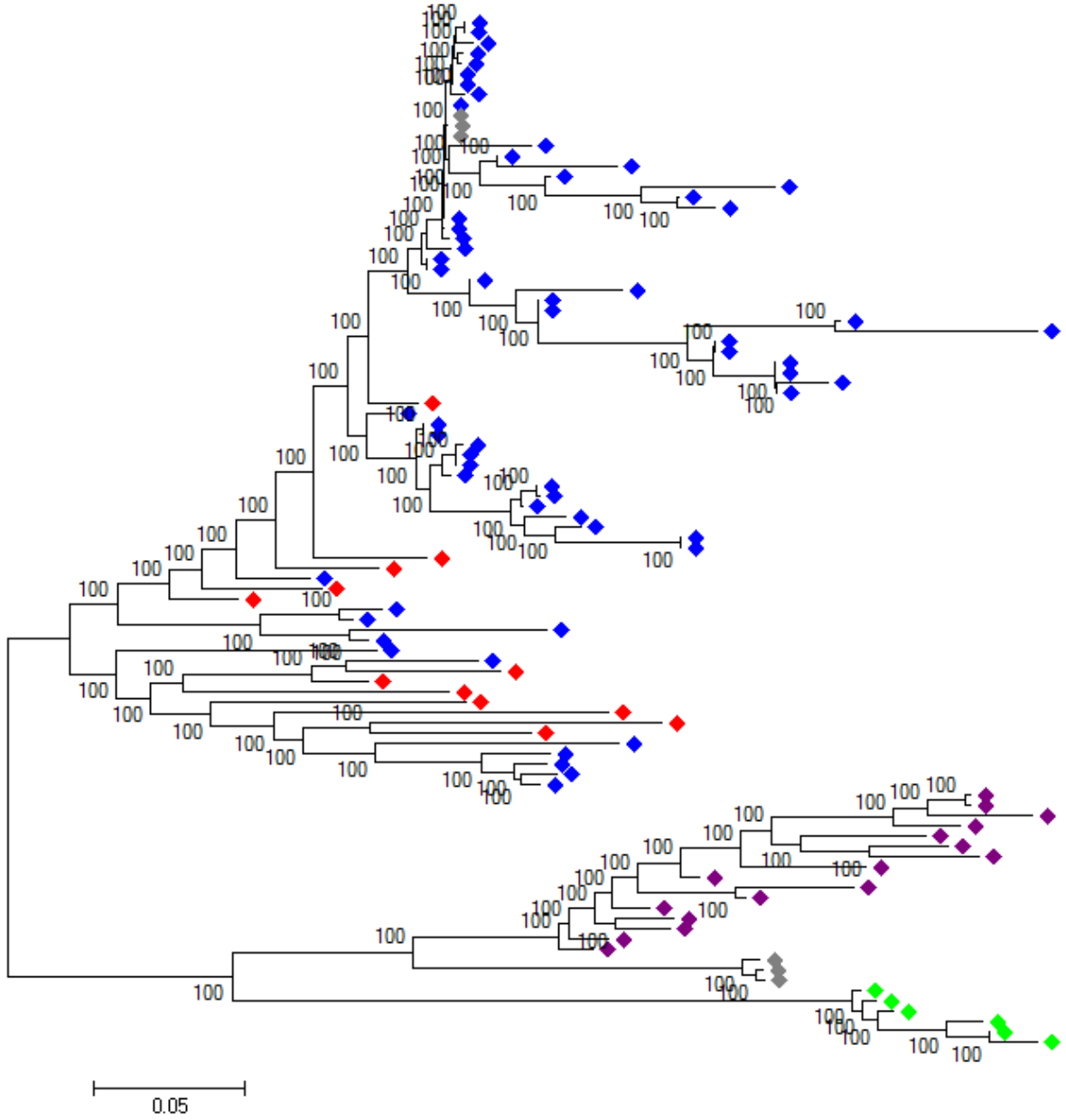


Figure S7.140 Chromosome 4: 23220001 - 23330001

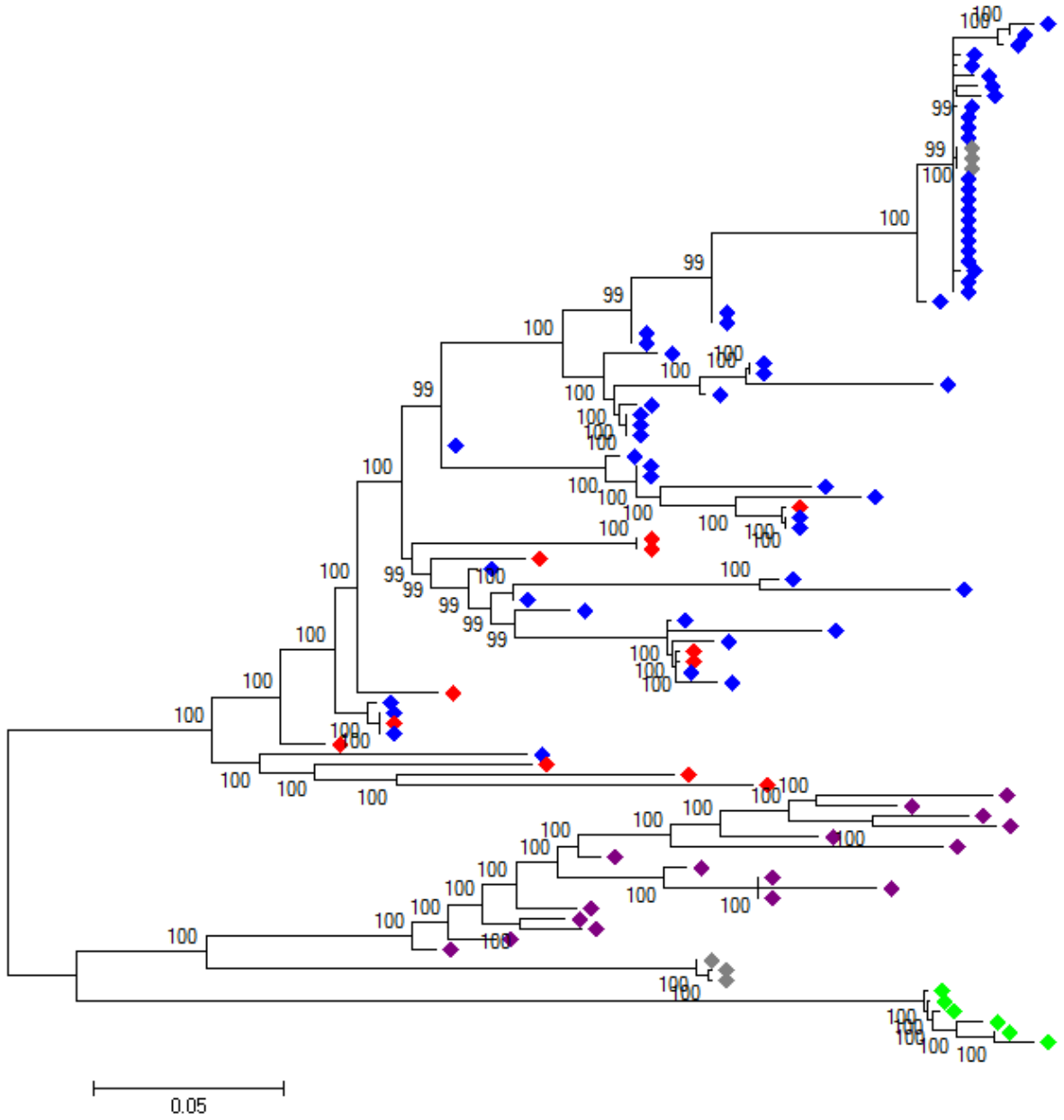


Figure S7.141 Chromosome 4: 23475001 - 23510001

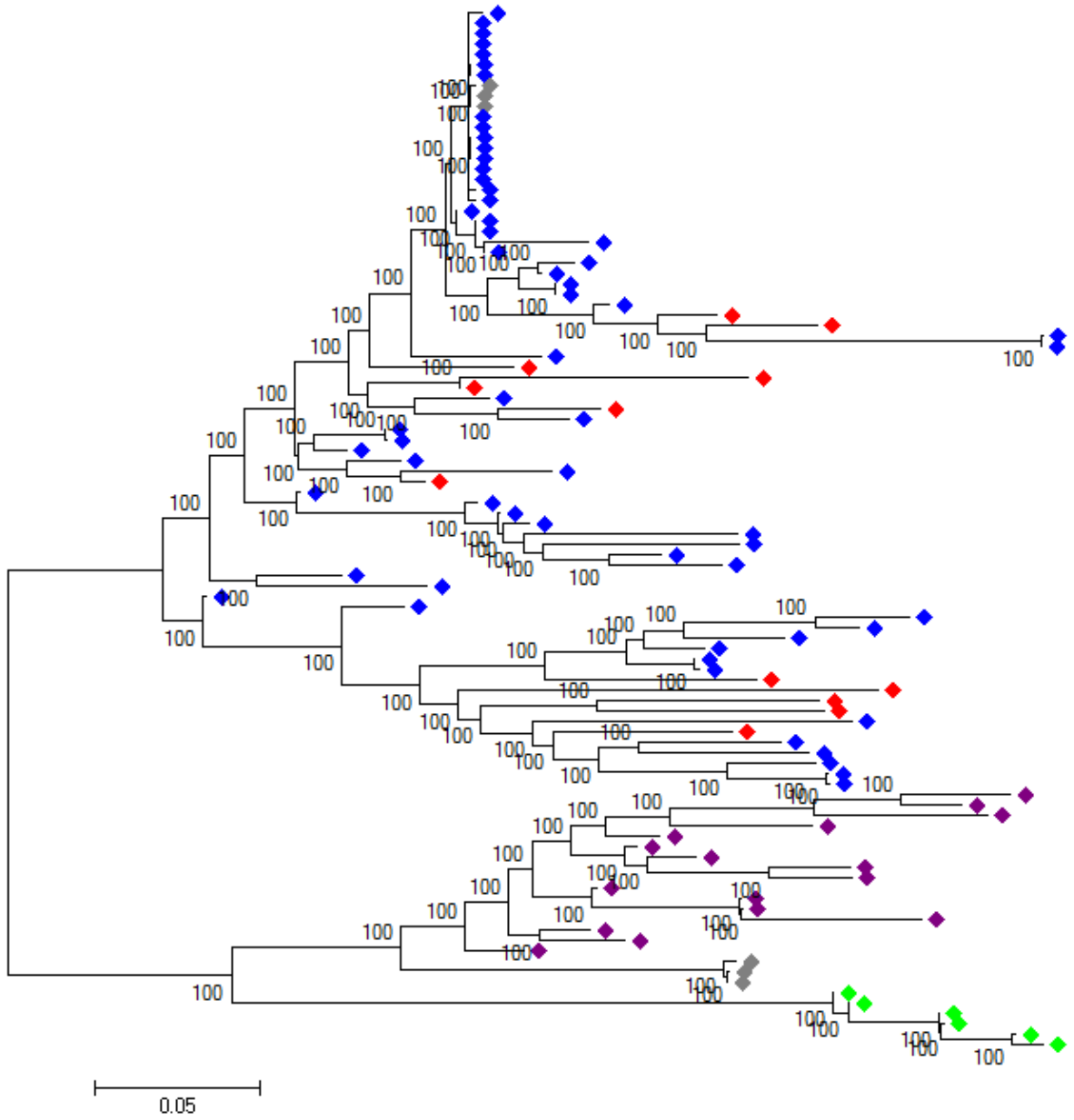


Figure S7.142 Chromosome 4: 23890001 - 23985001

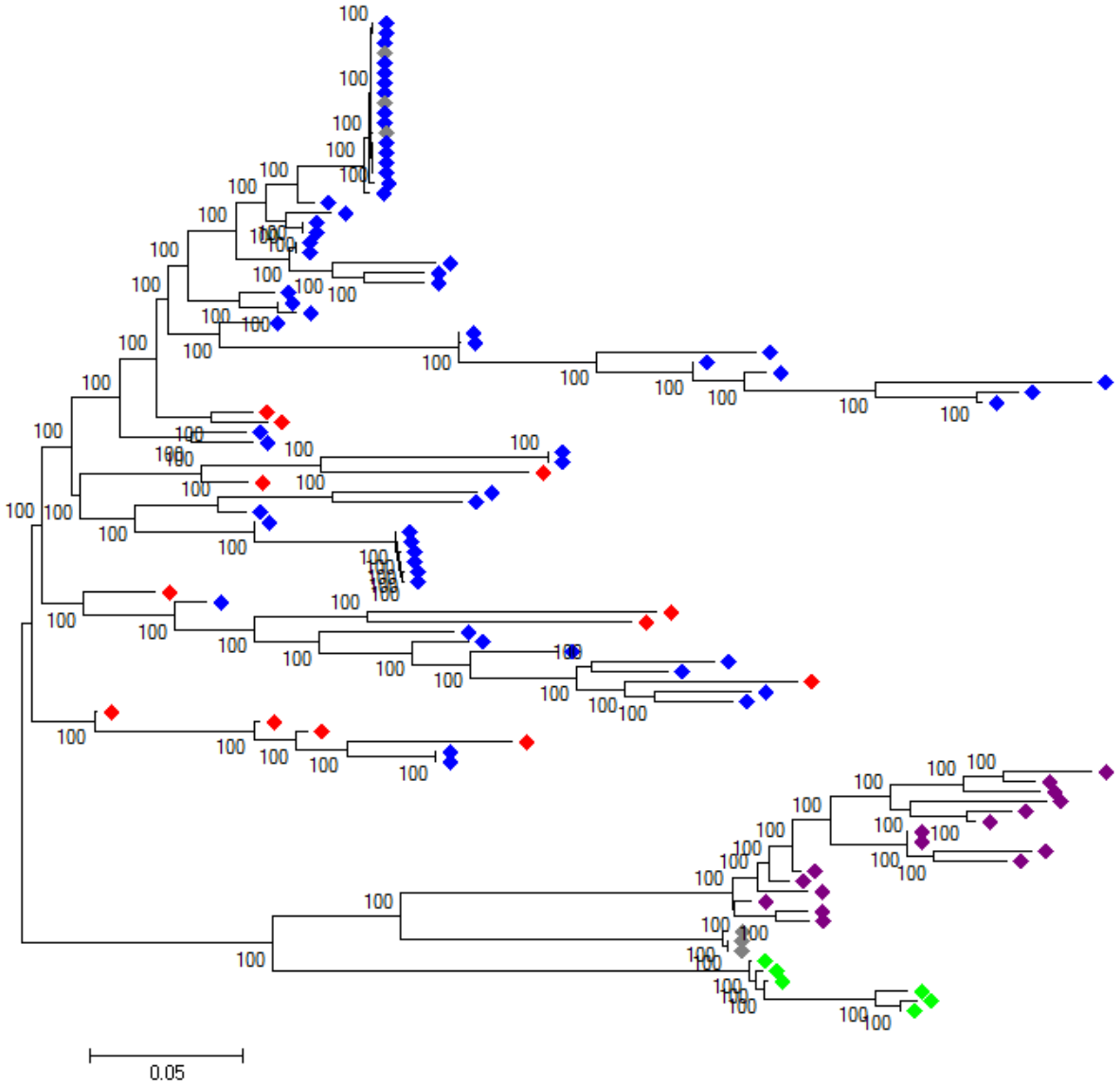


Figure S7.143 Chromosome 4: 24000001 - 24035001

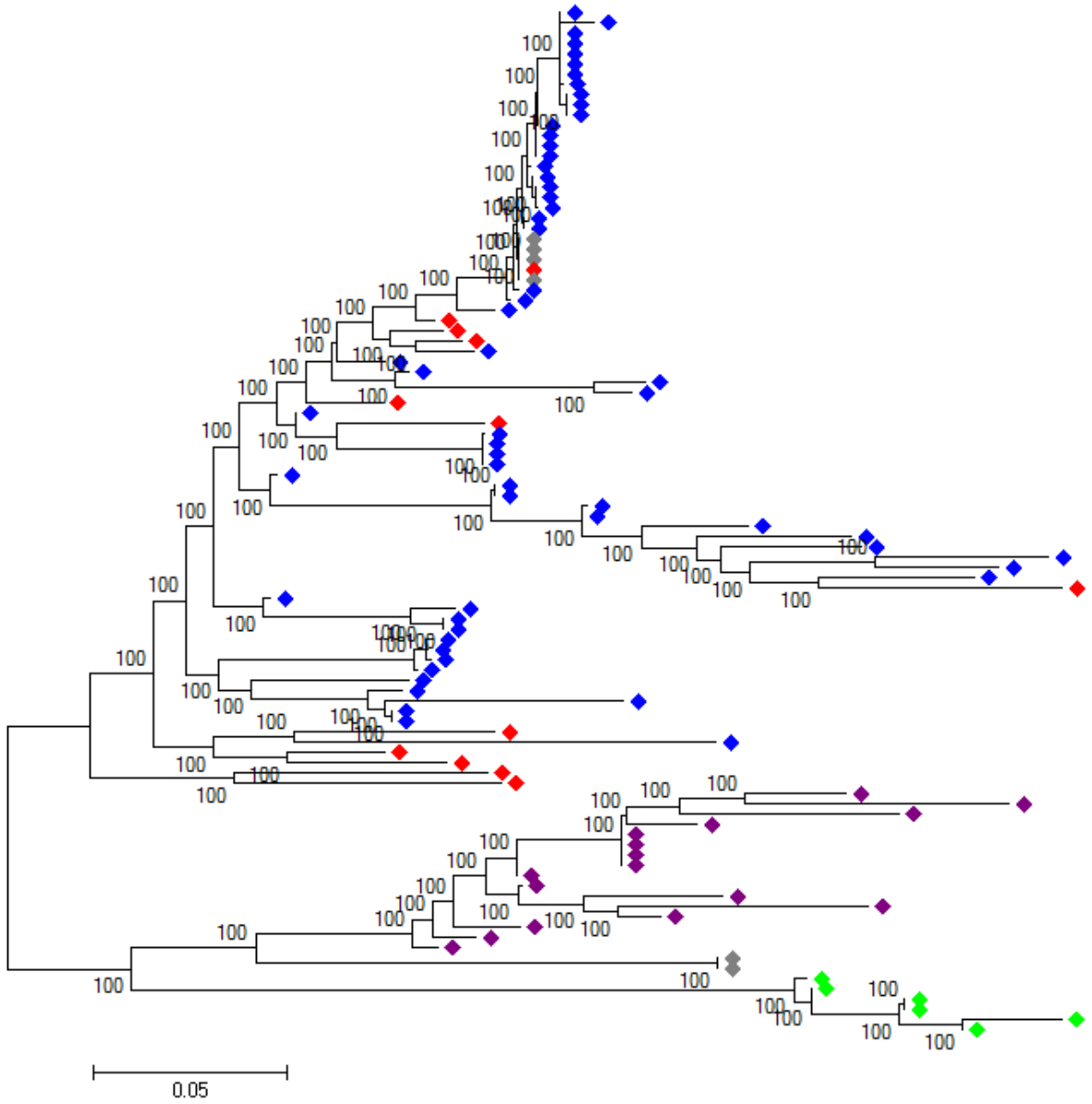


Figure S7.144 Chromosome 4: 25180001 - 25220001

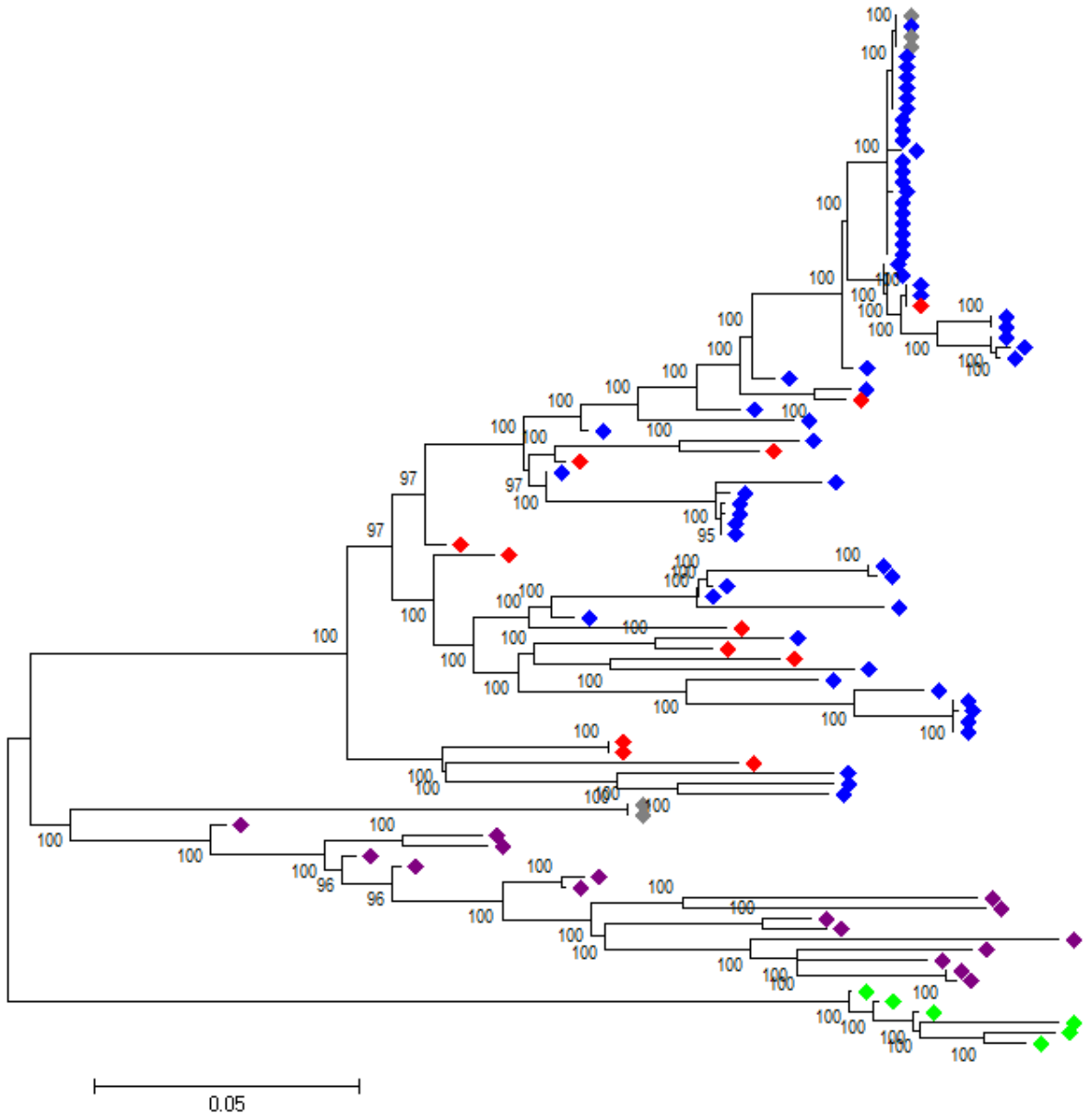


Figure S7.145 Chromosome 4: 25520001 - 25555001

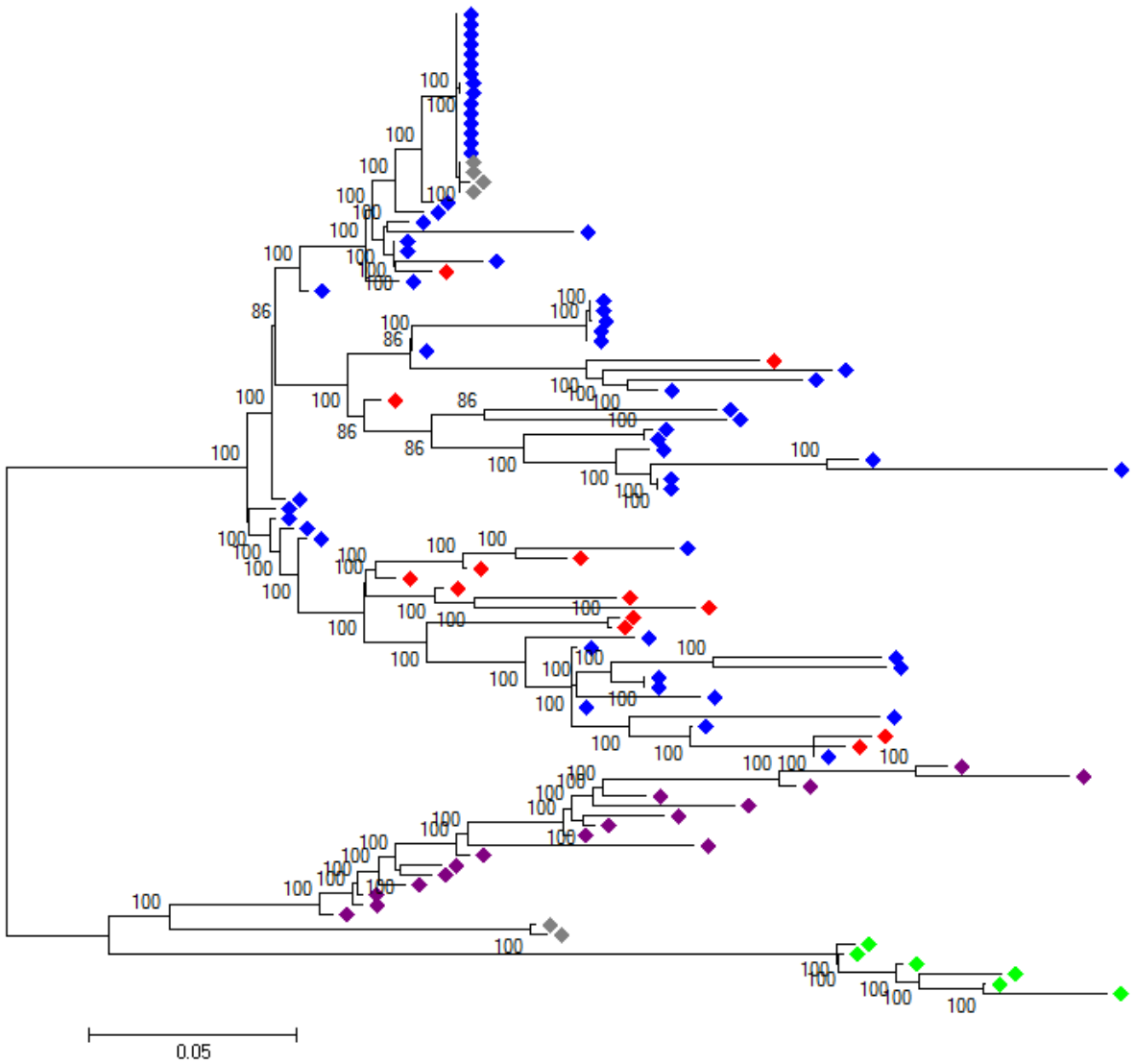


Figure S7.146 Chromosome 34: 25580001 - 25625001

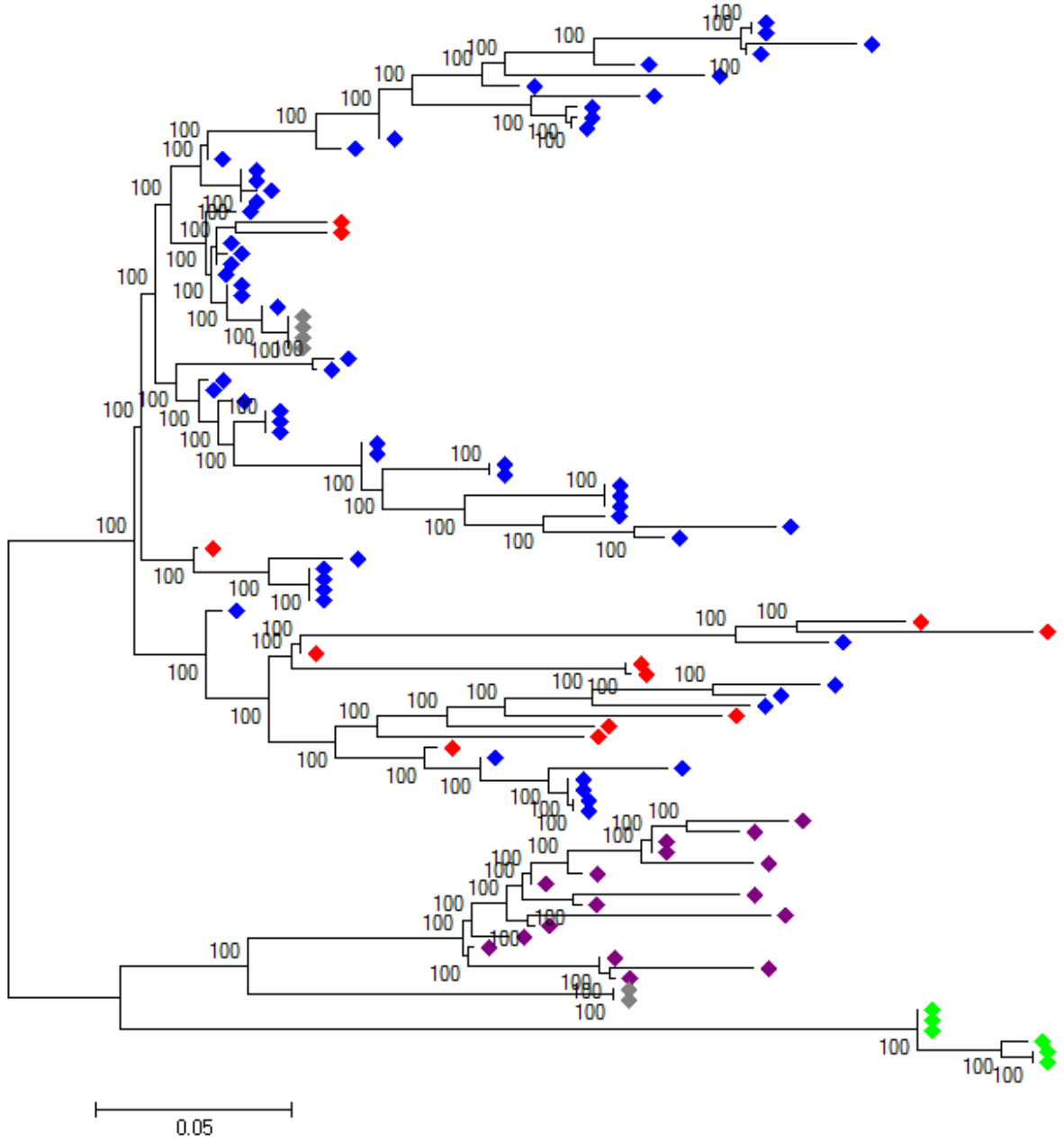


Figure S7.147 Chromosome 4: 26310001 - 26340001

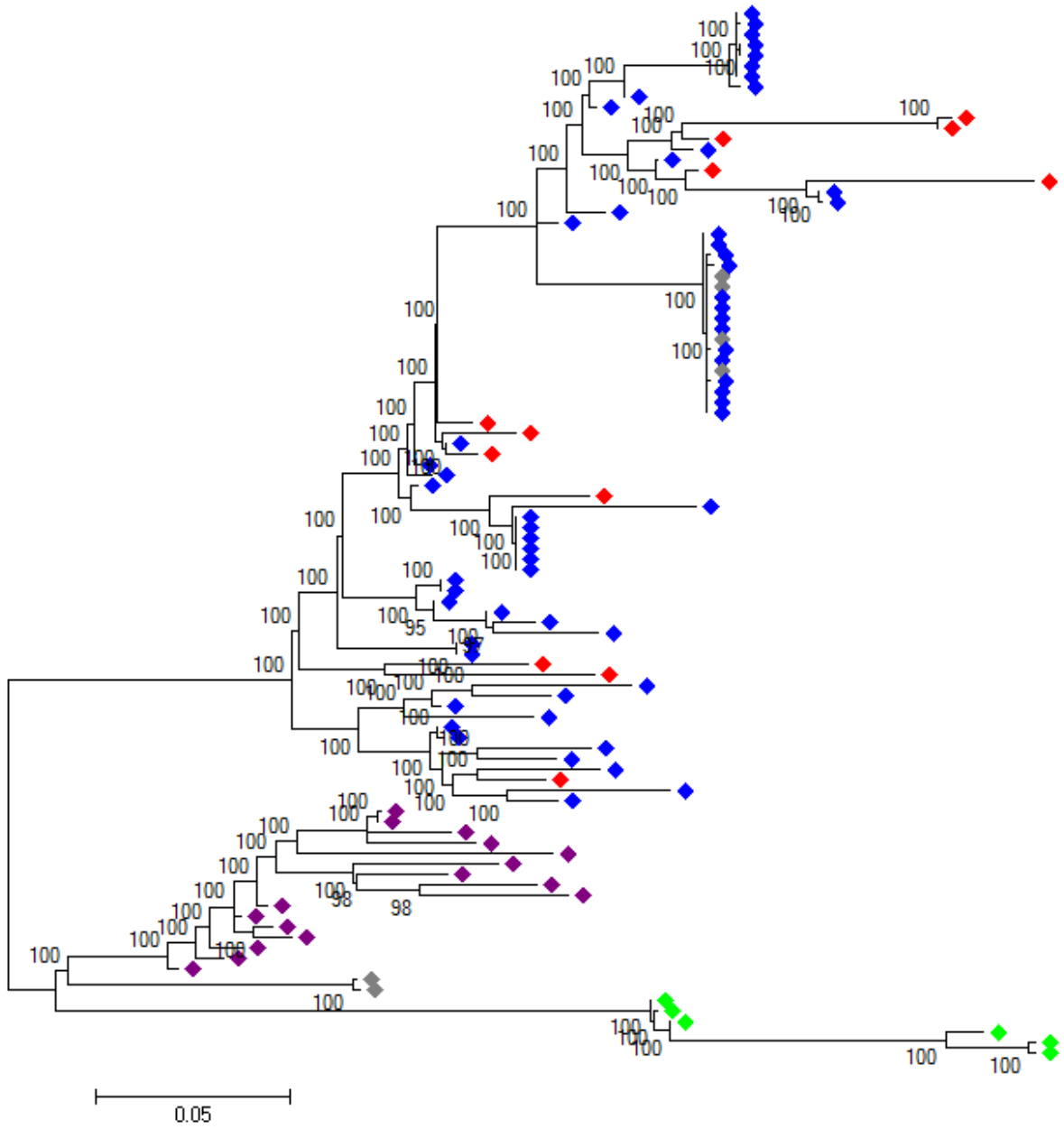


Figure S7.148 Chromosome 4: 26595001 - 26635001

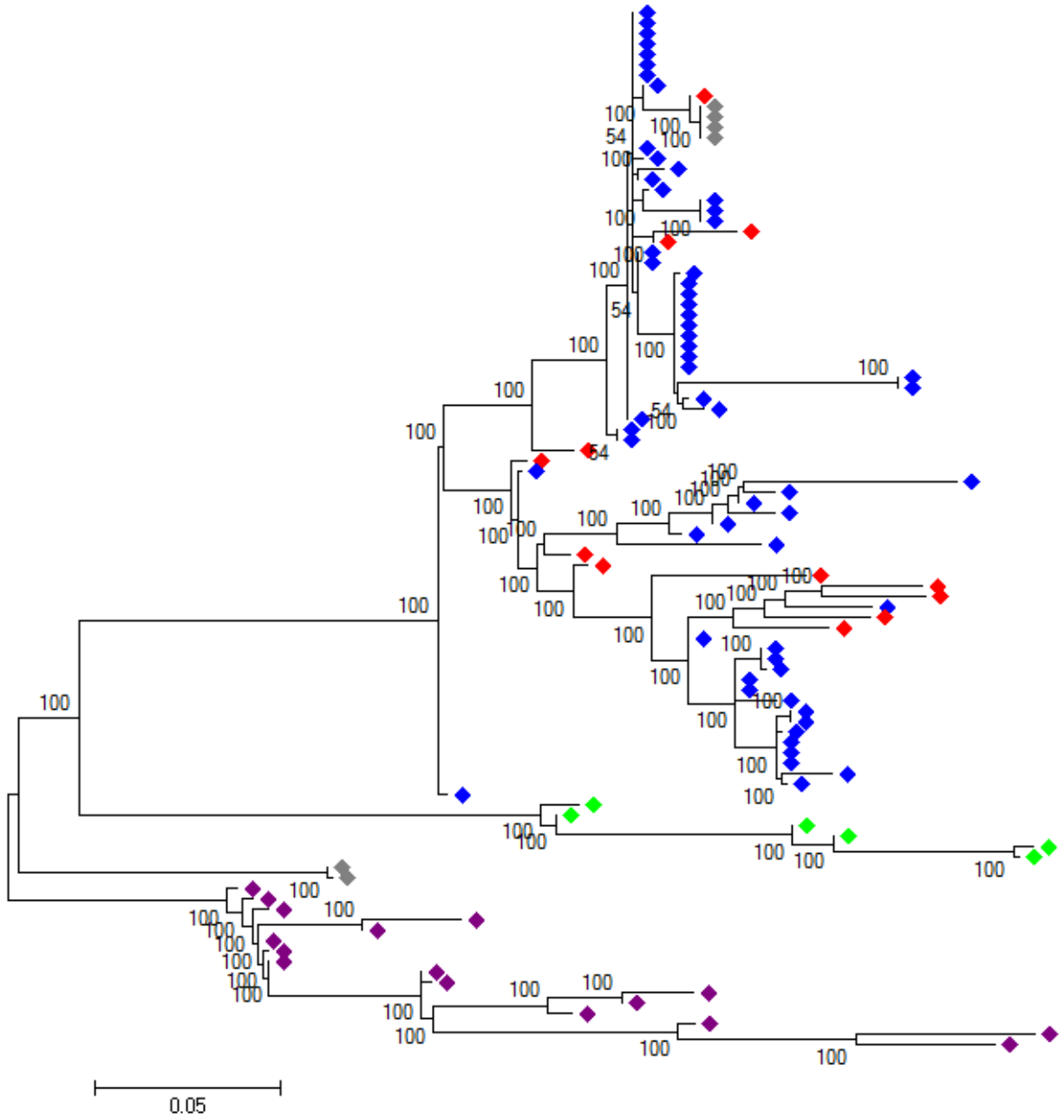


Figure S7.149 Chromosome 4: 26795001 - 26820001

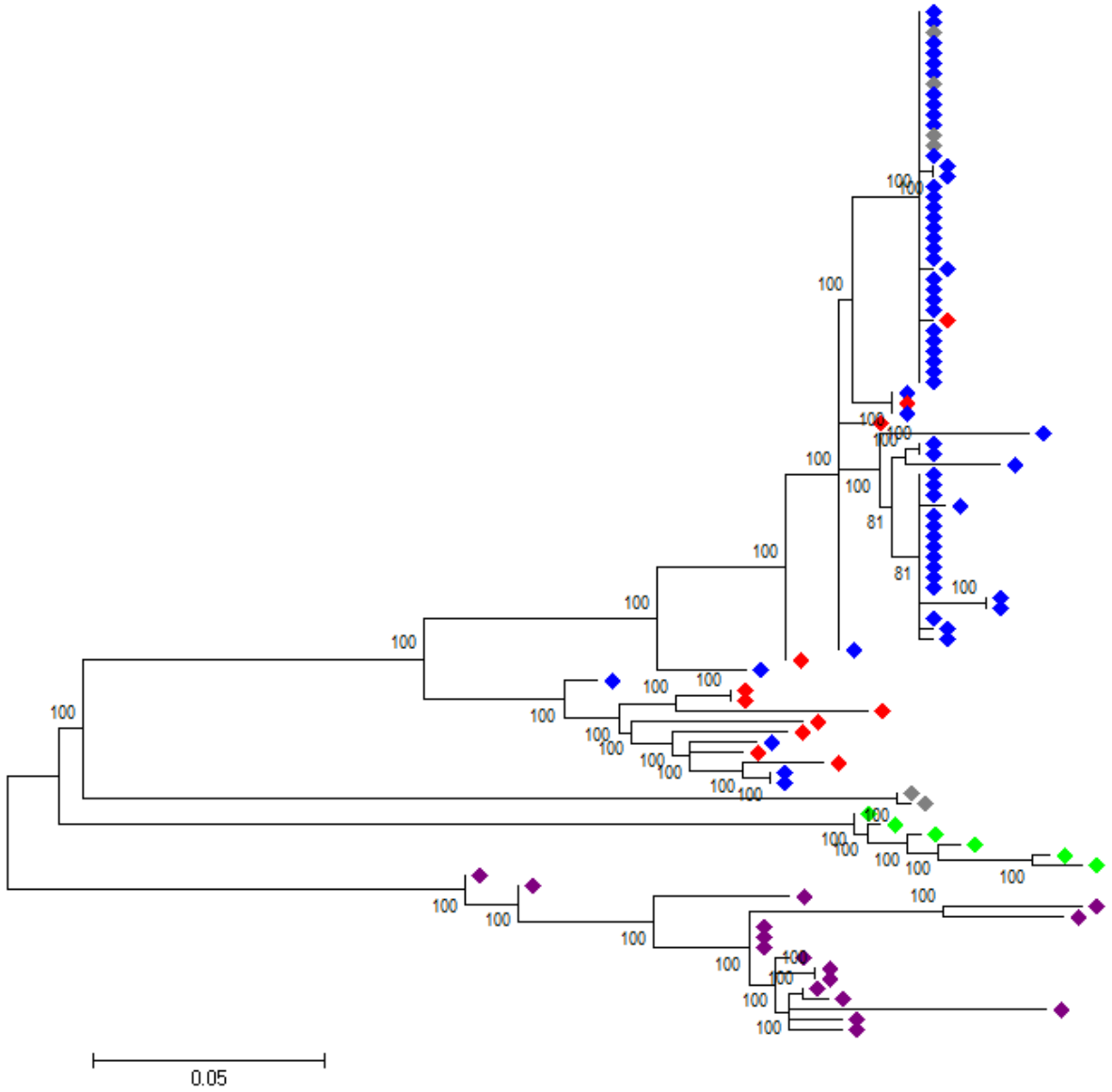


Figure S7.150 Chromosome 4: 26970001 - 27000001

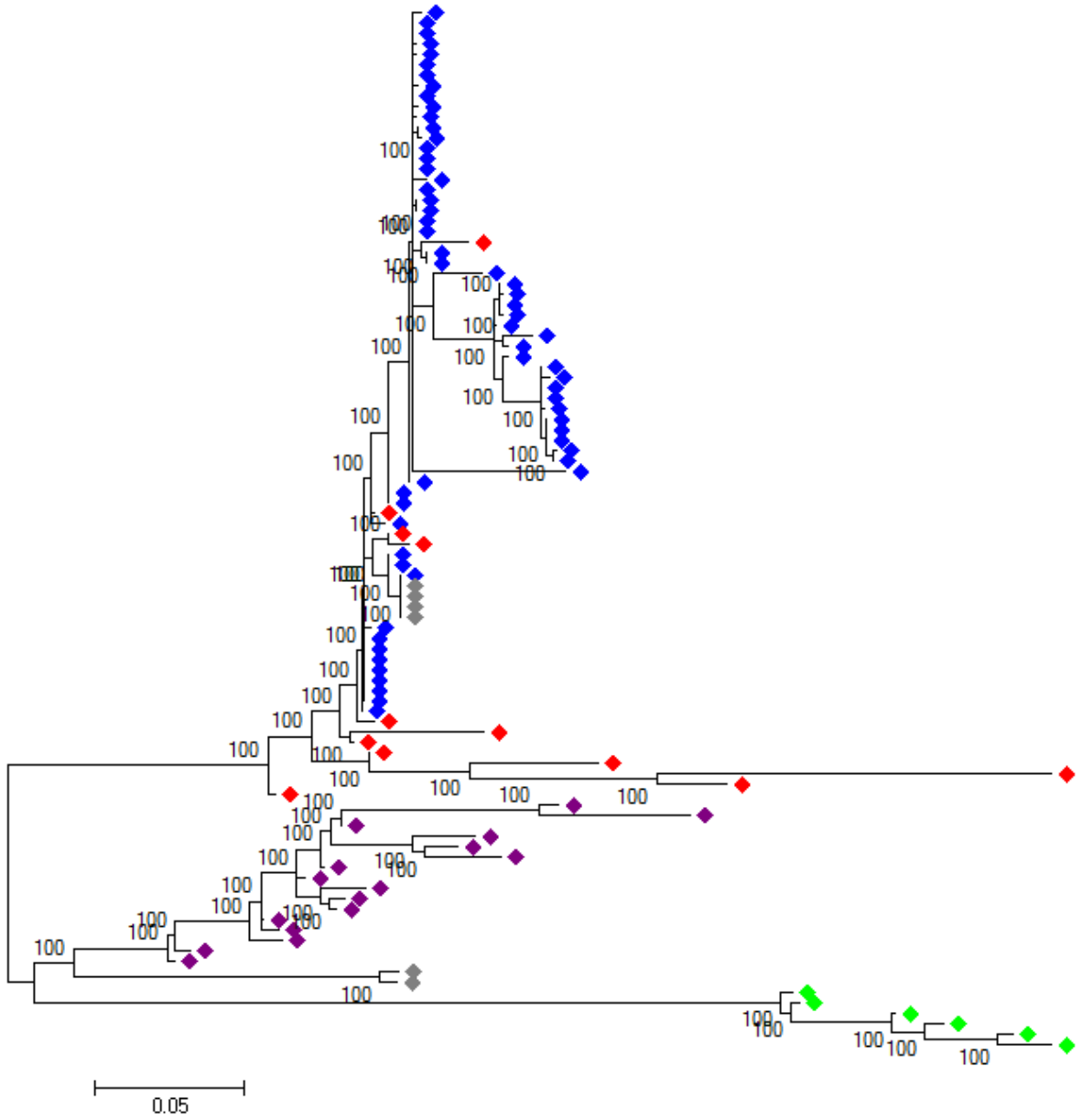


Figure S7.151 Chromosome 4: 27070001 - 27100001

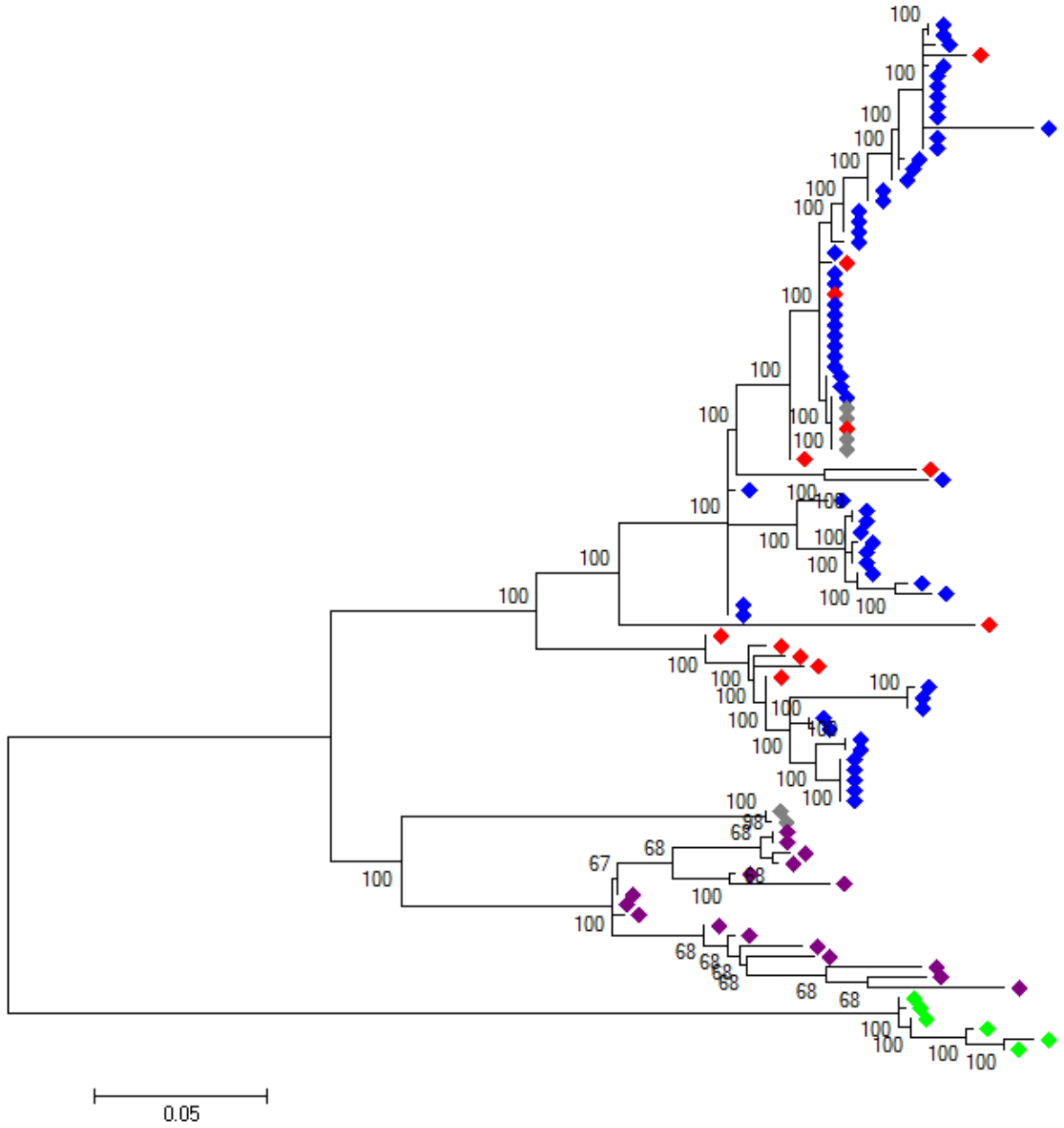


Figure S7.152 Chromosome 4: 27530001 - 27550001

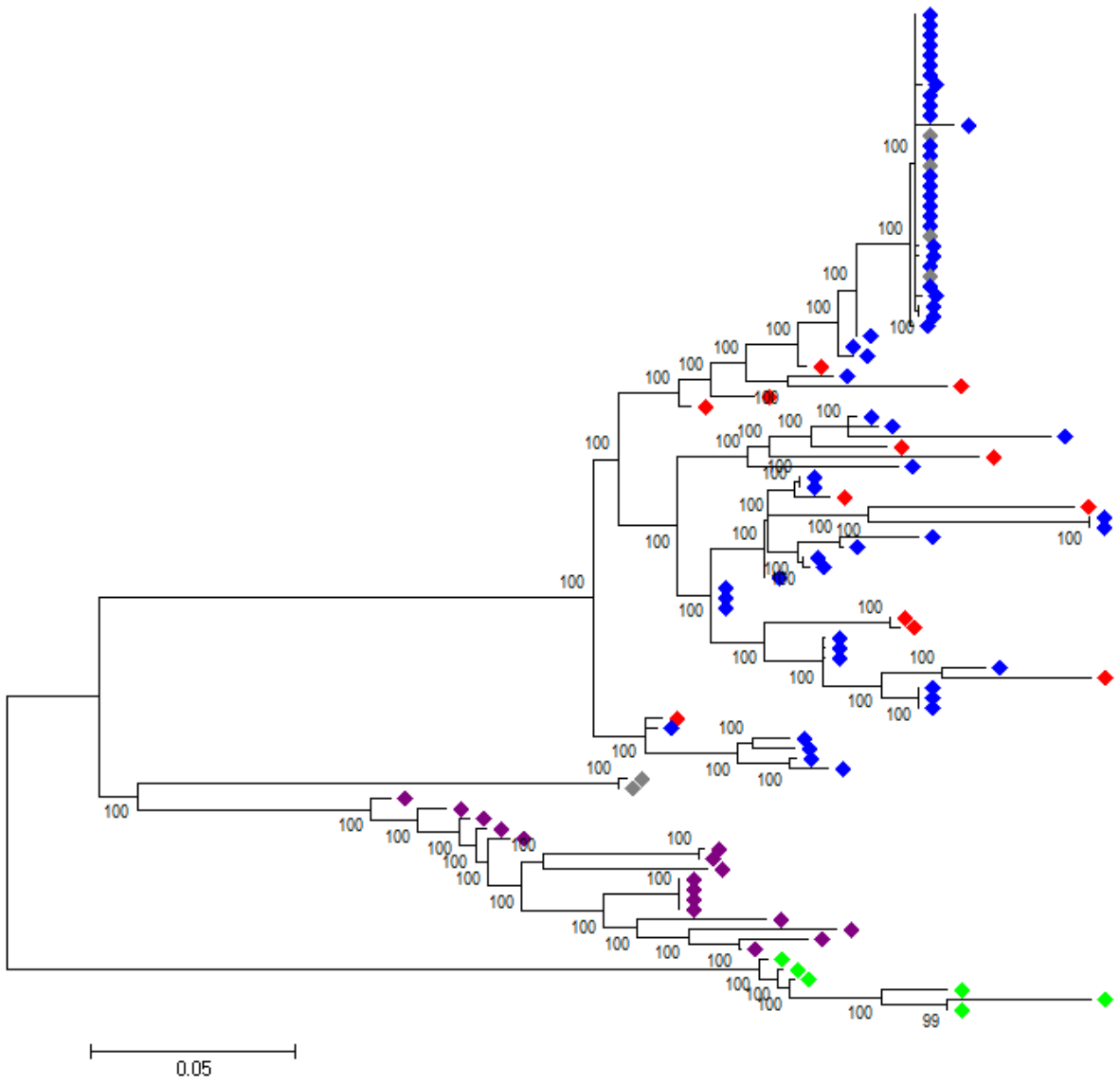


Figure S7.153 Chromosome 4: 27975001 - 28050001

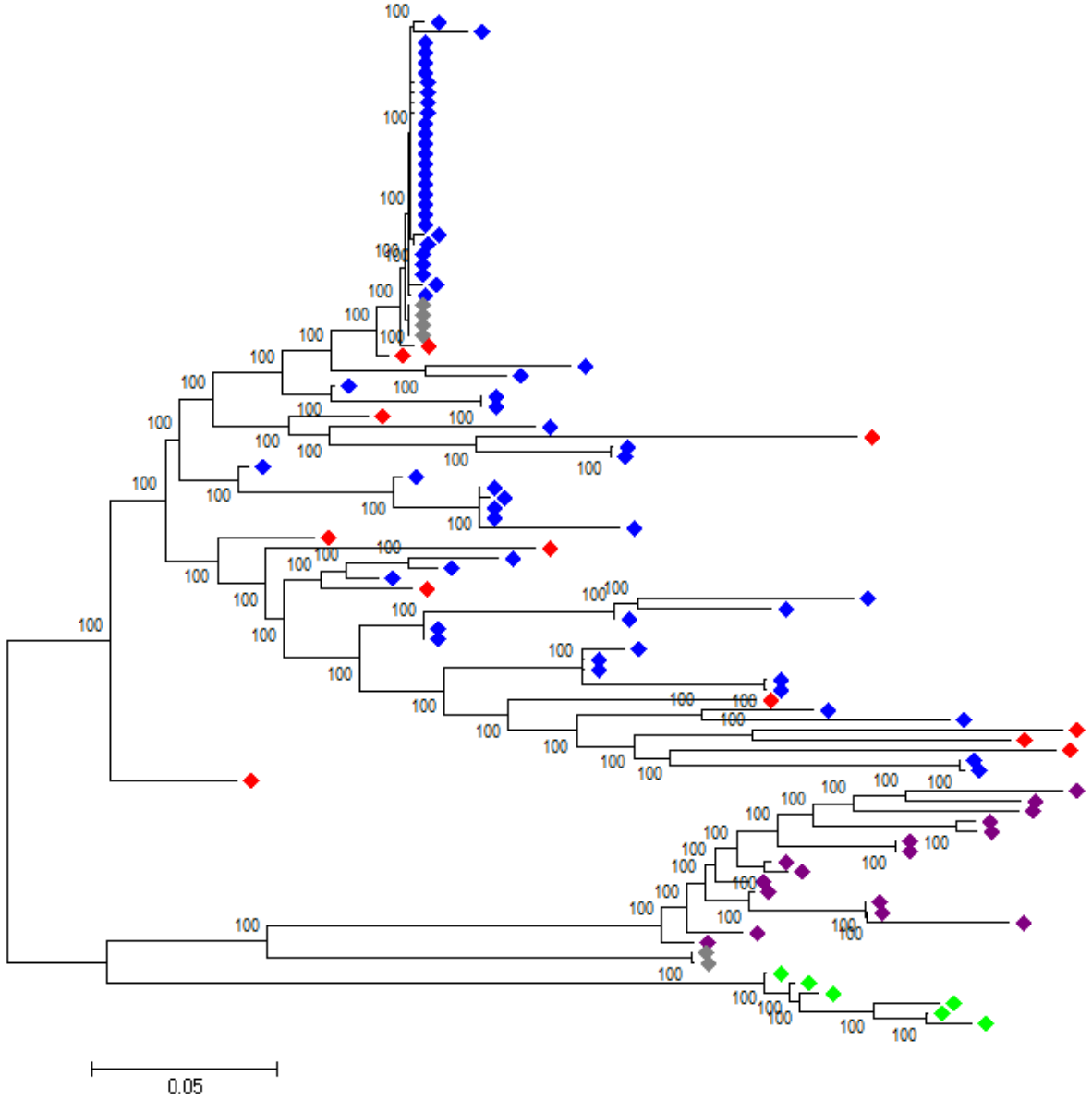


Figure S7.154 Chromosome 4: 28095001 - 28130001

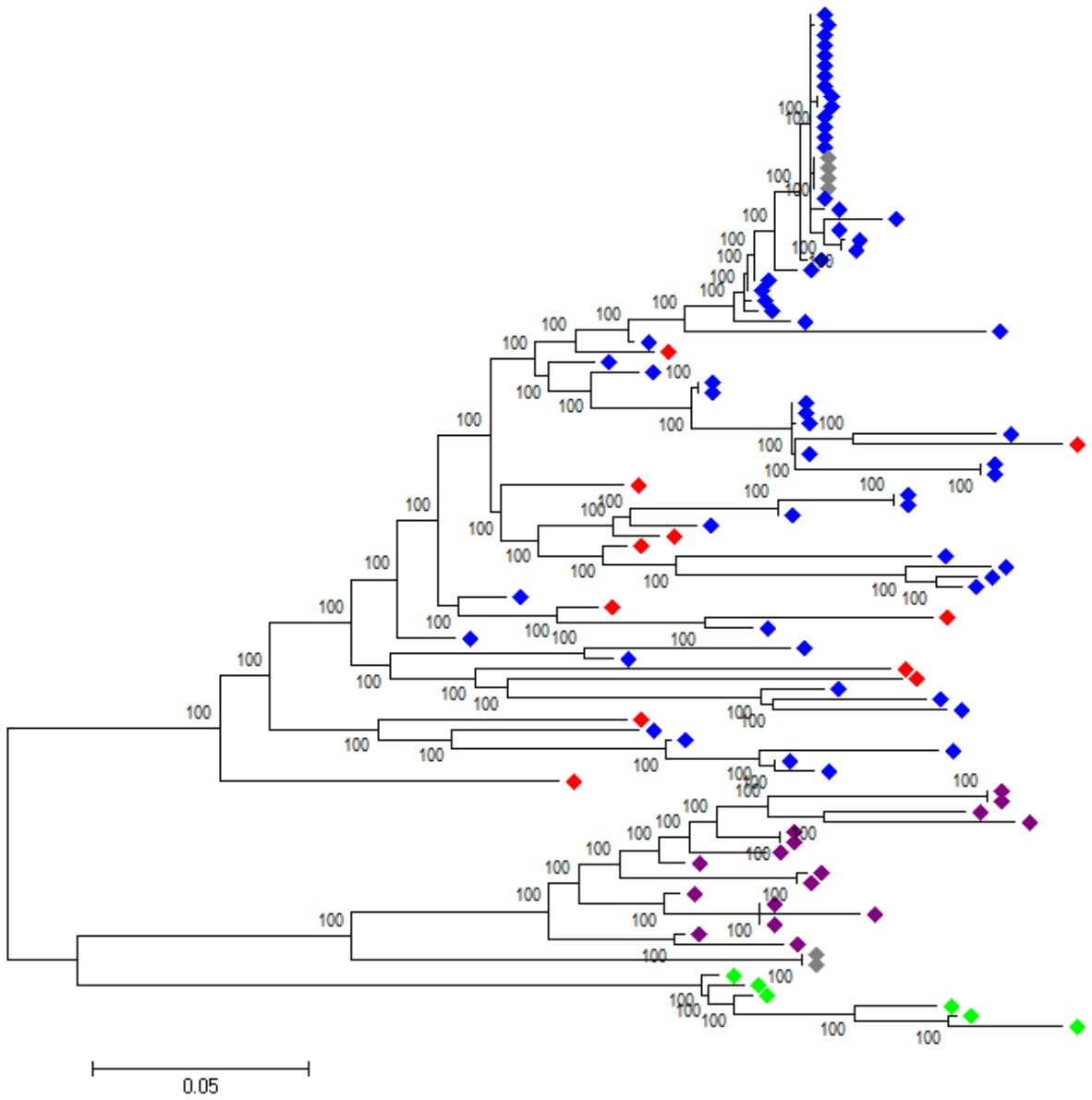


Figure S7.155 Chromosome 4: 28325001 - 28350001

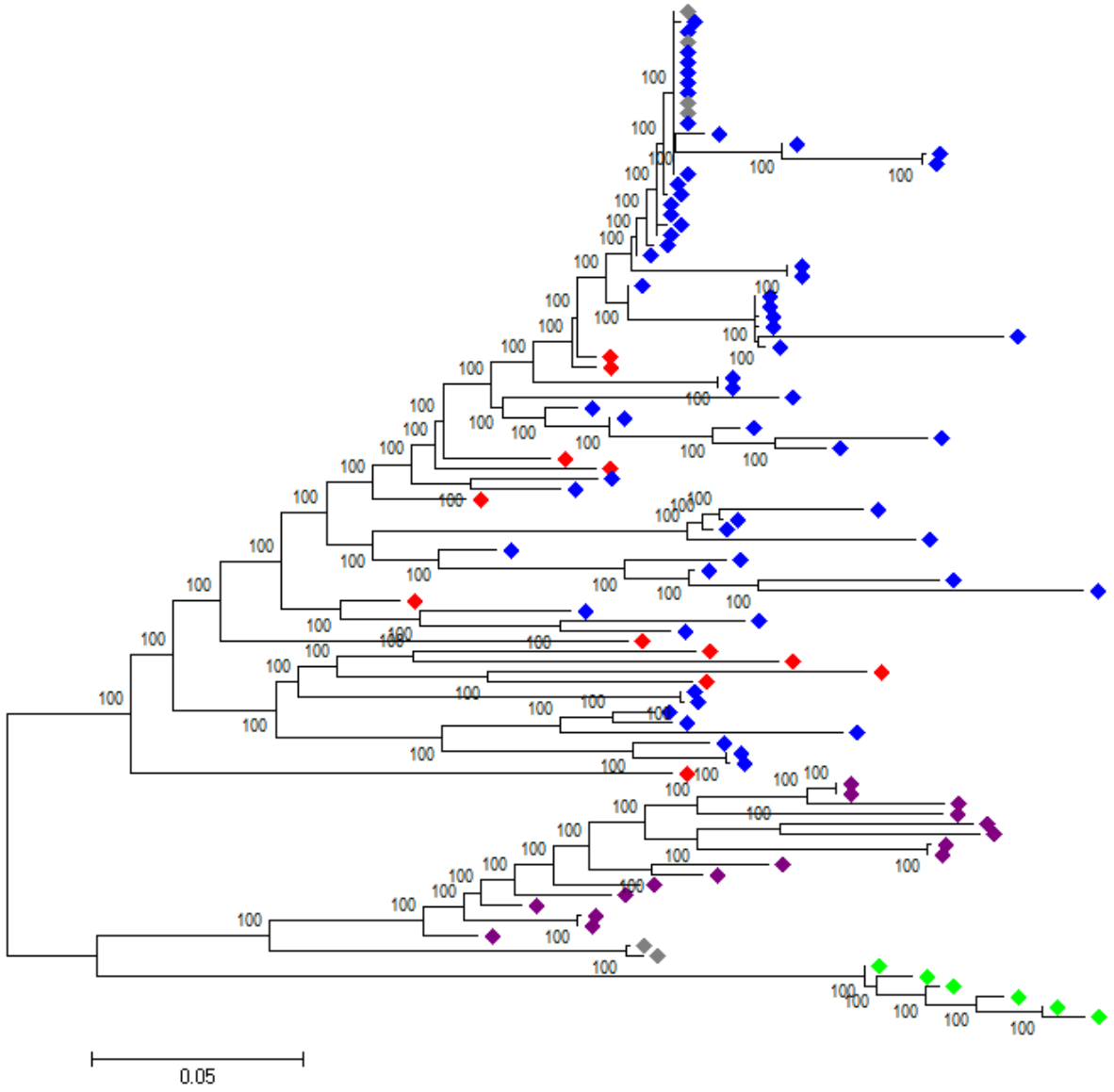


Figure S7.156 Chromosome 4: 28385001 - 28430001

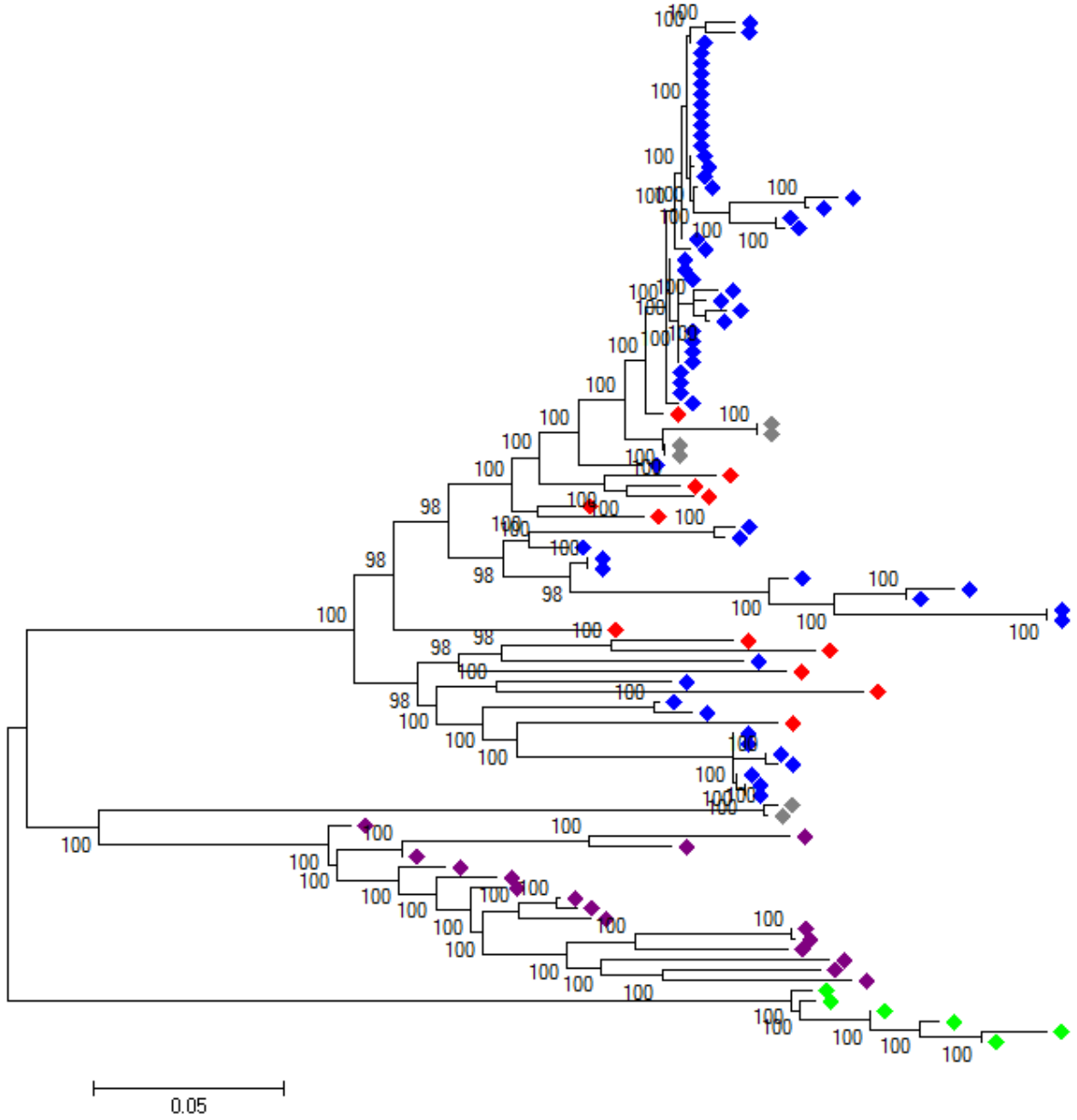


Figure S7.157 Chromosome 4: 29750001 - 29770001

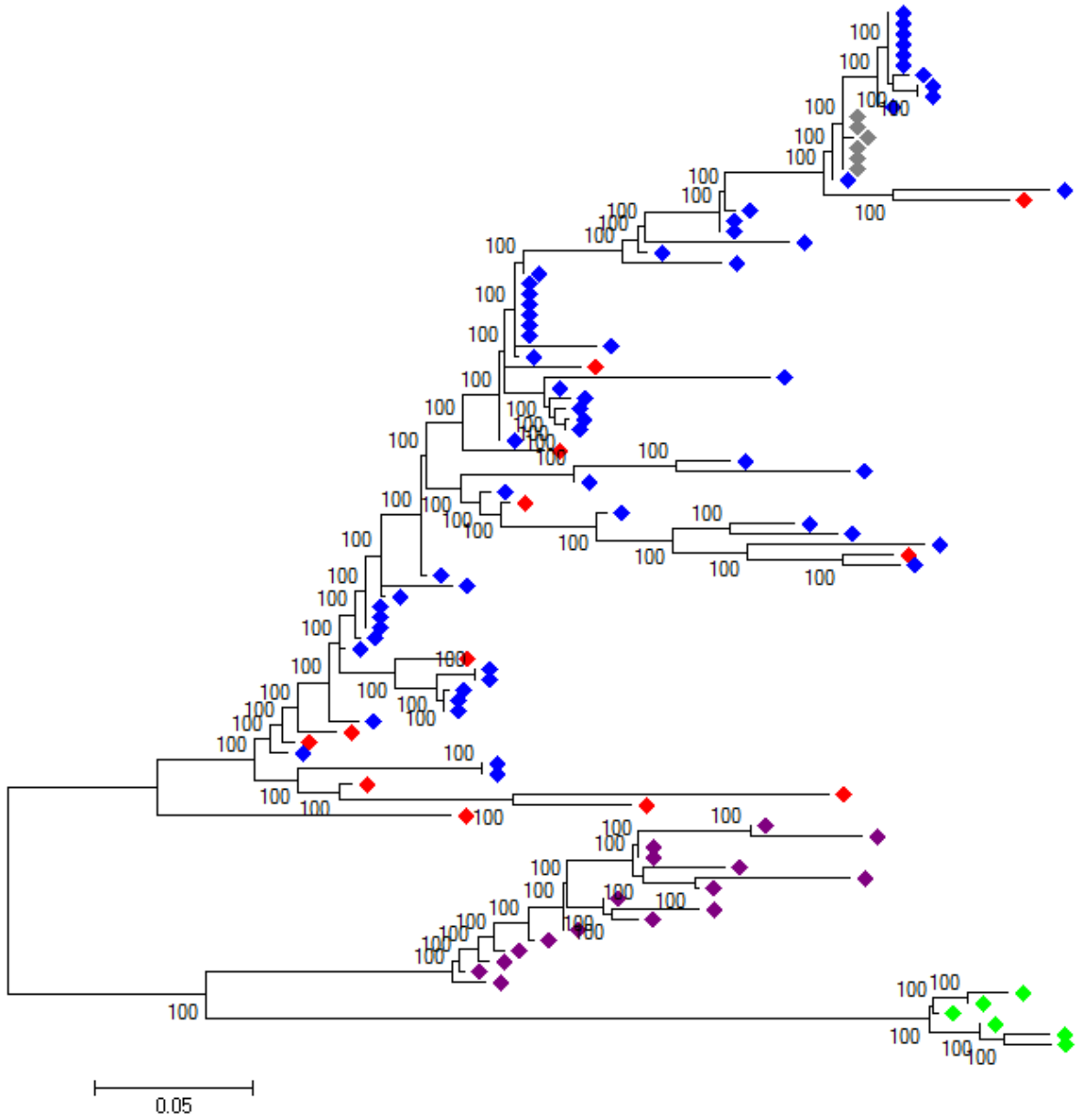


Figure S7.158 Chromosome 4: 31910001 - 31930001

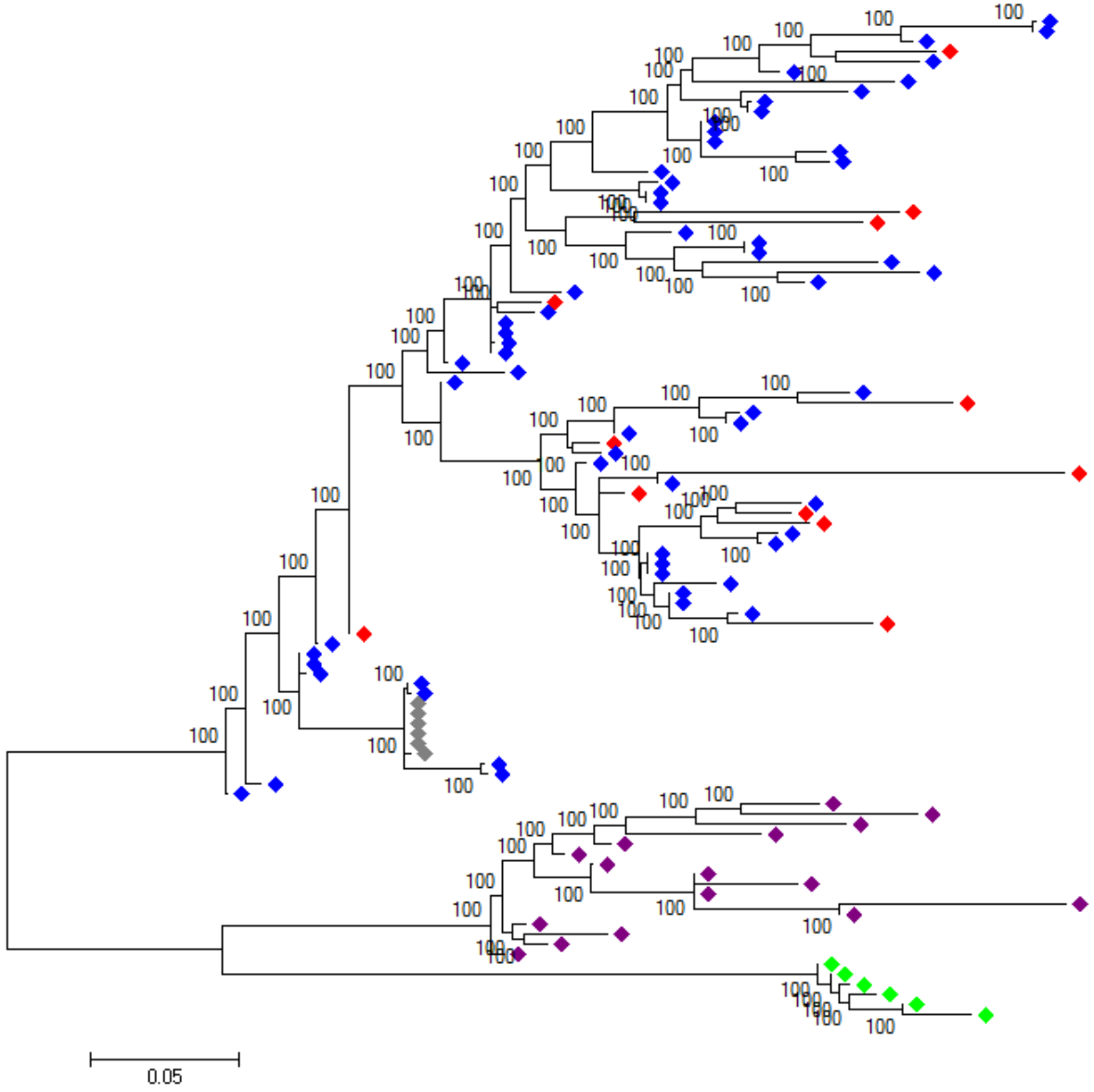


Figure S7.159 Chromosome 4: 32515001 - 32540001

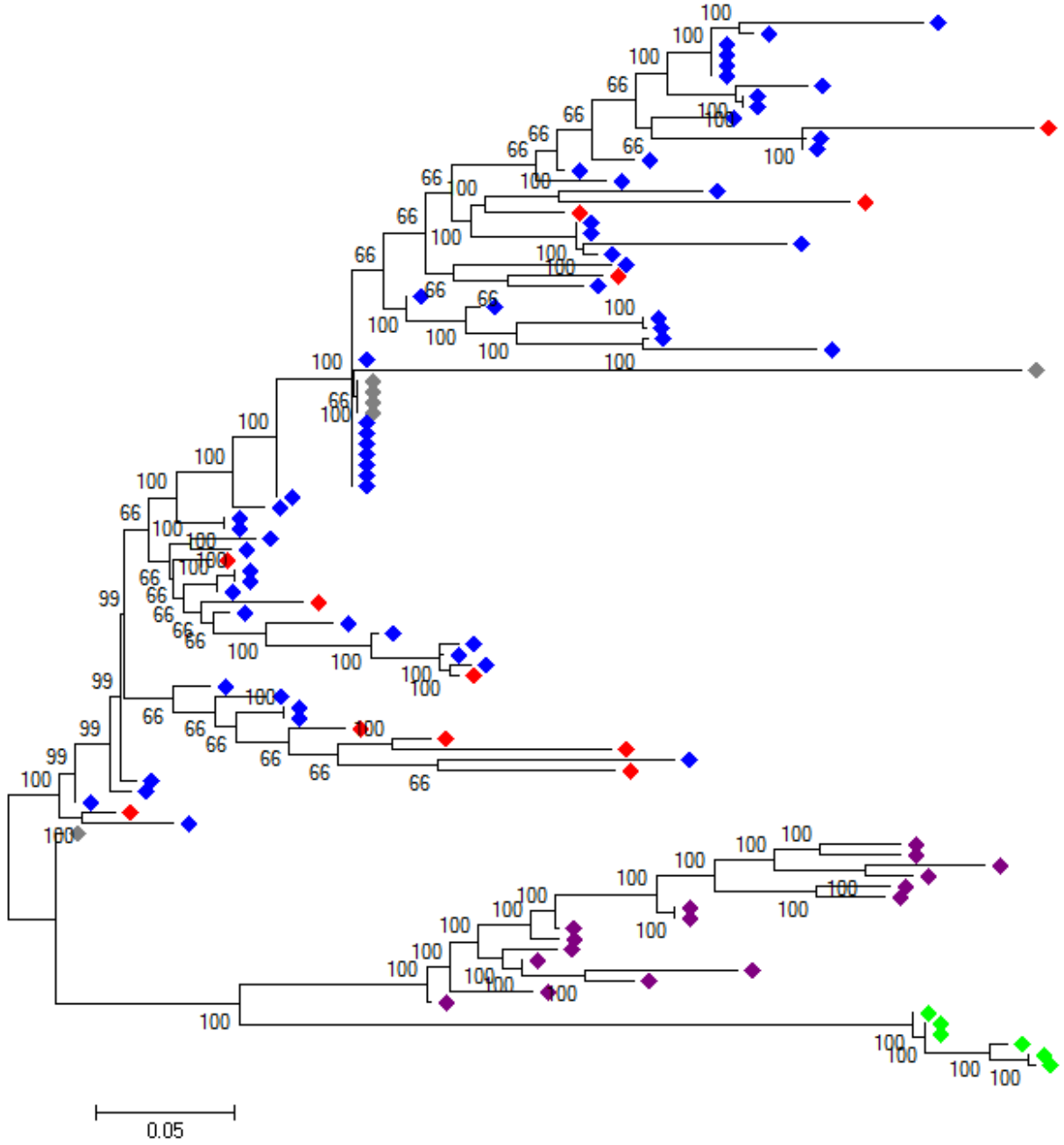


Figure S7.160 Chromosome 4: 32800001 - 32820001

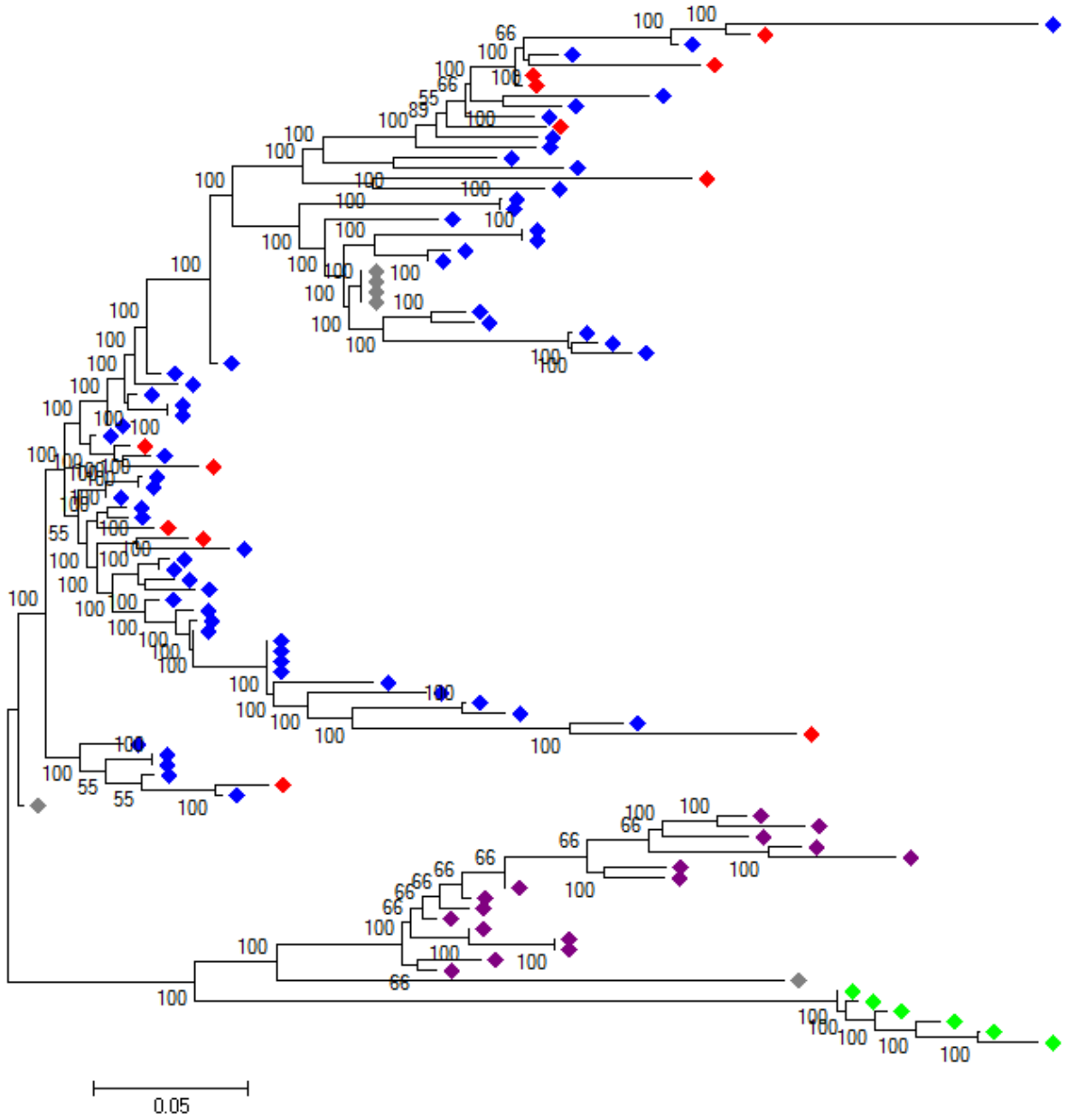


Figure S7.161 Chromosome 4: 32850001 - 32875001

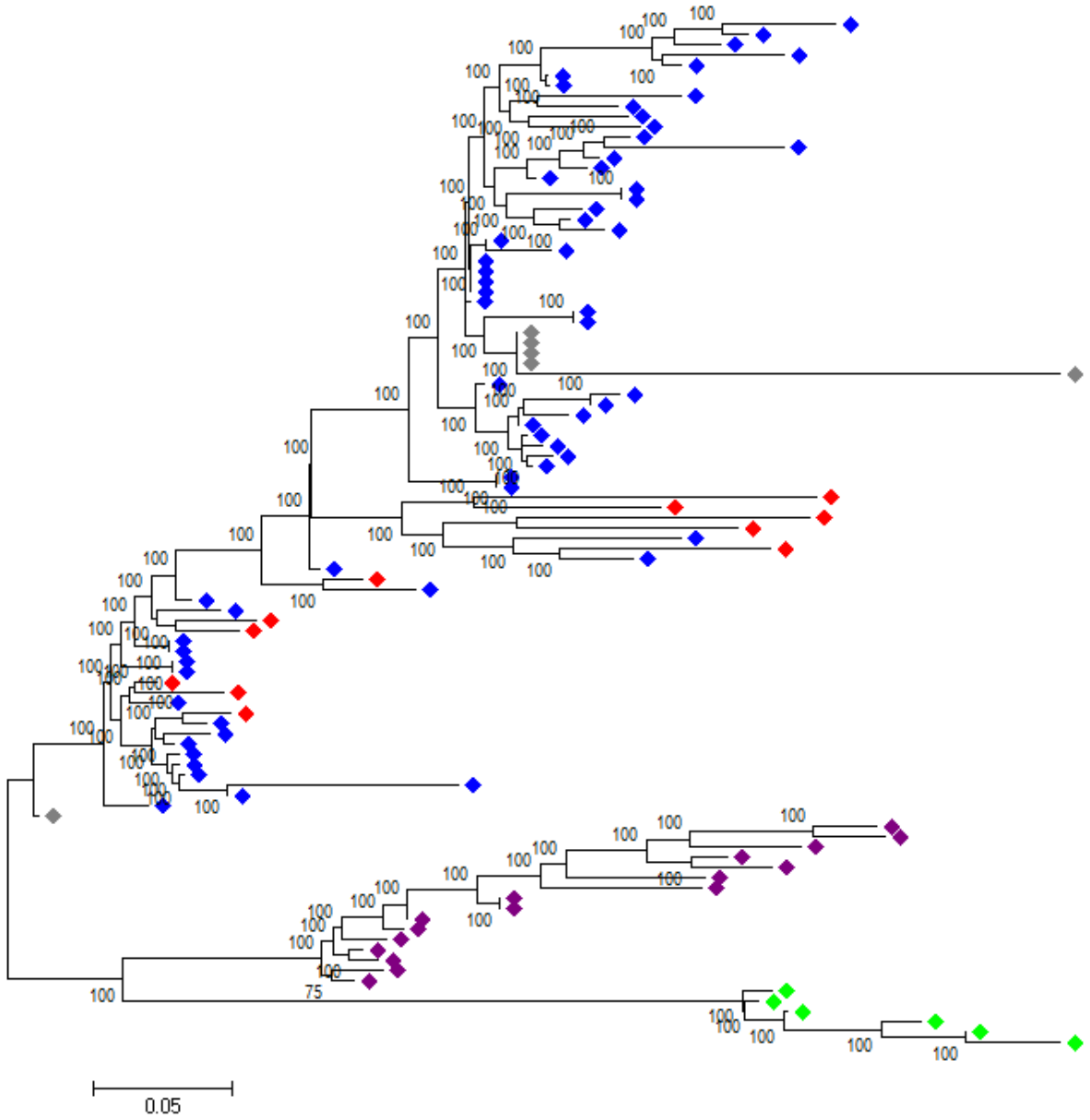


Figure S7.162 Chromosome 4: 33355001 - 33380001

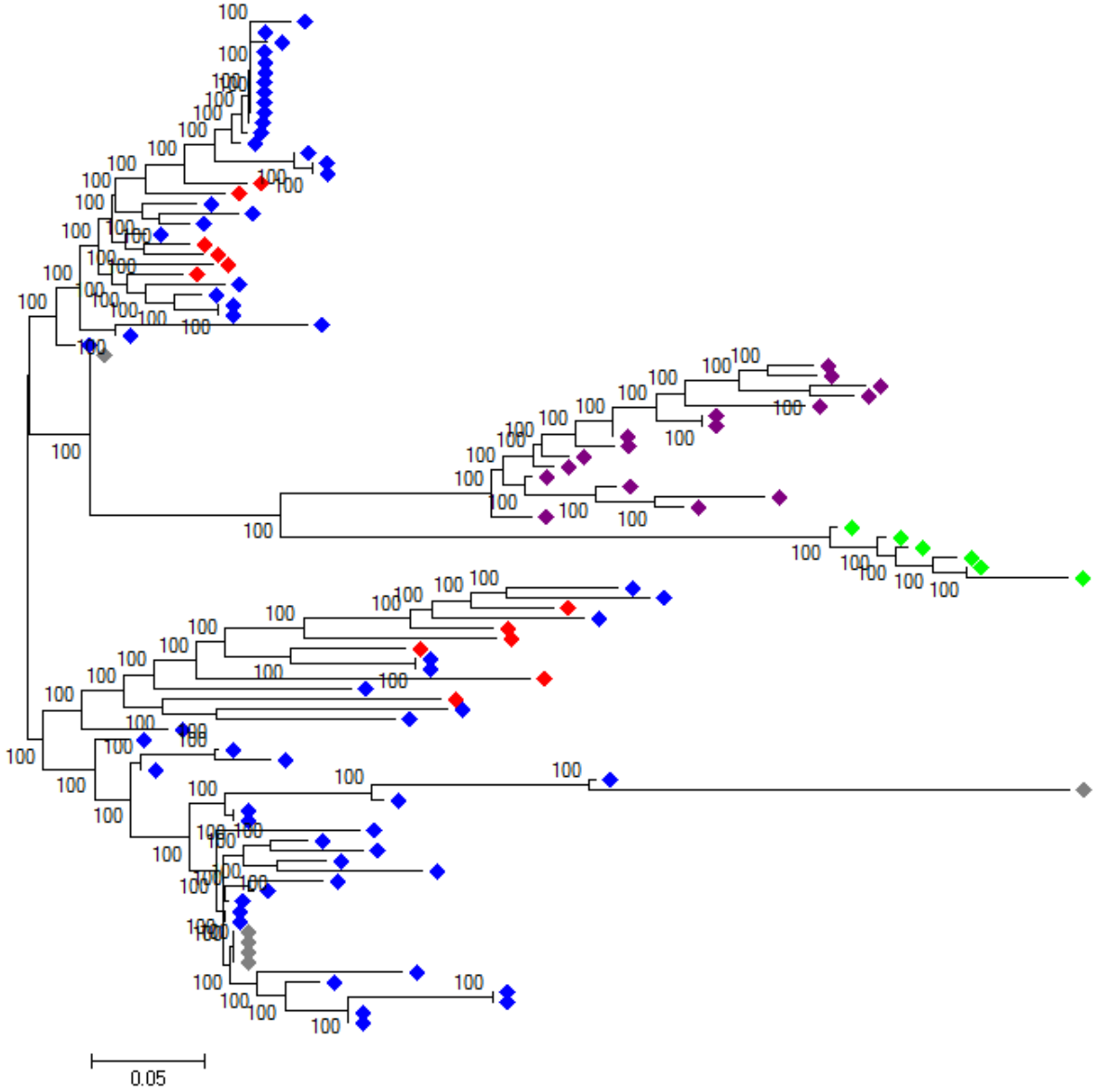


Figure S7.163 Chromosome 4: 33490001 - 33525001

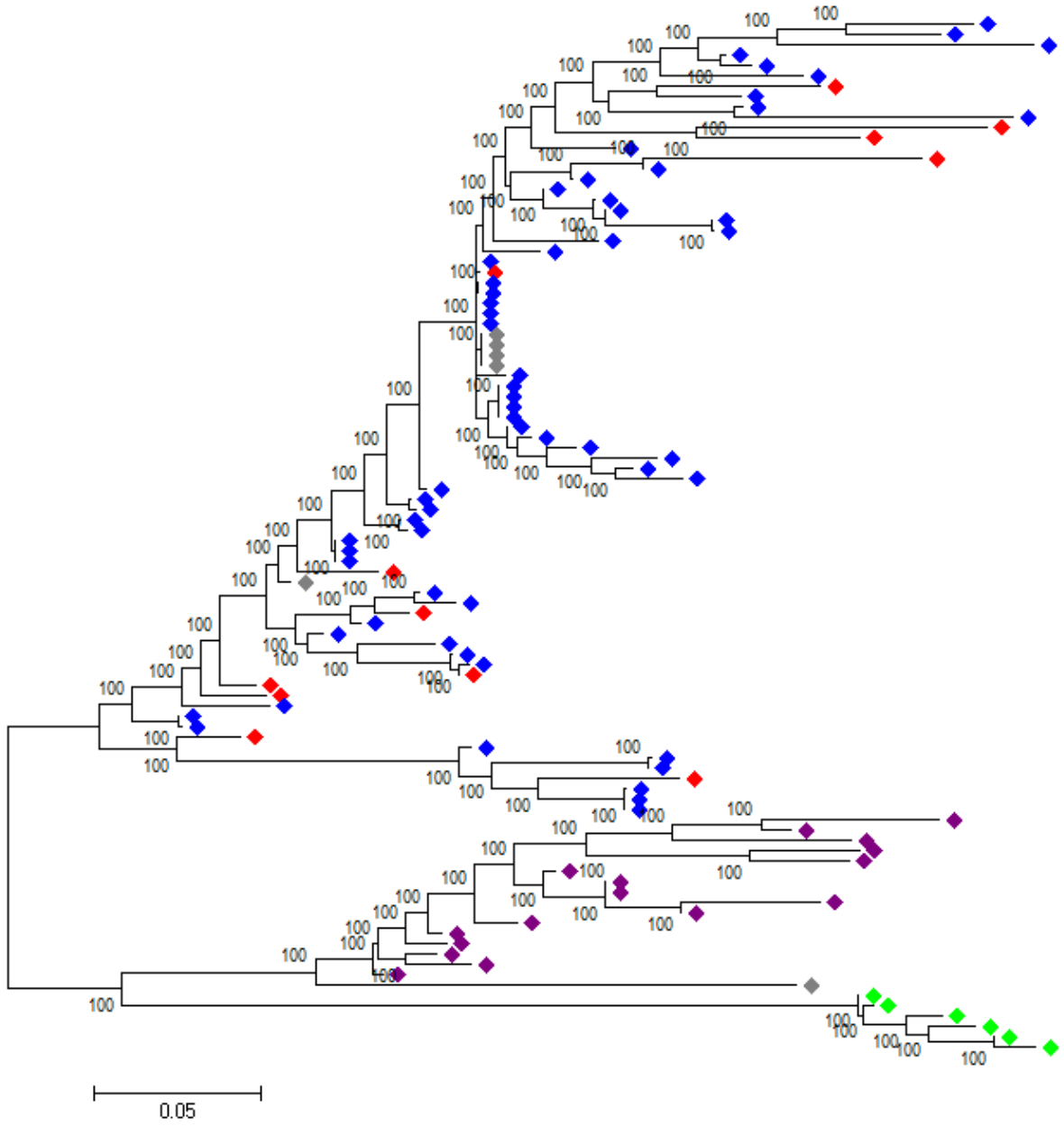


Figure S7.164 Chromosome 4: 33895001 - 33975001

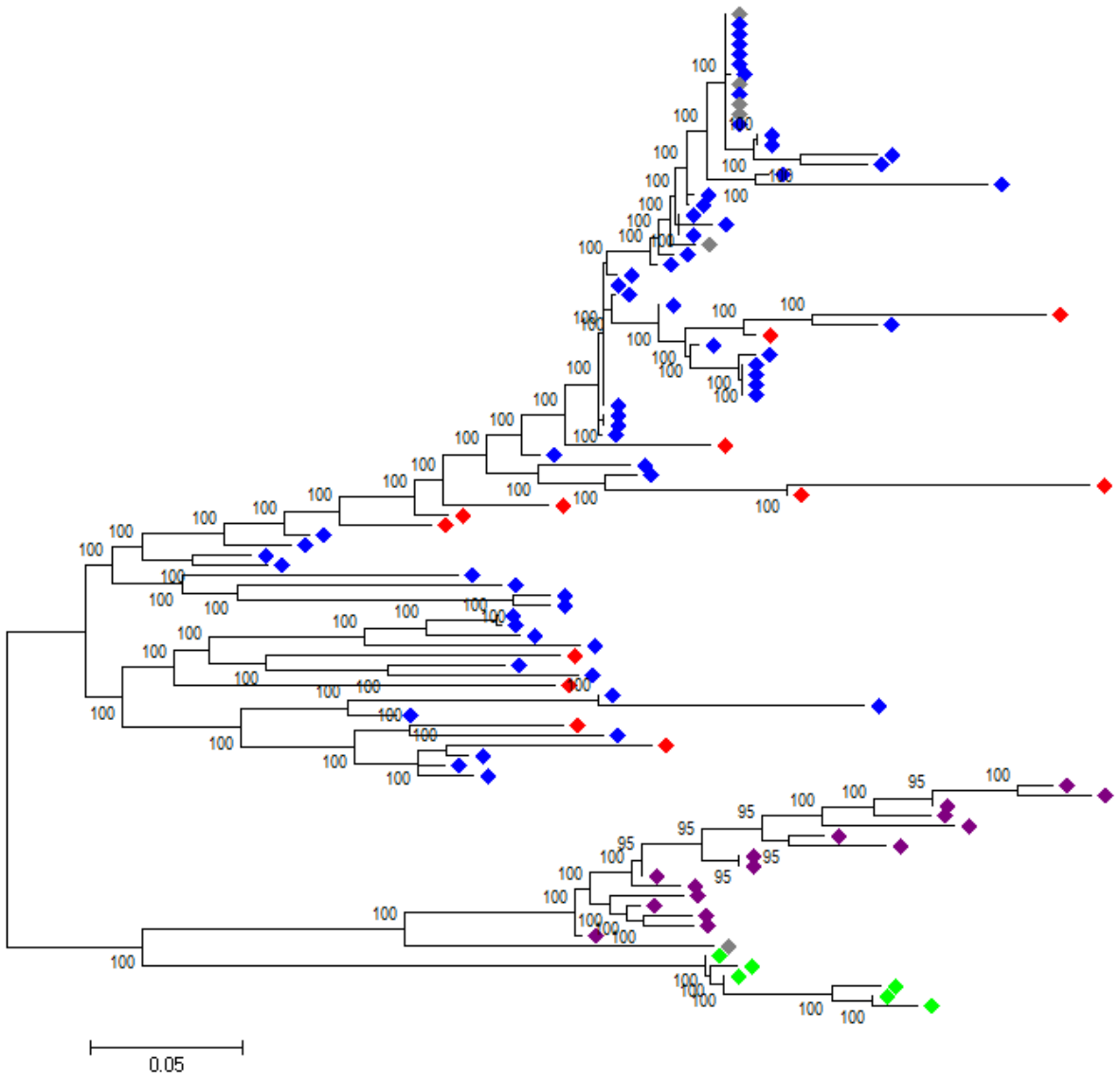


Figure S7.165 Chromosome 4: 33990001 - 34035001

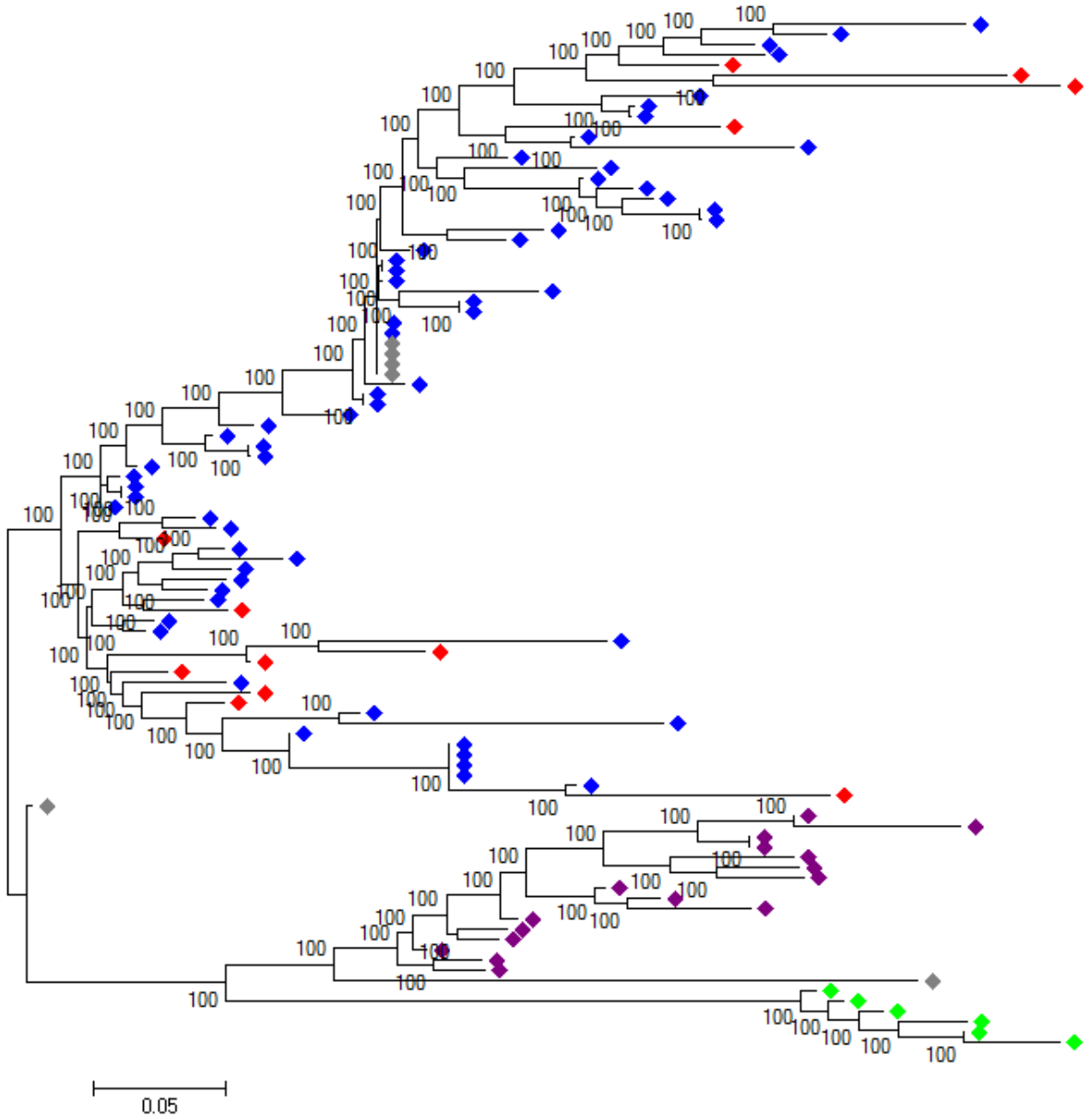


Figure S7.166 Chromosome 4: 34040001 - 34105001

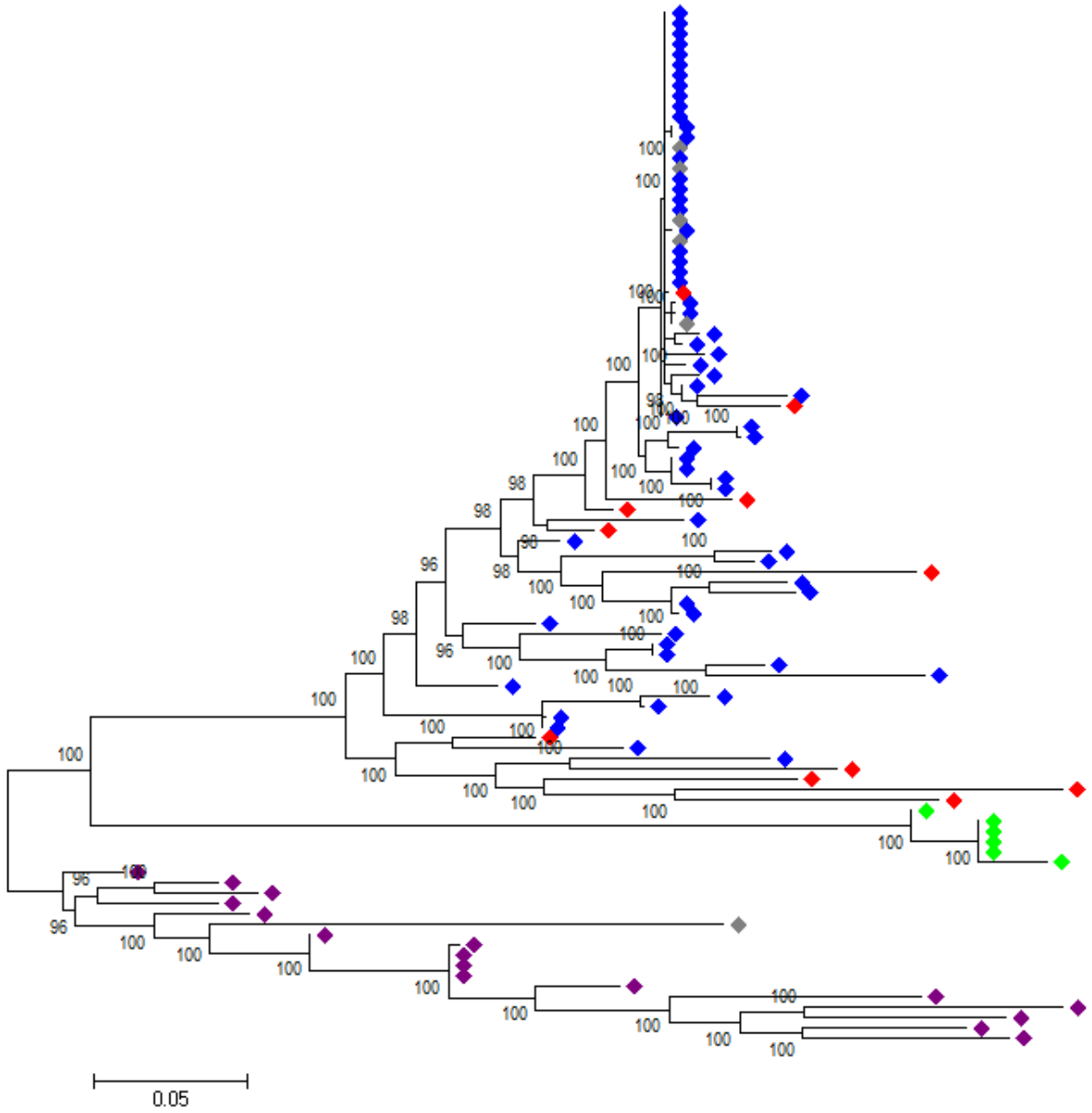


Figure S7.167 Chromosome 4: 39240001 - 39275001

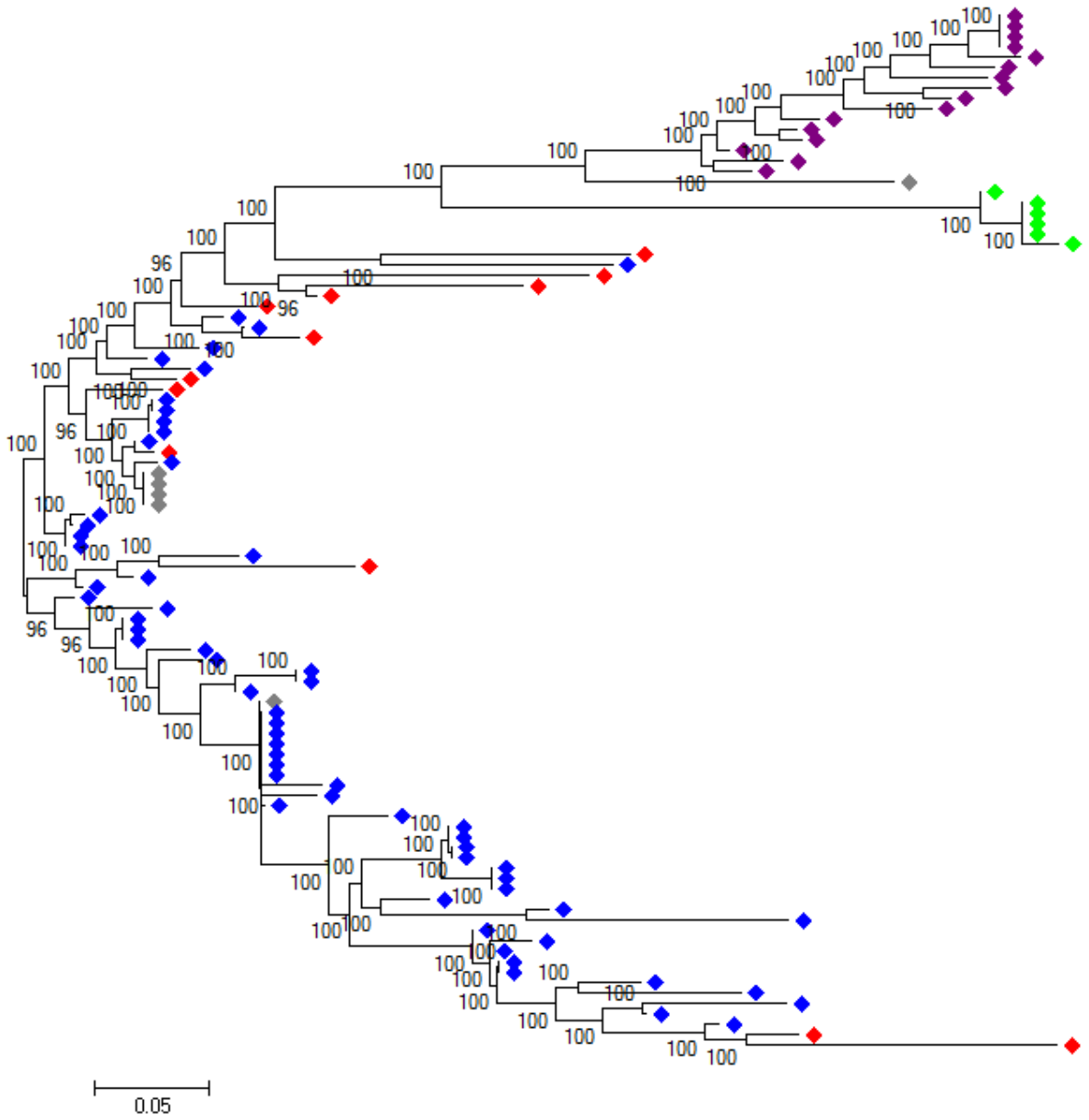


Figure S7.168 Chromosome 4: 39335001 - 39360001

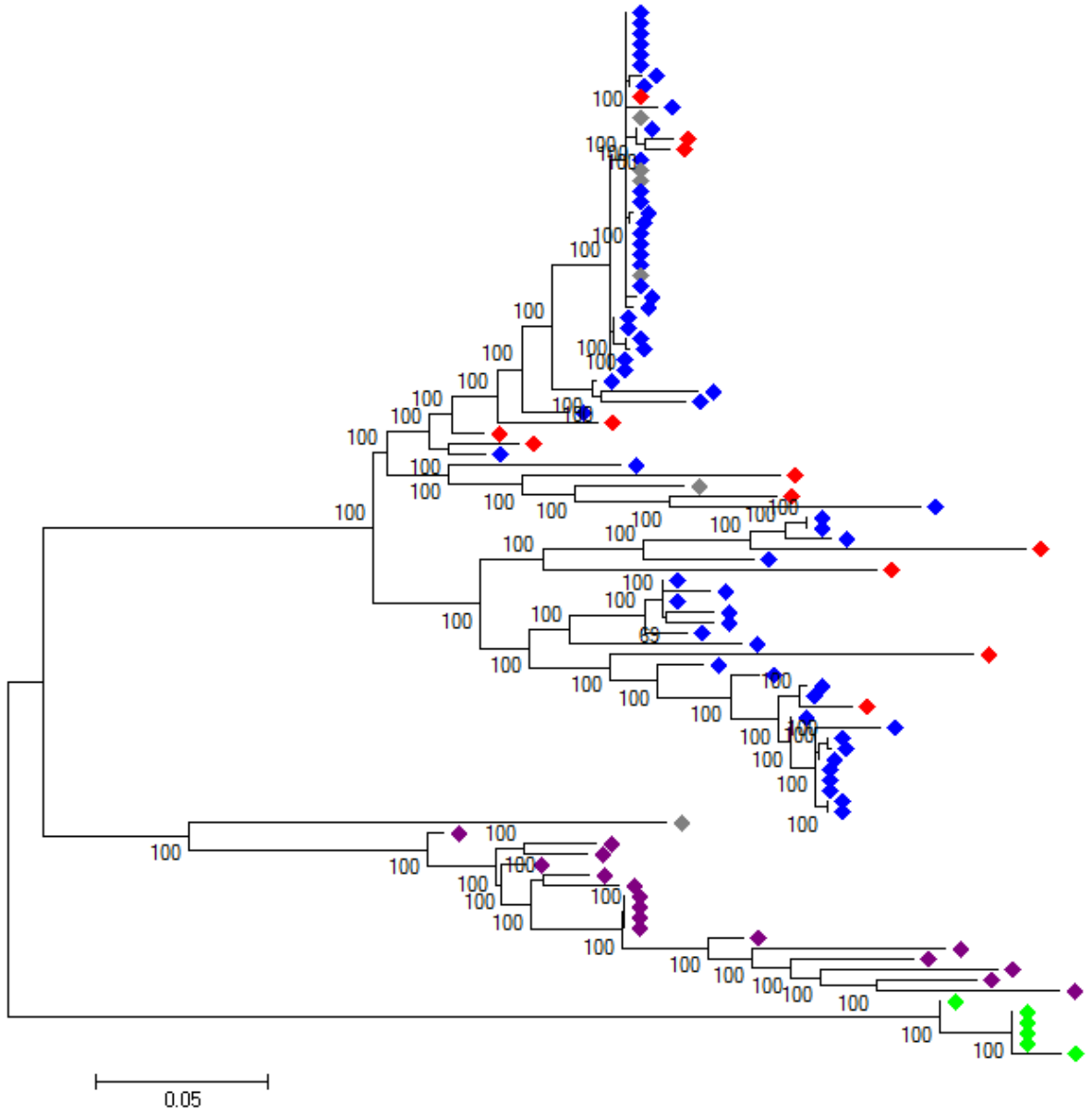


Figure S7.169 Chromosome 4: 39385001 - 39405001

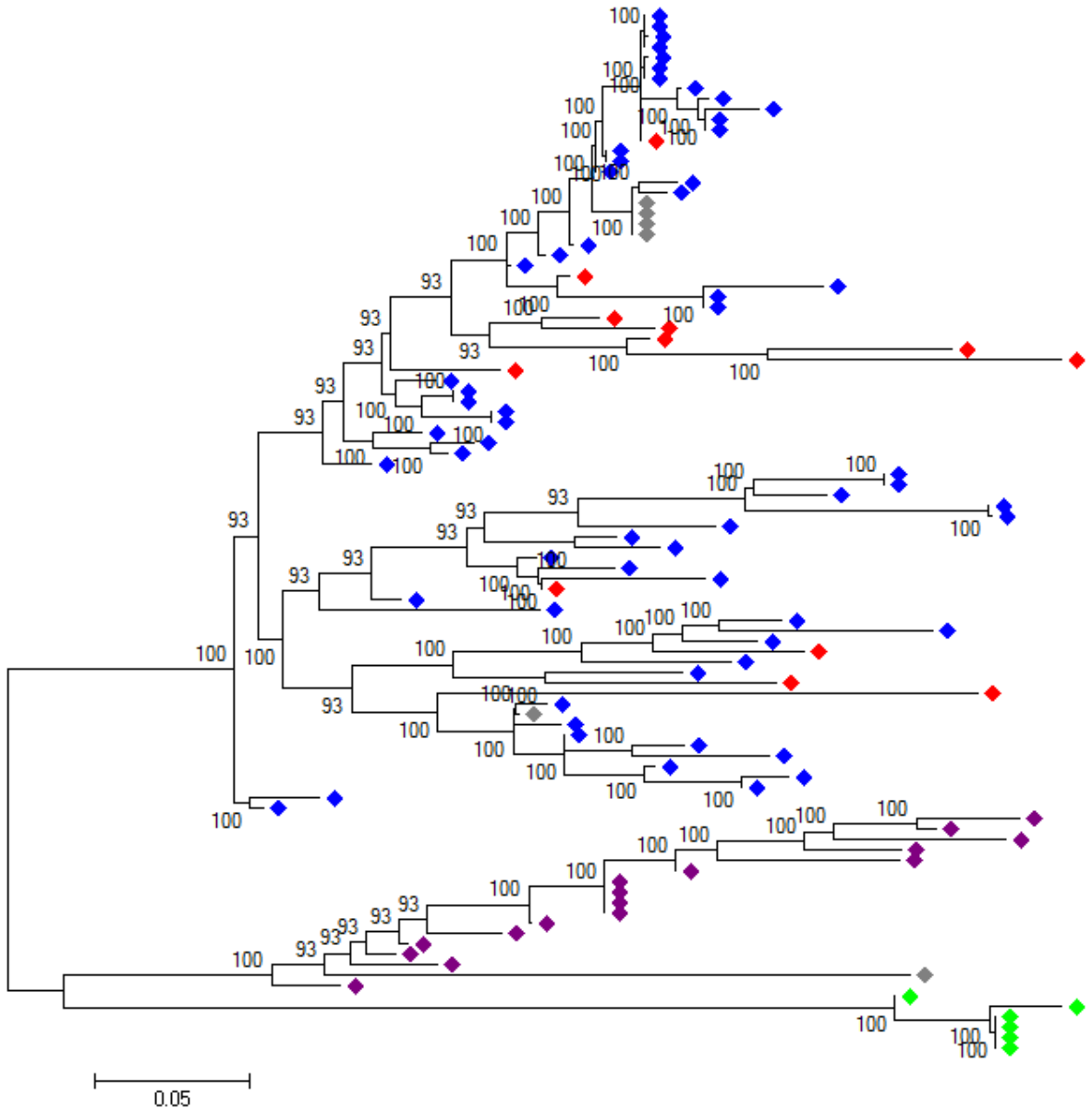


Figure S7.170 Chromosome 4: 39450001 - 3947500

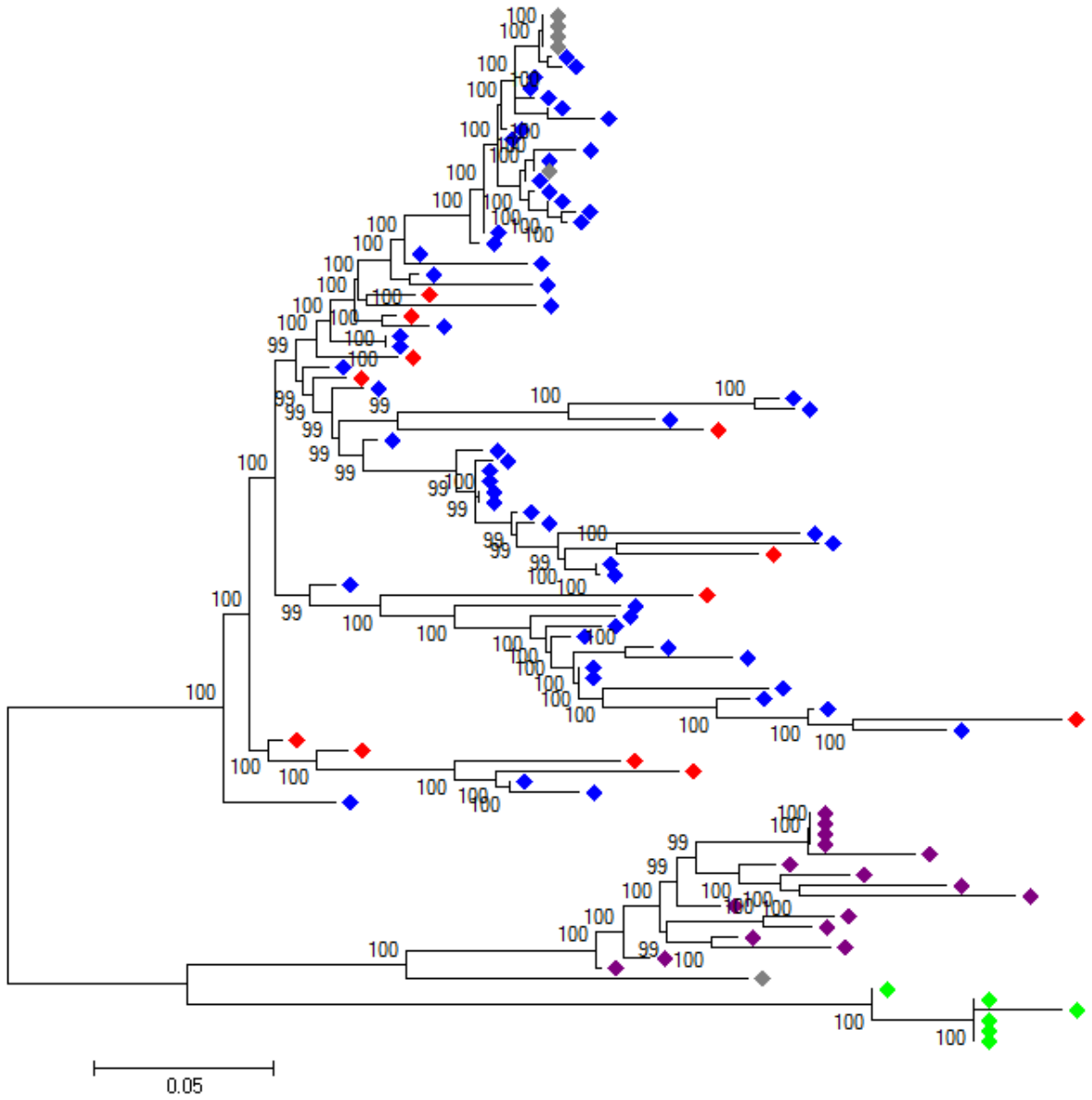


Figure S7.171 Chromosome 4: 39495001 - 39570001

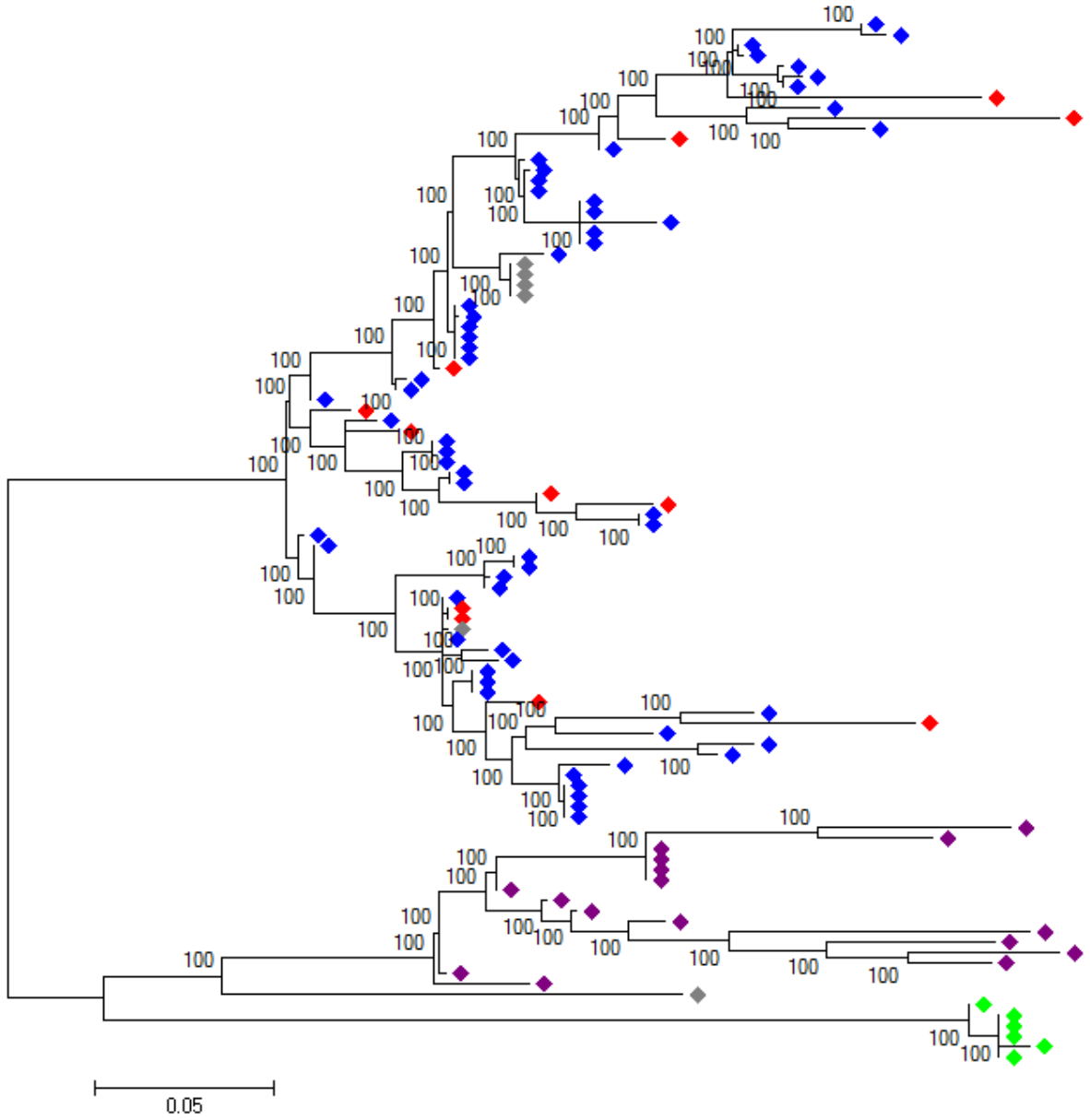


Figure S7.172 Chromosome 4: 39765001 - 39785001

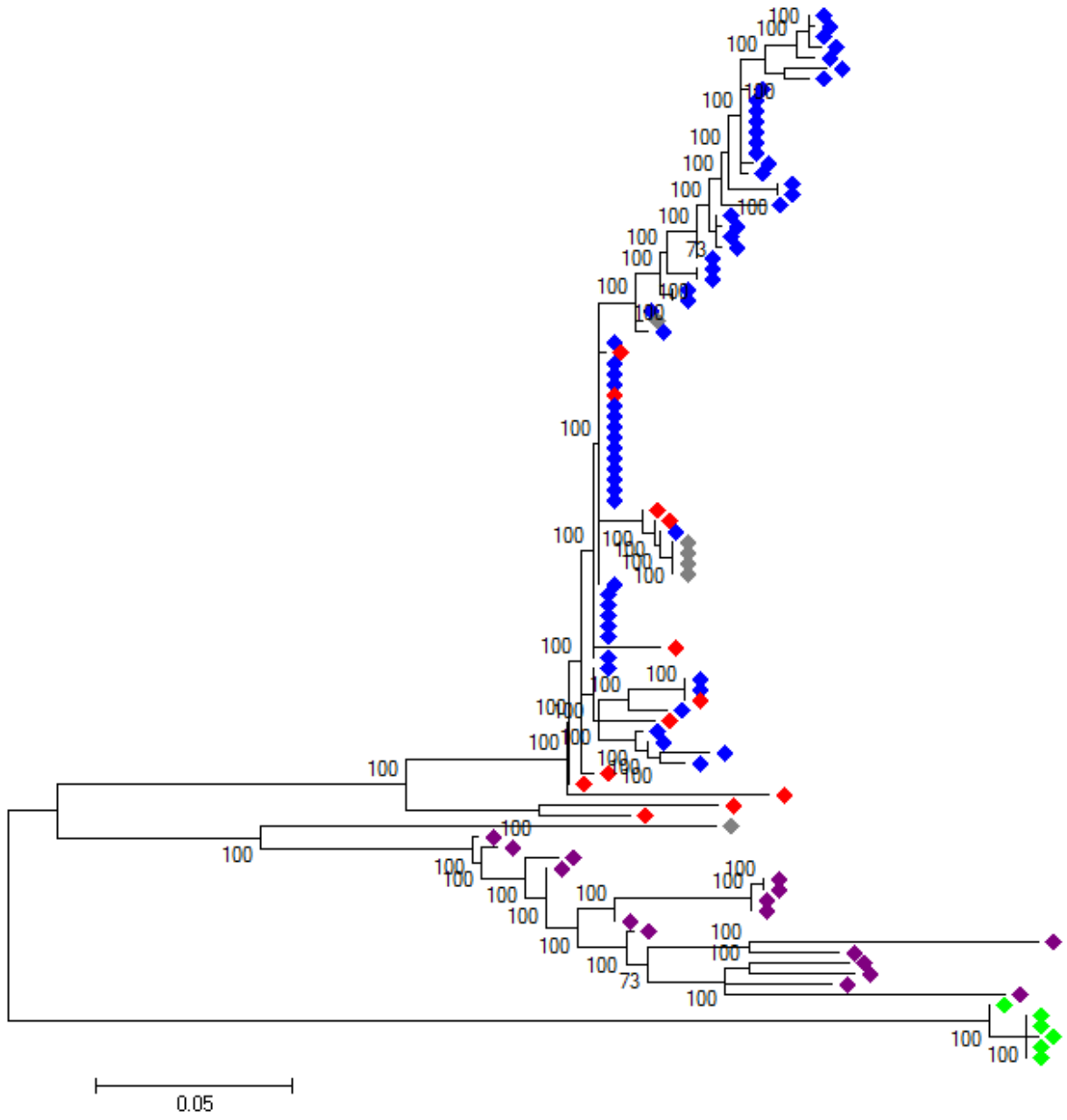


Figure S7.173 Chromosome 4: 40855001 - 40880001

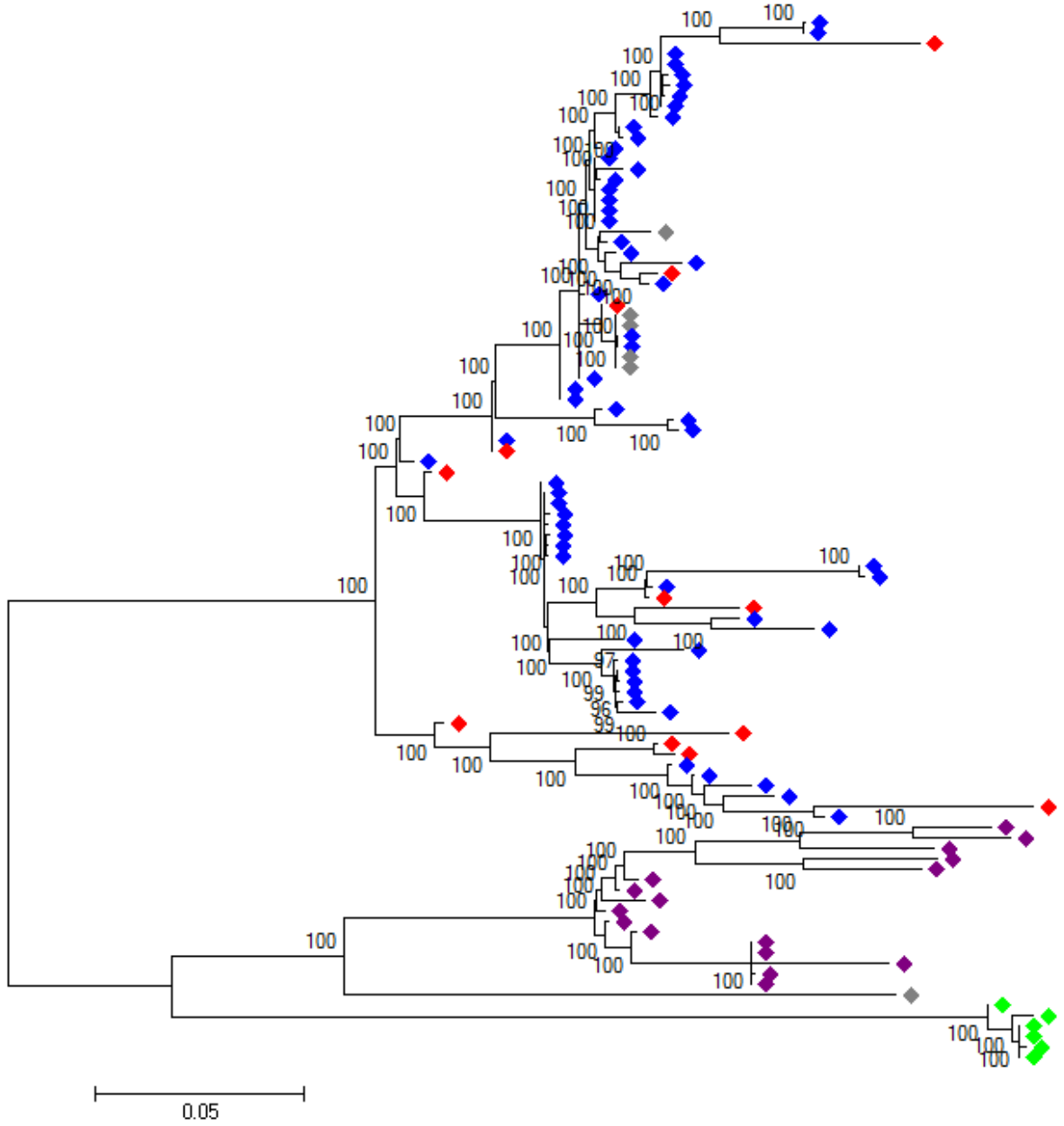


Figure S7.174 Chromosome 4: 40925001 - 40985001

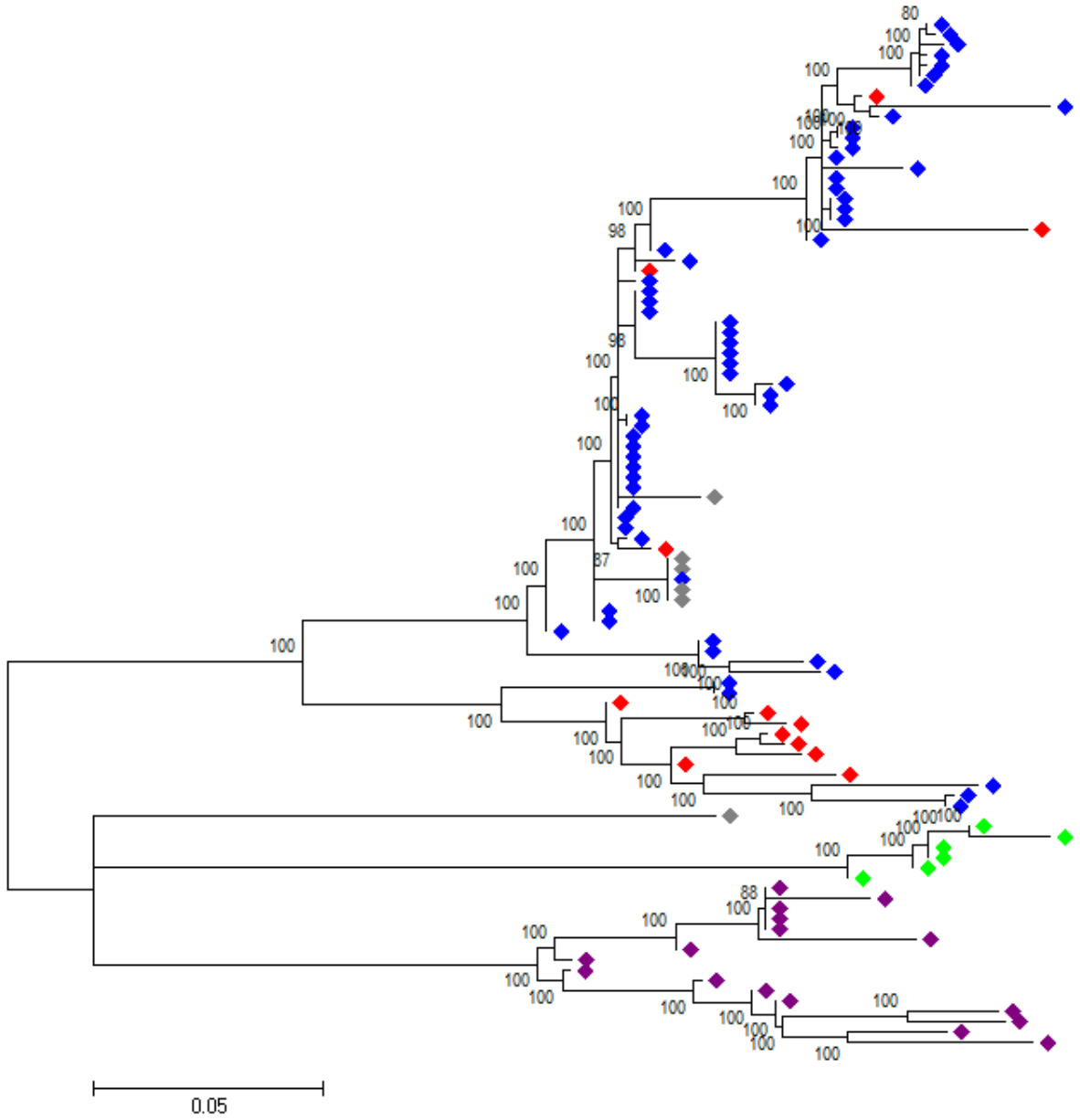


Figure S7.175 Chromosome 4: 41020001 - 41055001

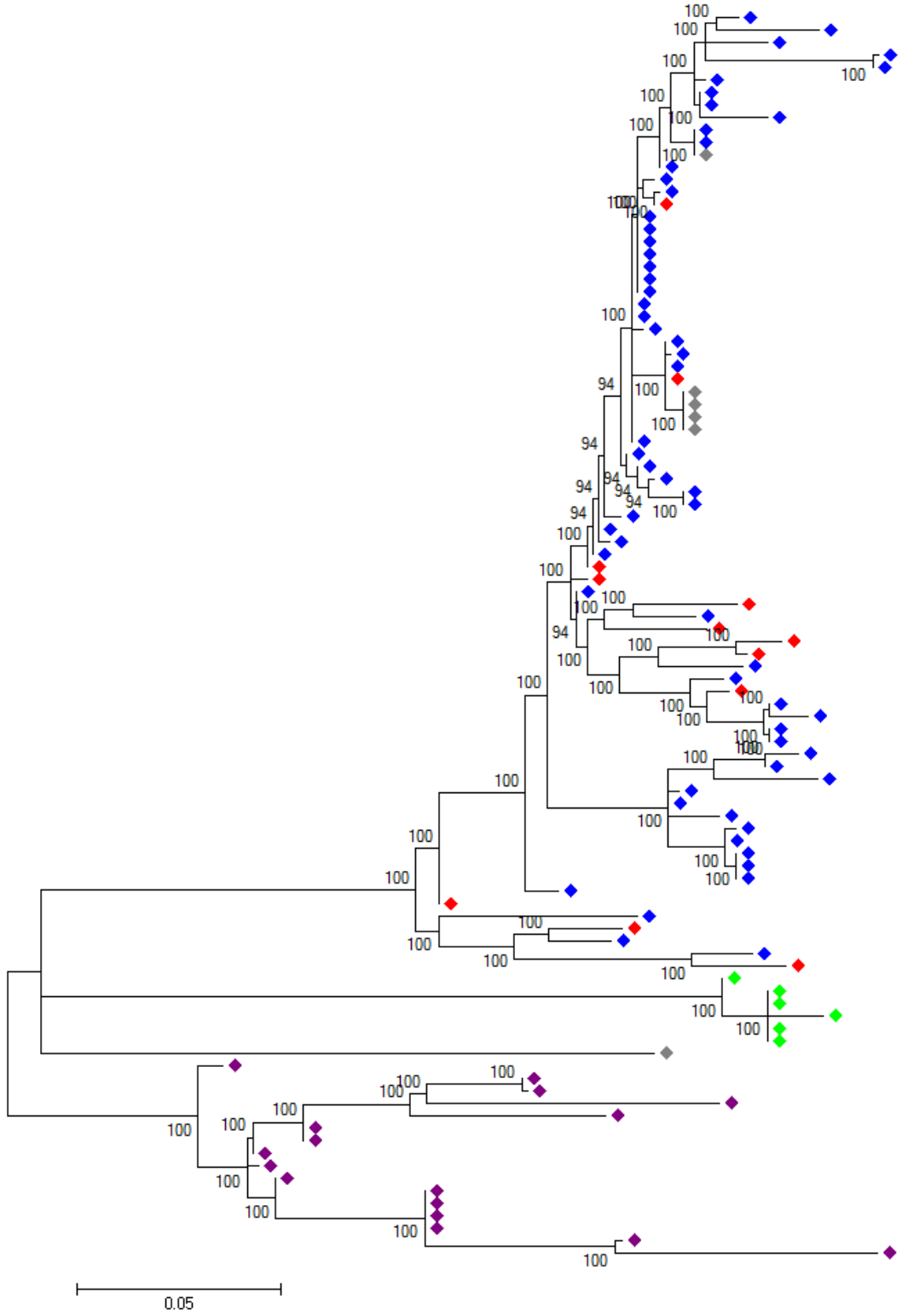


Figure S7.176 Chromosome 4: 41520001 - 41545001

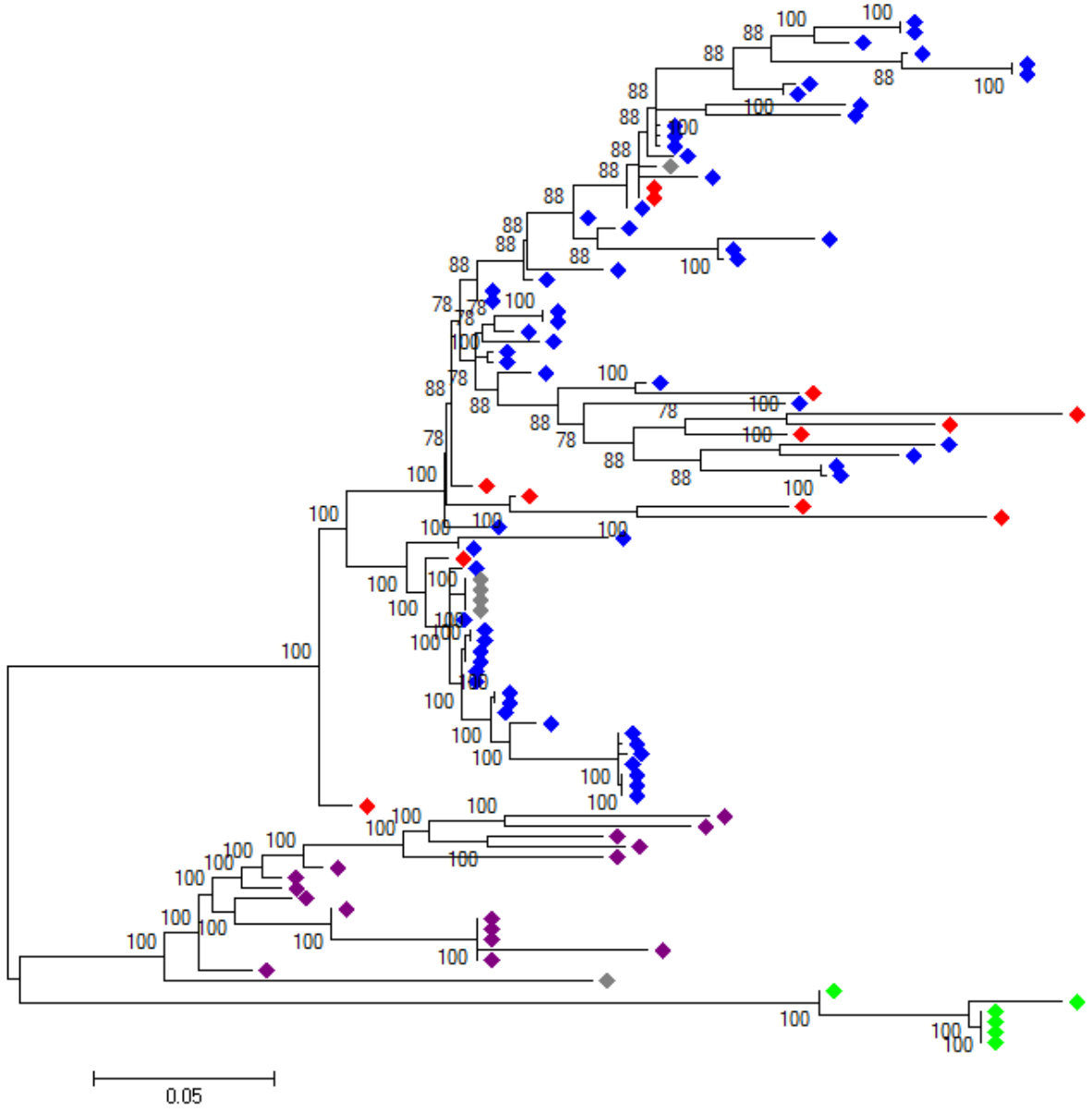


Figure S7.177 Chromosome 4: 41615001 - 41640001

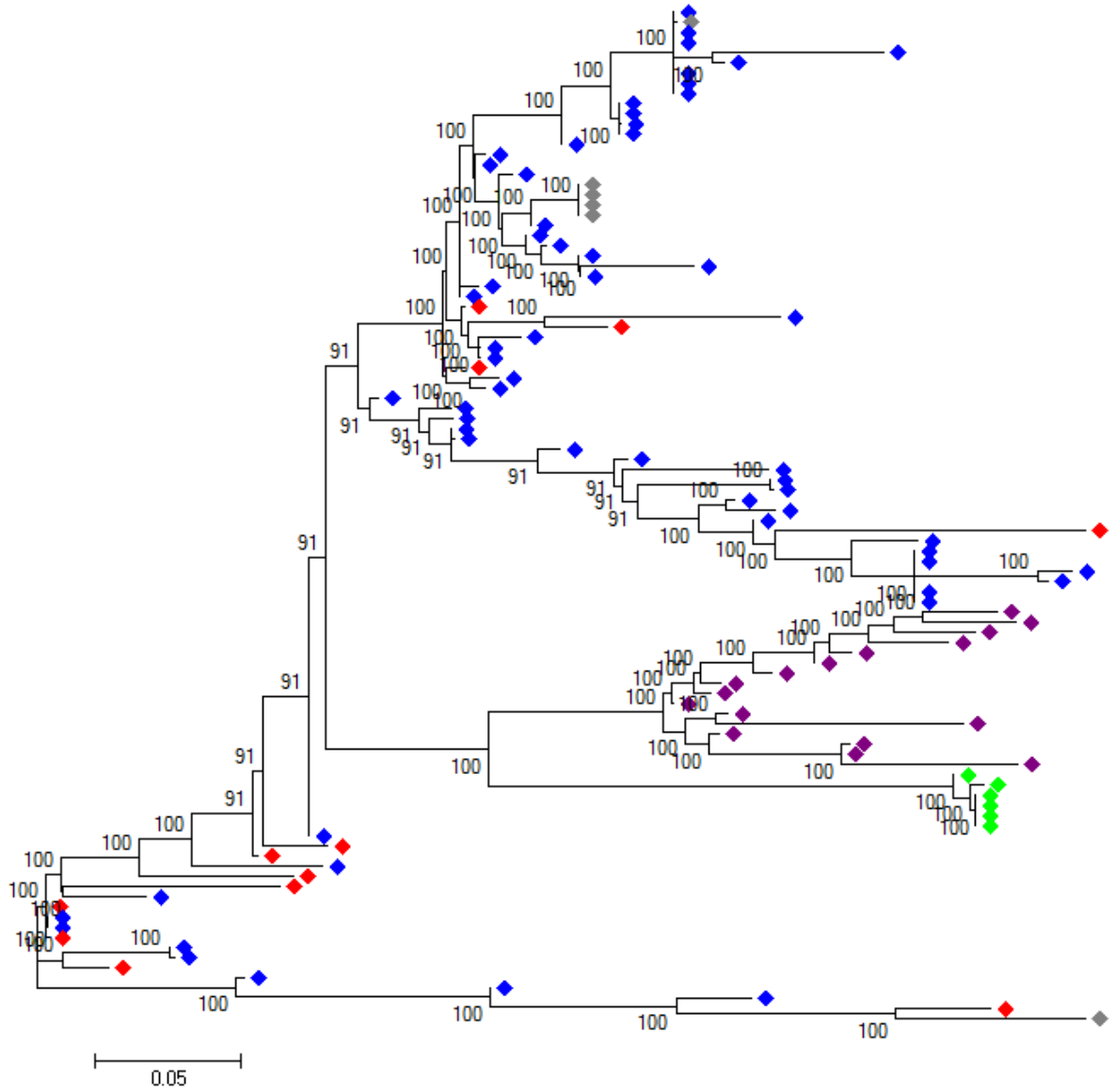


Figure S7.178 Chromosome 4: 41810001 - 41840001

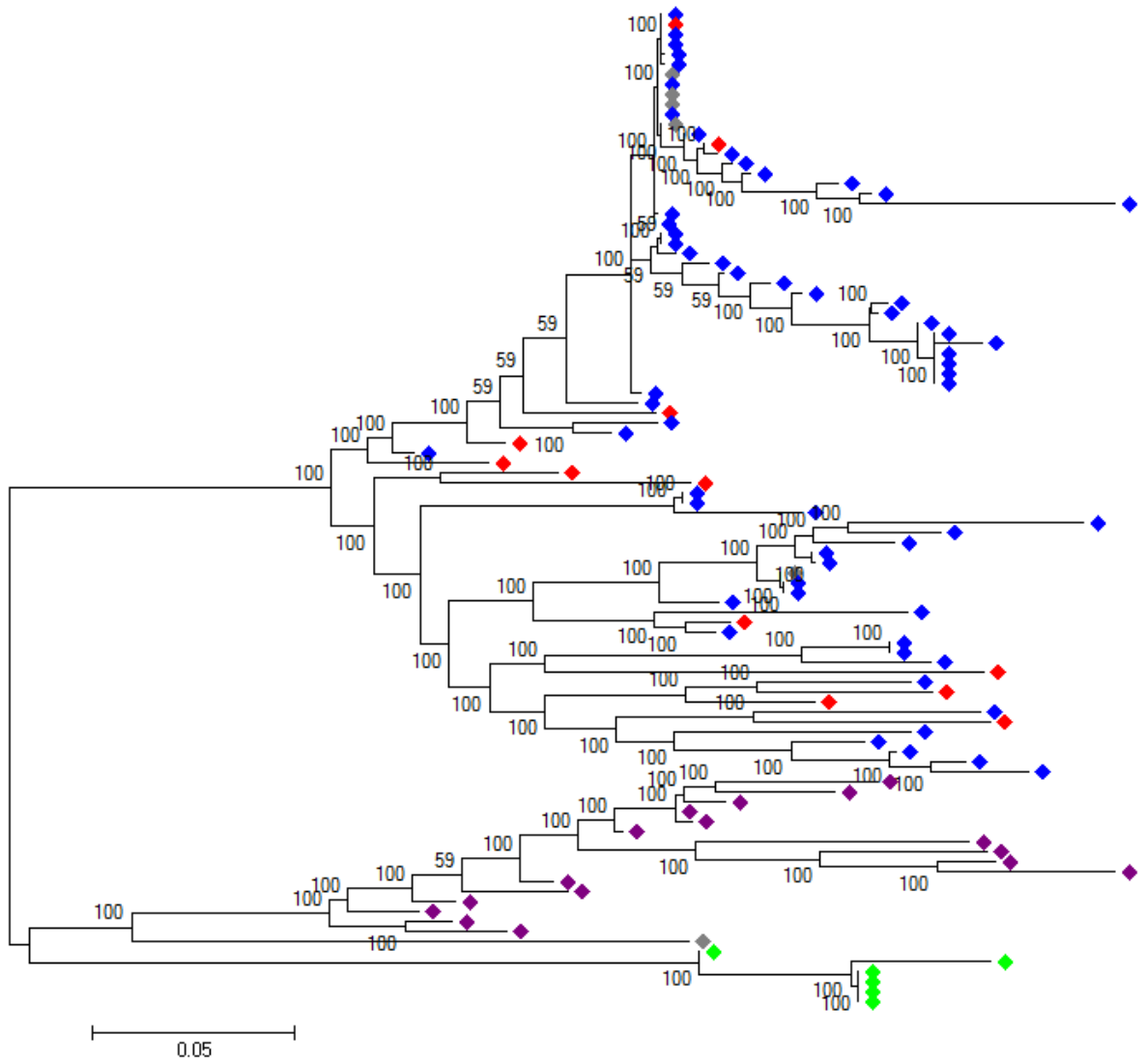


Figure S7.179 Chromosome 4: 42665001 - 42695001

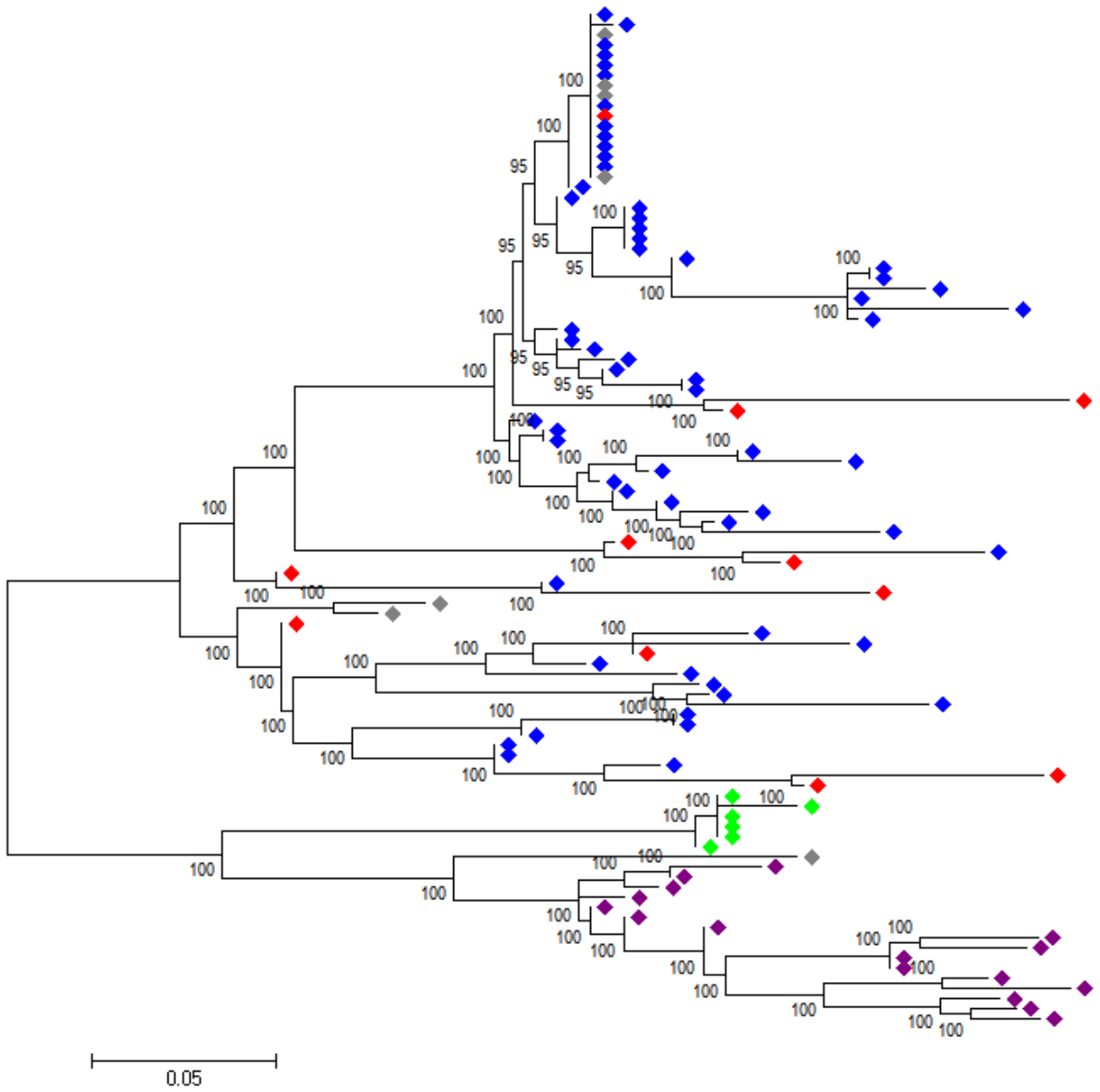


Figure S7.180 Chromosome 4: 42710001 - 42735001

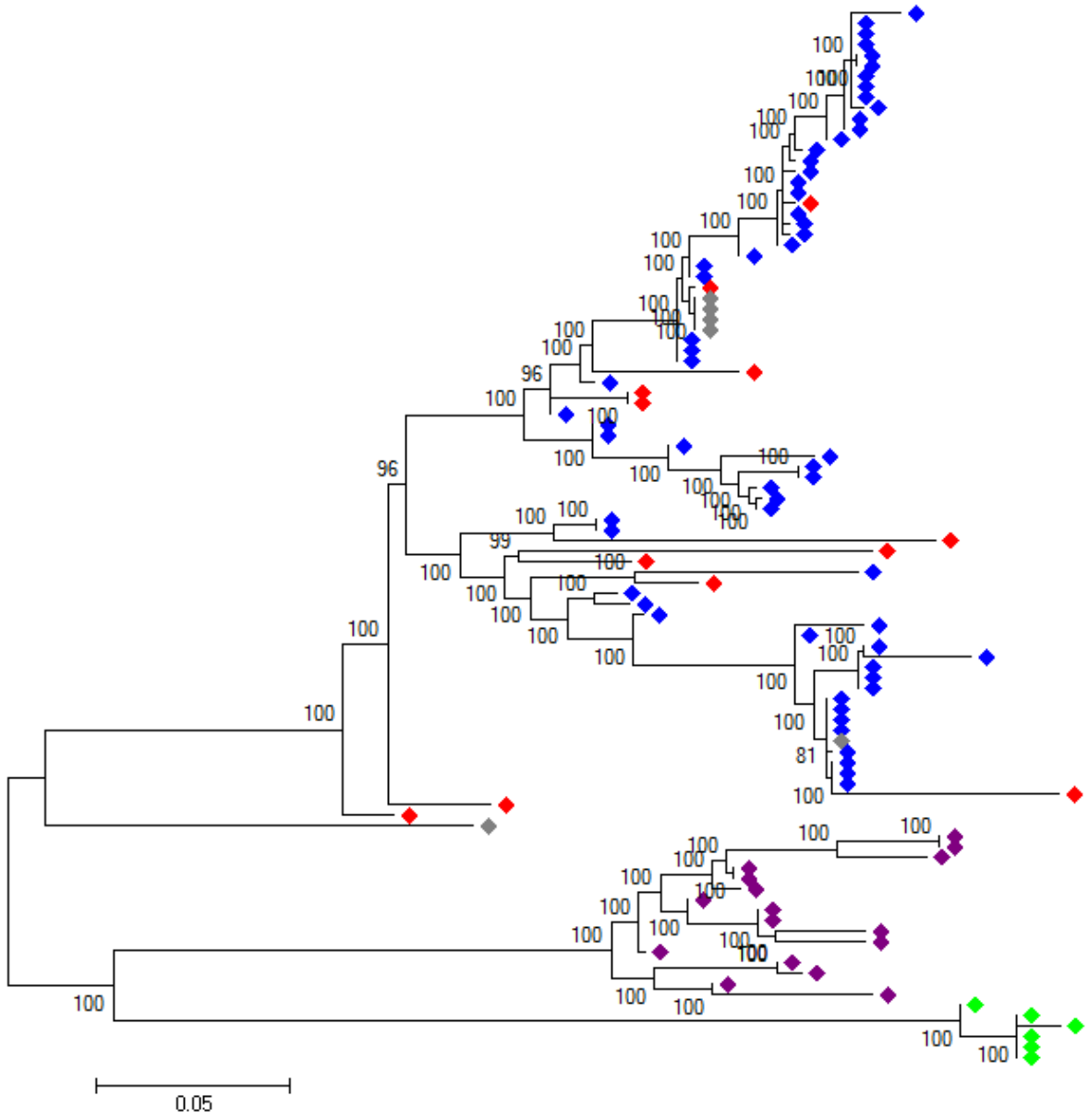


Figure S7.181 Chromosome 4: 42855001 - 42875001

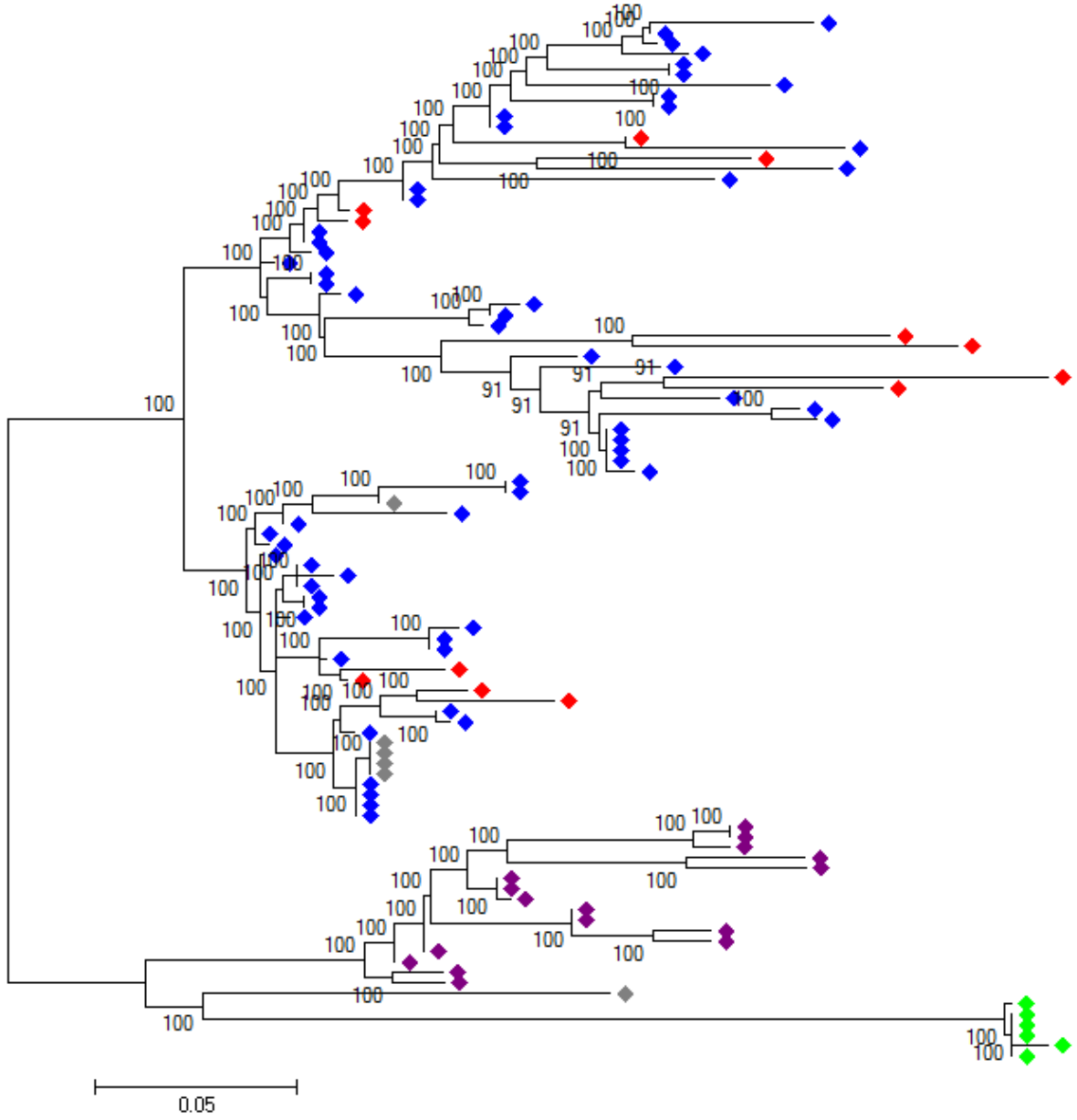


Figure S7.182 Chromosome 4: 42980001 - 43000001

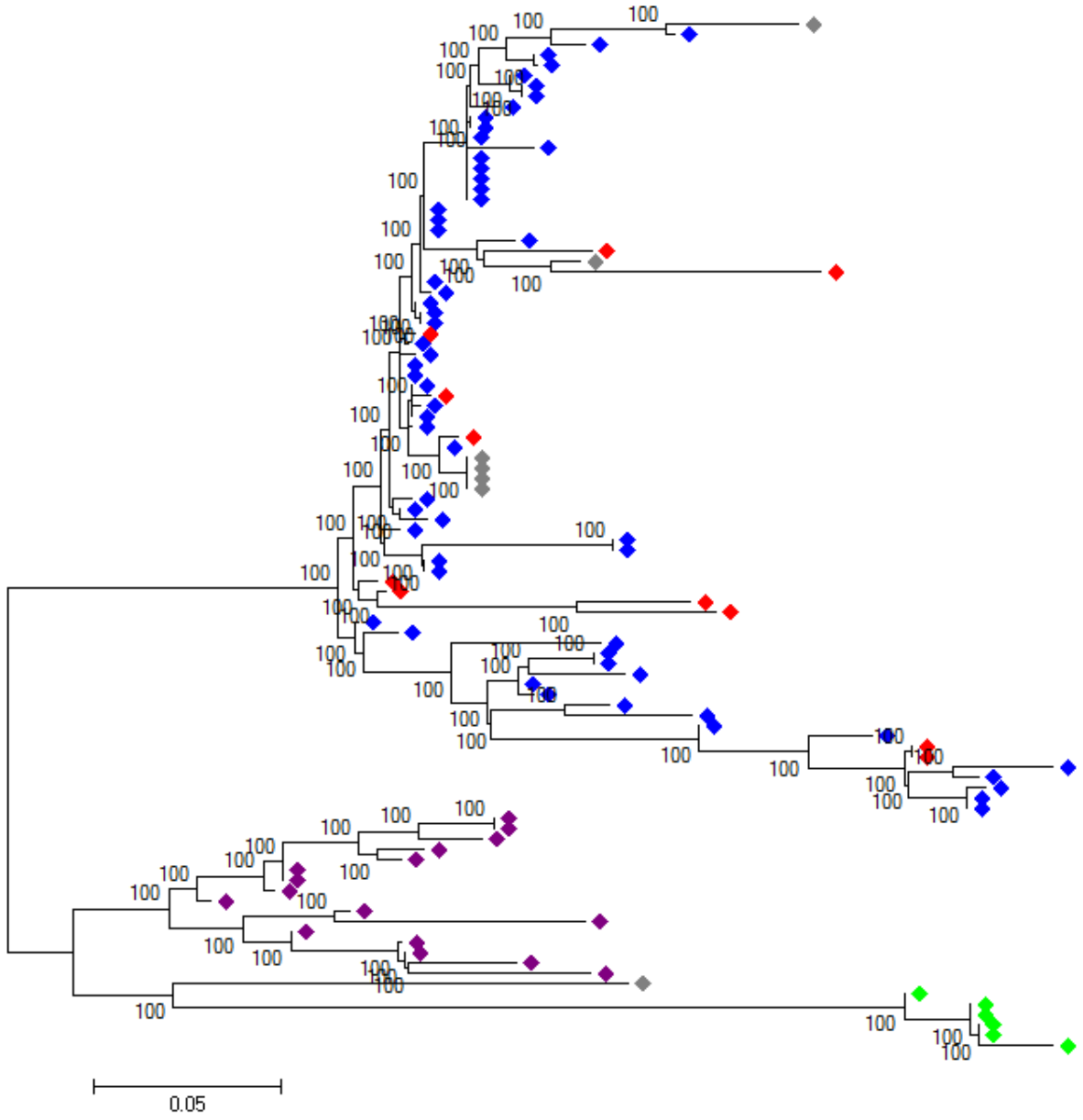


Figure S7.183 Chromosome 4: 43265001 - 43290001

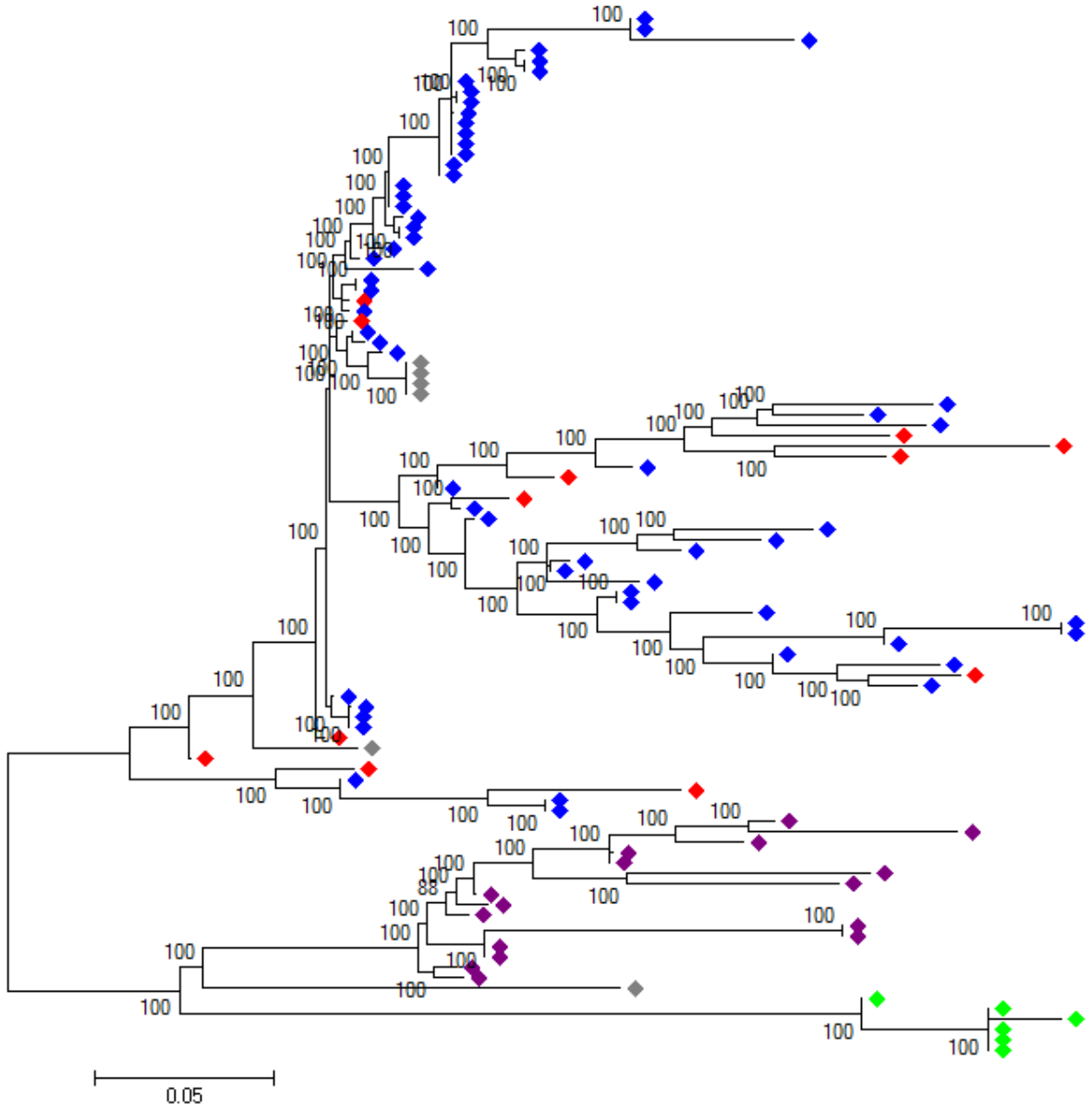


Figure S7.184 Chromosome 4: 43300001 - 43345001

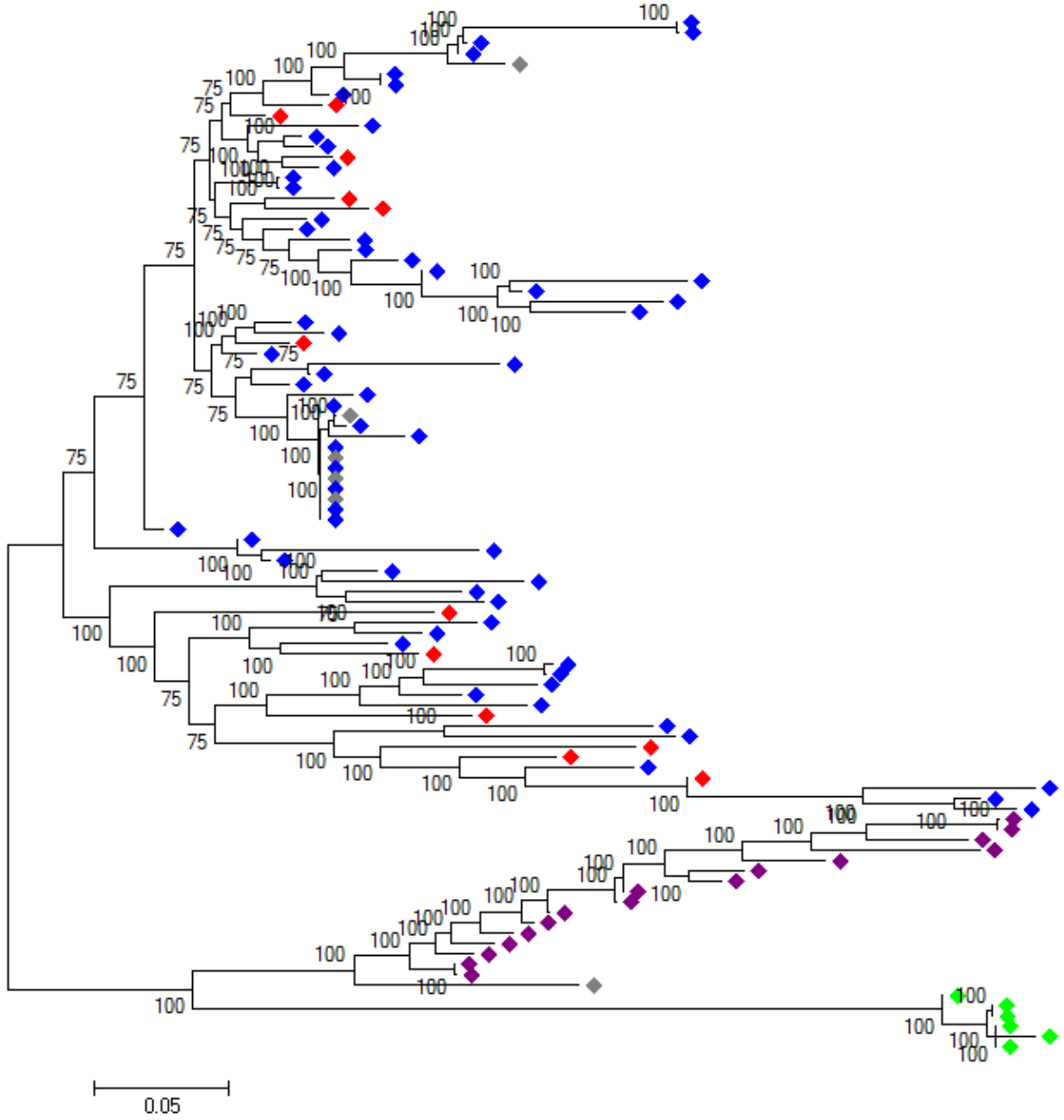


Figure S7.185 Chromosome 4: 43455001 - 43505001

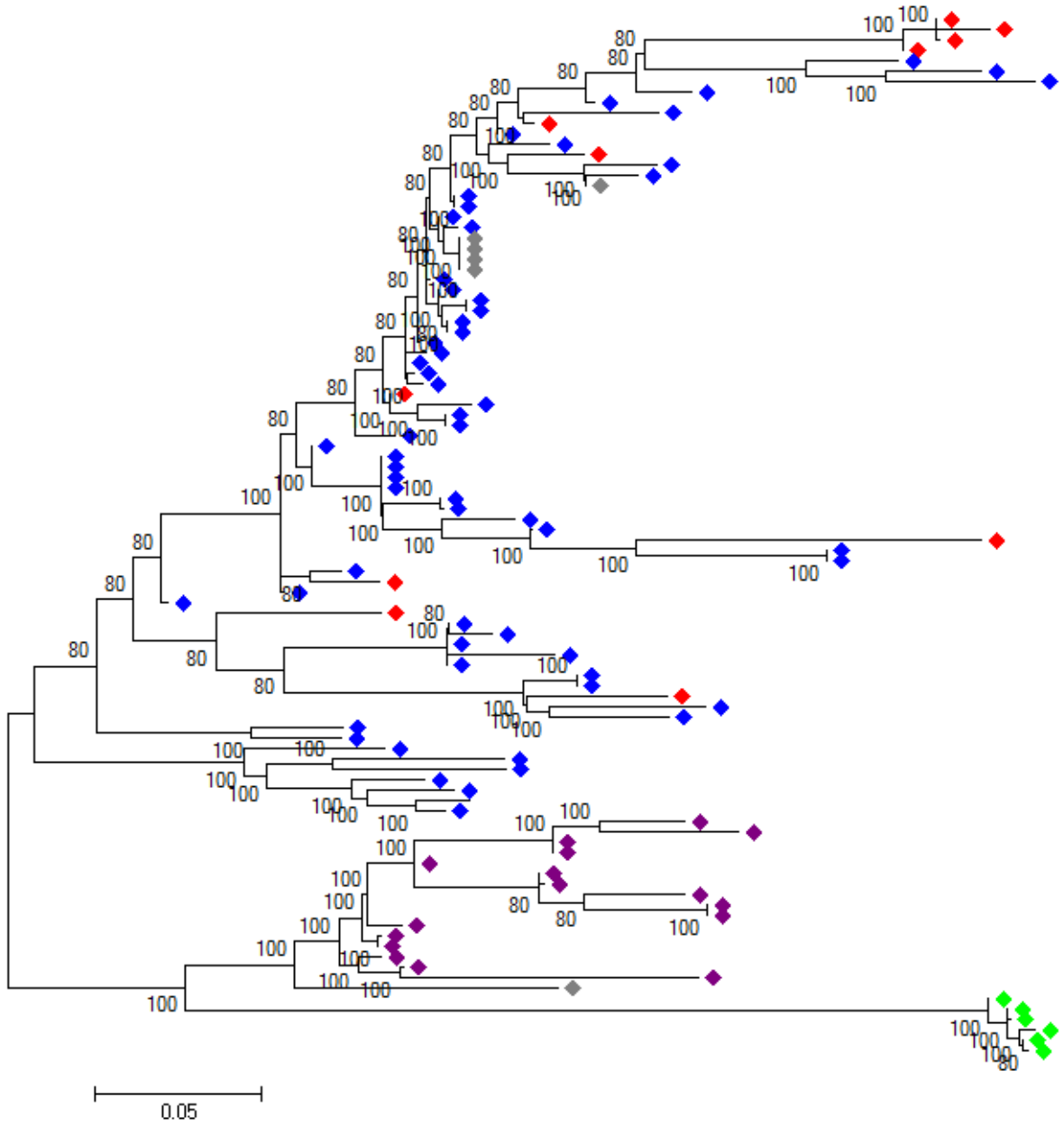


Figure S7.186 Chromosome 4: 43680001 - 43705001

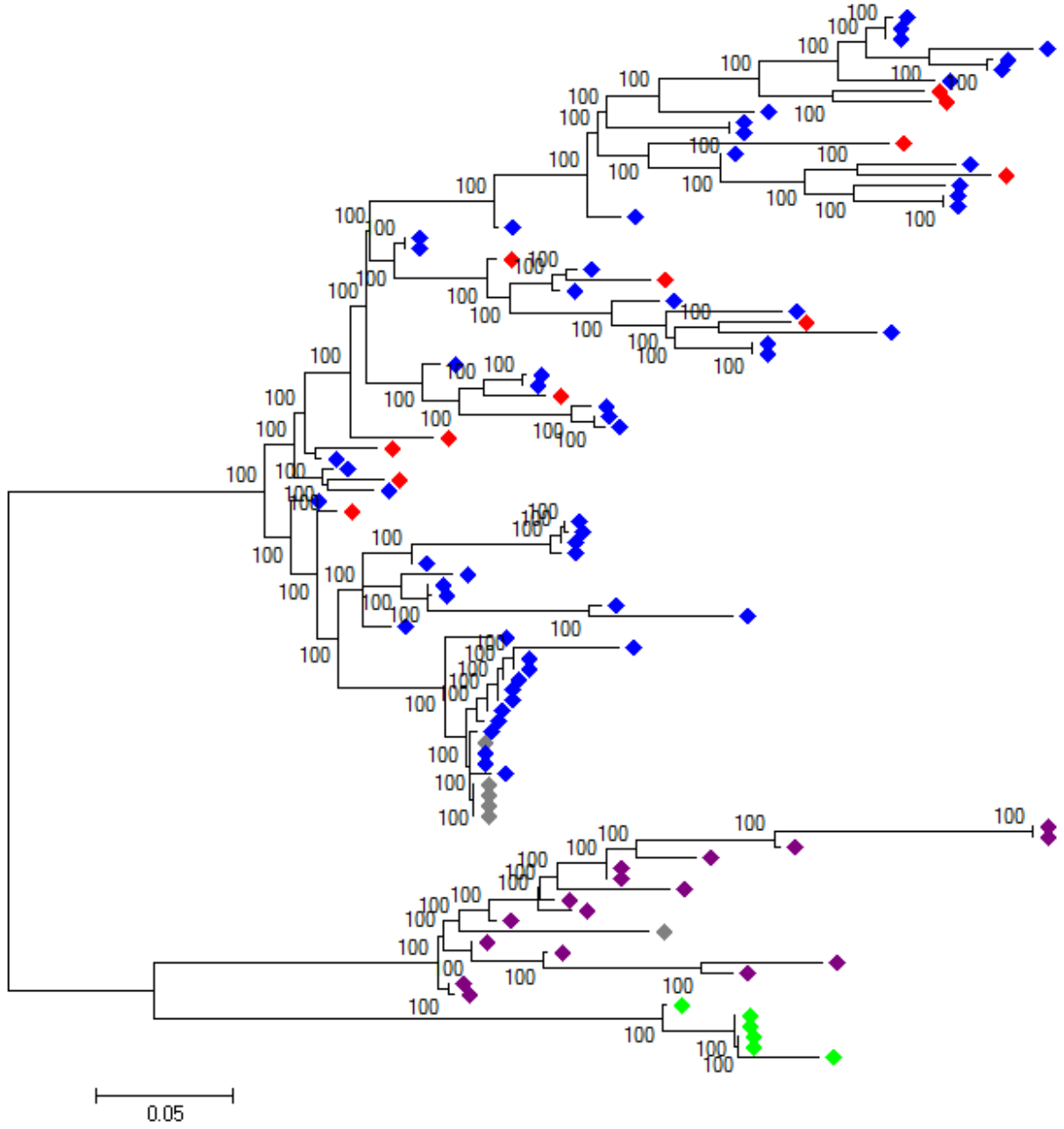


Figure S7.187 Chromosome 4: 43735001 - 43755001

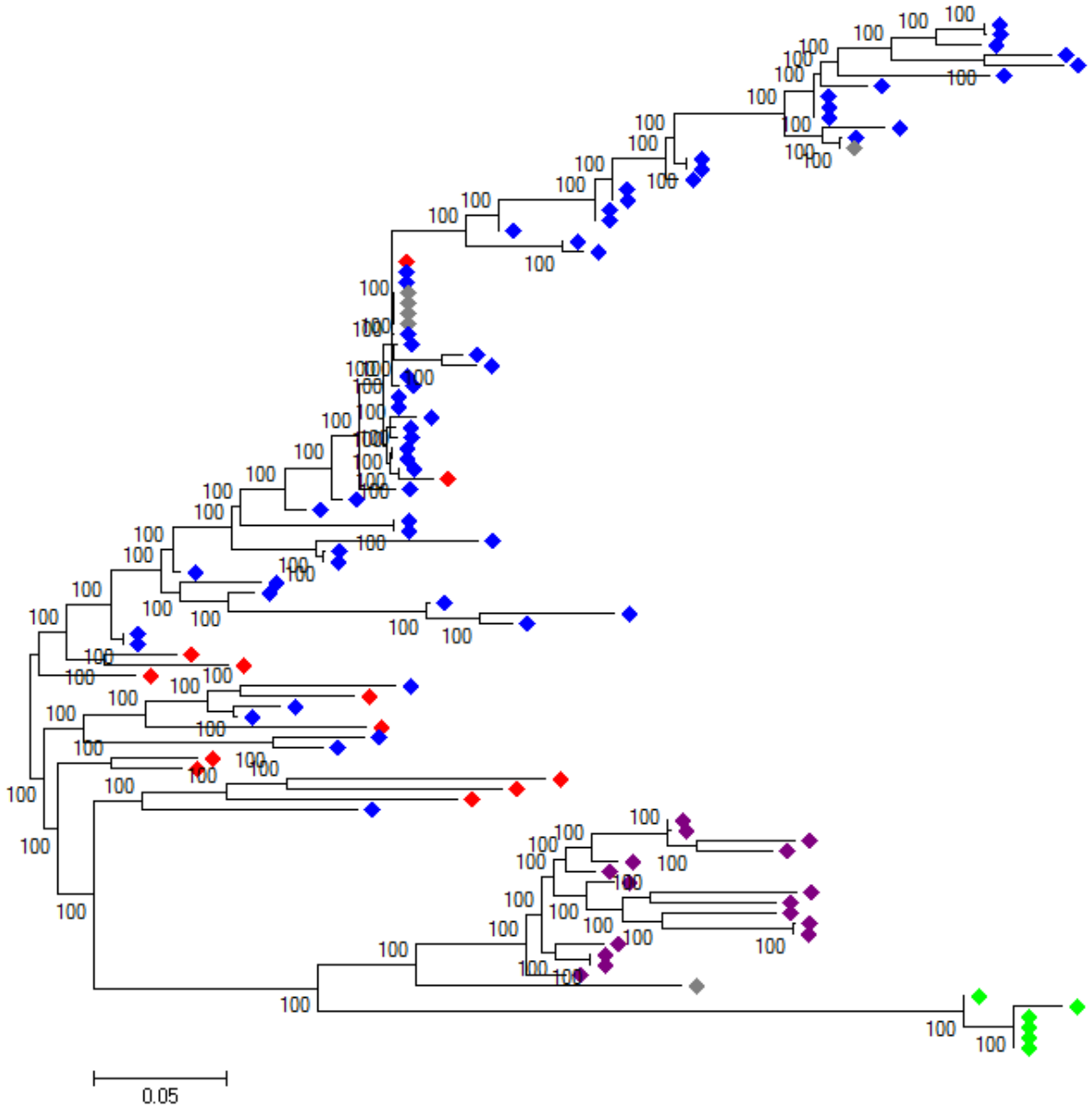


Figure S7.188 Chromosome 4: 43950001 - 44055001

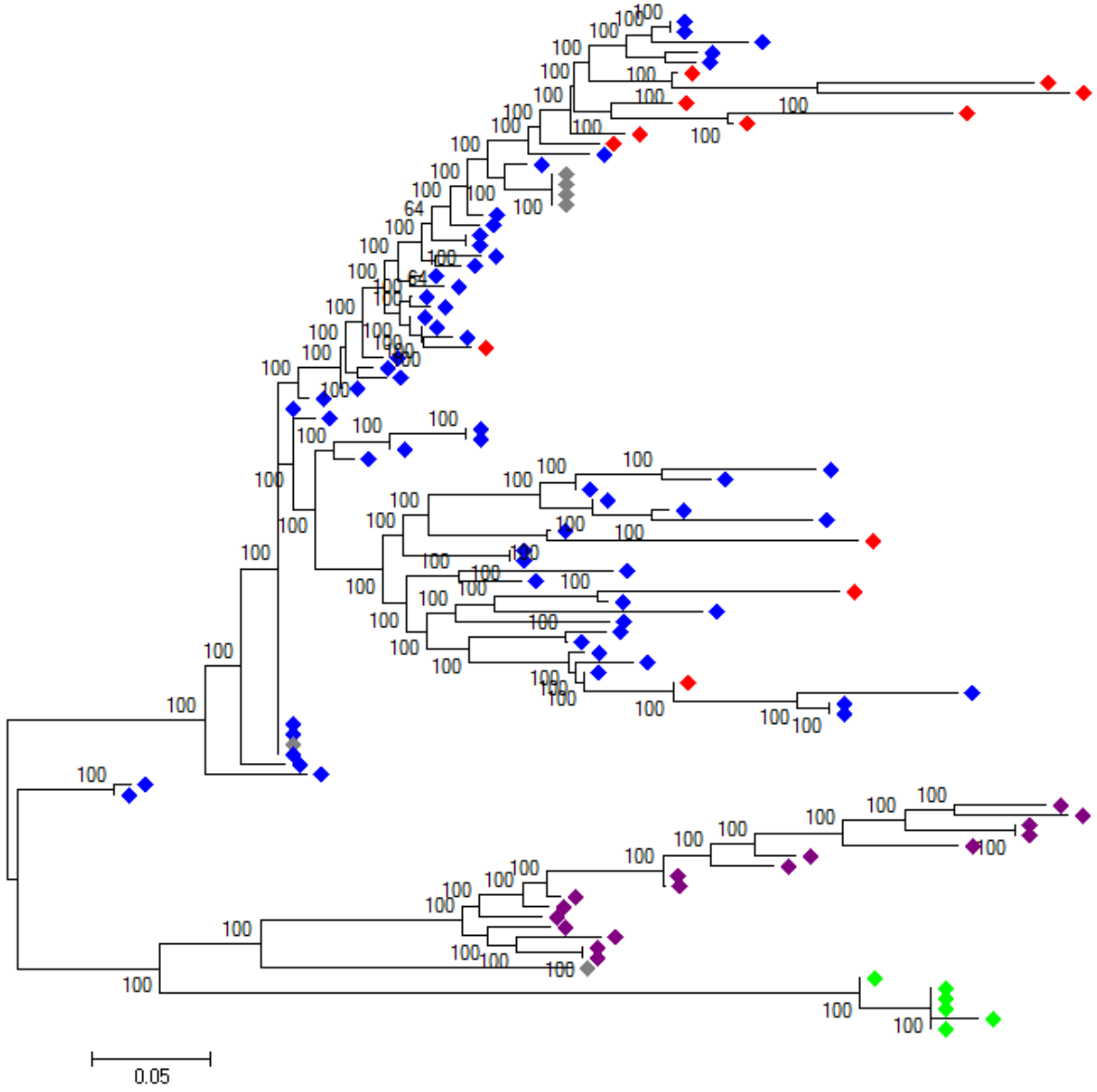


Figure S7.189 Chromosome 4: 44110001 - 44140001

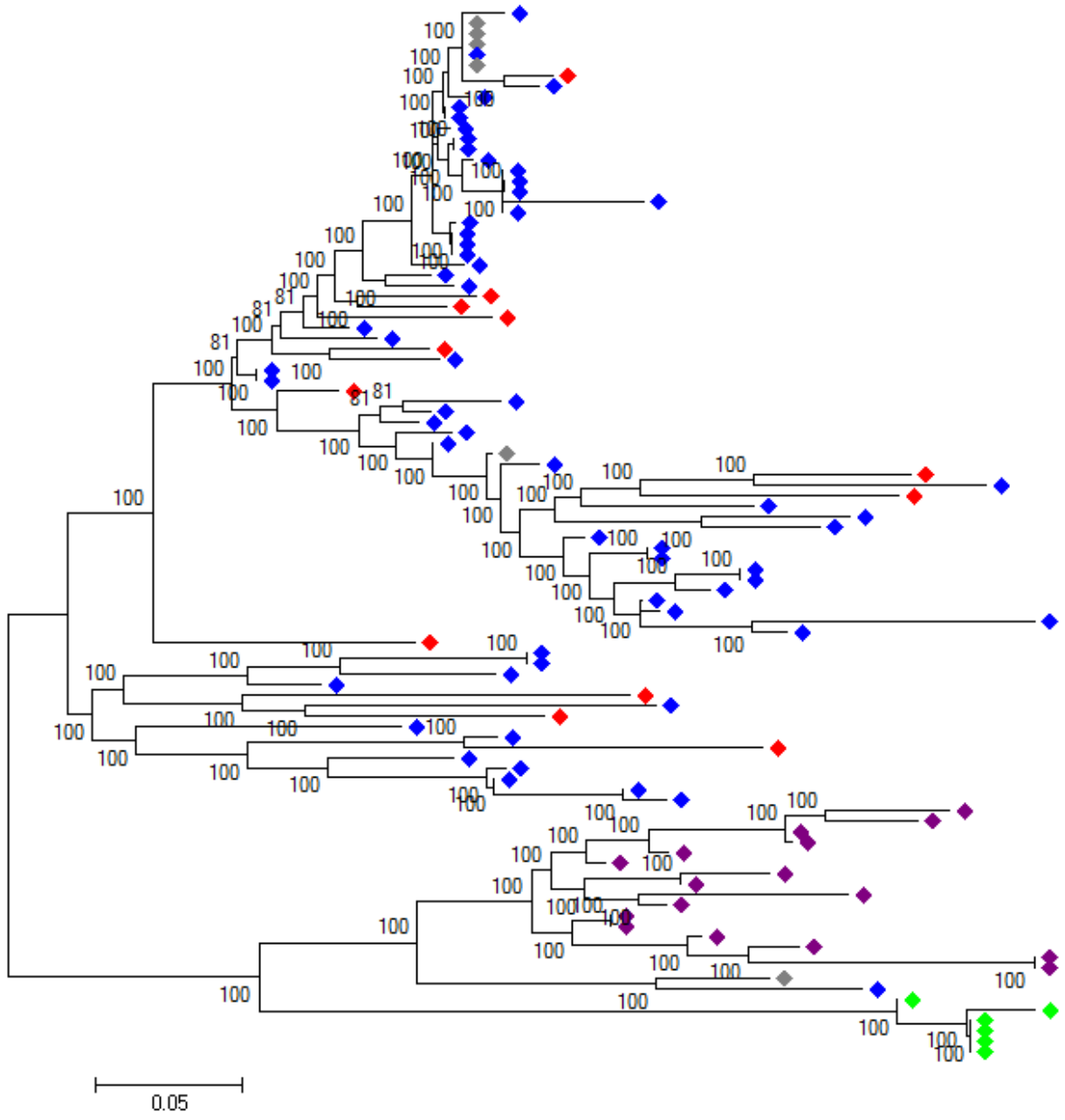


Figure S7.190 Chromosome 4: 44200001 - 44230001

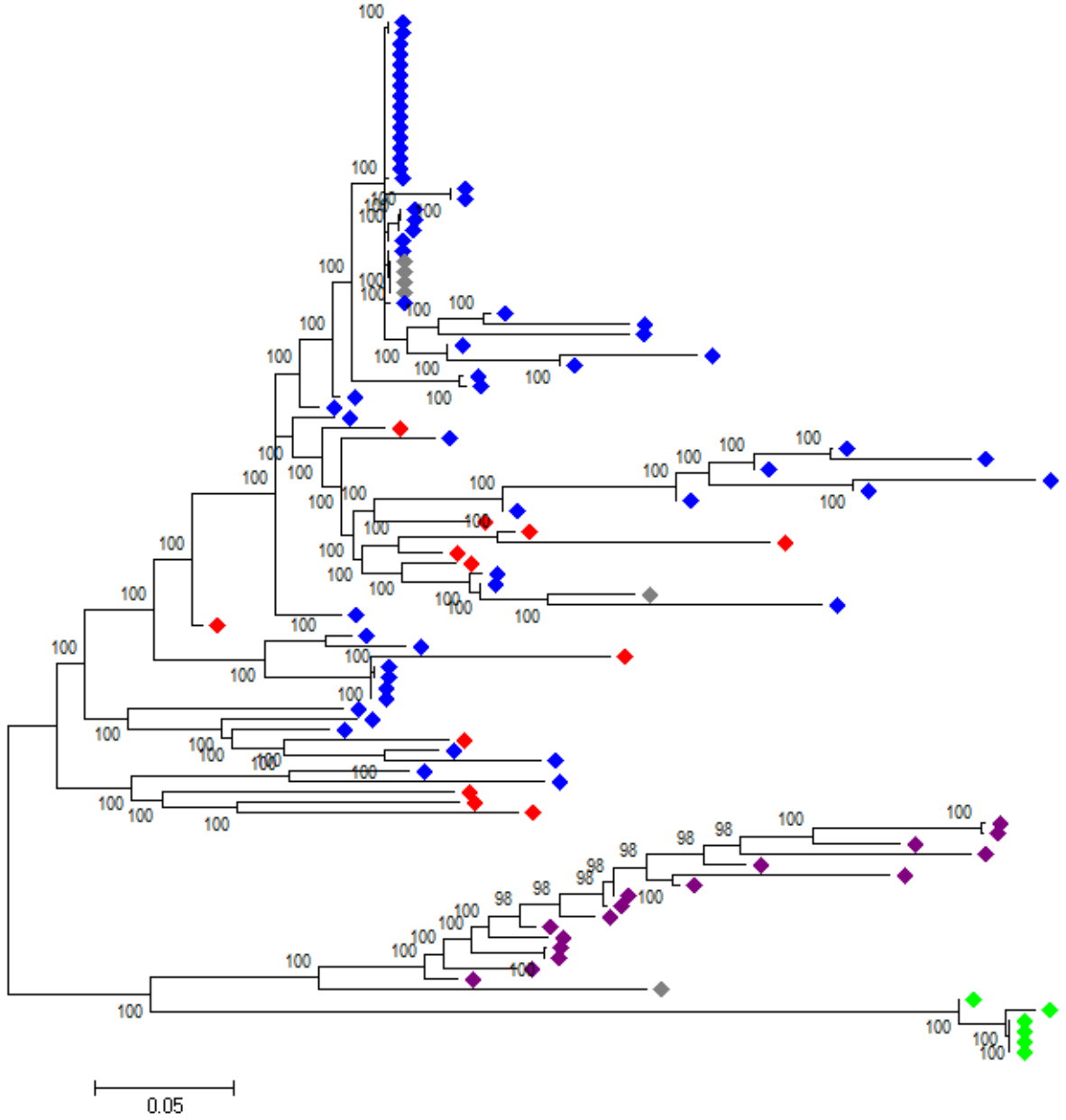


Figure S7.191 Chromosome 4: 44265001 - 44305001

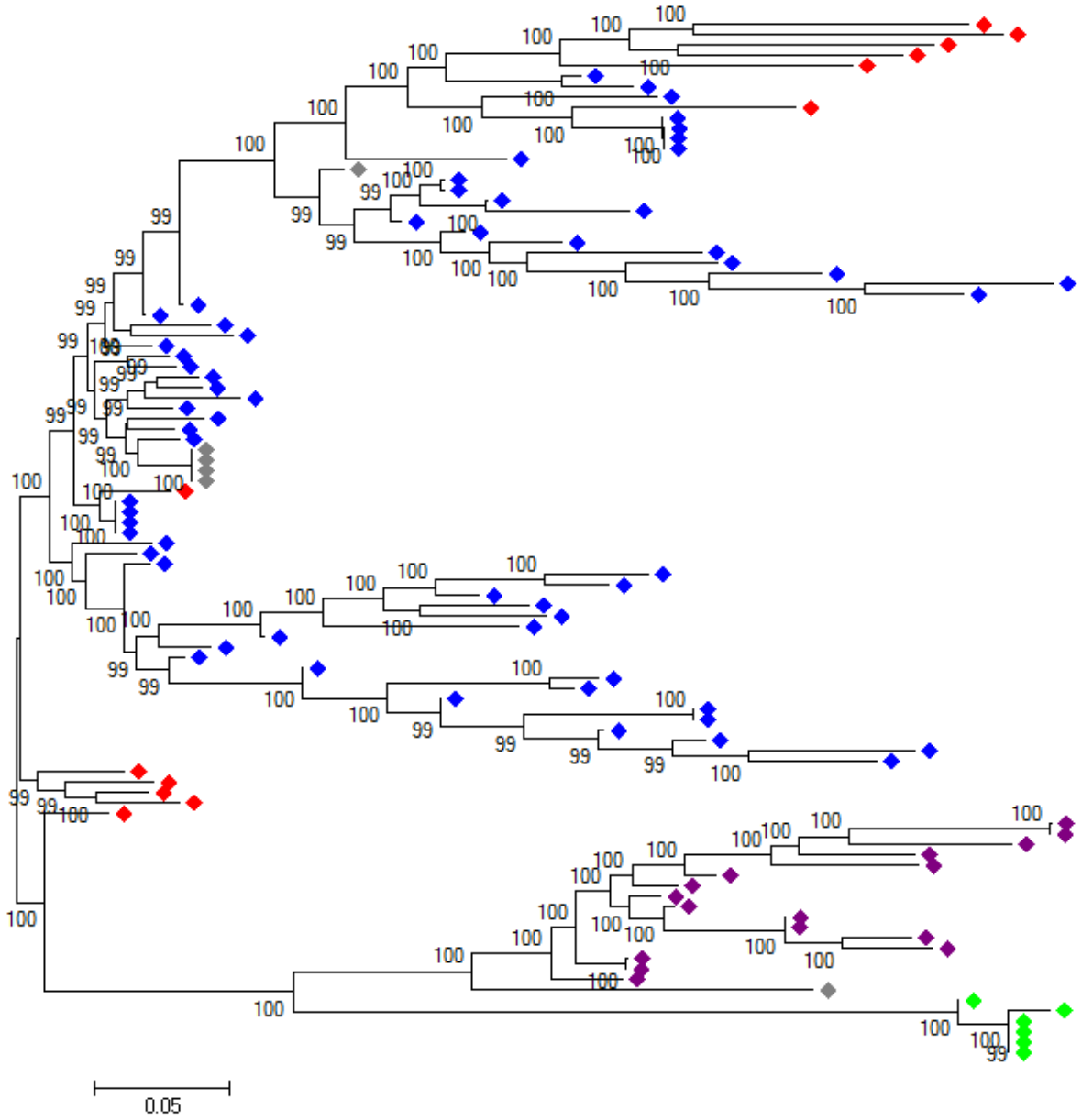


Figure S7.192 Chromosome 4: 44380001 - 44490001

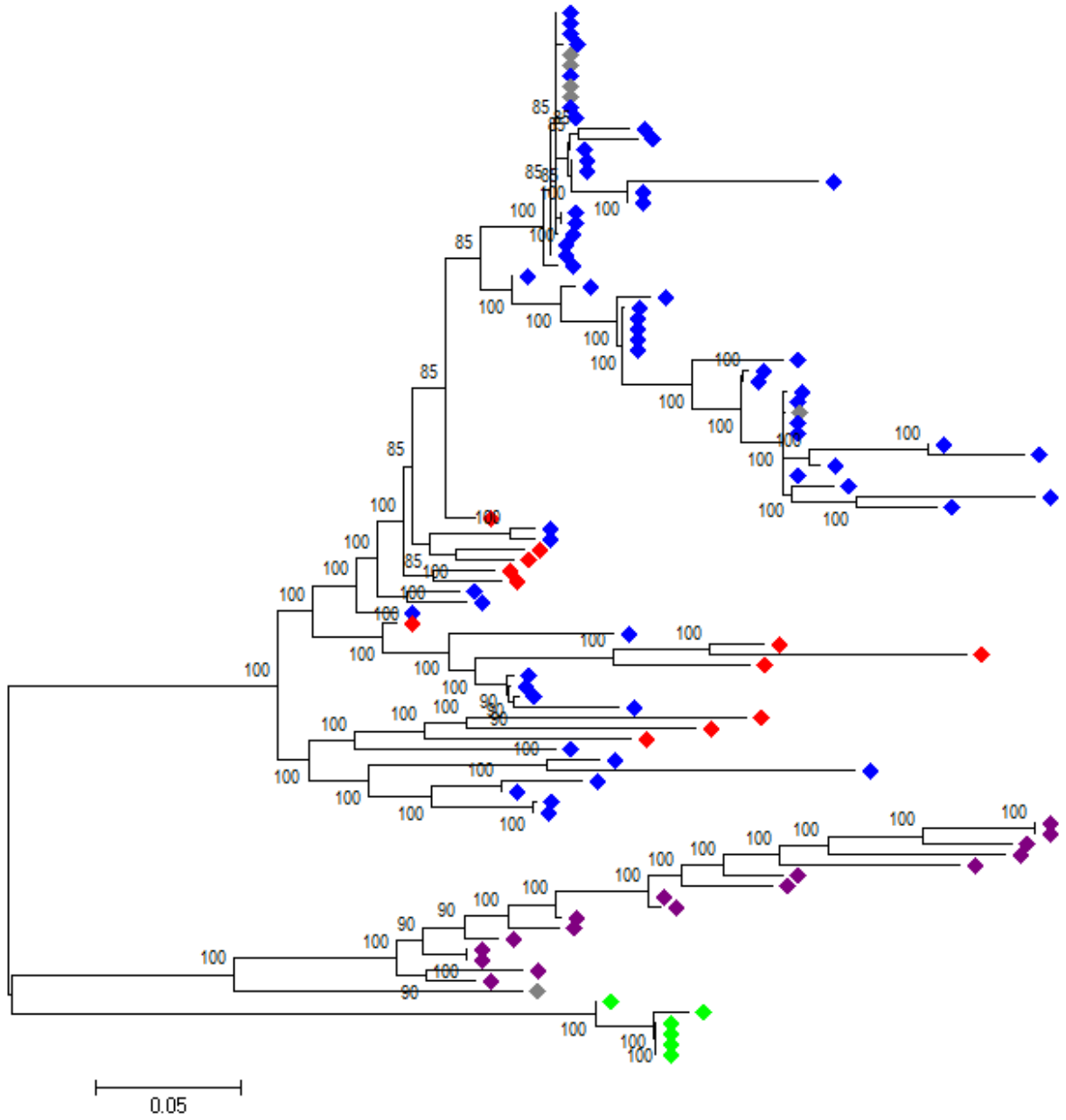


Figure S7.193 Chromosome 4: 44615001 - 44660001

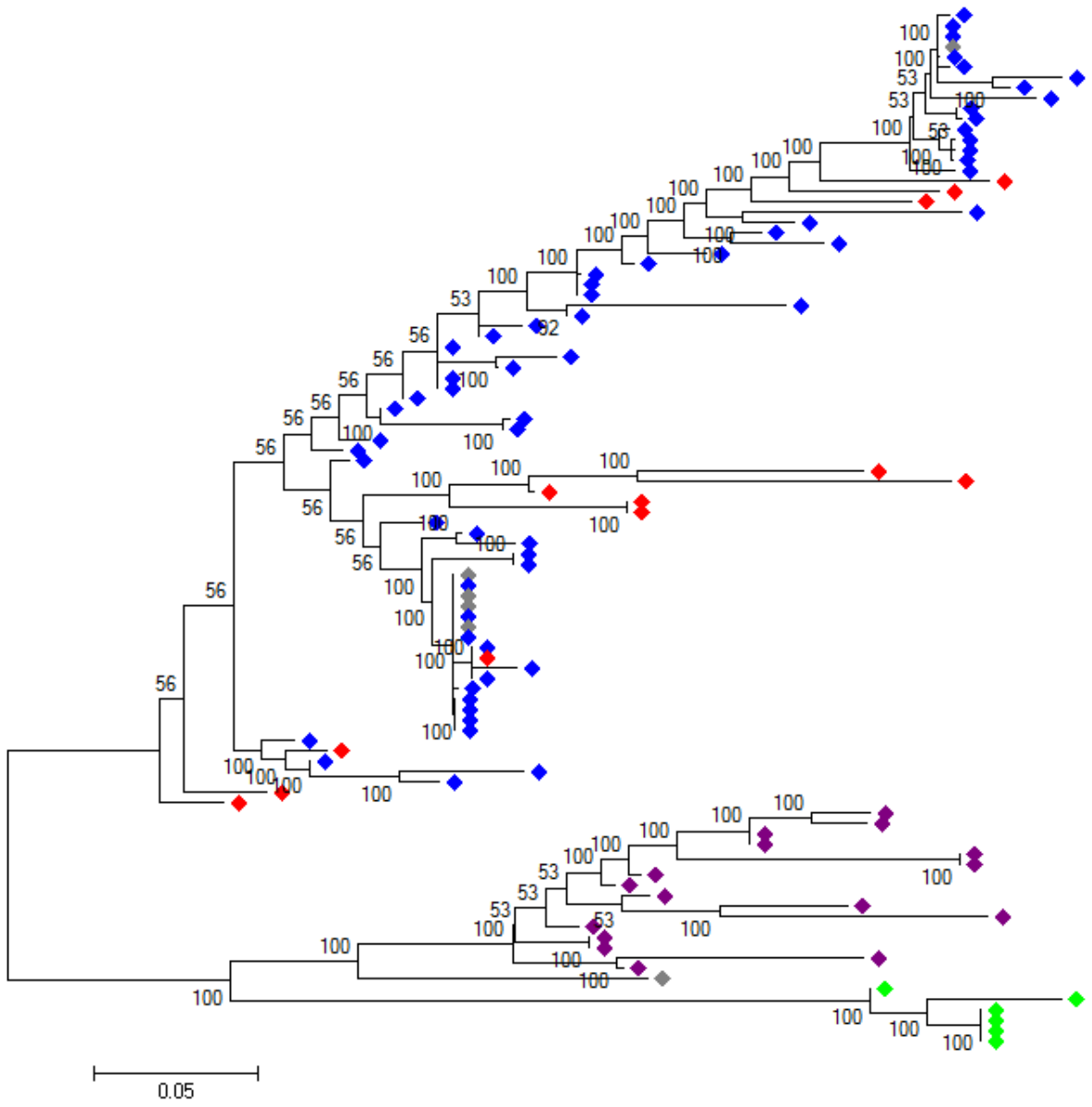


Figure S7.194 Chromosome 4: 44785001 - 44820001

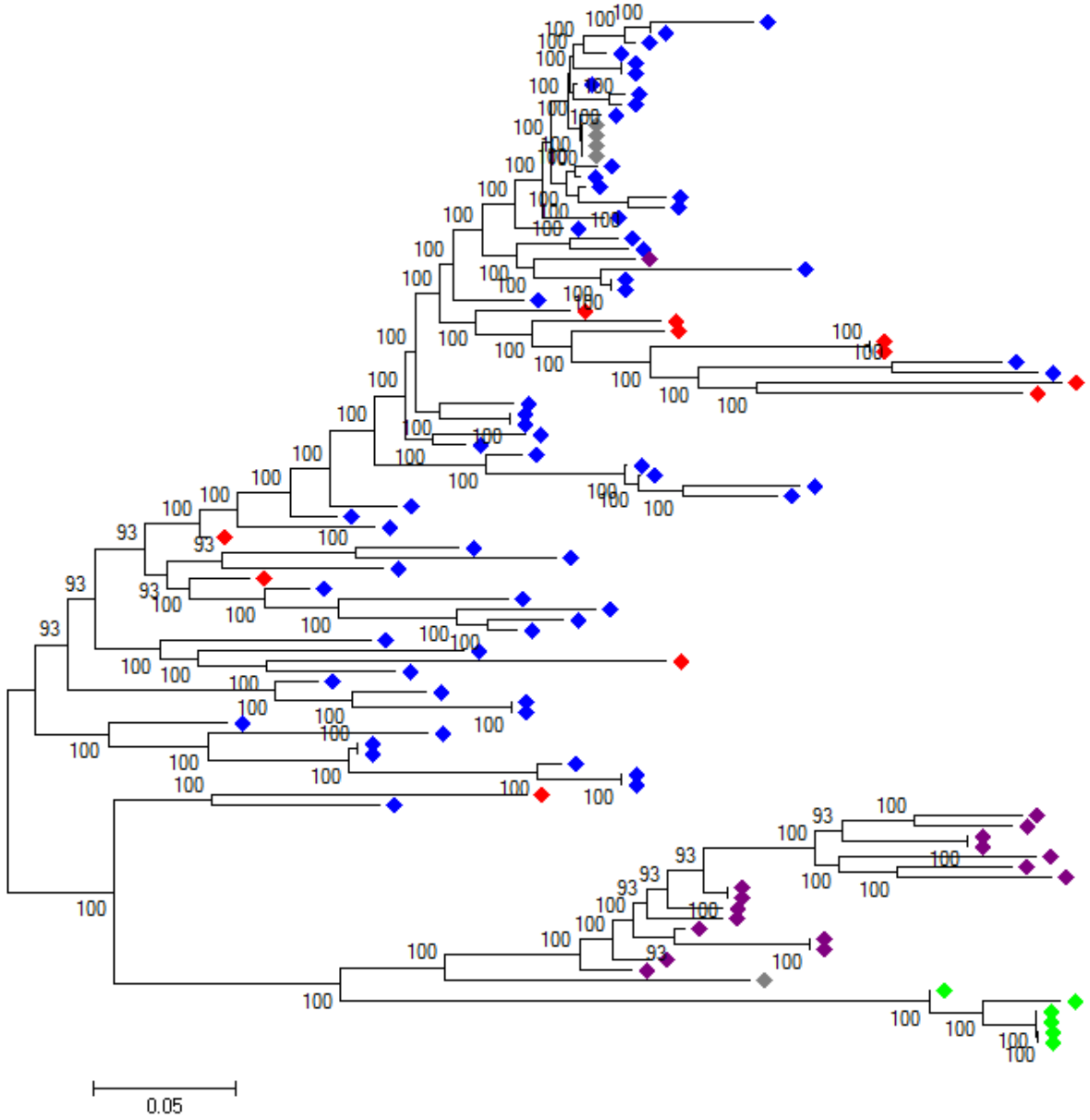


Figure S7.195 Chromosome 4: 45150001 - 45190001

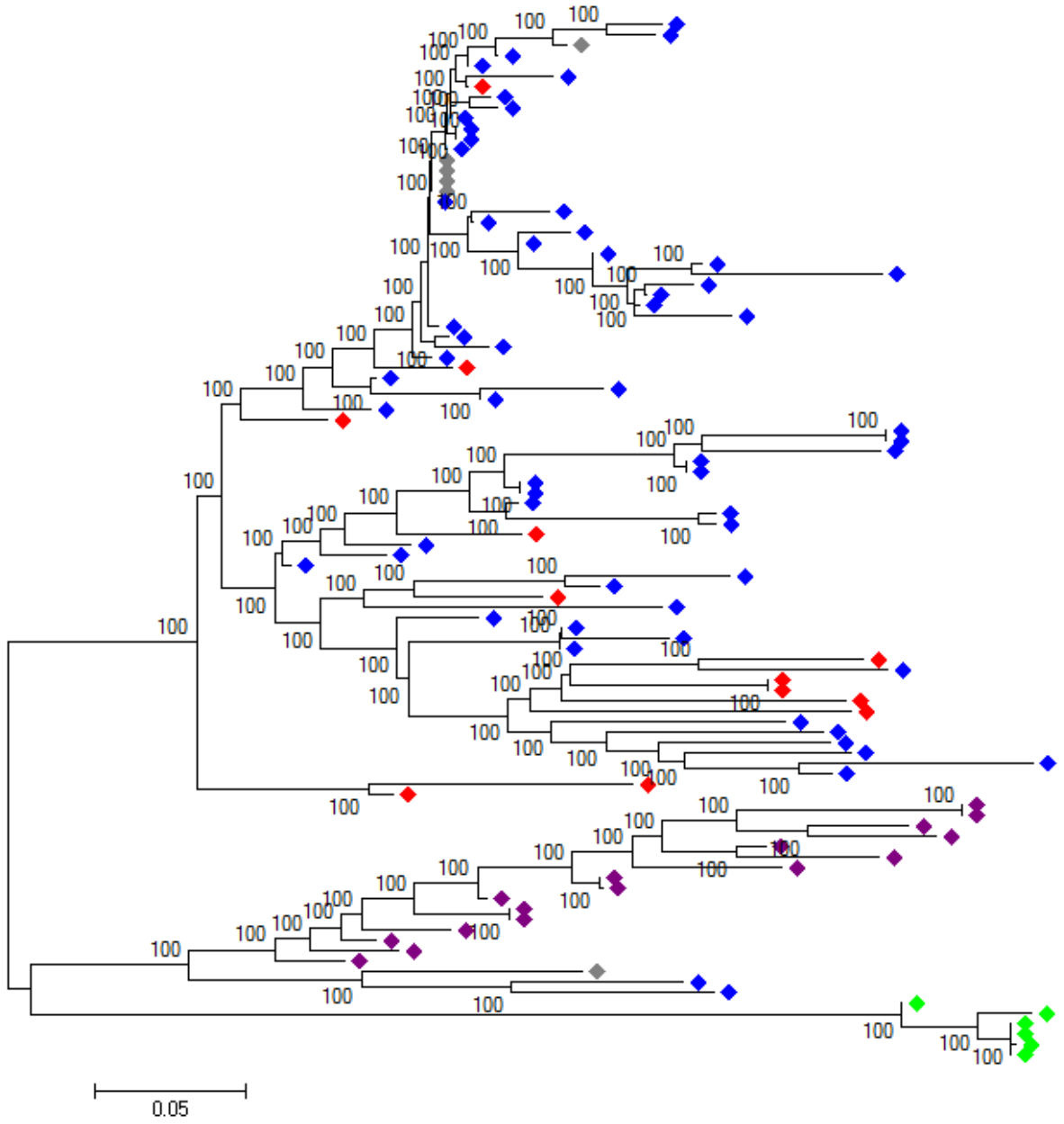


Figure S7.196 Chromosome 4: 45220001 - 45255001

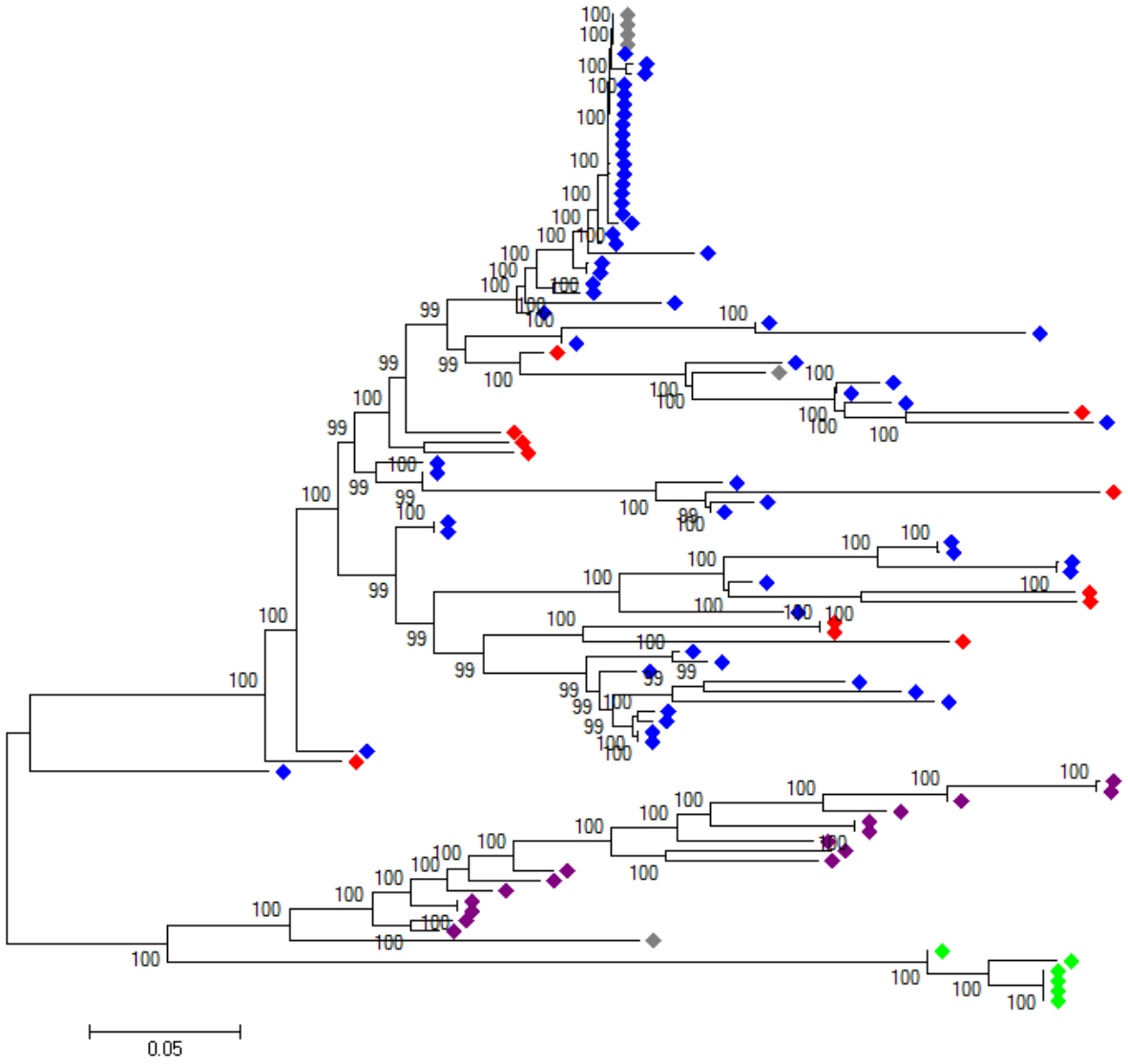


Figure S7.197 Chromosome 4: 45275001 - 45330001

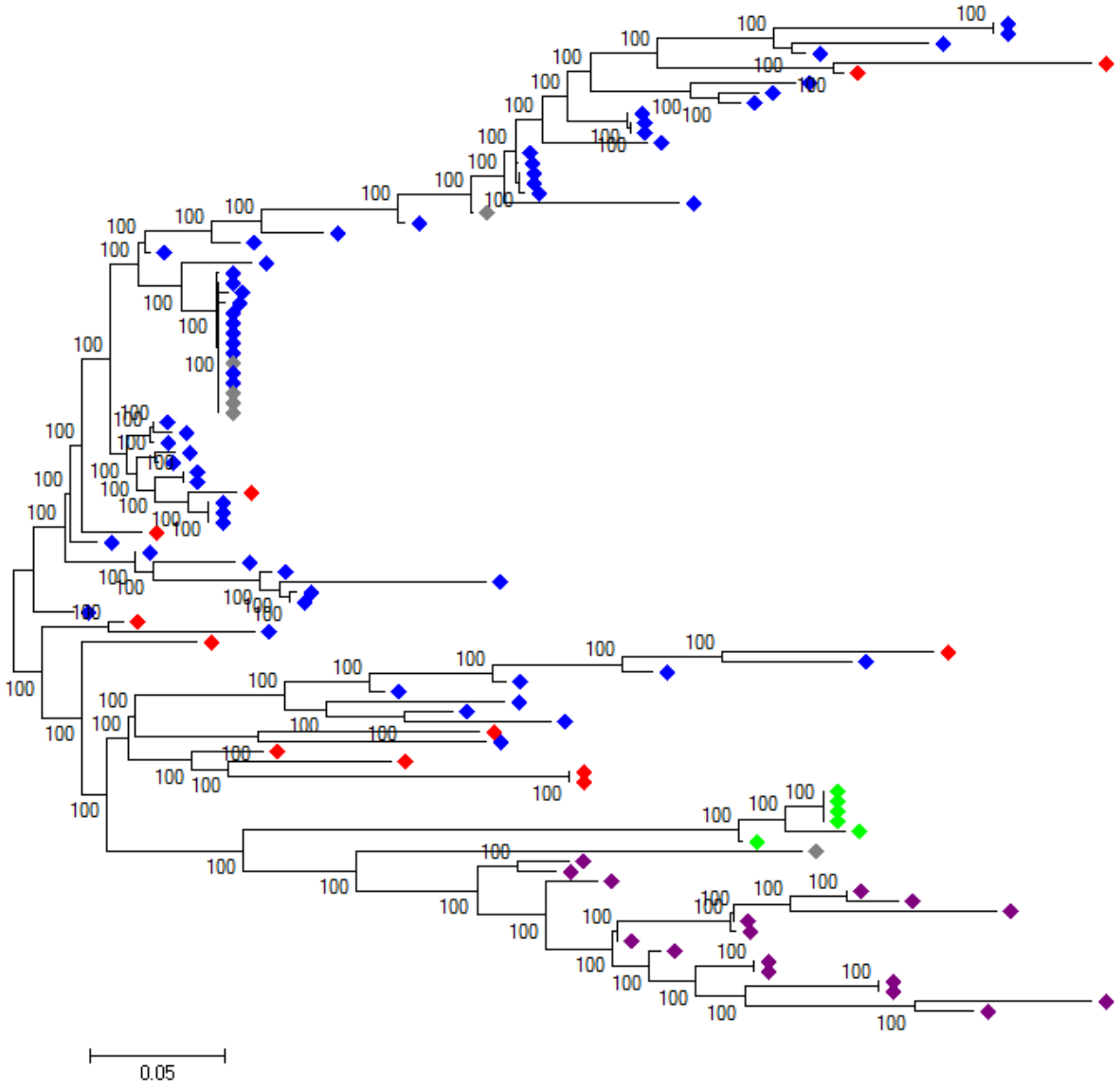


Figure S7.198 Chromosome 4: 45365001 - 45390001

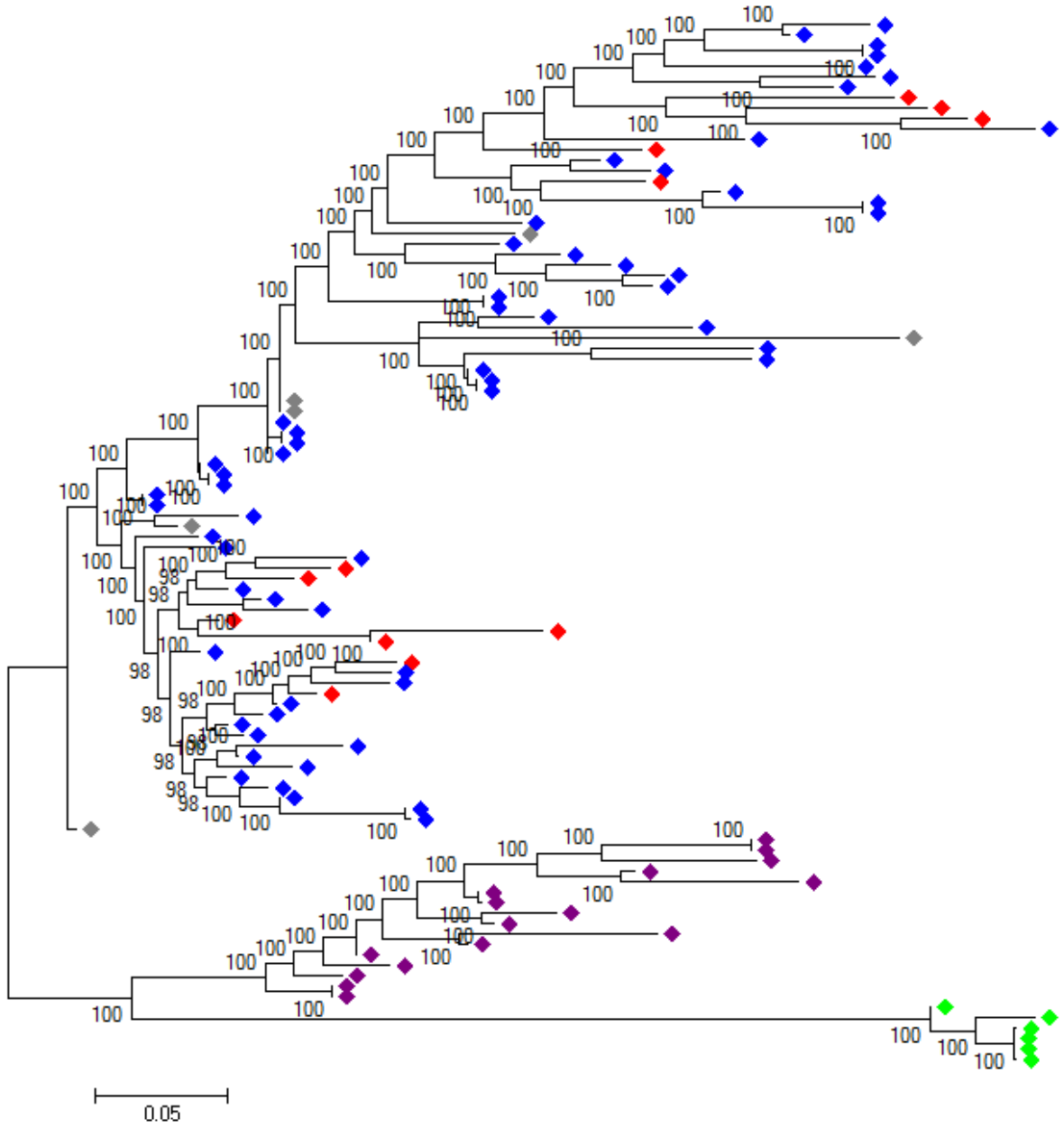


Figure S7.199 Chromosome 4: 46220001 - 46240001

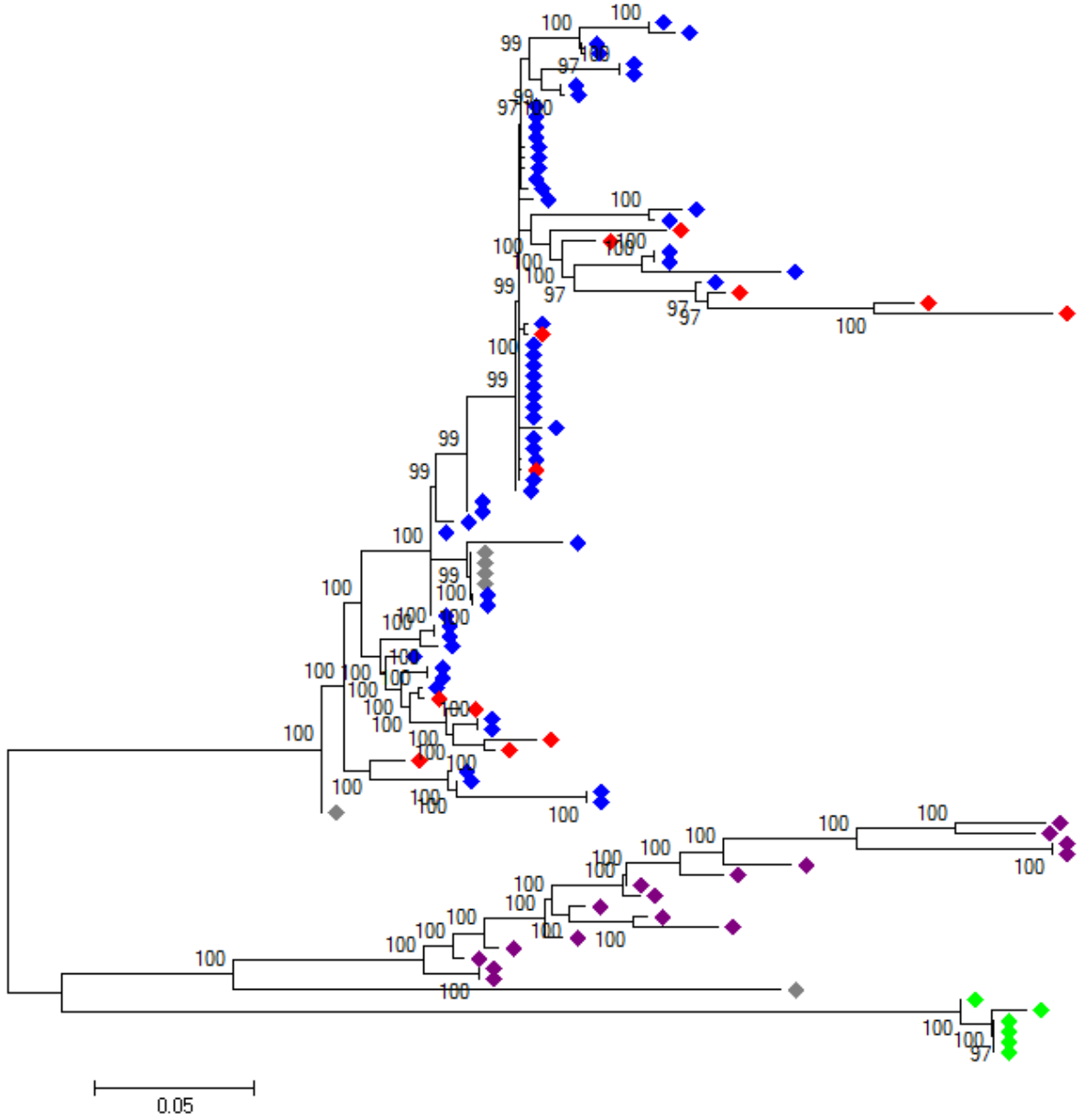


Figure S7.200 Chromosome 4: 46860001 - 46880001

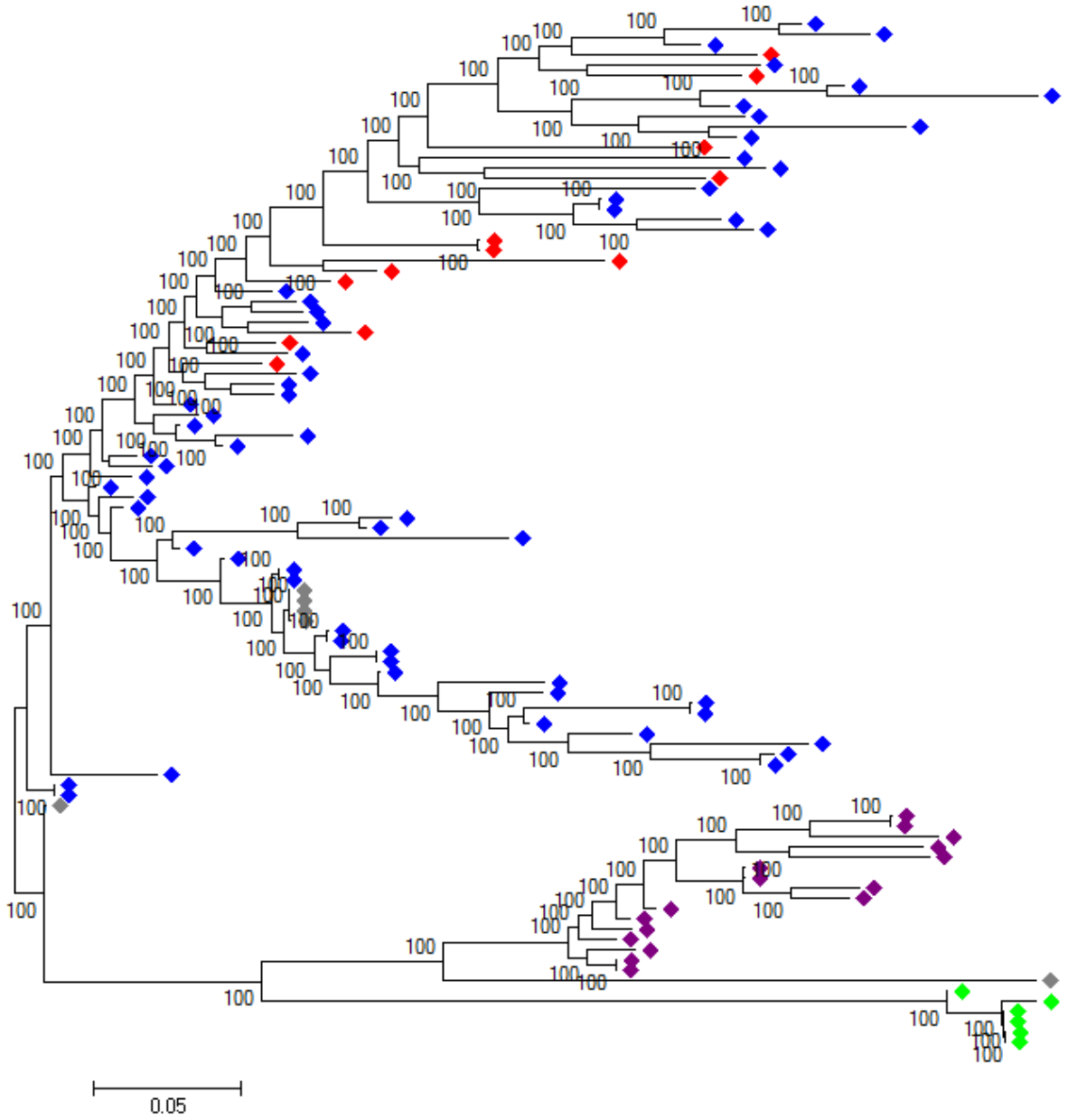


Figure S7.201 Chromosome 4: 47375001 - 47500001

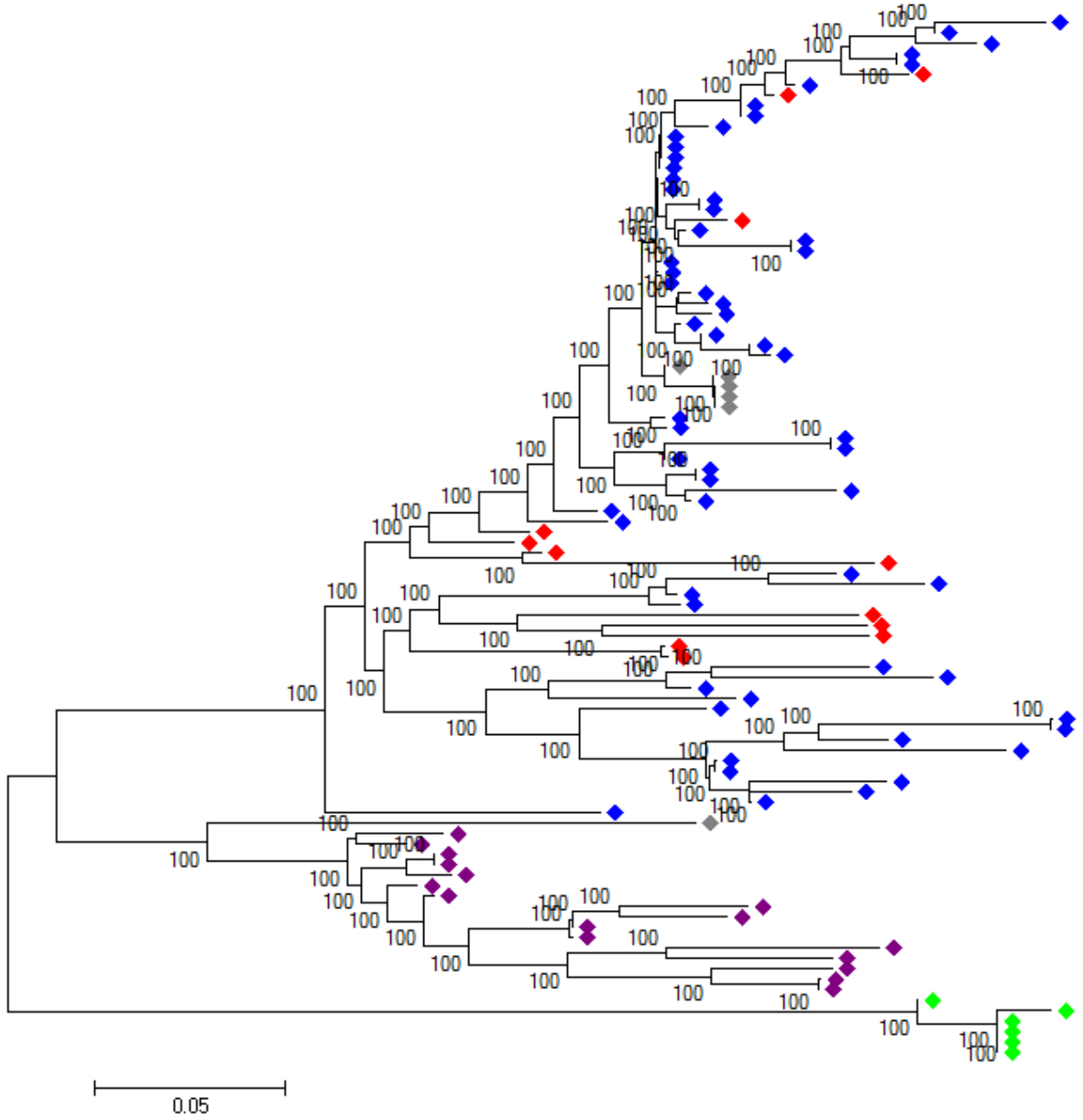


Figure S7.202 Chromosome 4: 47525001 - 47615001

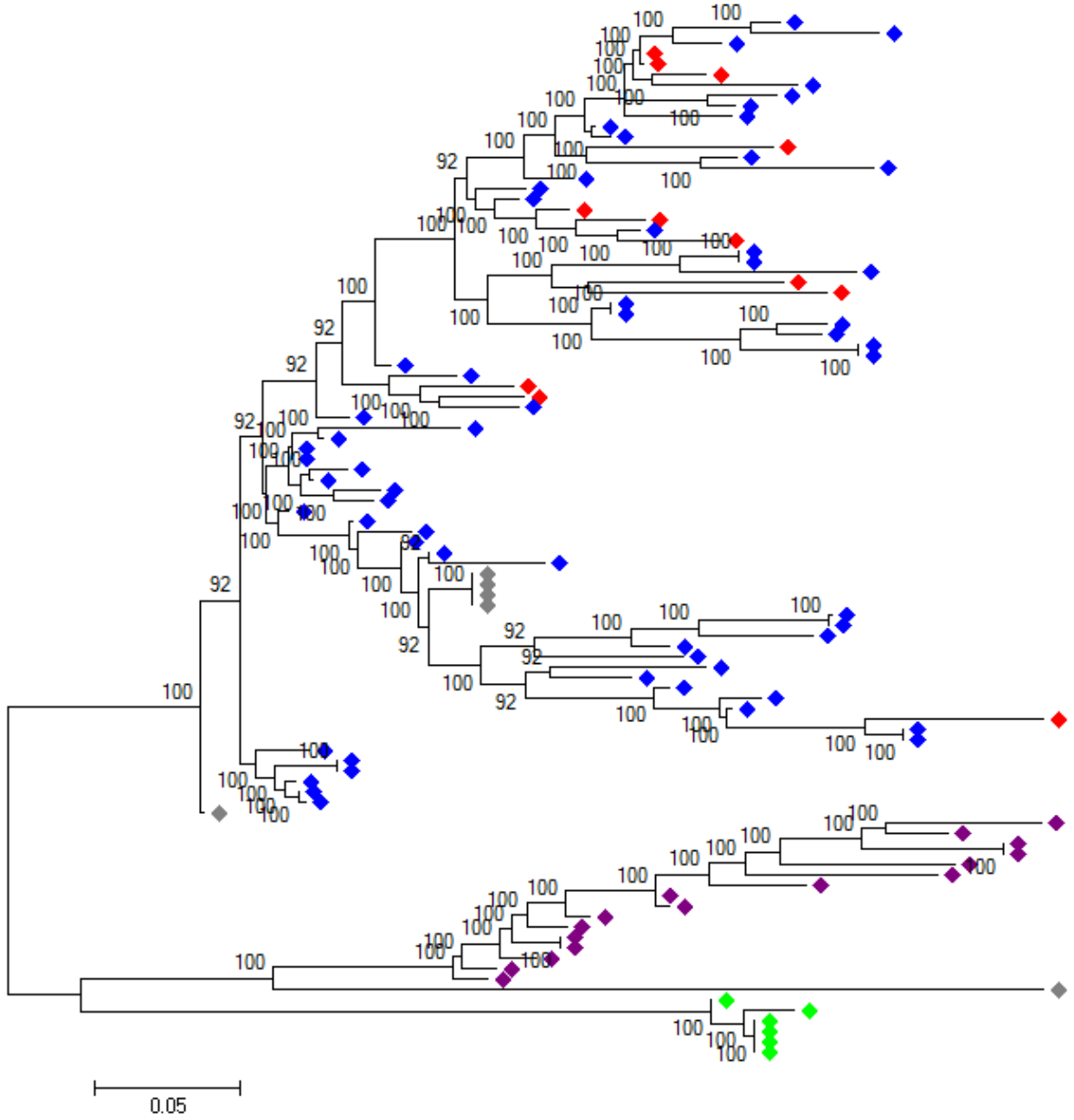


Figure S7.203 Chromosome 4: 47780001 - 47805001

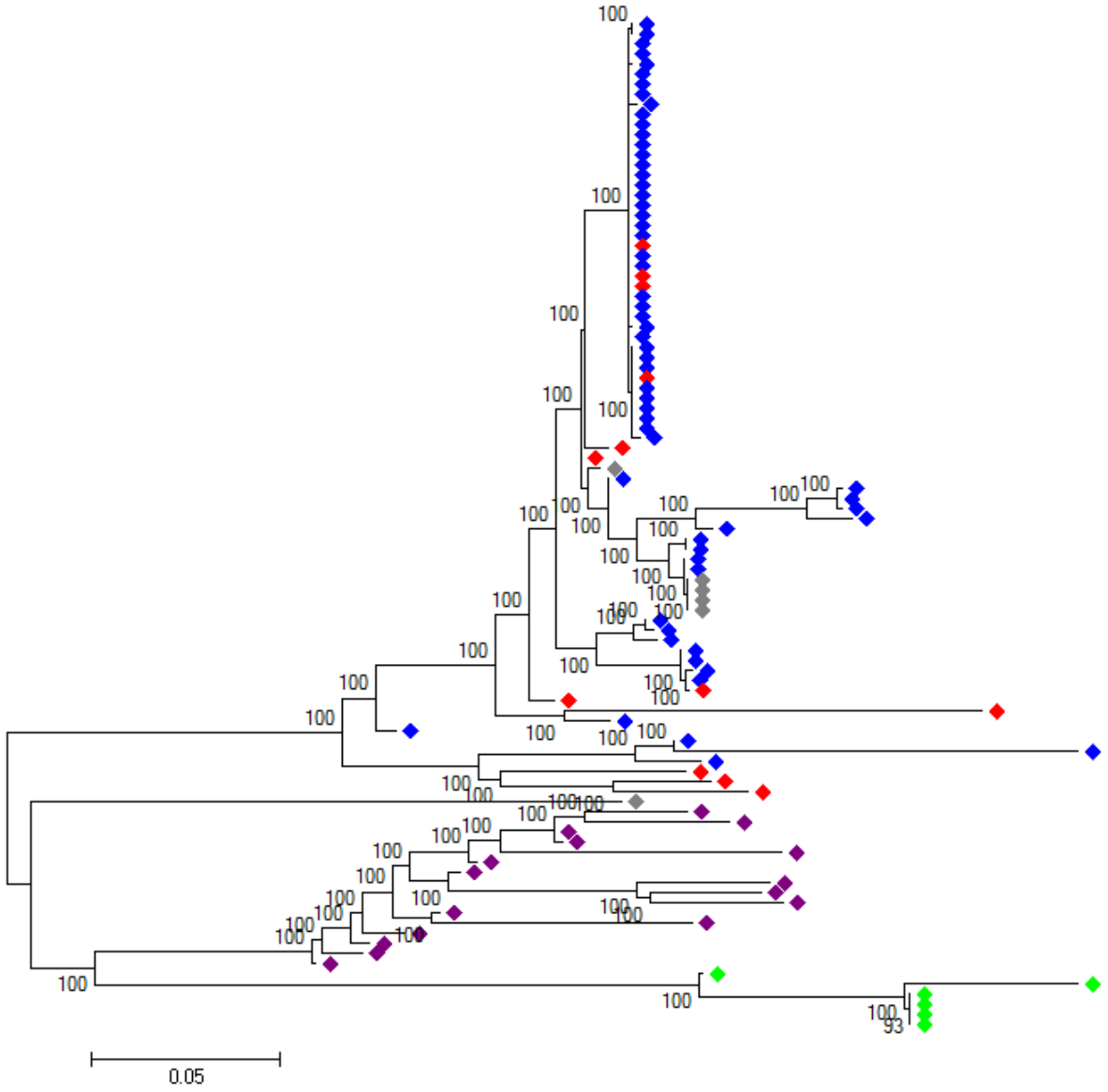


Figure S7.204 Chromosome 4: 48010001 - 48030001

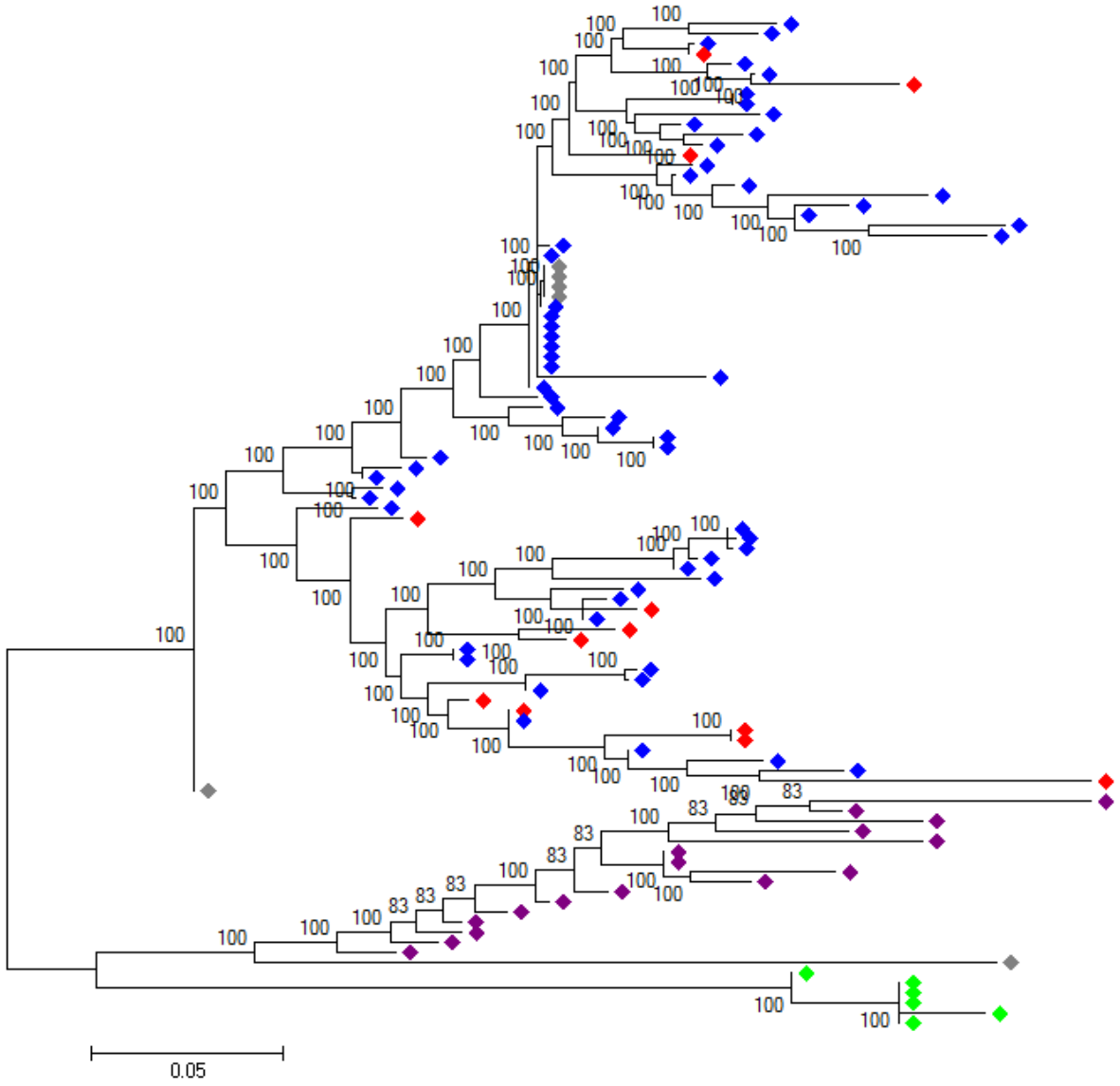


Figure S7.205 Chromosome 4: 48060001 - 48085001

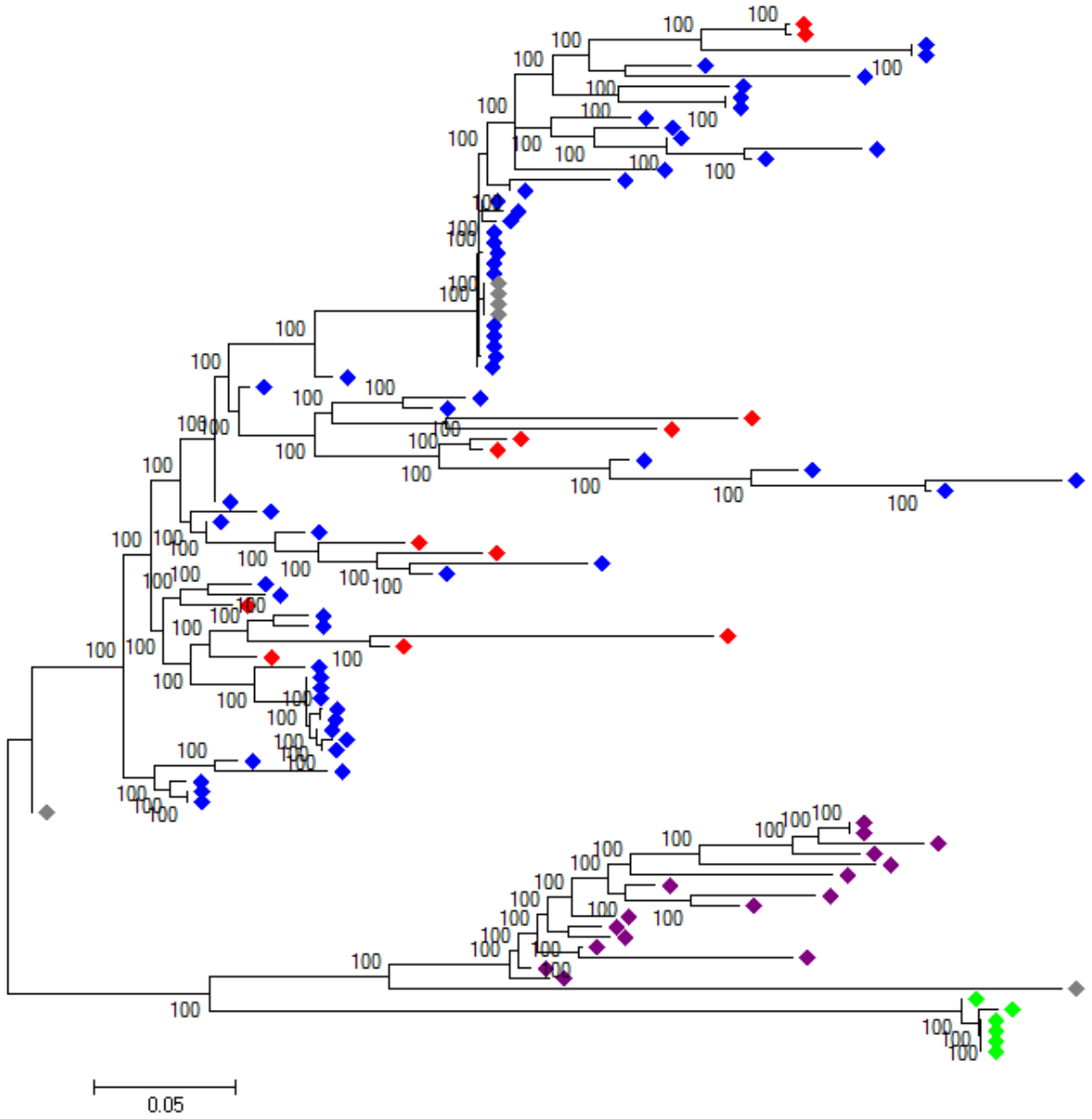


Figure S7.206 Chromosome 4: 48190001 - 48235001

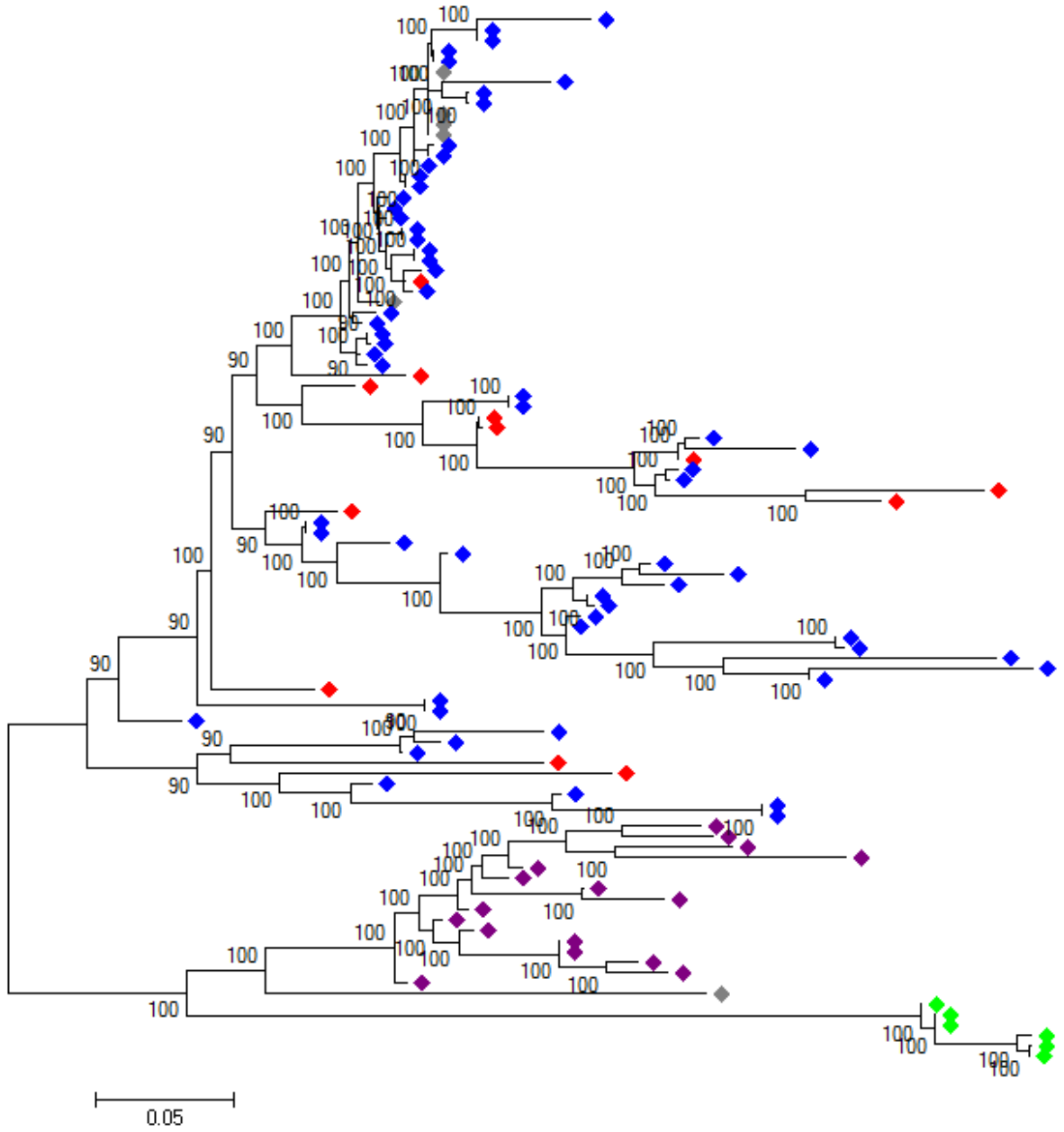


Figure S7.207 Chromosome 4: 48295001 - 48320001

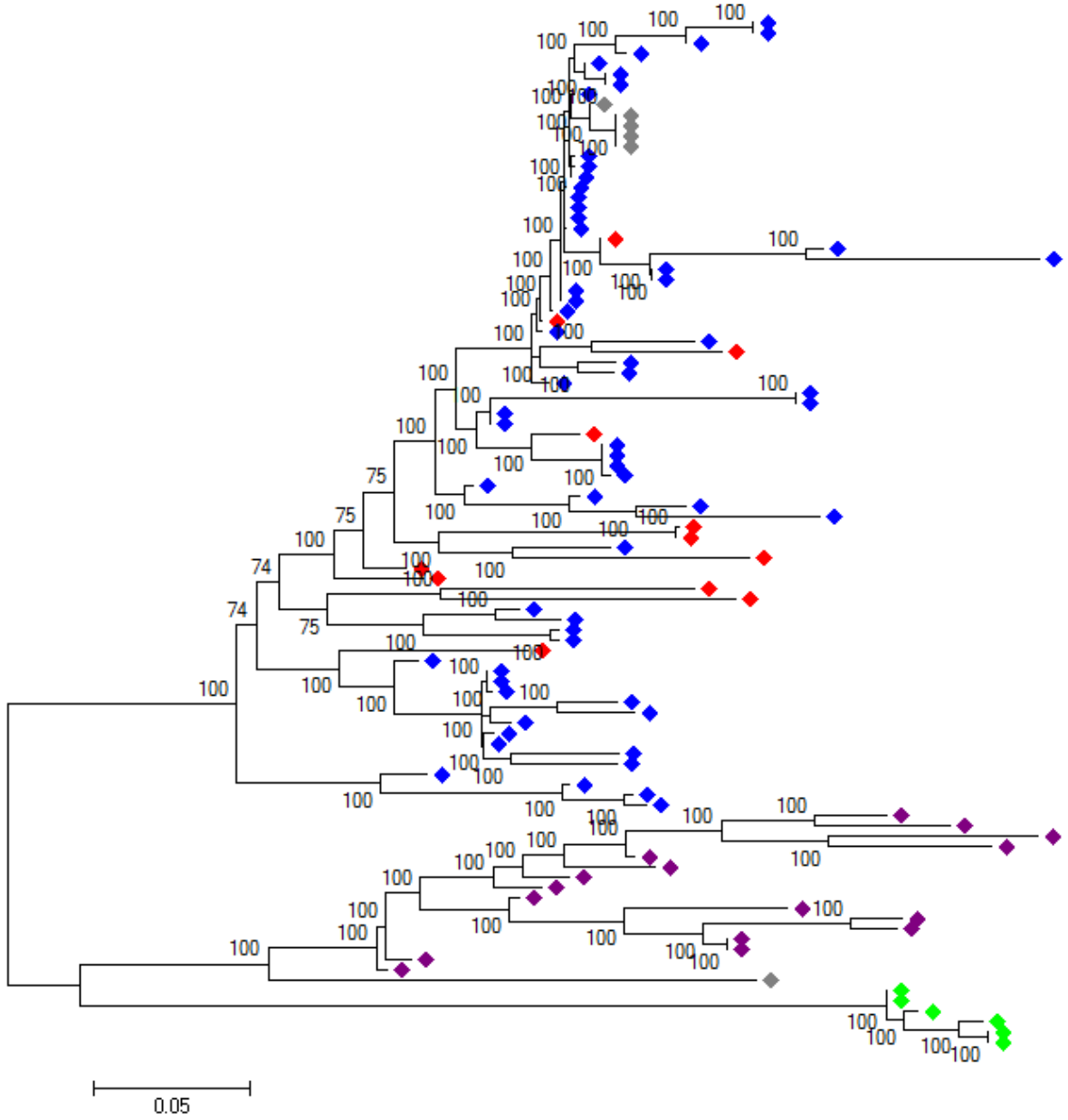


Figure S7.208 Chromosome 4: 48525001 - 48555001

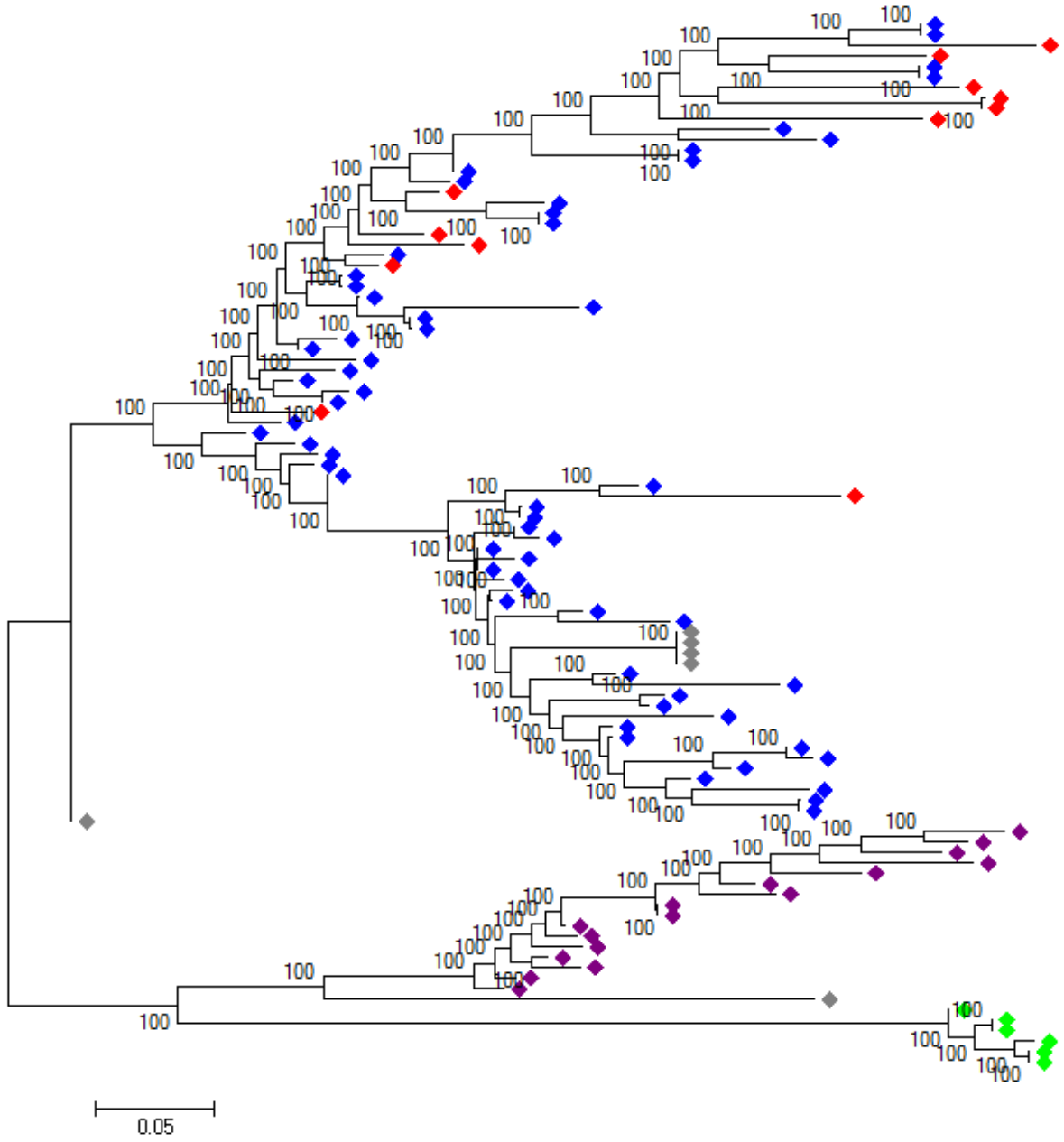


Figure S7.209 Chromosome 4: 48610001 - 48640001

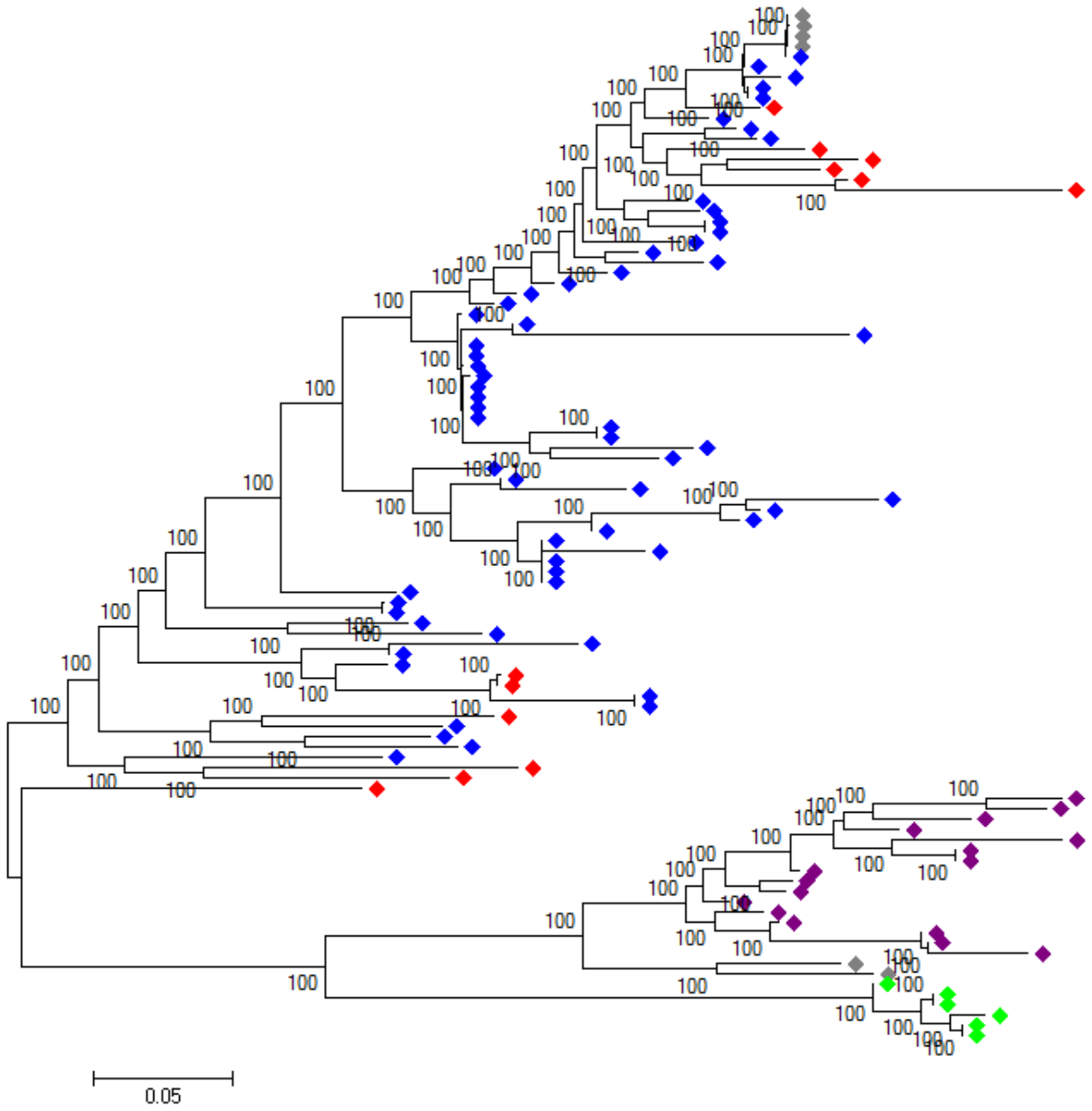


Figure S7.210 Chromosome 4: 49380001 - 49410001

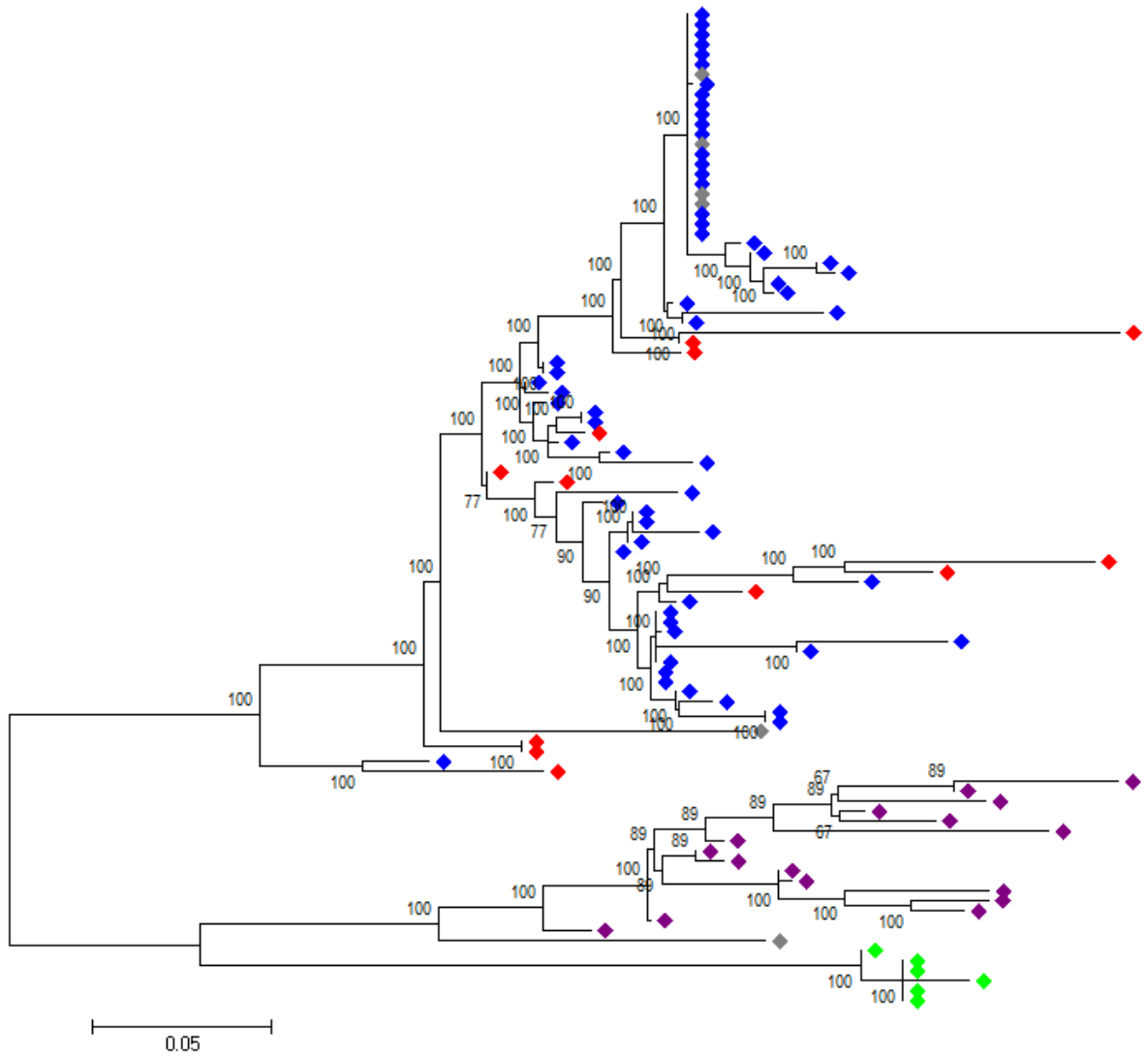


Figure S7.211 Chromosome 4: 49560001 - 49590001

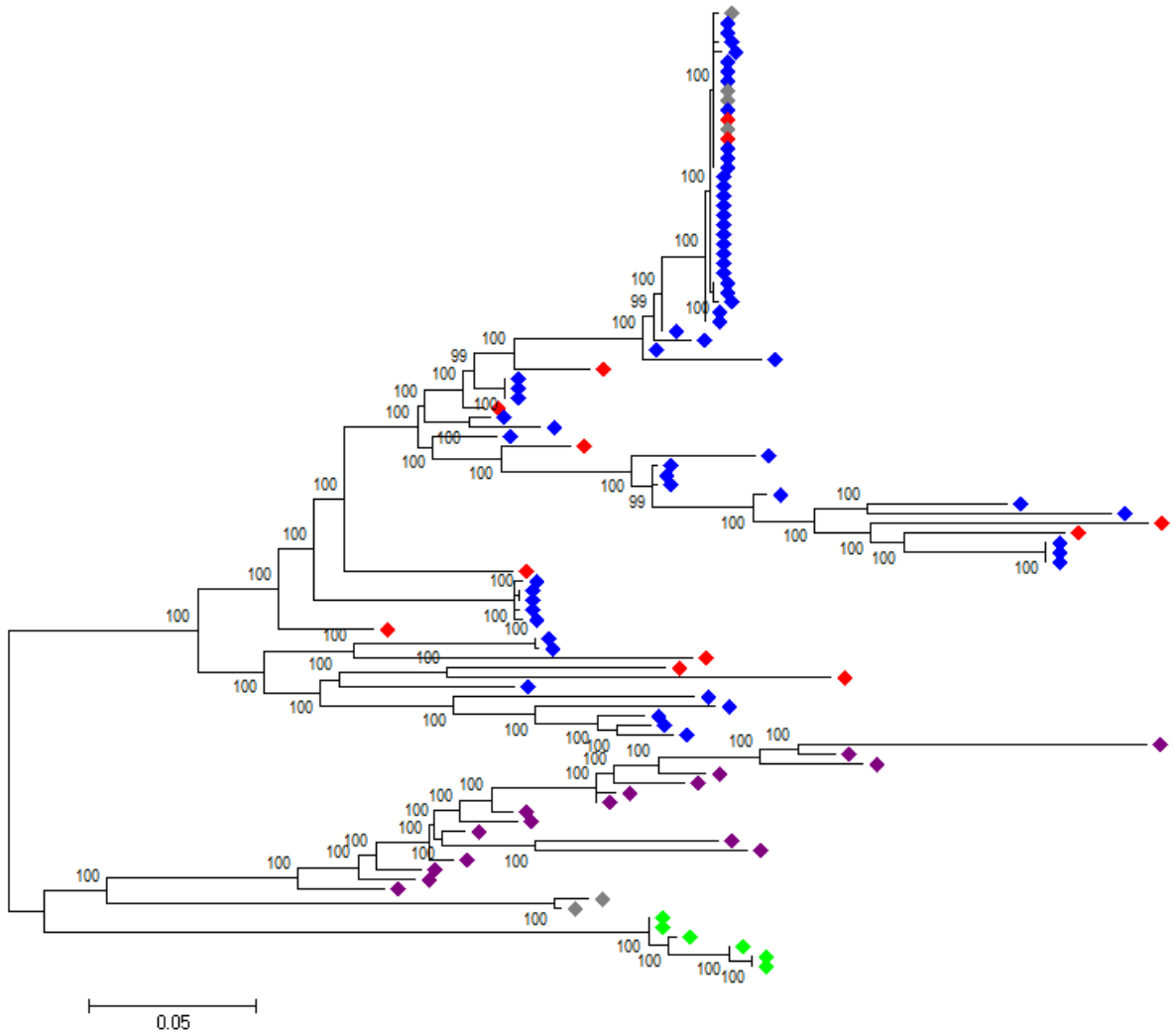


Figure S7.213 Chromosome 4: 49625001 - 49660001

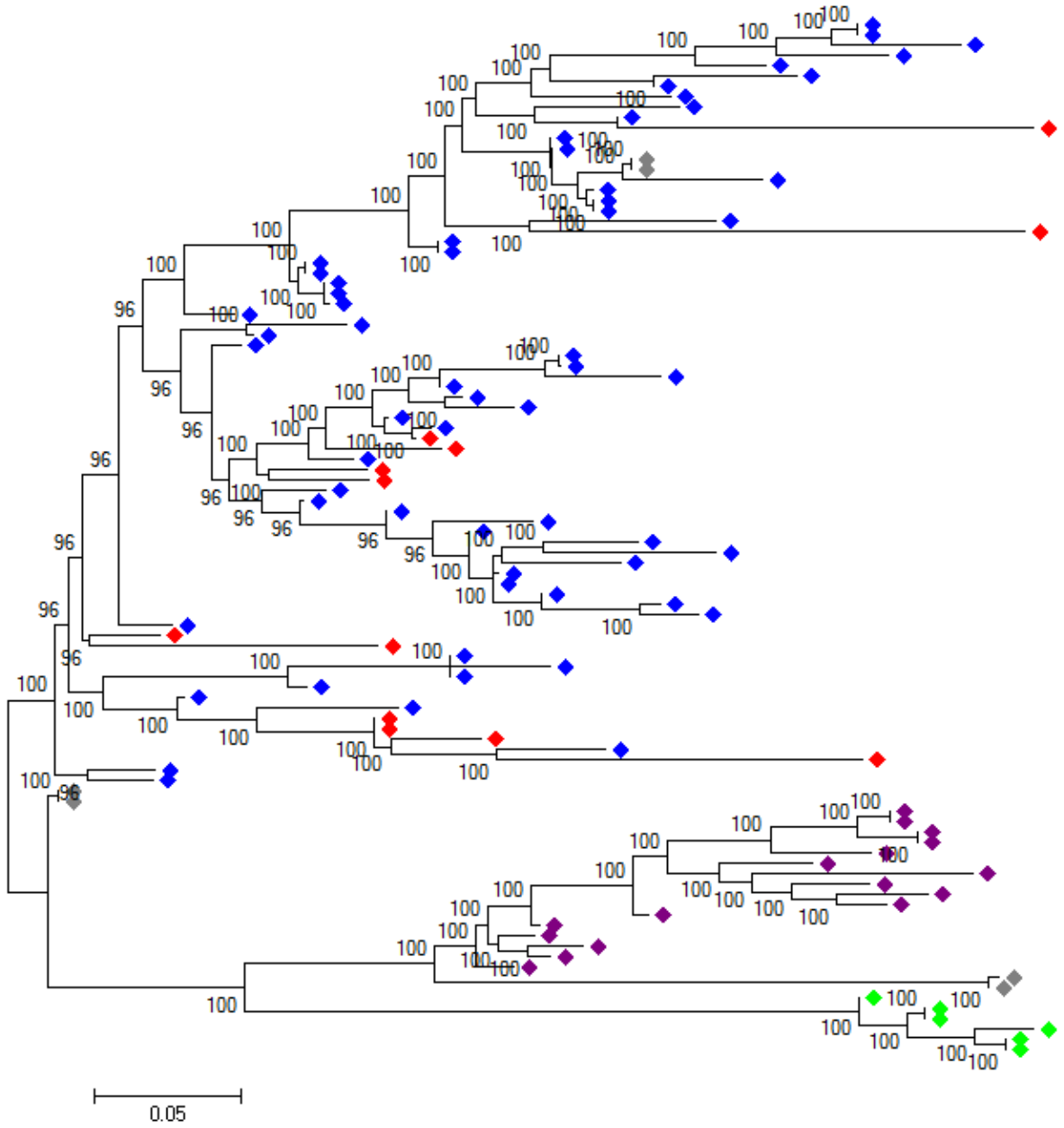


Figure S7.213 Chromosome 4: 50890001 - 50920001

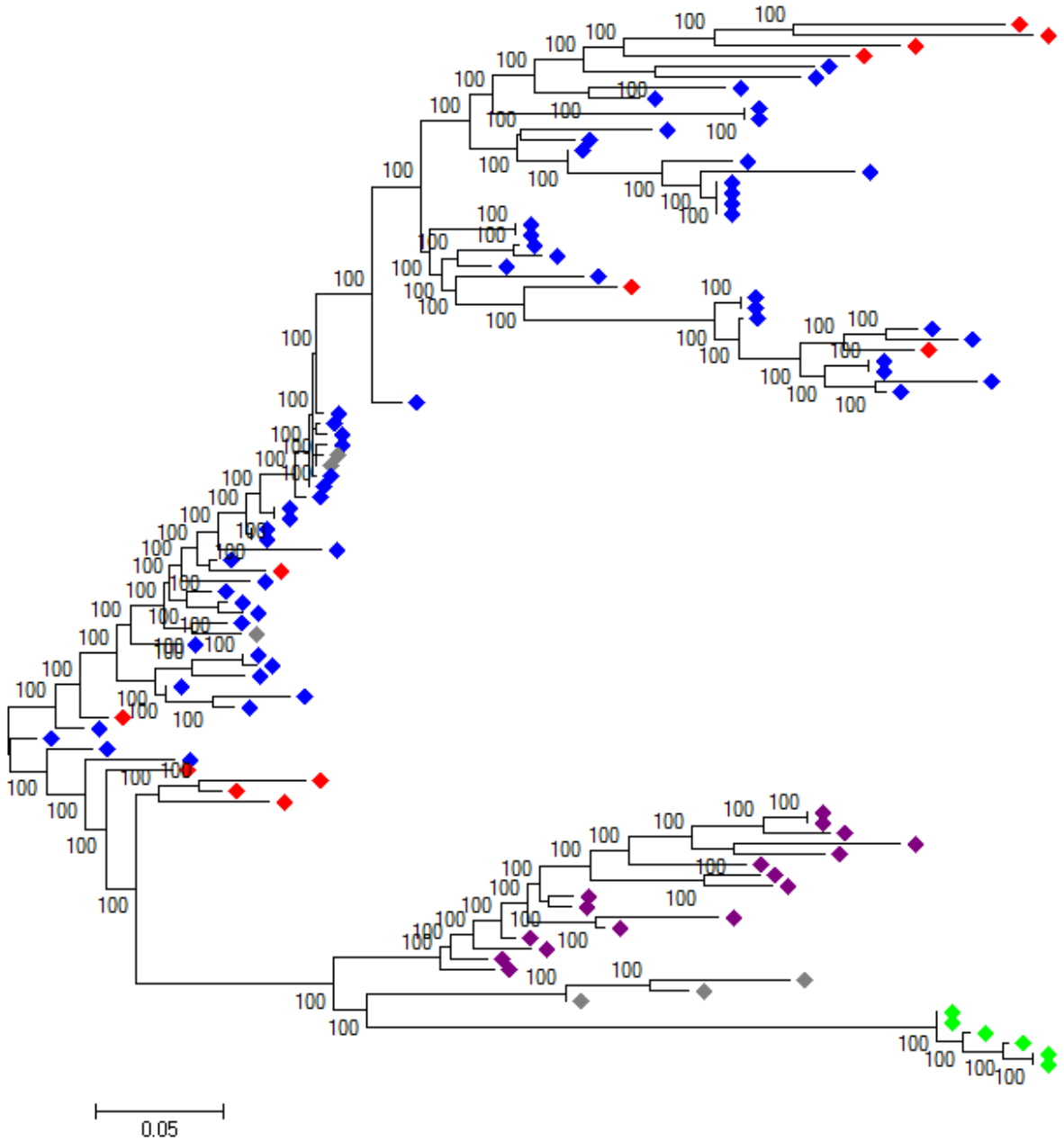


Figure S7.214 Chromosome 4: 75755001 - 75775001

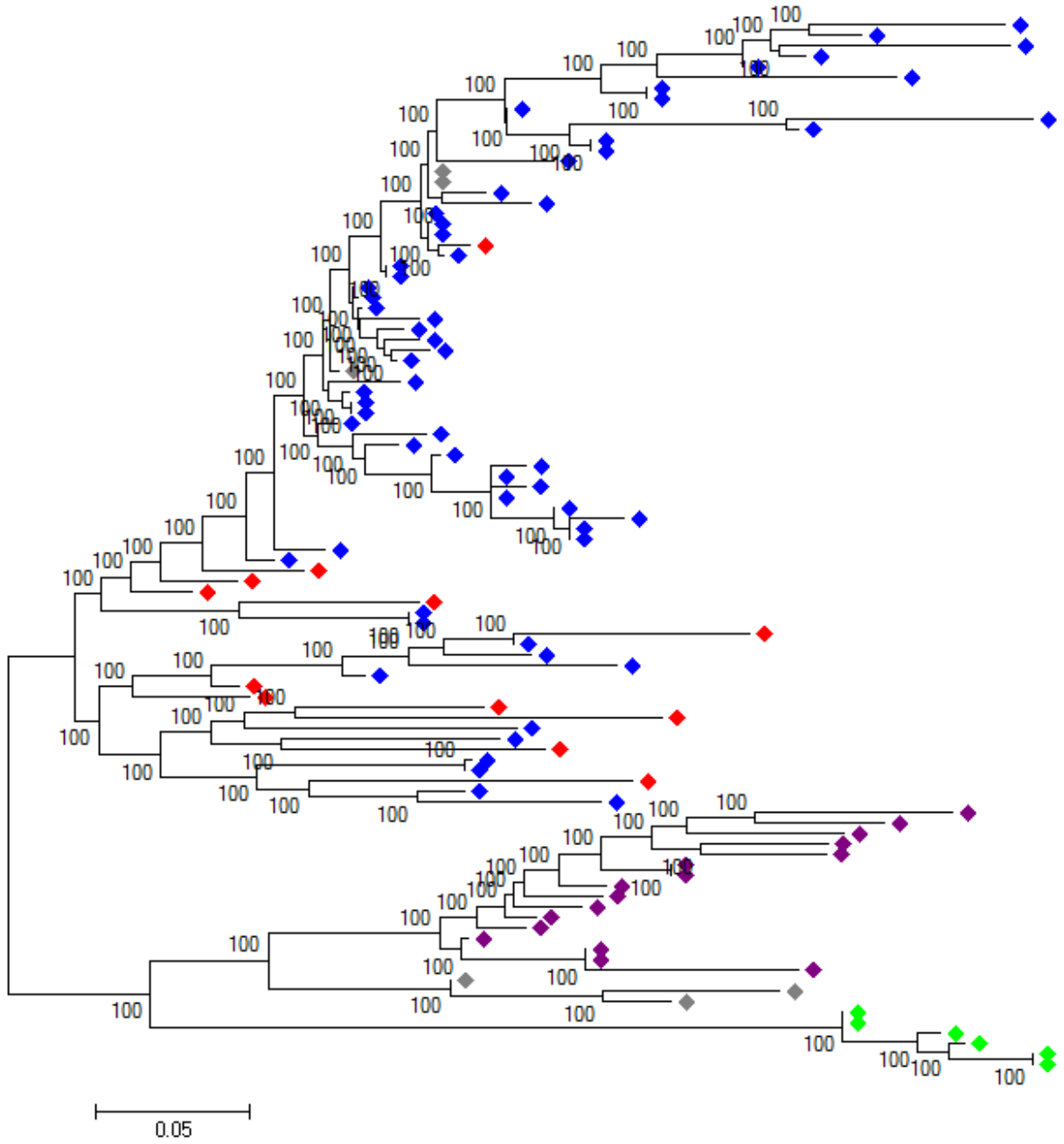


Figure S7.215 Chromosome 4: 75790001 - 75810001

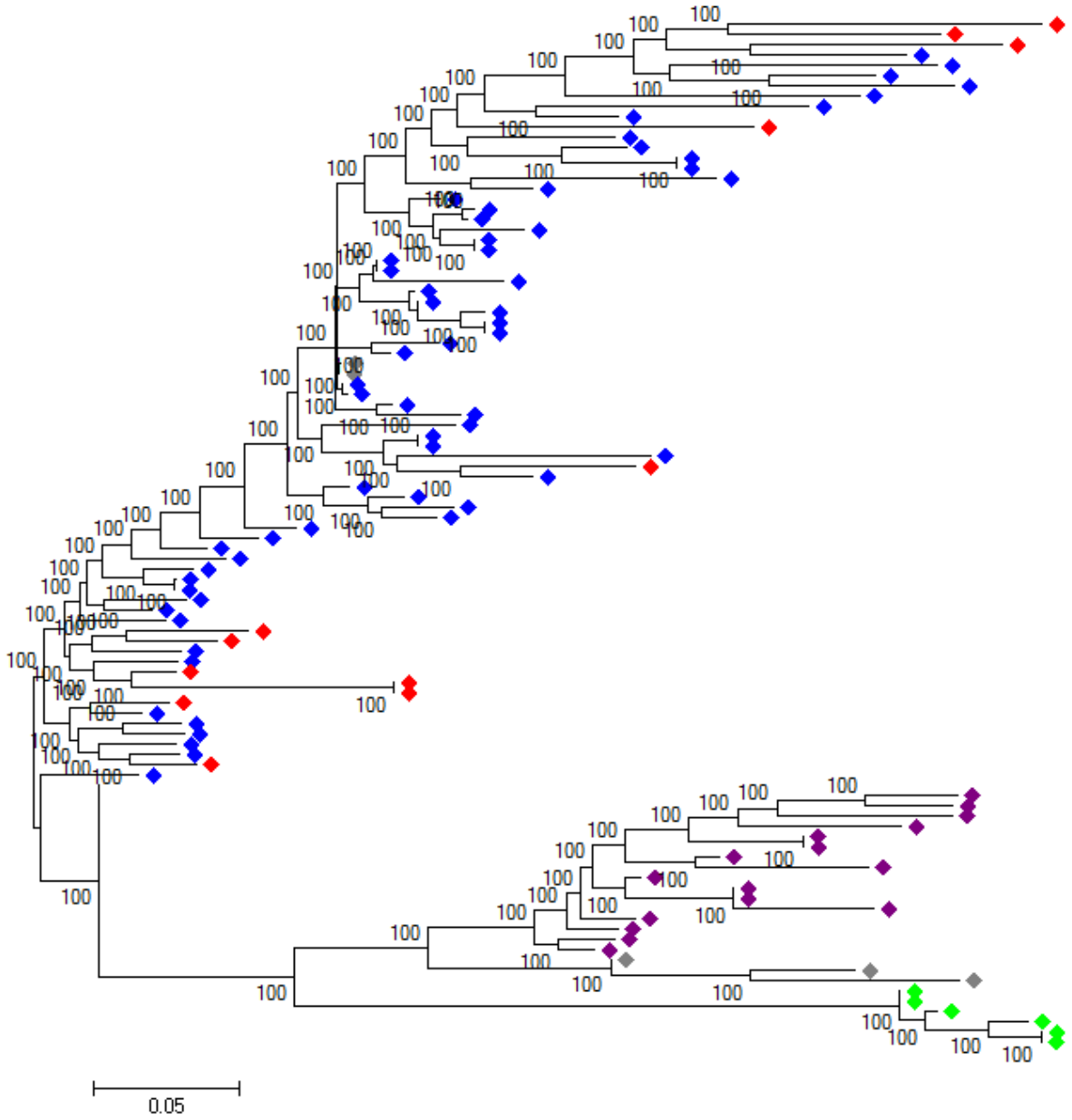


Figure S7.216 Chromosome 4: 75835001 - 75920001

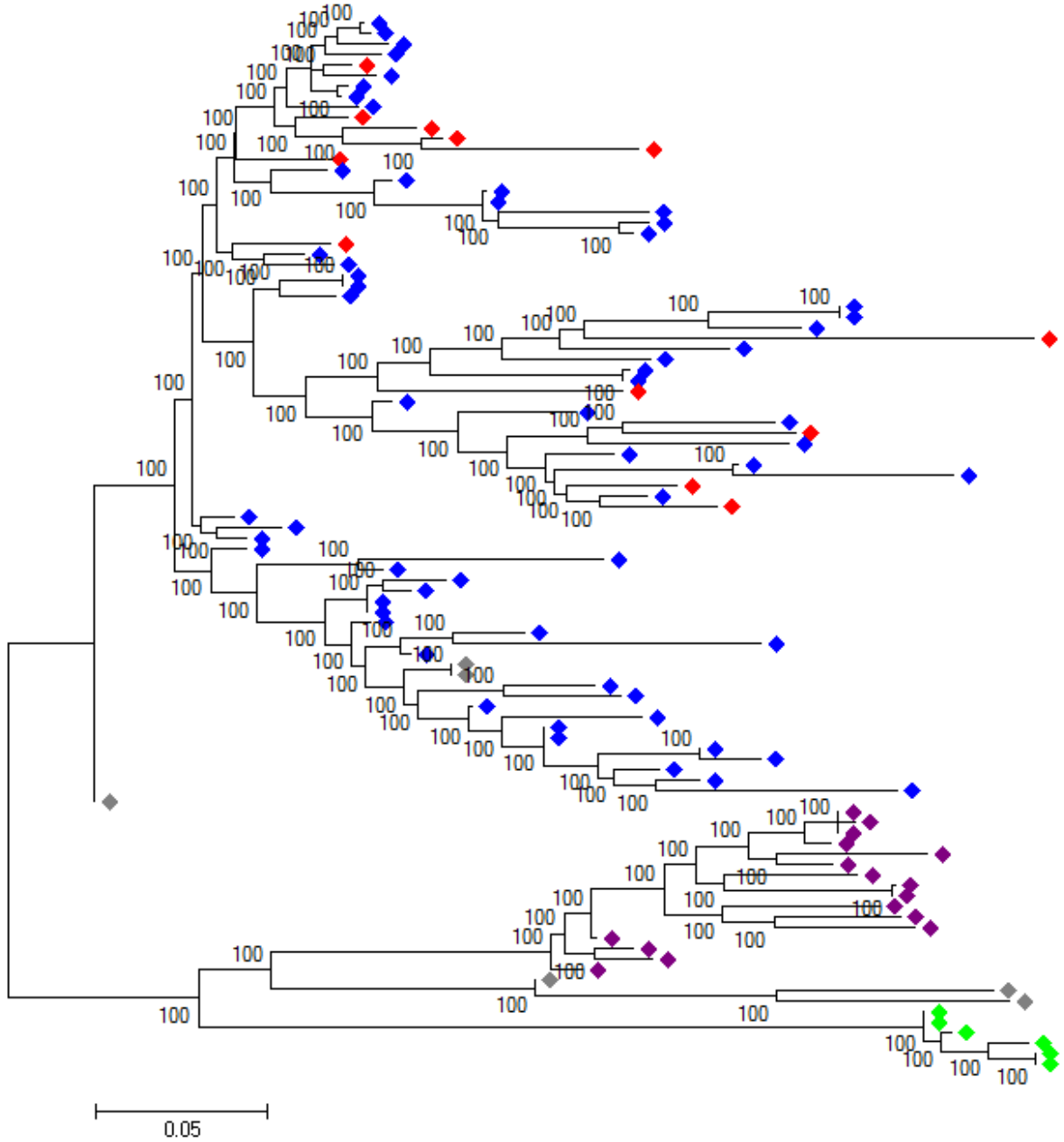


Figure S7.217 Chromosome 4: 76400001 - 76430001

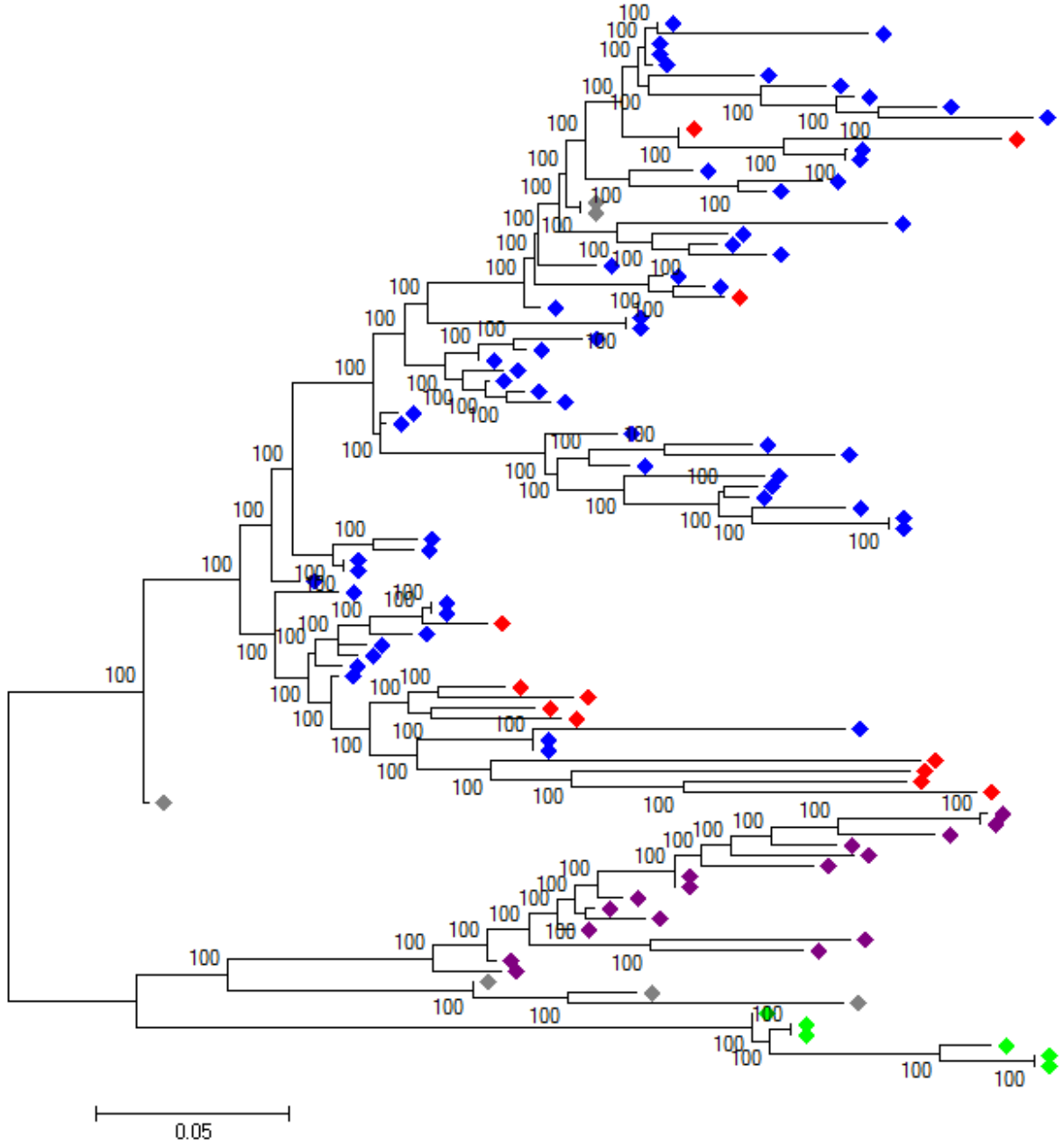


Figure S7.218 Chromosome 4: 76755001 - 76790001

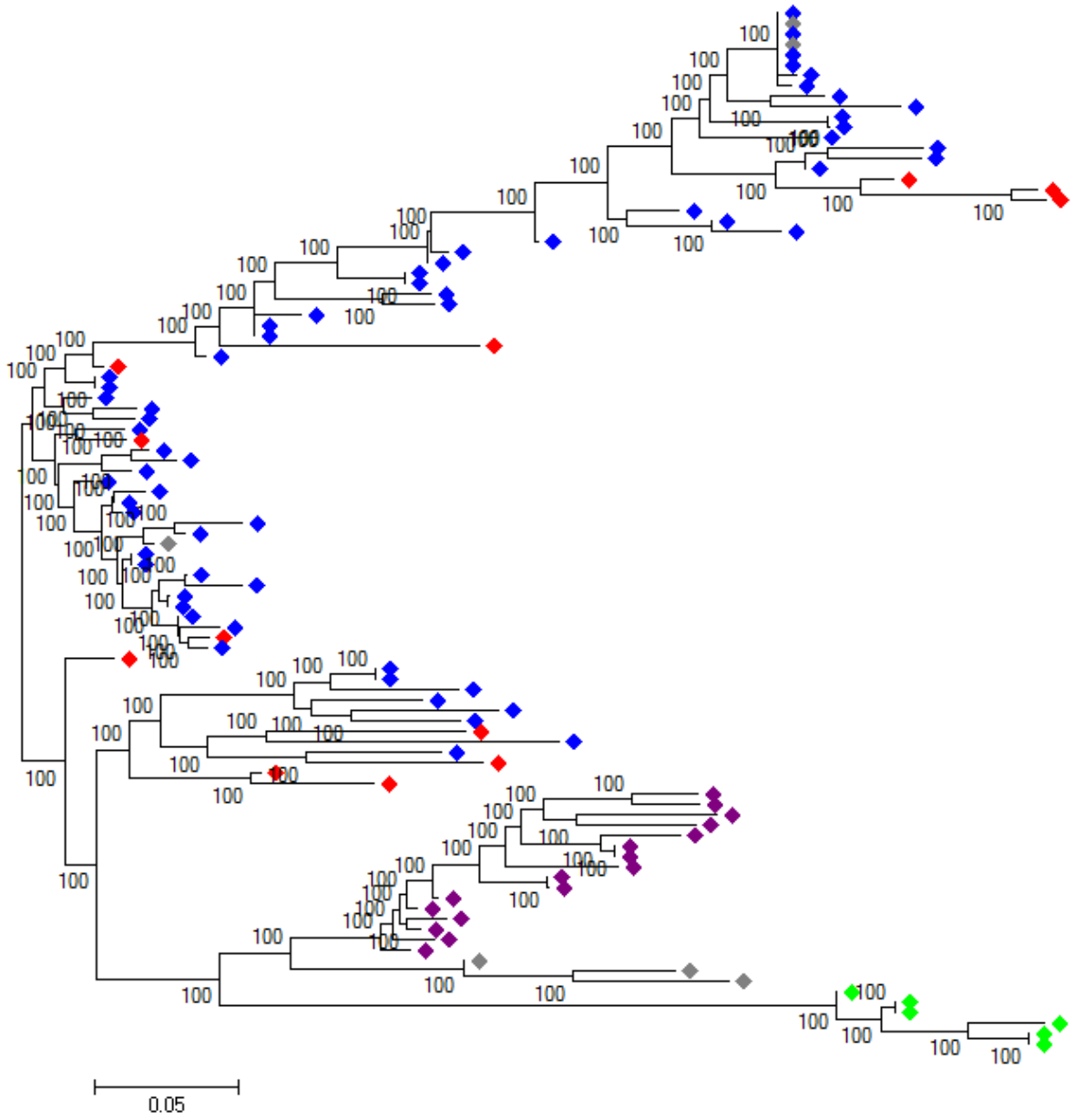


Figure S7.219 Chromosome 4: 76815001 - 76860001

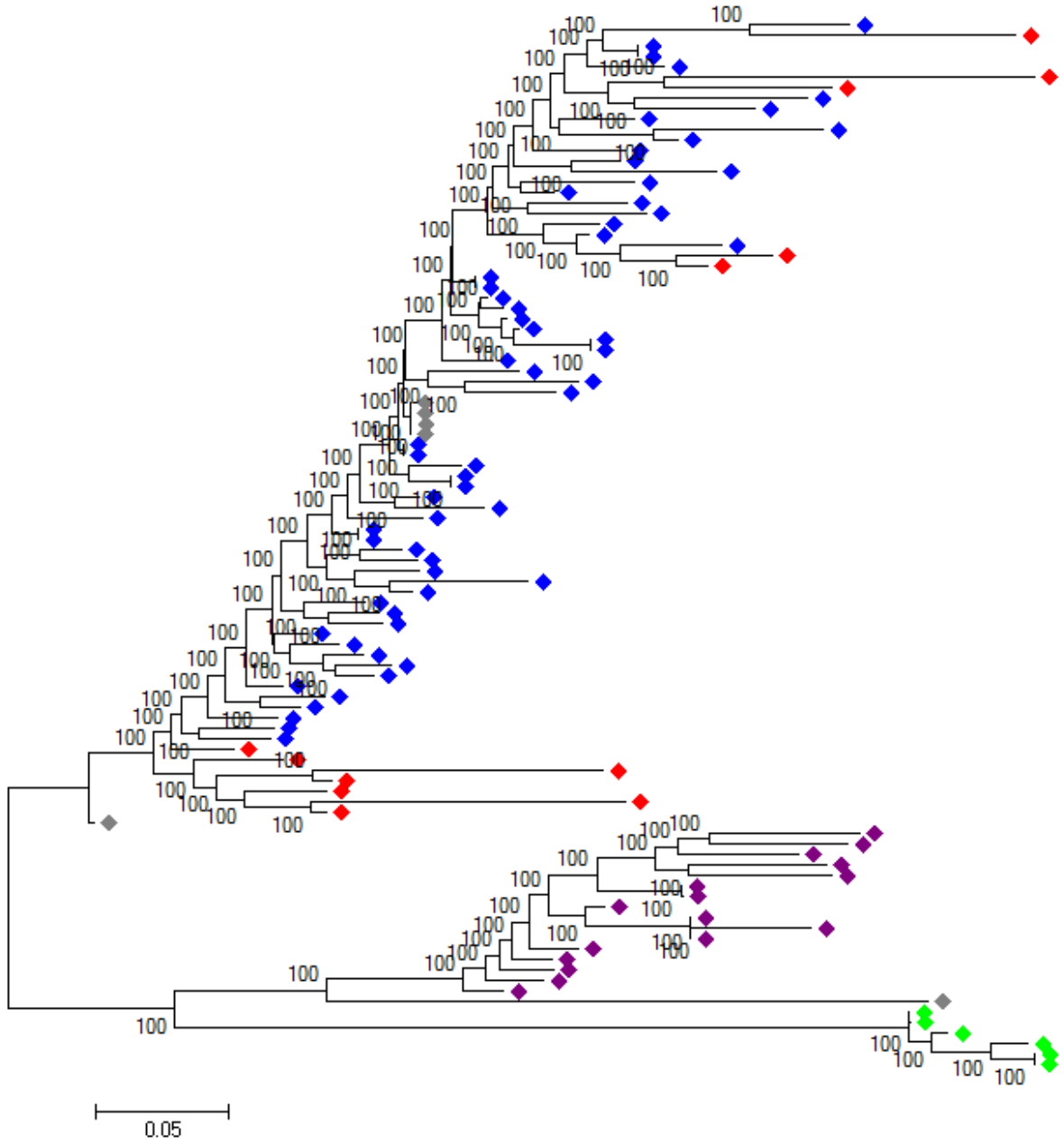


Figure S7.220 Chromosome 4: 76905001 - 77285001

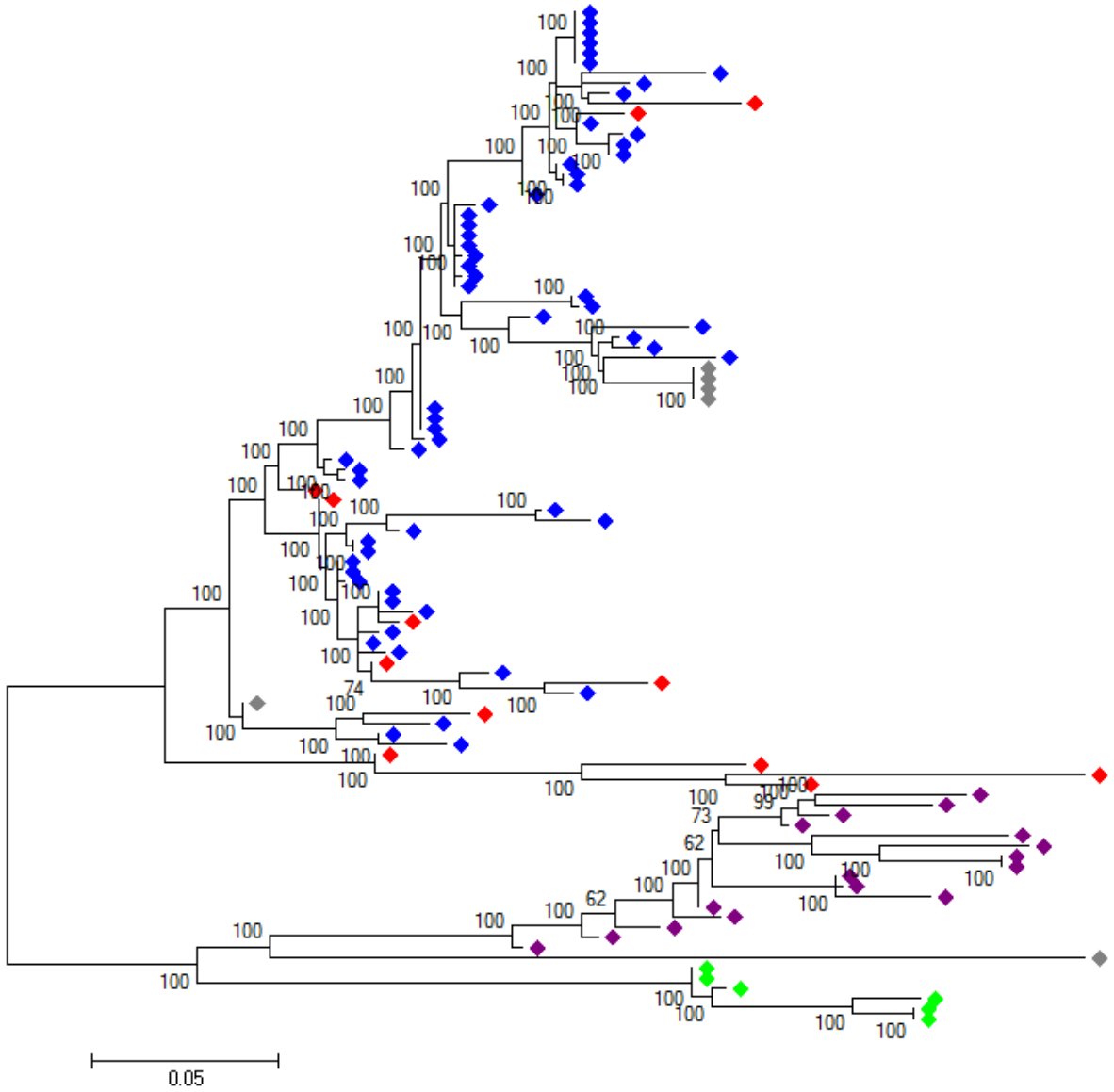


Figure S7.221 Chromosome 4: 77355001 - 77375001

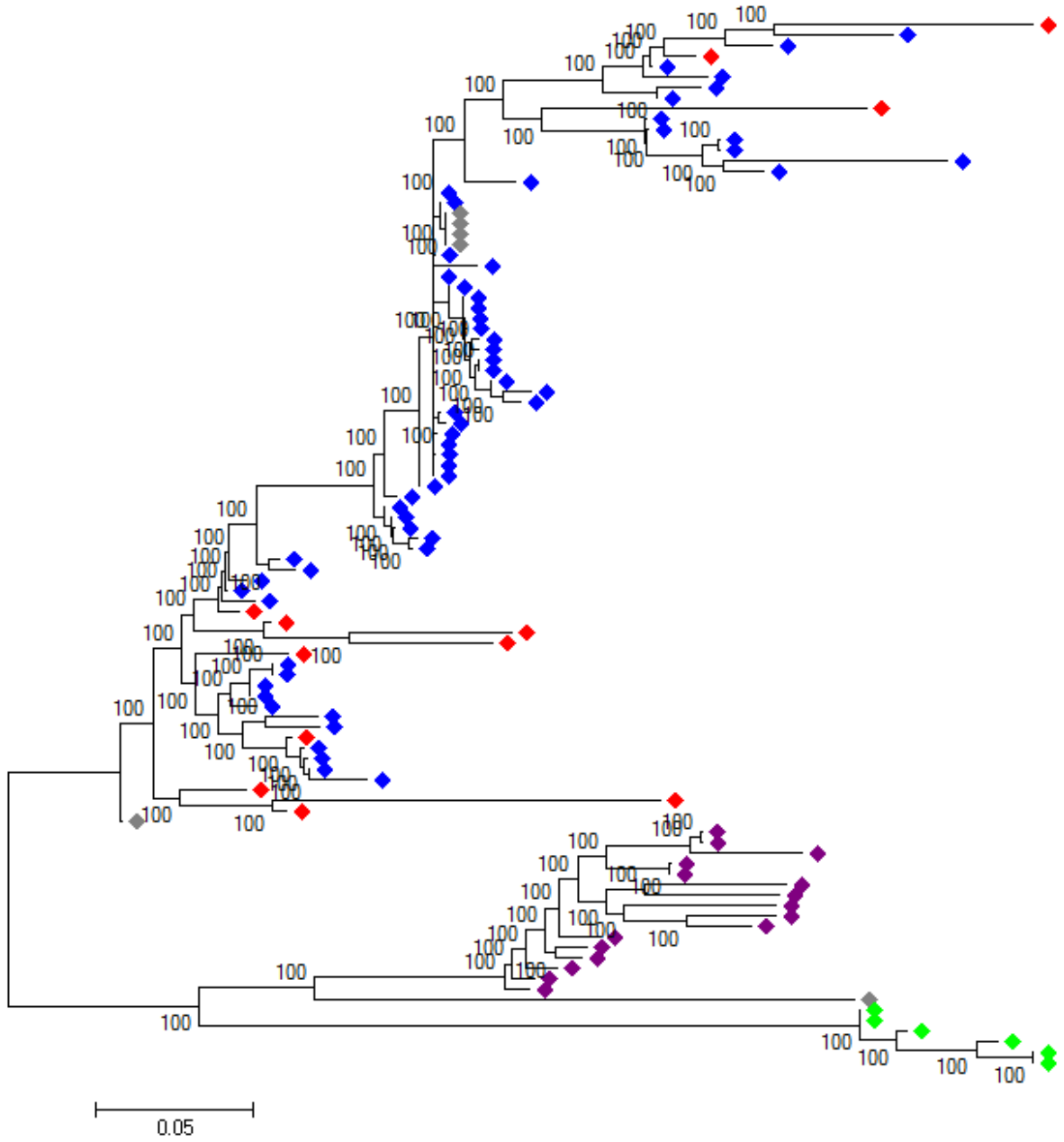


Figure S7.222 Chromosome 4: 77405001 - 77555001

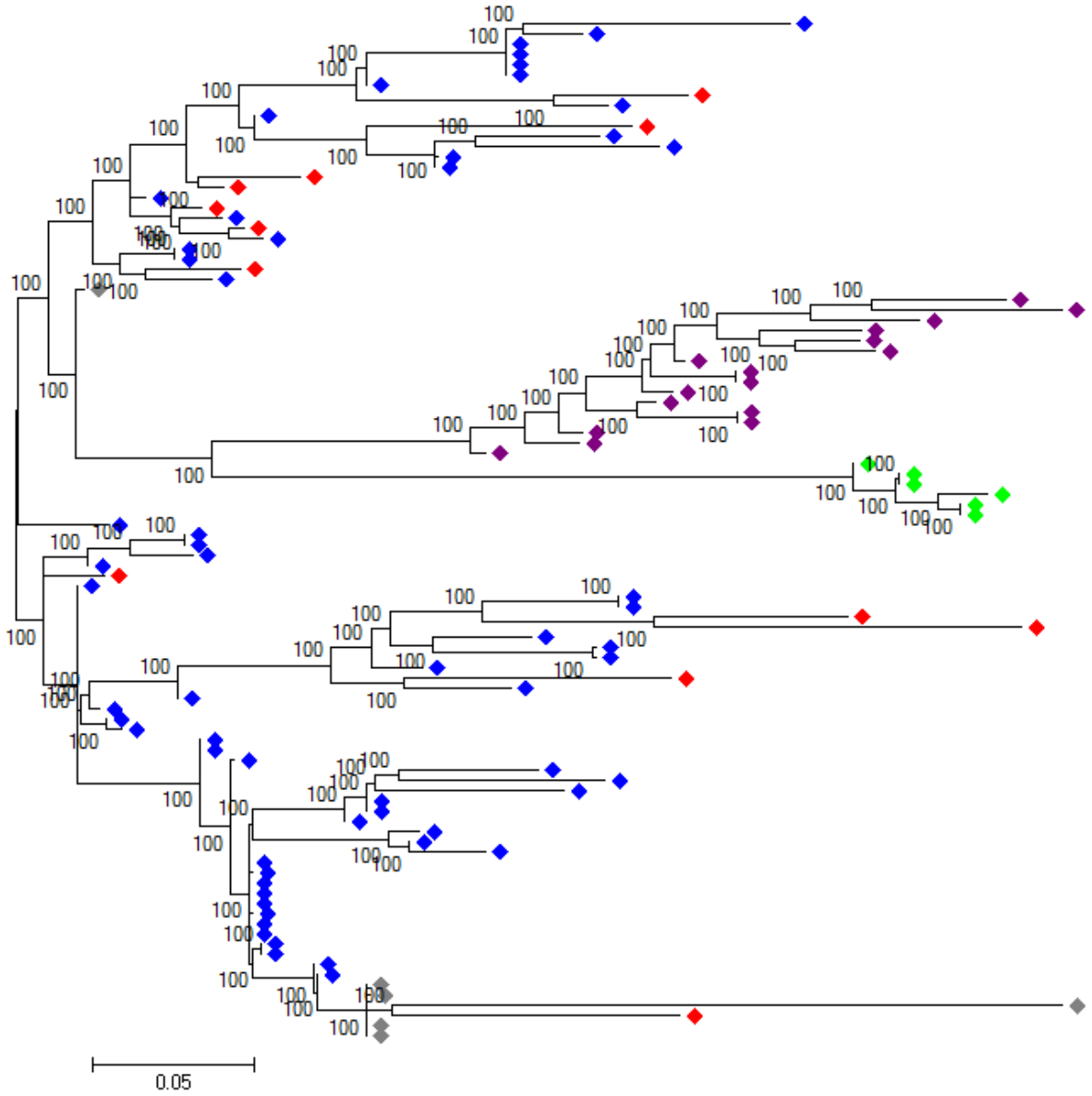


Figure S7.223 Chromosome 4: 77600001 - 77625001

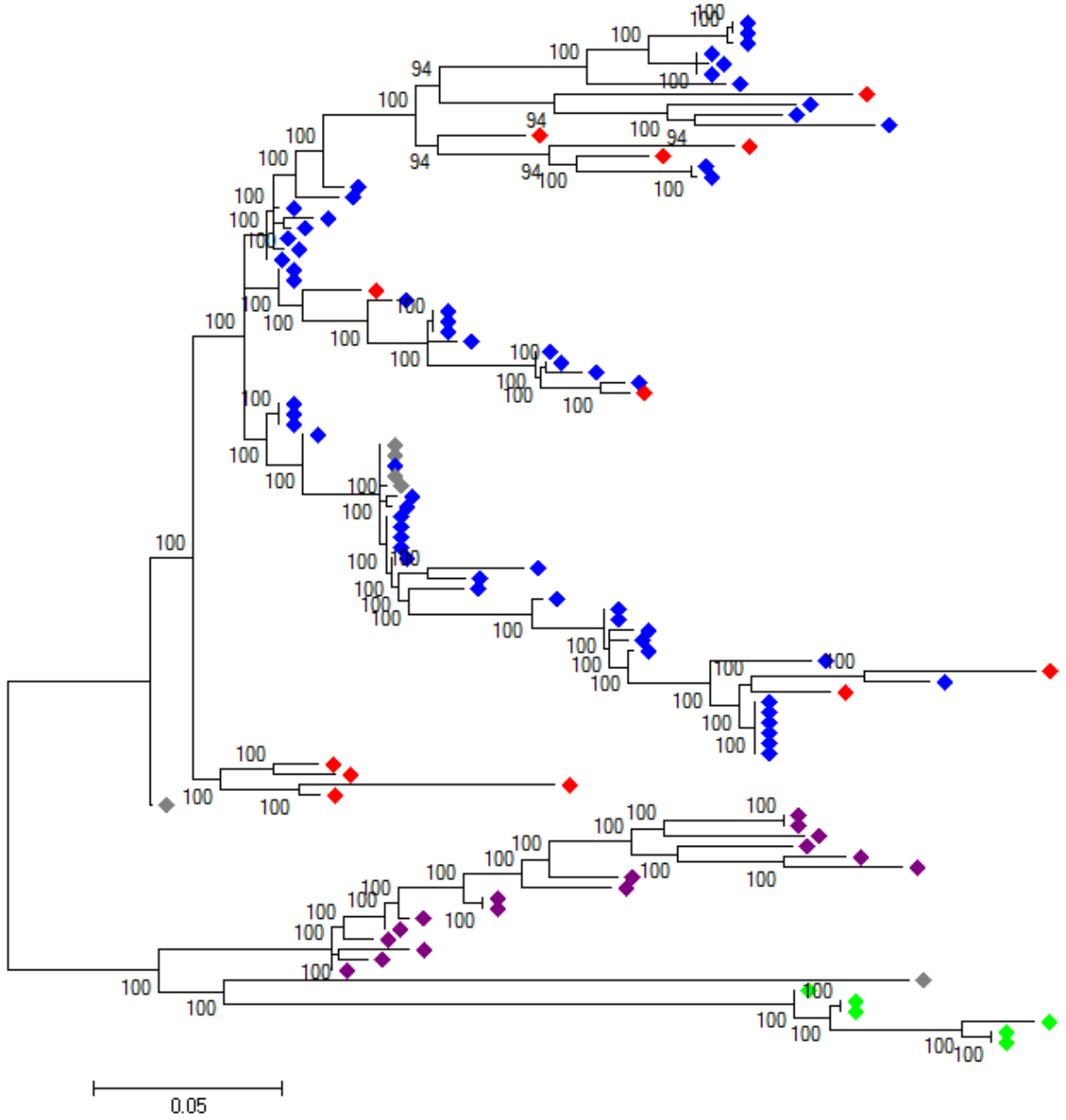


Figure S7.224 Chromosome 4: 77685001 - 77705001

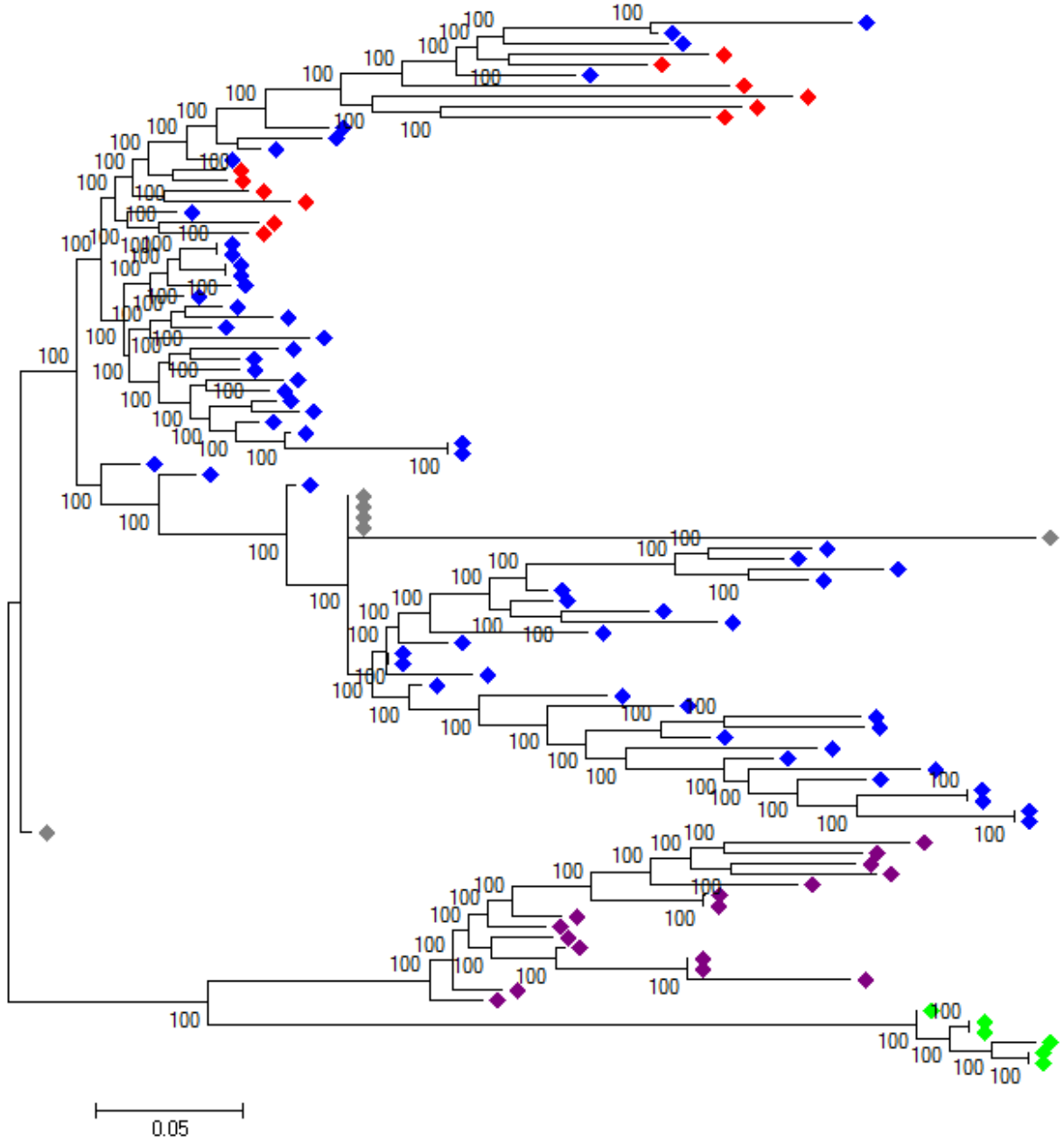


Figure S7.225 Chromosome 4: 77910001 - 78060001

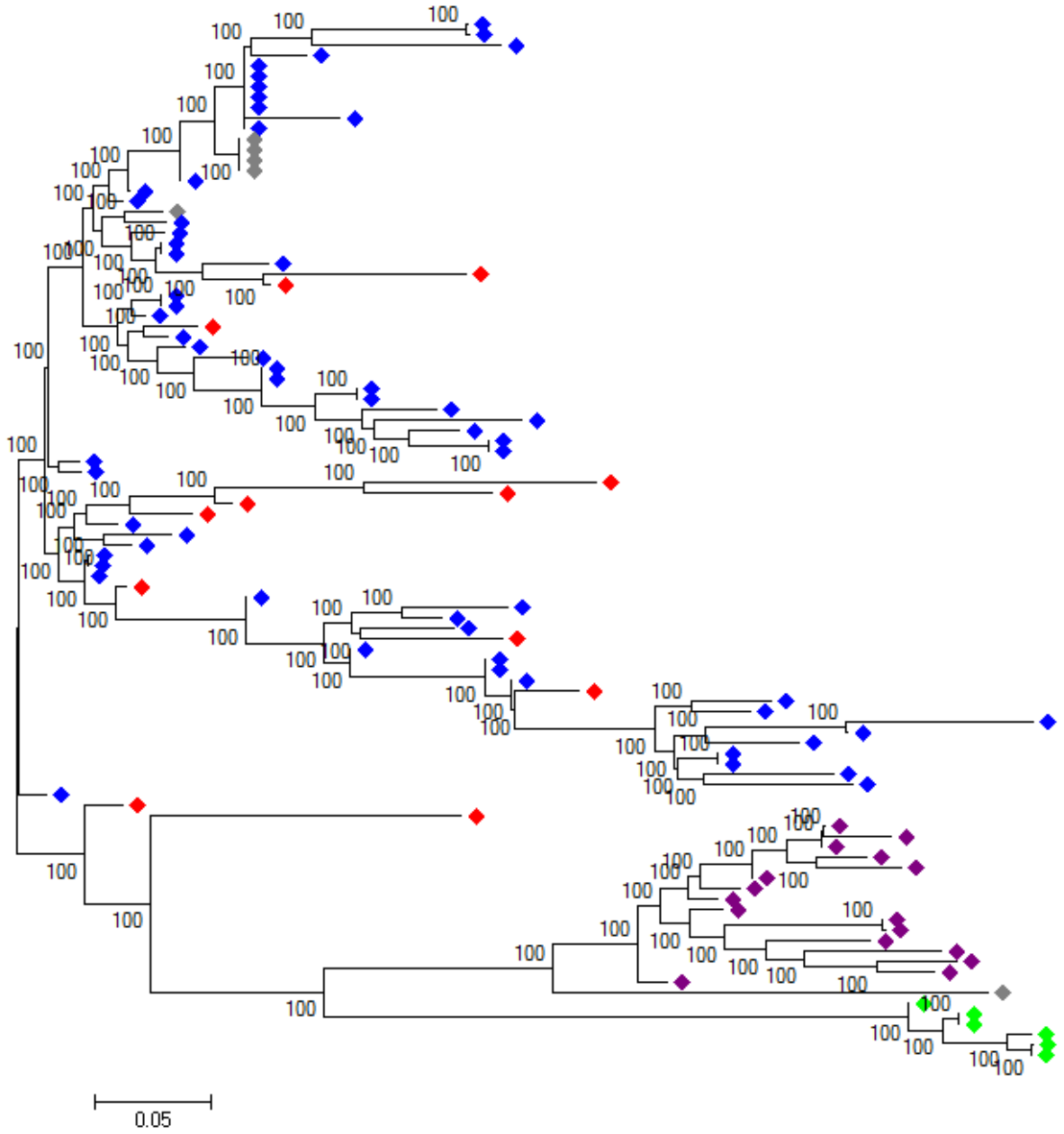


Figure S7.226 Chromosome 4: 78095001 - 78125001

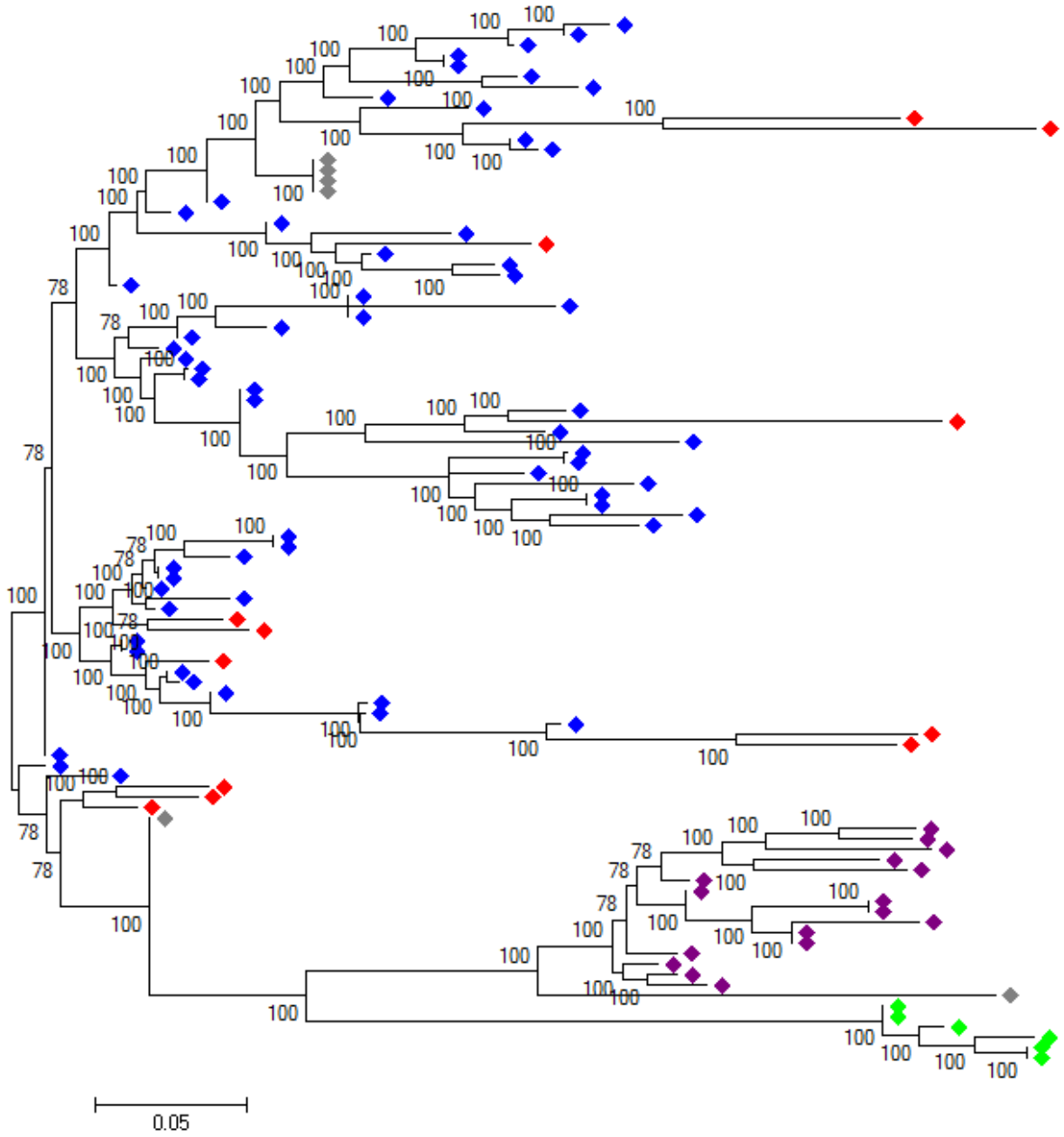


Figure S7.227 Chromosome 4: 78200001 - 78220001

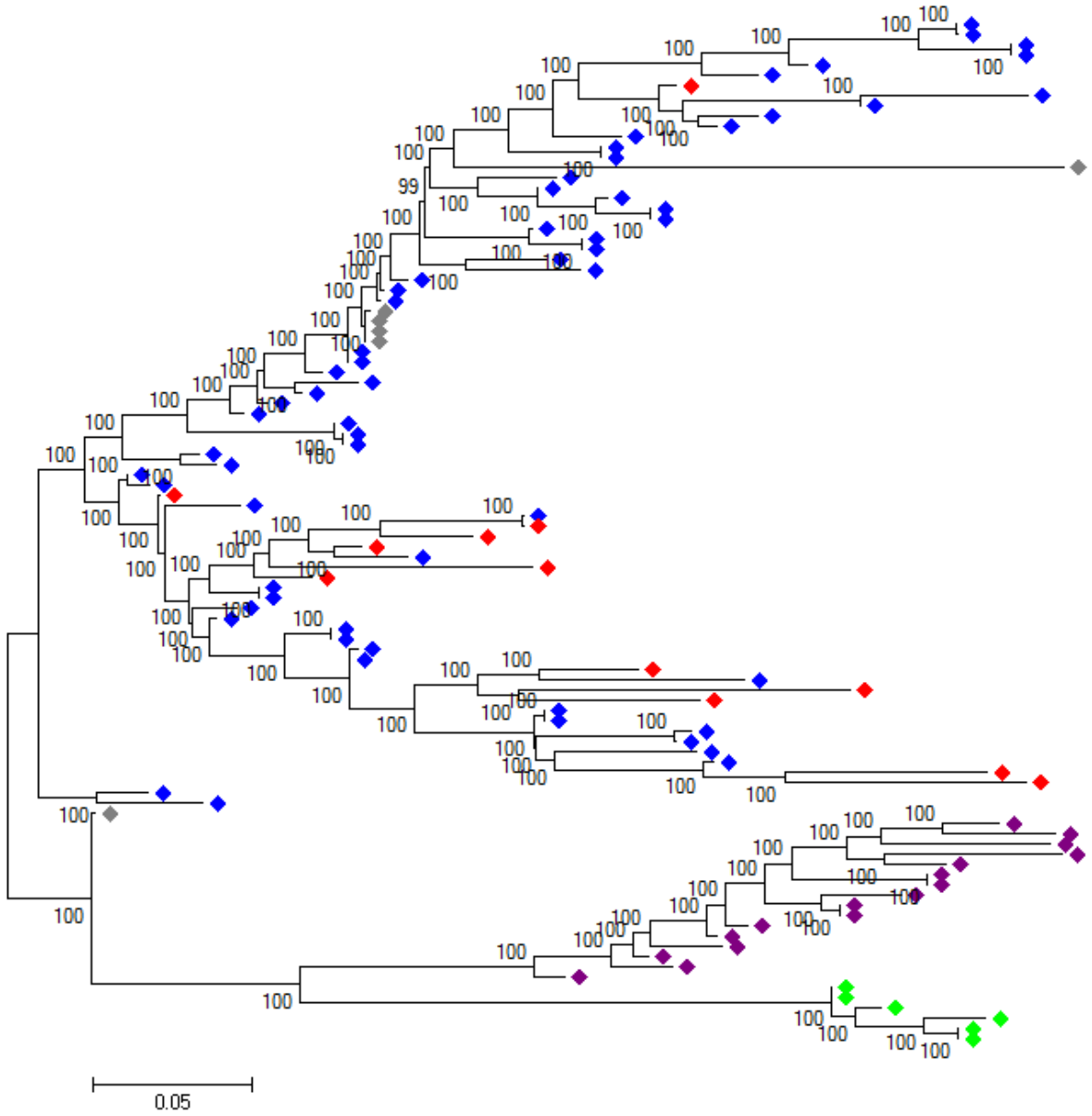


Figure S7.228 Chromosome 4: 78255001 - 78300001

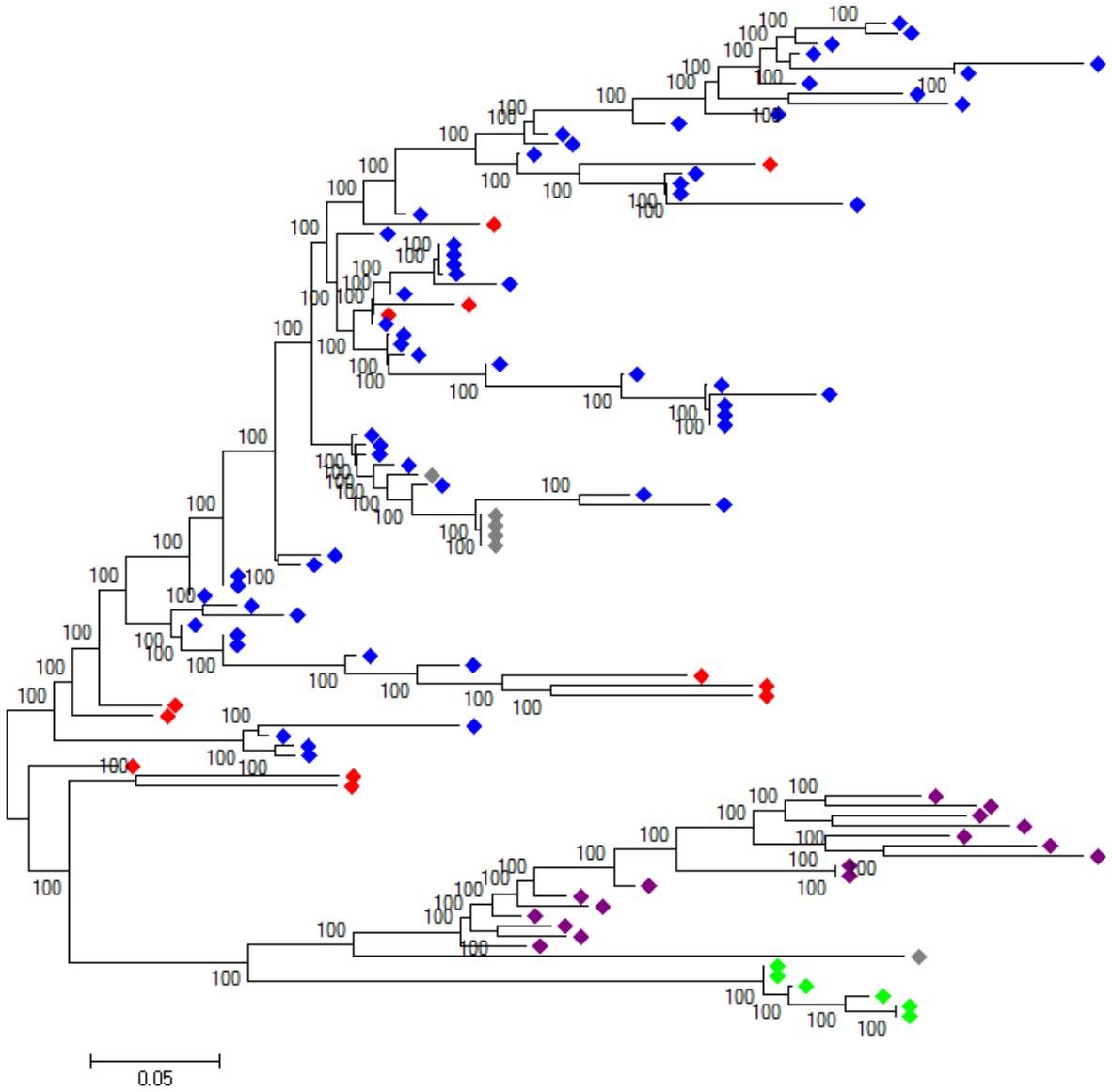


Figure S7.229 Chromosome 4: 78375001 - 78405001

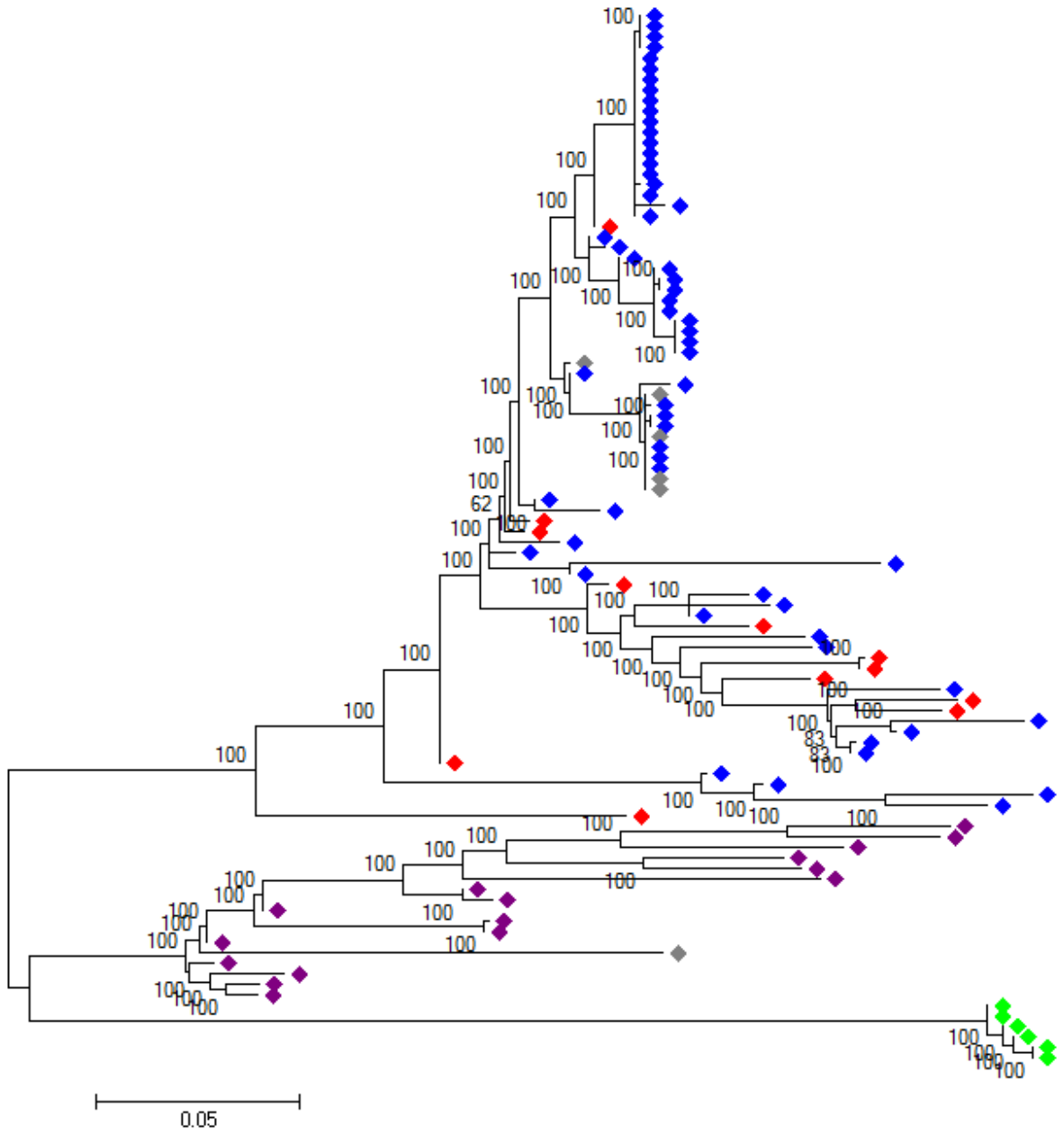


Figure S7.230 Chromosome 4: 78495001 - 78515001

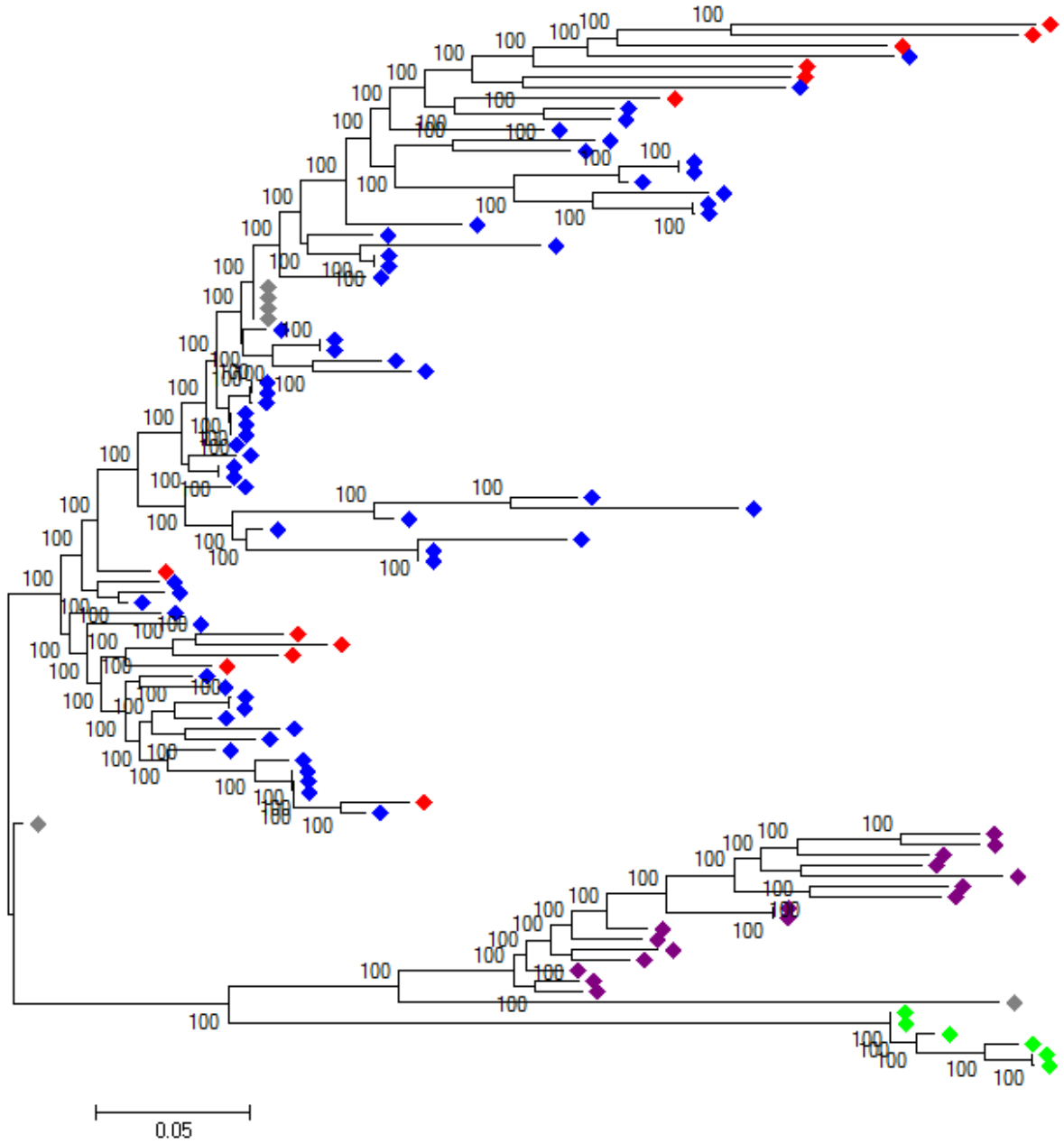


Figure S7.231 Chromosome 4: 78605001 - 78710001

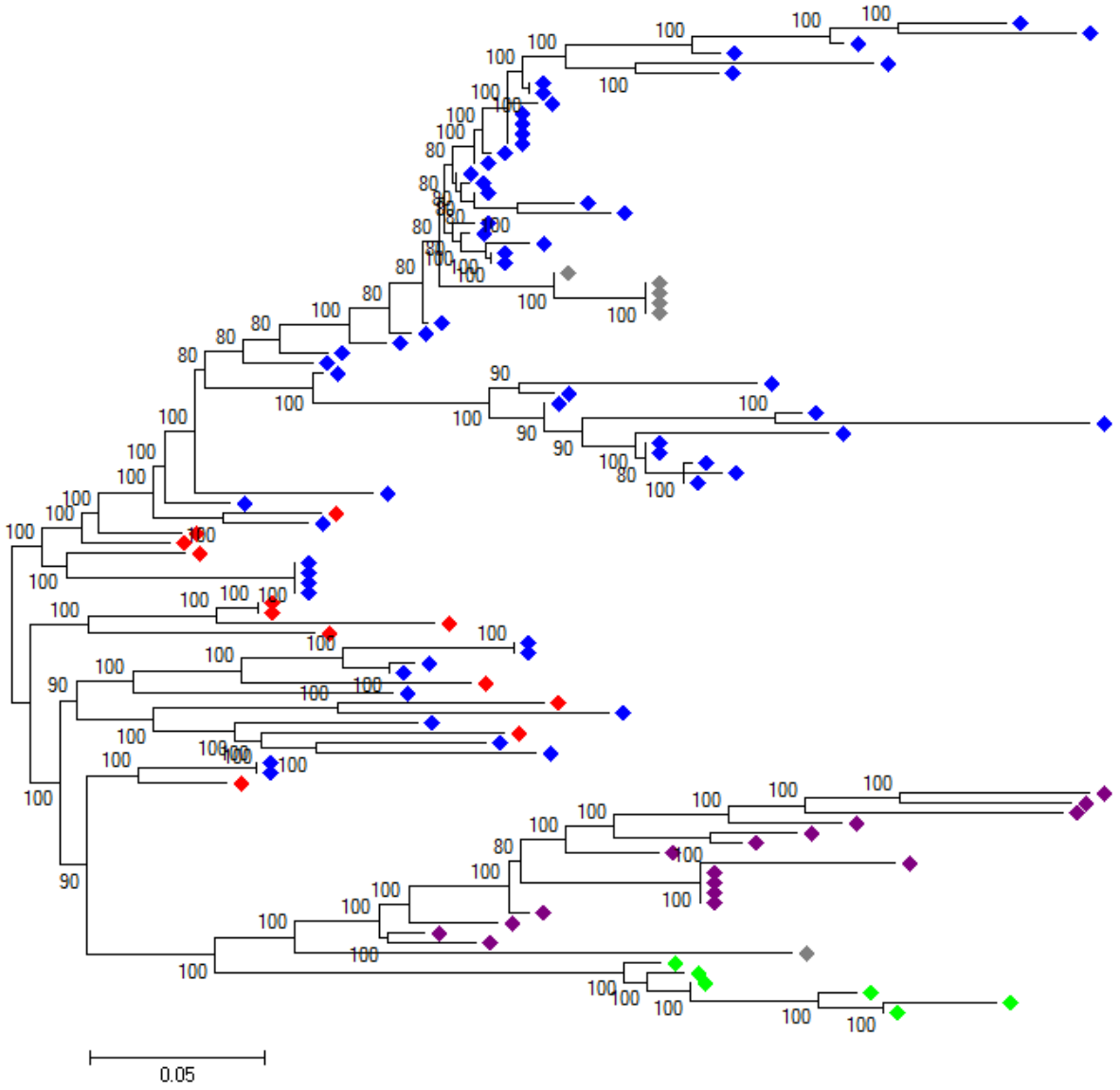


Figure S7.232 Chromosome 6: 210001 - 230001

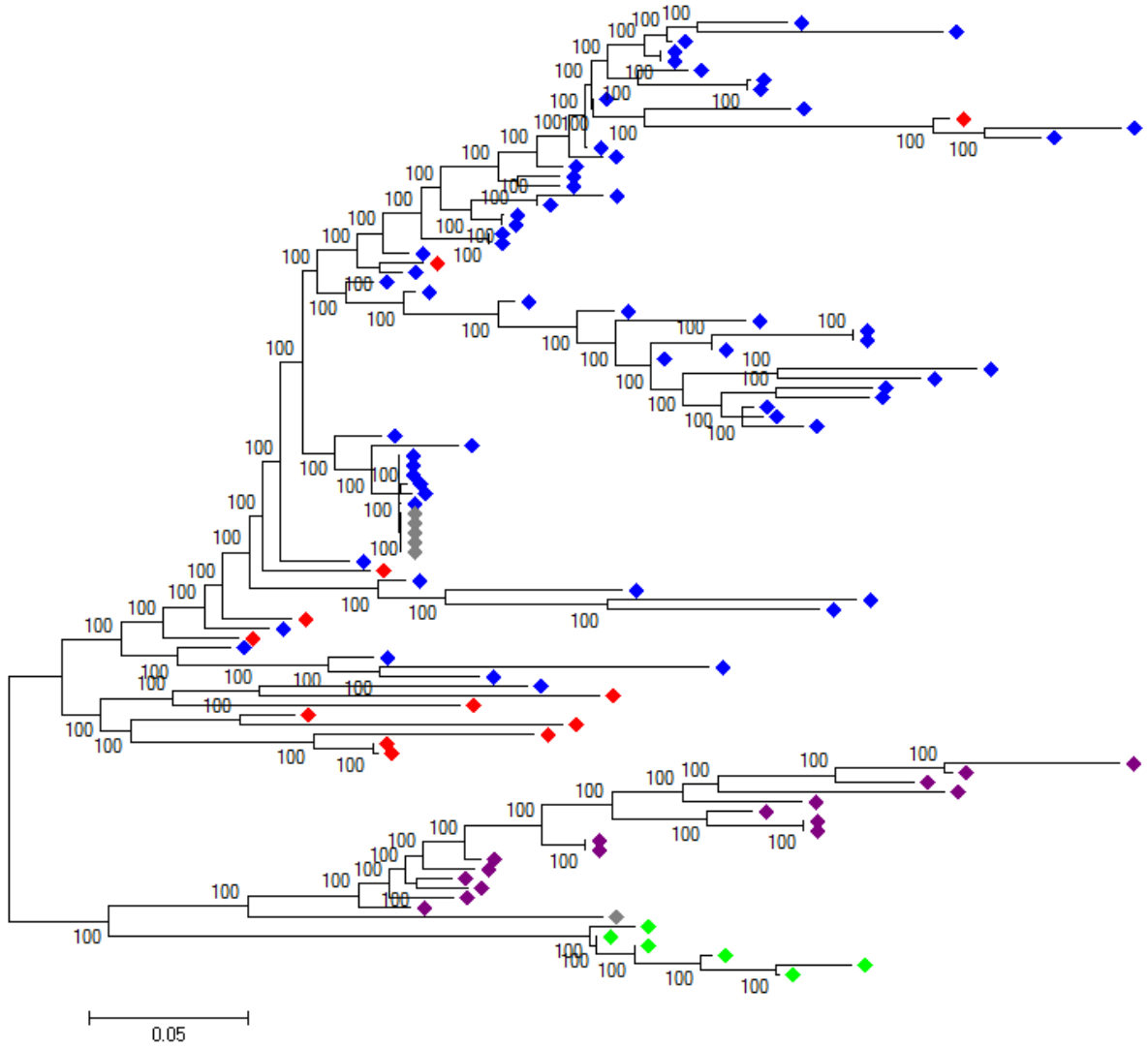


Figure S7.233 Chromosome 6: 680001 - 730001

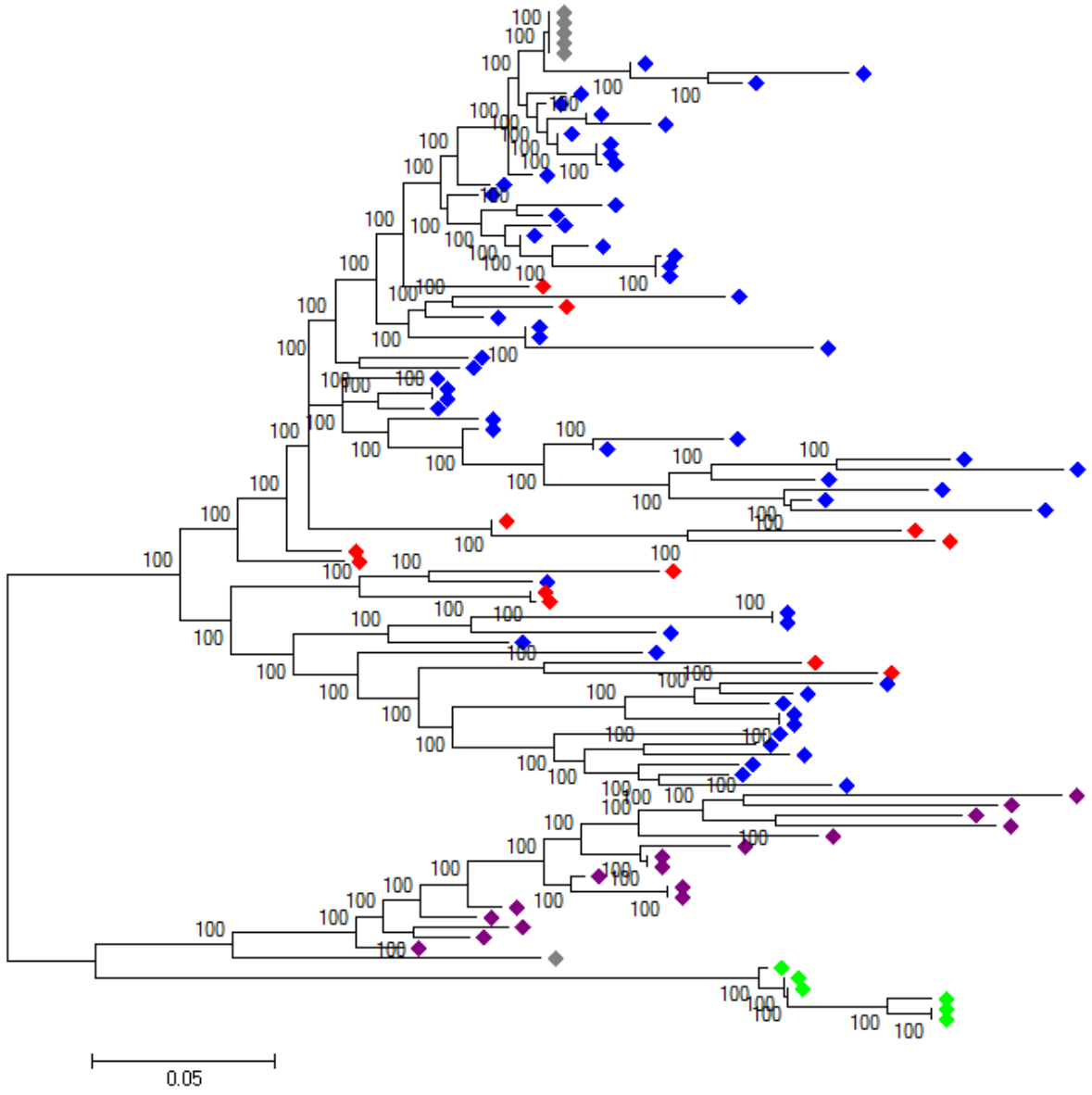


Figure S7.234 Chromosome 6: 790001 - 810001

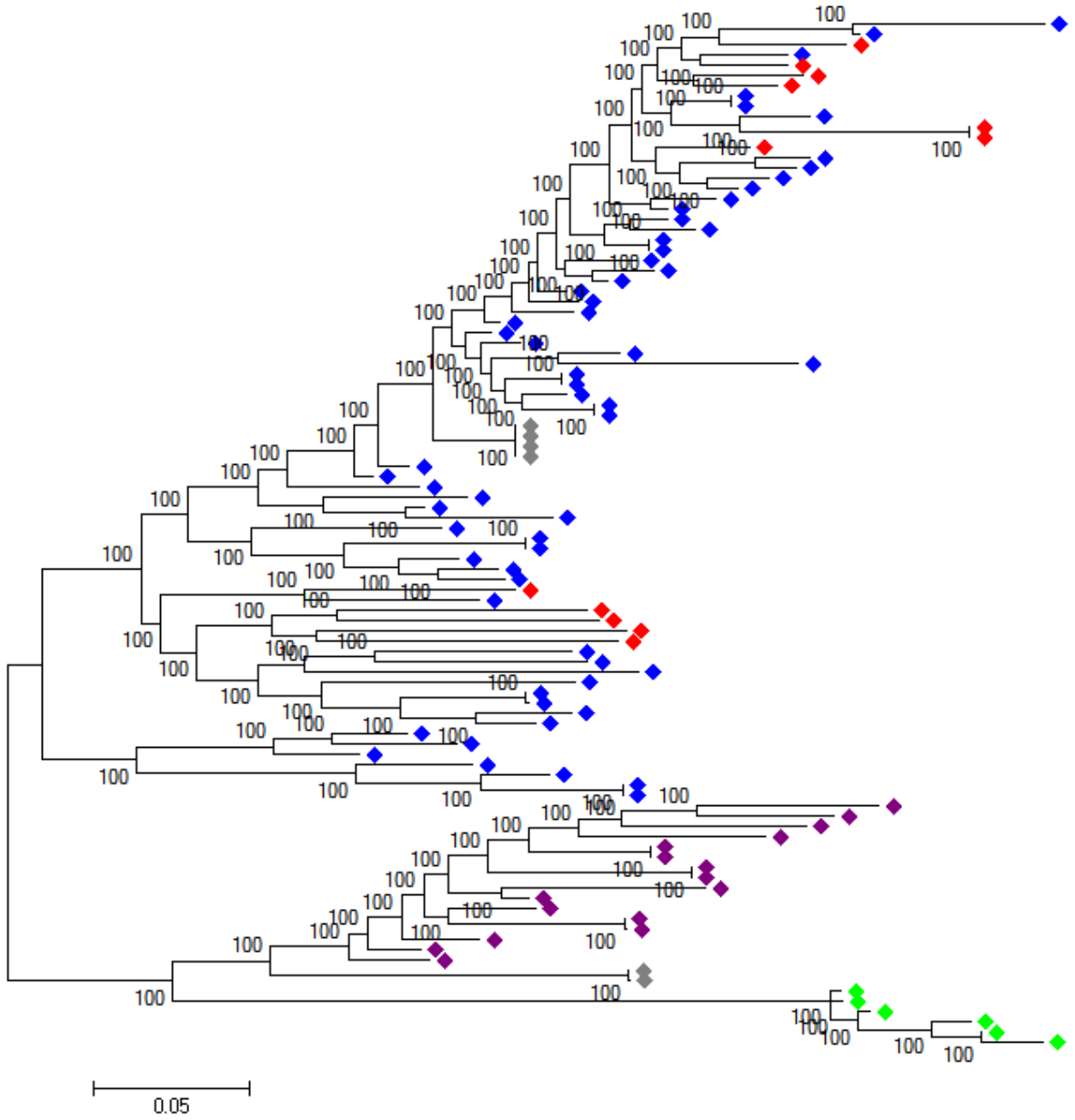


Figure S7.235 Chromosome 6: 1320001 - 1400001

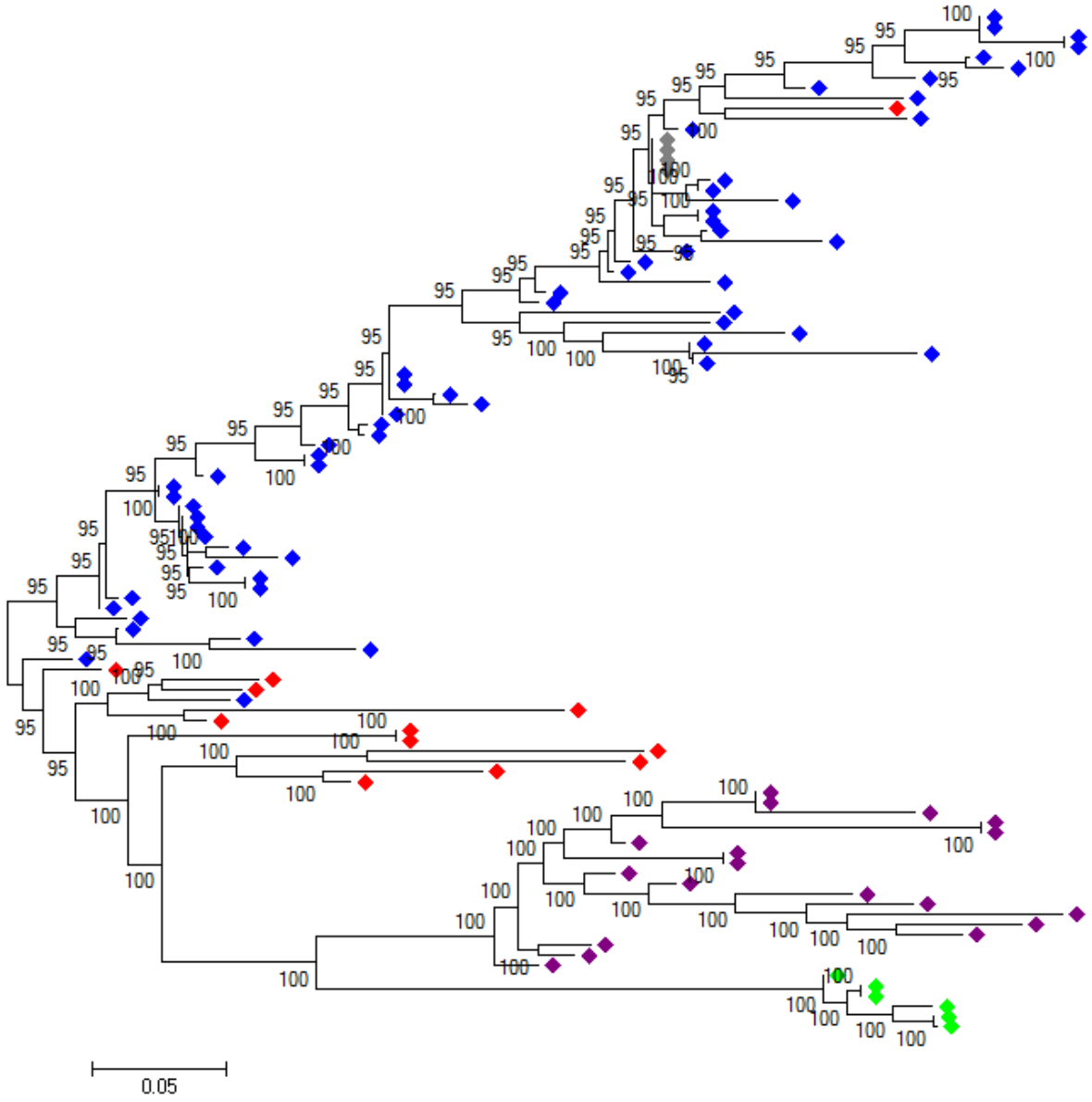


Figure S7.236 Chromosome 6: 1470001 - 1500001

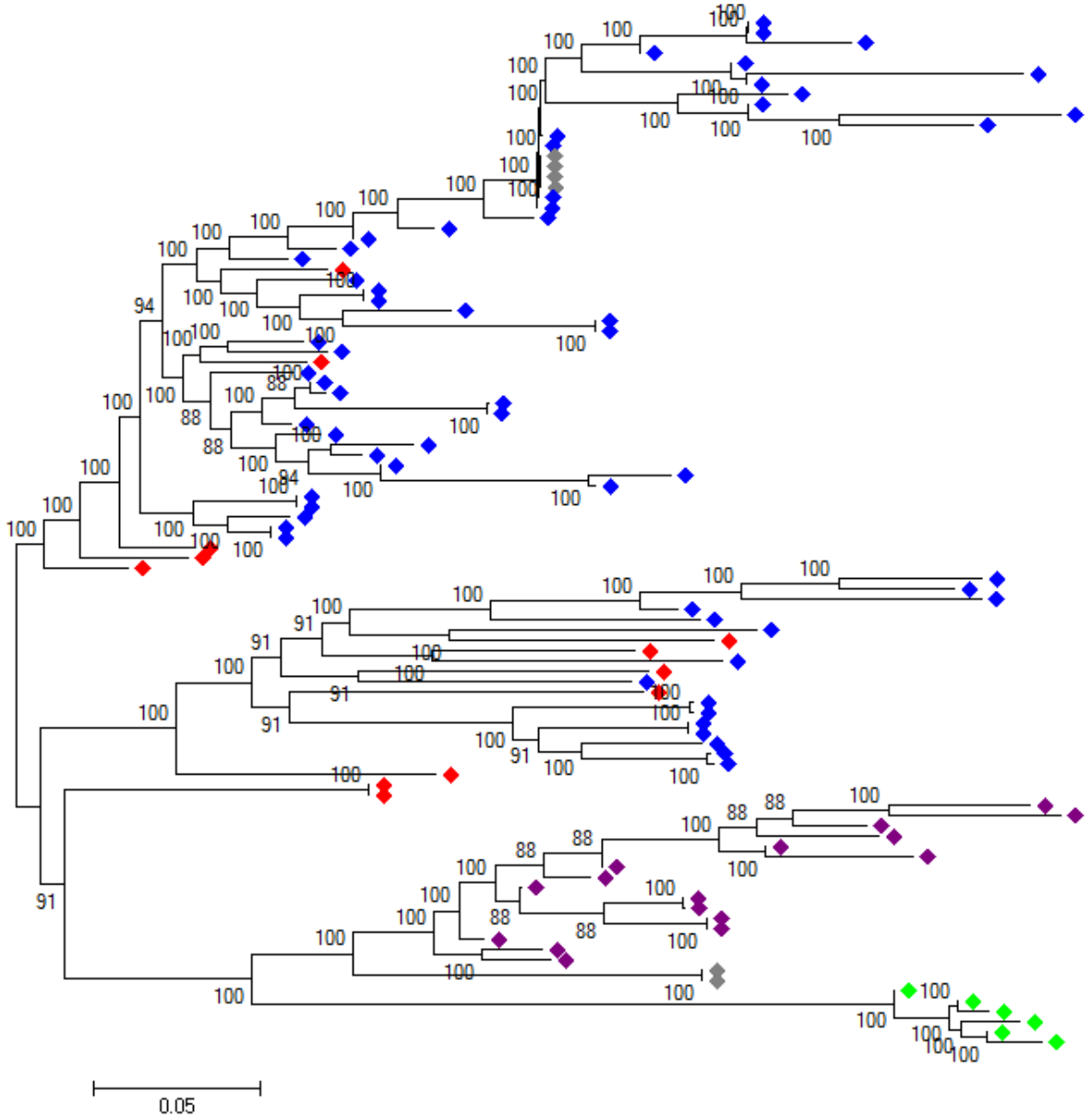


Figure S7.237 Chromosome 6: 1565001 - 1620001

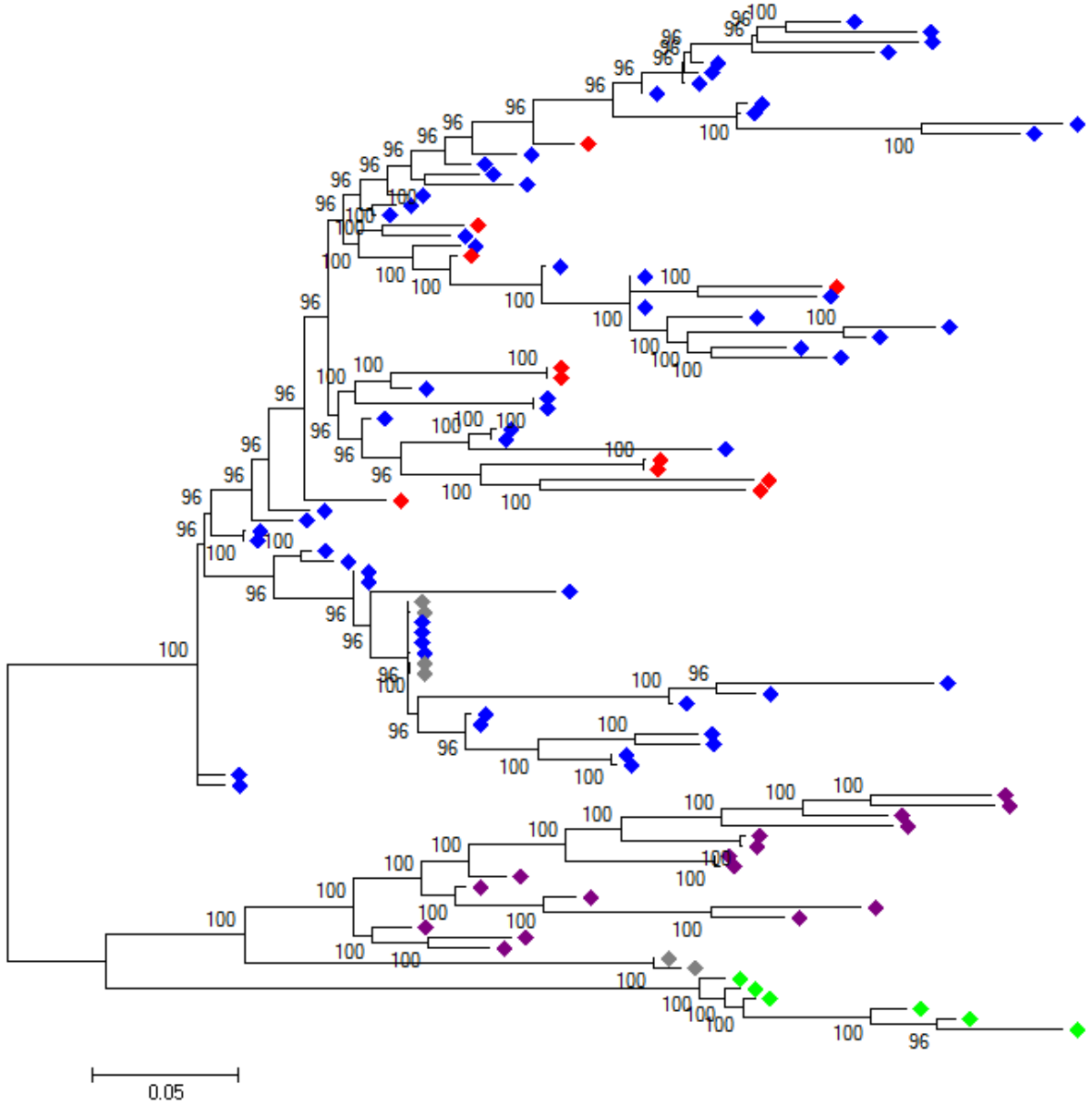


Figure S7.238 Chromosome 6: 1660001 - 1685001

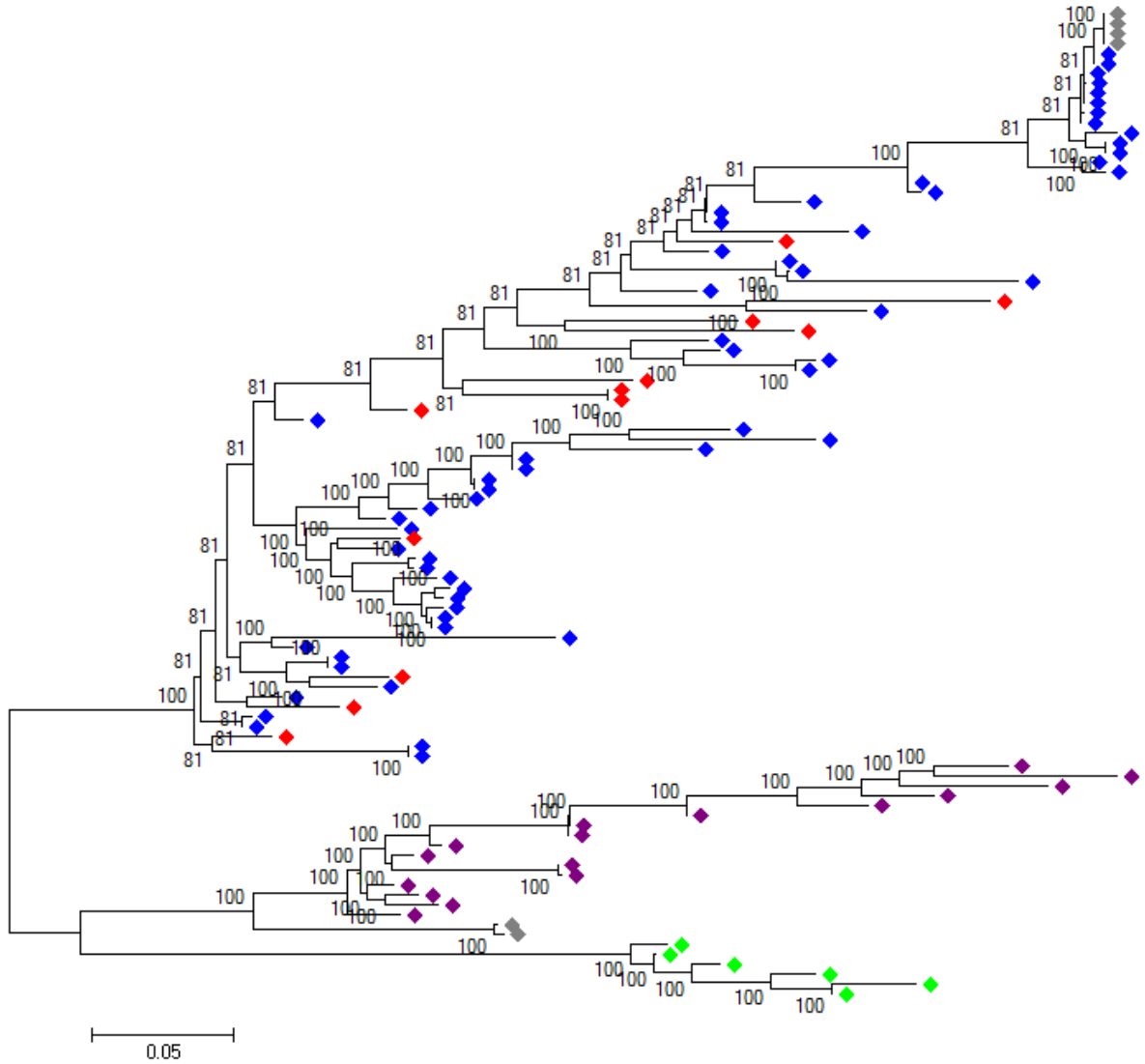


Figure S7.239 Chromosome 6: 1695001 - 1725001

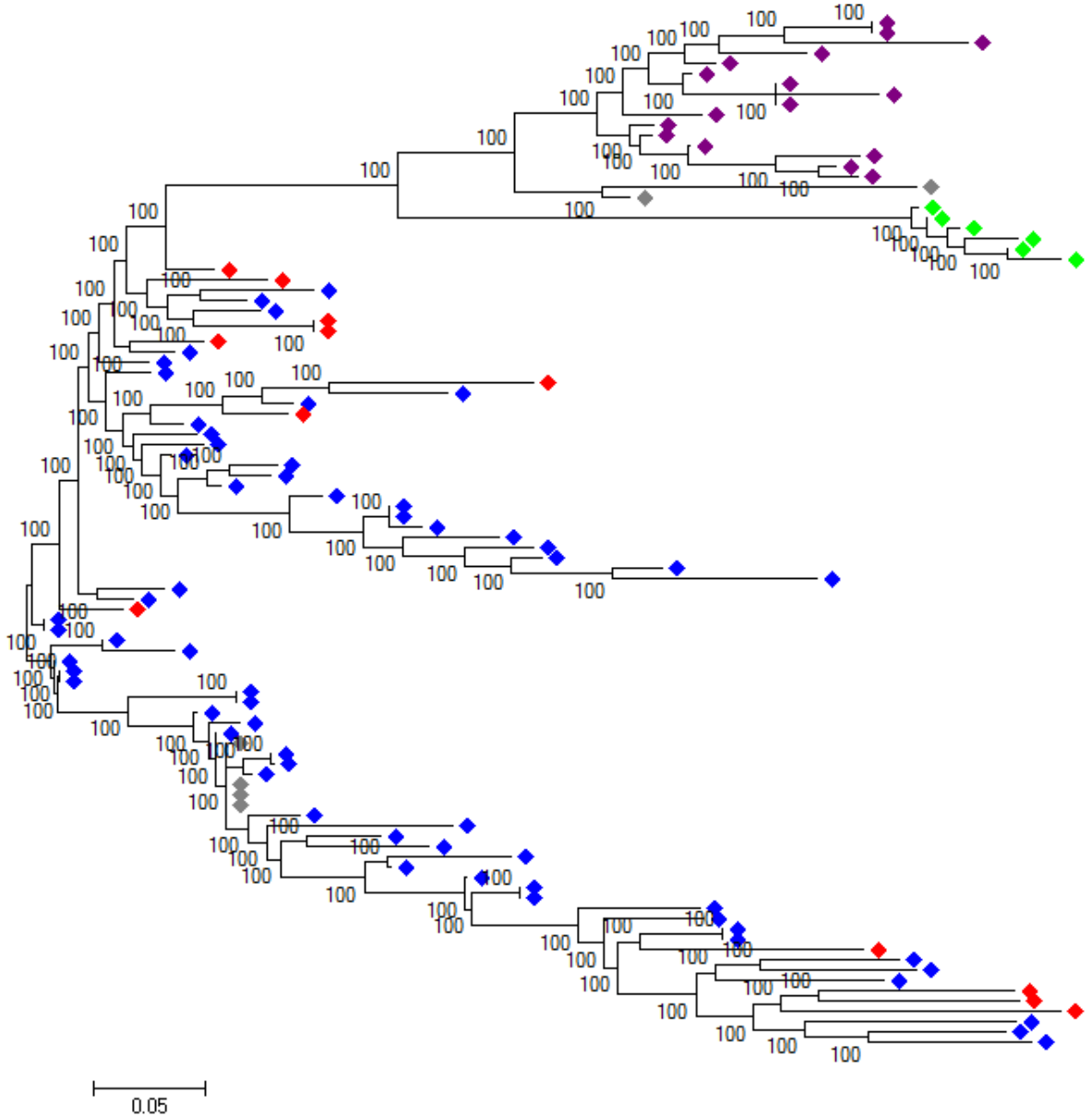


Figure S7.240 Chromosome 6: 1785001 - 1815001

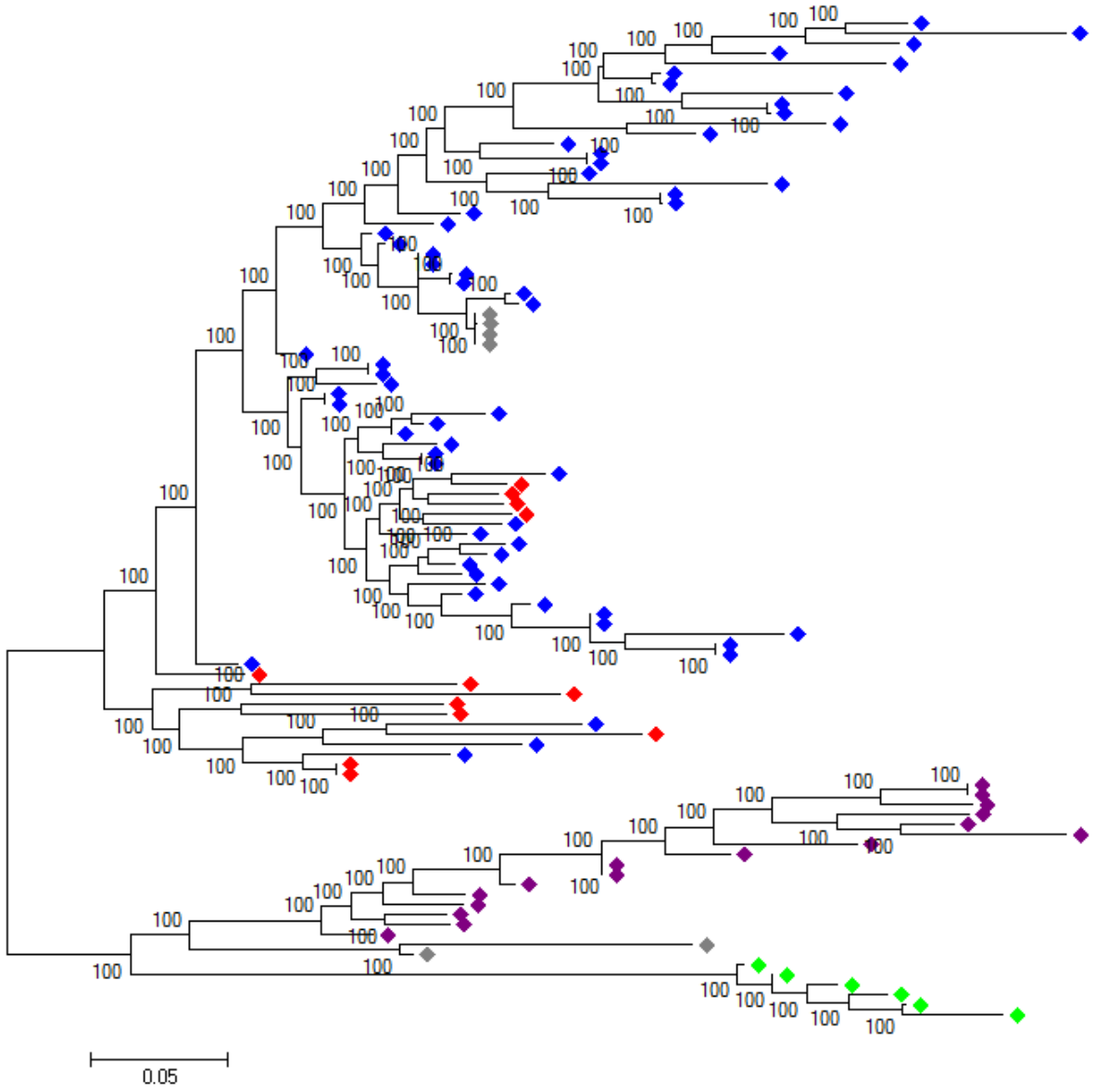


Figure S7.241 Chromosome 6: 1900001 - 1955001

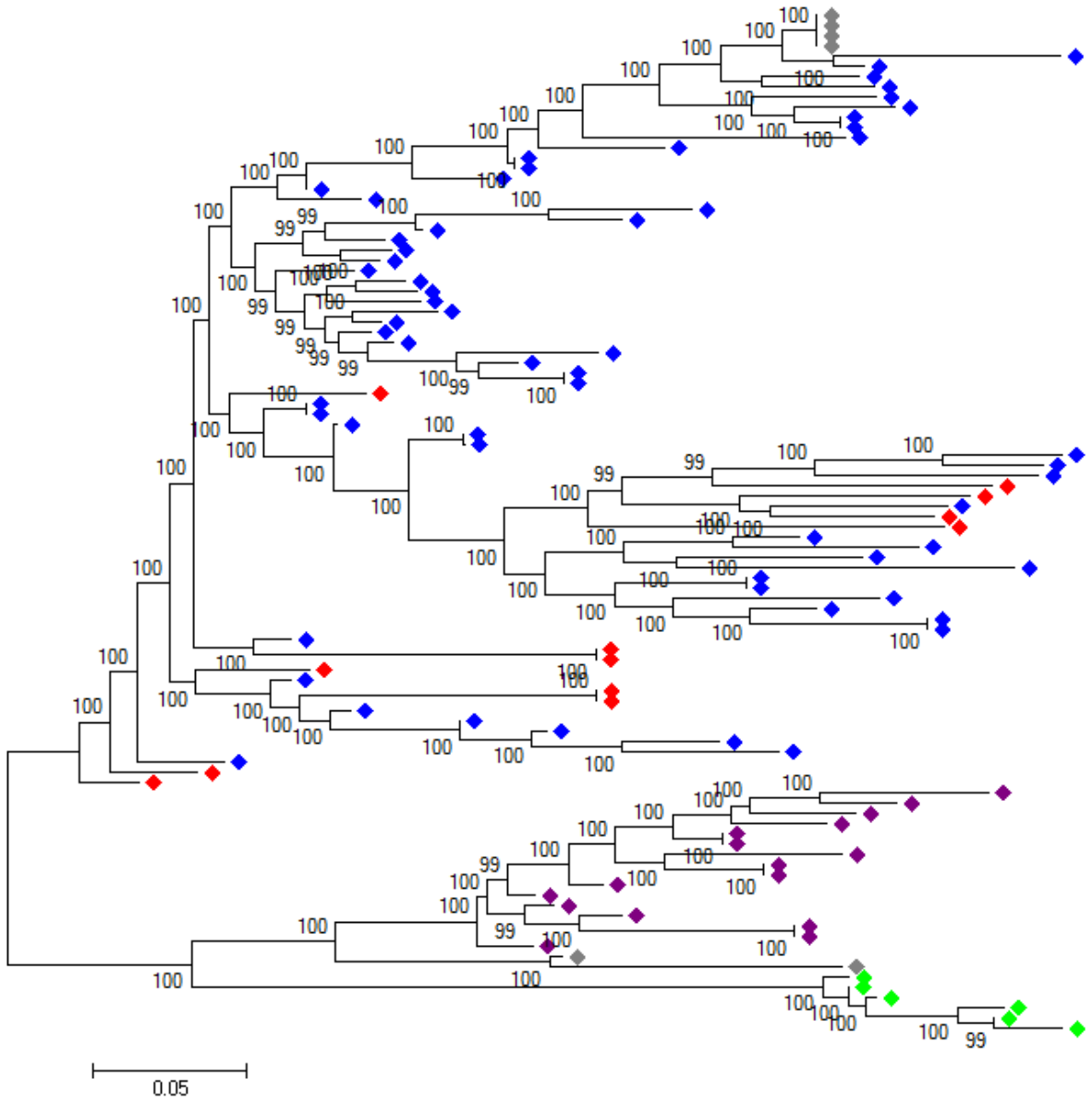


Figure S7.242 Chromosome 6: 2035001 - 2070001

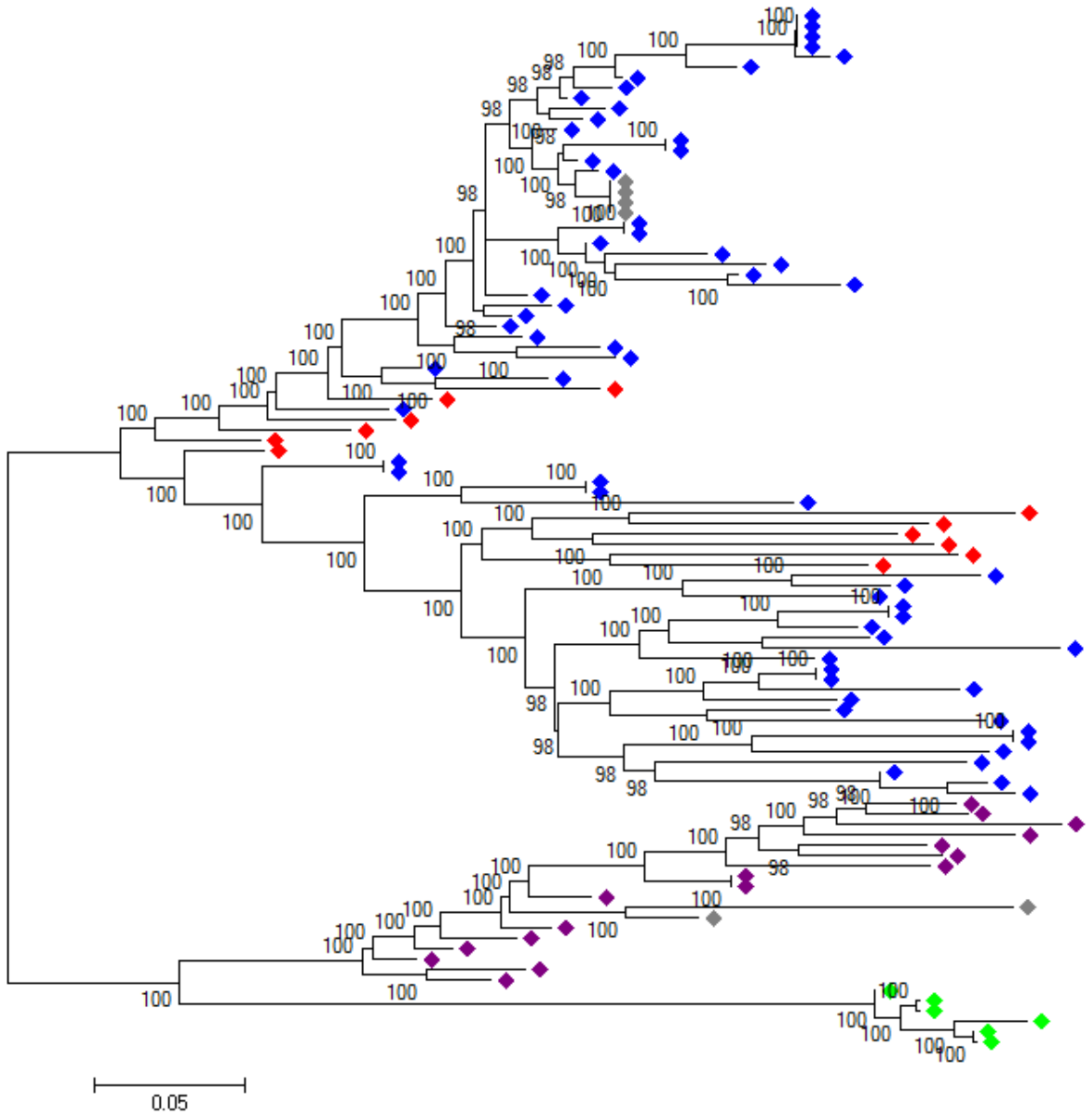


Figure S7.243 Chromosome 6: 2510001 - 2530001

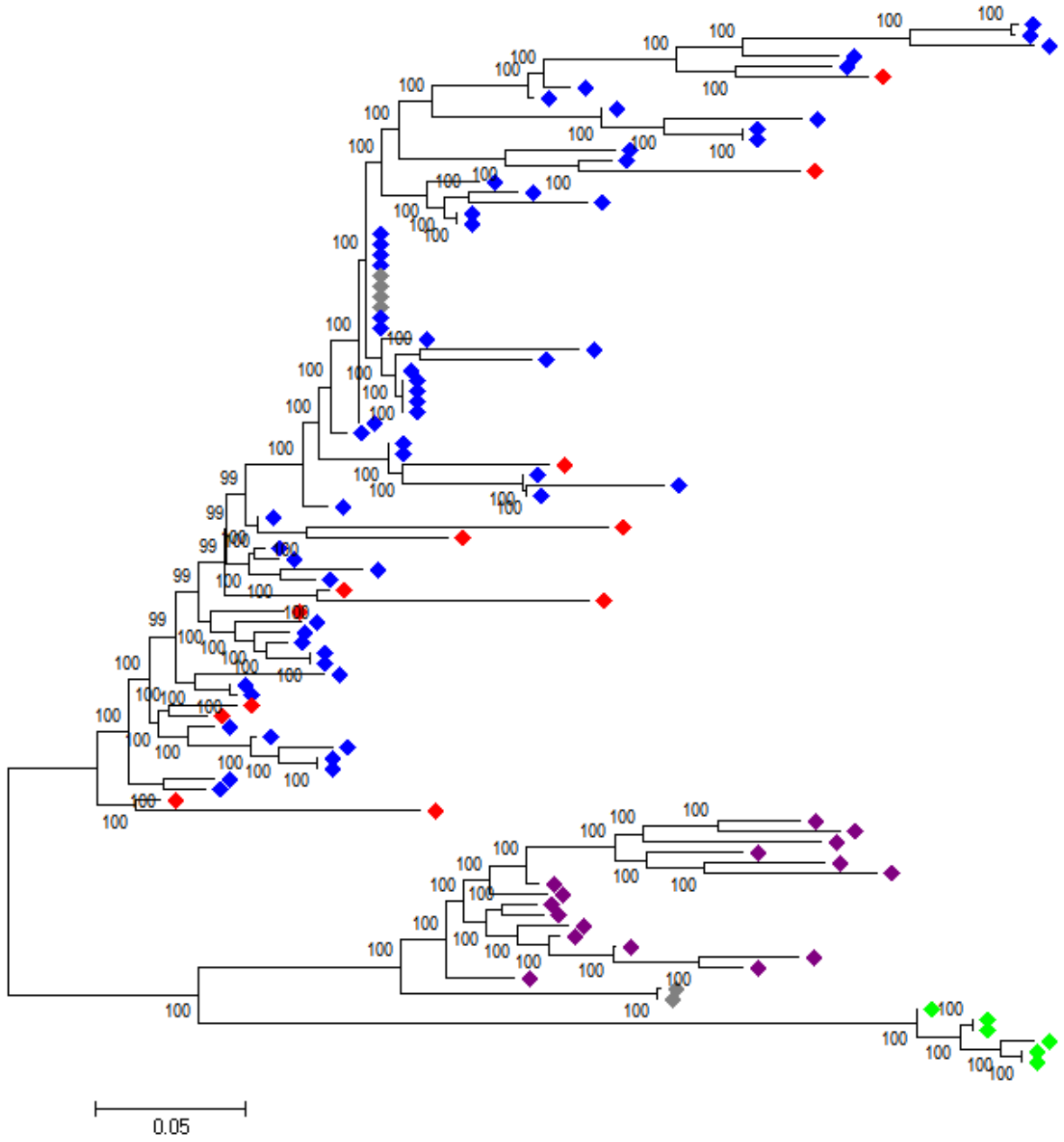


Figure S7.244 Chromosome 6: 3055001 - 3075001

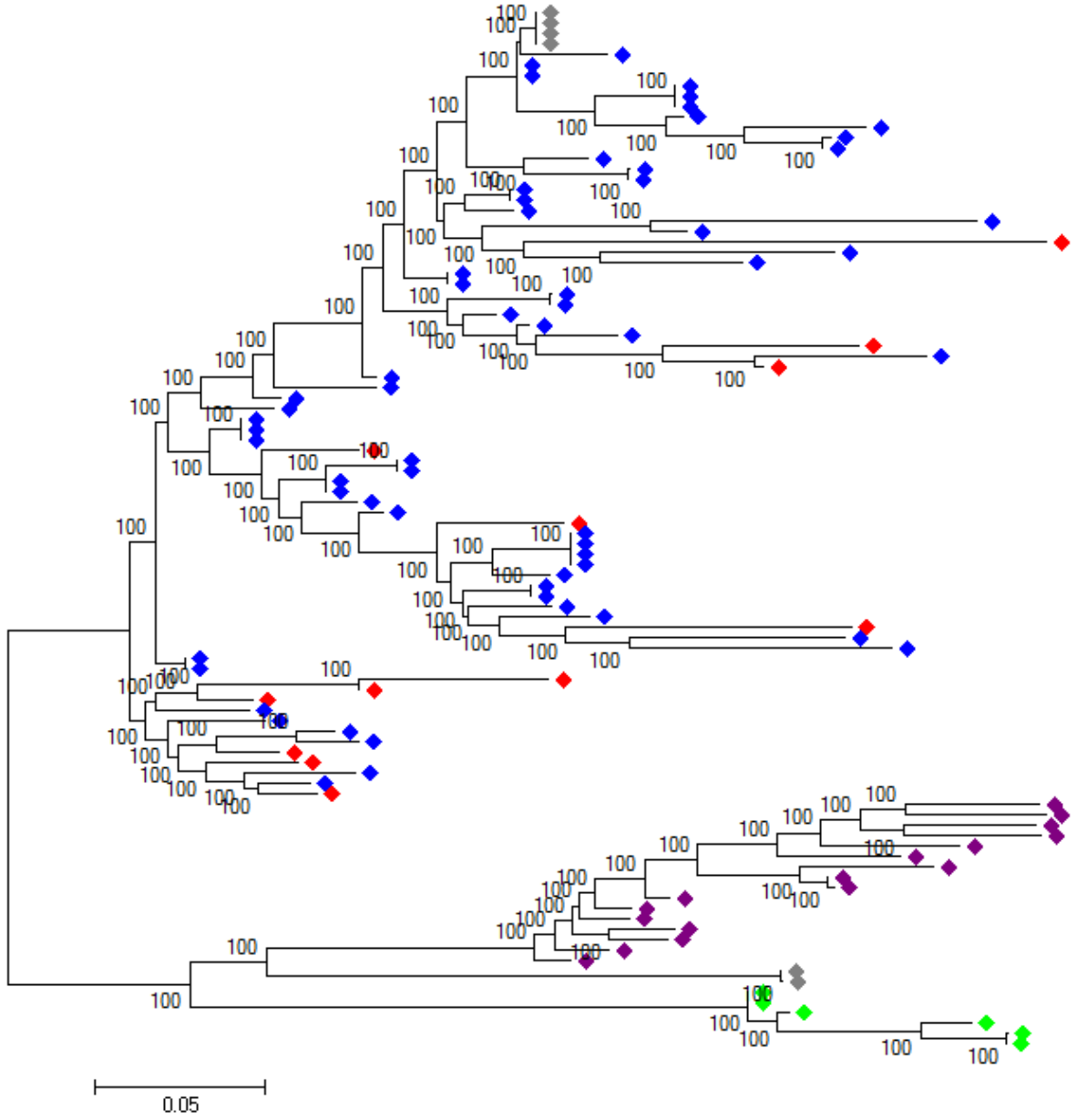


Figure S7.245 Chromosome 6: 3375001 - 3415001

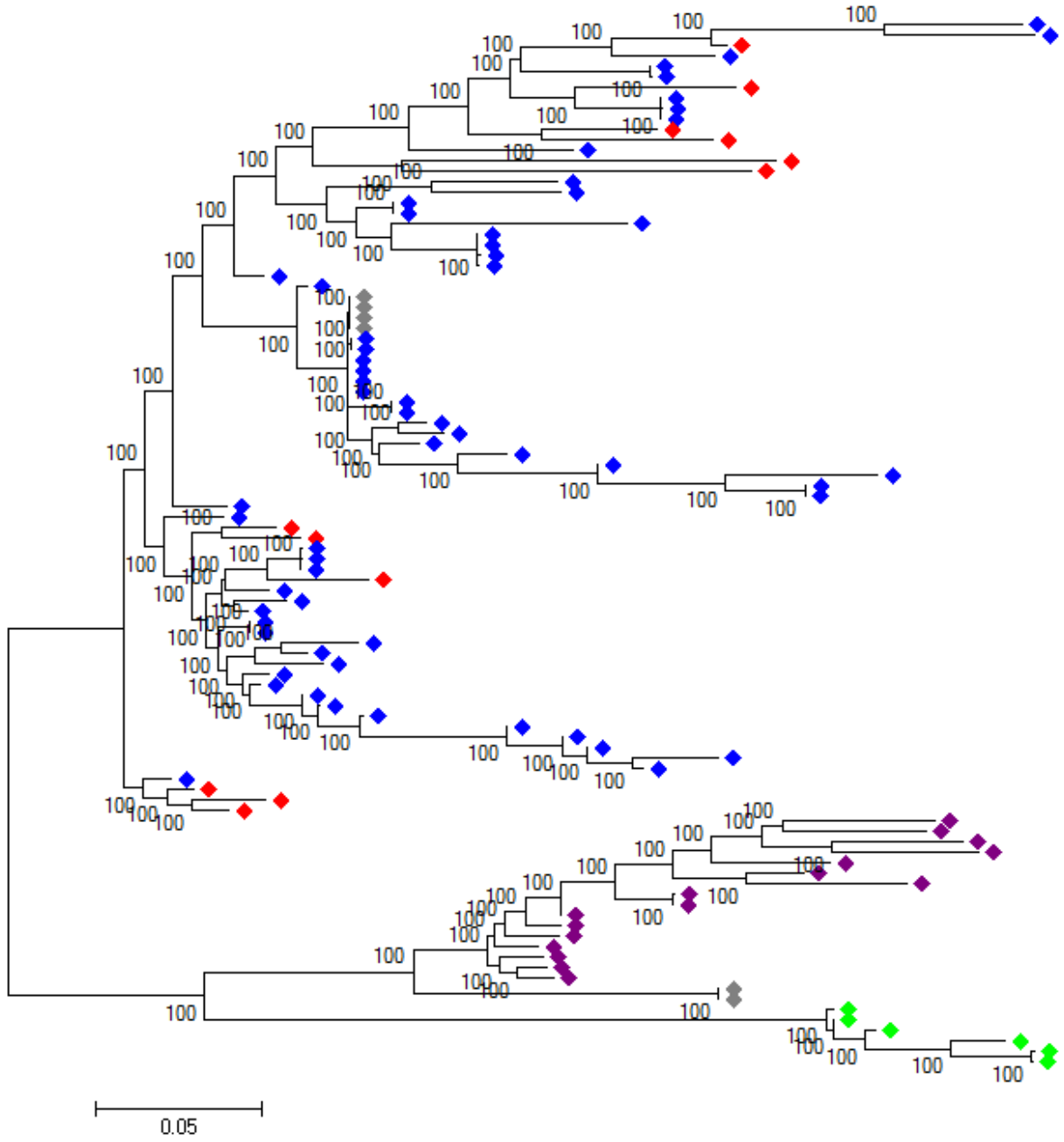


Figure S7.246 Chromosome 6: 3430001 - 3485001

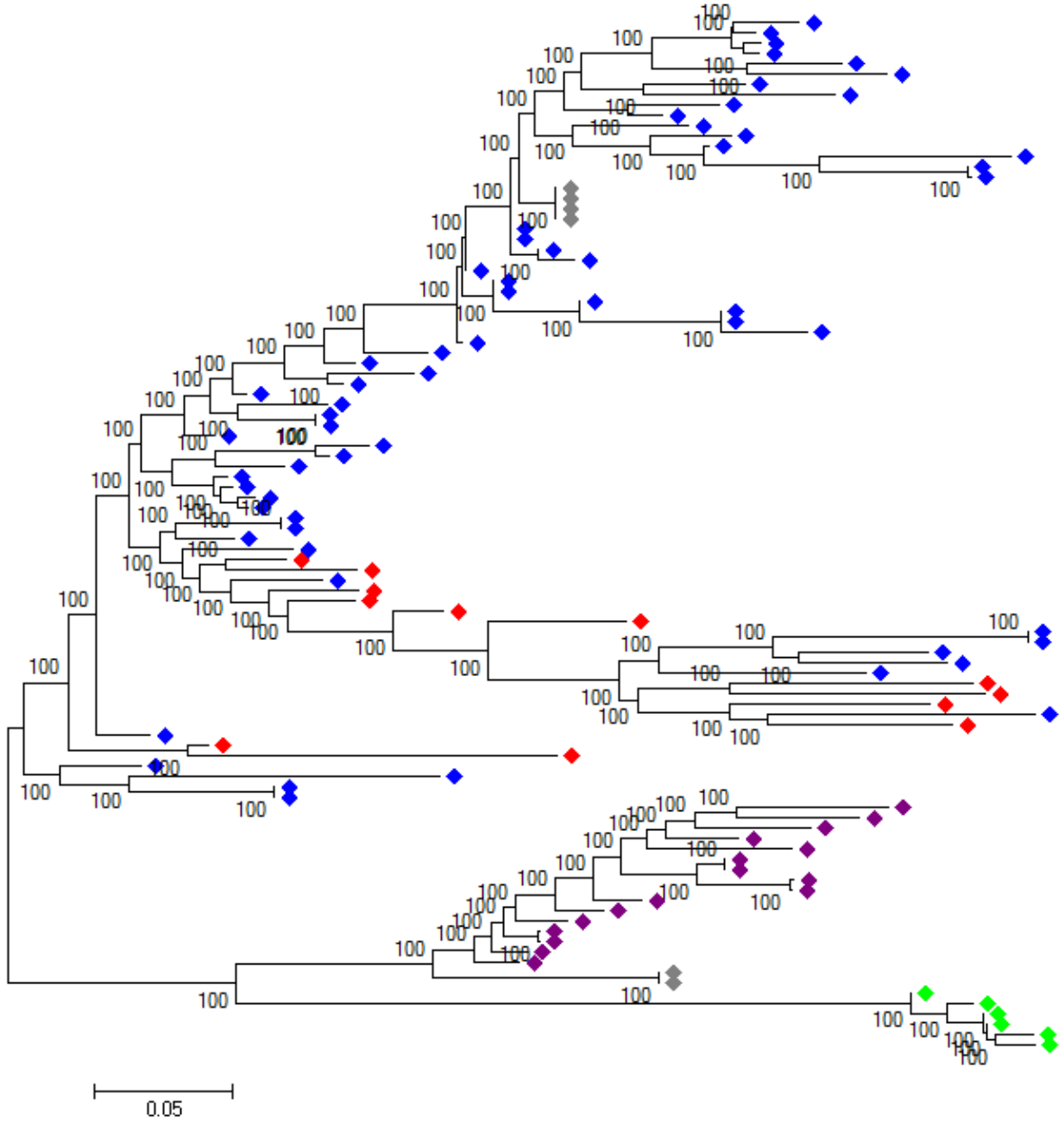


Figure S7.247 Chromosome 6: 4125001 - 4170001

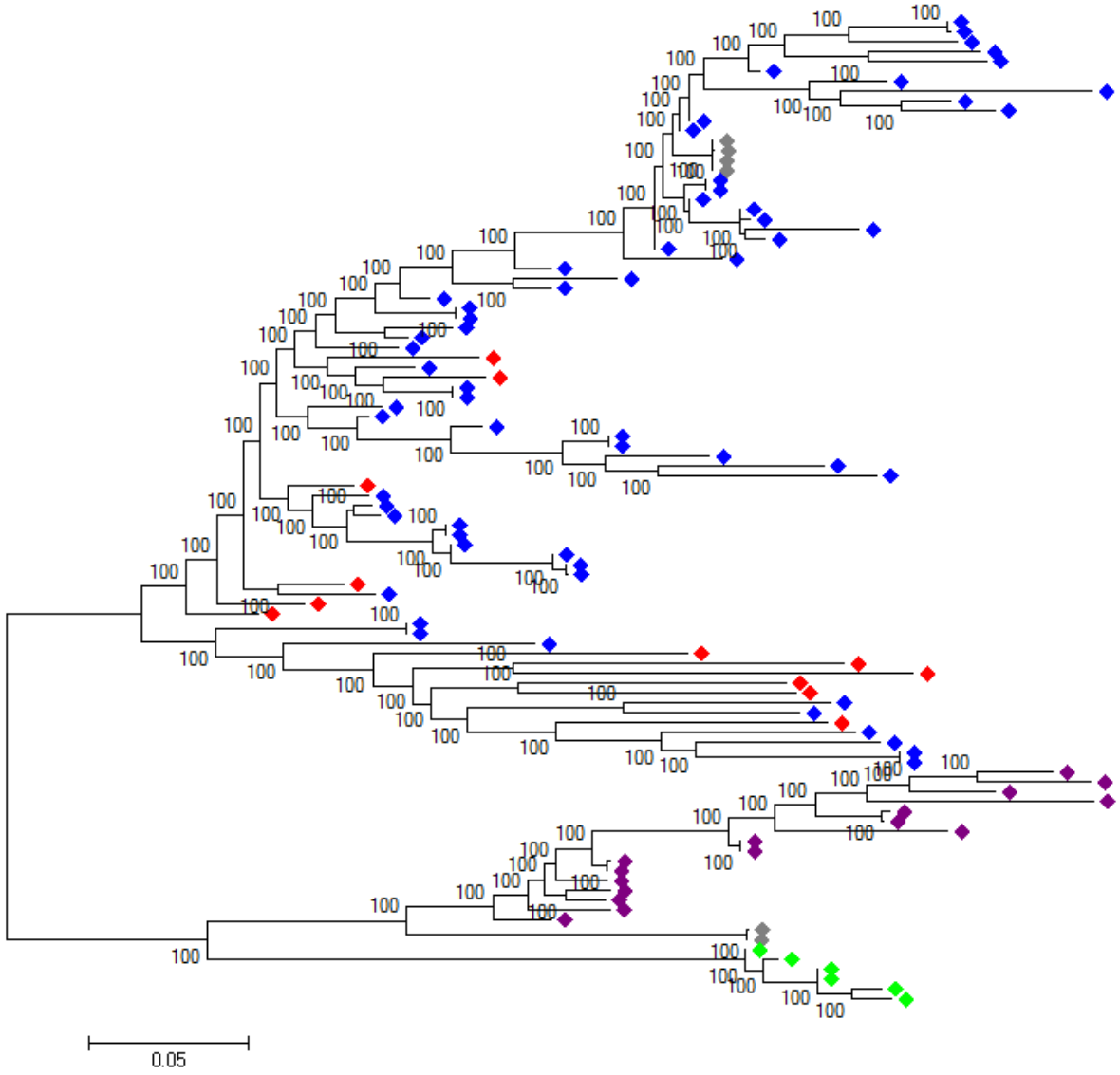


Figure S7.248 Chromosome 6: 4465001 - 4525001

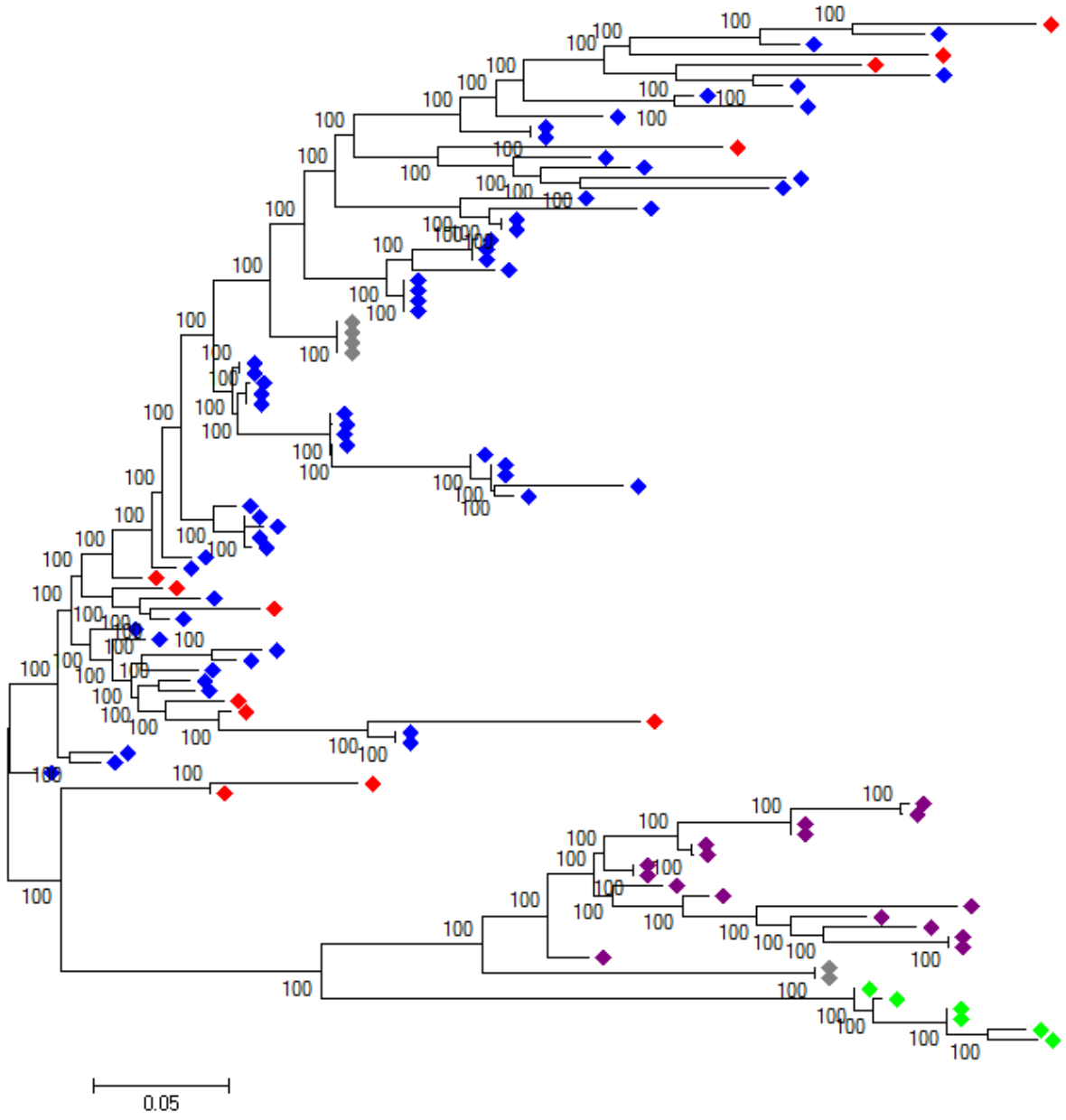


Figure S7.249 Chromosome 6: 4930001 - 4950001

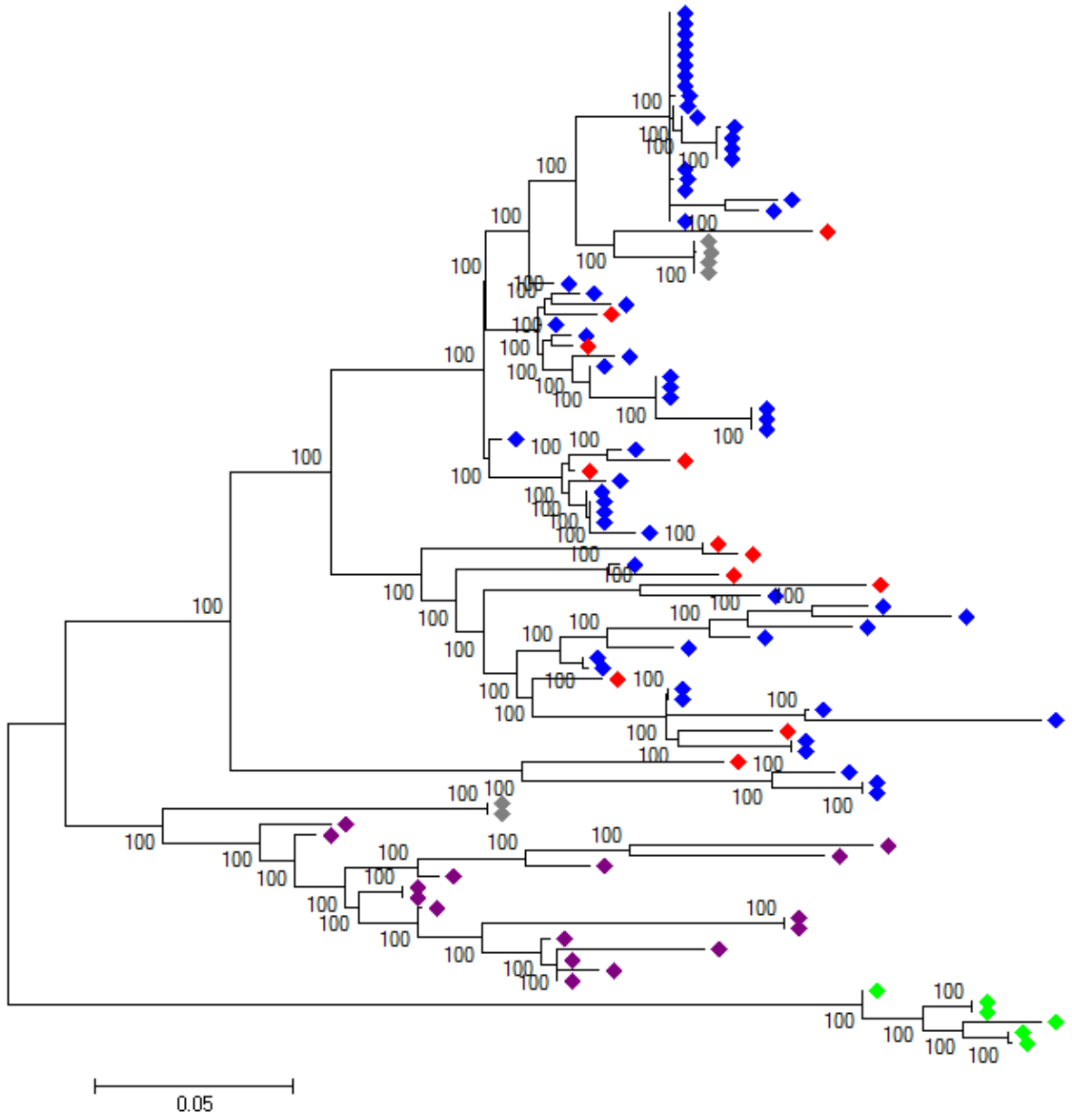


Figure S7.250 Chromosome 6: 6150001 - 6175001

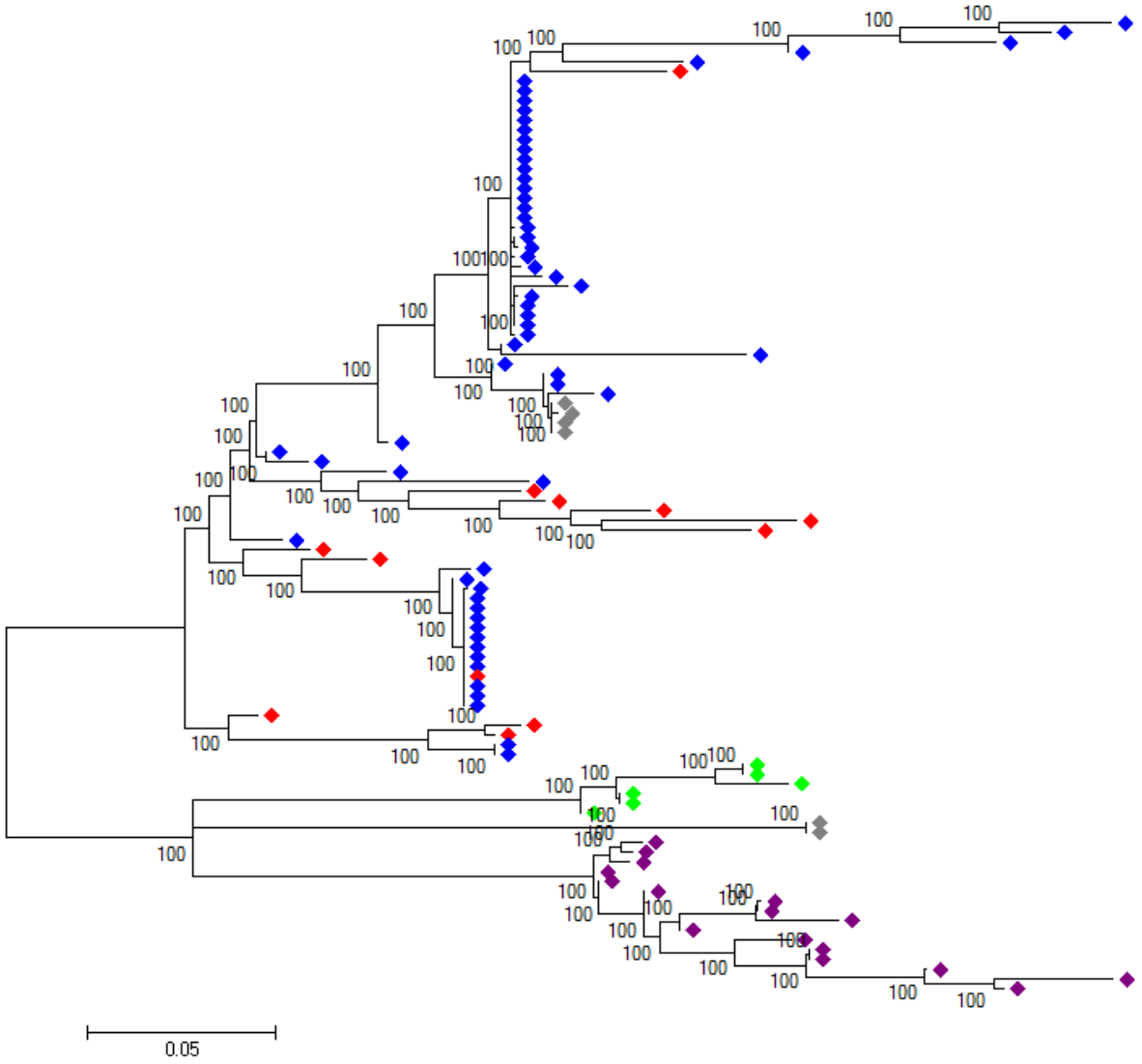


Figure S7.251 Chromosome 6: 6525001 - 6550001

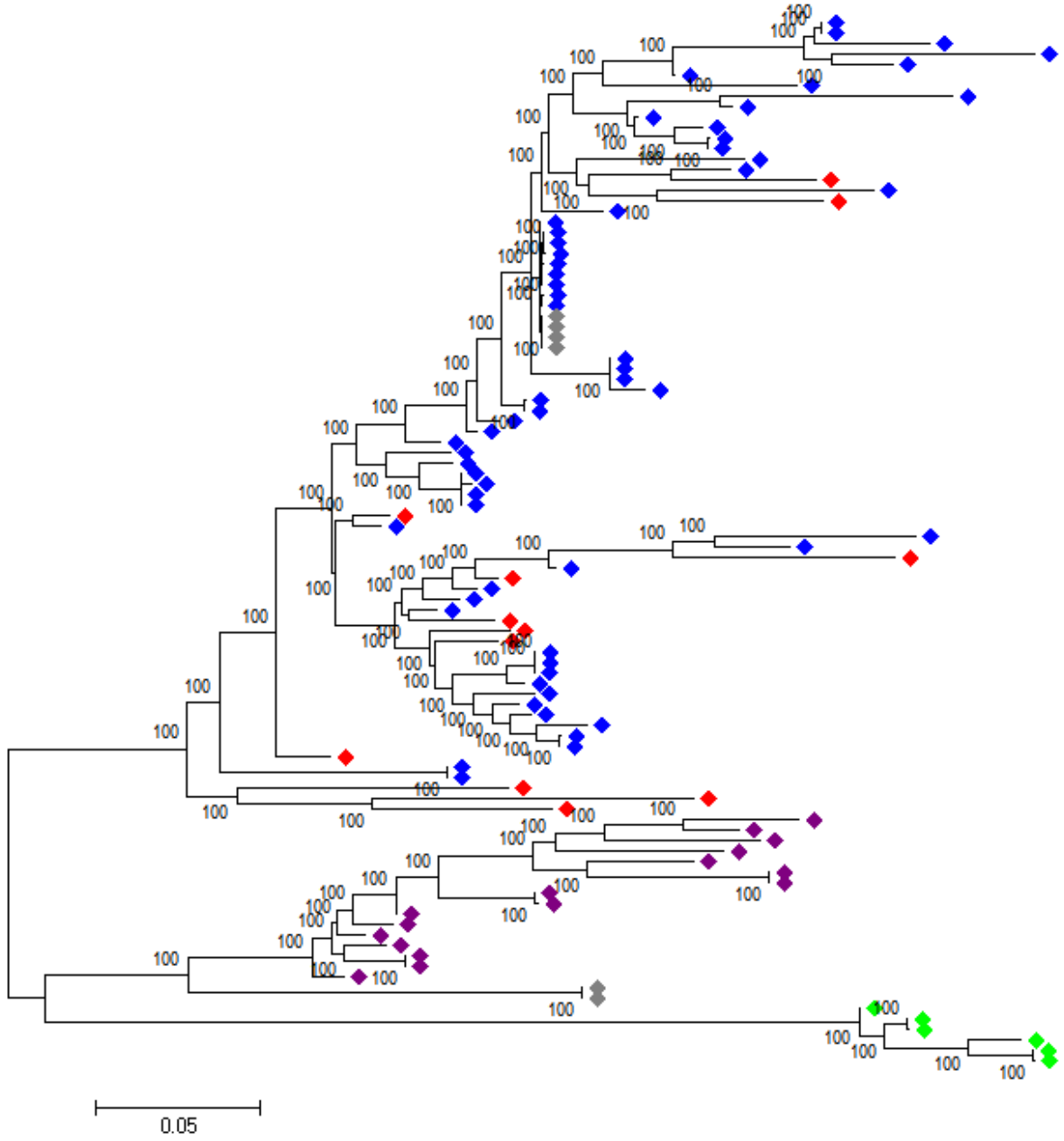


Figure S7.252 Chromosome 6: 6660001 - 6765001

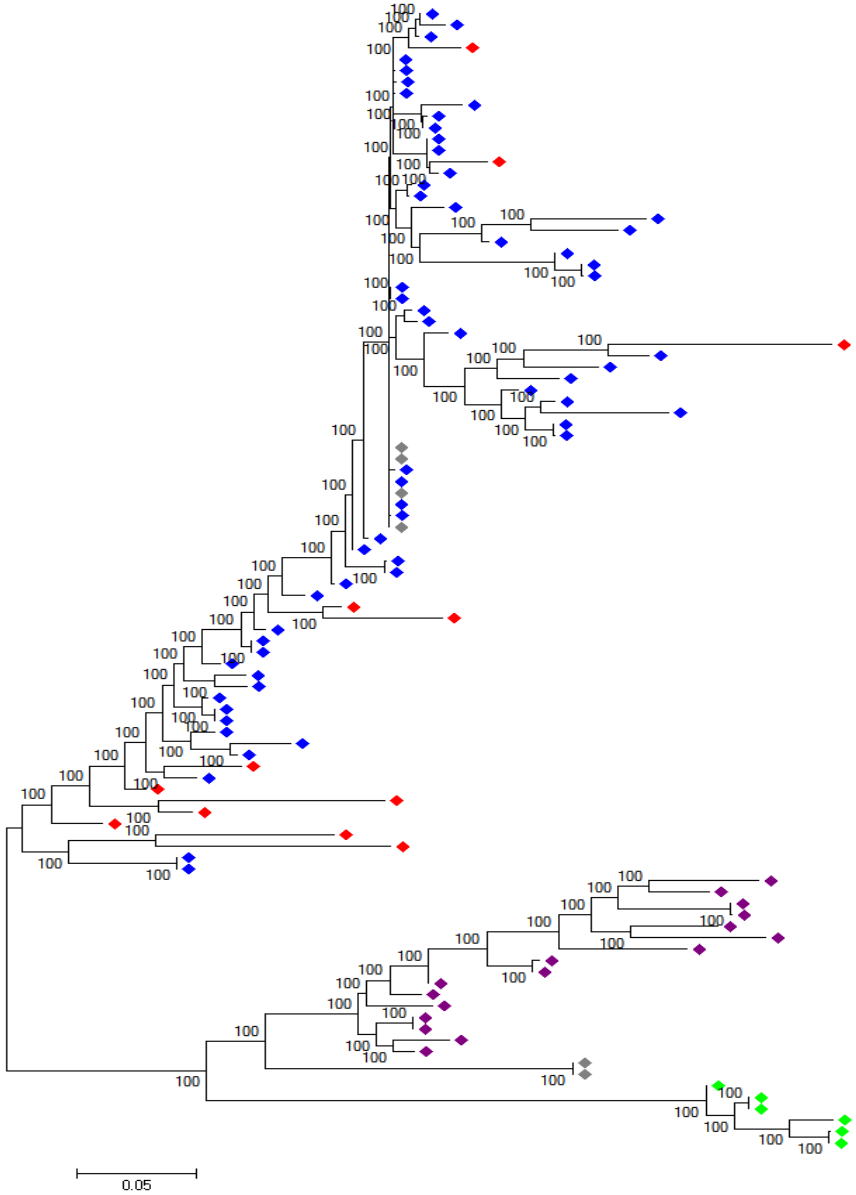


Figure S7.253 Chromosome 6: 6855001 - 6915001

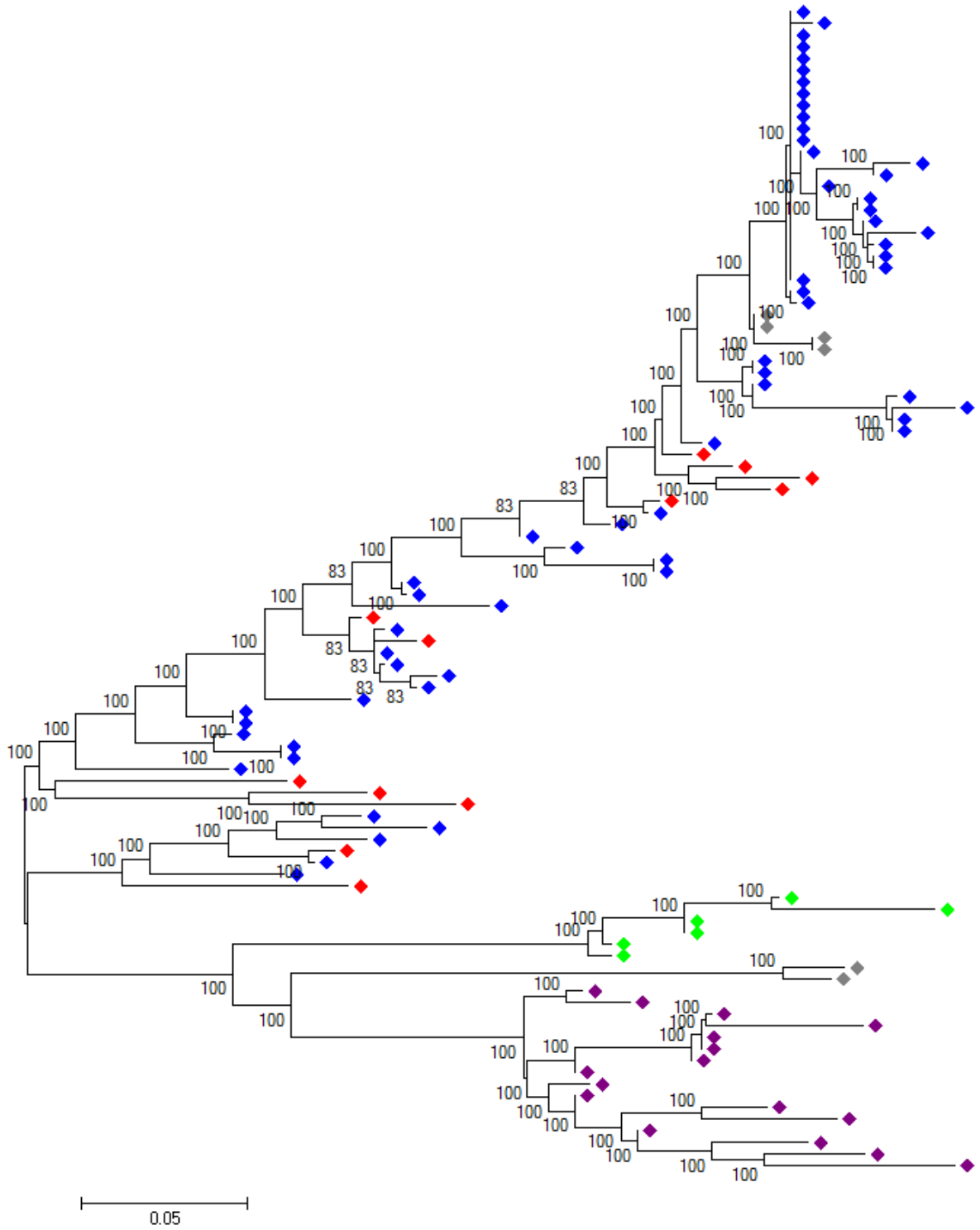


Figure S7.354 Chromosome 8: 22575001 - 22595001

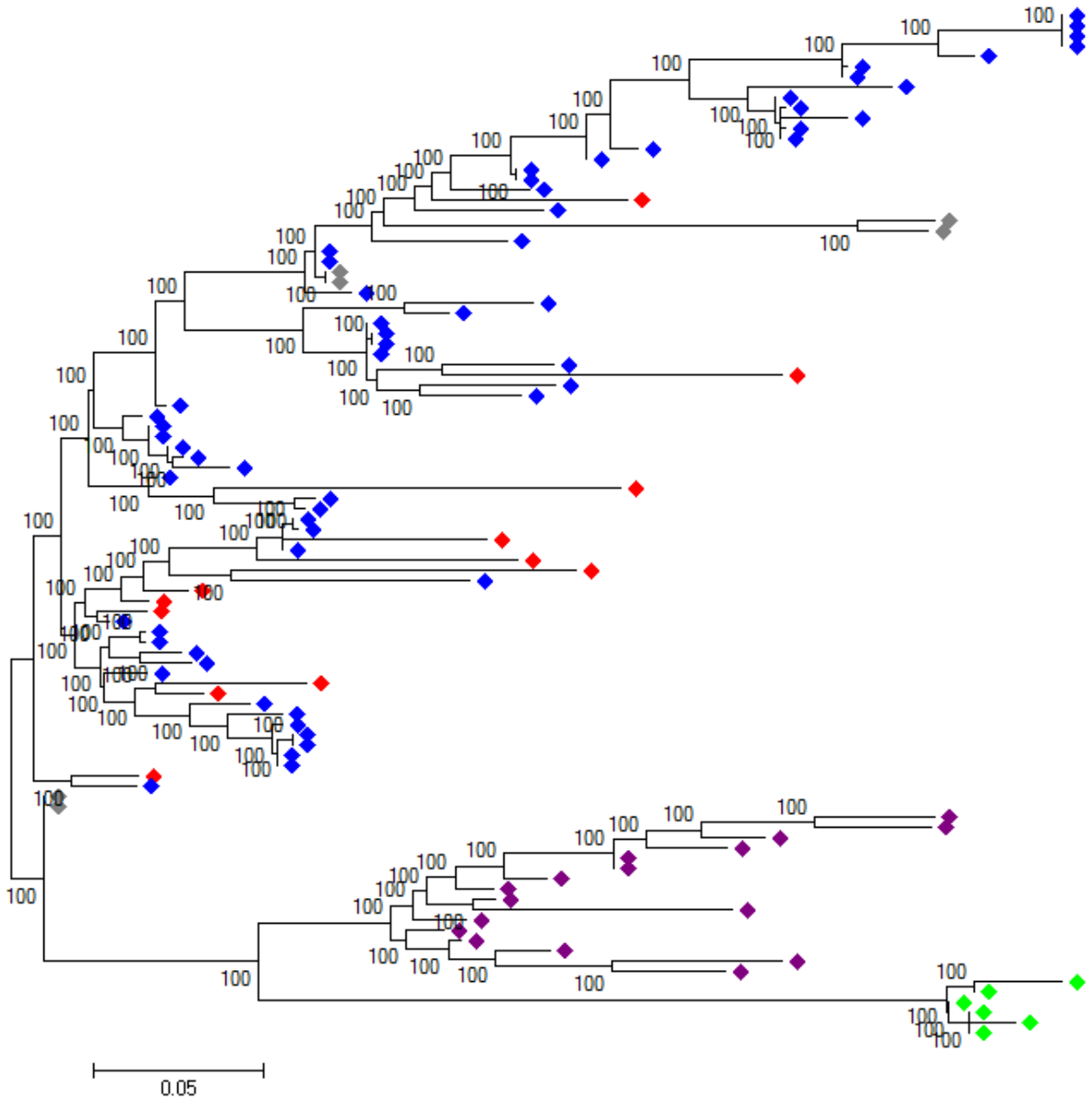


Figure S7.255 Chromosome 8: 22700001 - 22720001

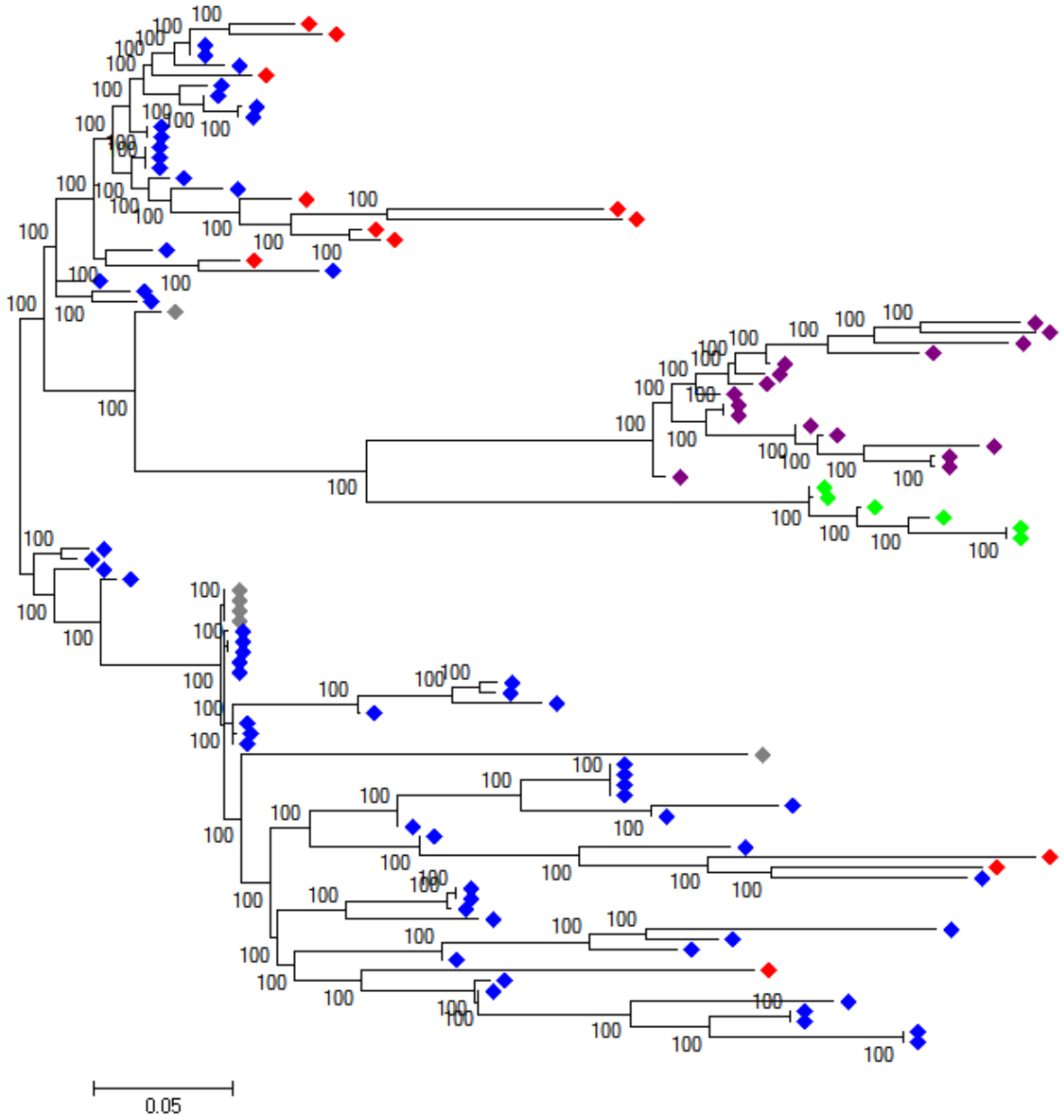


Figure S7.256 Chromosome 8: 27525001 - 27545001

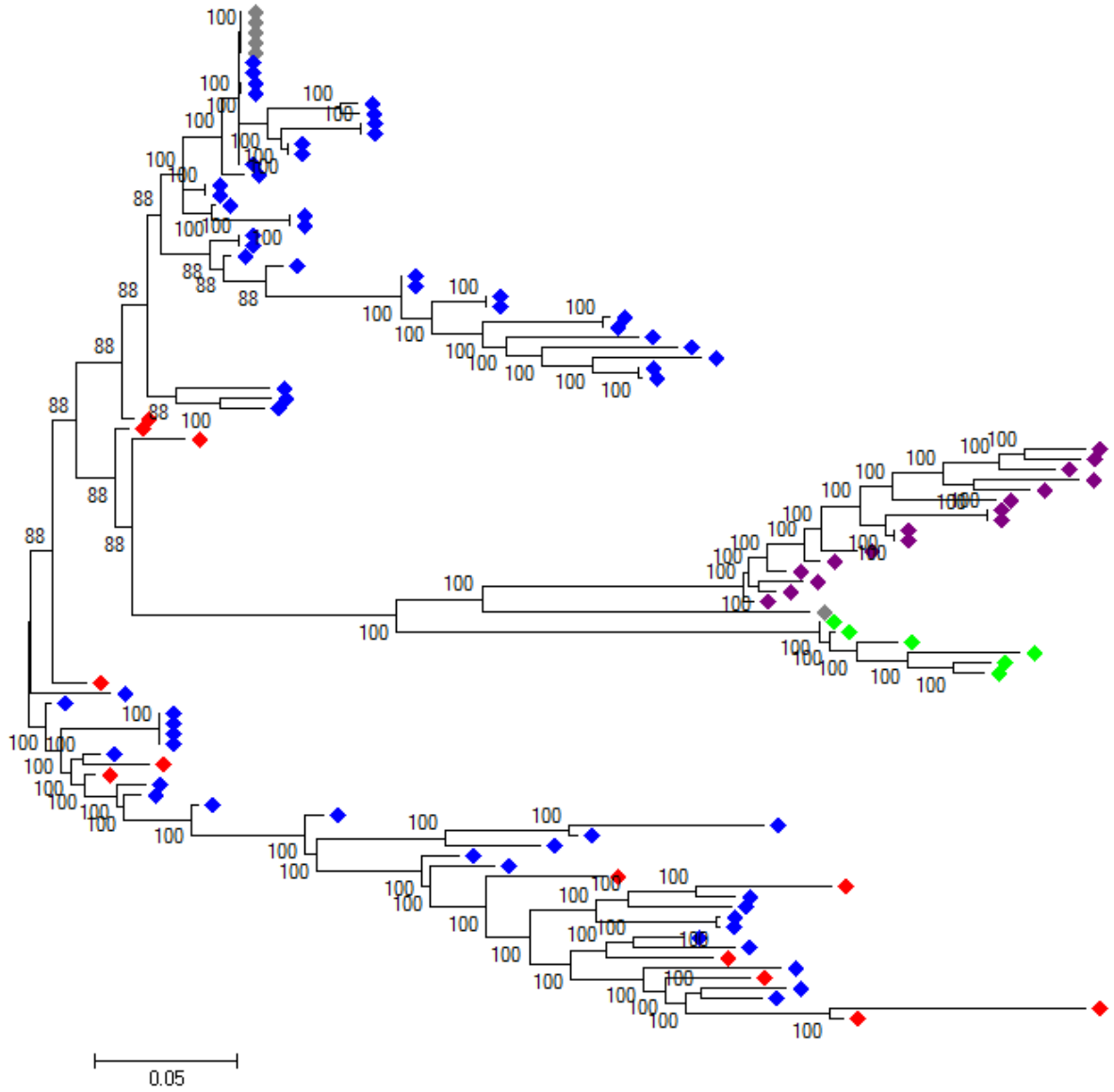


Figure S7.257 Chromosome 11: 7030001 - 705000

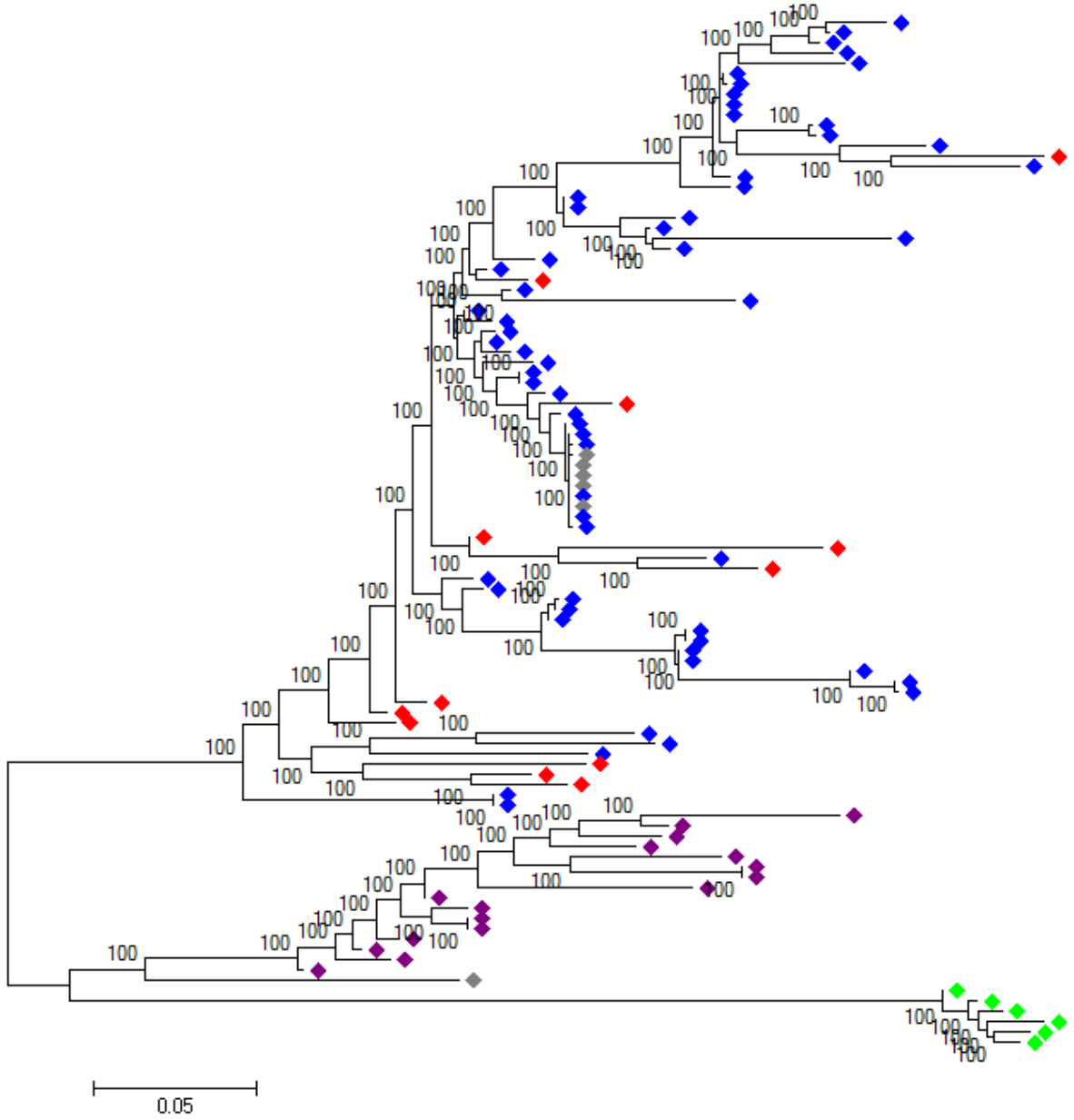


Figure S7.258 Chromosome 11: 7385001 - 7405001

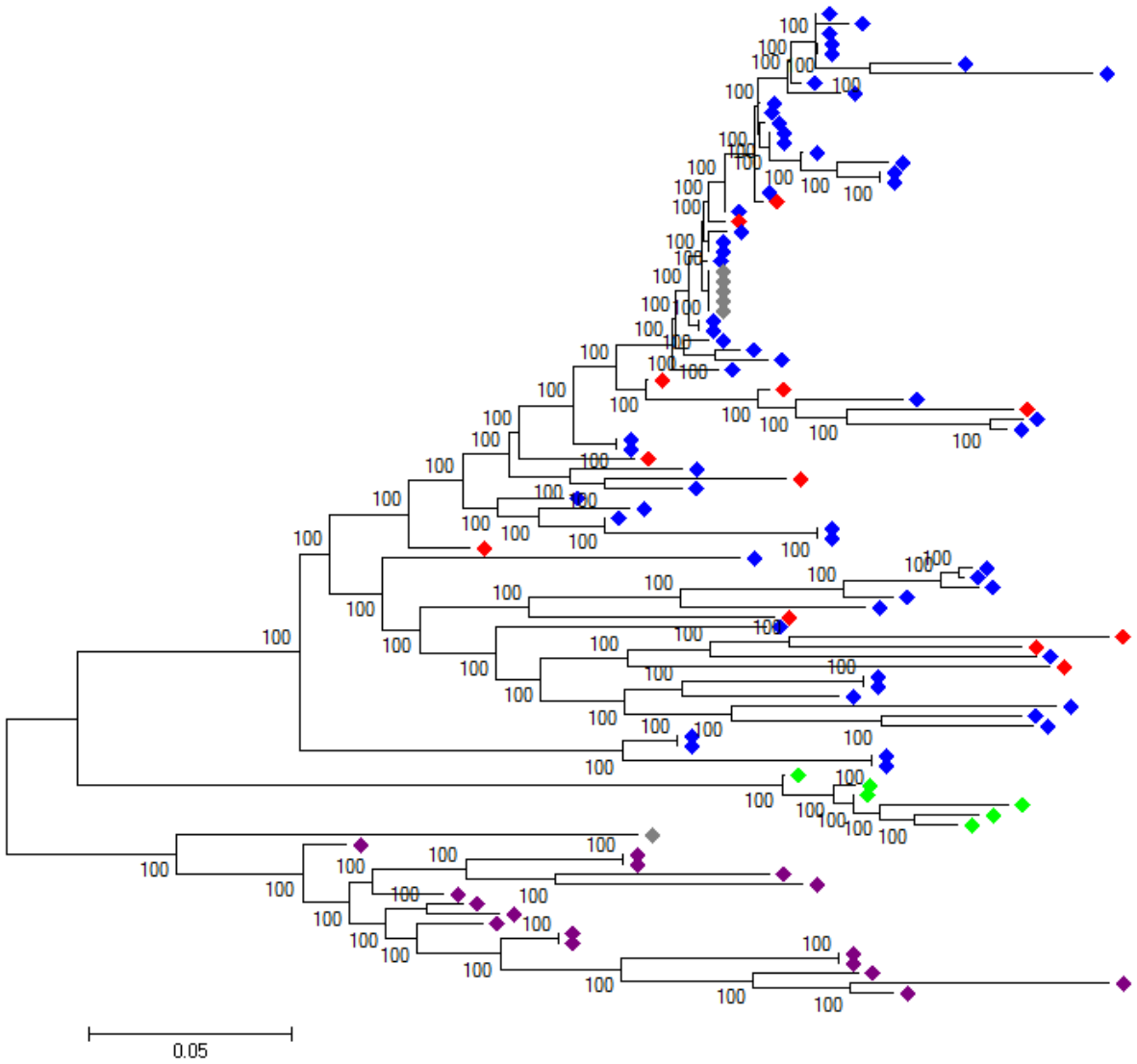


Figure S7.259 Chromosome 11: 7645001 - 7680001

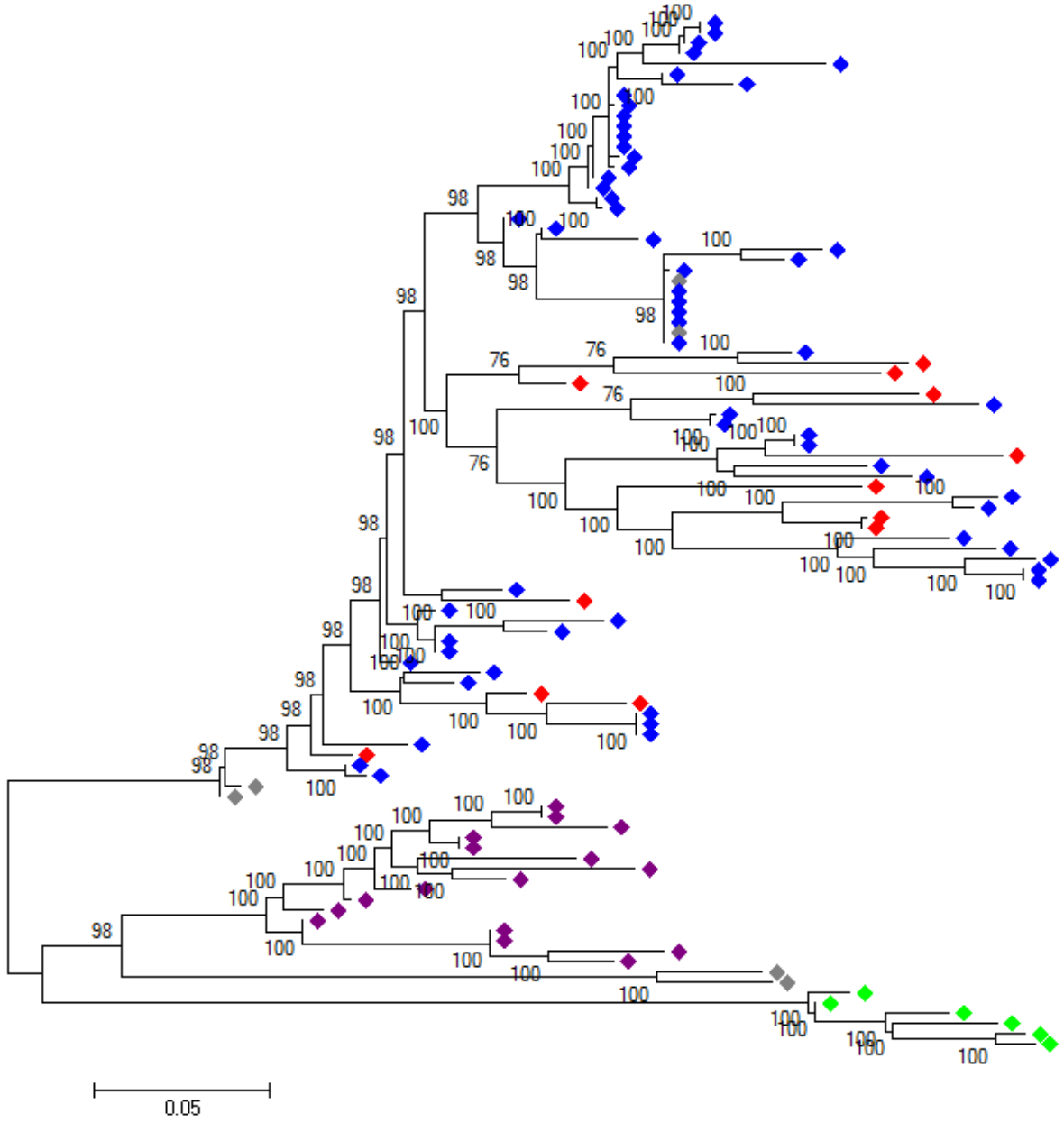


Figure S7.260 Chromosome 11: 9175001 - 9195001

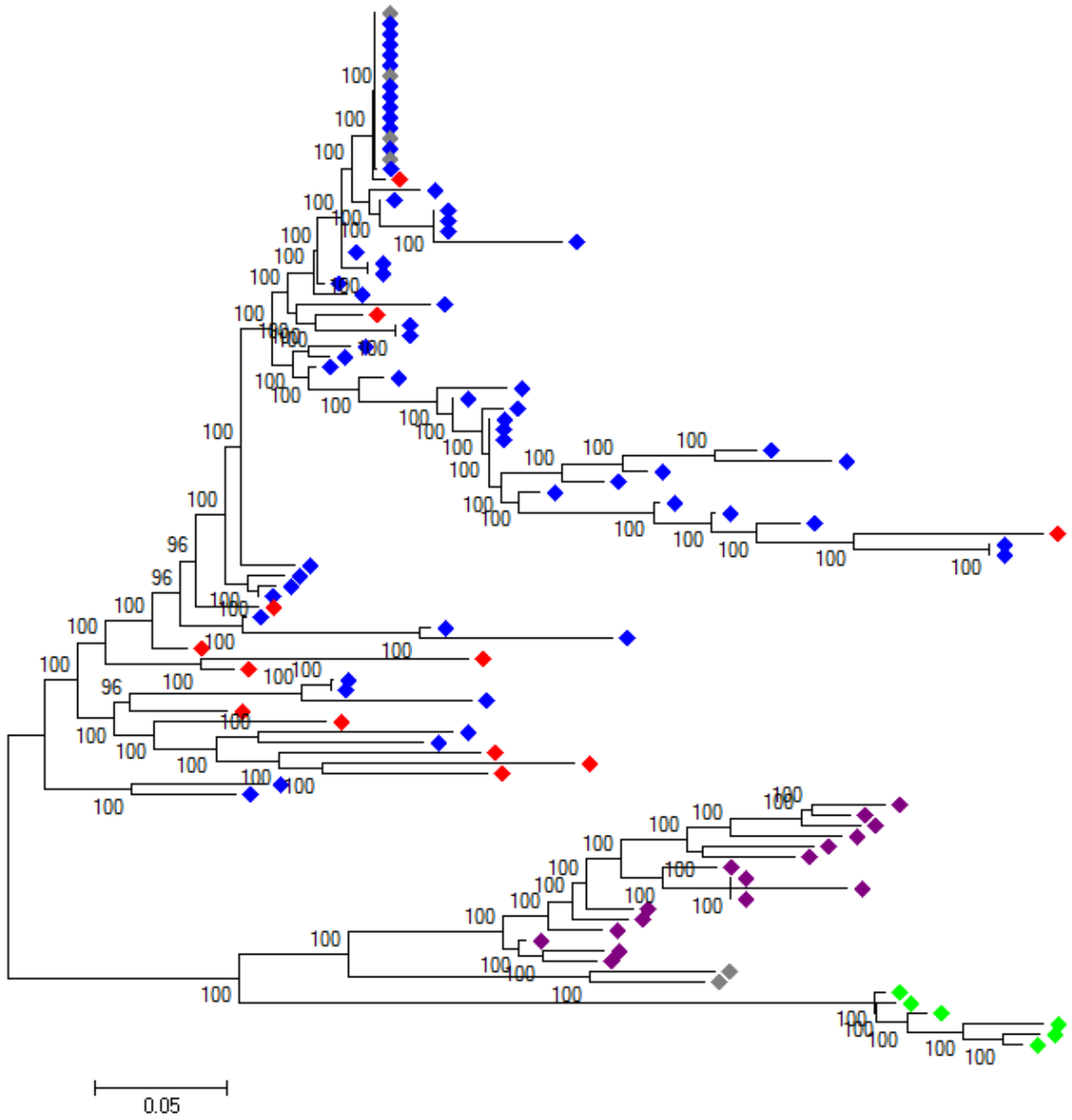


Figure S7.261 Chromosome 11: 9355001 - 9395001

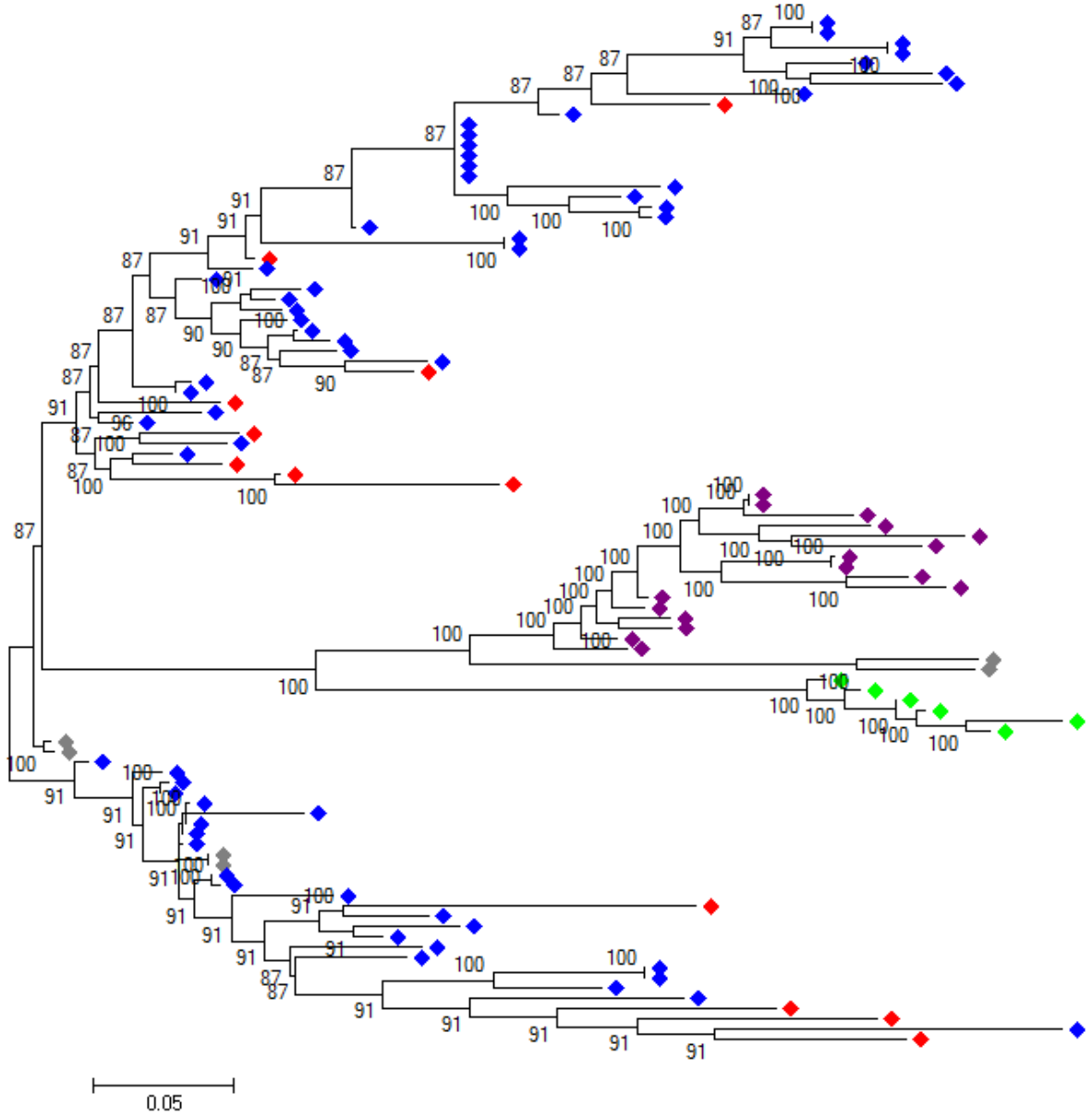


Figure S7.262 Chromosome 11: 9780001 - 9800001

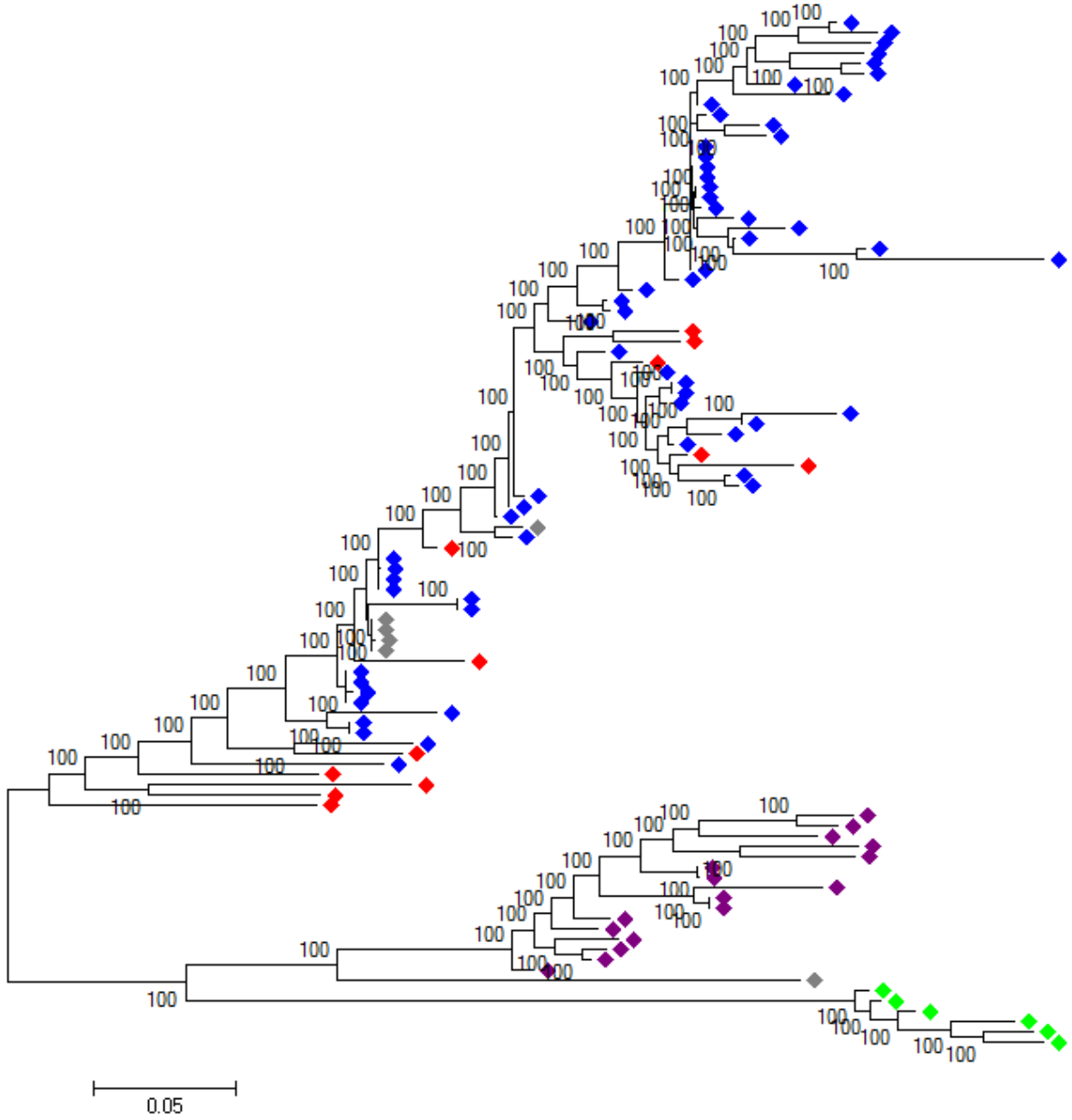


Figure S7.263 Chromosome 11: 10520001 - 10585001

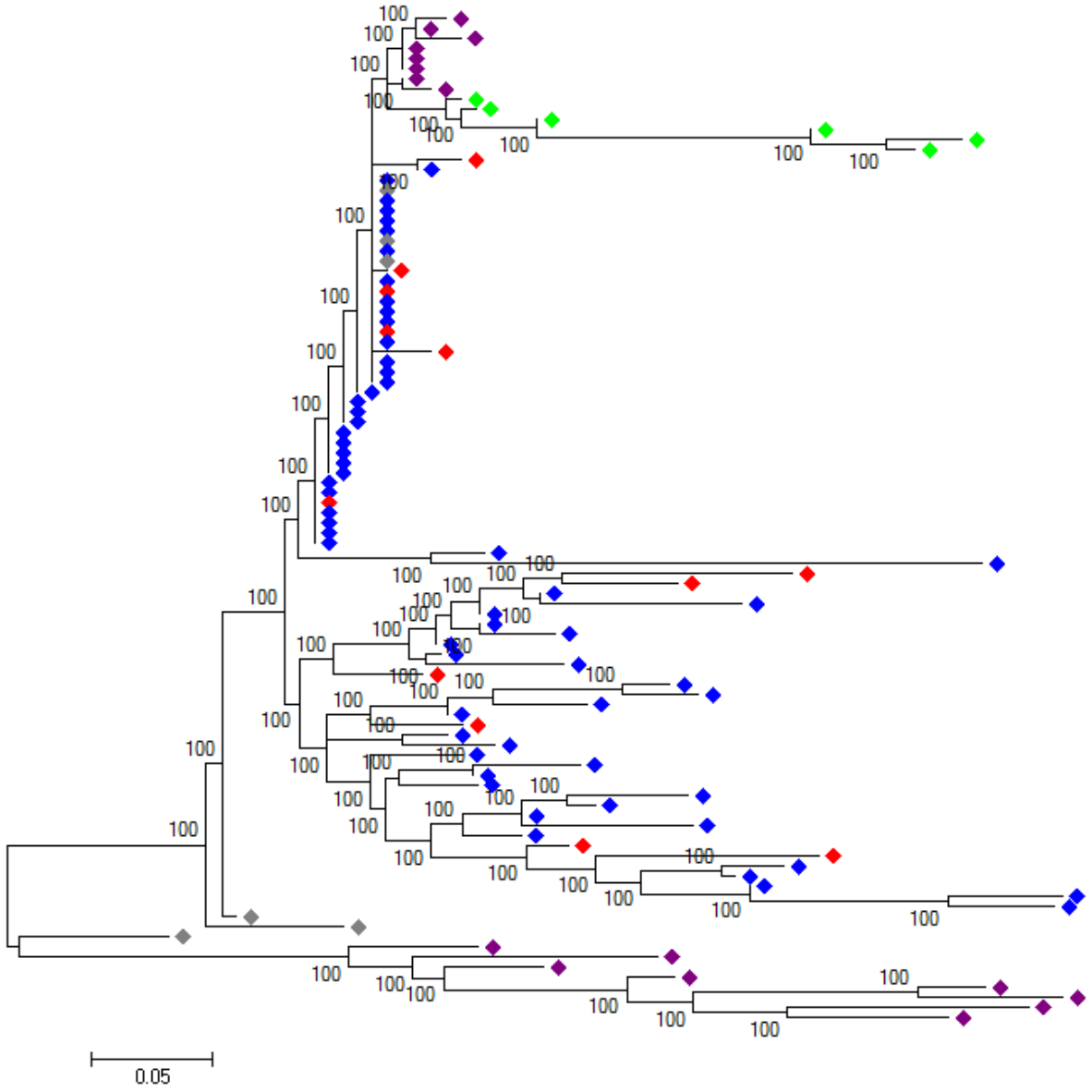


Figure S7.264 Chromosome 13: 45001 - 70001

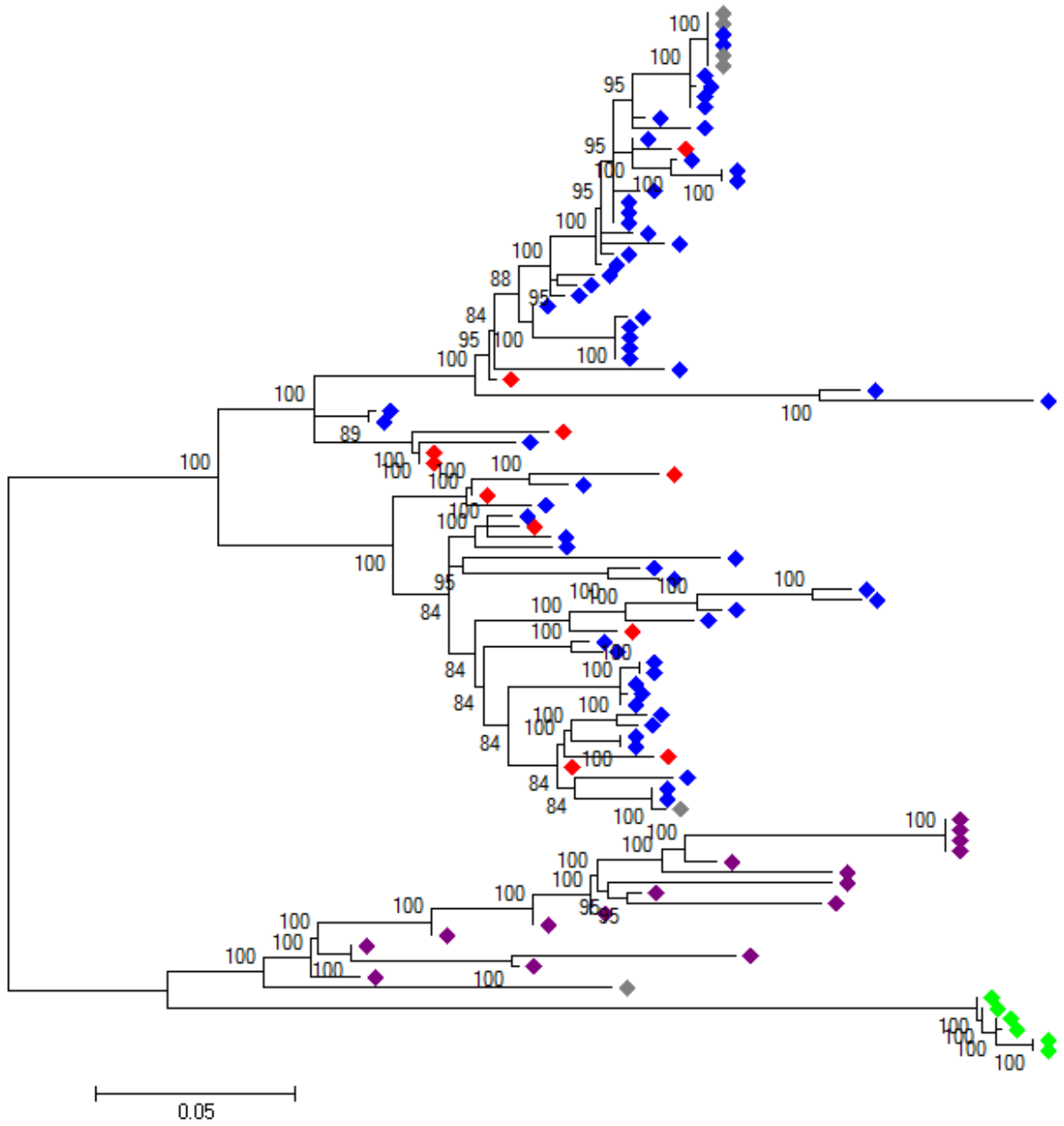


Figure S7.265 Chromosome 13: 150001 - 170001

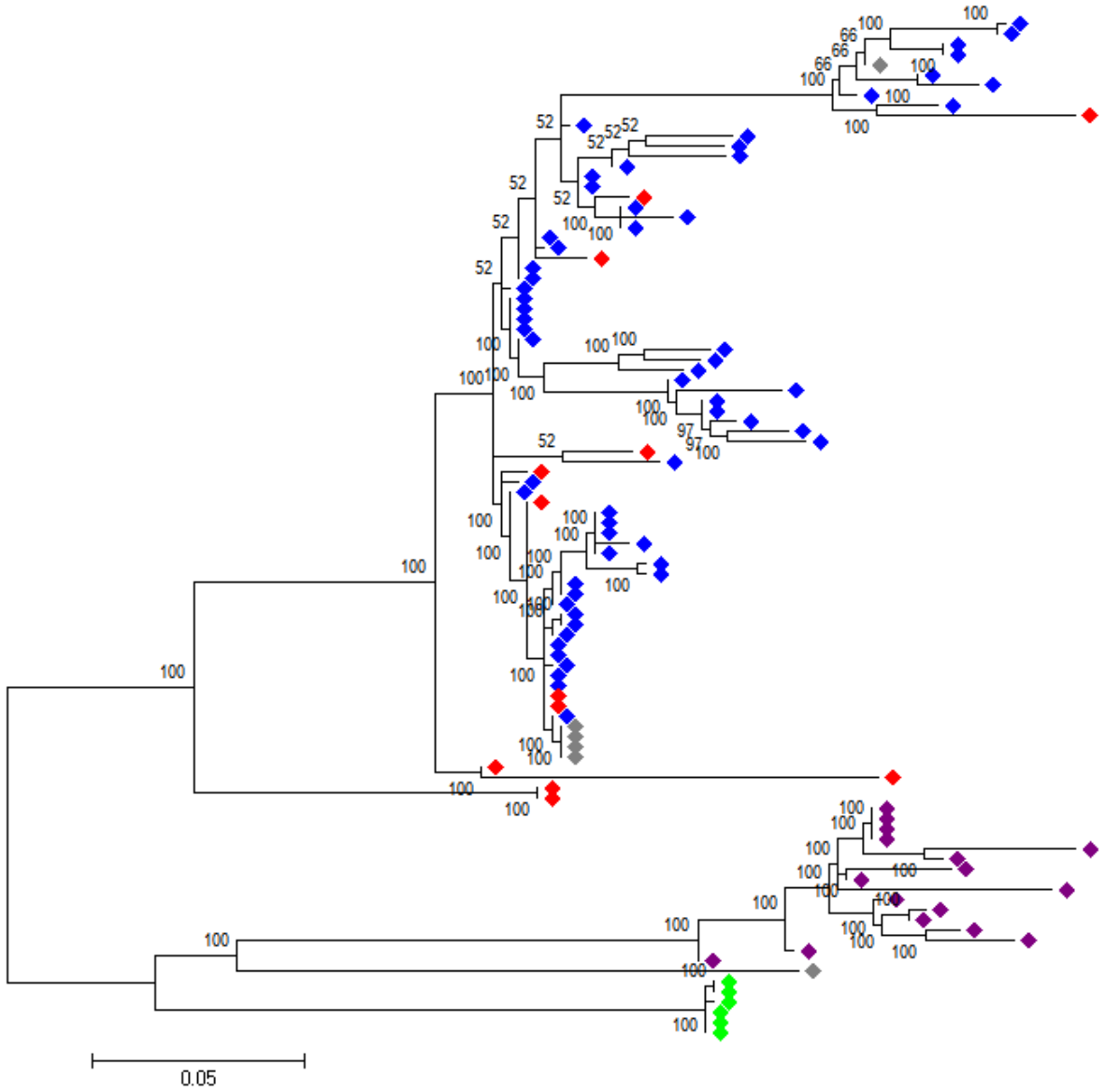


Figure S7.266 Chromosome 13: 235001 - 255001

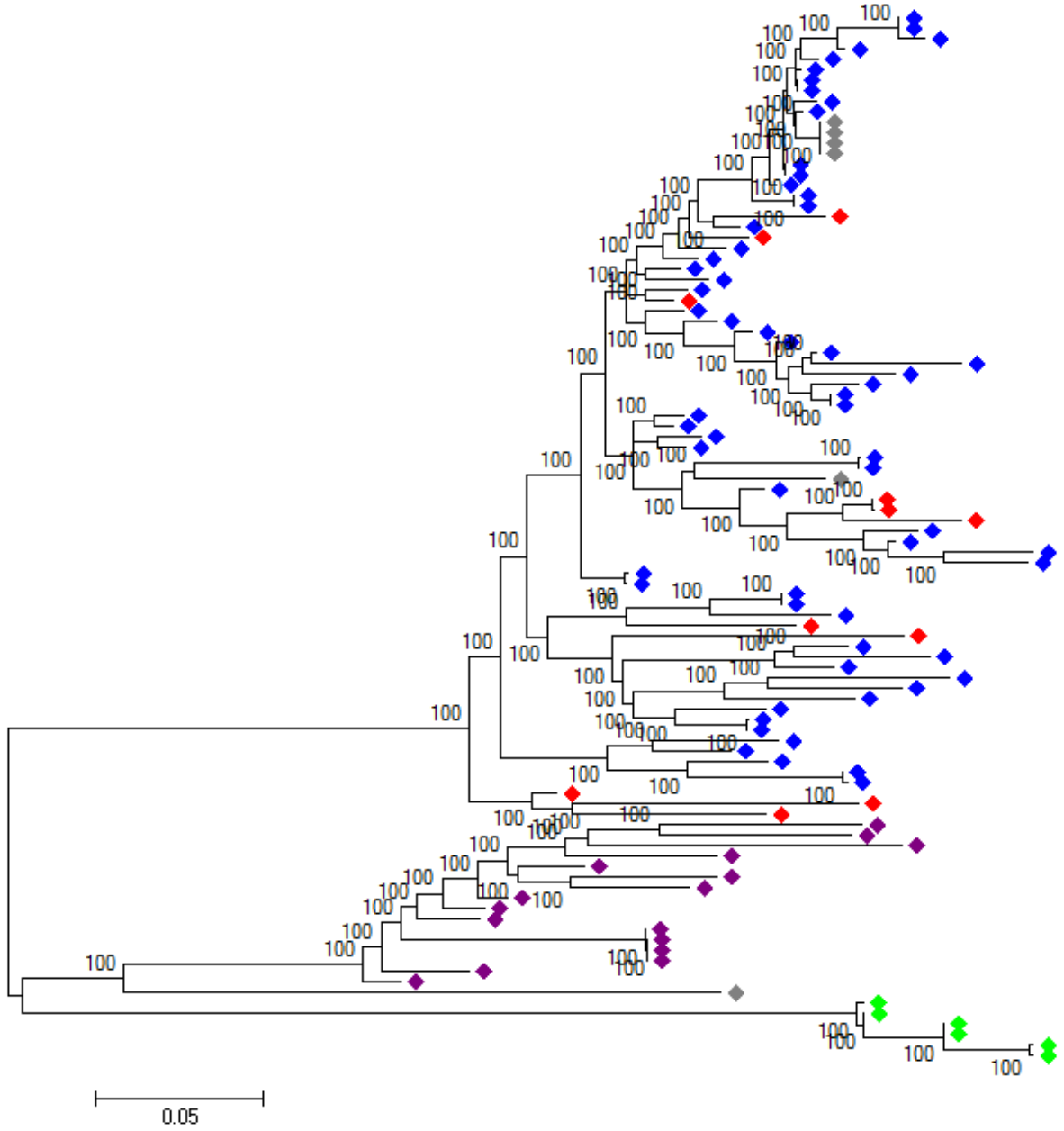


Figure S7.267 Chromosome 13: 430001 - 490001

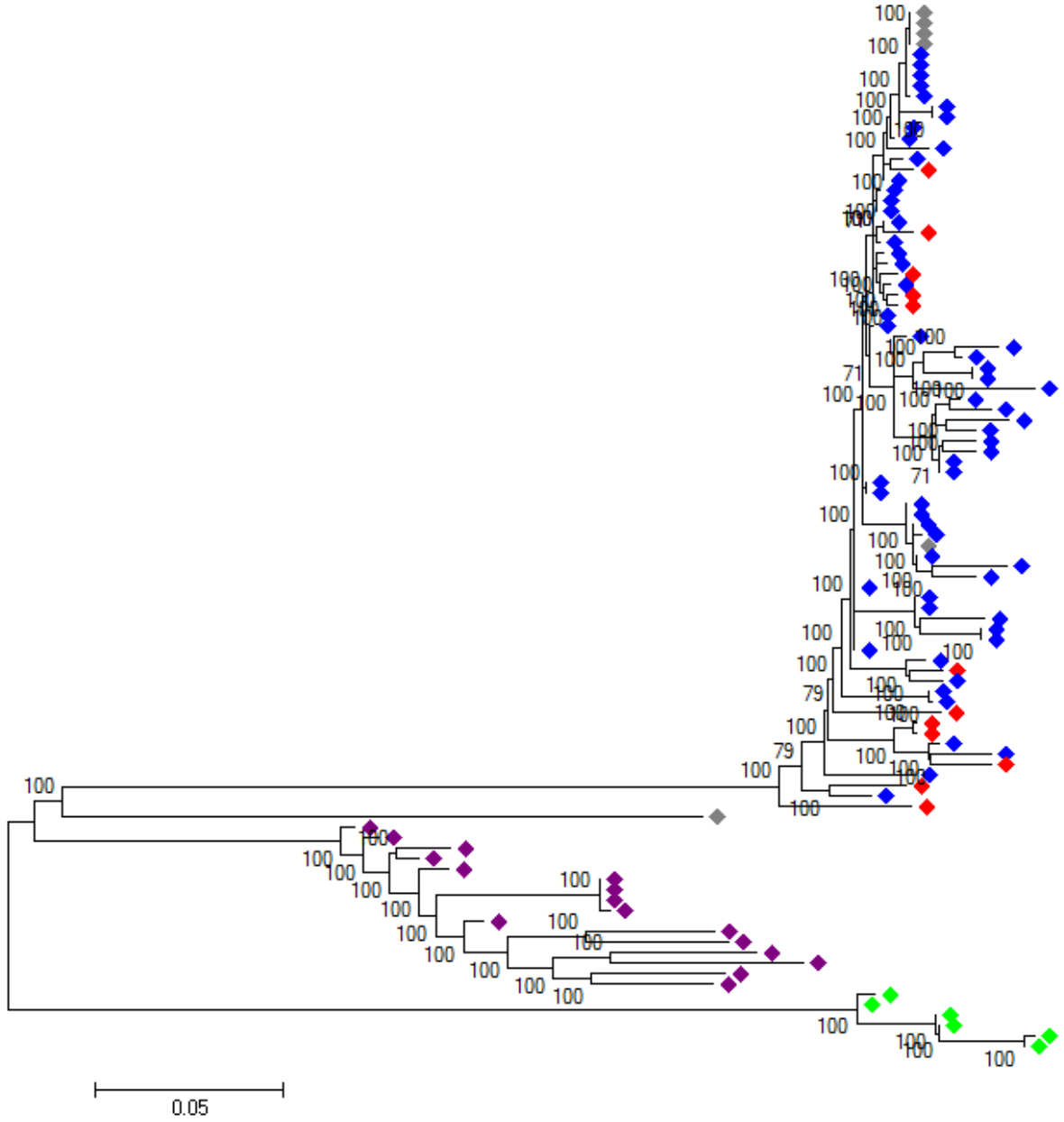


Figure S7.268 Chromosome 13: 610001 - 645001

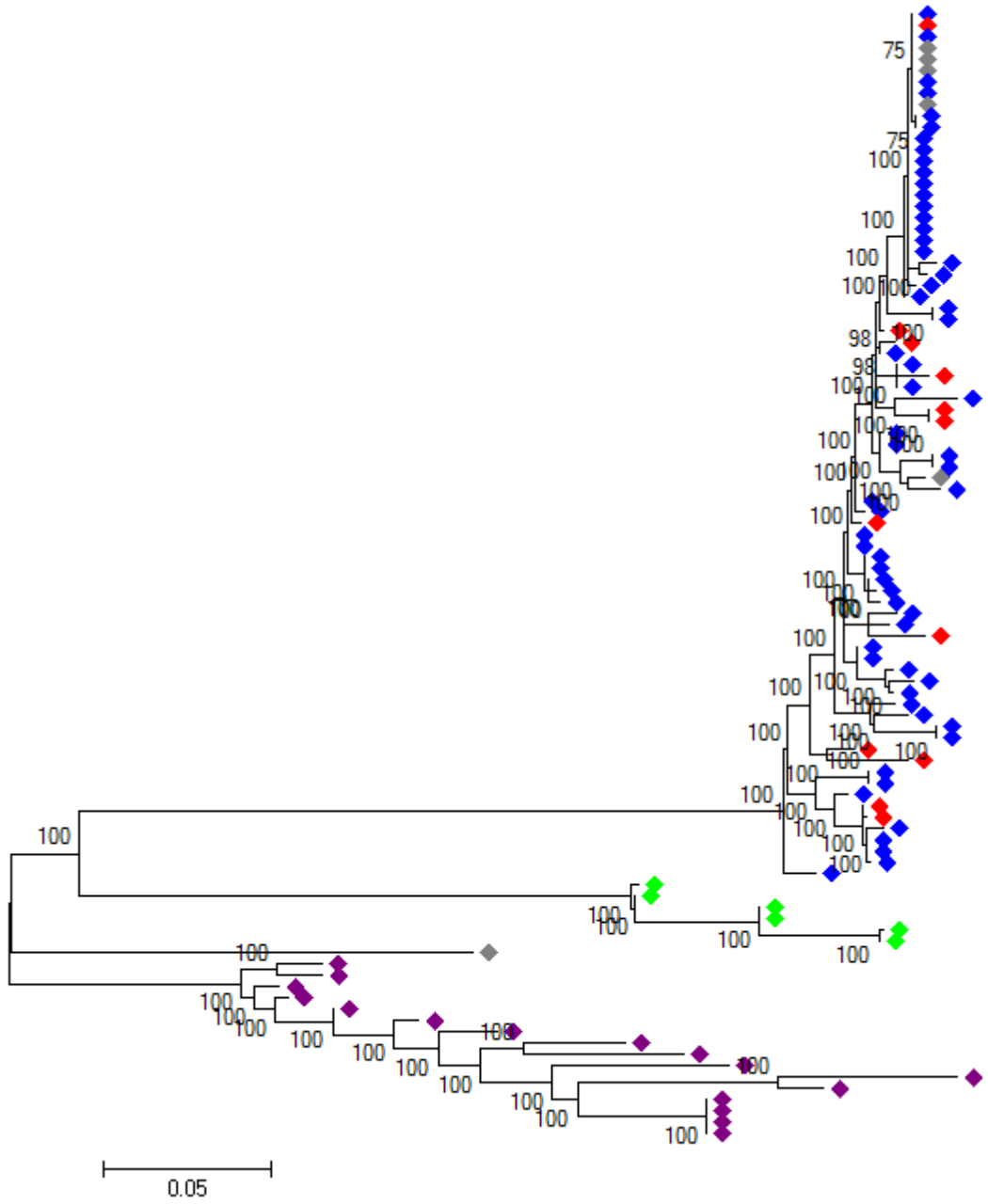


Figure S7.269 Chromosome 13: 780001 - 820001

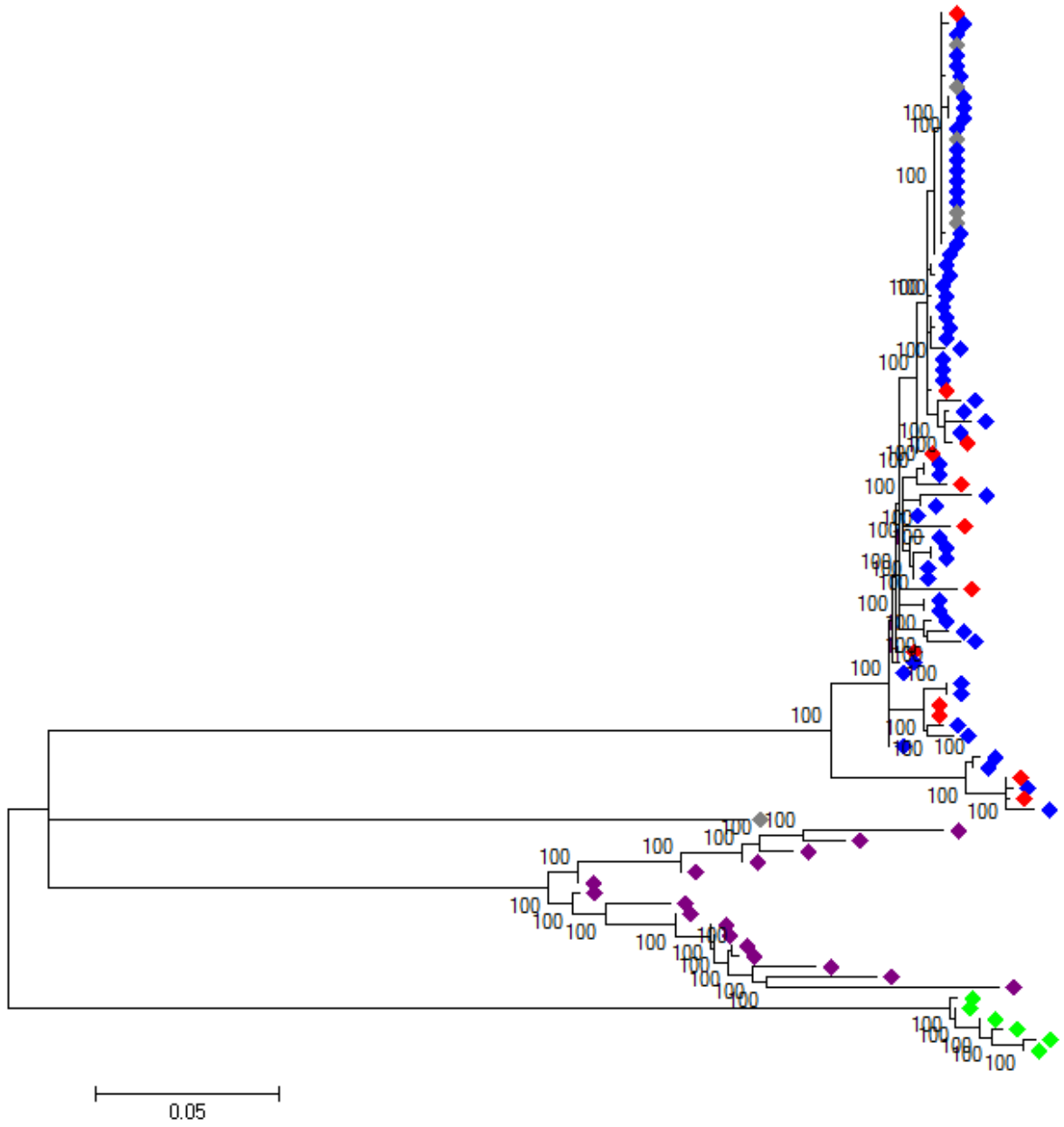


Figure S7.270 Chromosome 13: 910001 - 965001

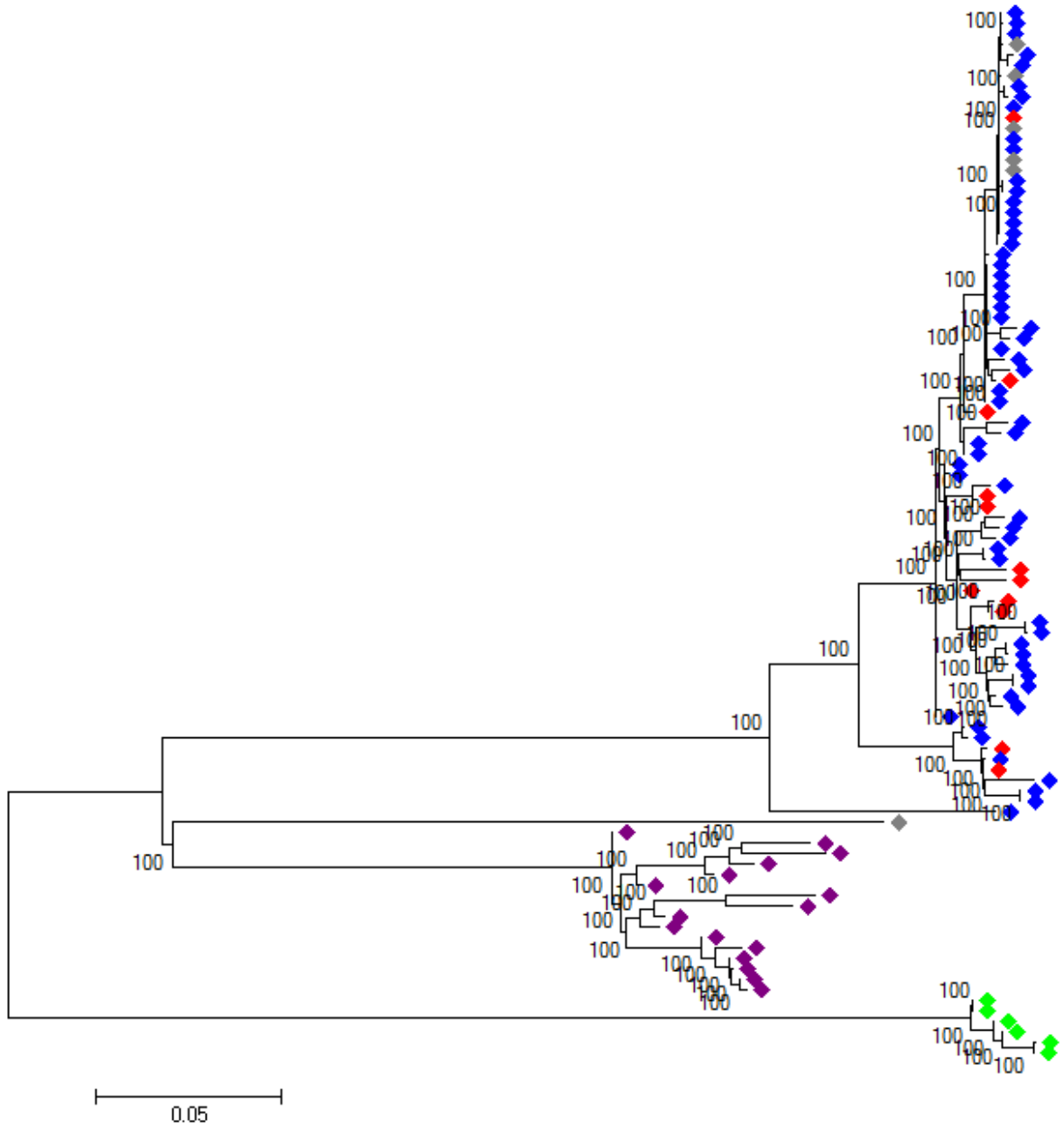


Figure S7.271 Chromosome 13: 975001 - 1090001

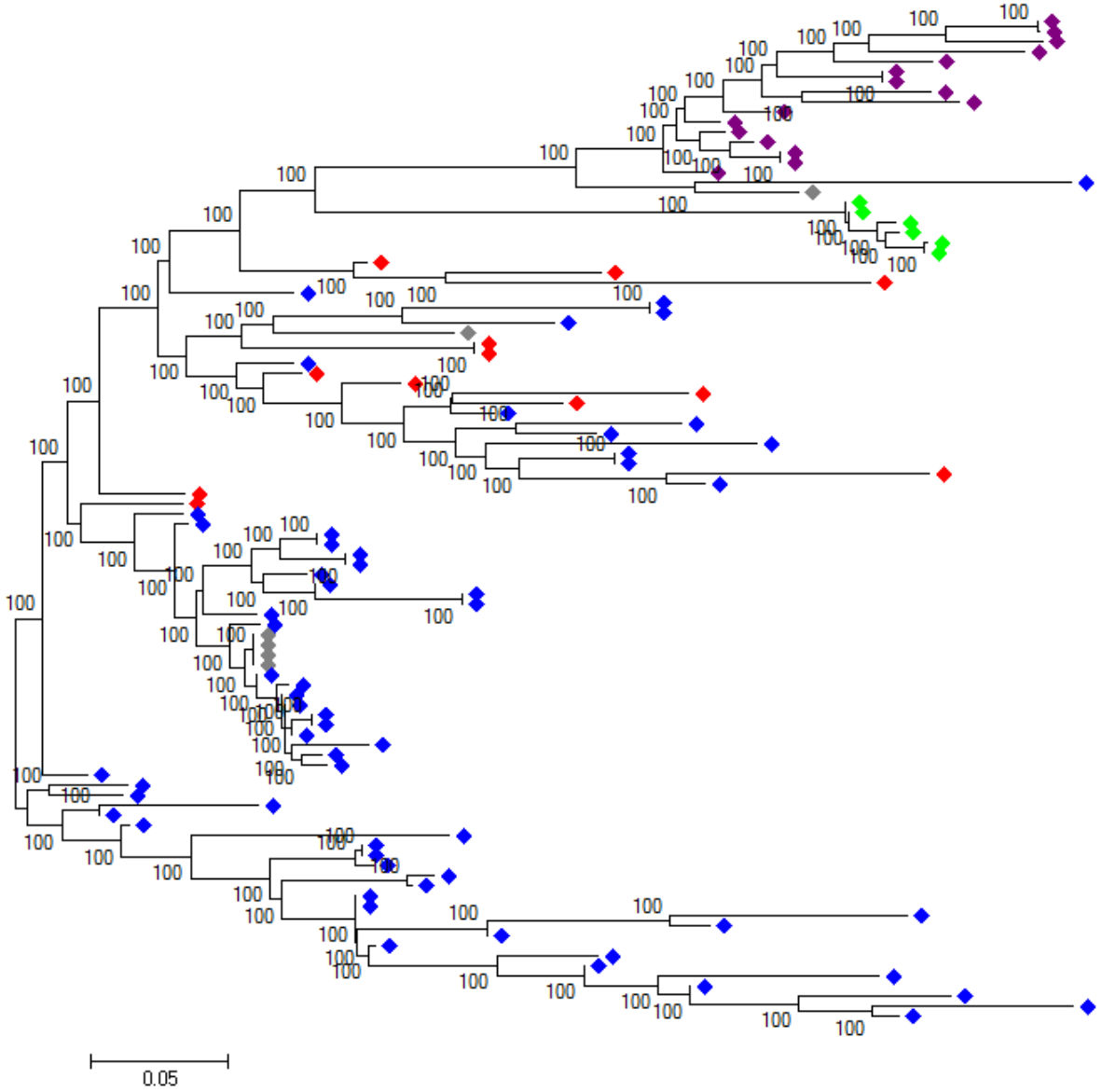


Figure S7.272 Chromosome 13: 2275001 - 2295001

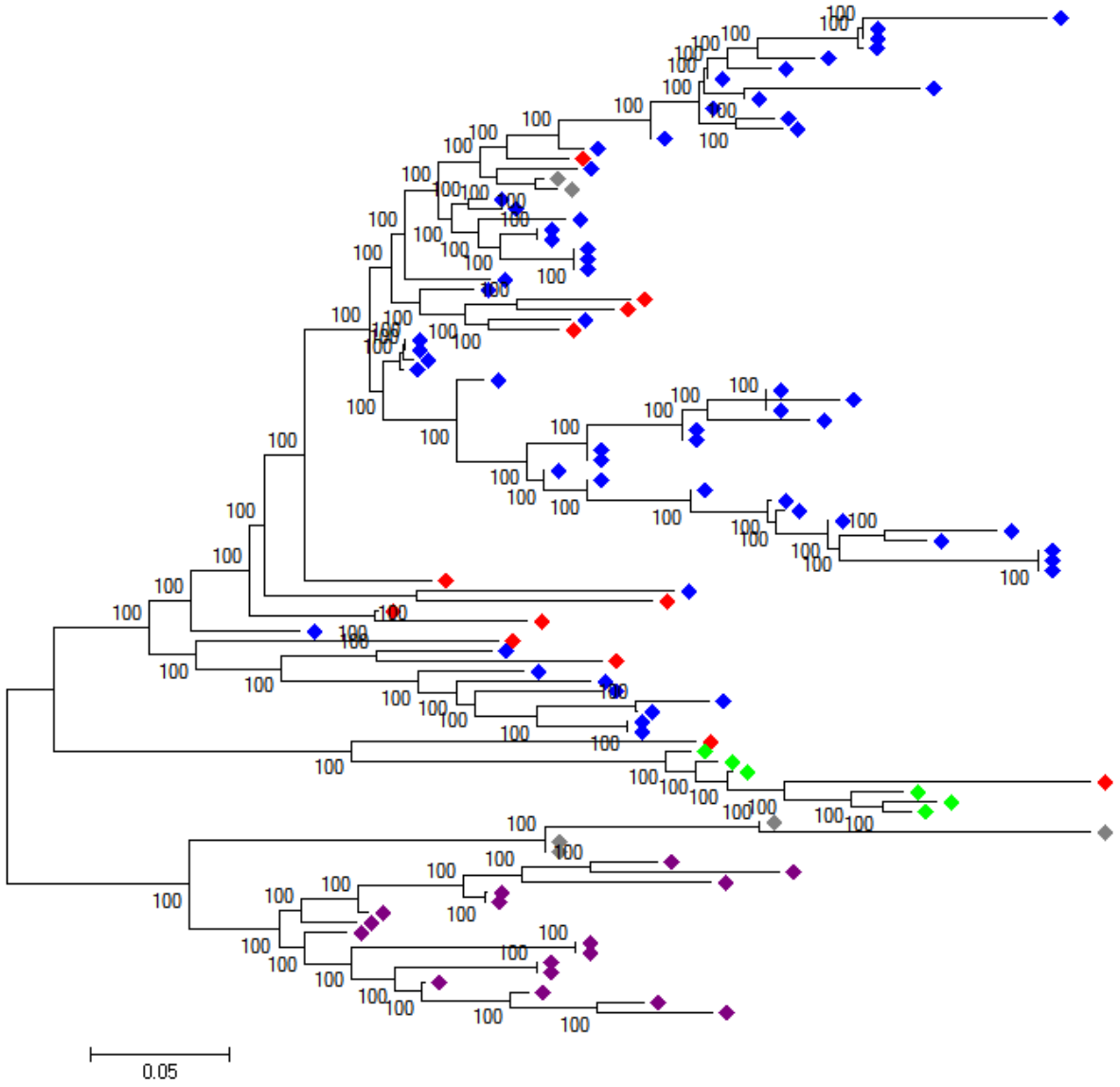


Figure S7.273 Chromosome 17: 5410001 - 5450001

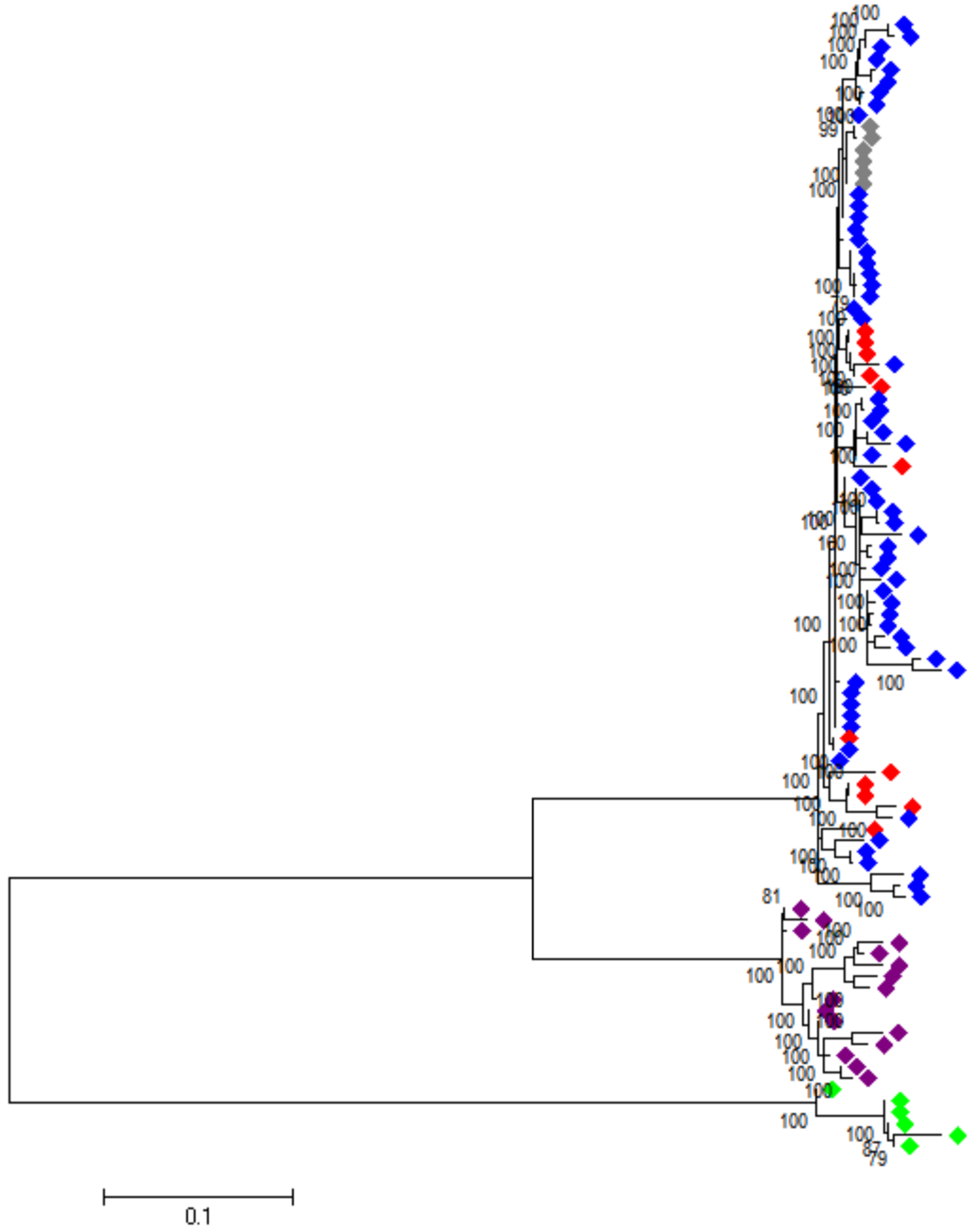


Figure S7.274 Chromosome 18: 95001 - 145001

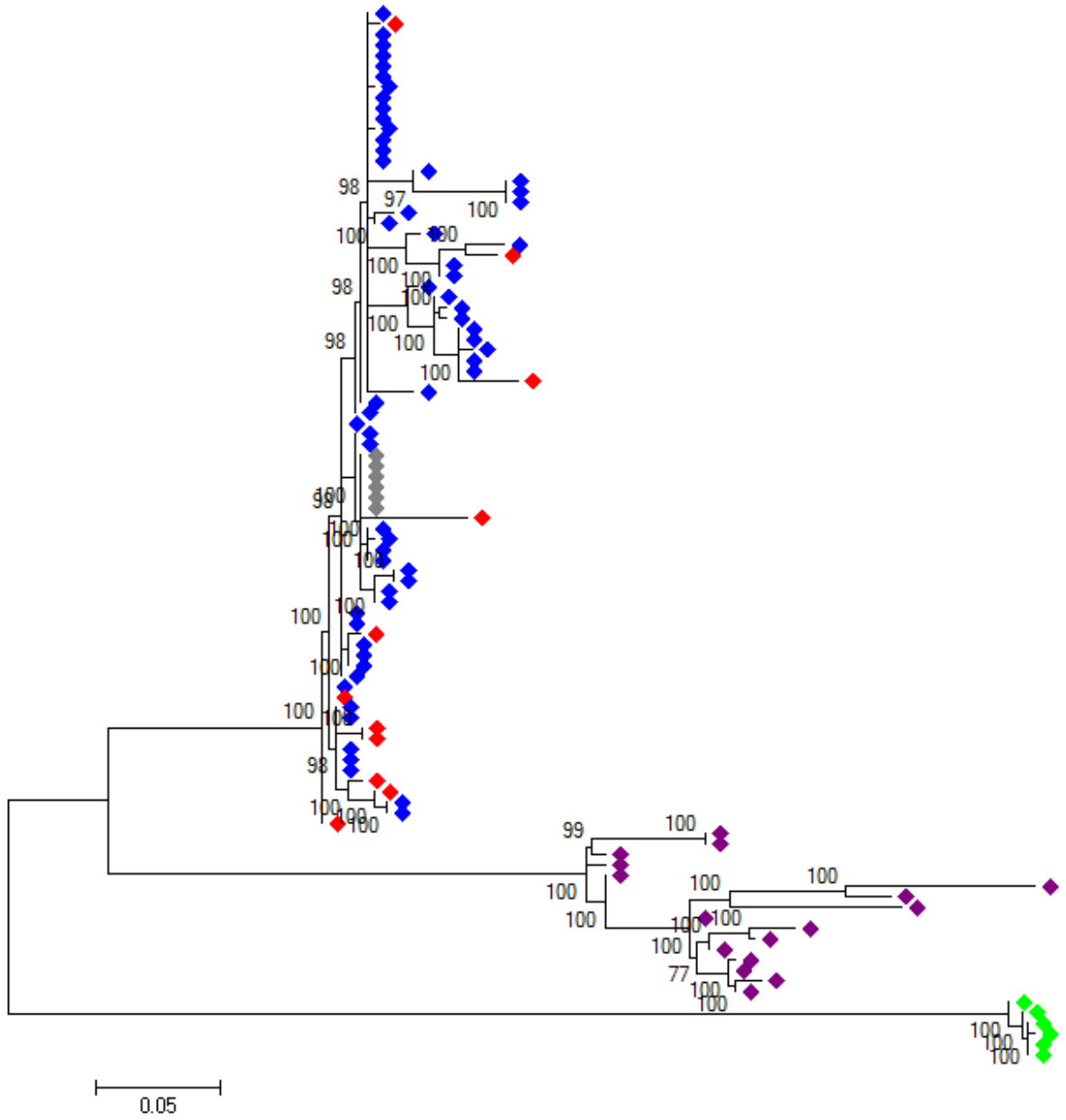


Figure S7.275 Chromosome 18: 180001 - 200001

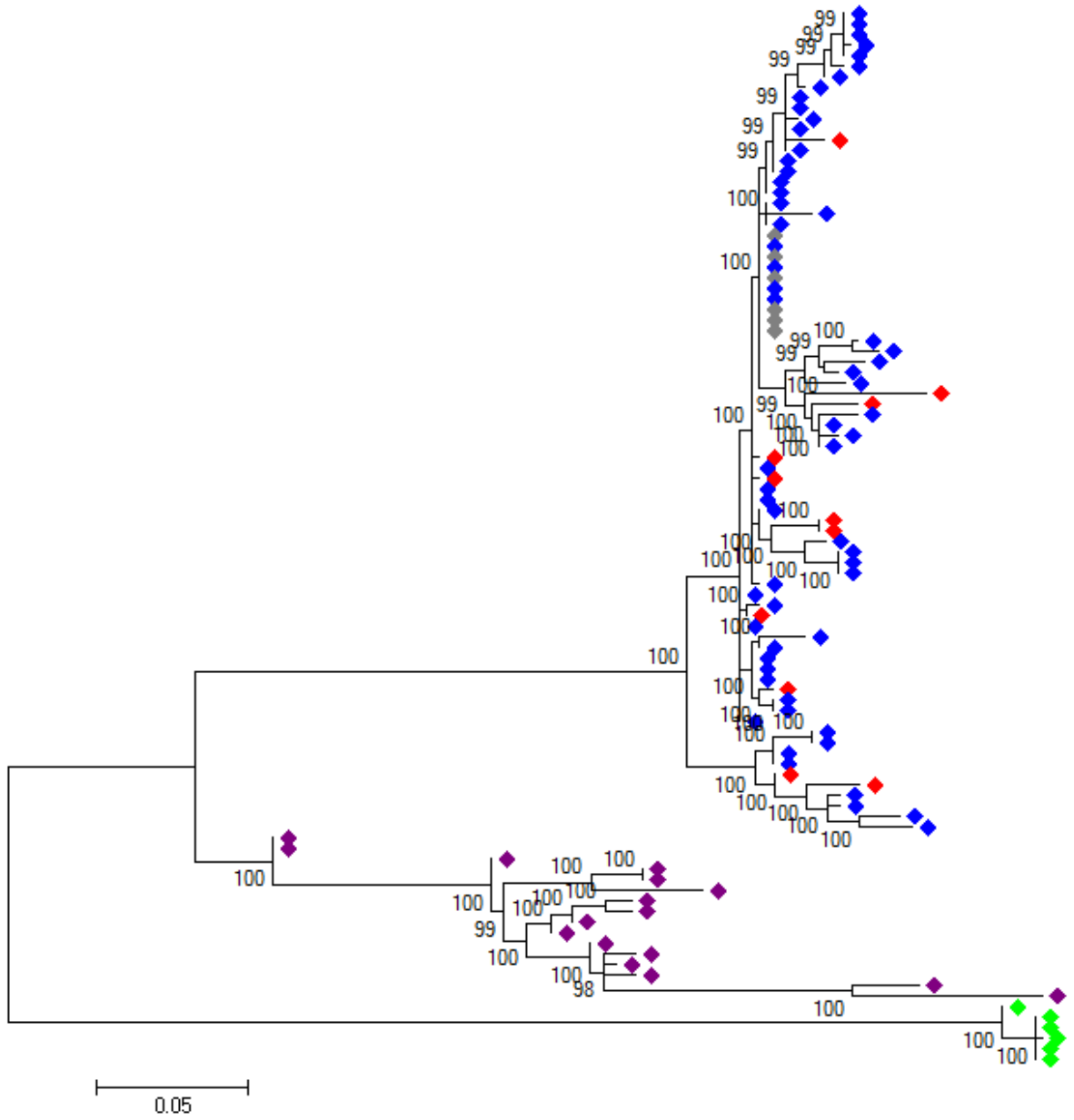


Figure S7.276 Chromosome 18: 205001 - 225001

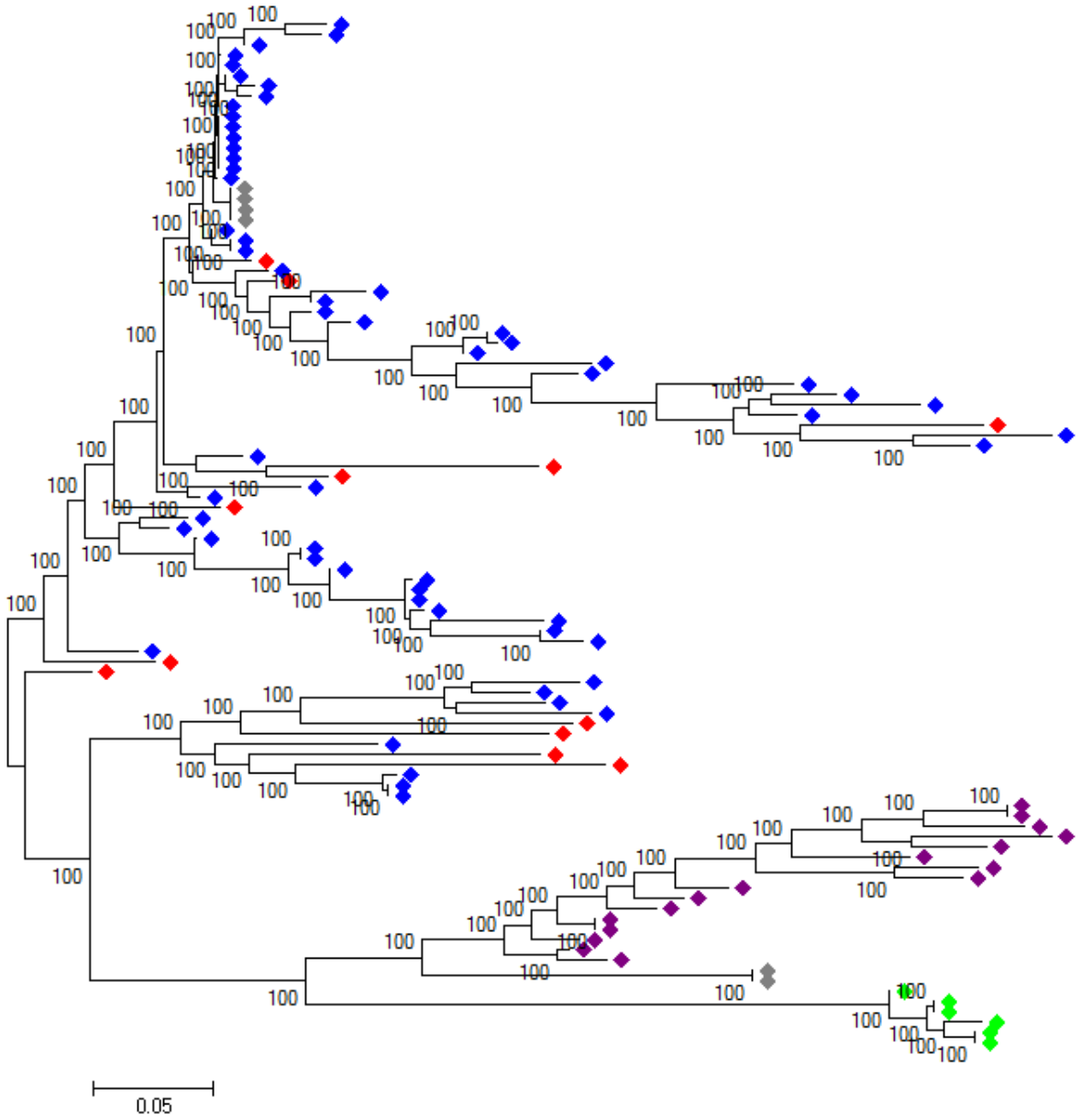


Figure S7.277 Chromosome 19: 425001 - 480001

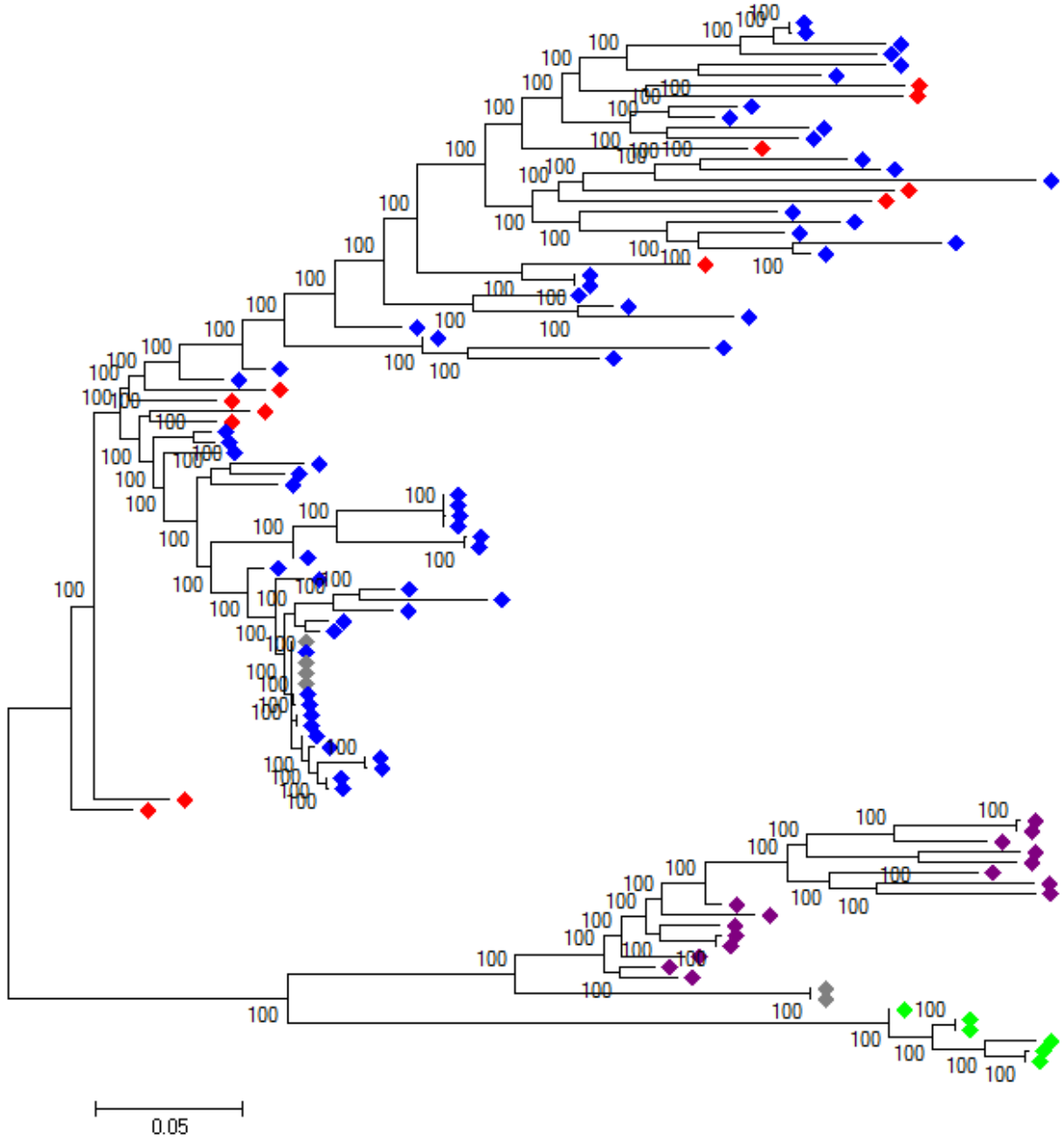


Figure S7.278 Chromosome 19: 505001 - 545001

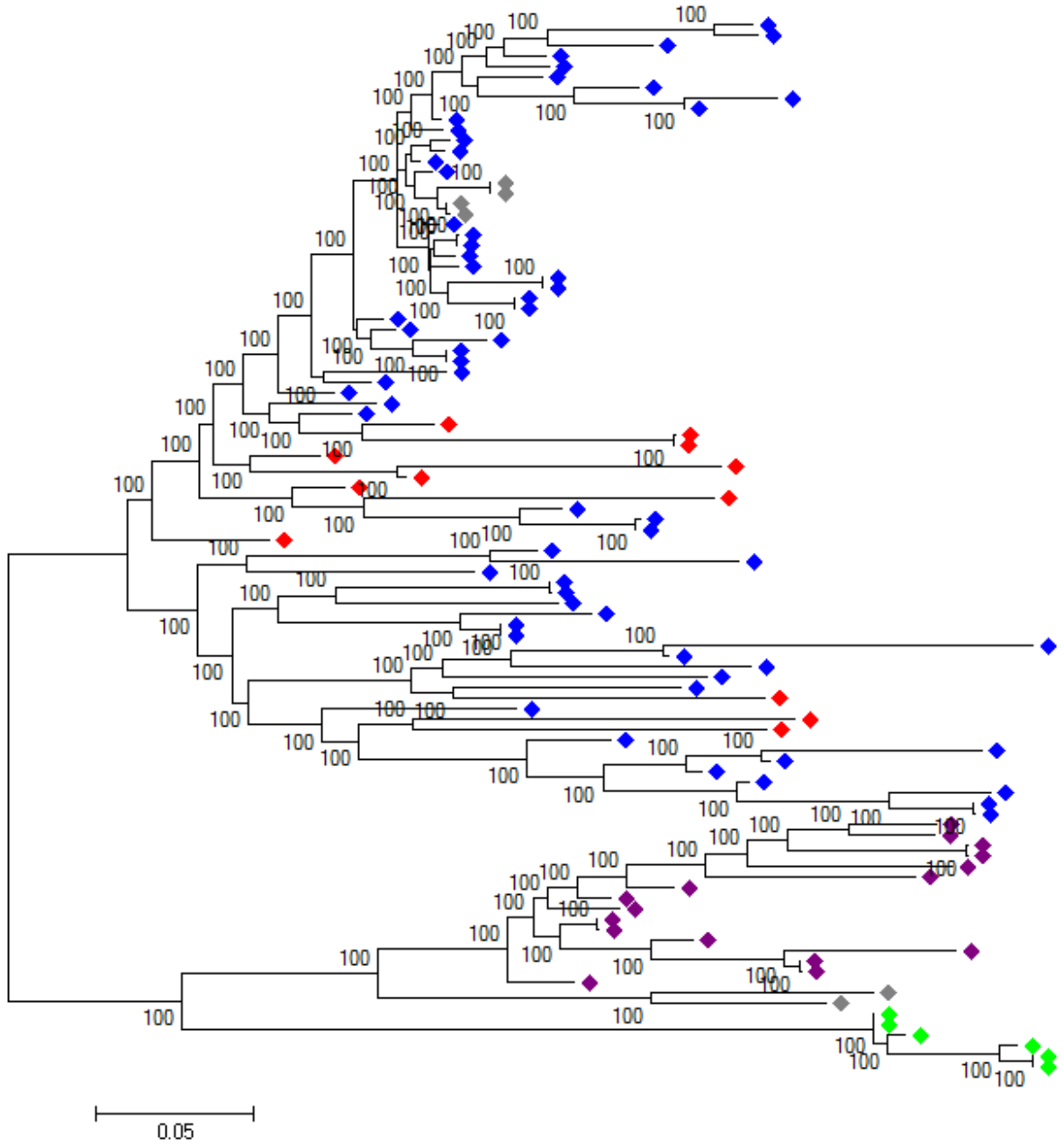


Figure S7.279 Chromosome 19: 1405001 - 1445001

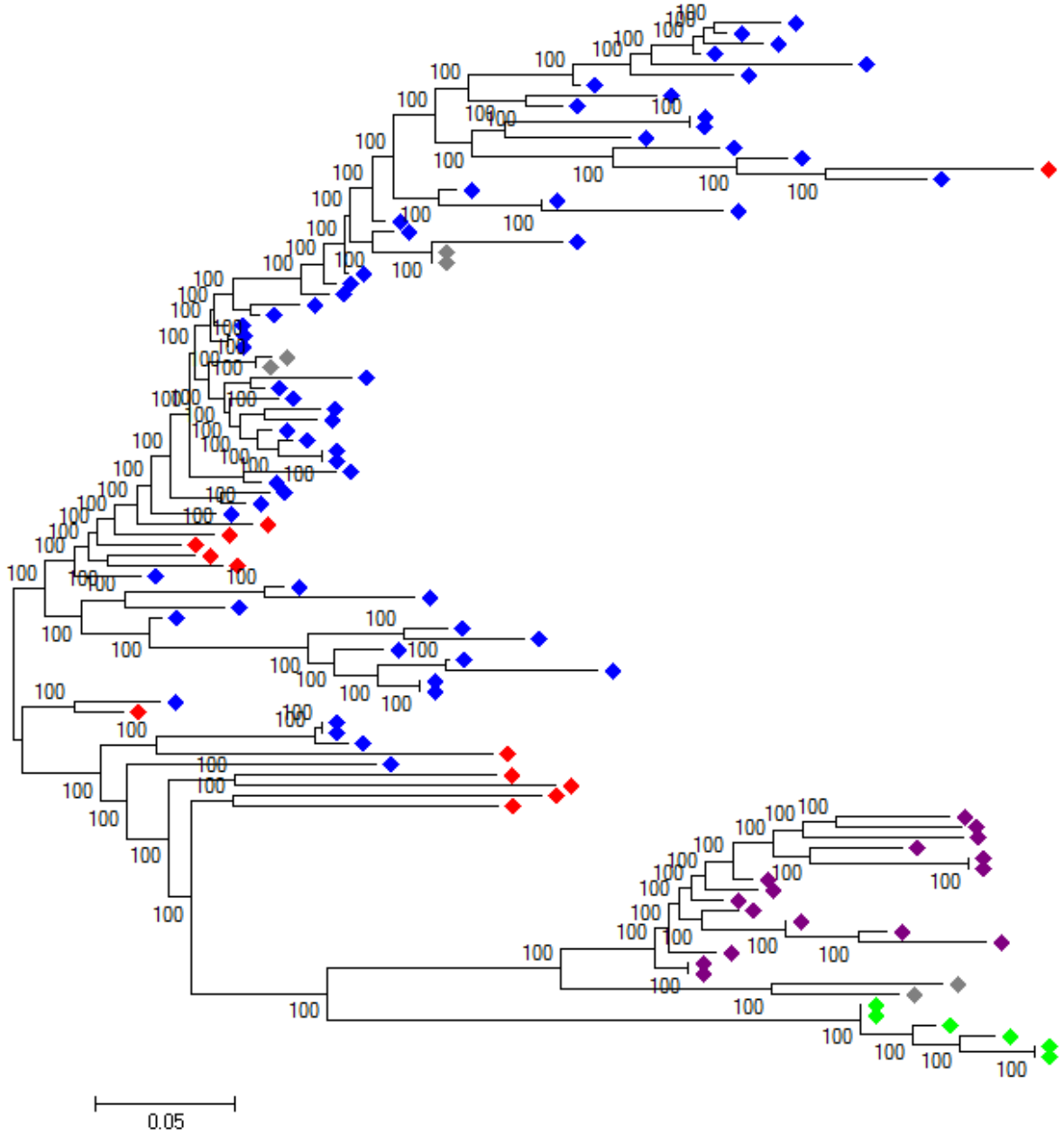


Figure S7.280 Chromosome 19: 1520001 - 1595001

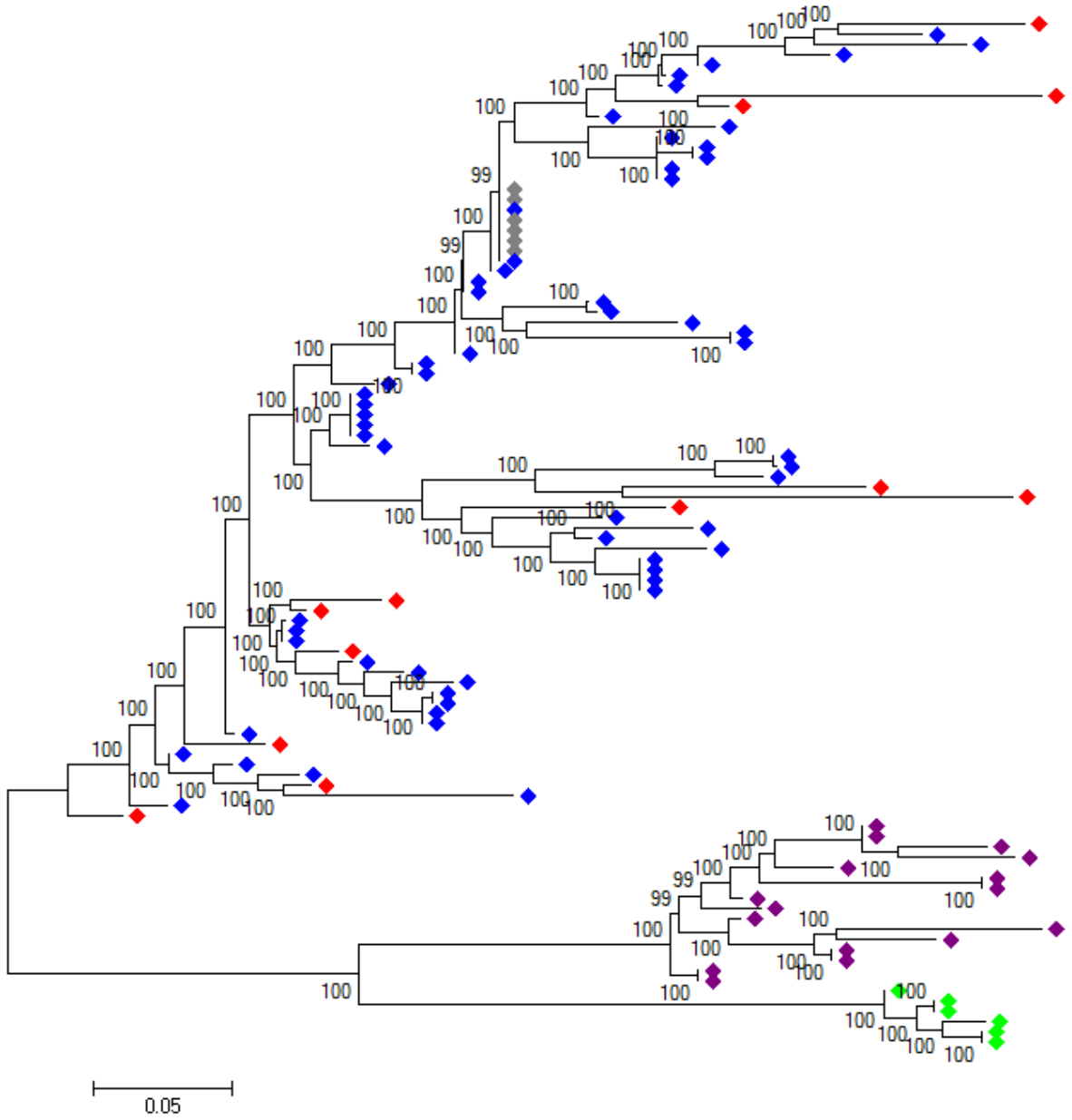


Figure S7.281 Chromosome 19: 2965001 - 2990001

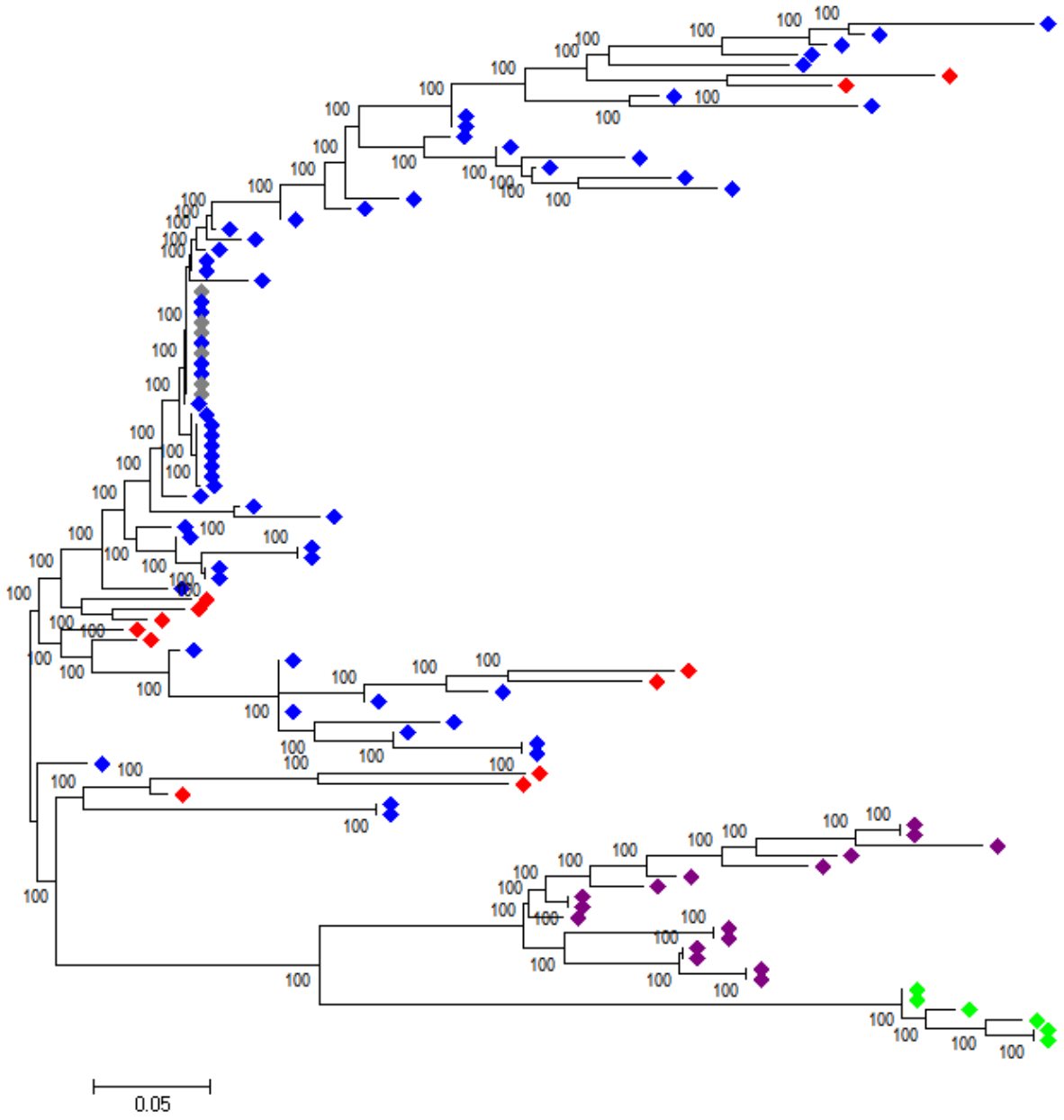


Figure S7.282 Chromosome 19: 3345001 - 3365001

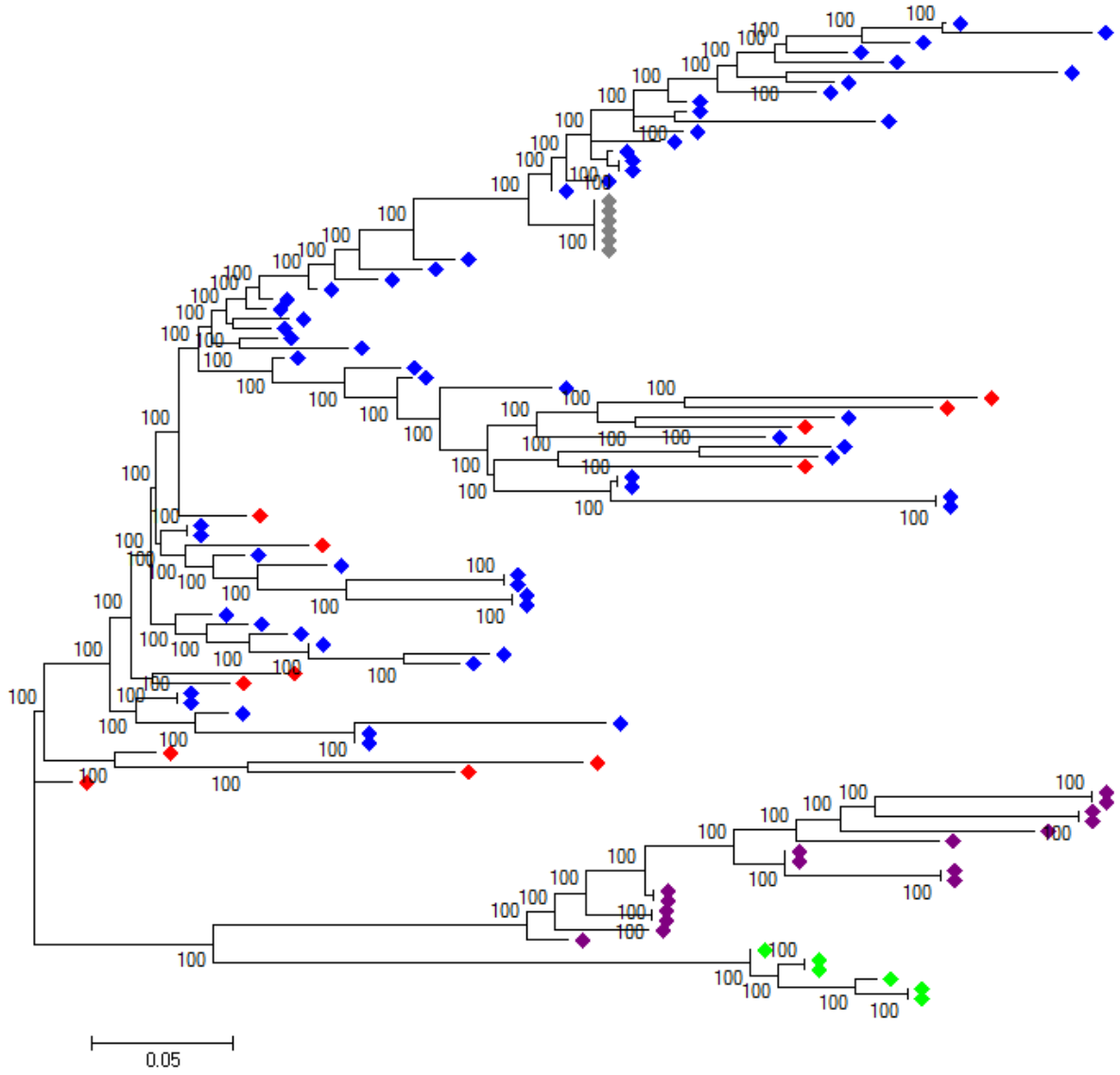


Figure S7.283 Chromosome 19: 4030001 - 4060001

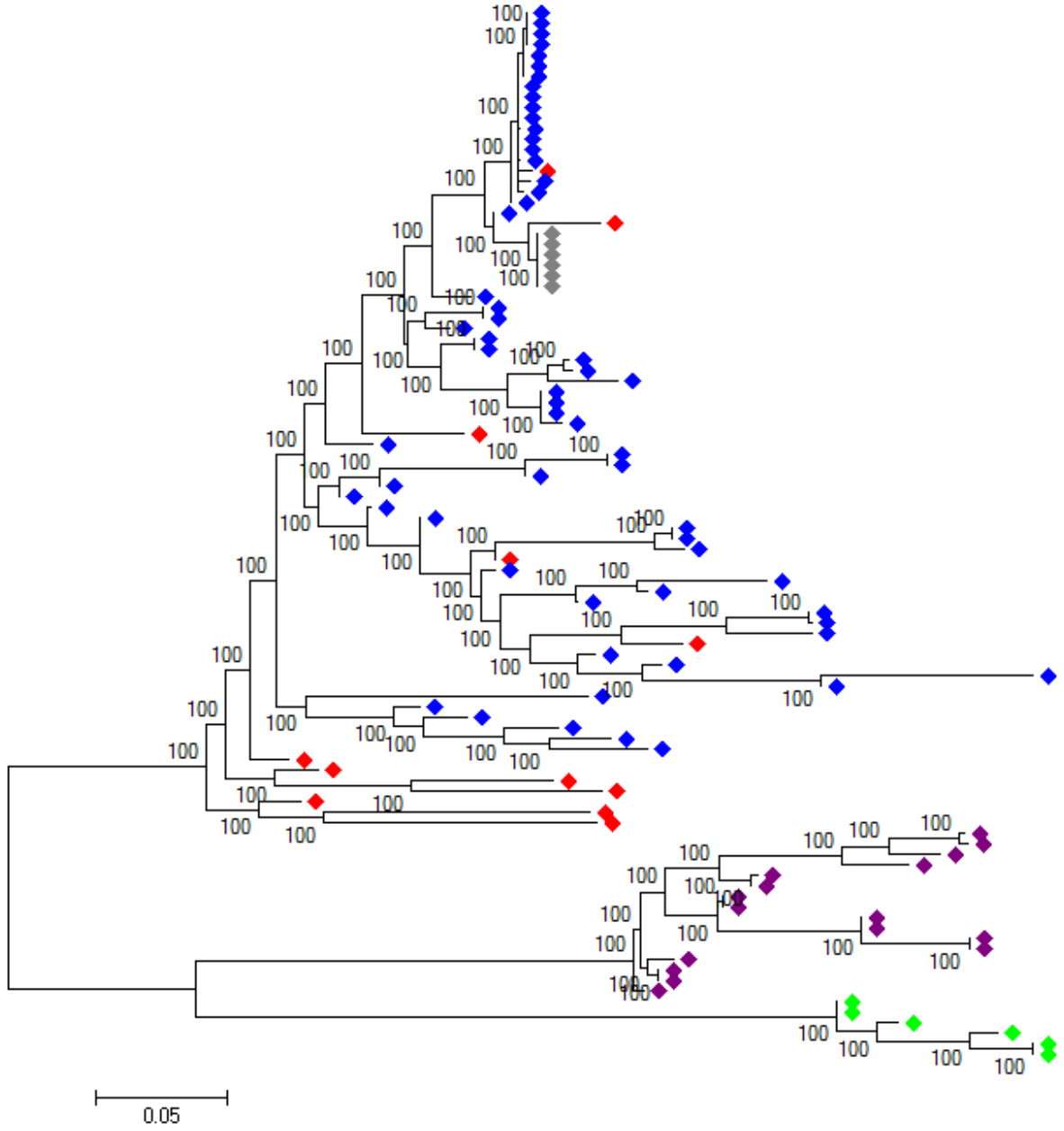


Figure S7.284 Chromosome 19: 4515001 - 4545001

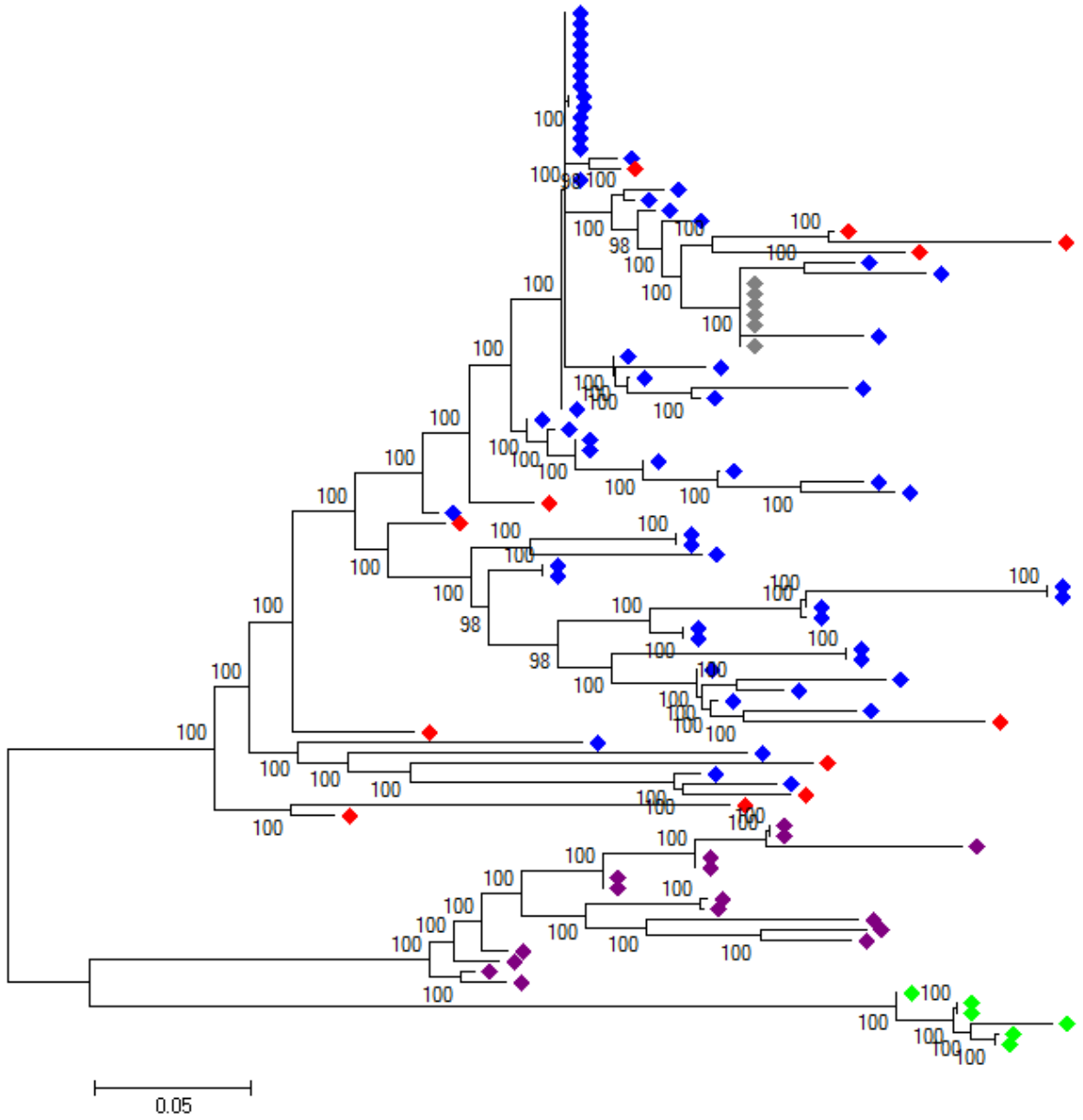


Figure S7.285 Chromosome 19: 4595001 - 4625001

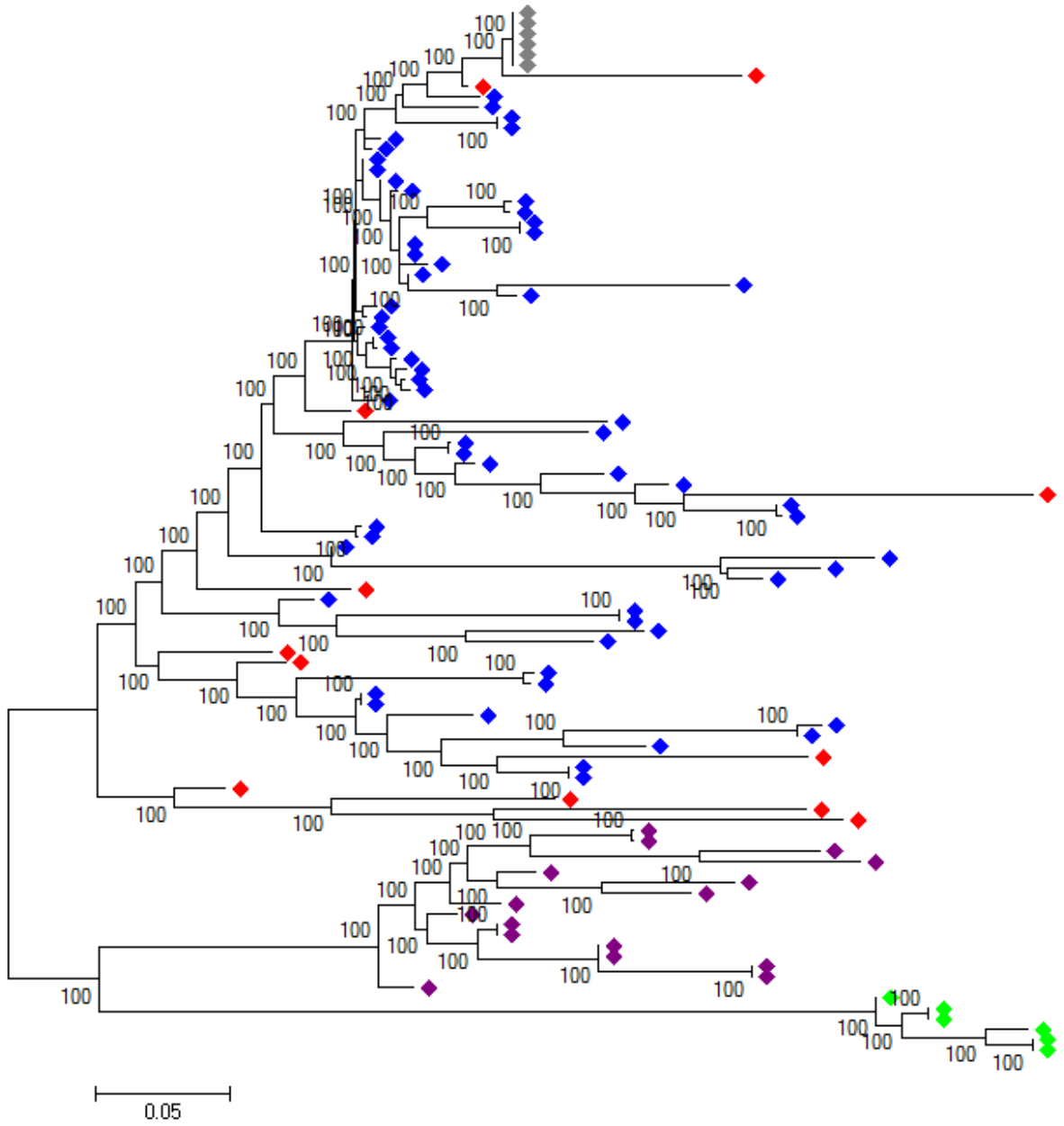


Figure S7.286 Chromosome 19: 4695001 - 4725001

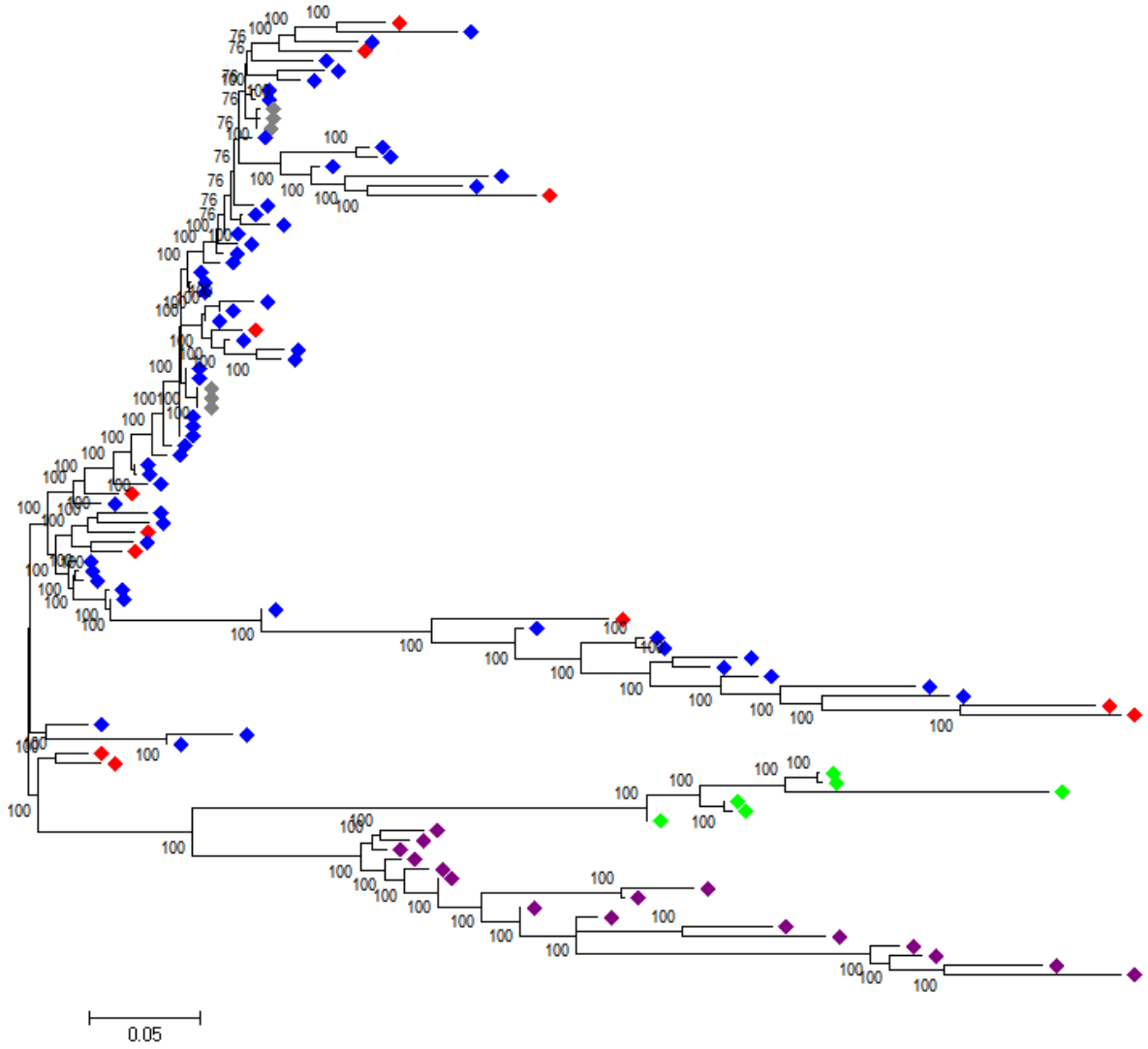


Figure S7.287 Chromosome 19: 4805001 - 4835001

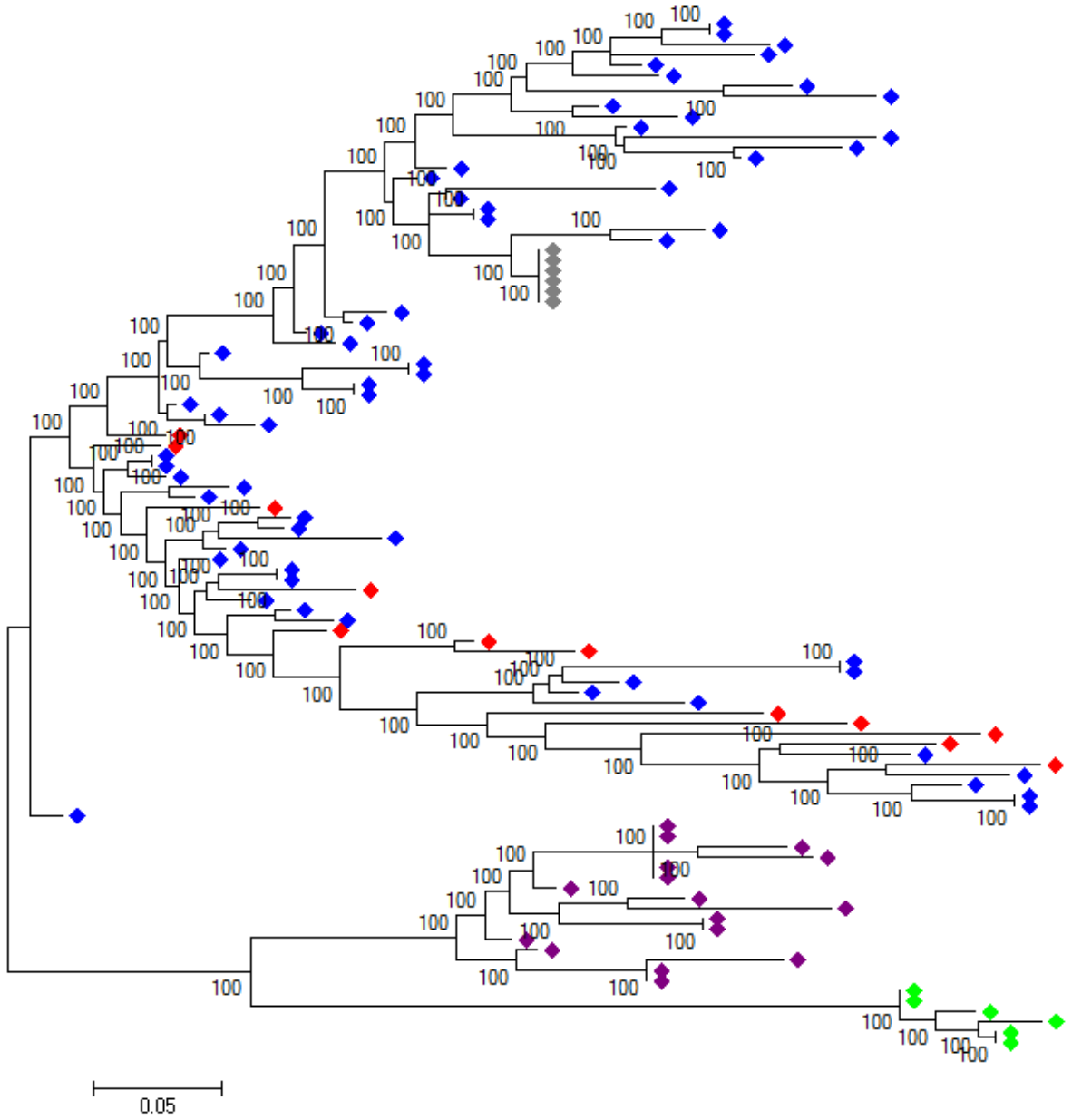


Figure S7.288 Chromosome 19: 5145001 - 5165001

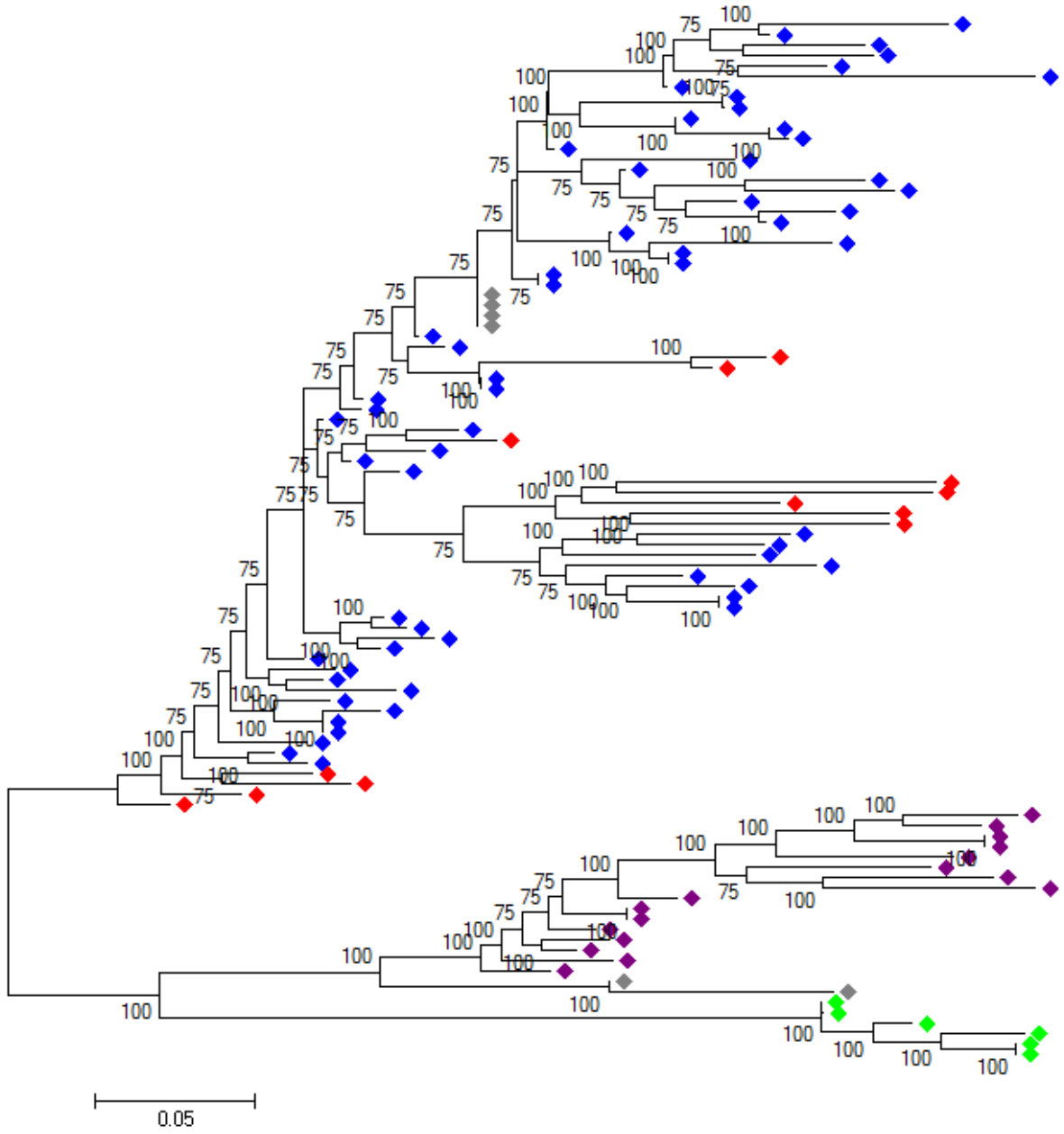


Figure S7.289 Chromosome 20: 11075001 - 11110001

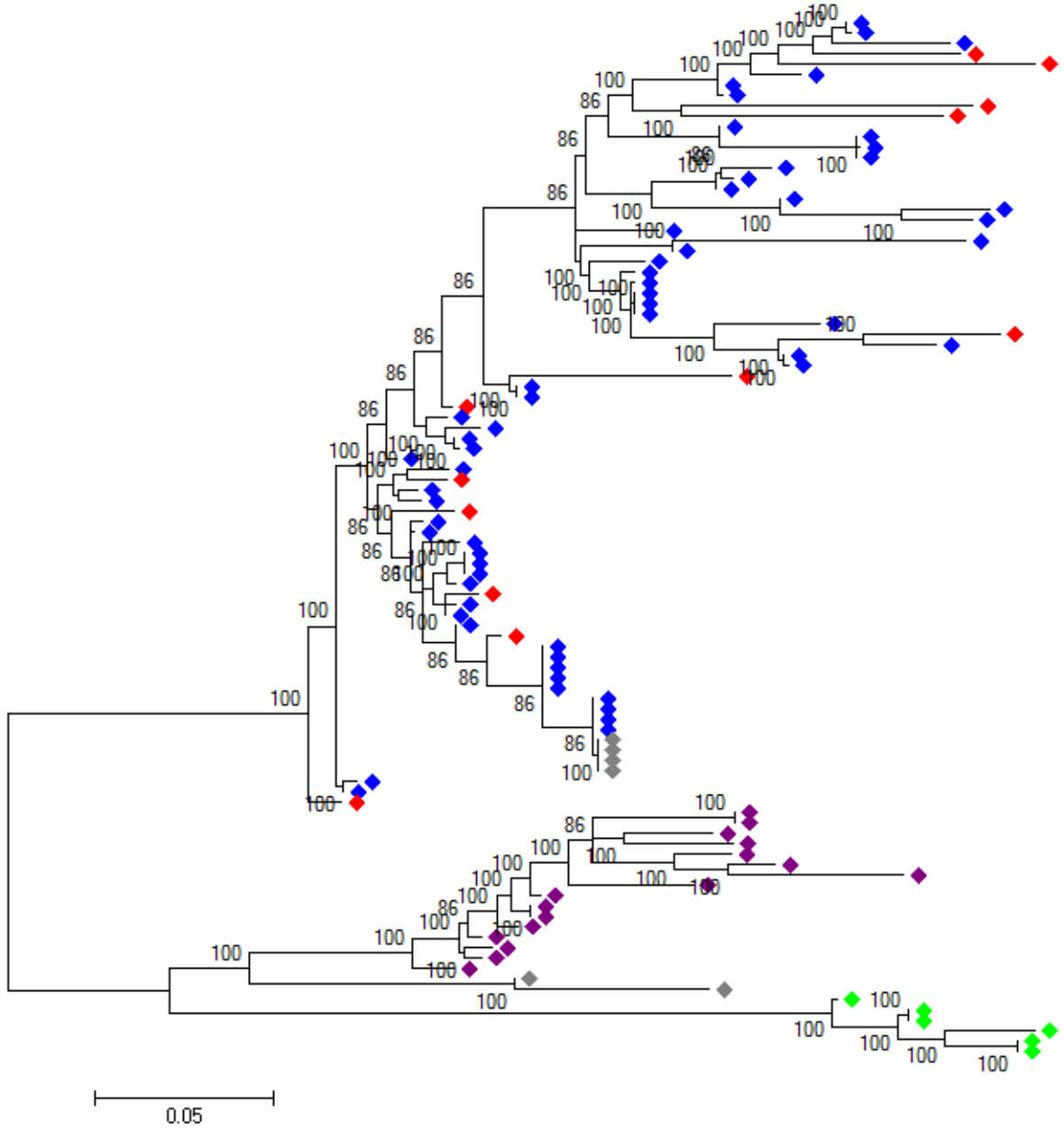


Figure S7.290 Chromosome 20: 11130001 - 11150001

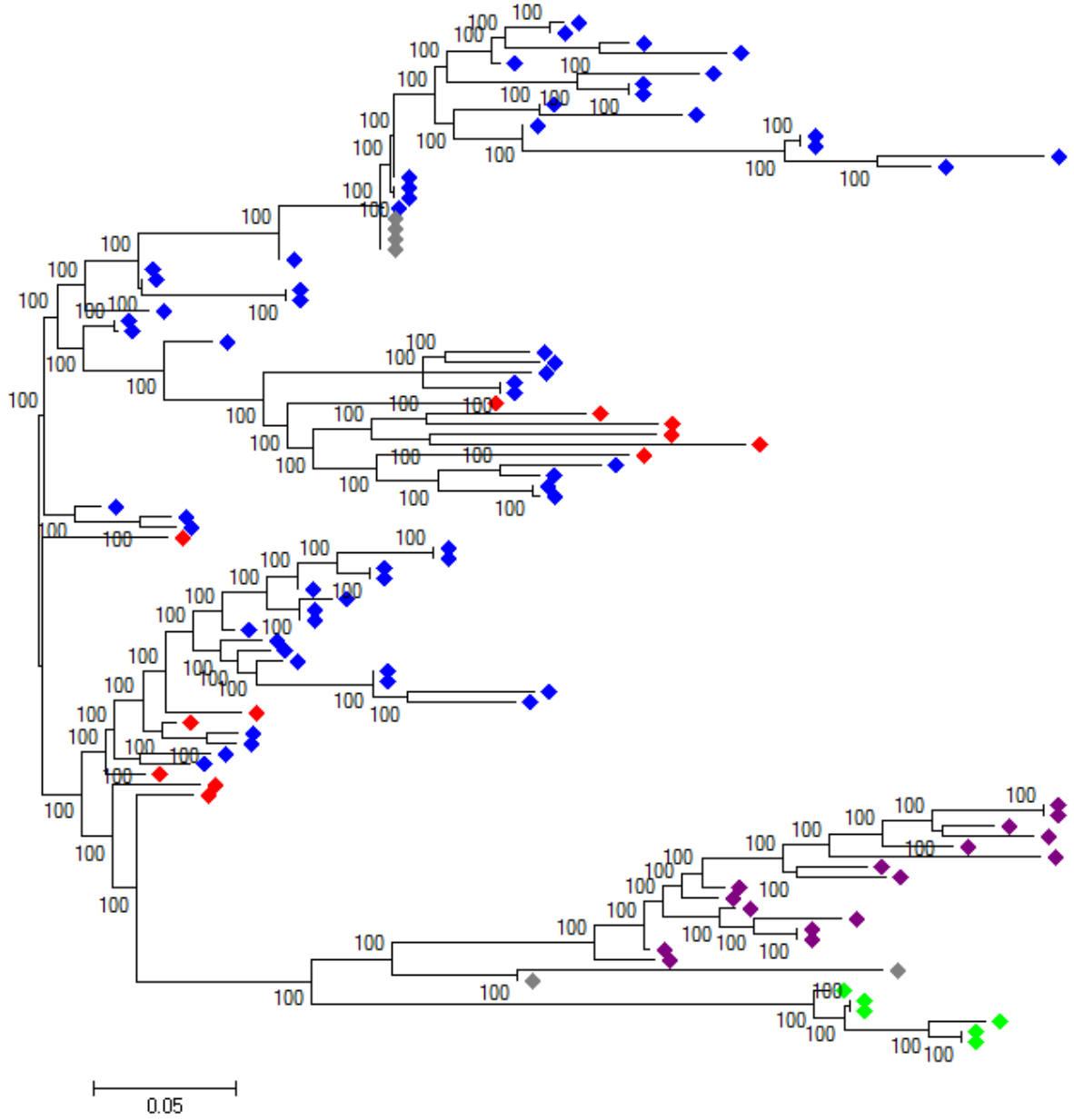


Figure S7.291 Chromosome 20: 11420001 - 11440001

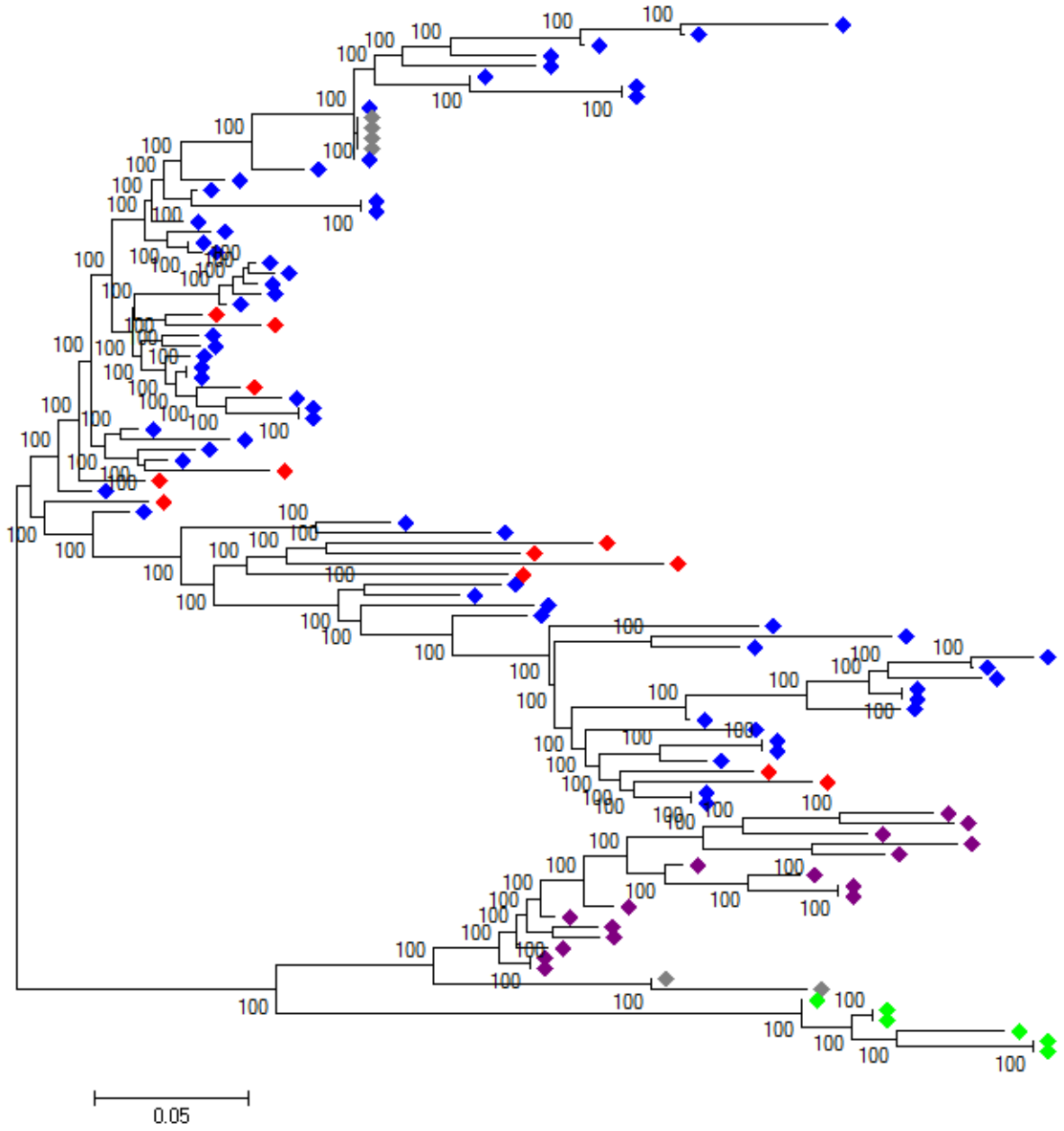


Figure S7.292 Chromosome 20: 11785001 - 11815001

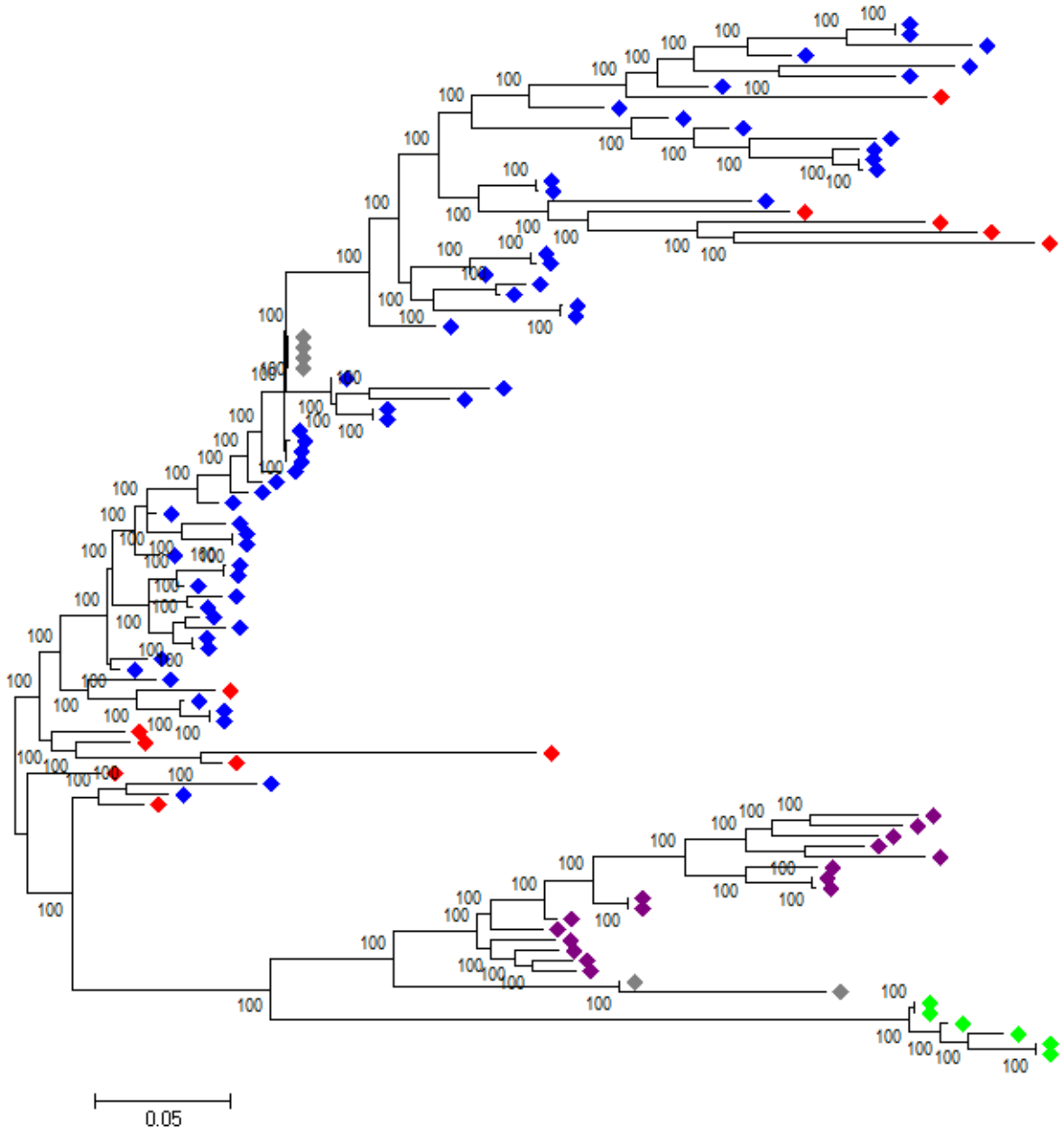


Figure S7.293 Chromosome 20: 11990001 - 12025001

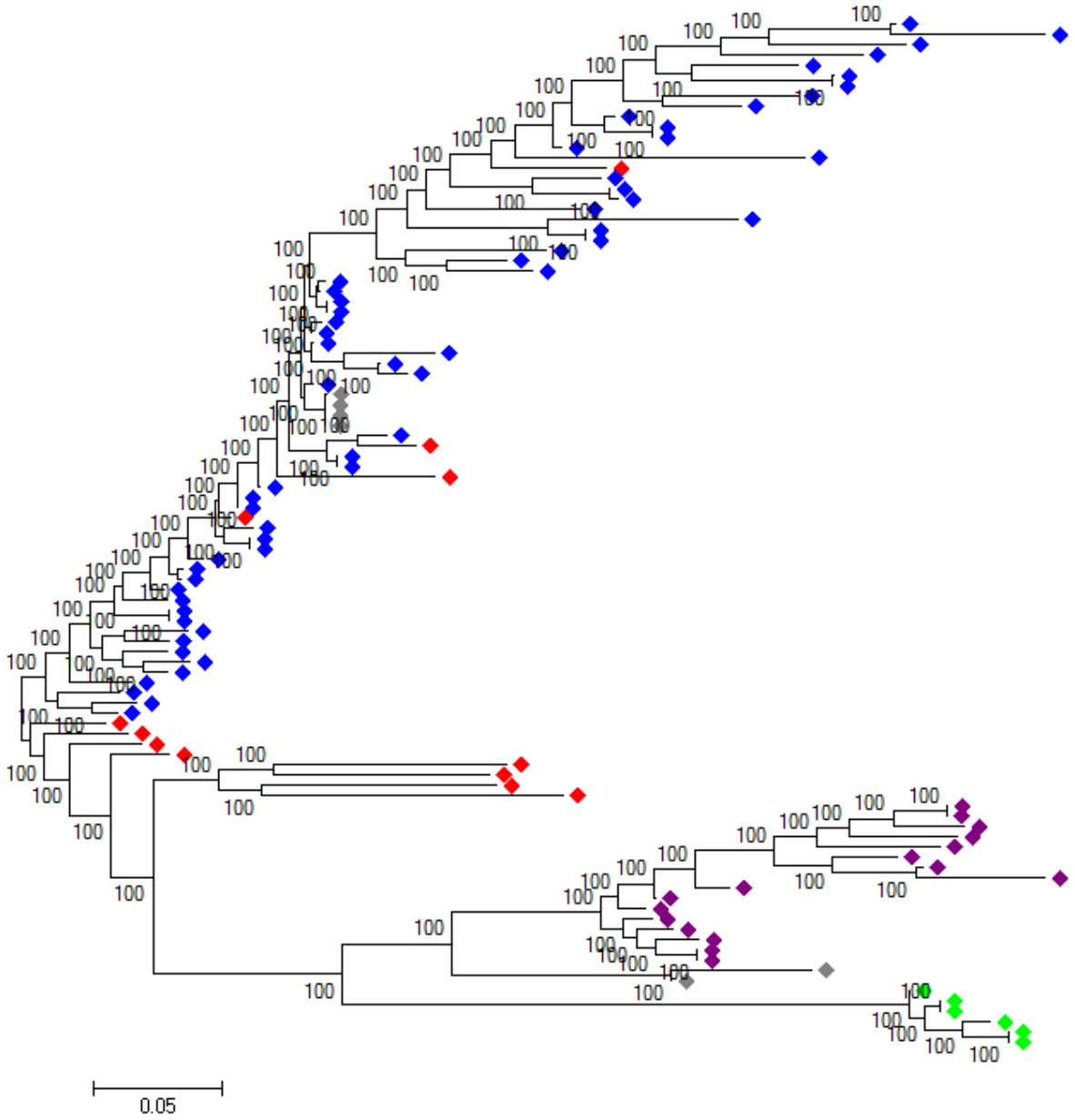


Figure S7.294 Chromosome 20: 12200001 - 12315001

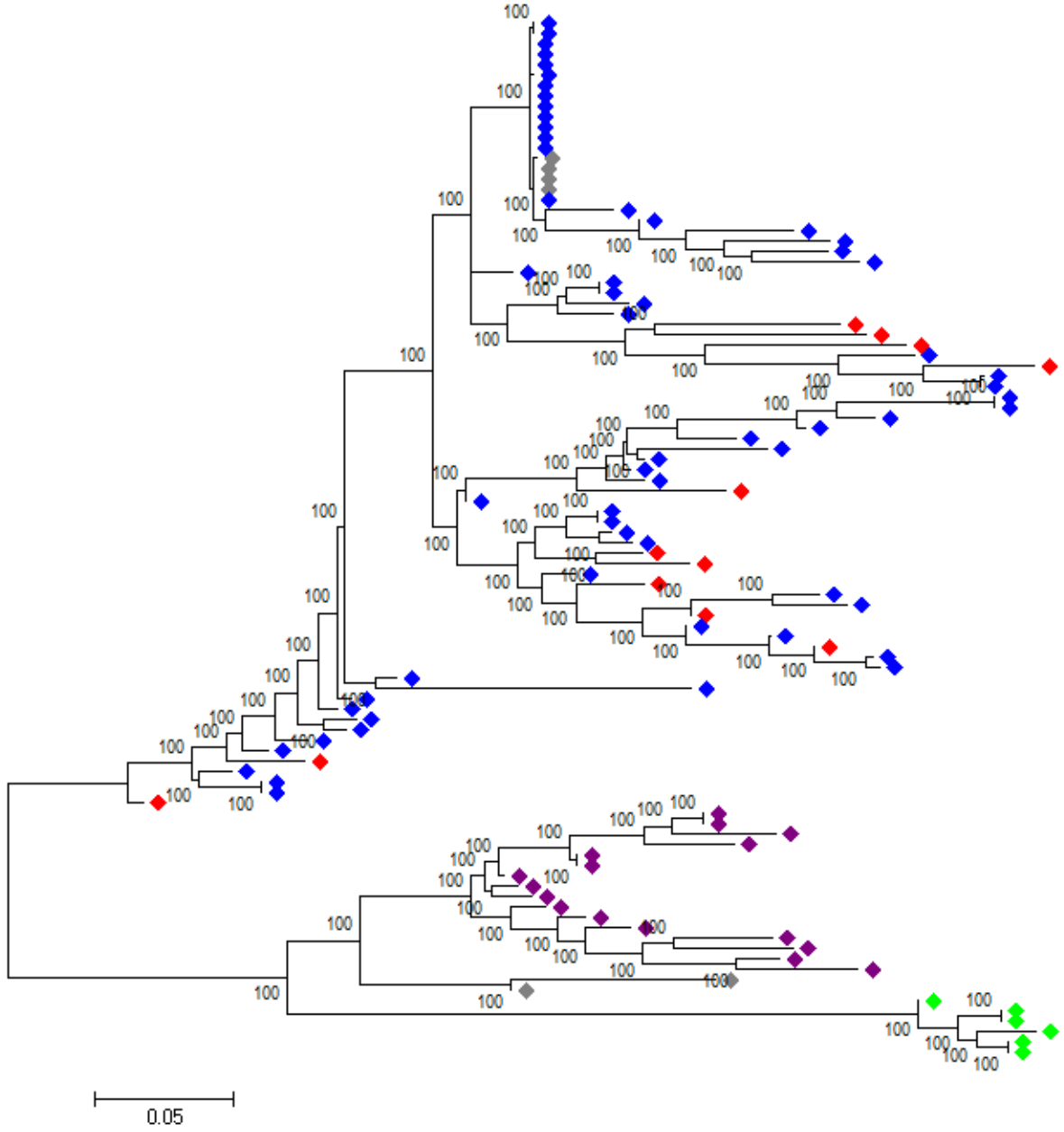


Figure S7.295 Chromosome 20: 12355001 - 12375001

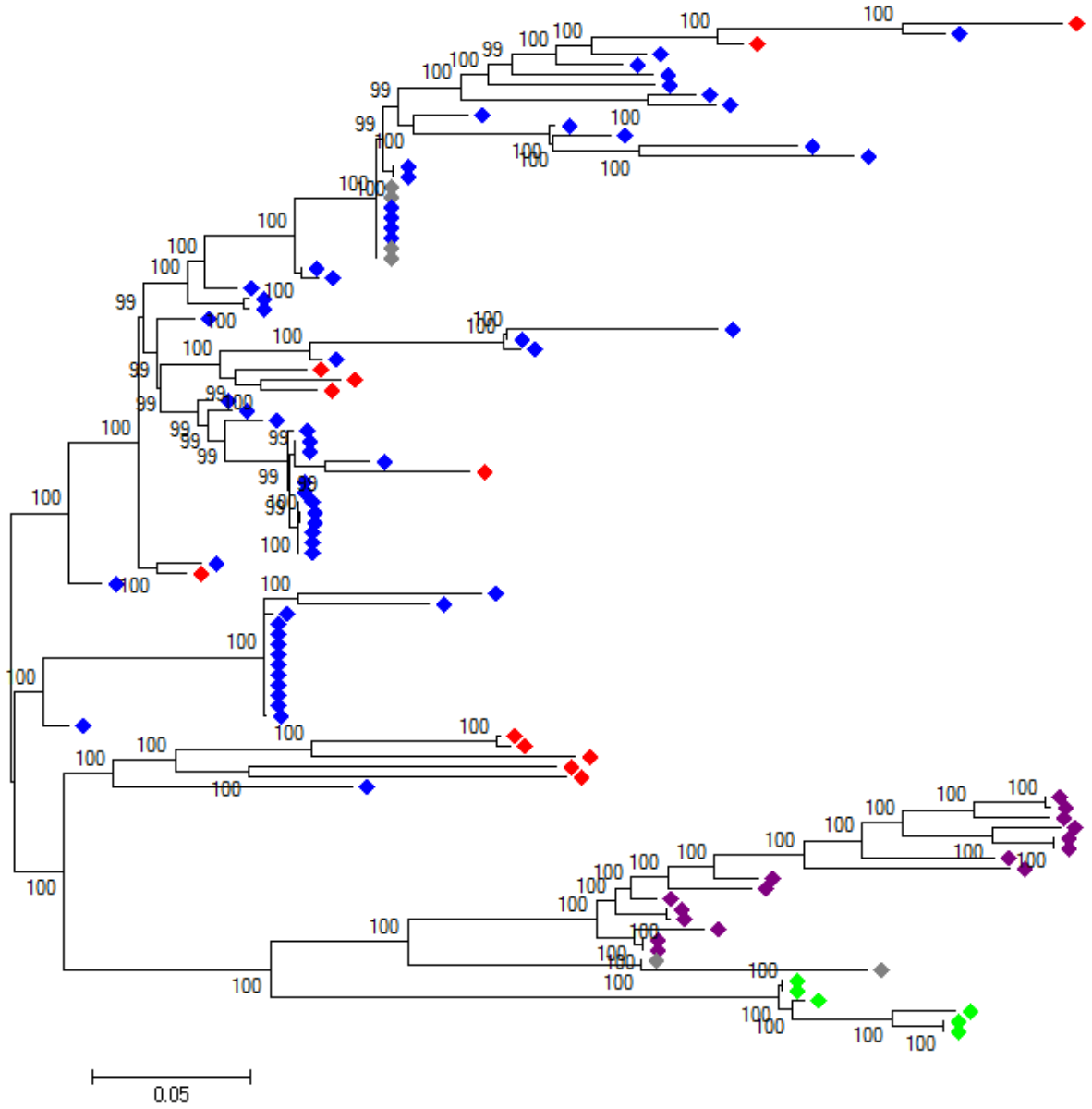


Figure S7.296 Chromosome 20: 12415001 - 12440001

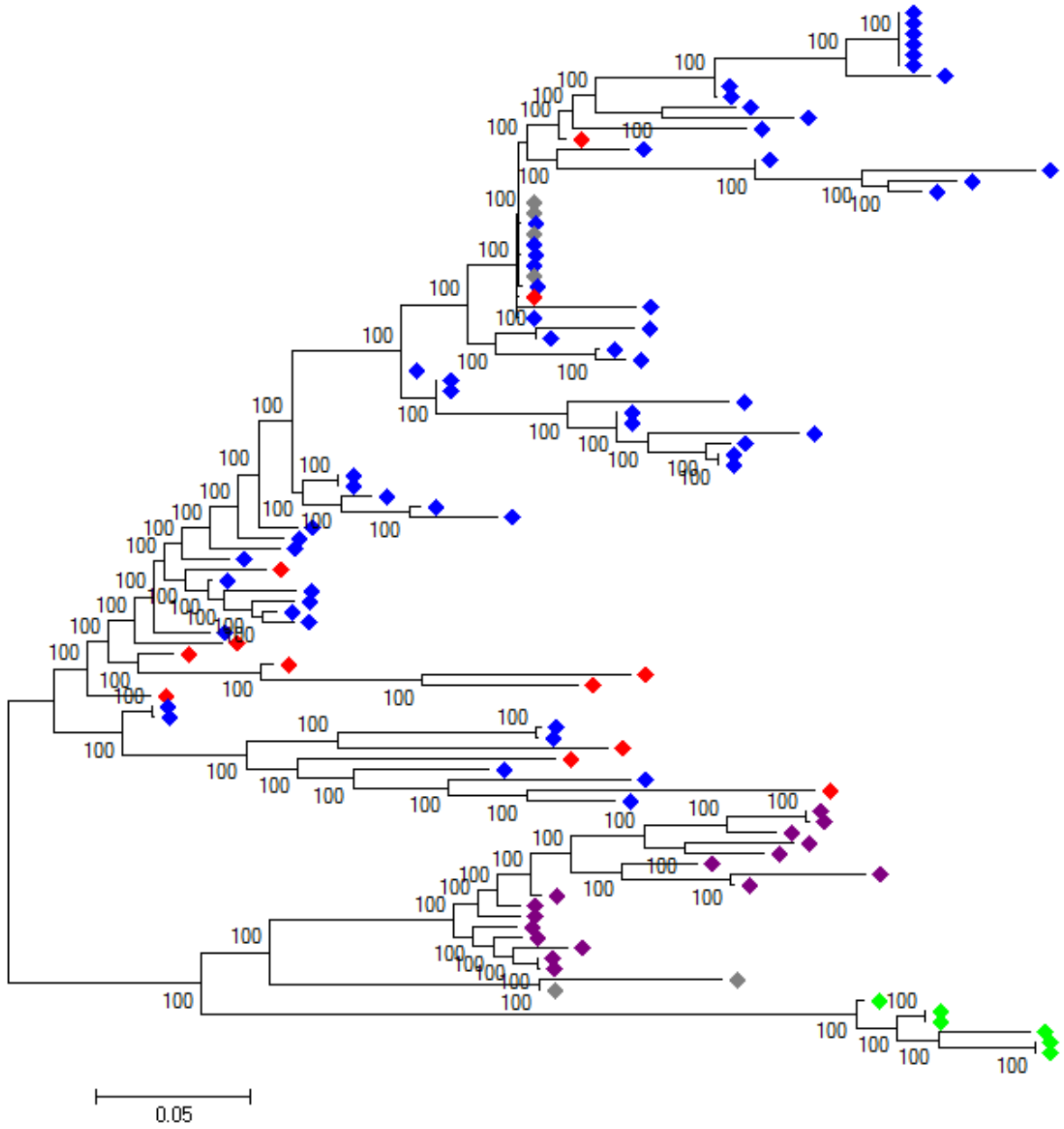


Figure S7.297 Chromosome 20: 12525001 - 12560001

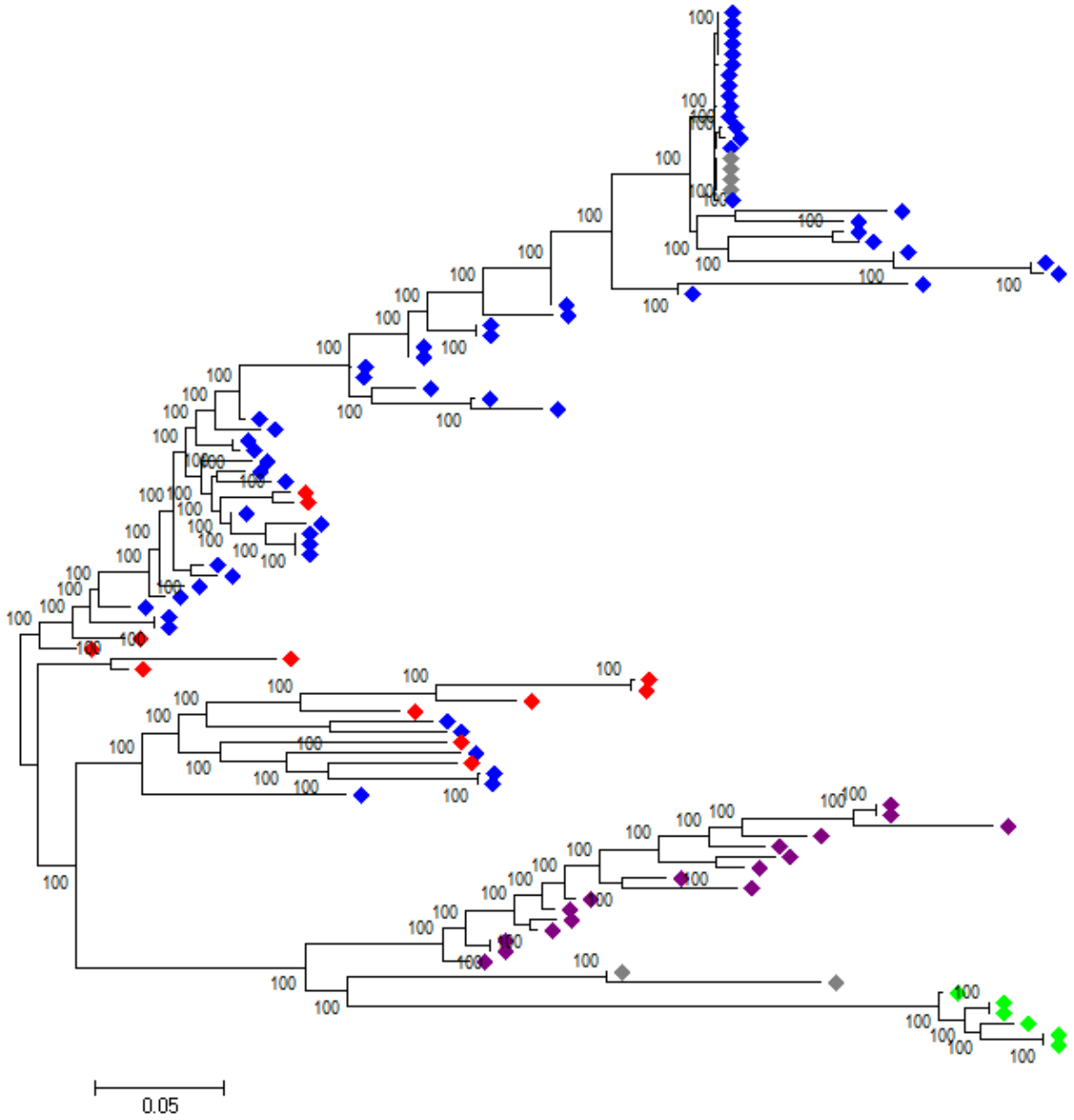


Figure S7.298 Chromosome 20: 12600001 - 12635001

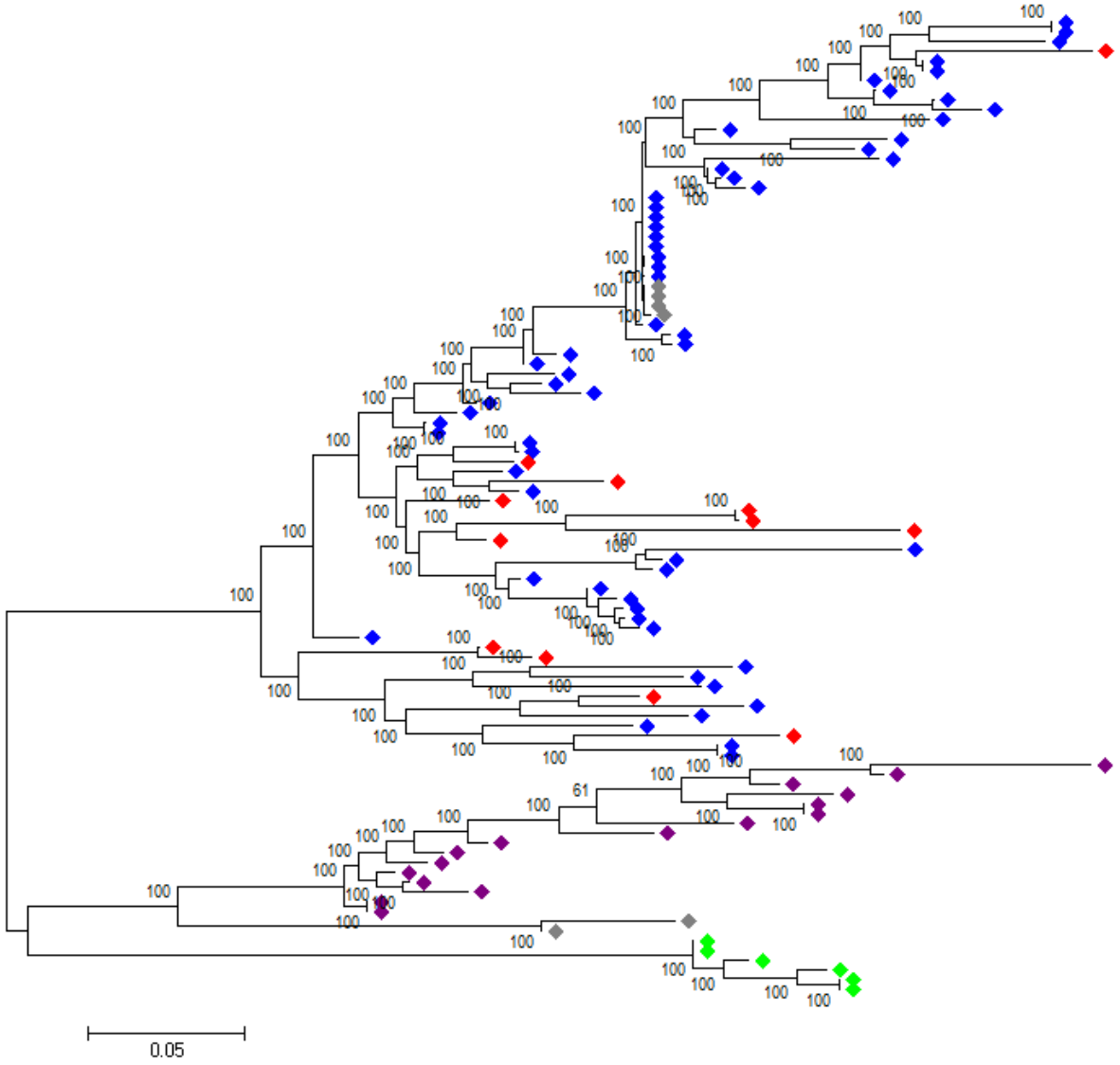


Figure S7.299 Chromosome 20: 12655001 - 12690001

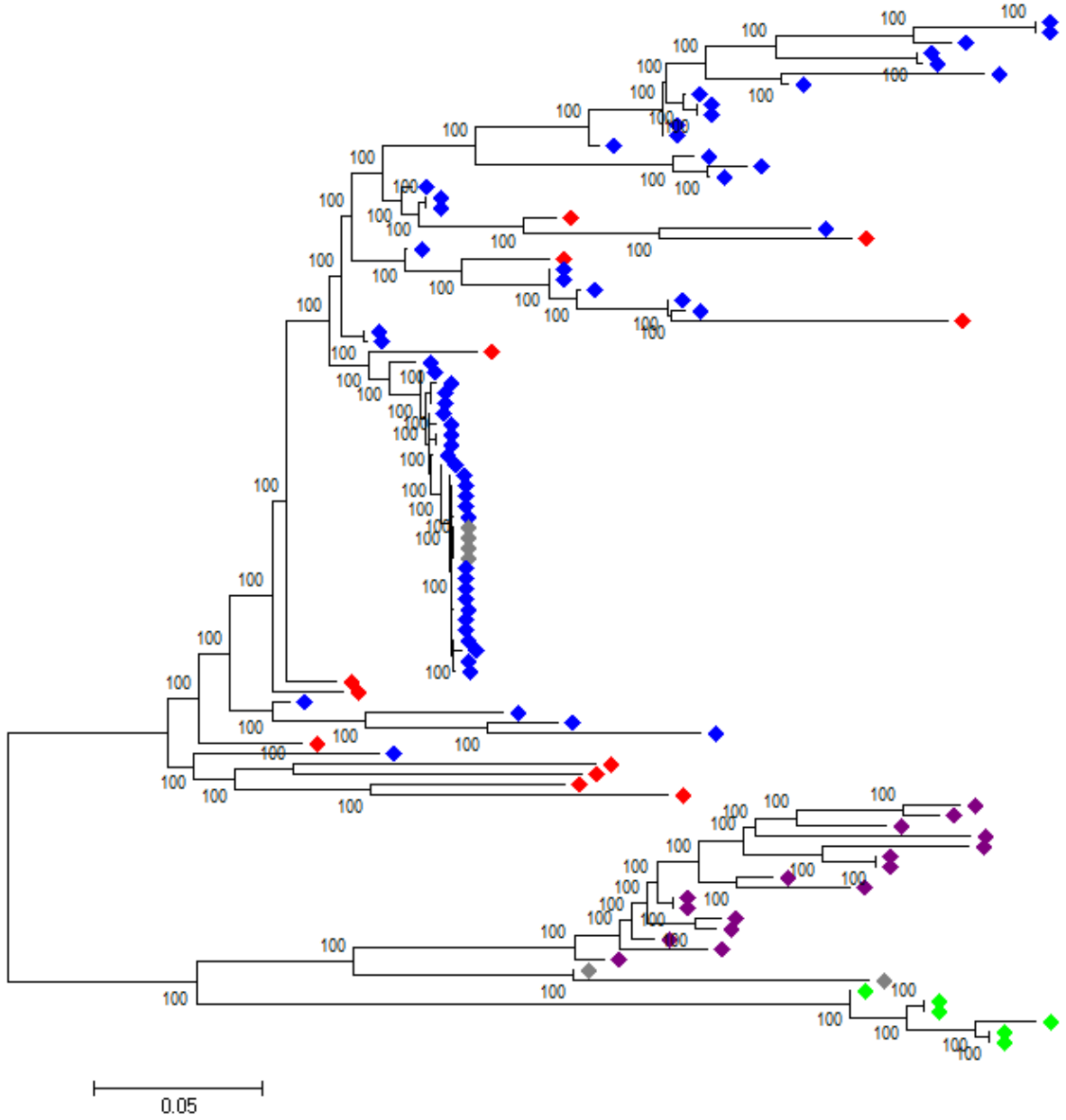


Figure S7.300 Chromosome 20: 12720001 - 12755001

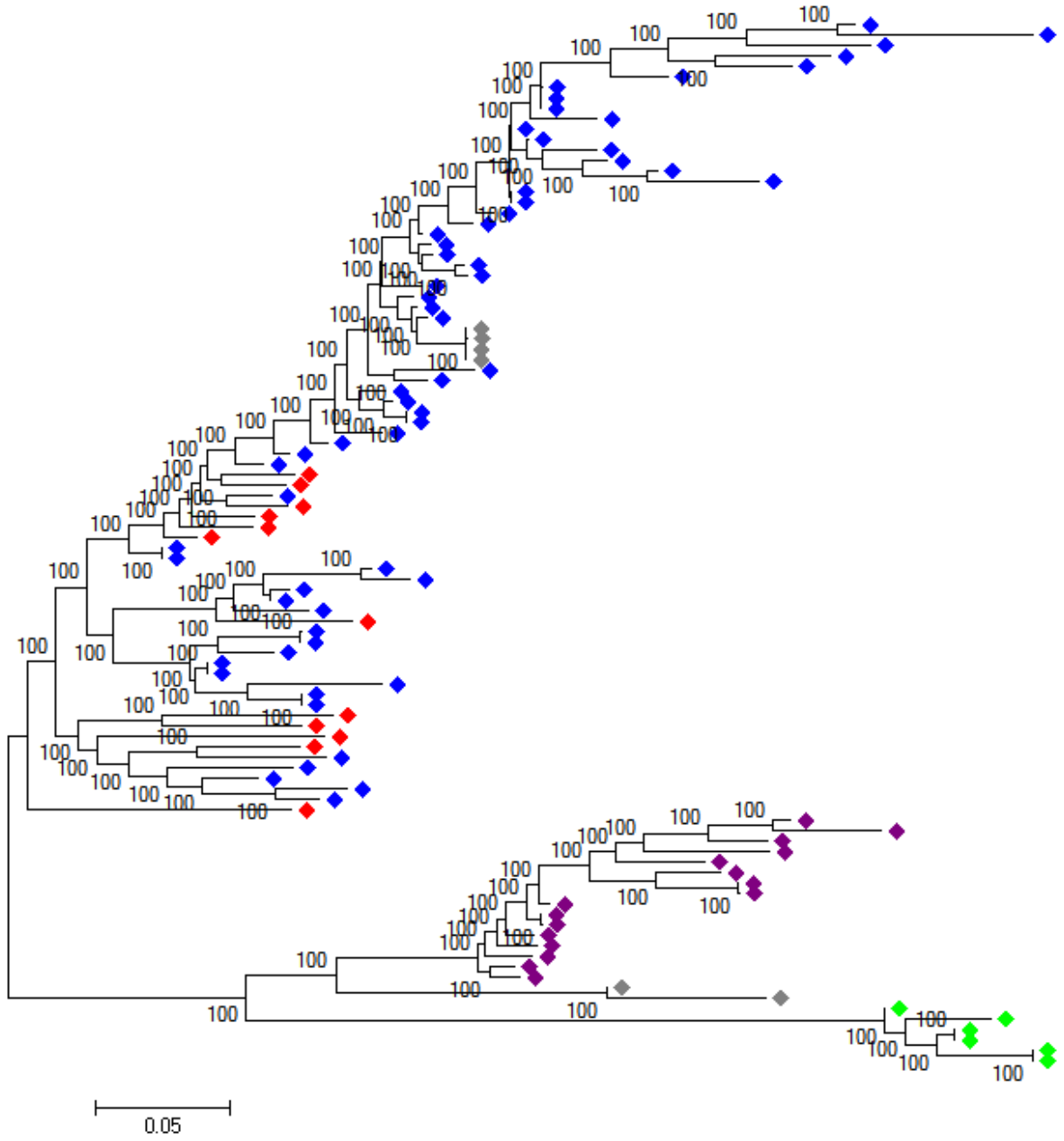


Figure S7.301 Chromosome 20: 12800001 - 12905001

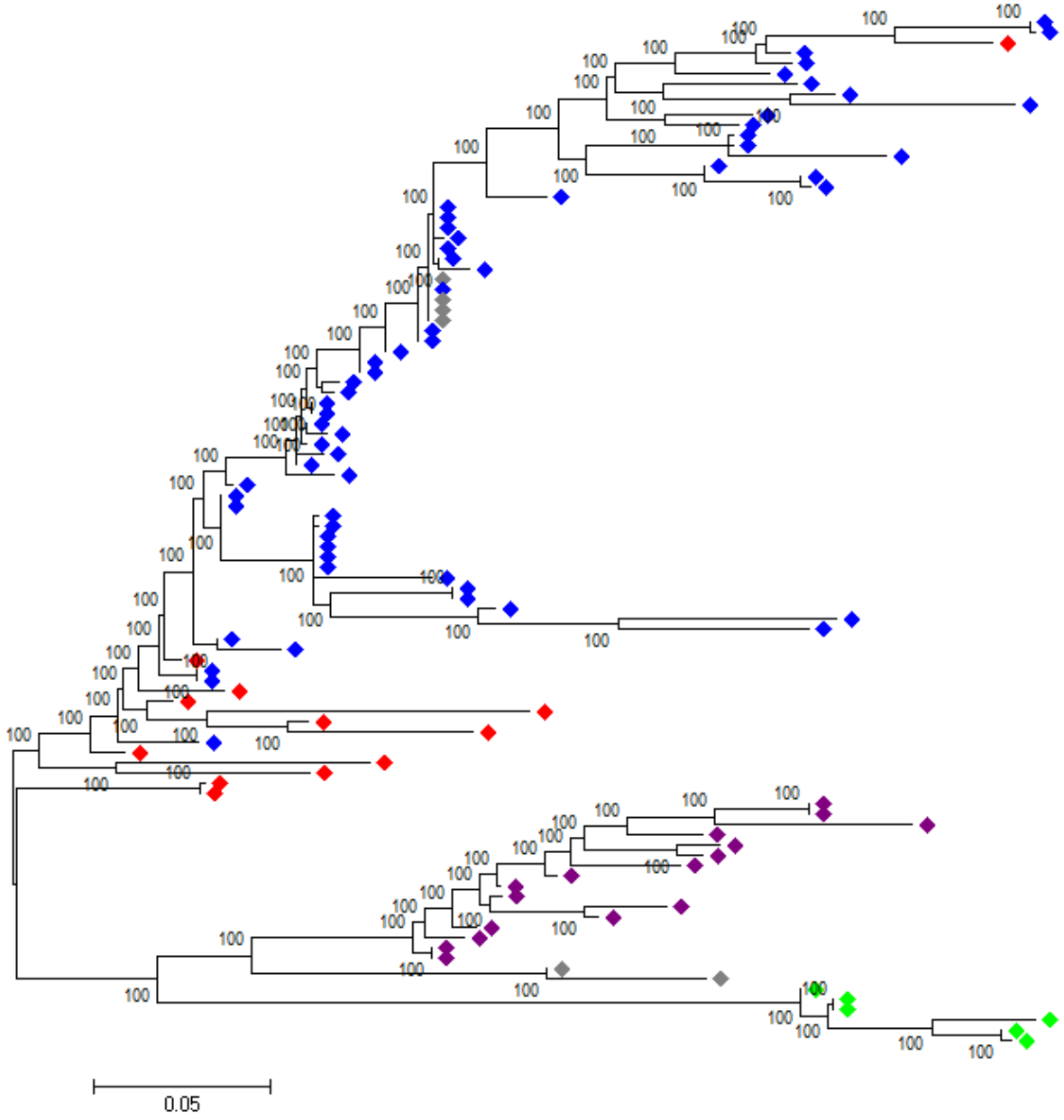


Figure S7.302 Chromosome 20: 13095001 - 13115001

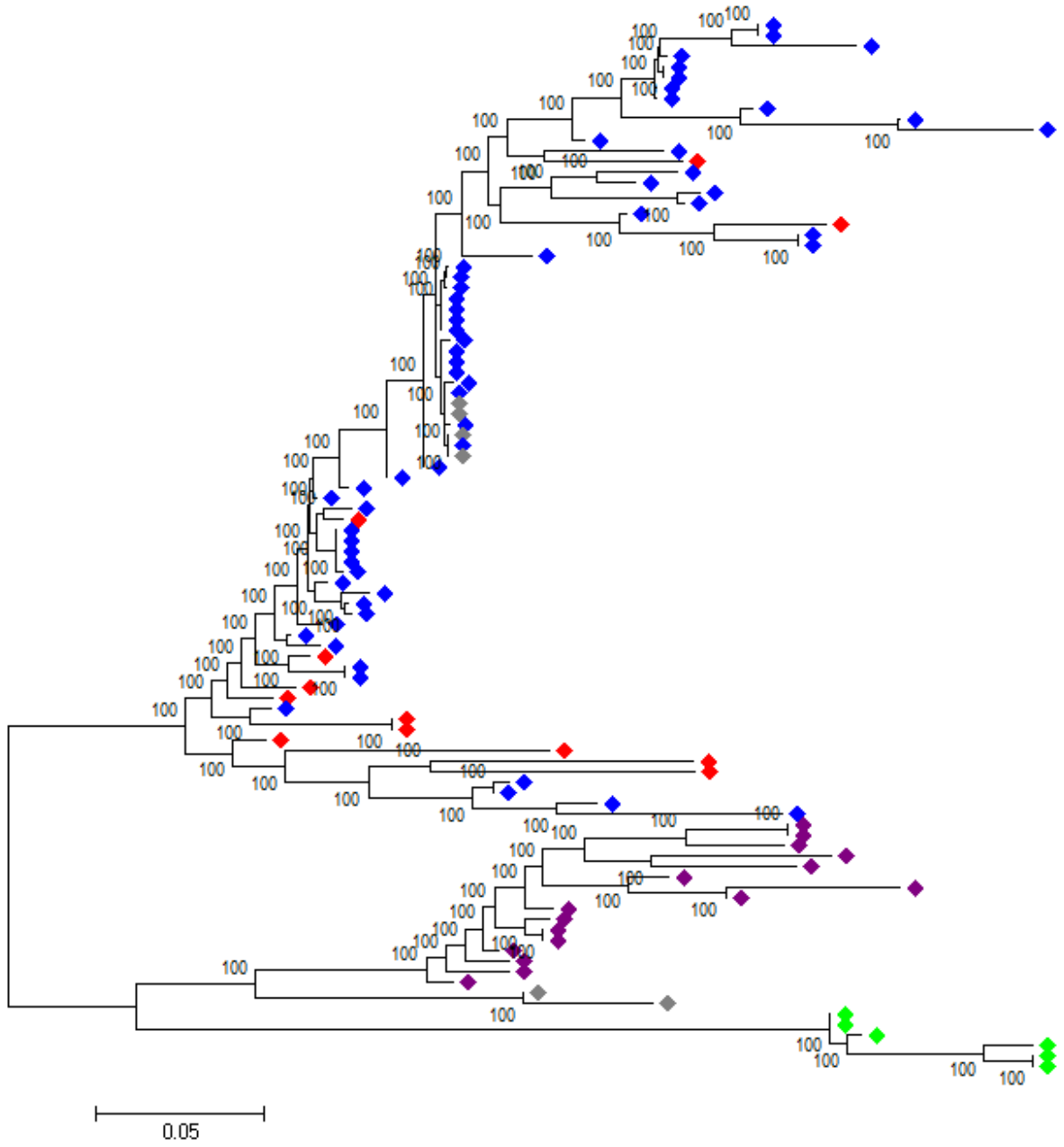


Figure S7. 303 Chromosome 20: 13165001 - 13210001

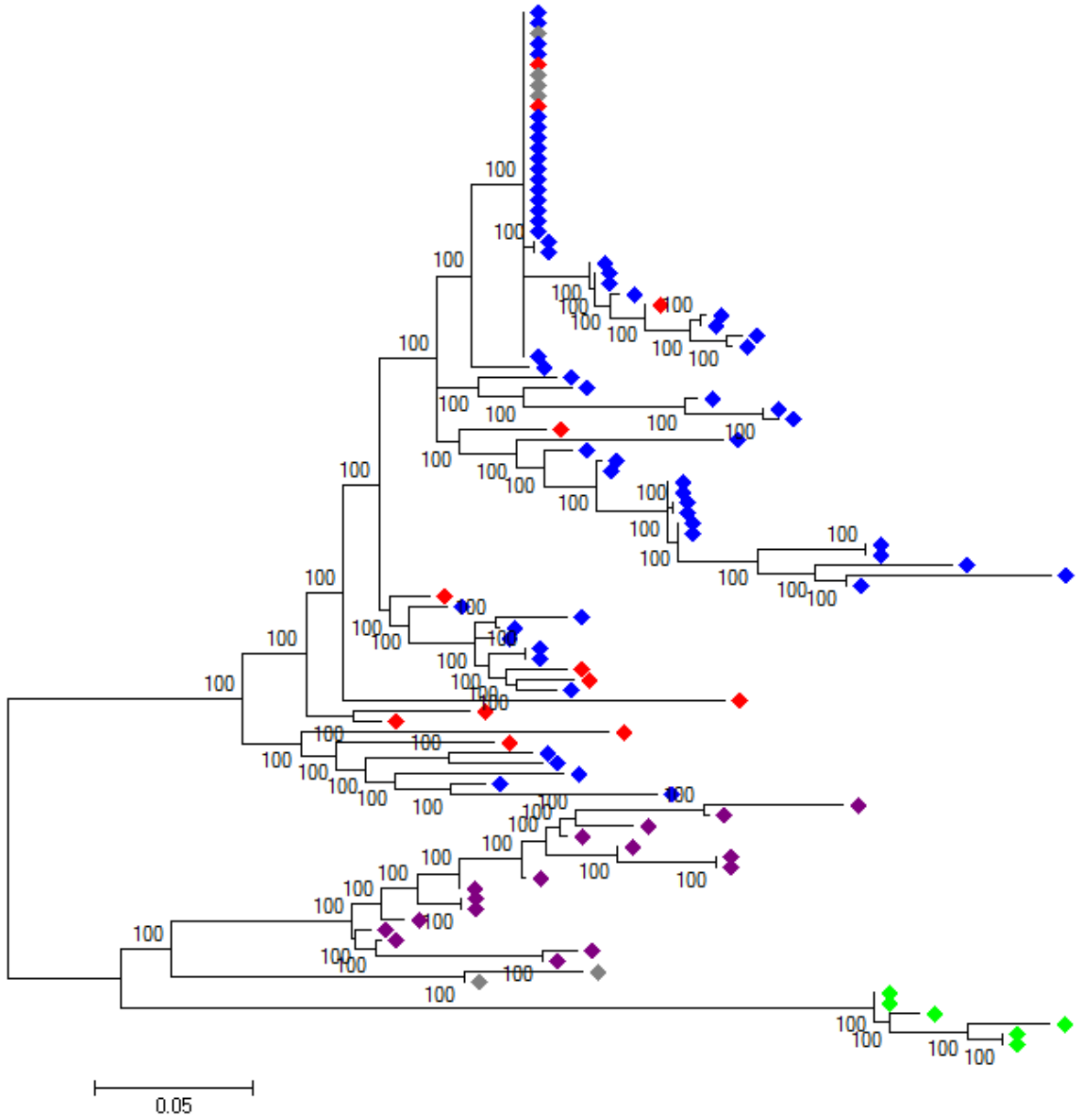


Figure S7. 304 Chromosome 20: 13245001 - 13265001

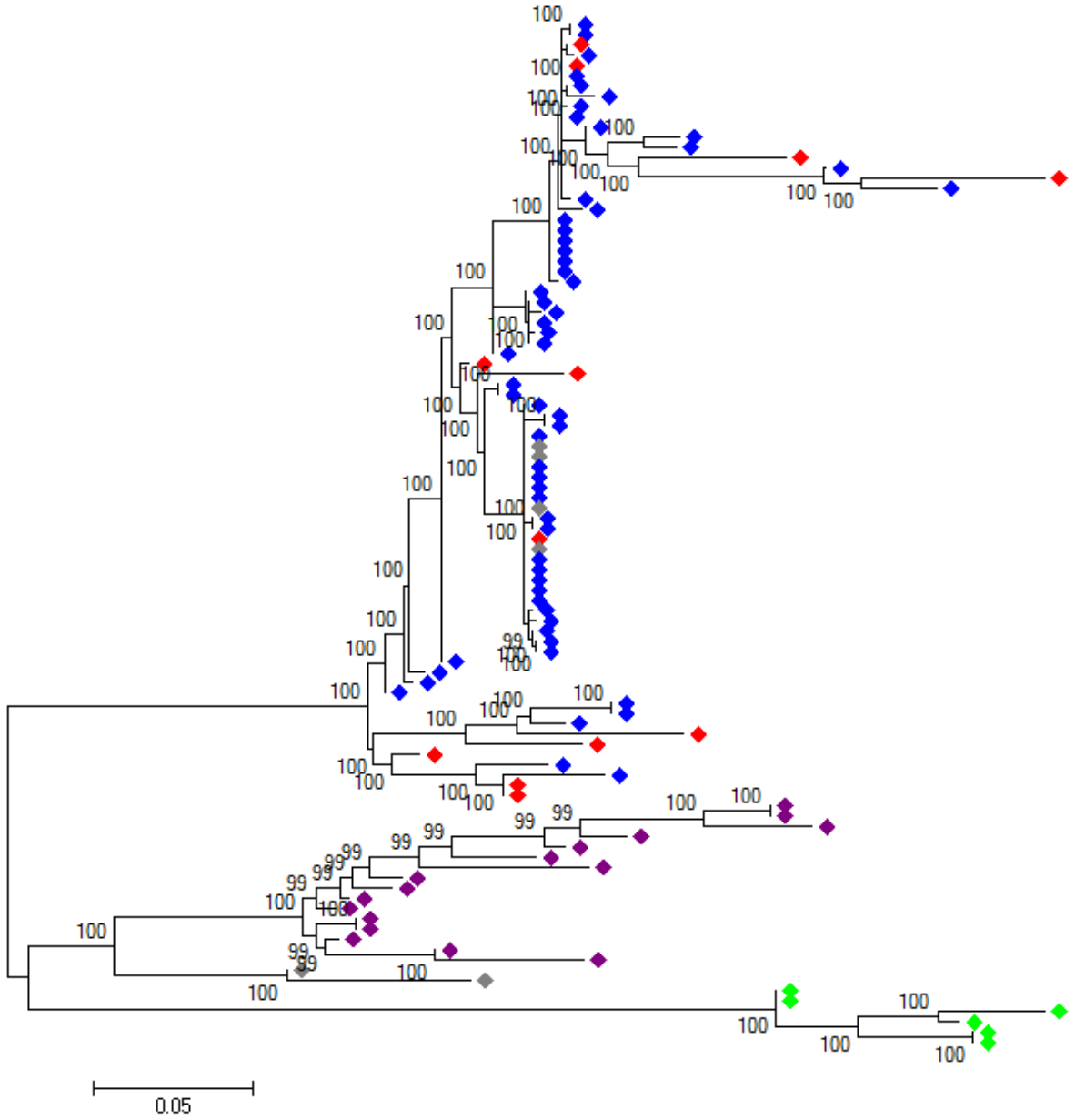


Figure S7. 305 Chromosome 20: 13565001 - 13595001

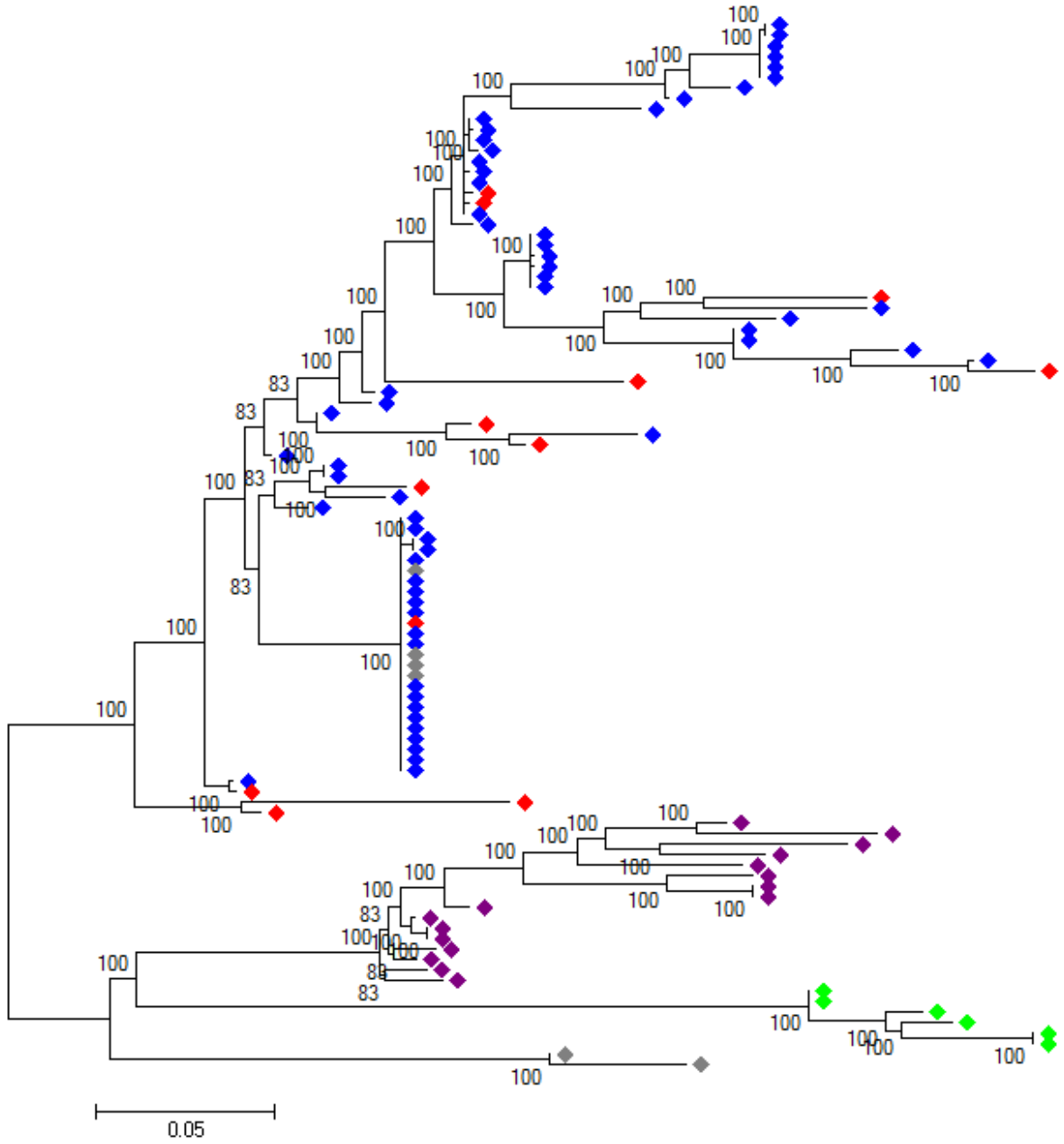


Figure S7. 306 Chromosome 20: 13610001 - 13630001

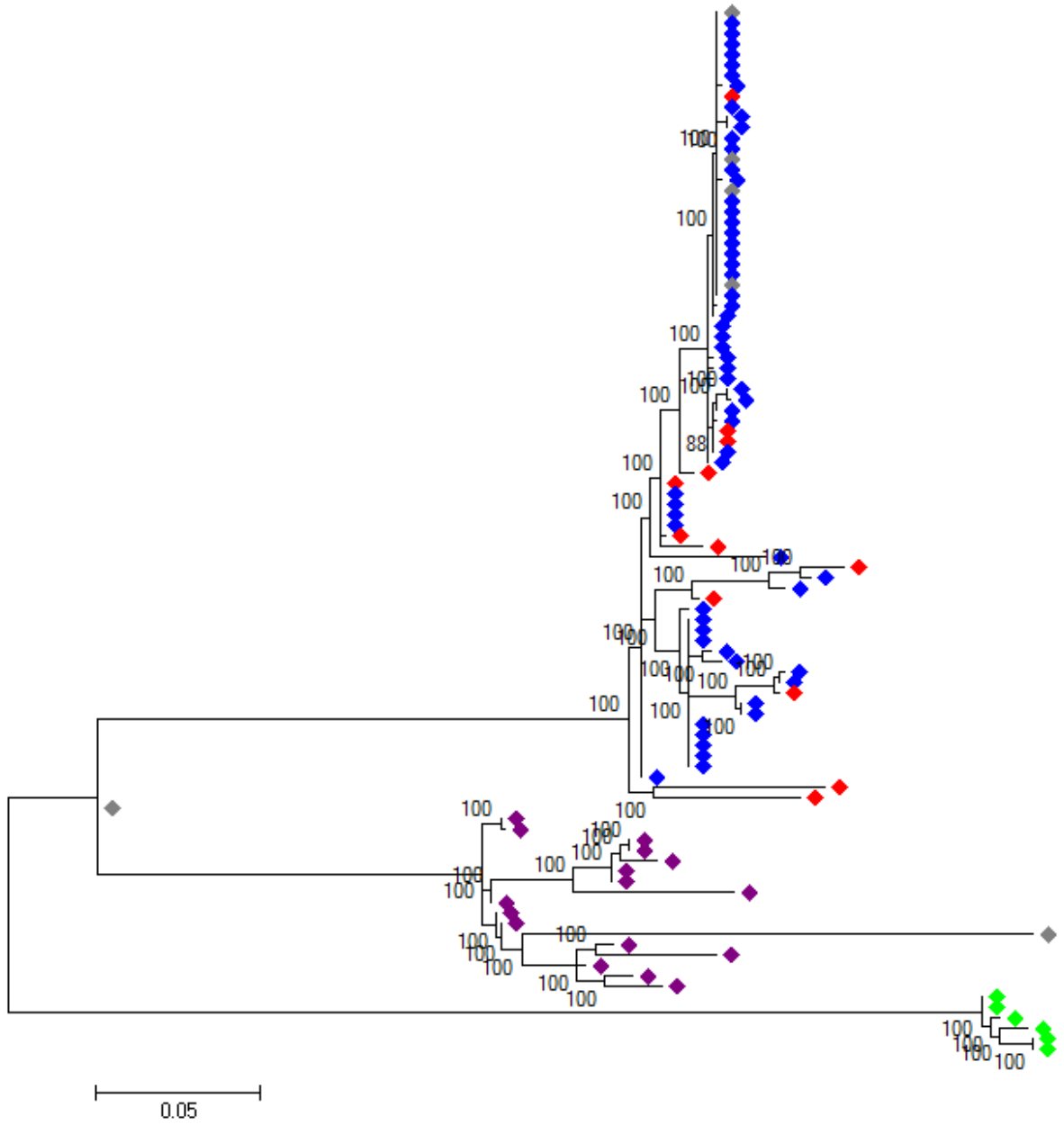


Figure S7. 307 Chromosome 20: 13855001 - 13890001

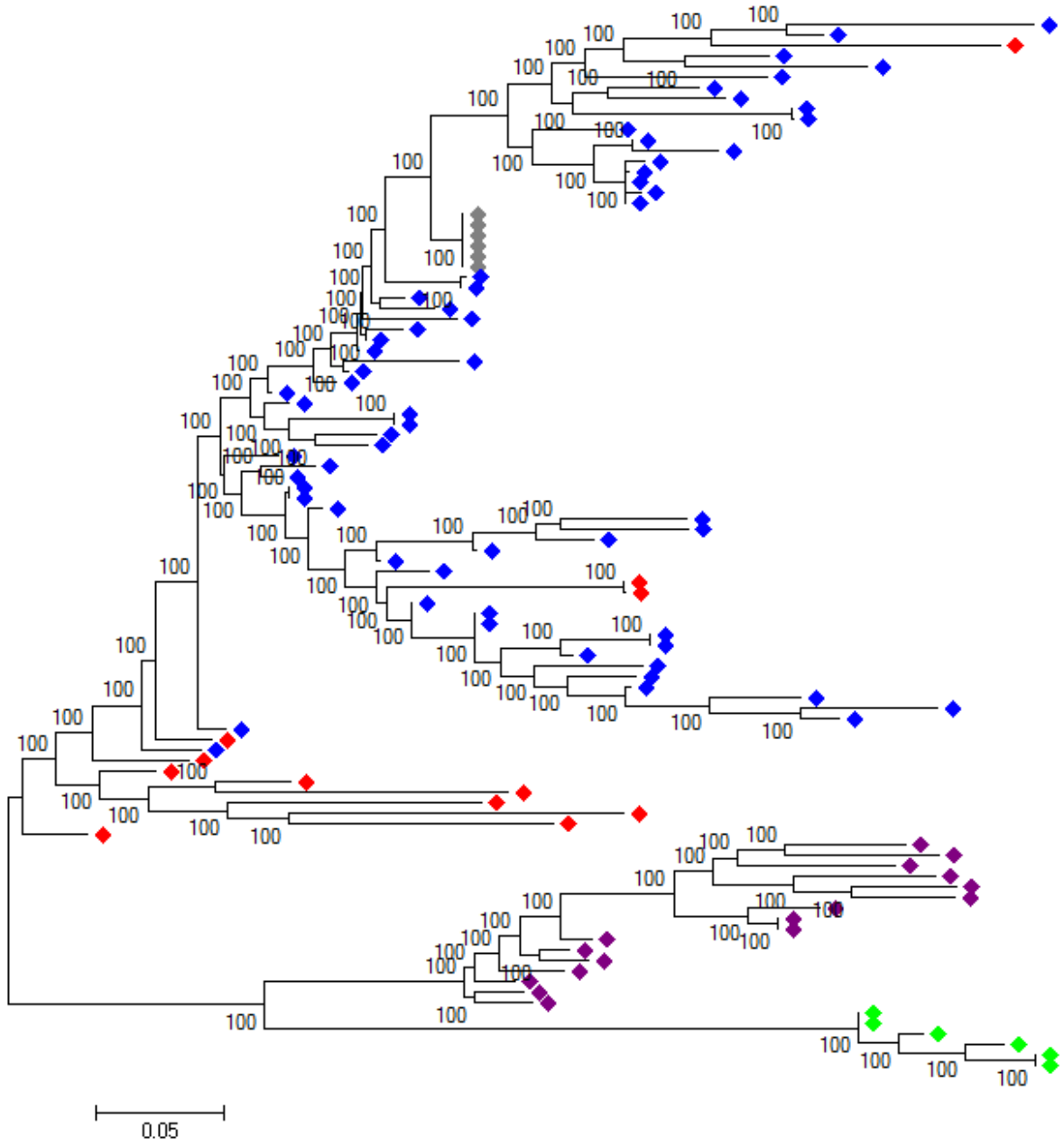


Figure S7. 308 Chromosome 21: 2695001 - 273000

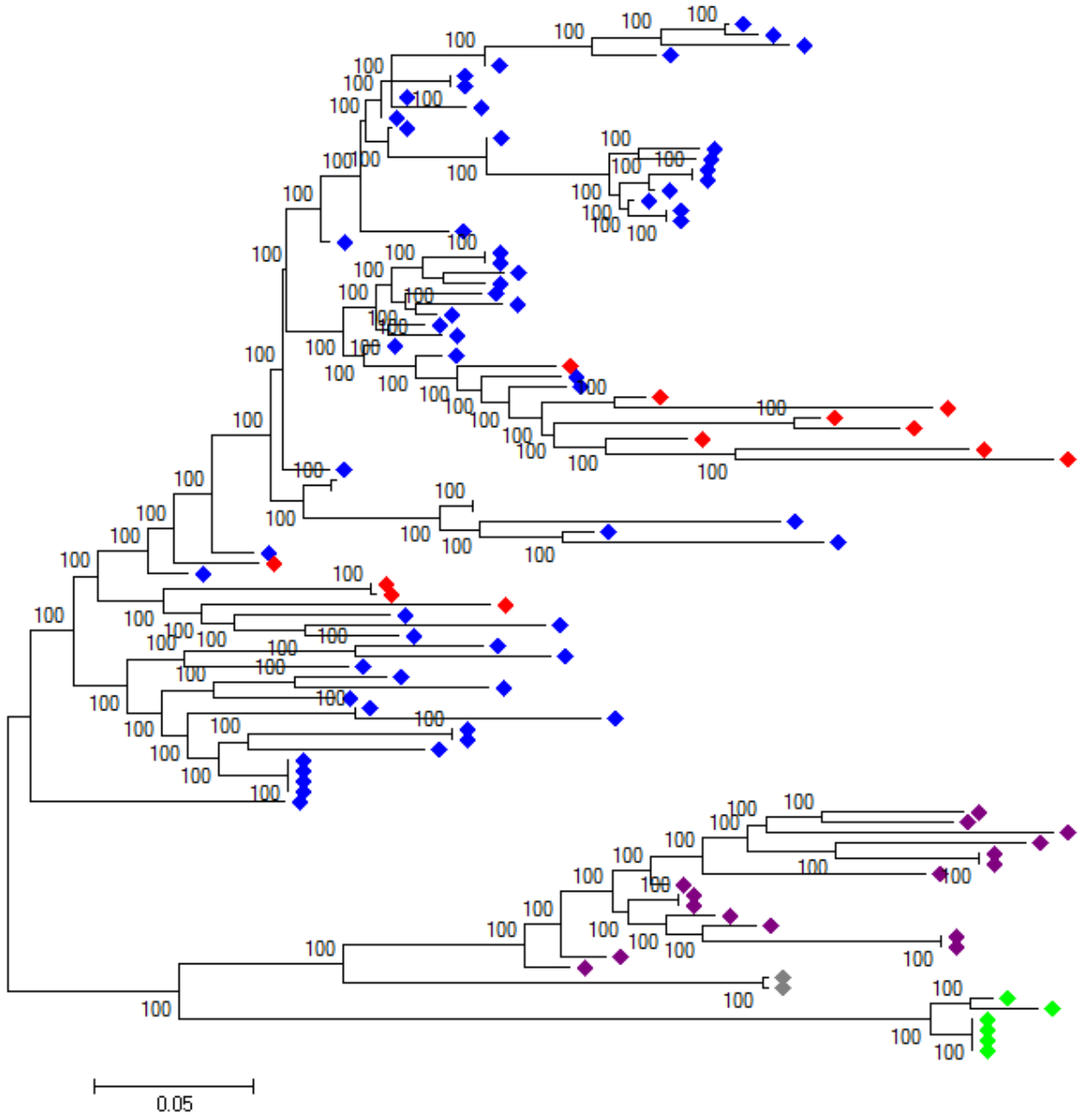


Figure S7. 309 Chromosome 21: 3785001 - 3805001

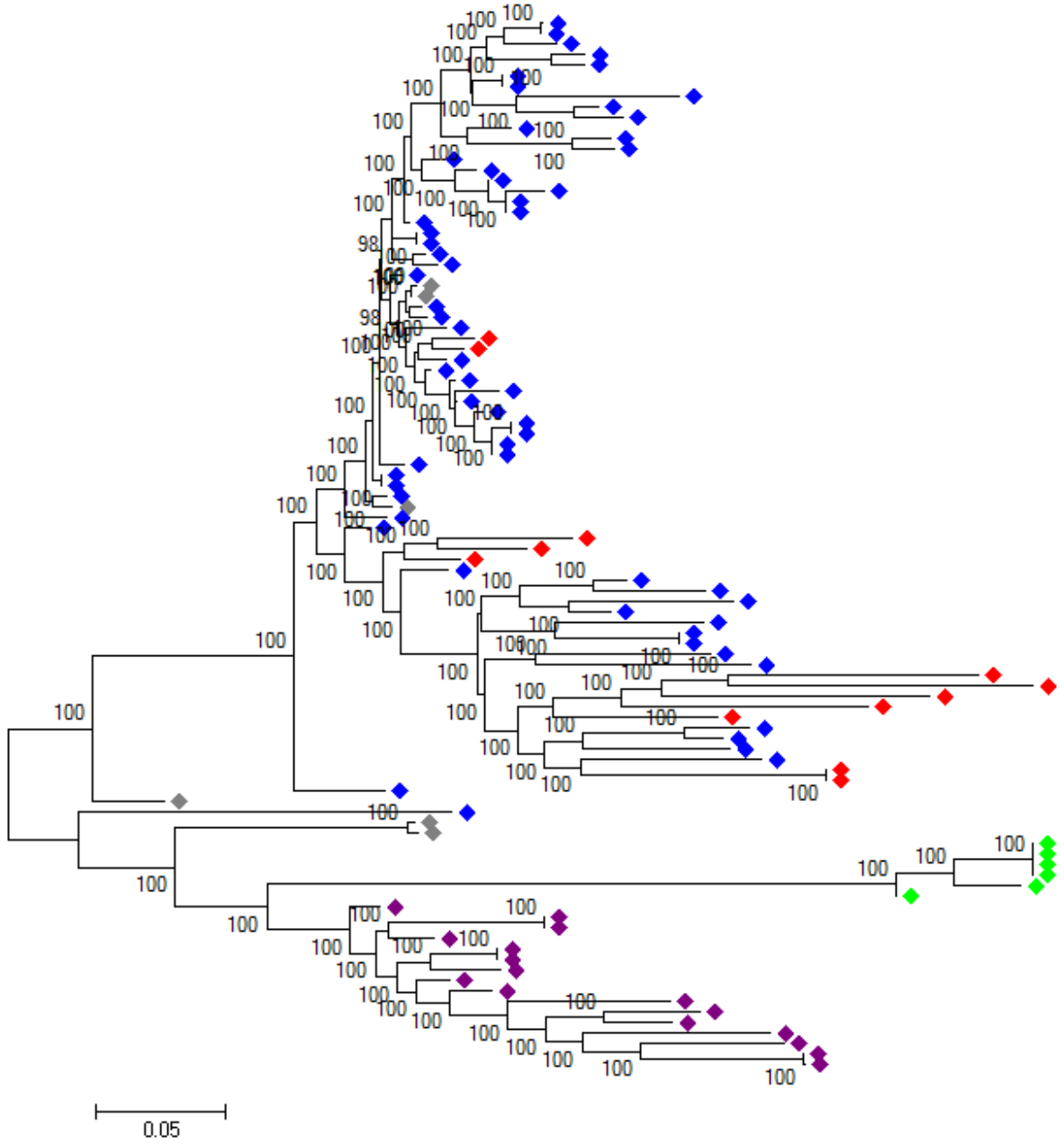


Figure S7. 310 Chromosome 21: 3810001 - 3855001

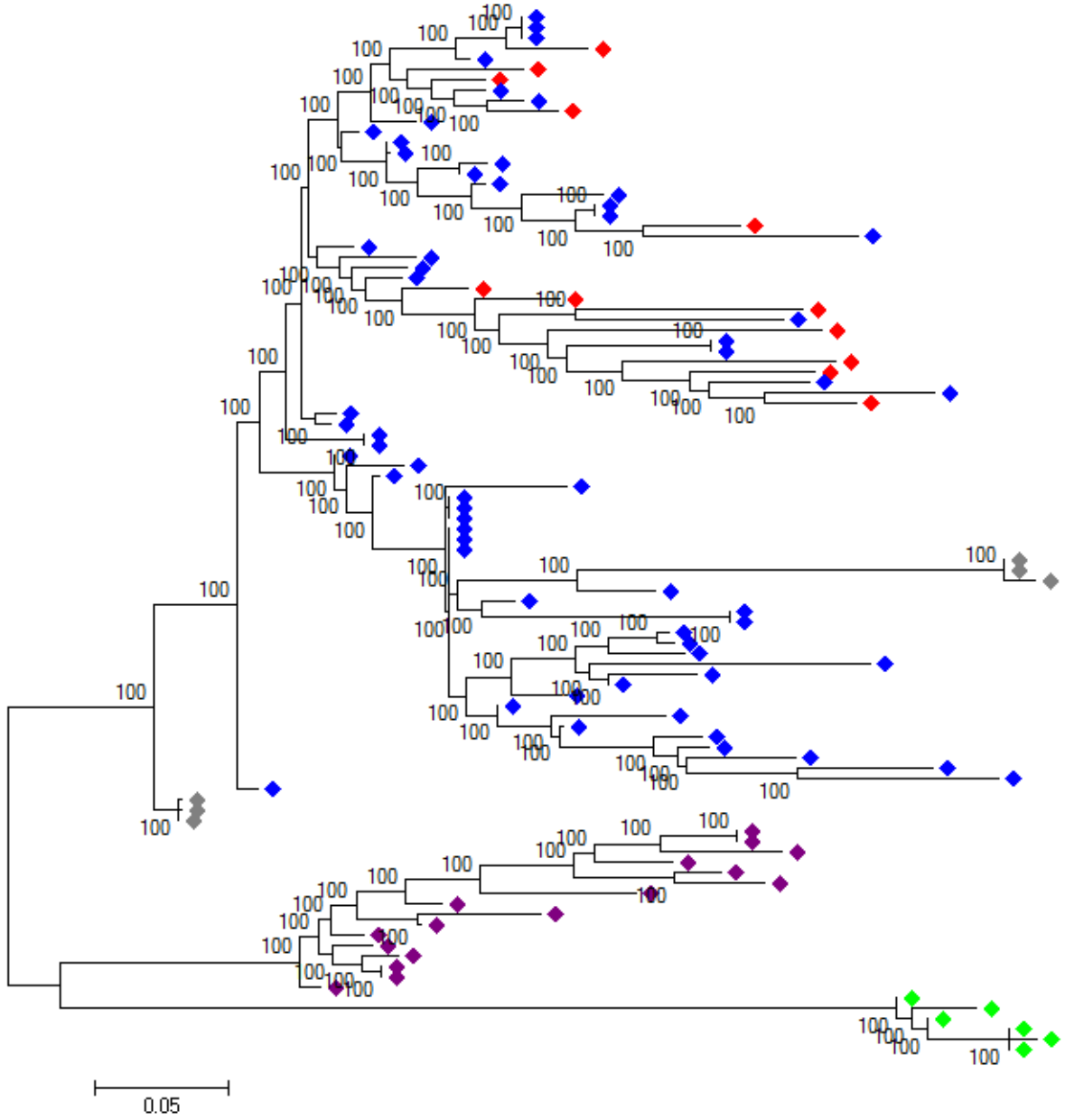


Figure S7. 311 Chromosome 21: 4275001 - 4295001

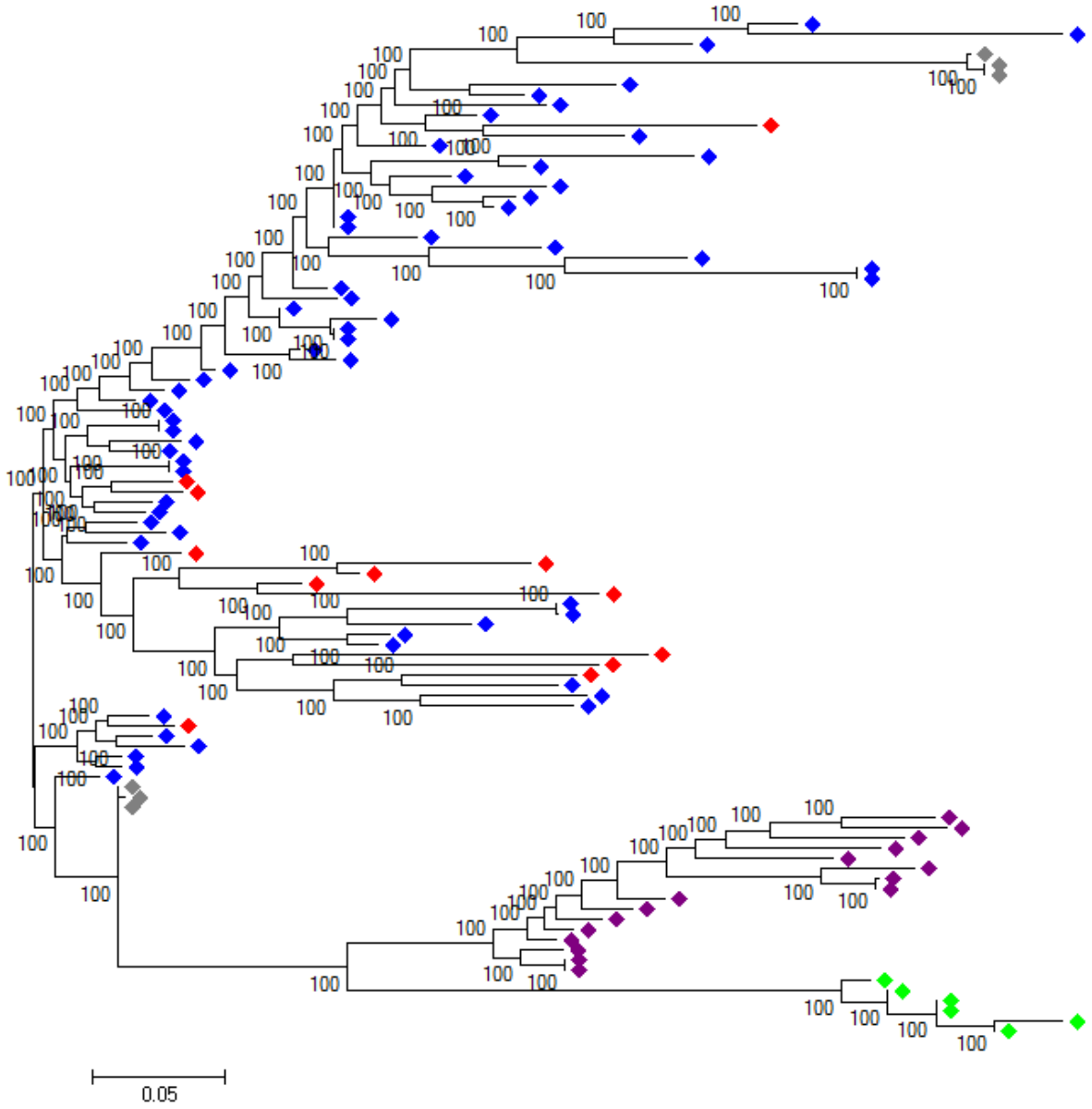


Figure S7. 312 Chromosome 21: 4320001 - 4355001

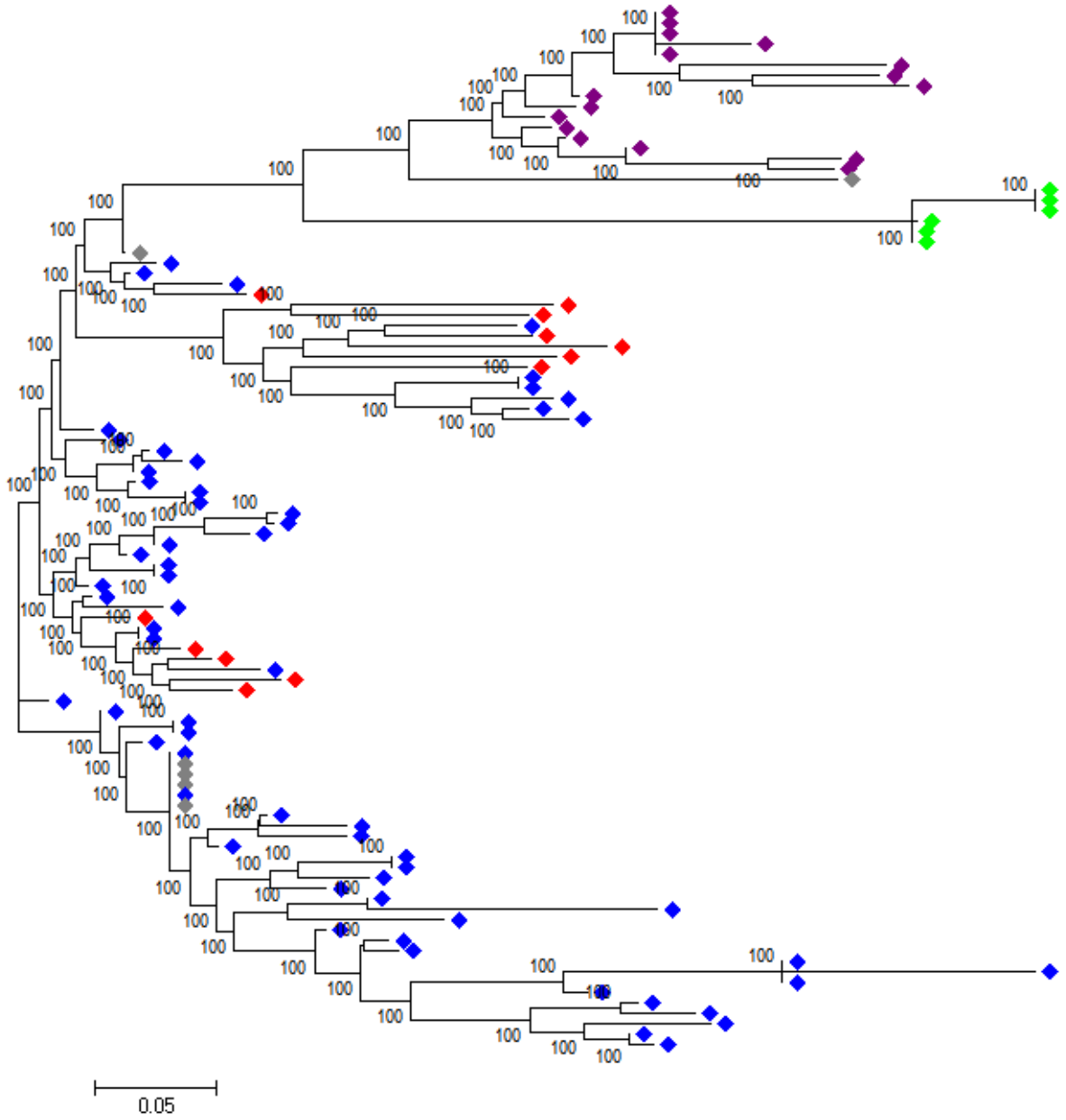


Figure S7. 313 Chromosome 26: 1870001 - 1890001

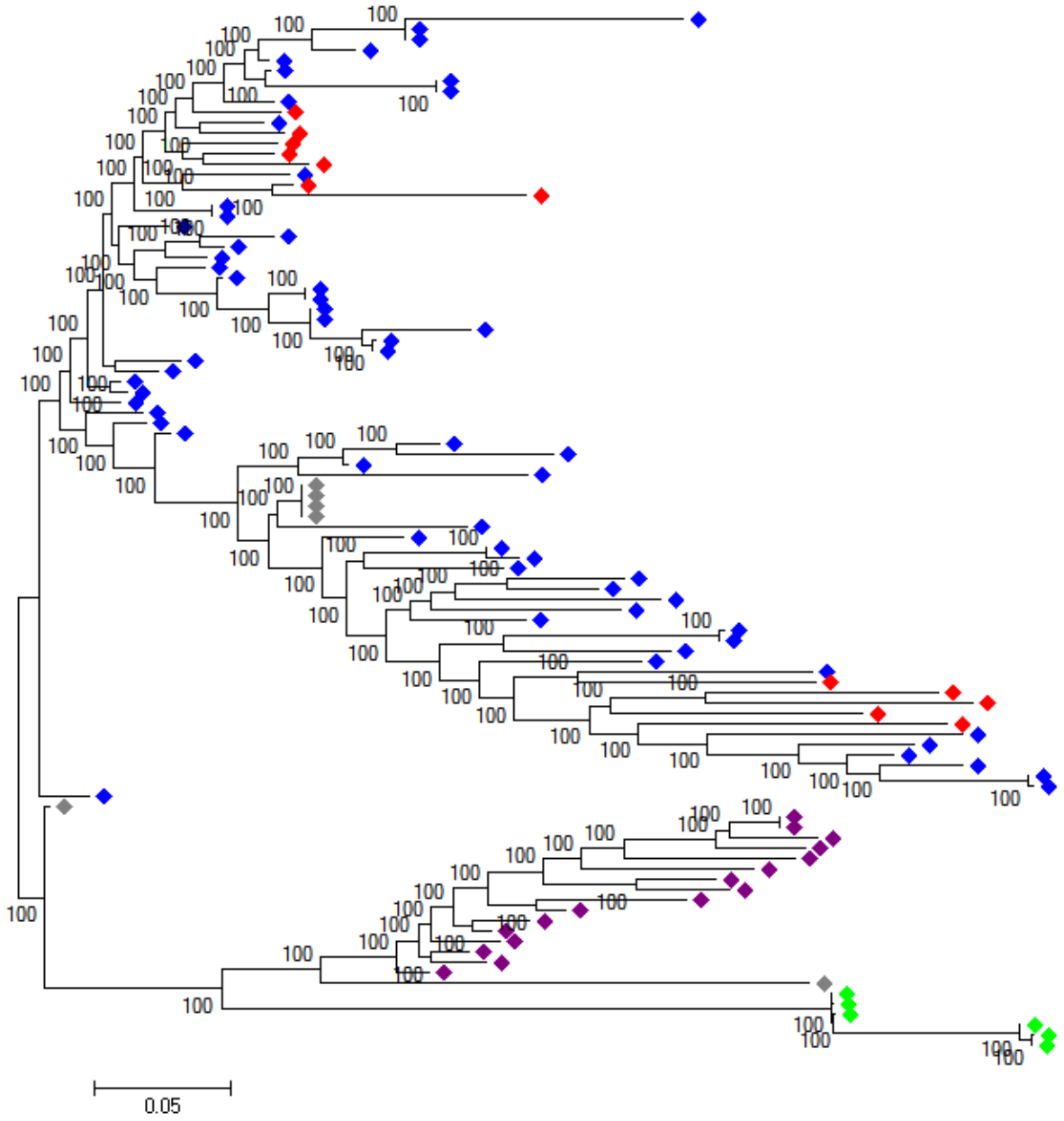


Figure S7. 314 Chromosome 26: 2040001 - 2075001

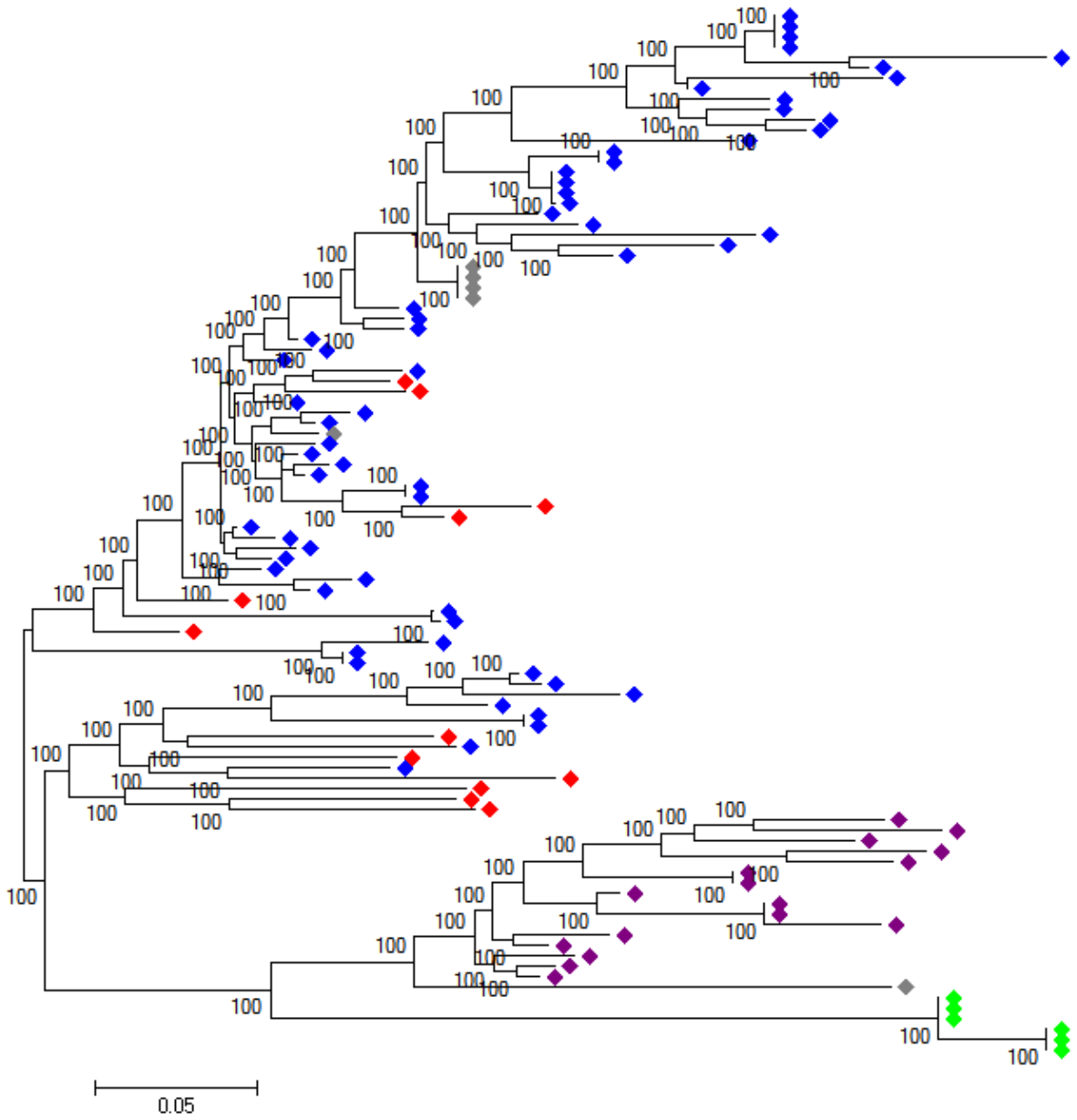


Figure S7. 315 Chromosome 26: 2120001 - 2145001

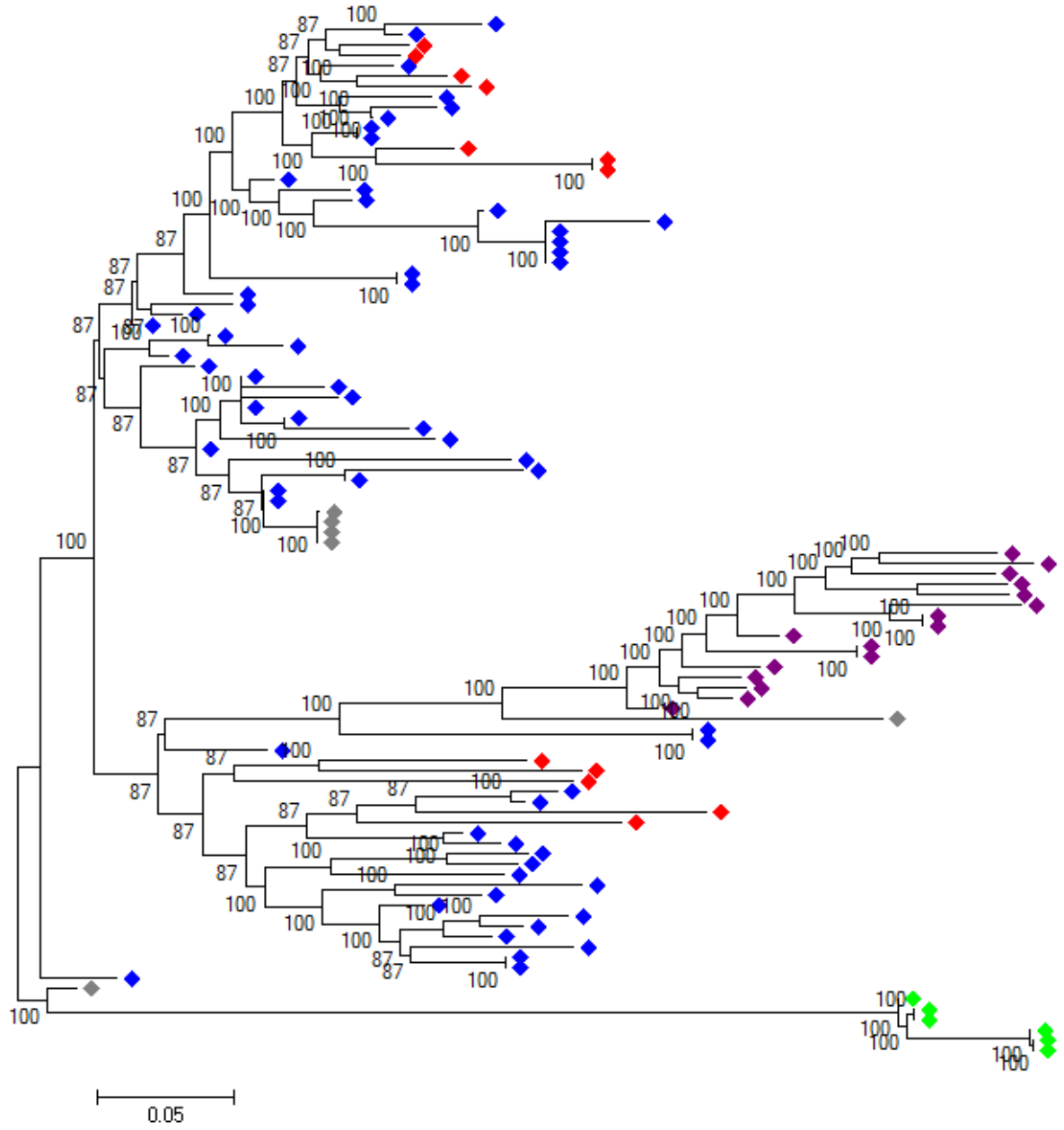


Figure S7. 316 Chromosome 26: 2170001 - 2190001

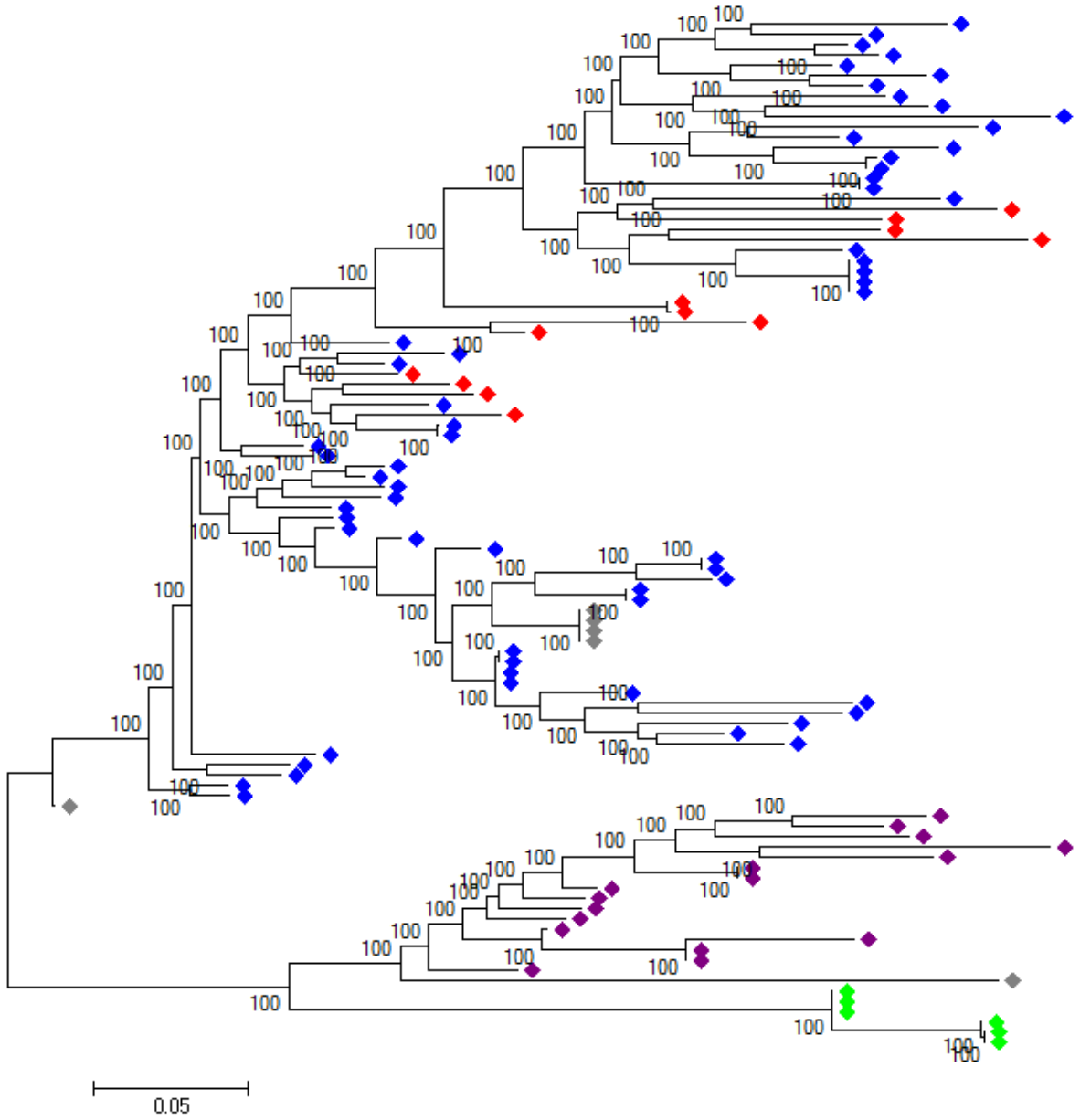


Figure S7. 317 Chromosome 26: 2215001 - 2245001

Figures S8. The regions reported in this section are peaks above the genome threshold in the grey junglefowl - domestic chicken ABBA/BABA analysis, but the haplotype trees show inconclusive relationships.

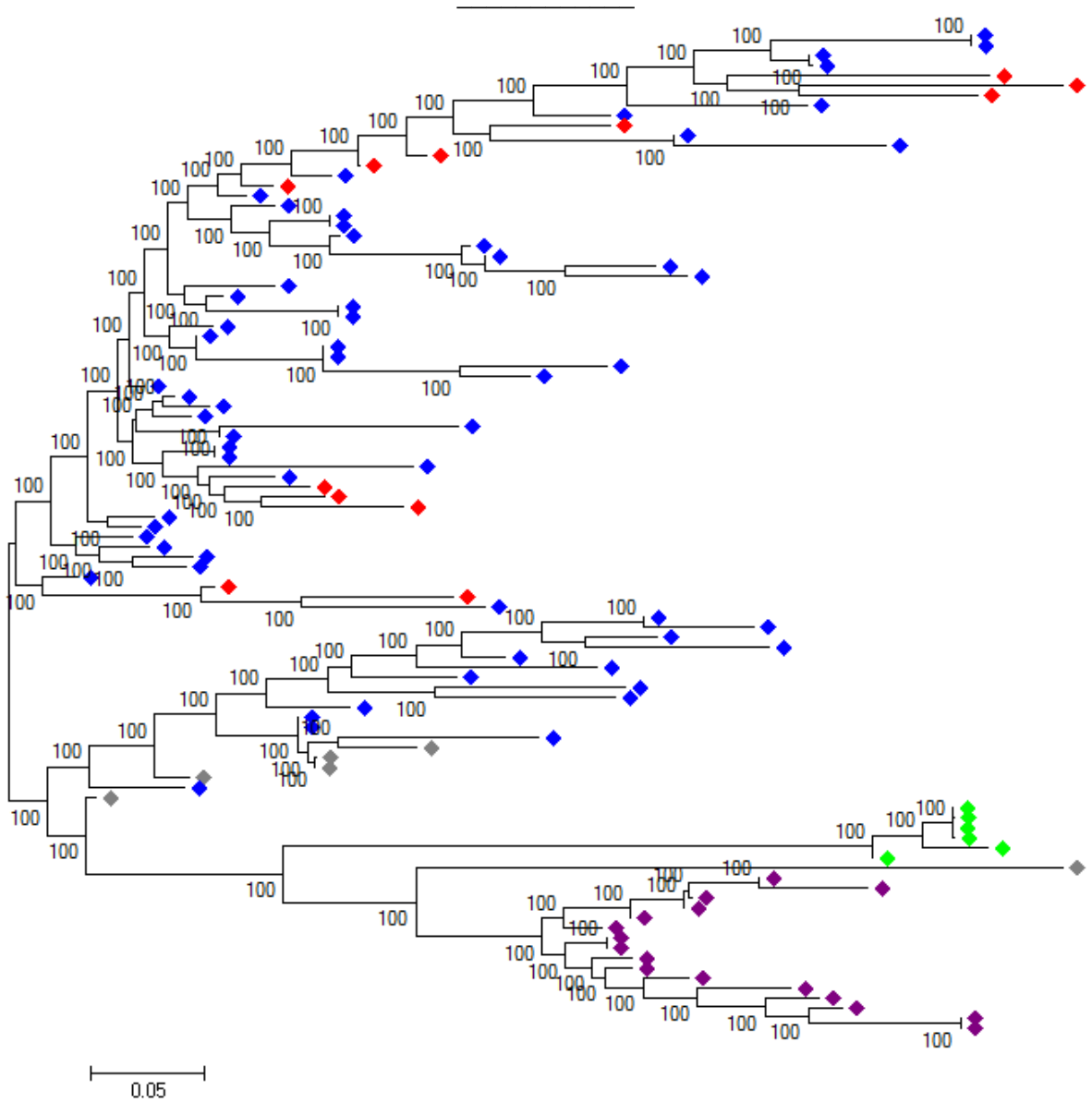


Figure S8.1. Haplotype based maximum likelihood tree of grey junglefowl into domestic chicken or domestic chicken into grey junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 50-kb fragment size that extends from 45865001 to 45915001 bp on chromosome 4. The tree is inconclusive as it suggests either introgression of domestic chicken into grey junglefowl or introgression of grey into domestic chicken.

Figures S9. Haplotype relationships analysis for candidate regions of introgression from Ceylon junglefowl into domestic chicken. ABBA – BABA Ceylon junglefowl - domestic chicken.

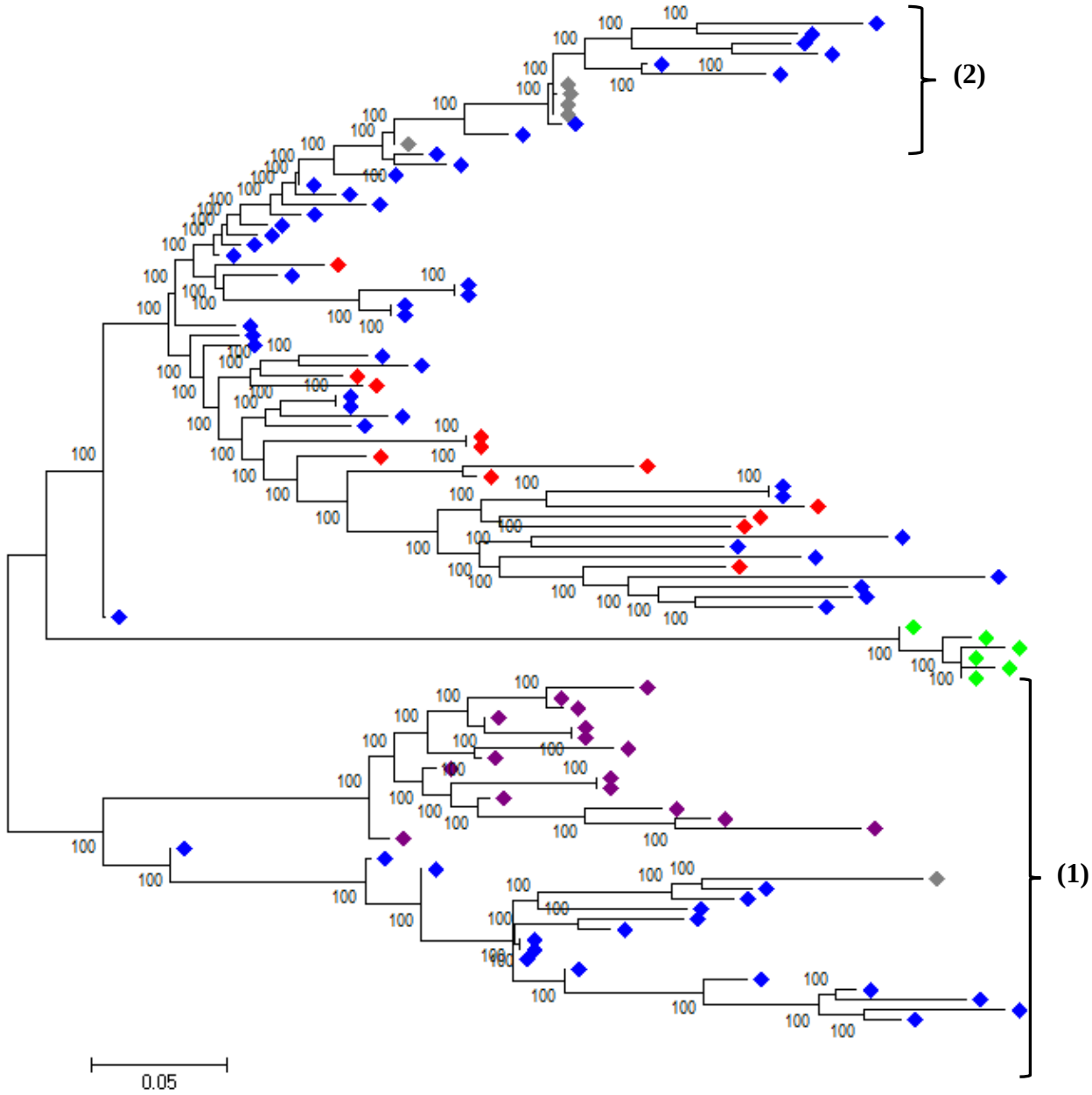


Figure S9.1. Haplotype based maximum likelihood tree for two patterns of introgression for a region of 30-kb (15525001 to 15555001 bp) on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of Ceylon junglefowl haplotype to samples from Ethiopia chicken as well as Sri Lanka4, followed by an introgression of domestic chicken into grey junglefowl. (**2)** Introgression of domestic chicken into grey junglefowl.

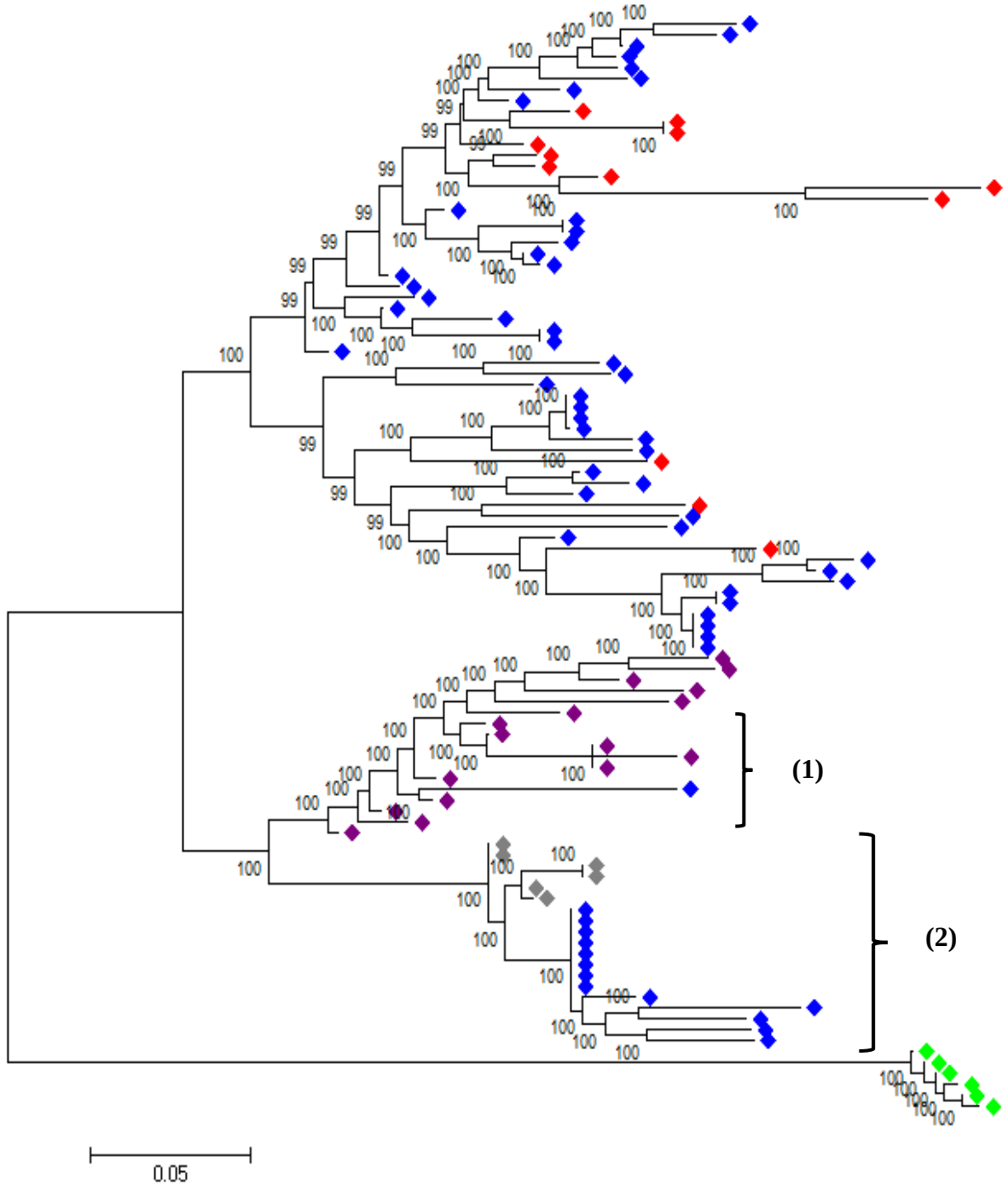


Figure S9.2. Haplotype based maximum likelihood tree of 35-kb fragment size that extends from 25800001 to 25835001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl**. (1**) Evidence of introgression from Ceylon junglefowl into a single Sri Lanka domestic bird. (**2**) The same region has also been introgressed from the grey junglefowl into several domestic chickens.

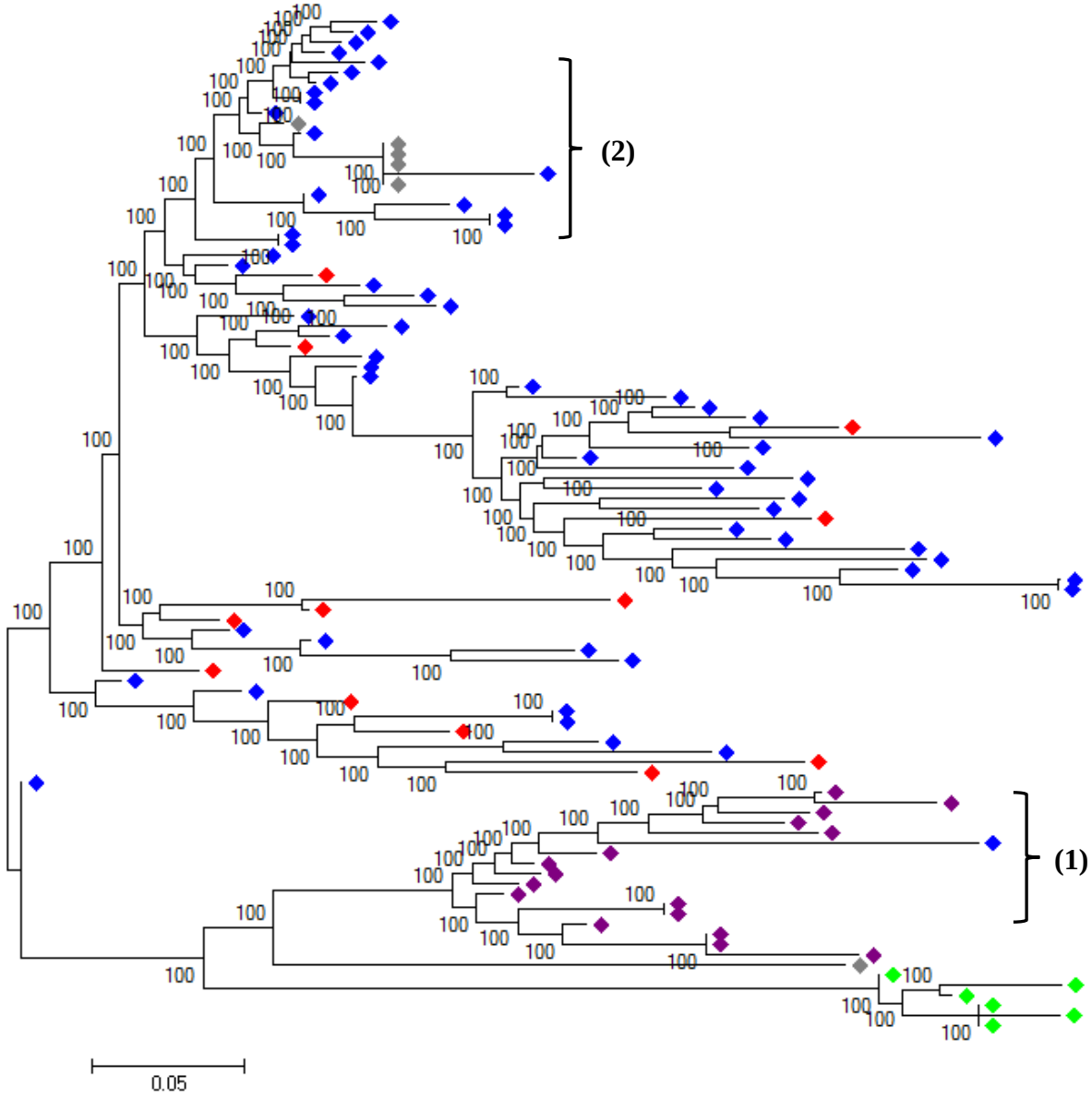


Figure S9.3. Haplotype based maximum likelihood tree for two patterns of introgression. 75-kb fragment size that extend from 145735001 to 145810001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)**. Introgression of Ceylon junglefowl haplotype to Sri Lanka16. **(2)** Introgression of domestic chicken into grey junglefowl.

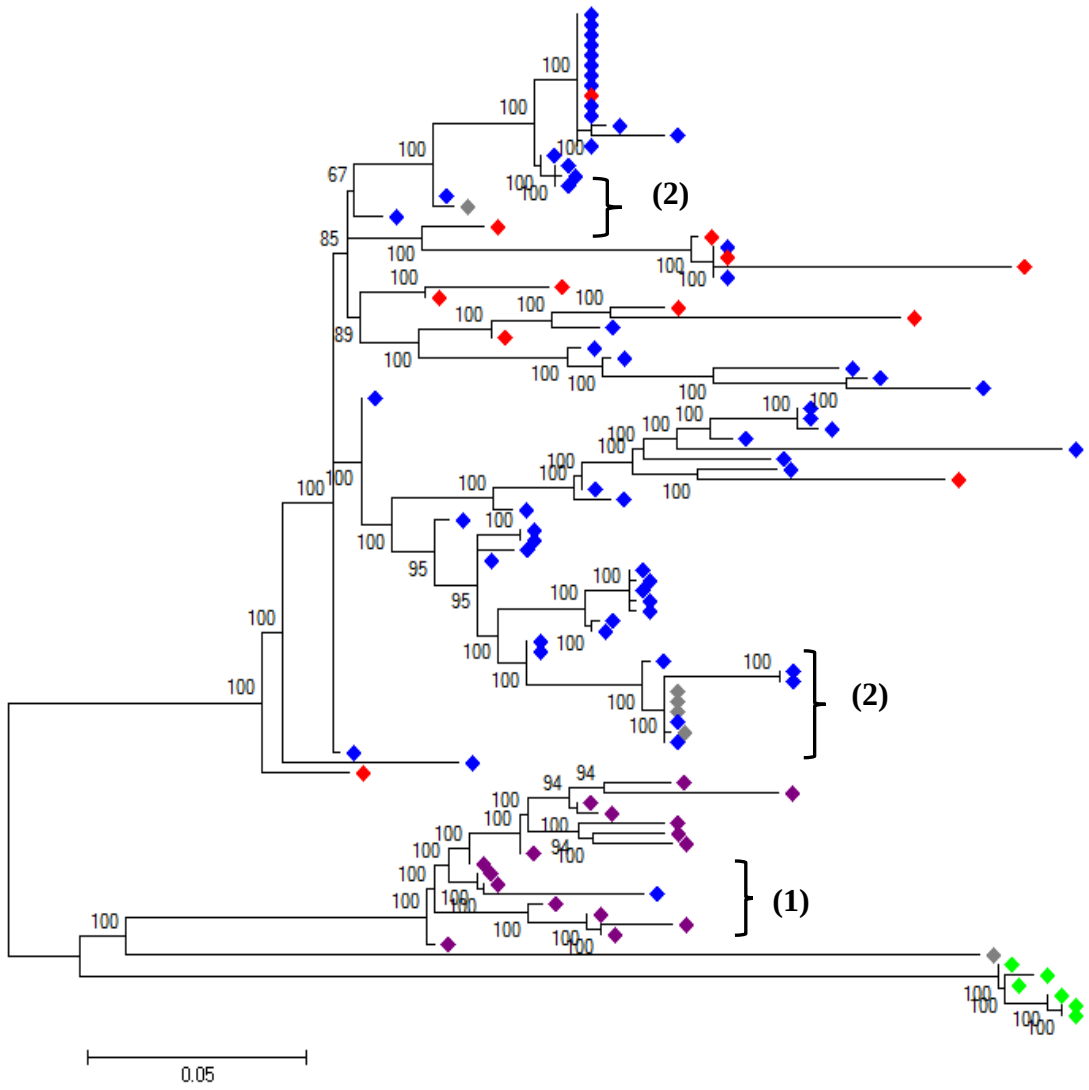


Figure S9.4. Haplotype based maximum likelihood tree illustrating introgression of domestic into grey as well as of Ceylon junglefowl into domestic chicken. The region is on chromosome 1 (146075001 - 146095001 bp) and it is 20-kb in length. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. (**1**) A single introgressed haplotype from Ceylon junglefowl was present in a domestic chicken from Sri Lanka16. (**2**) Domestic chicken – red junglefowl introgression into grey junglefowl, with at least two introgression events as supported by the presence of grey junglefowl haplotypes embedded in domestic chicken in two different parts of the tree.

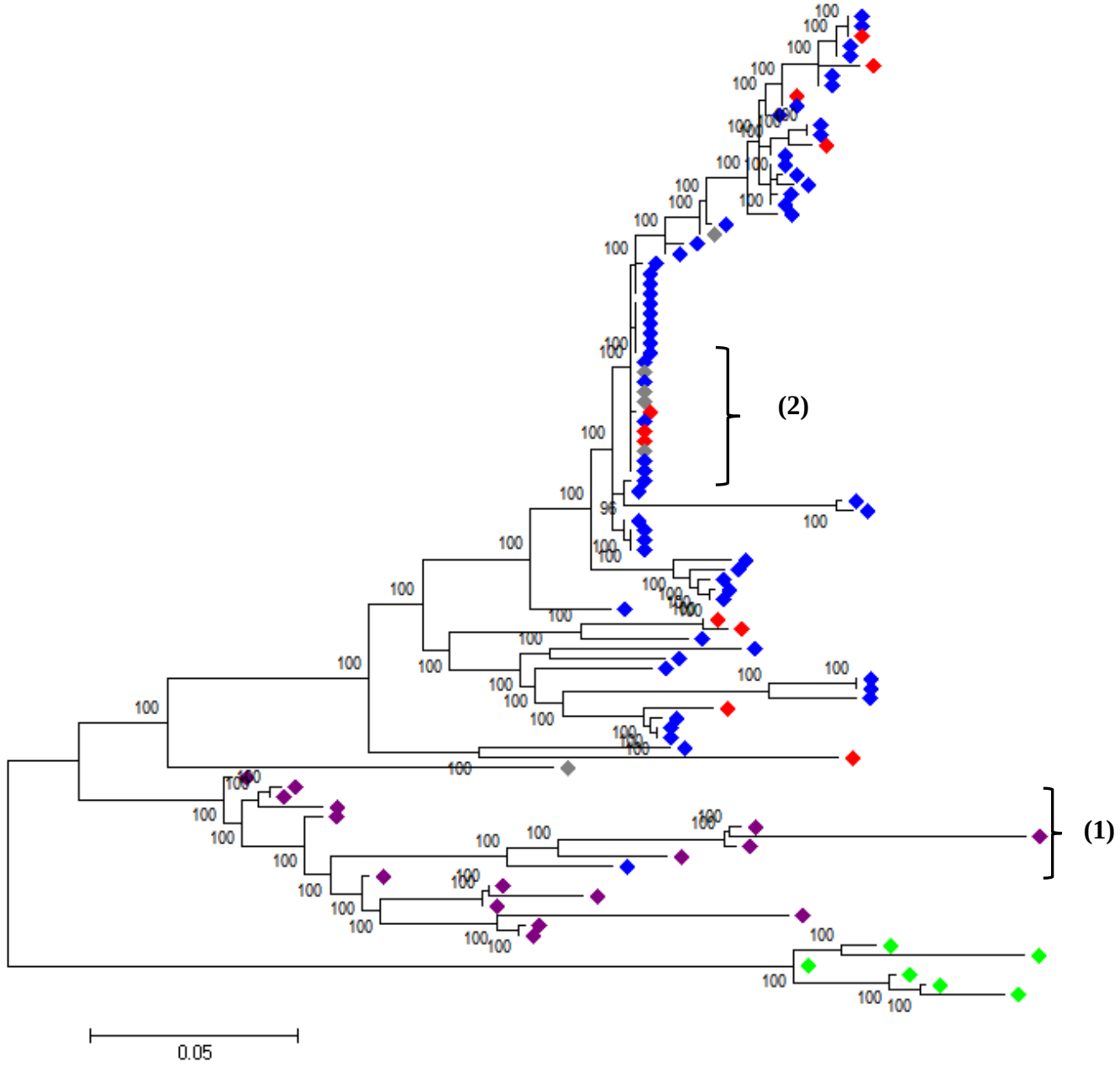


Figure S9.5. Haplotype based maximum likelihood tree illustrating introgression of domestic into grey as well as of Ceylon junglefowl into domestic chicken. The region is on chromosome 1 (146175001 – 146195001 bp) and it is 20-kb in length. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. (**1**) Domestic chicken – red junglefowl introgression into grey junglefowl. (**2**) A single introgressed haplotype from Ceylon junglefowl was present in a domestic chicken from Sri Lanka16.

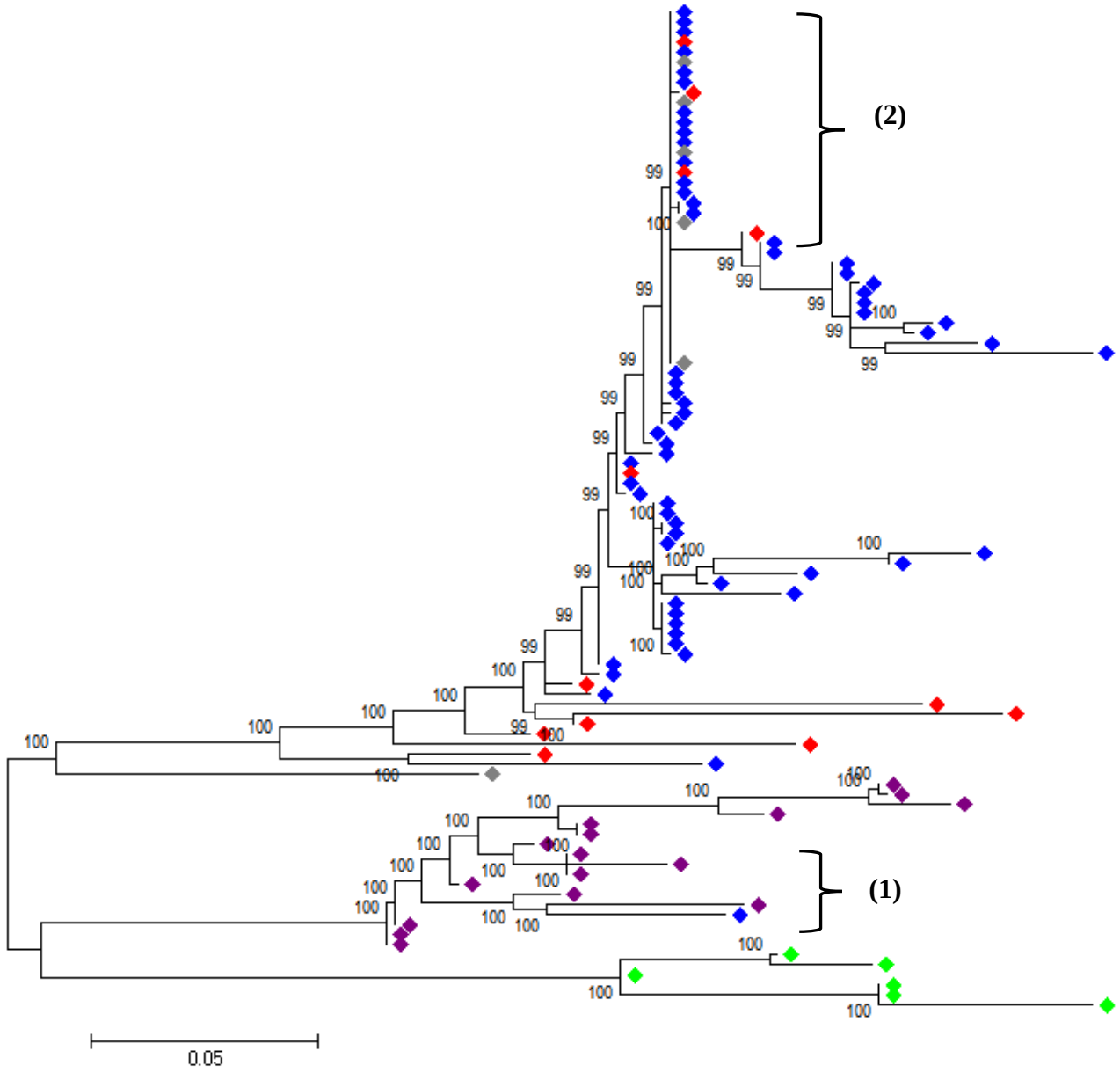


Figure S9.6. Haplotype based maximum likelihood tree illustrating introgression of domestic into grey as well as of Ceylon junglefowl into domestic chicken. The region is on chromosome 1 (146225001 – 146245001 bp) and it is 20-kb in length. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. (**1**) Domestic chicken – red junglefowl introgression into grey junglefowl. (**2**) A single introgressed haplotype from Ceylon junglefowl was present in a domestic chicken from Sri Lanka16

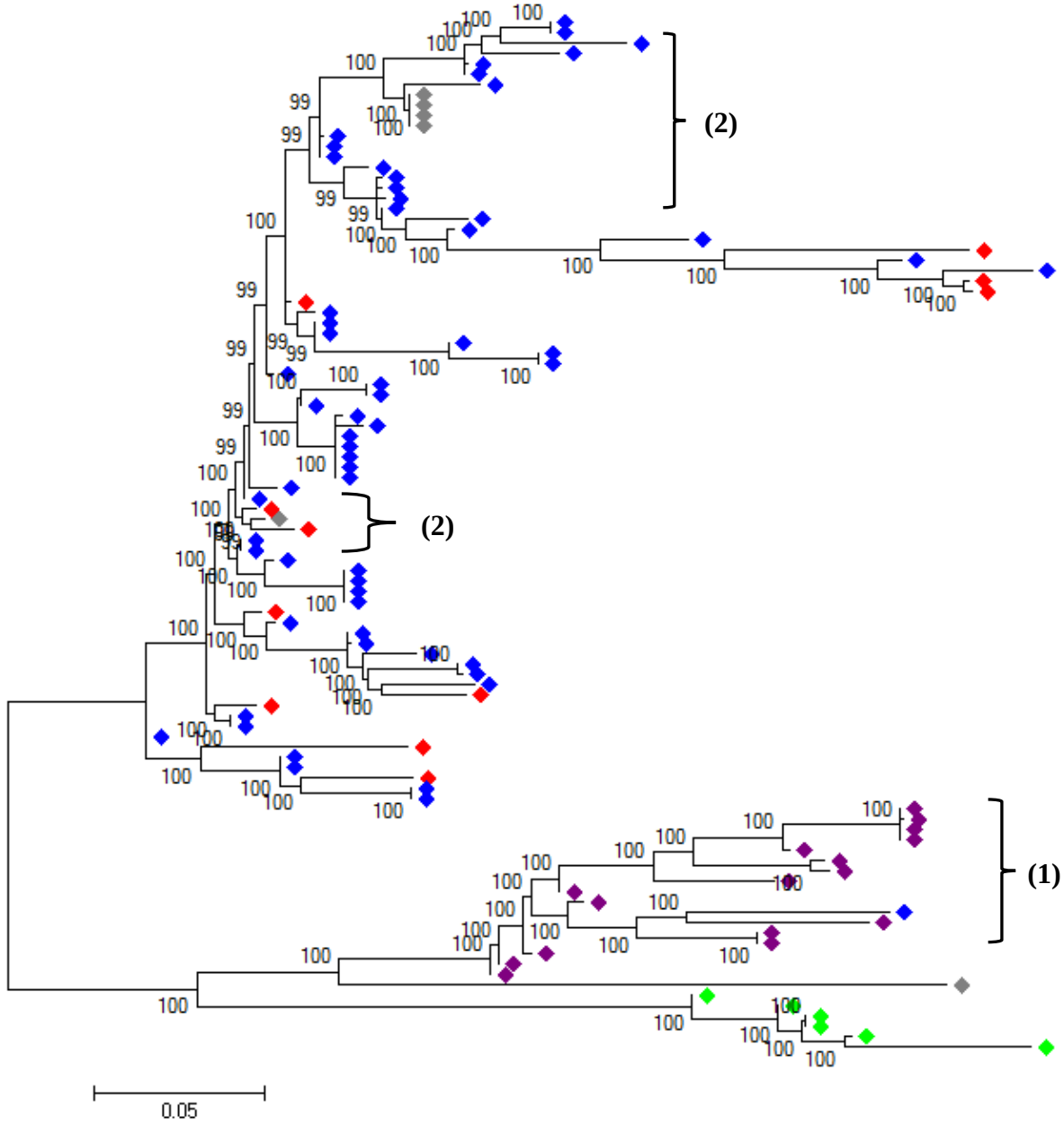


Figure S9.7. Haplotype based maximum likelihood tree for two patterns of introgression based. 25-kb fragment size that extends from 146460001 to 146485001 bp on chromosome **1.** ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl**. (1)** Introgression of a Ceylon junglefowl haplotype into Sri Lanka16. **(2)** Introgression of domestic chicken into grey junglefowl.

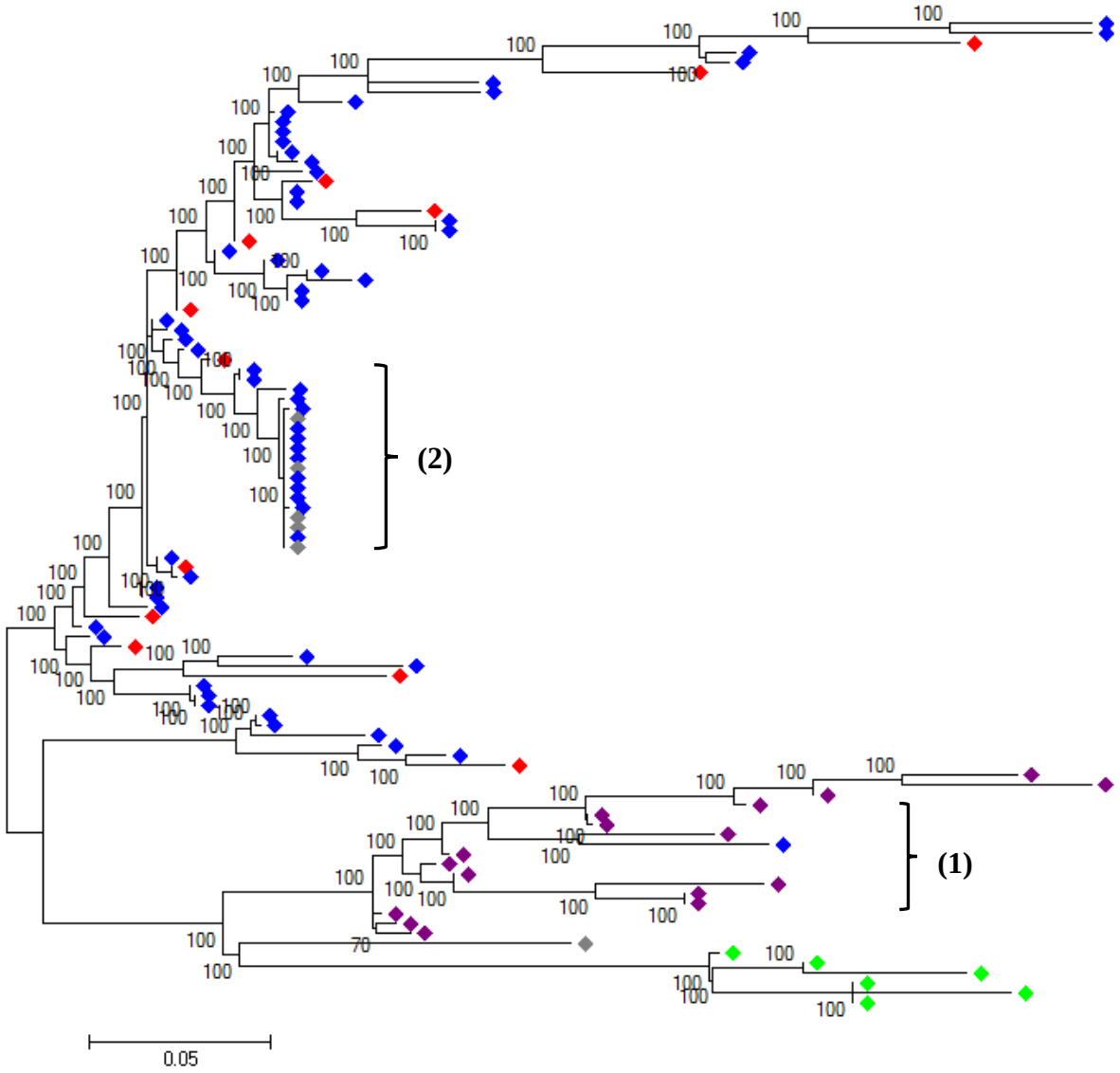


Figure S9.8. Haplotype based maximum likelihood tree for two patterns of introgression. 20-kb fragment size that extend from 146505001 to 146525001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of Ceylon junglefowl haplotype to Sri Lanka16. **(2)** Introgression of domestic chicken into grey junglefowl.

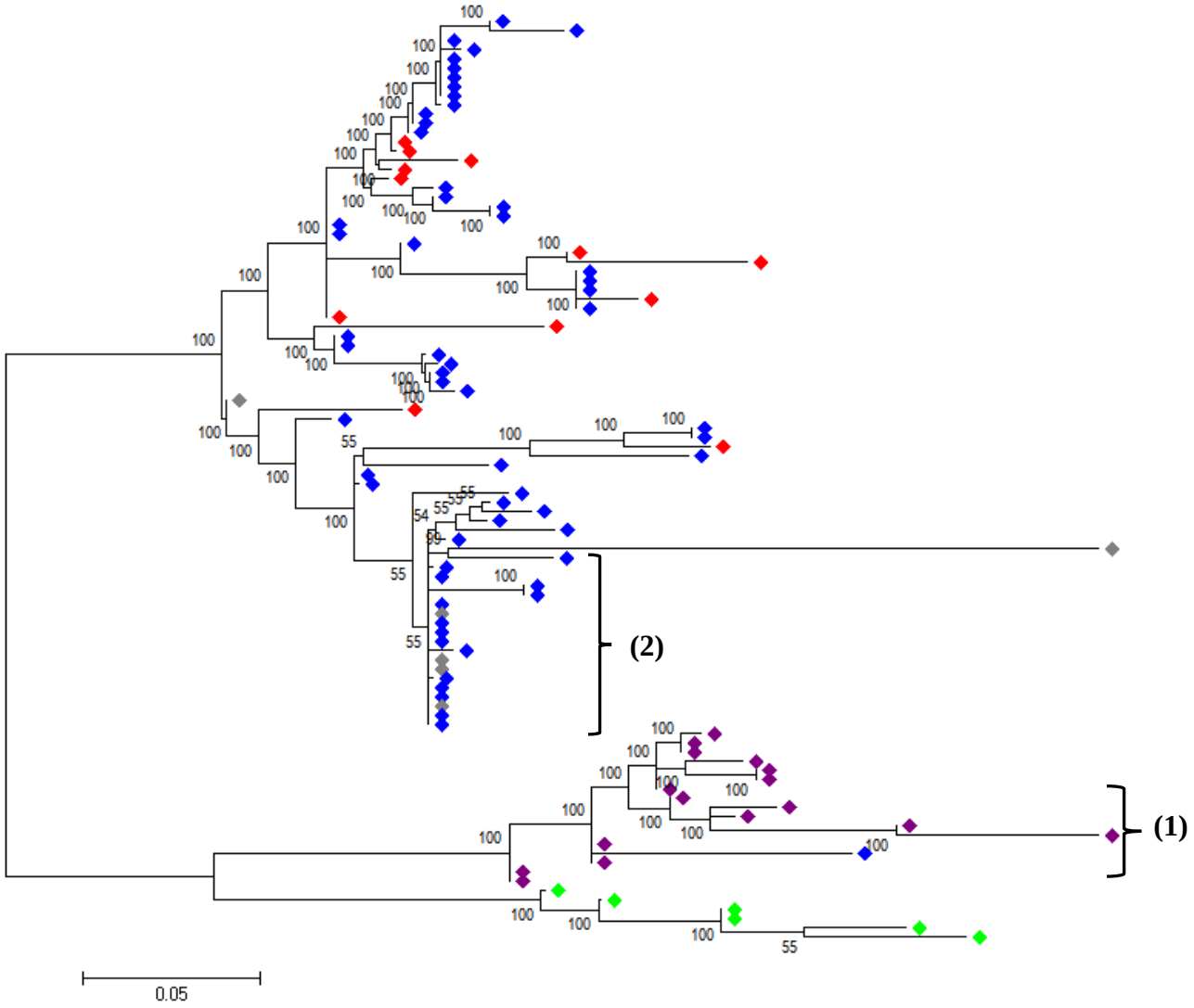


Figure S9.9. Haplotype based maximum likelihood tree for two patterns of introgression. 25-kb fragment size that extend from 146705001 to 146730001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of Ceylon junglefowl haplotype to Sri Lanka16. **(2)** Introgression of domestic chicken into grey junglefowl.

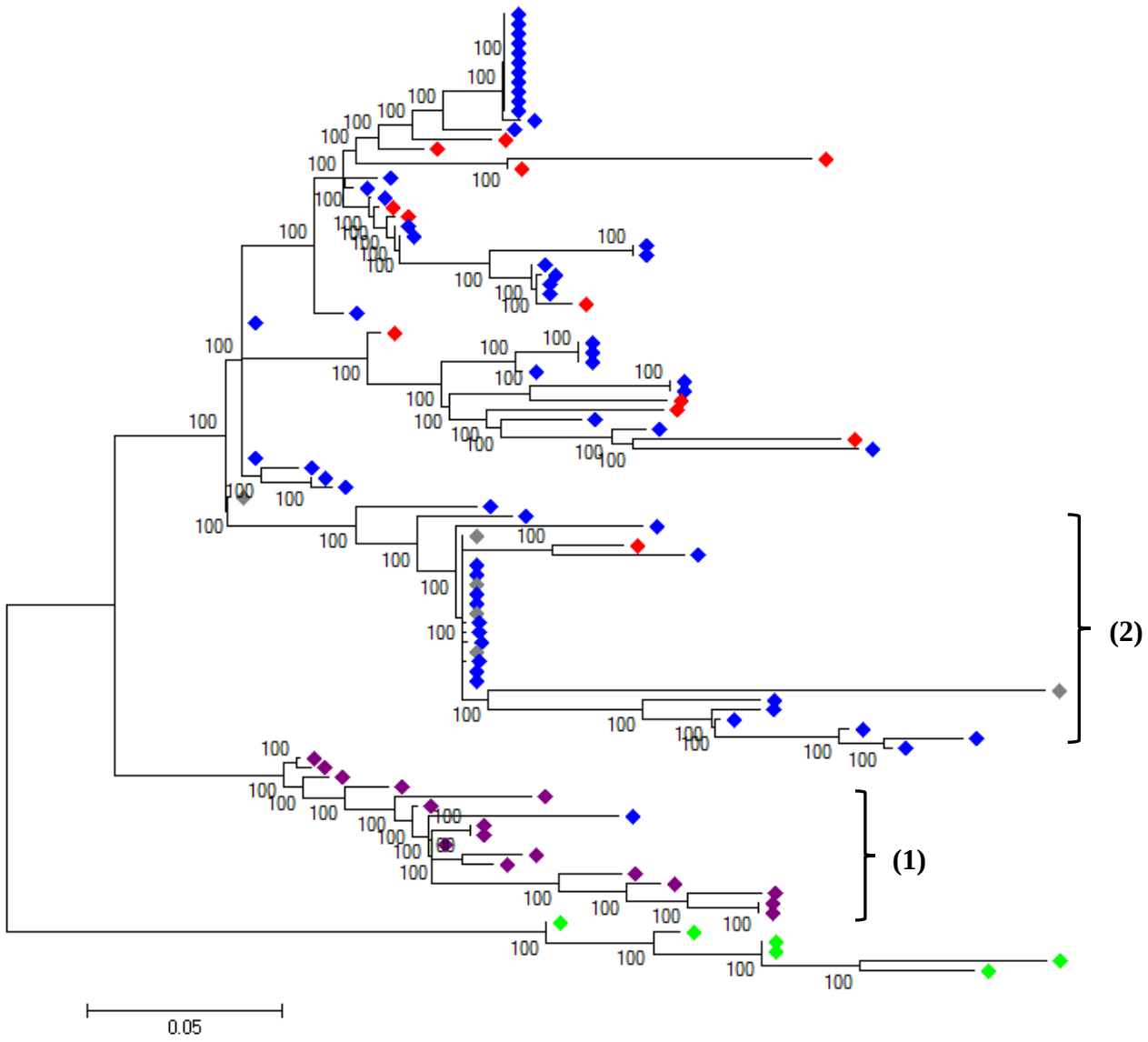


Figure S9.10. Haplotype based maximum likelihood tree for two patterns of introgression. 40-kb fragment size that extend from 146765001 to 146805001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Region of introgression of Ceylon junglefowl haplotype to Sri Lanka16. **(2)** Introgression of domestic chicken into grey junglefowl.

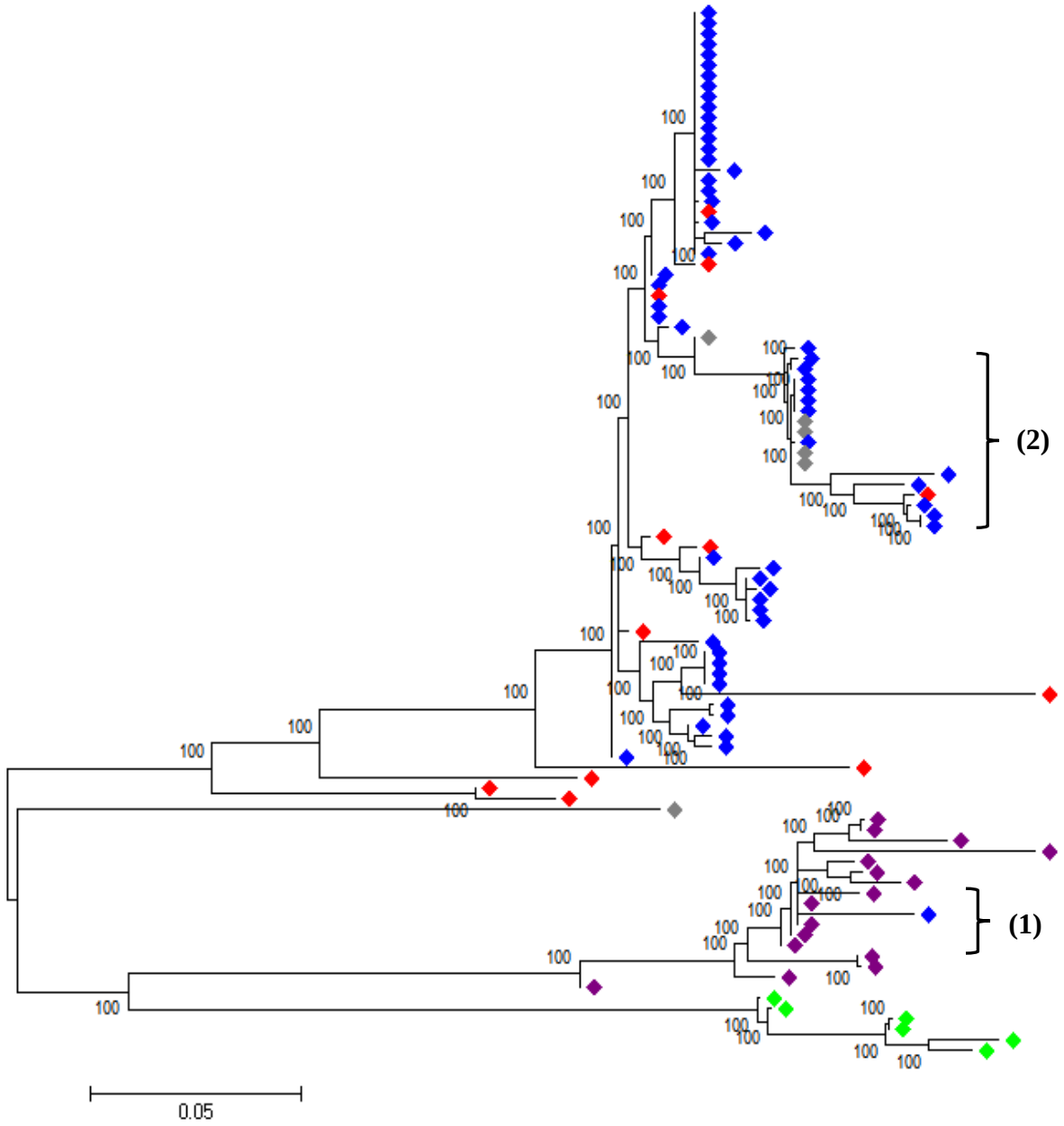


Figure S9.11. Haplotype based maximum likelihood tree for two patterns of introgression for a region of 40-kb fragment size that extend from 146965001 to 147005001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression from Ceylon junglefowl into domestic Sri Lanka16 sample. **(2)** Introgression of domestic chicken into grey junglefowl.

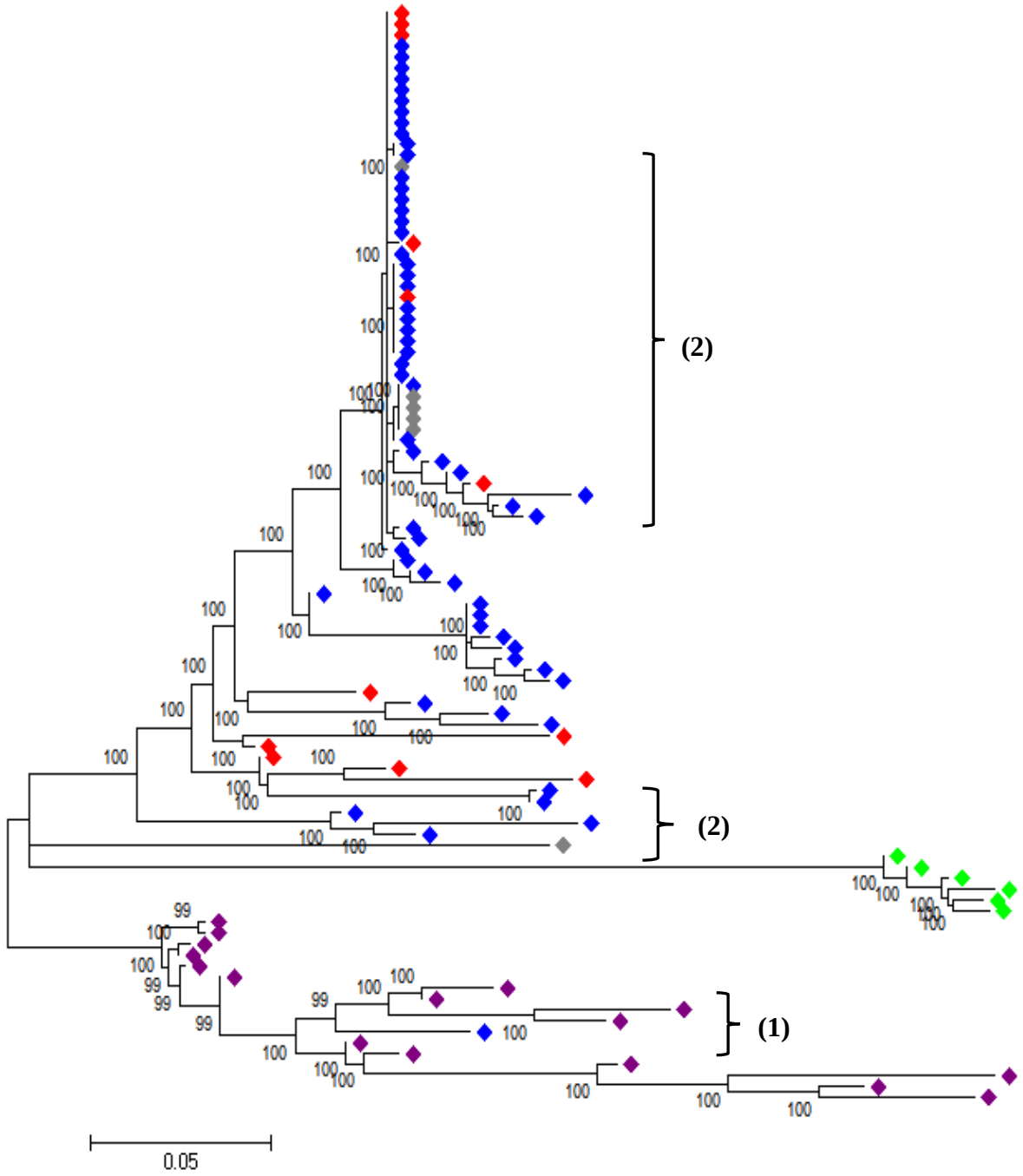


Figure S9.12. Haplotype based maximum likelihood tree for two patterns of introgression based on the 20-kb fragment size that extend from 147060001 to 147080001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression from Ceylon junglefowl to domestic chicken sample (Sri Lanka16). **(2)** Introgression of domestic chicken into grey junglefowl.

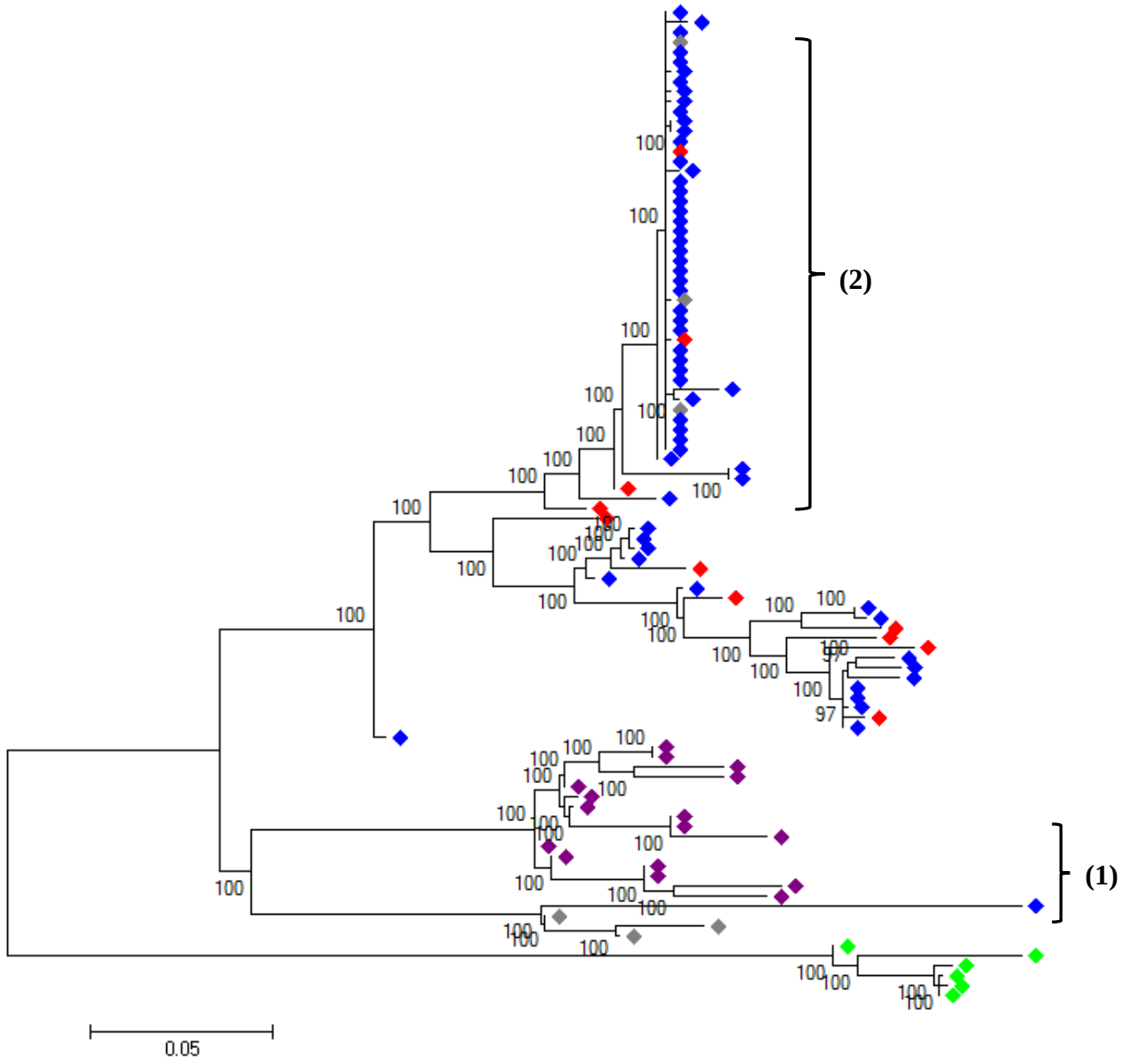


Figure S9.13. Haplotype based maximum likelihood tree for two patterns of introgression. 30-kb fragment size that extends from 147605001 to 147635001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression from Ceylon junglefowl into domestic chicken Saudi Arabia1 sample. **(2)** Introgression of domestic chicken into grey junglefowl.

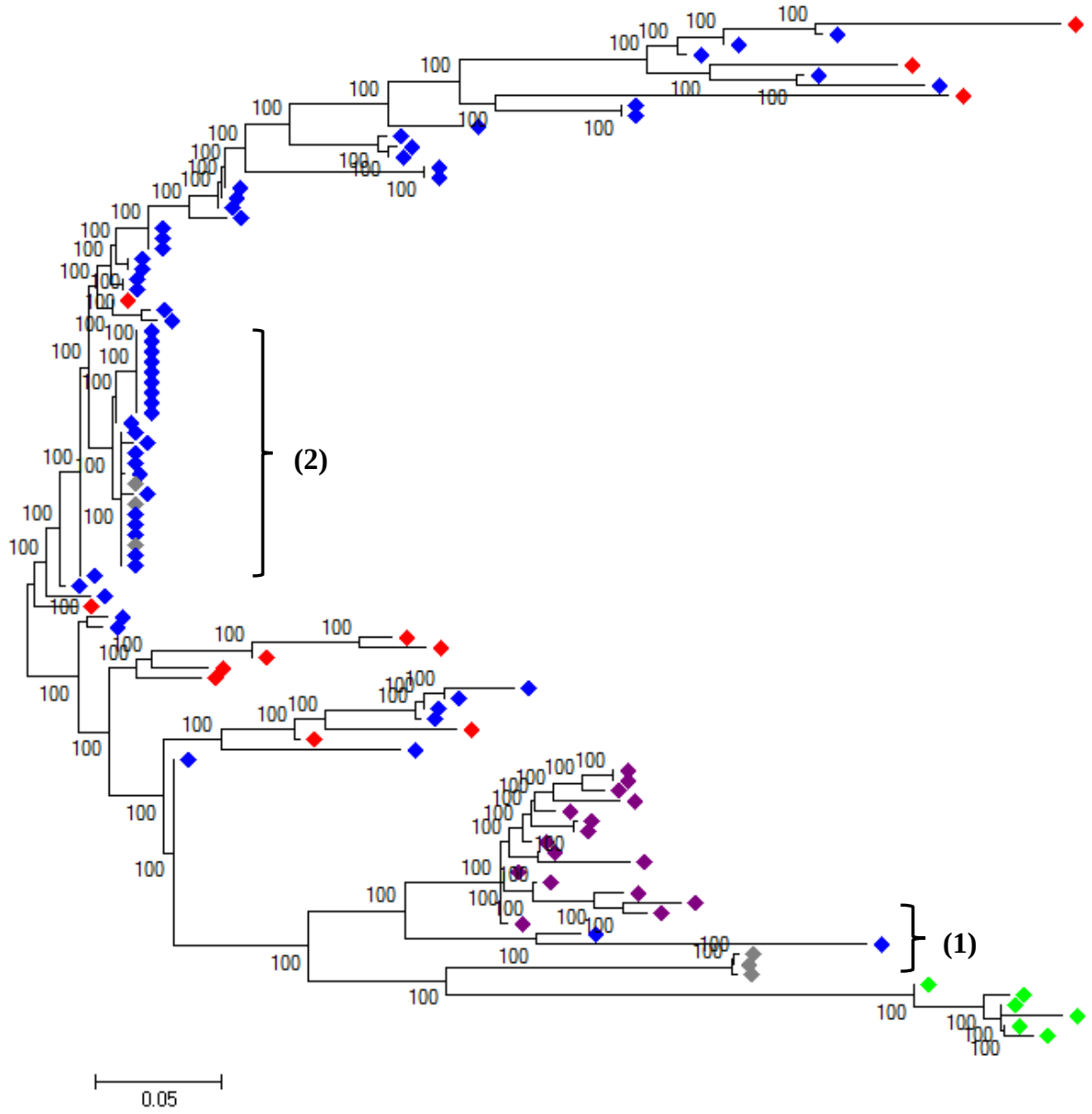


Figure S9.14. Haplotype based maximum likelihood tree for two patterns of introgression. 20-kb fragment size that extends from 23965001 to 23985001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression from Ceylon junglefowl into one haplotype of Sri Lanka4 and HA2A10B samples of domestic chicken. **(2)** Introgression of domestic chicken into grey junglefowl.

Figure S10. Haplotypes relationships analysis for candidate regions of introgression from grey junglefowl into domestic chicken. ABBA – BABA test Ceylon junglefowl - domestic chicken.

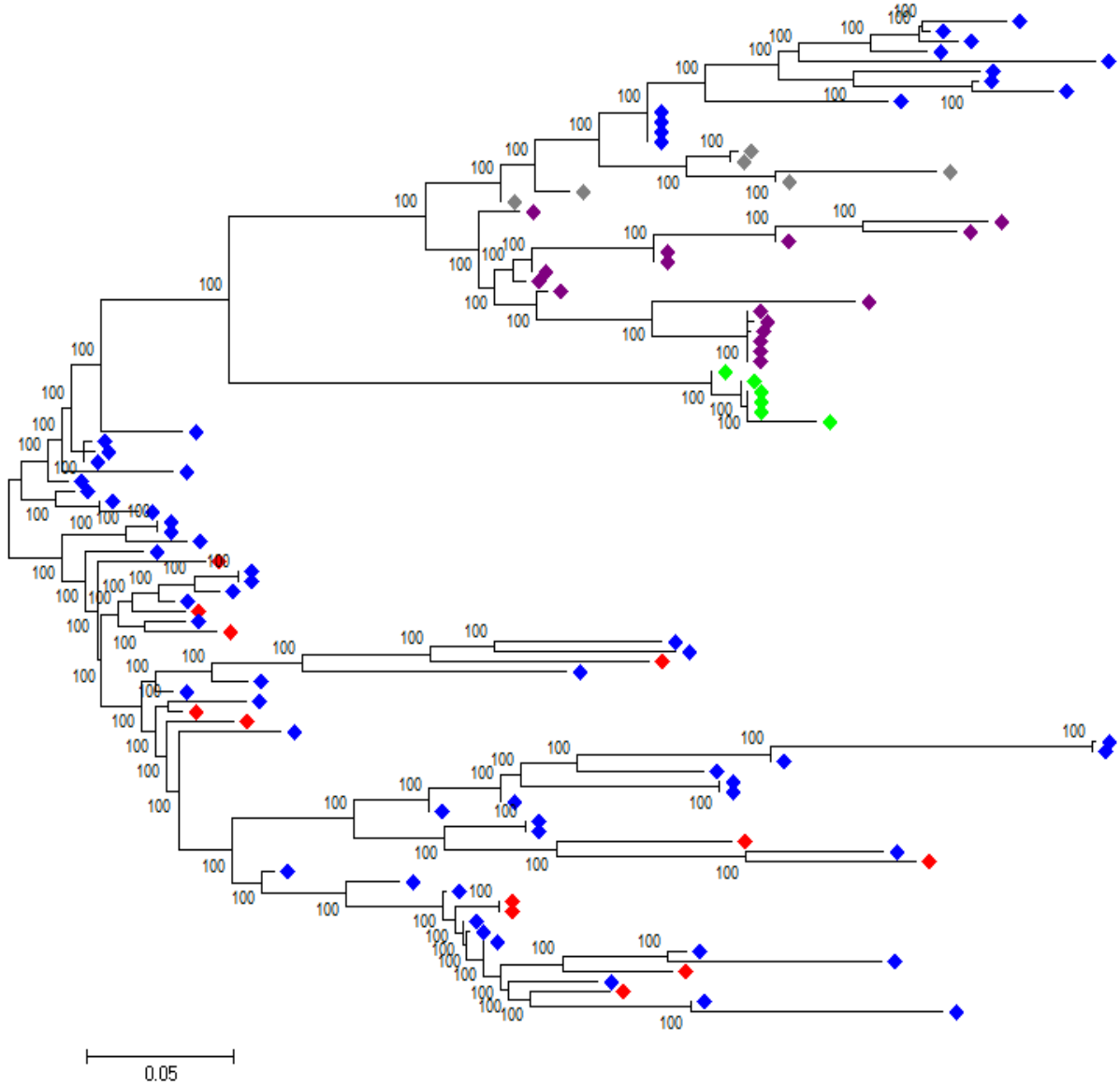


Figure S10.1. Haplotype based maximum likelihood tree for a region of introgression from grey junglefowl into domestic chicken on chromosome 1, 25-kb region, position 13070001 - 13095001 bp. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl.

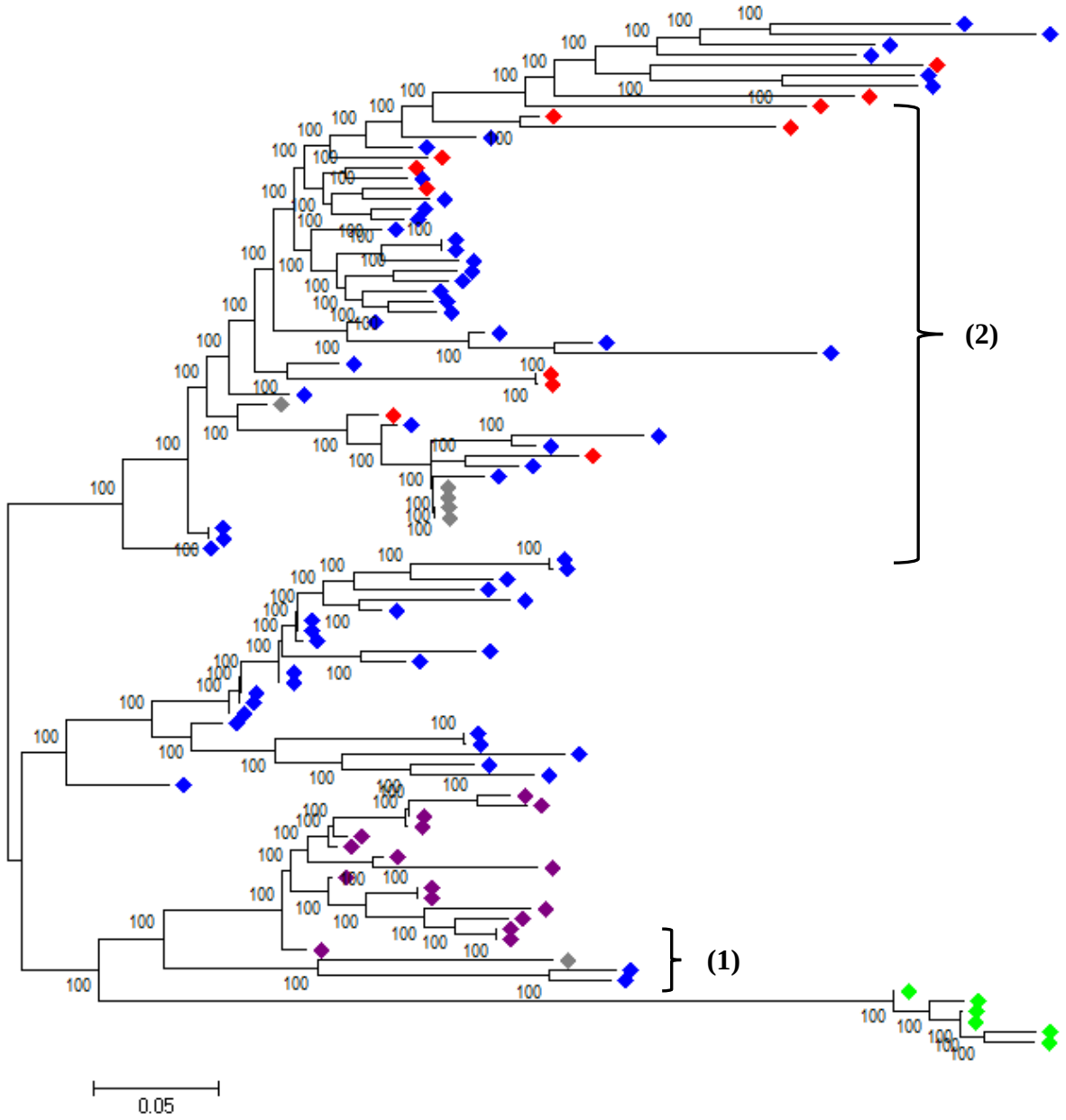


Figure S10.2. Haplotype based maximum likelihood tree for two patterns of introgression based on the 55-kb fragment size that extends from 14950001 to 15005001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Domestic haplotypes of grey junglefowl introgressed origin (Sri Lanka2 and Sri Lanka16 samples). (**2**) Grey junglefowl haplotypes of domestic - red junglefowl origin.

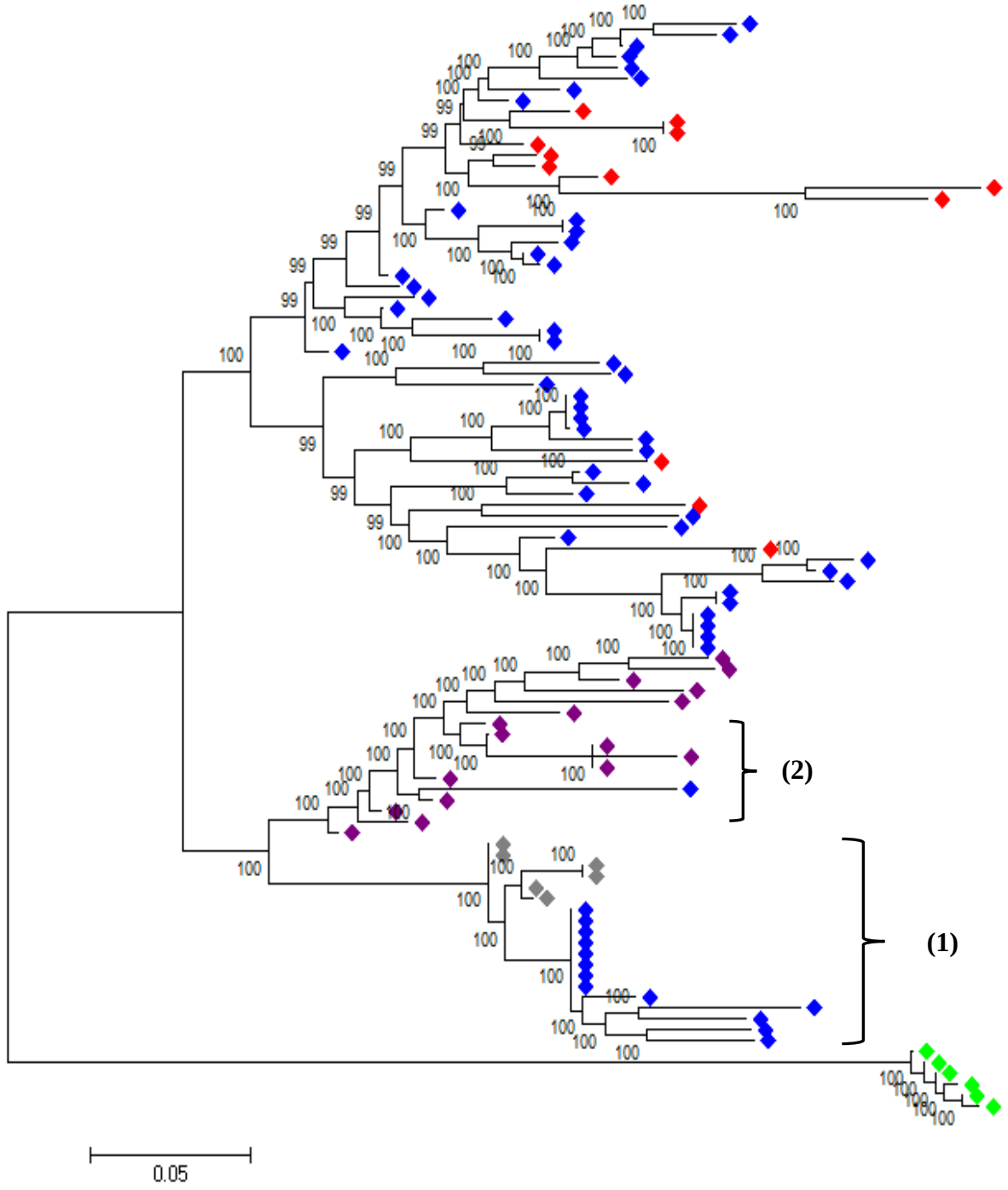


Figure S10.3. Haplotype based maximum likelihood tree for a region of introgression from grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 35-kb fragment size that extends from 25800001 to 25835001 bp on chromosome 1. **(1)** Introgression of grey into domestic. (**2**). One haplotype of a Sri Lanka domestic chicken likely originates from the introgression of Ceylon junglefowl into domestic chicken

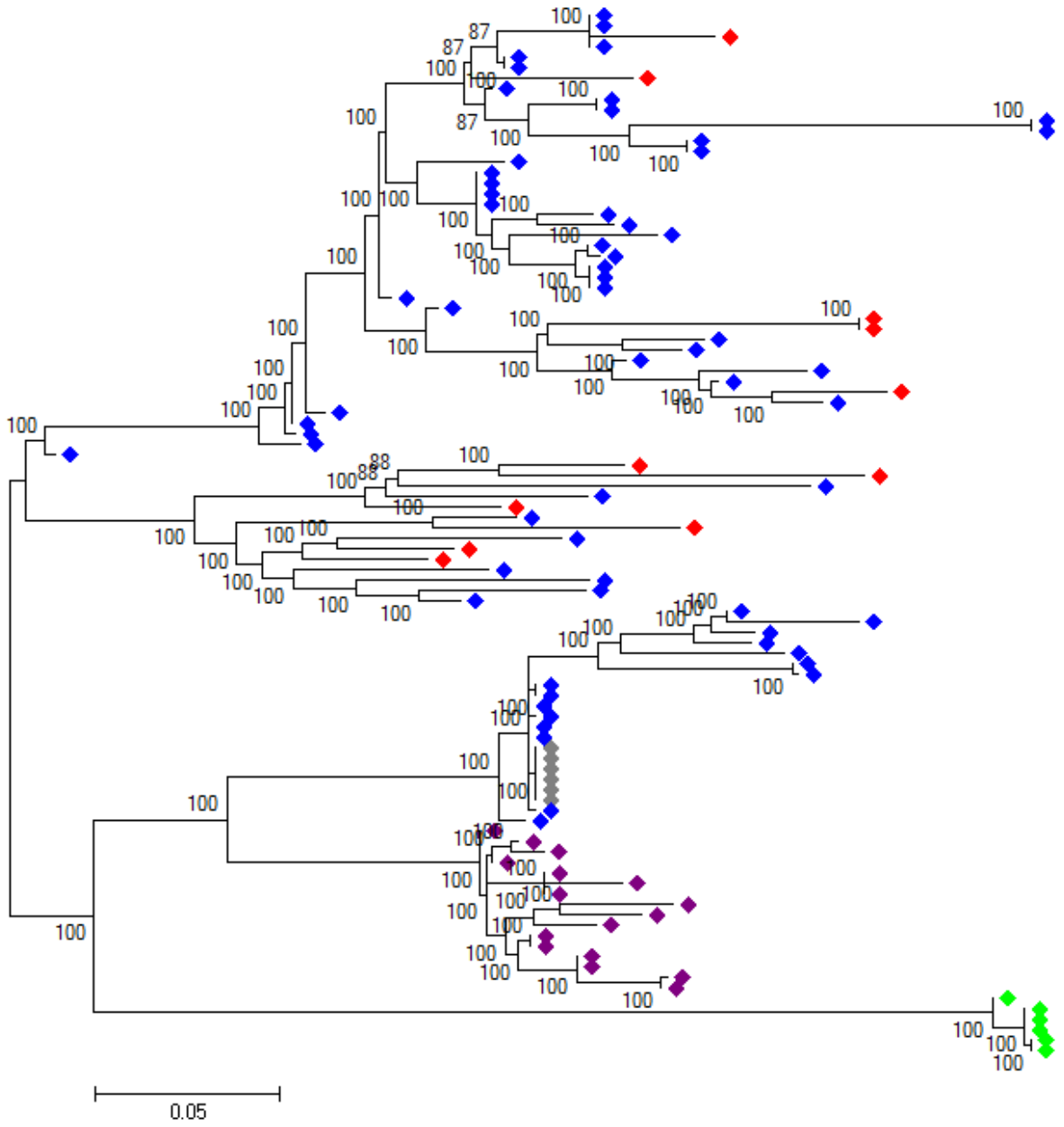


Figure S10.4. Haplotype based maximum likelihood tree for a region of introgression from grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 30-kb fragment size that extends from 52835001 to 52865001 bp on chromosome 1.

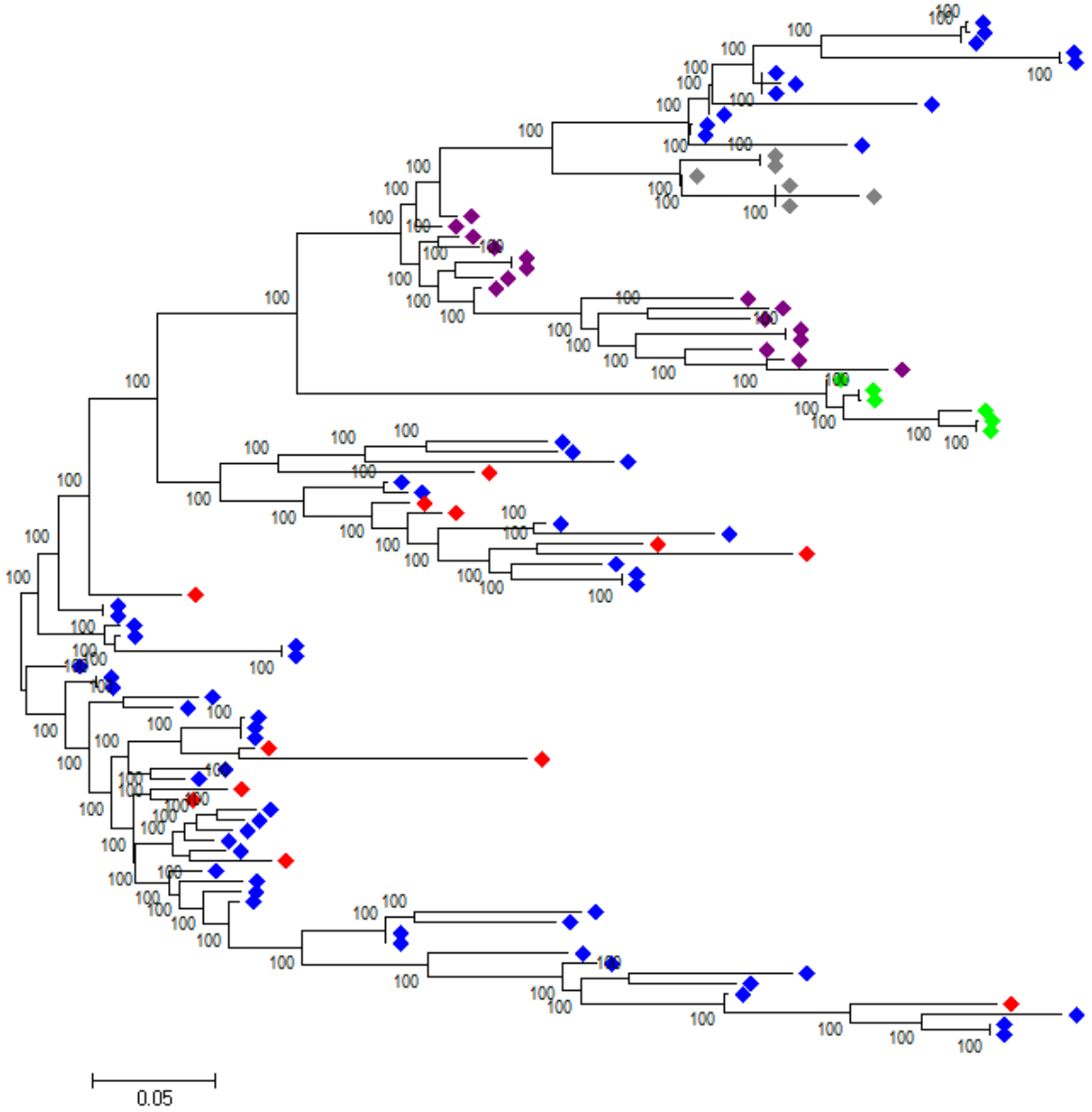


Figure S10.5. Haplotype based maximum likelihood tree for a candidate region of introgression from grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 40-kb fragment size that extends from 90490001 to 90530001 bp on chromosome 1.

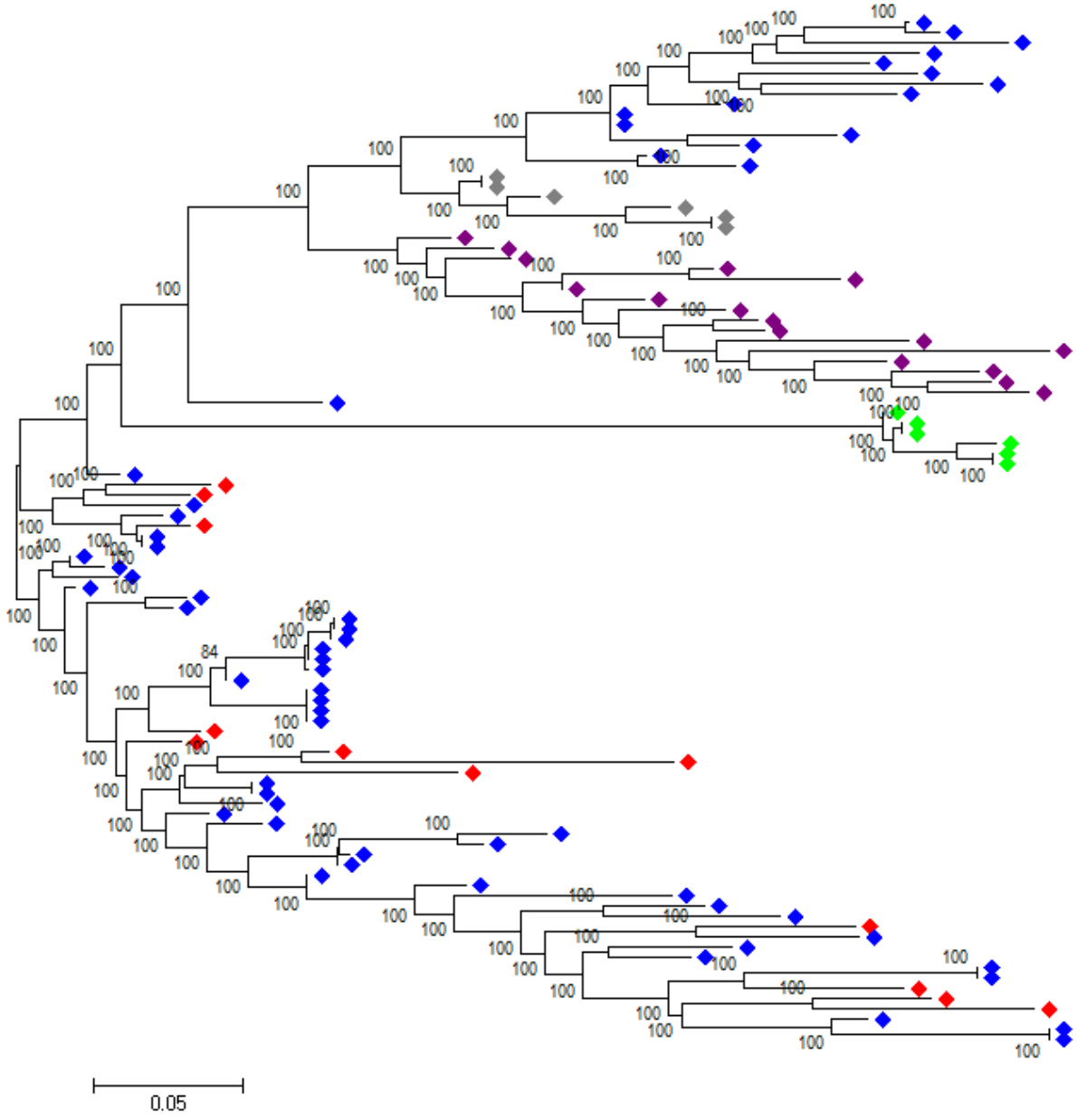


Figure S10.6. Haplotype based maximum likelihood tree for candidate region of introgression from grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 192750001 to 192770001 bp on chromosome 1.

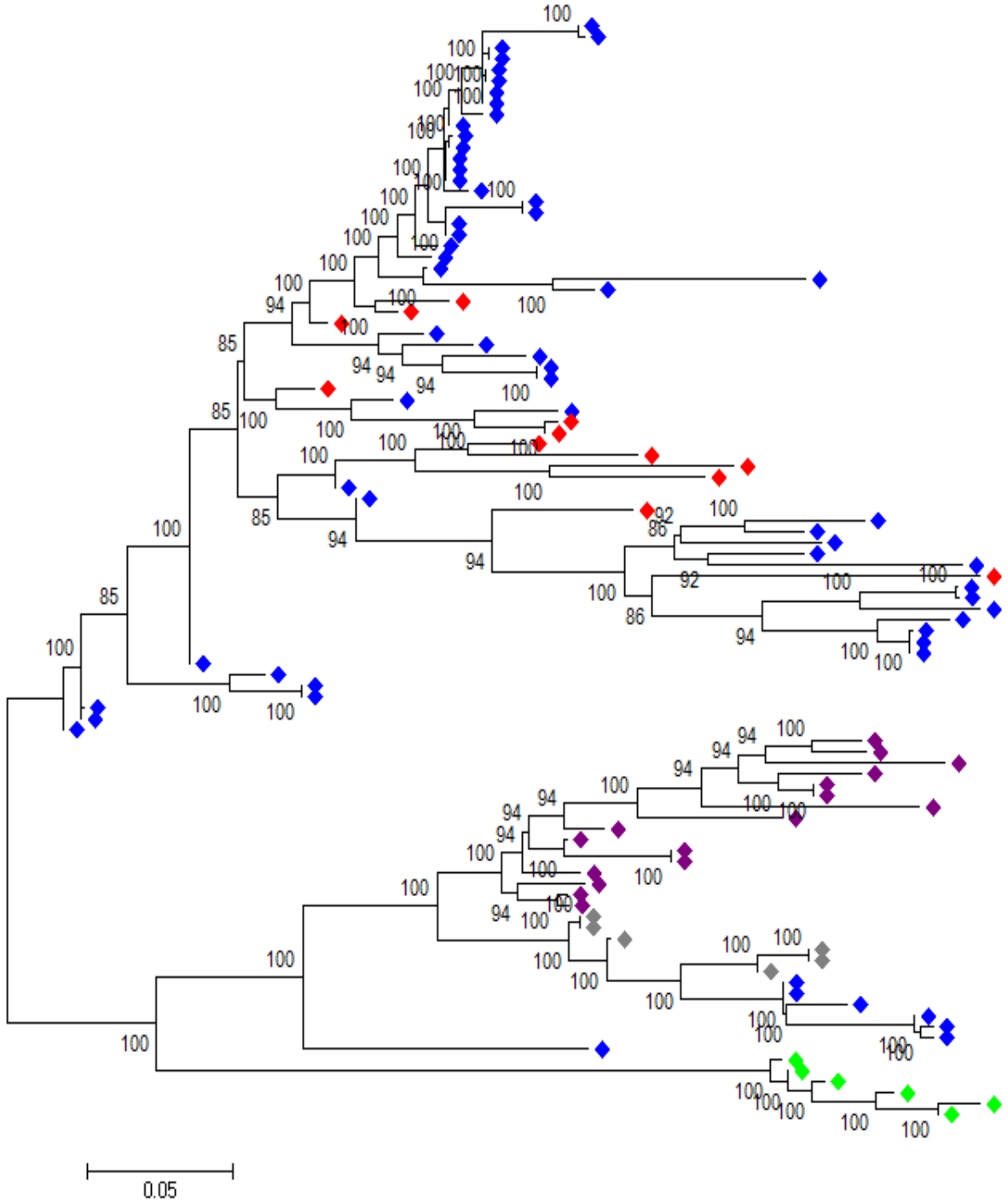


Figure S10.7. Haplotype based maximum likelihood tree for a region of introgression from grey junglefowl into domestic chicken based. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This is a 20-kb fragment size that extends from 79185001 to 79205001 bp on chromosome 2.

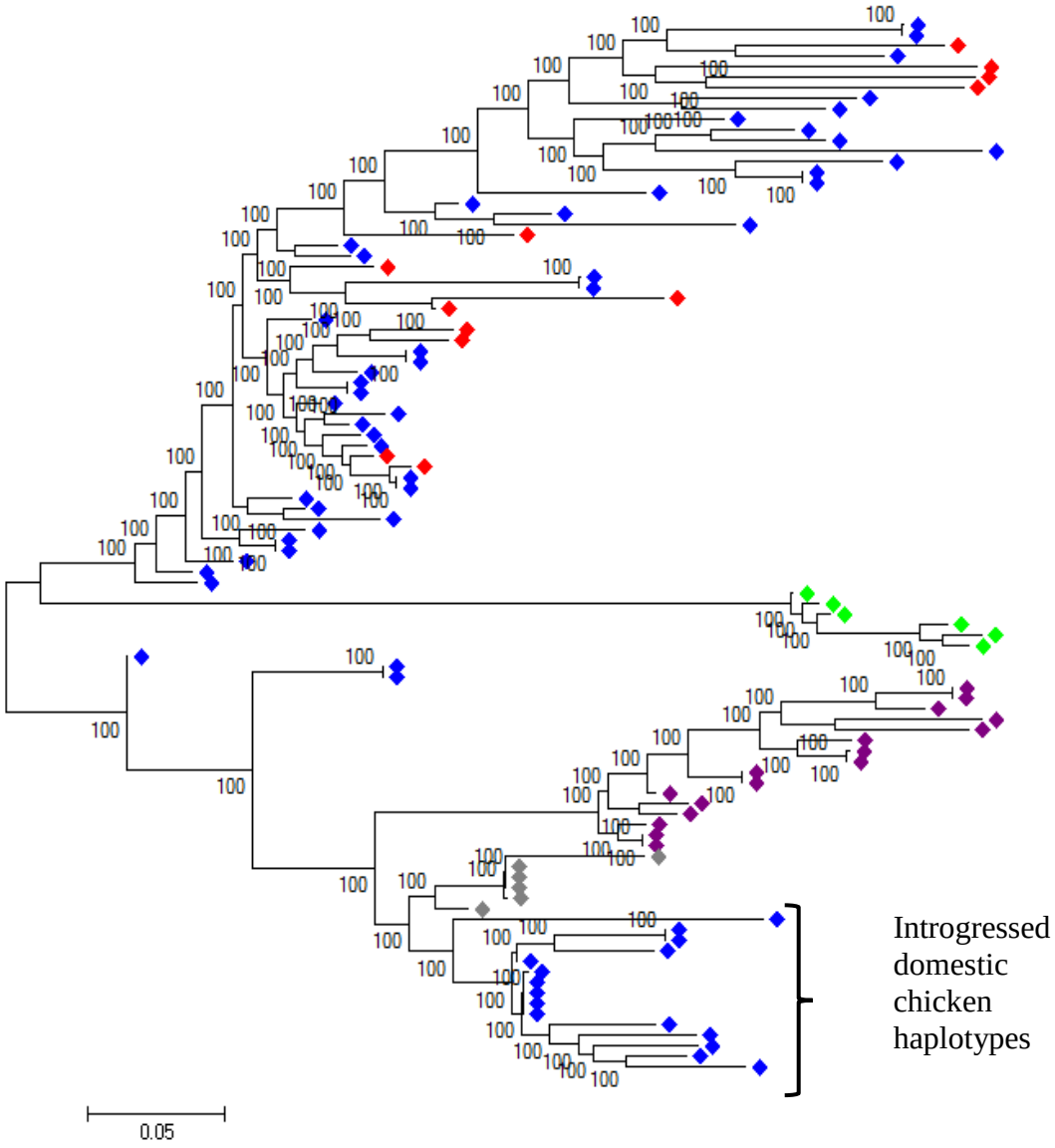


Figure S10.8. Haplotype based maximum likelihood tree illustrating introgression of grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The region is on chromosome 2 (115180001 115295001 bp) and it is 115-kb in length. The six grey junglefowl haplotypes are found at the root of 15 domestic chicken haplotypes. This is the longest region of introgression of grey into domestic chicken identified in this study.

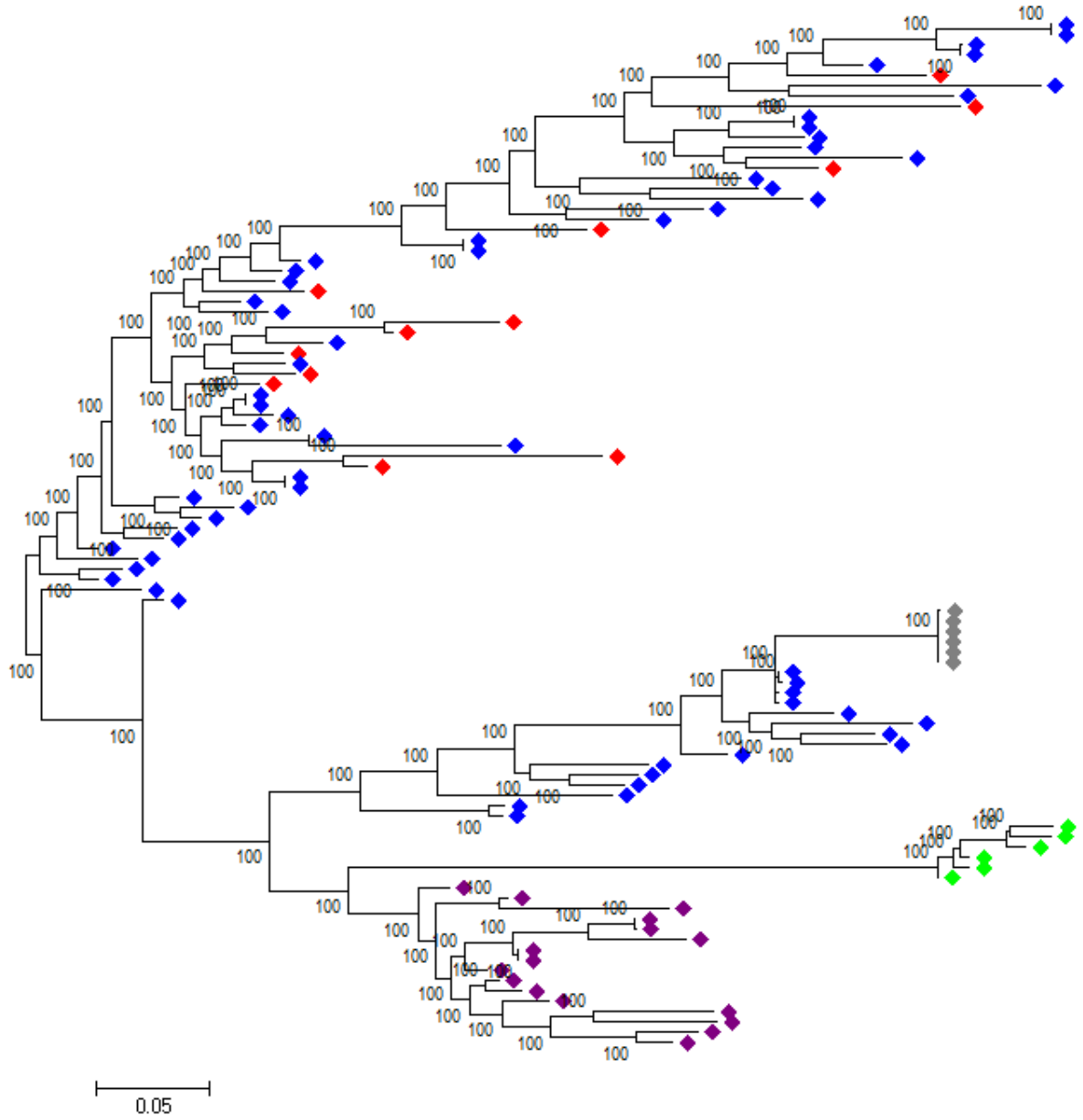


Figure S10.9. Haplotype based maximum likelihood tree for a region of introgression from grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 35-kb fragment size that extends from 106610001 to 106645001 bp on chromosome 3.

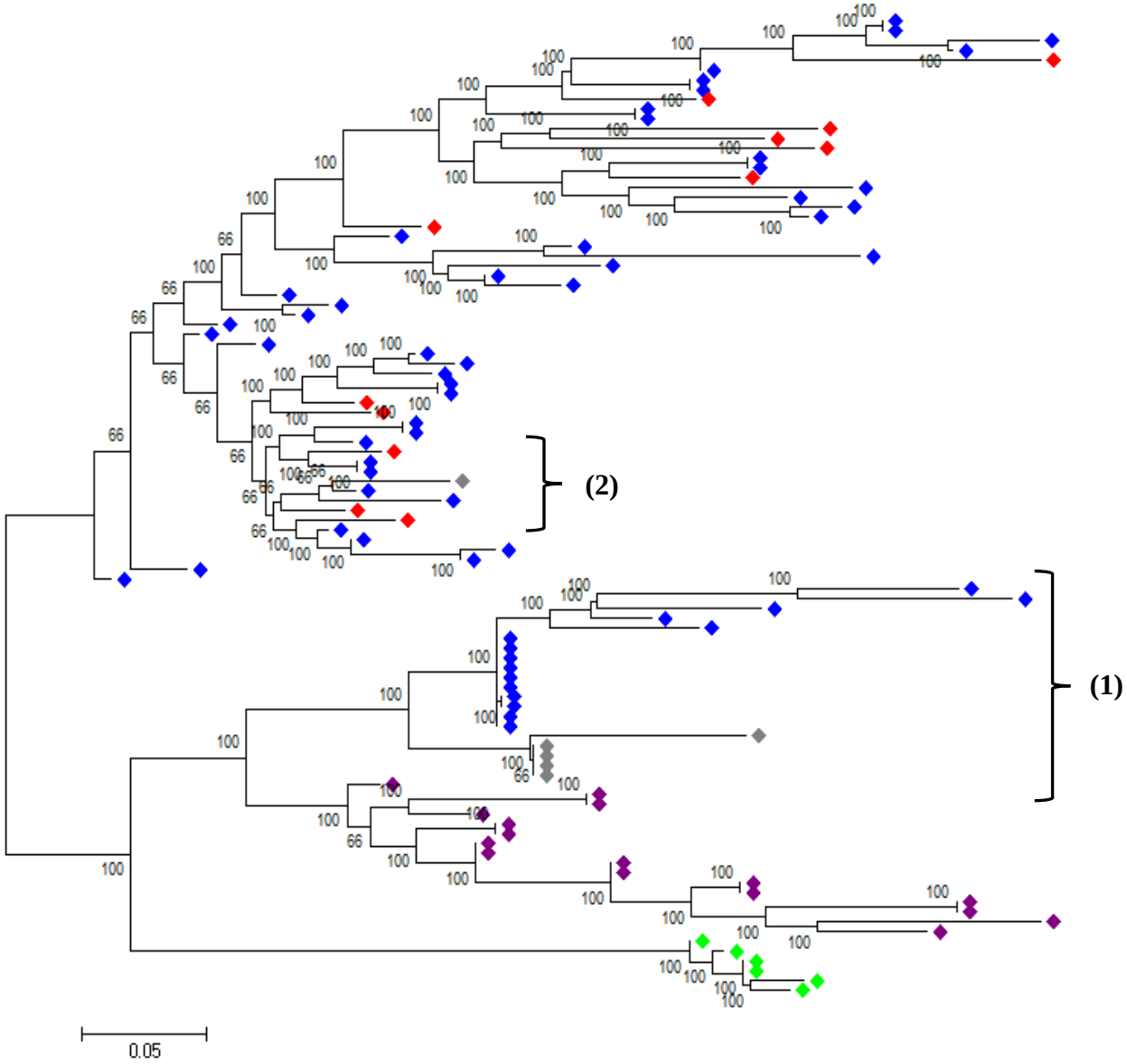


Figure S10.10. Haplotype based maximum likelihood tree for two patterns of introgression on a region of 30-kb that extends from 2195001 to 2225001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of grey junglefowl haplotype to one haplotype from Sri Lanka1. (**2**) Introgression of domestic chicken into grey junglefowl.

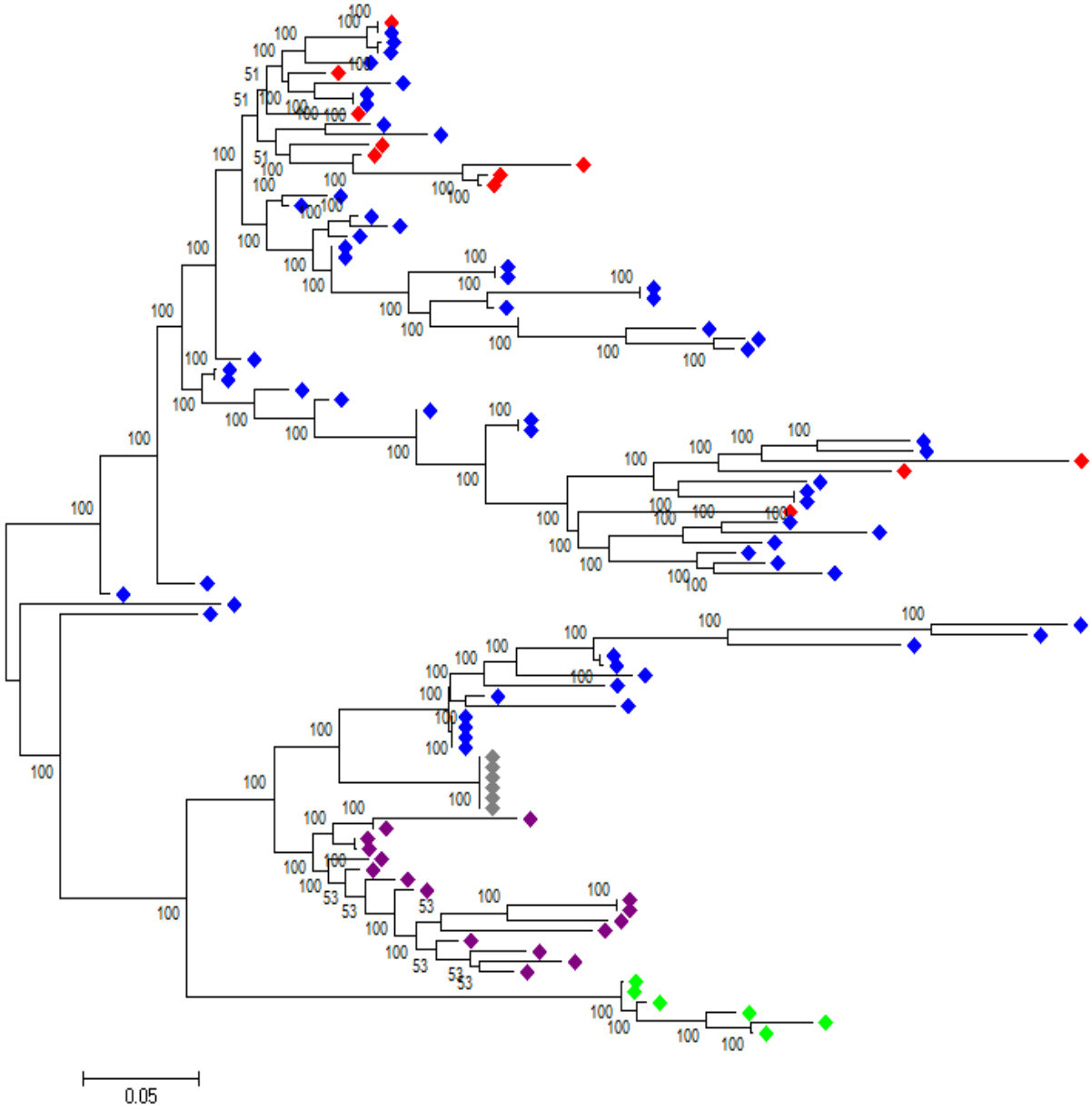


Figure S10.11. Haplotype based maximum likelihood tree for a region of introgression from grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 50-kb fragment size that extends from 12010001 to 12060001 bp on chromosome 4.

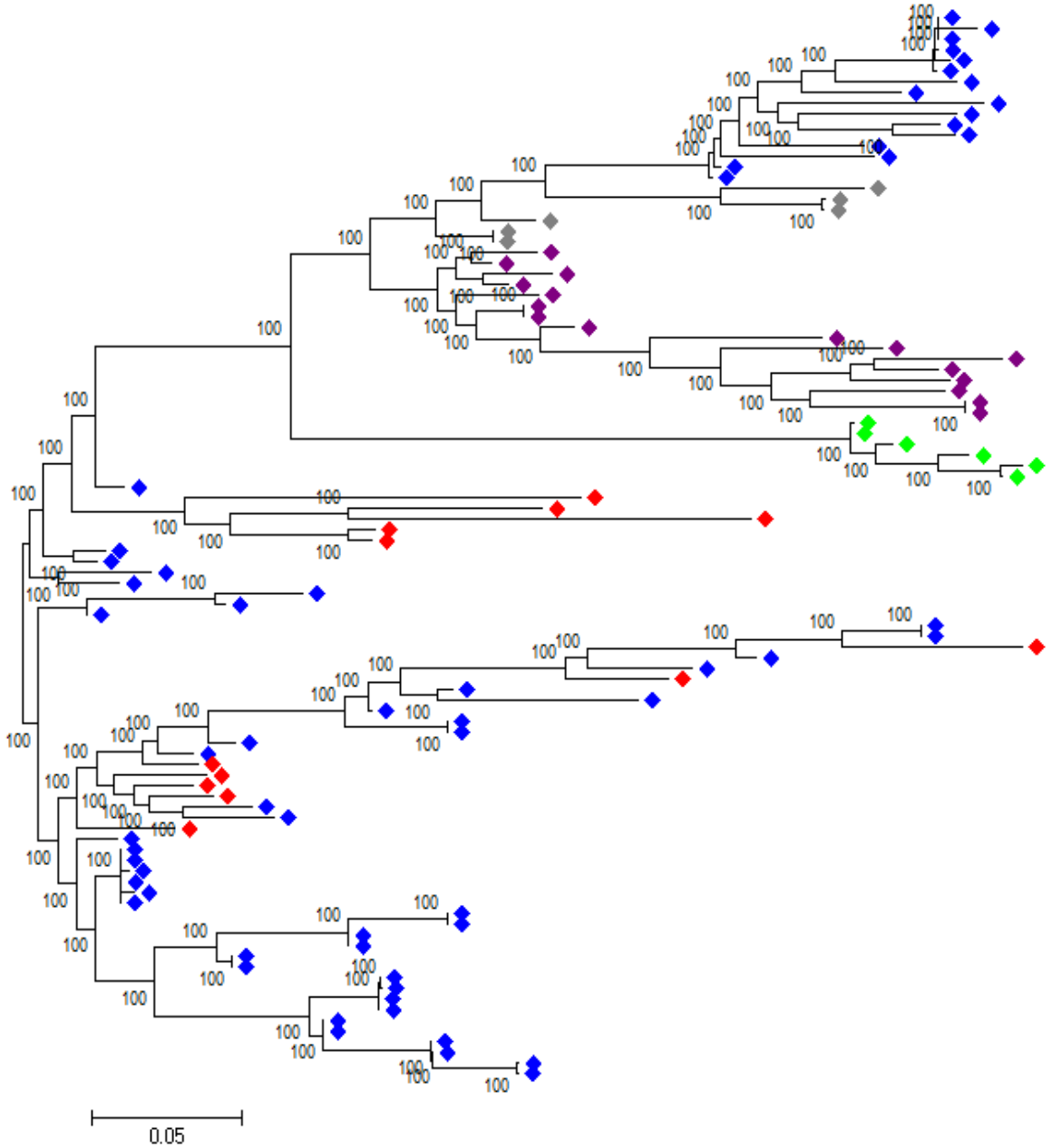


Figure S10.12. Haplotype based maximum likelihood tree for a region of introgression from grey junglefowl into domestic chicken based. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 50-kb fragment size that extends from 13130001 to 13180001 bp on chromosome 4.

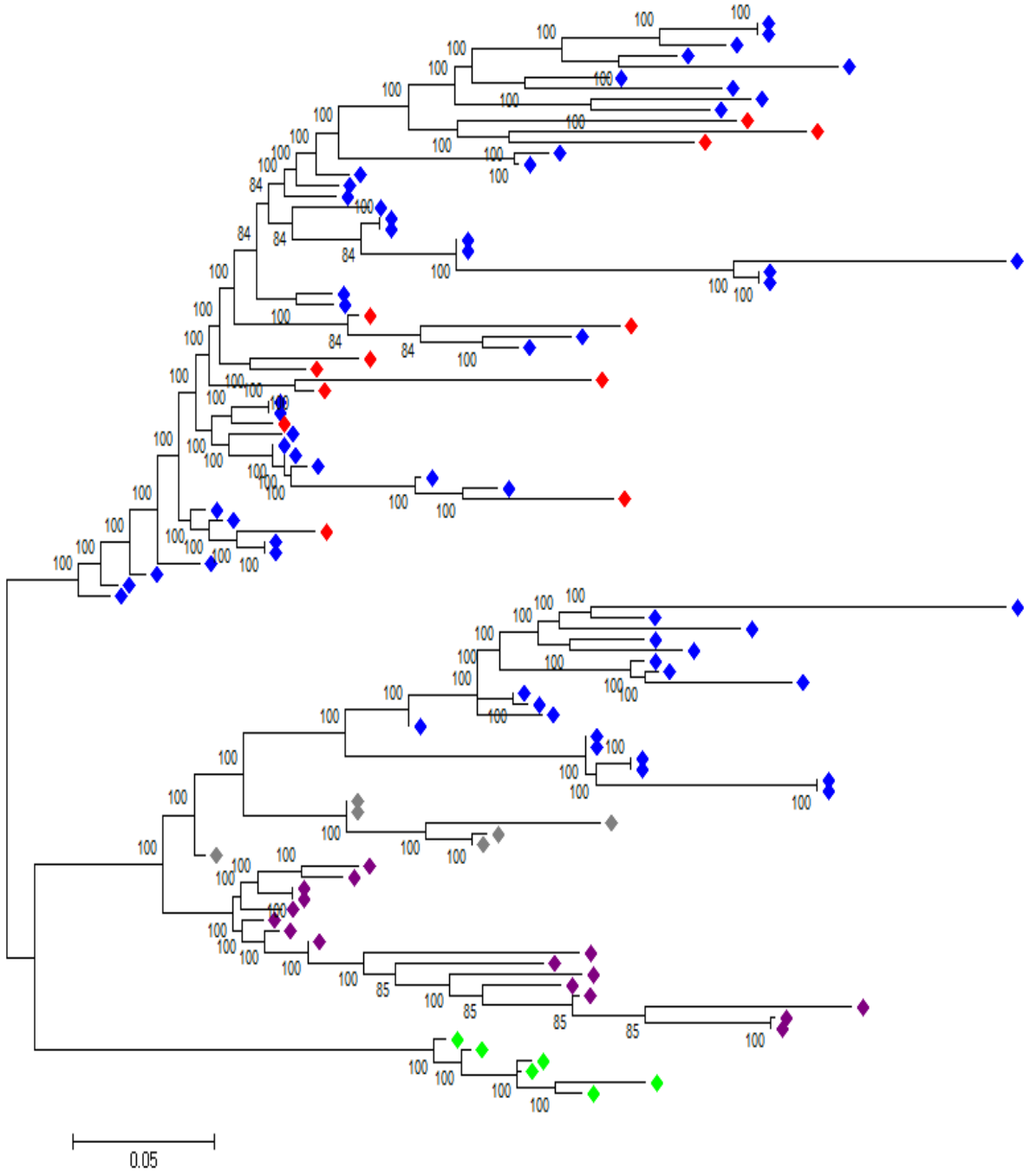


Figure S10.13. Haplotype based maximum likelihood tree for a region of introgression from grey junglefowl into domestic chicken based. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 13265001 to 13285001 bp on chromosome 4.

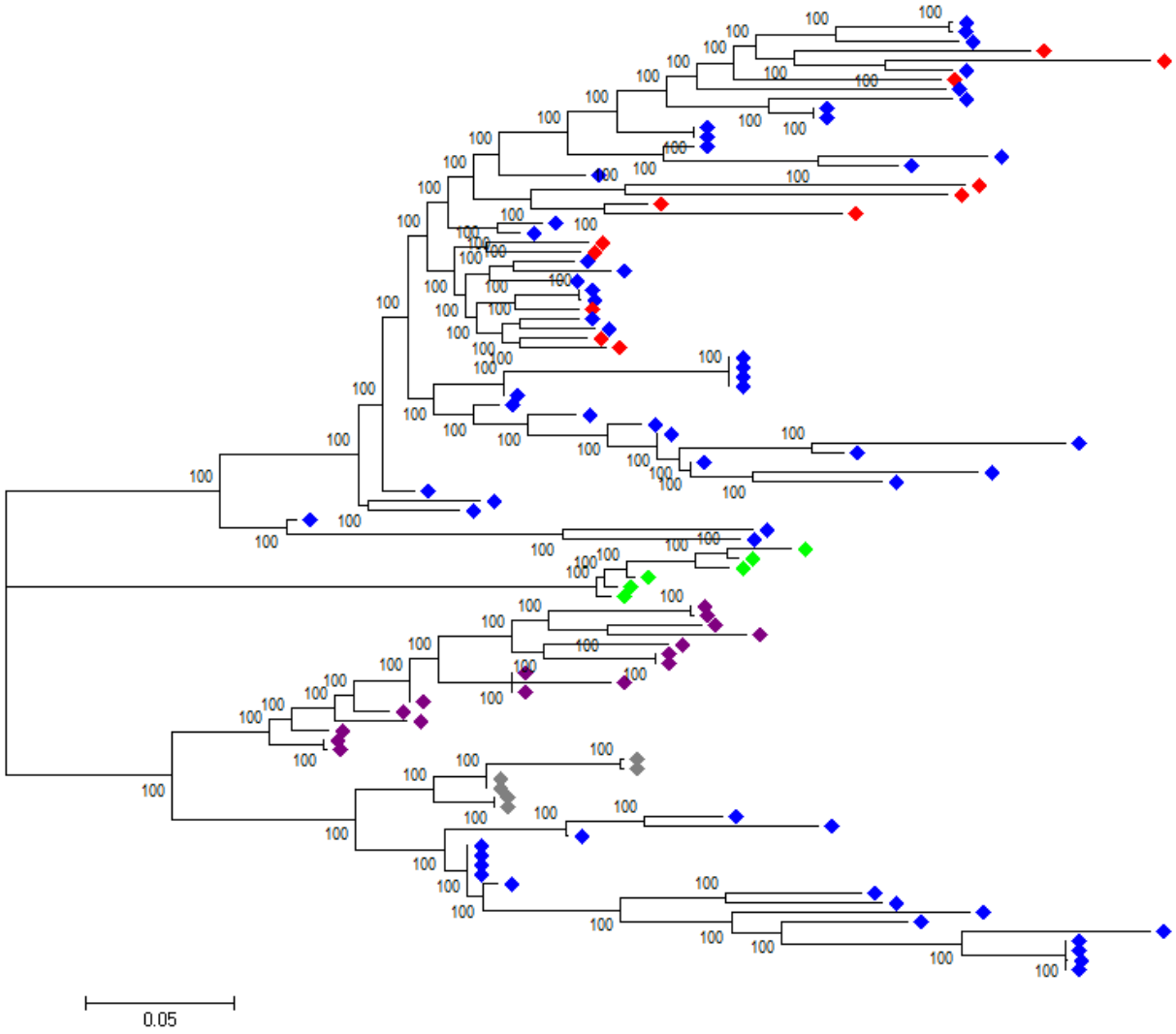


Figure S10.14. Haplotype based maximum likelihood tree of introgression of grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 30-kb fragment size that extends from 22795001 to 22825001 bp on chromosome 7.

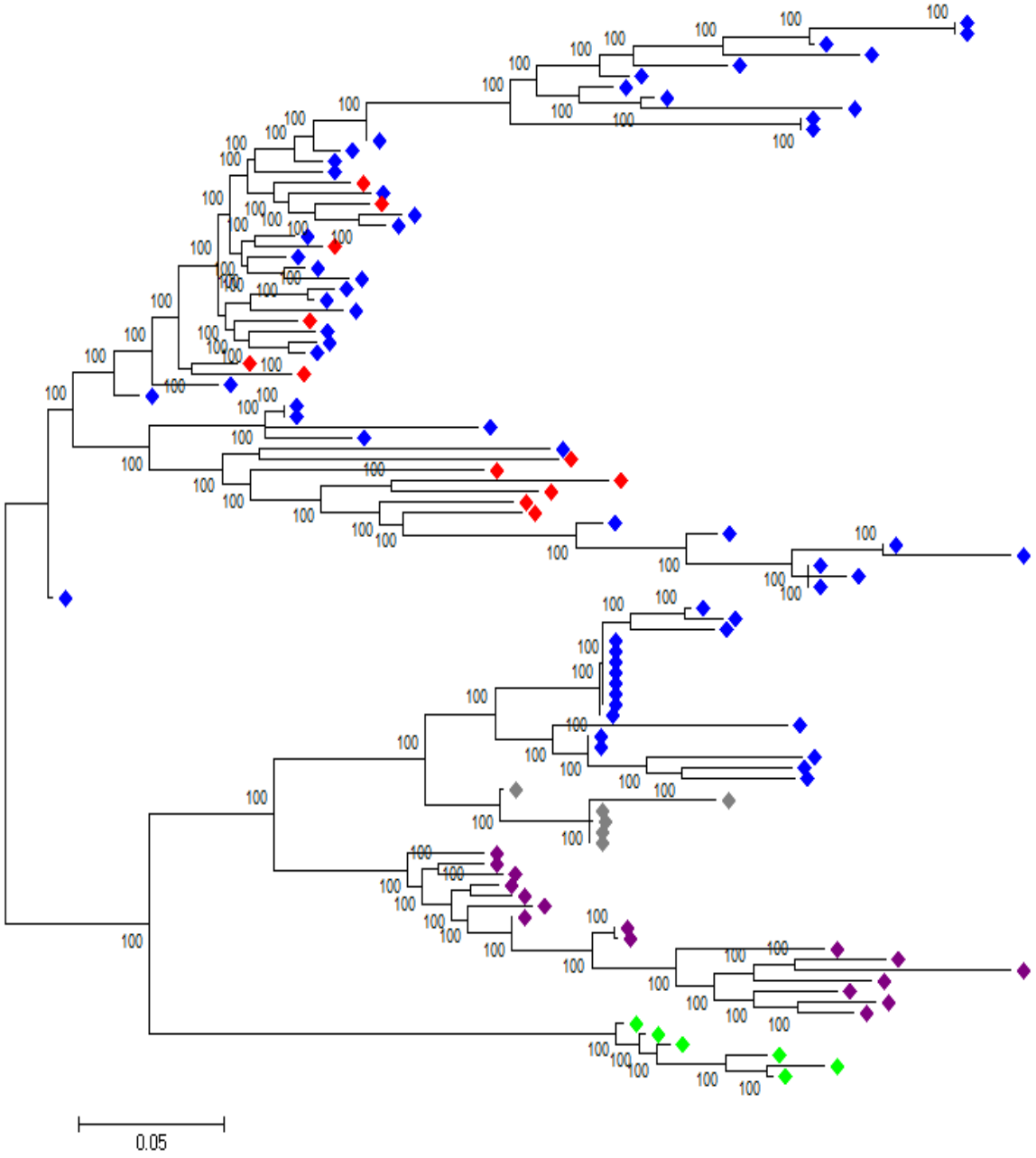


Figure S10.15. Haplotype based maximum likelihood tree of introgression of grey junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 14570001 to 14590001 bp on chromosome 11.

Figures S11. Haplotype relationships analysis for candidate regions of introgression from domestic chicken into Ceylon junglefowl. ABBA – BABA Ceylon junglefowl domestic chicken.

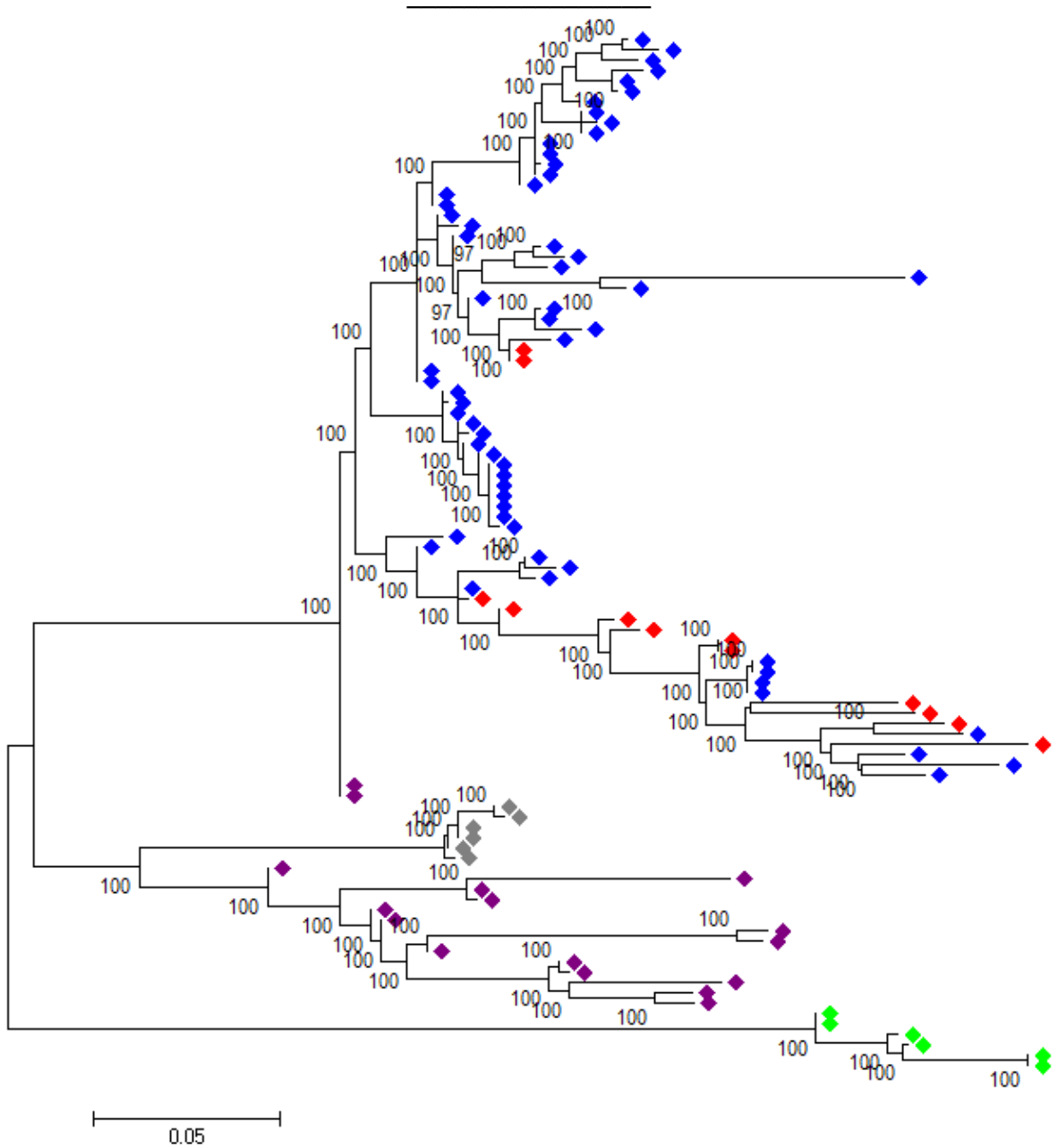


Figure S11.1. Haplotype based maximum likelihood tree for a candidate region (30-kb length) of introgression position 76905001 to 76935001 bp on chromosome 2. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. Haplotypes relationships analysis suggests domestic chicken introgression in Ceylon junglefowl, which however are at the root of the domestic/red junglefowl lineage.

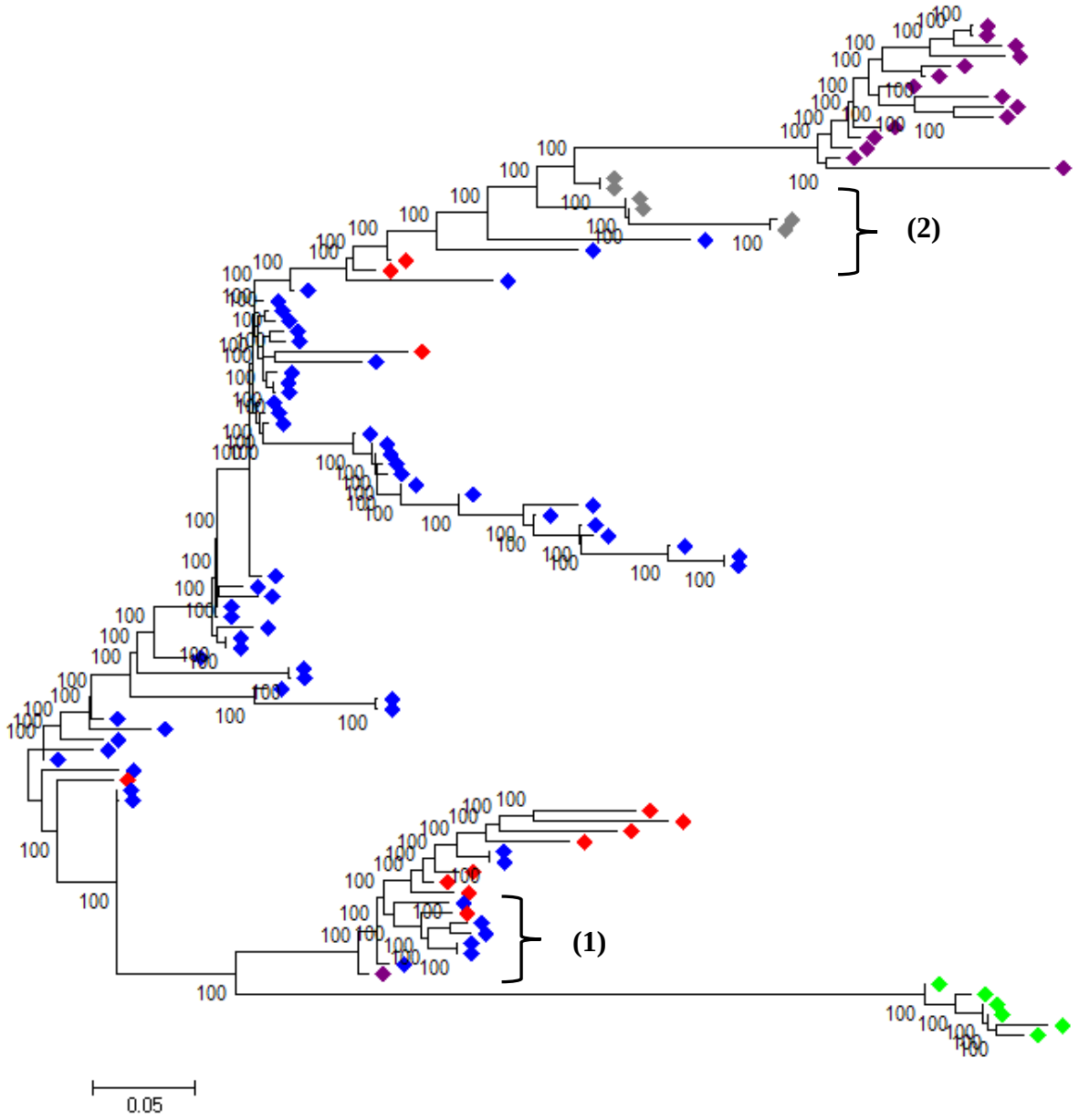


Figure S11.2. Haplotype based maximum likelihood tree constructed for the 70-kb (6120001 - 6190001 bp) region of Ceylon junglefowl – domestic chicken introgression on chromosome 24. The region includes the 30-kb candidate region (6130001 - 6160000 bp) on chromosome 24 significant in the grey junglefowl – domestic chicken introgression analysis (**Figure S5.23**). ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl**. (1)** A Ceylon junglefowl haplotype of likely domestic chicken origin. **(2)** Introgressed domestic chicken haplotypes of grey junglefowl or Ceylon junglefowl into domestic chicken.

Figure S12. The regions reported in this section show peaks that above the genome threshold in the Ceylon domestic – chicken ABBA/BABA analysis, but the trees show inconclusive relationships between haplotypes

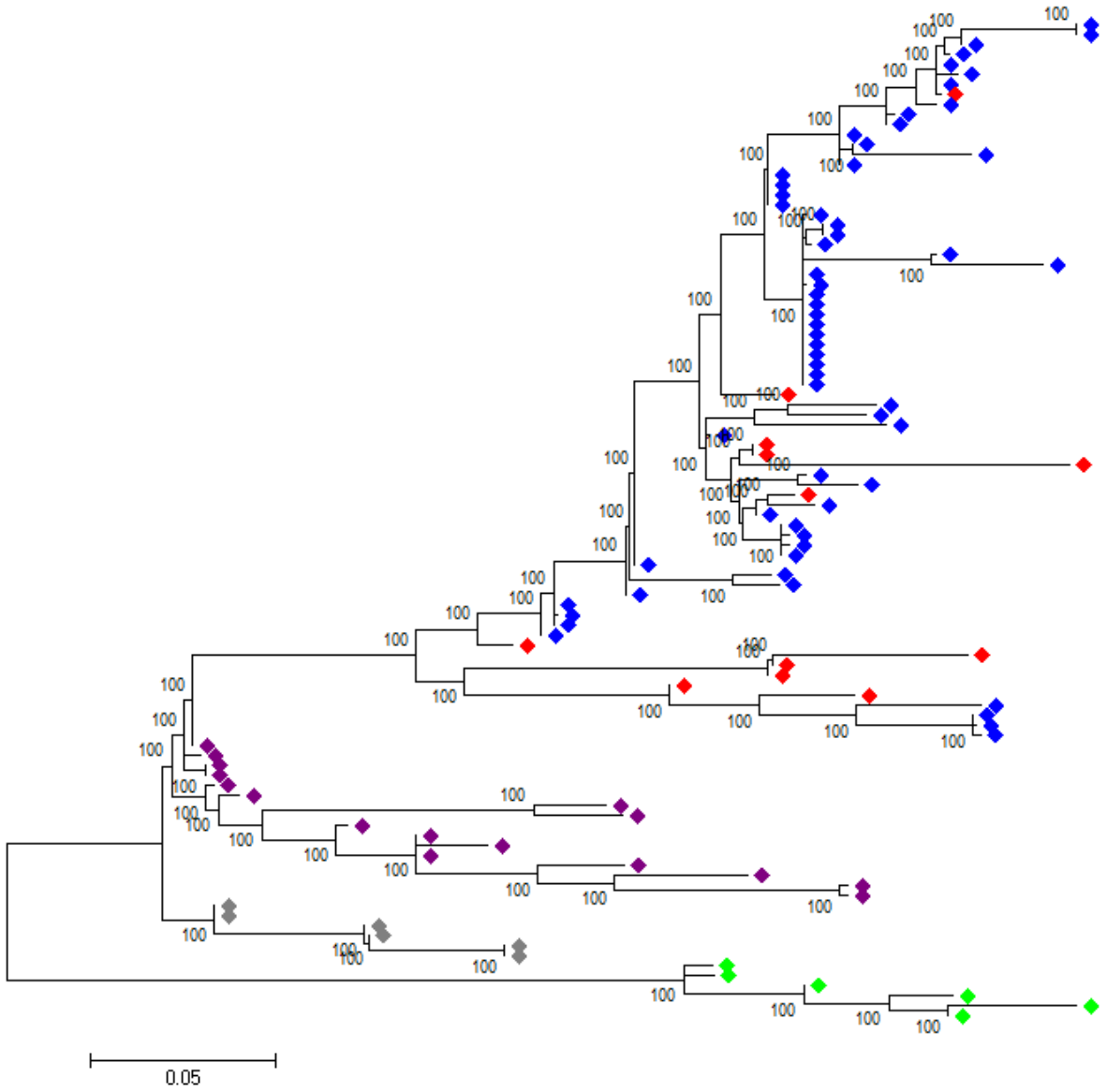


Figure S12.1. A 20-kb candidate region of introgression with fragment size that extends from 23740001 to 23760001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. ABBA-BABA Ceylon junglefowl - domestic chicken test show significant peak but haplotypes relationships analysis could not determine the donor species.

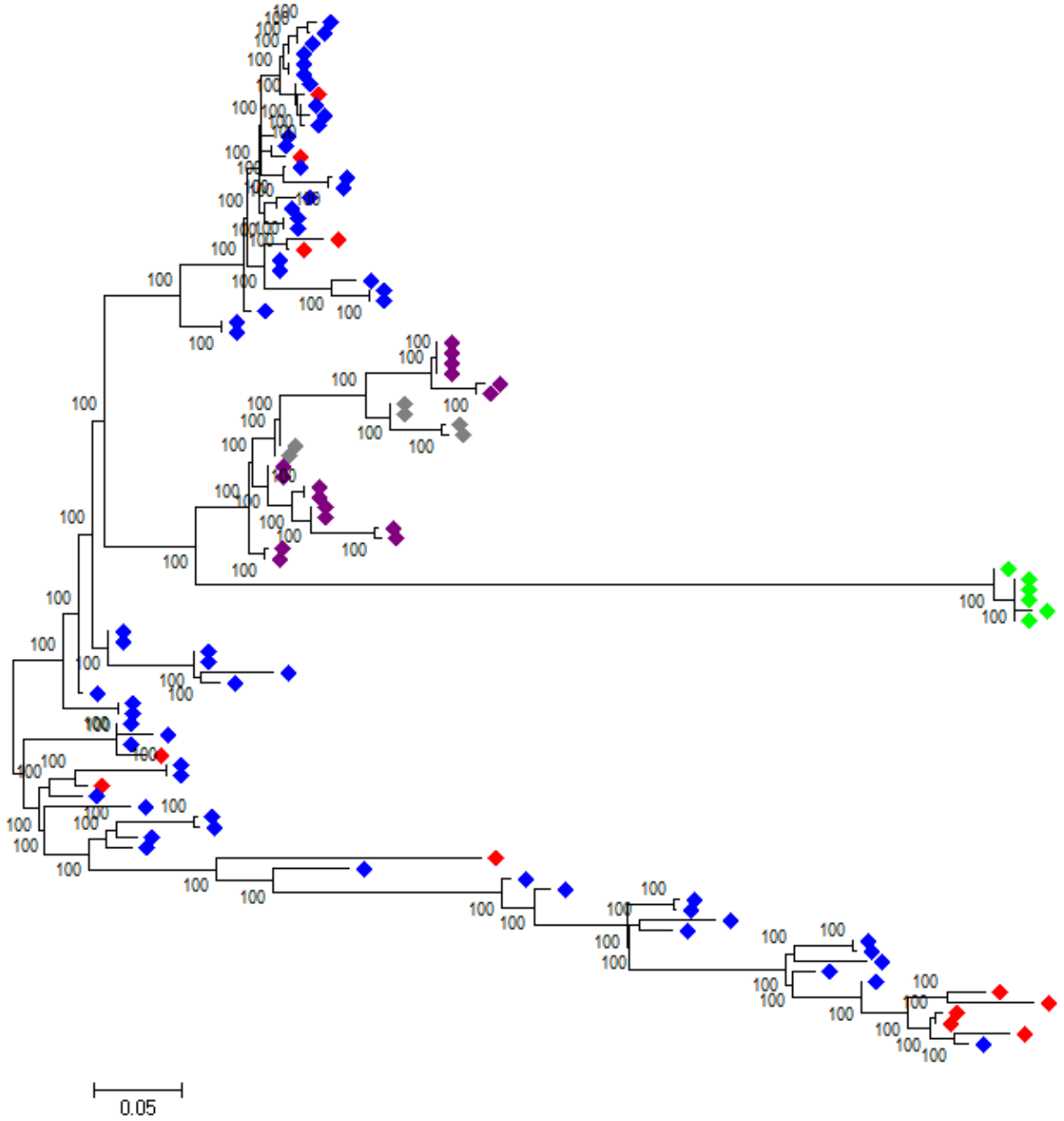


Figure S12.2. A 20-kb candidate region of introgression with fragment size that extends from 77730001 to 77750001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. ABBA-BABA Ceylon junglefowl - domestic chicken test show significant peak but haplotypes relationships analysis suggests shared ancestry

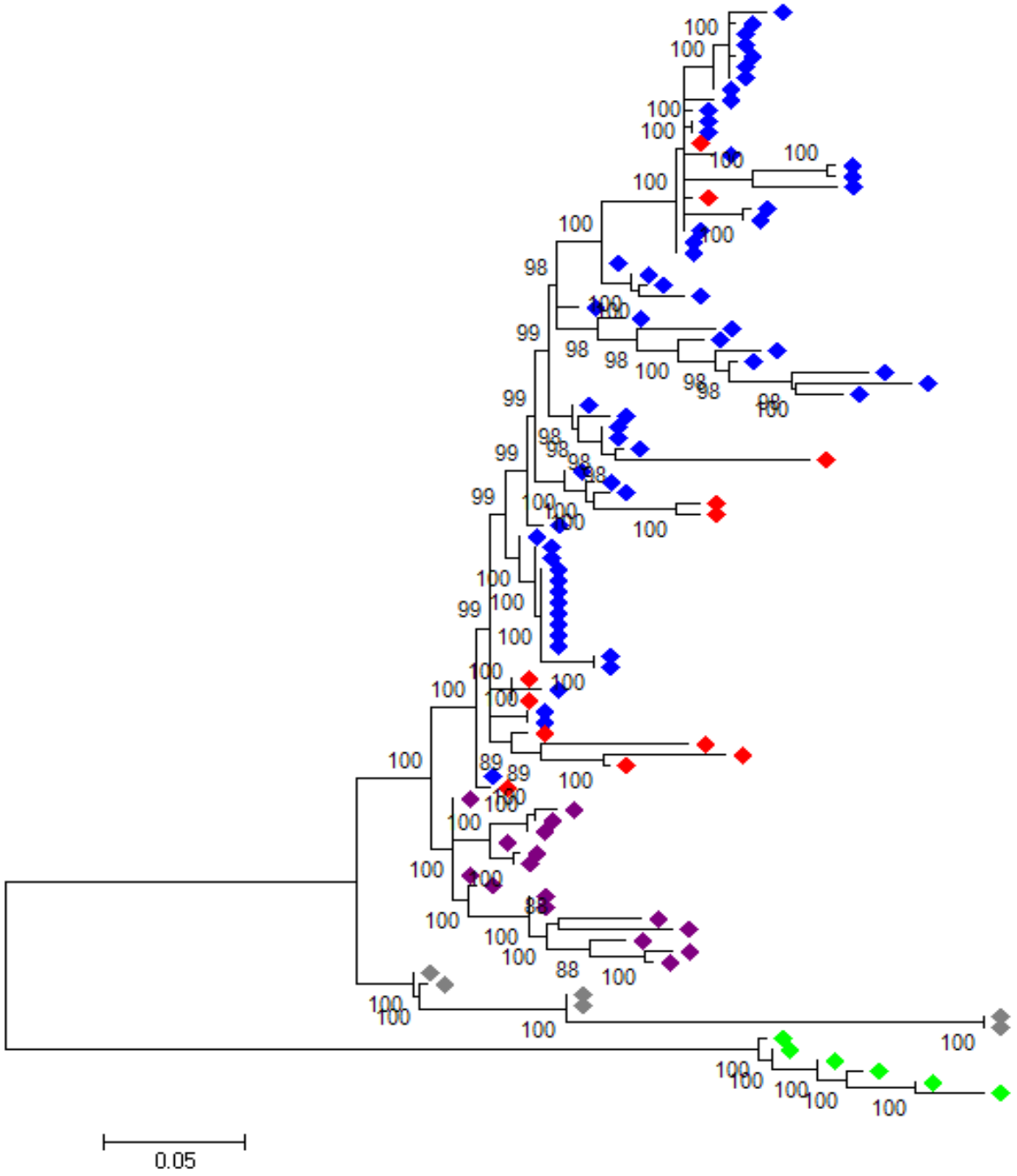


Figure S12.3. A 25-kb candidate region of introgression with fragment size that extends from 81835001 to 81860001 bp on chromosome 2. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. ABBA-BABA Ceylon junglefowl - domestic chicken test show significant peak but haplotypes relationships analysis suggests that this introgression is present in all domestic chicken and wild red junglefowl samples; an unlikely event considering the current geographic distribution of the wild Ceylon and red junglefowl.

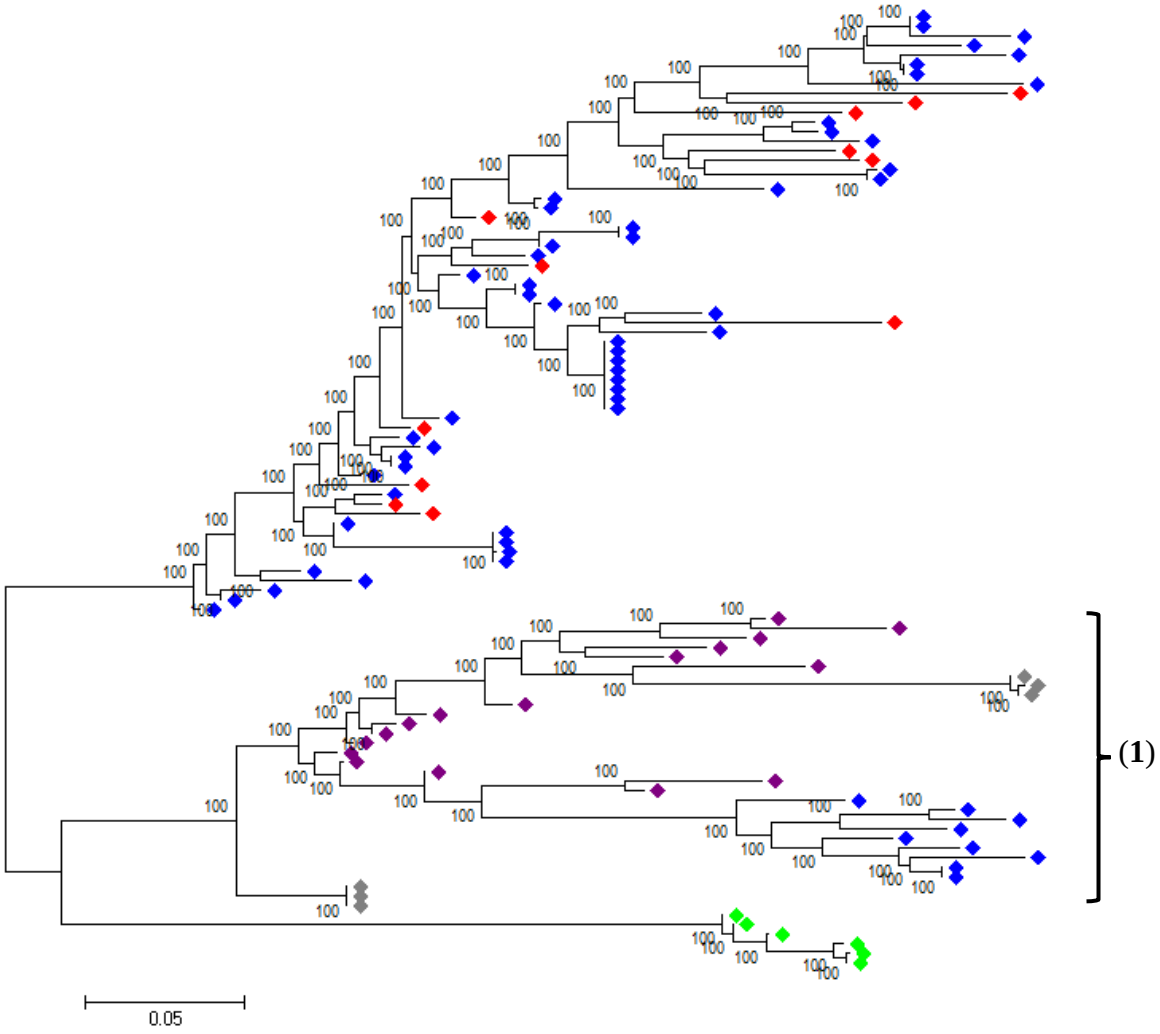


Figure S12.4. A 20-kb candidate region of introgression with fragment size that extends from 185510001 to 185530001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The haplotype analysis could not resolve the donor species between the Ceylon and grey junglefowls

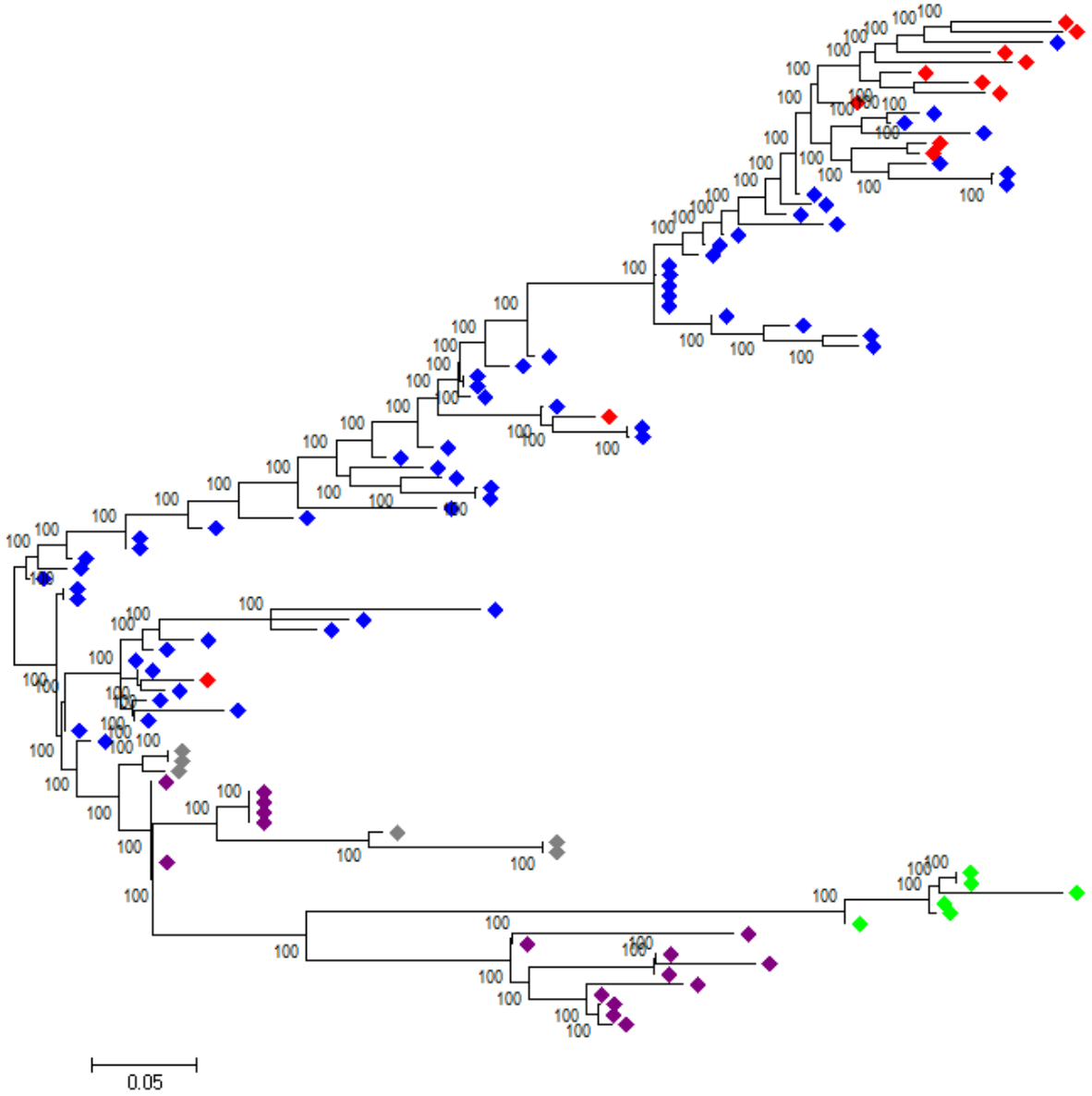


Figure S12.5. A 95-kb candidate region of introgression with fragment size that extends from 100455001 to 100550001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The haplotype analysis could not resolve the donor species between the Ceylon and grey junglefowls

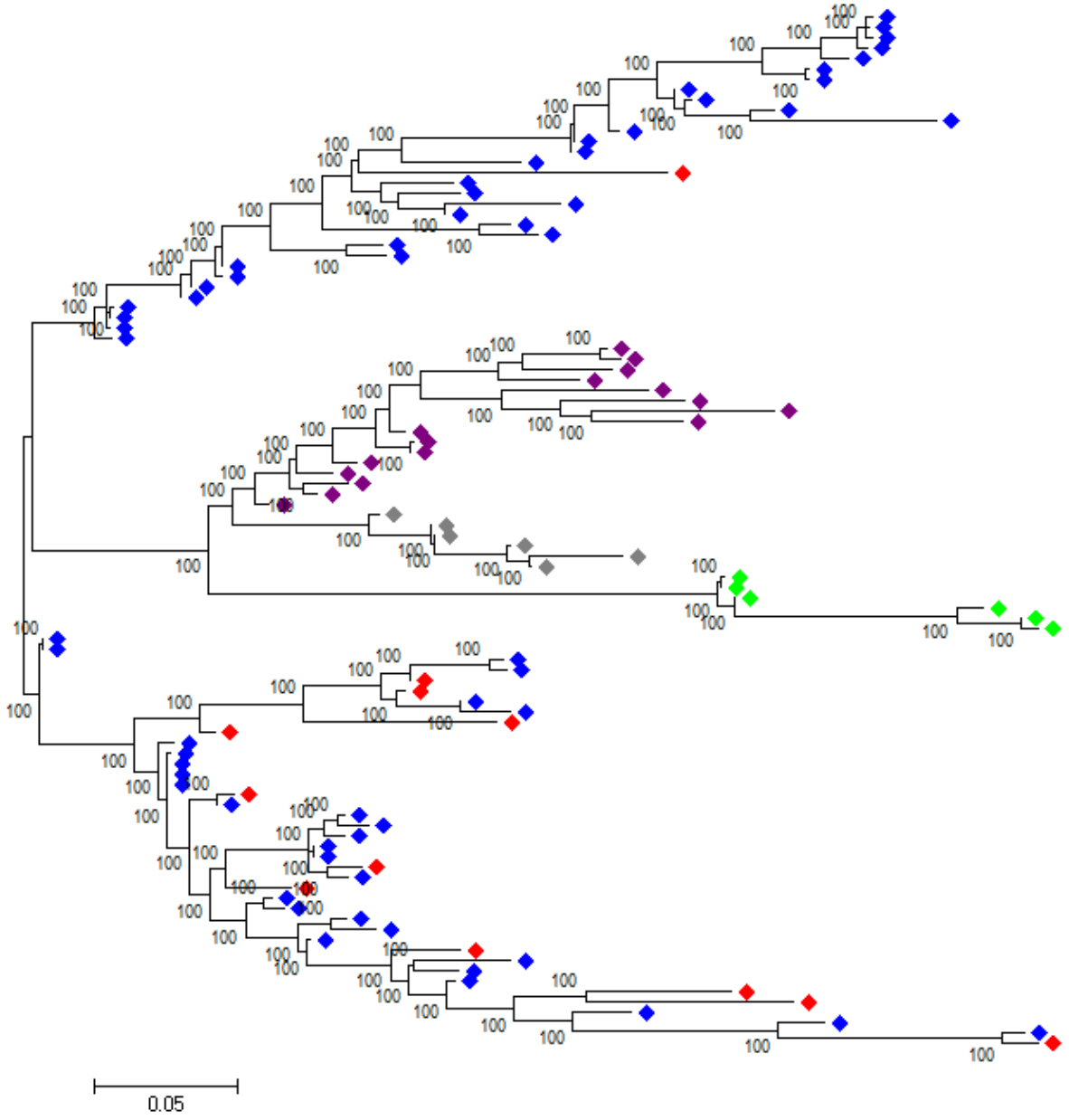


Figure S12.6. A 30-kb candidate region of introgression with fragment size that extends from 186290001 to 186320001 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The haplotype analysis suggests region of shared ancestry

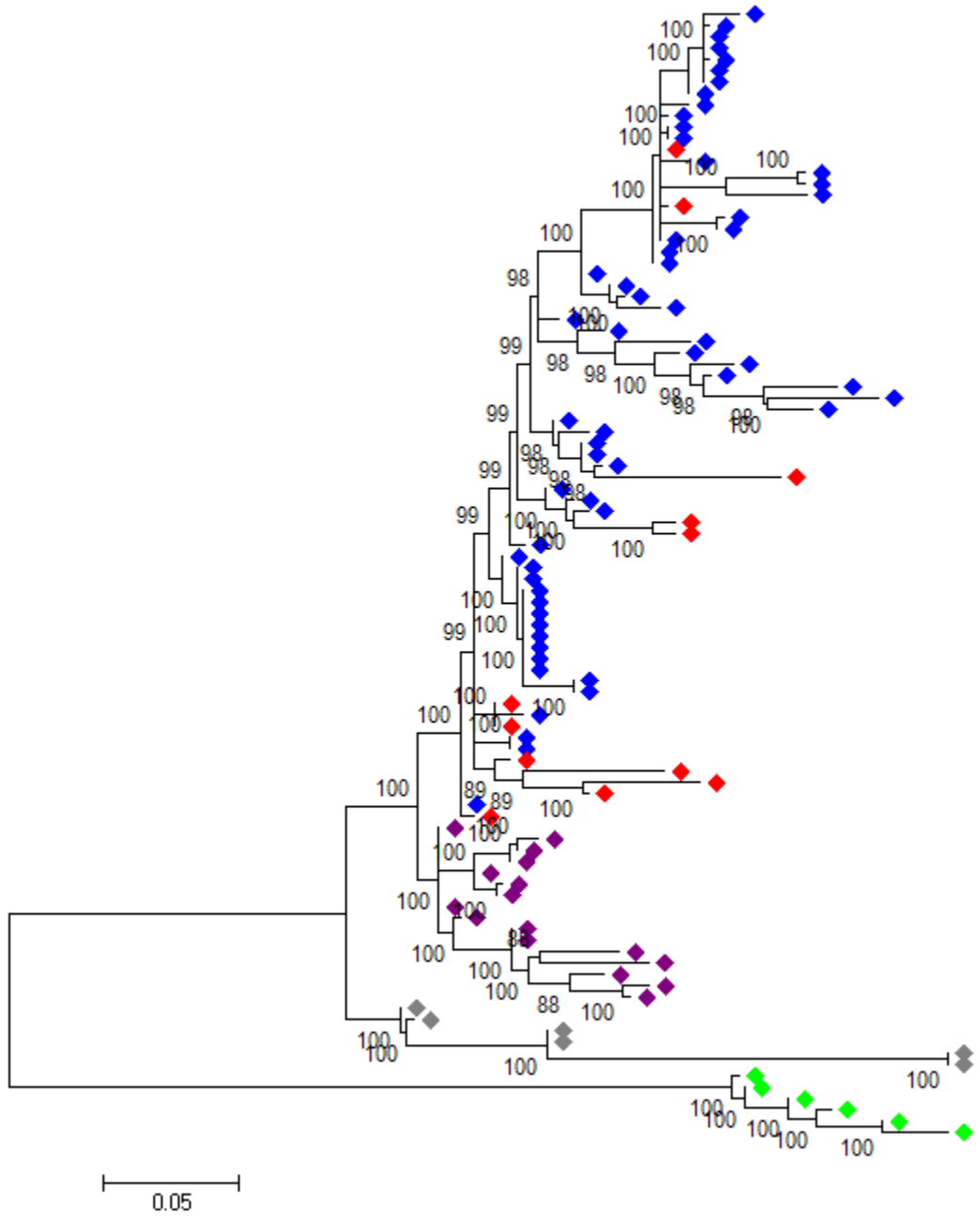


Figure S12.7. A 25-kb candidate region of introgression with fragment size that extends from 81835001 to 81860001 bp on chromosome 2. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The haplotype analysis suggests region of shared ancestry

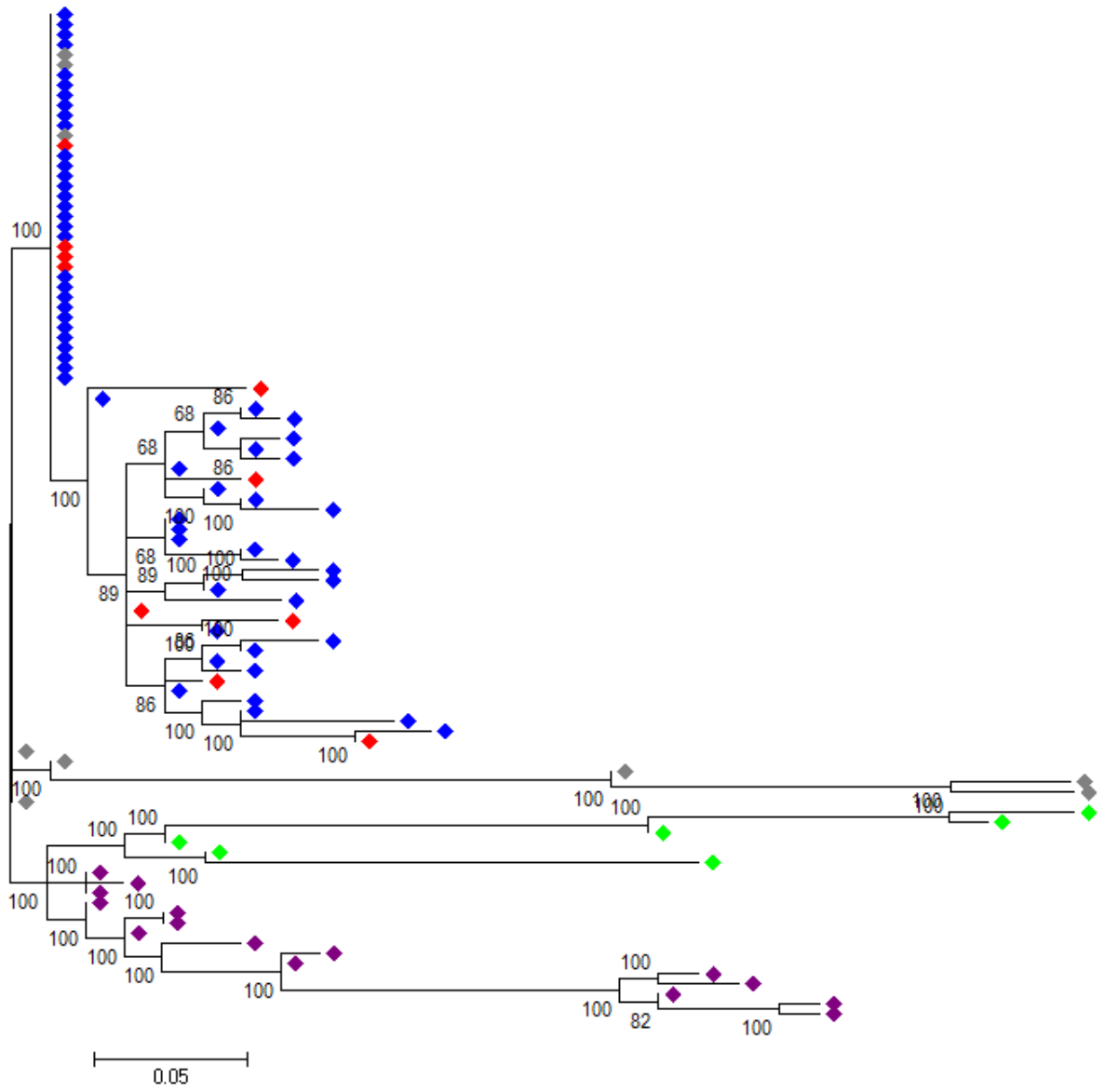


Figure S12.8. A 35-kb candidate region of introgression with fragment size that extends from 142360001 to 142395001 bp on chromosome 2. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The haplotype analysis suggests region of ambiguity

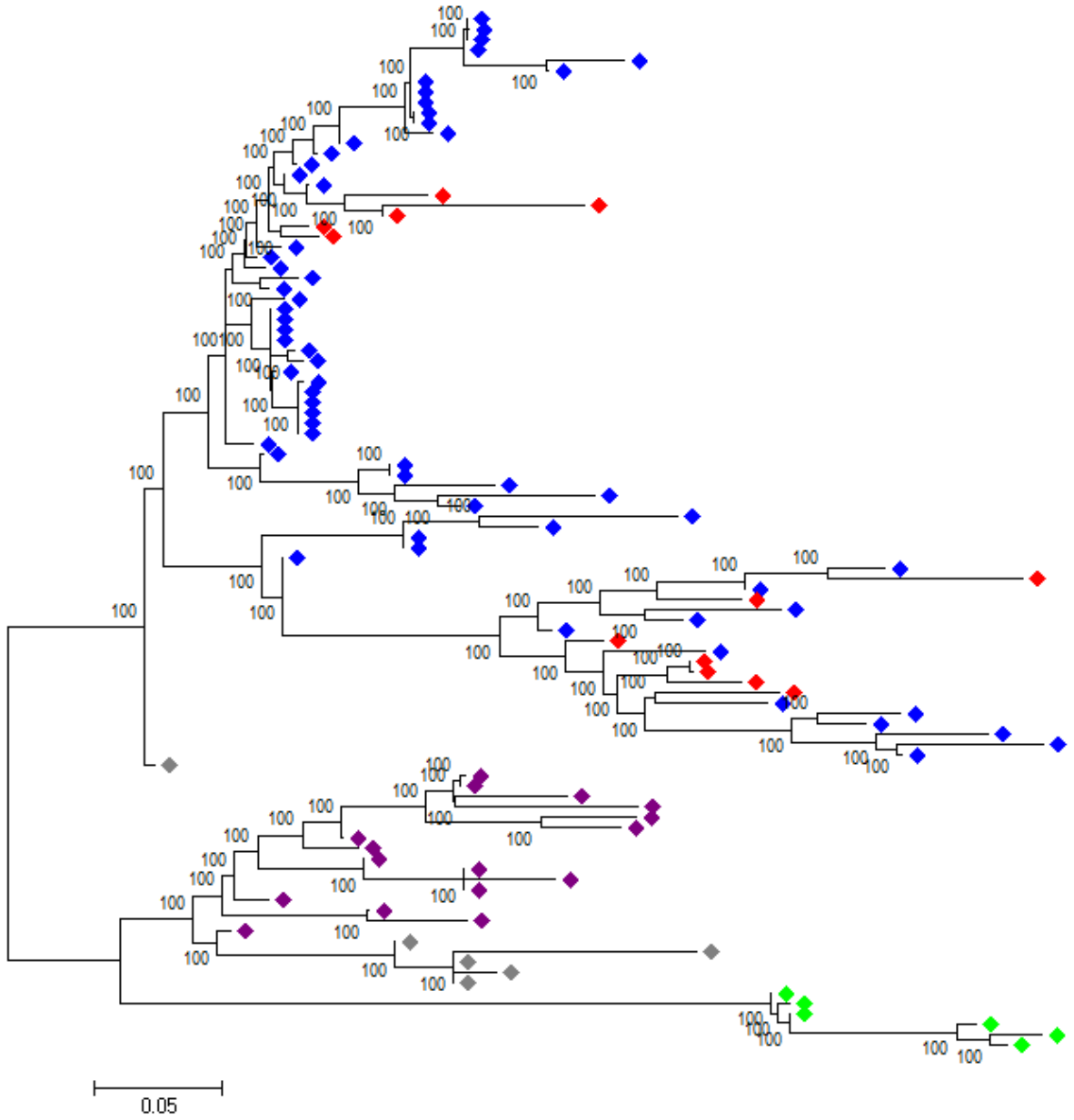


Figure S12.9. A 20-kb candidate region of introgression with fragment size that extends from 12170001 to 12190001 bp on chromosome 3. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The haplotype analysis suggests region of ambiguity

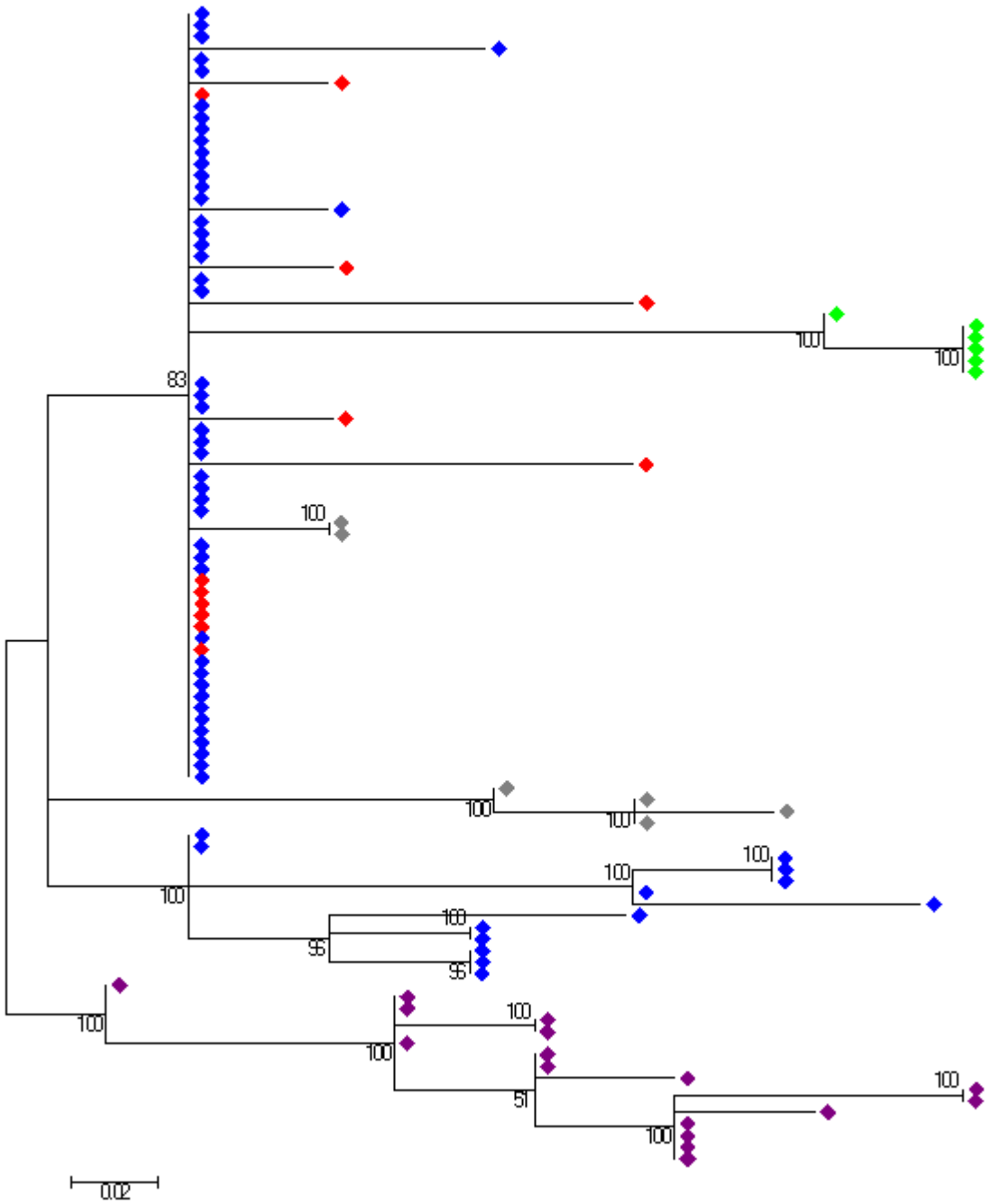


Figure S12.10. A 20-kb candidate region of introgression with fragment size that extends from 73505001 to 73525001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The haplotype analysis suggests region of ambiguity

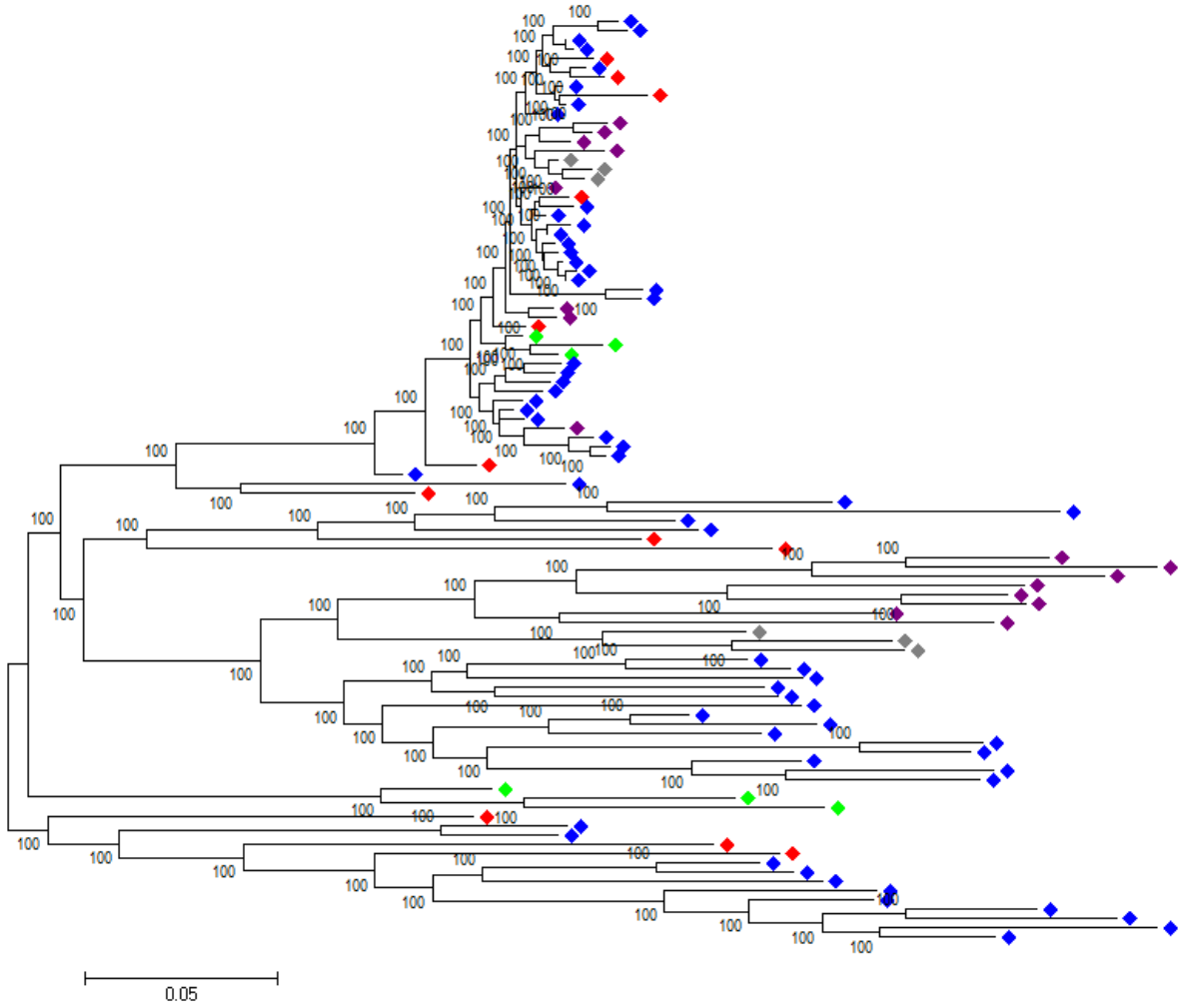


Figure S12.11. A 35-kb candidate region of introgression with fragment size that extends from 85380001 to 85415001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The haplotype analysis suggests region of ambiguity

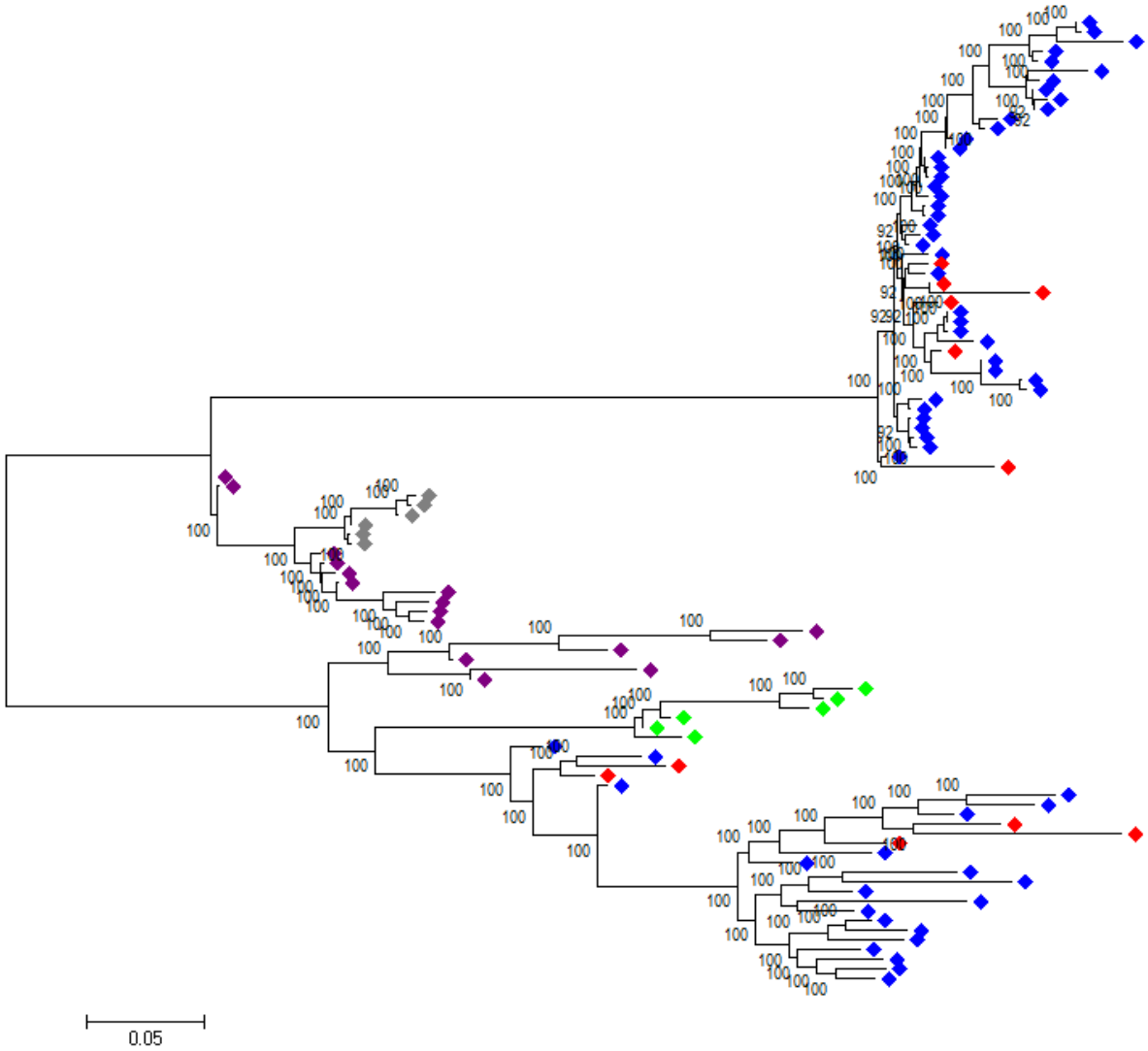


Figure S12.12. A 65-kb region with fragment size that extends from 13625001 to 13690001 bp on chromosome 14. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. ABBA-BABA test shows a significant peak between Ceylon junglefowl and domestic chicken but haplotype relationships analysis suggests rather a pattern of introgression from green junglefowl into domestic chicken and wild red junglefowl samples.

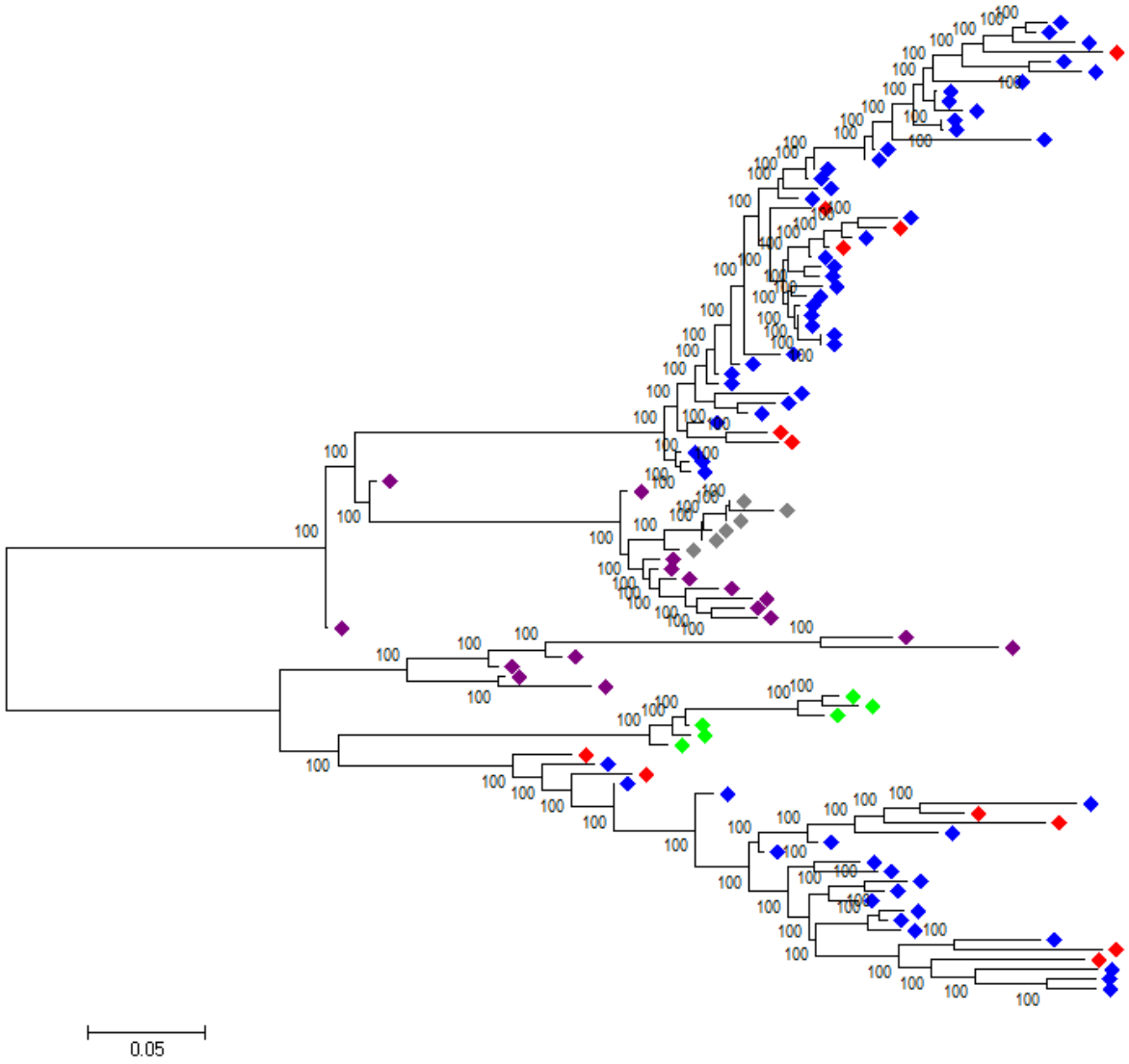


Figure S12.13. A 60-kb region with fragment size that extends from 14255001 to 14315001 bp on chromosome 14. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. ABBA-BABA Ceylon junglefowl - domestic chicken test shows significant peak between Ceylon junglefowl and domestic chicken but haplotypes relationship analysis suggests rather an introgression from the green junglefowl.

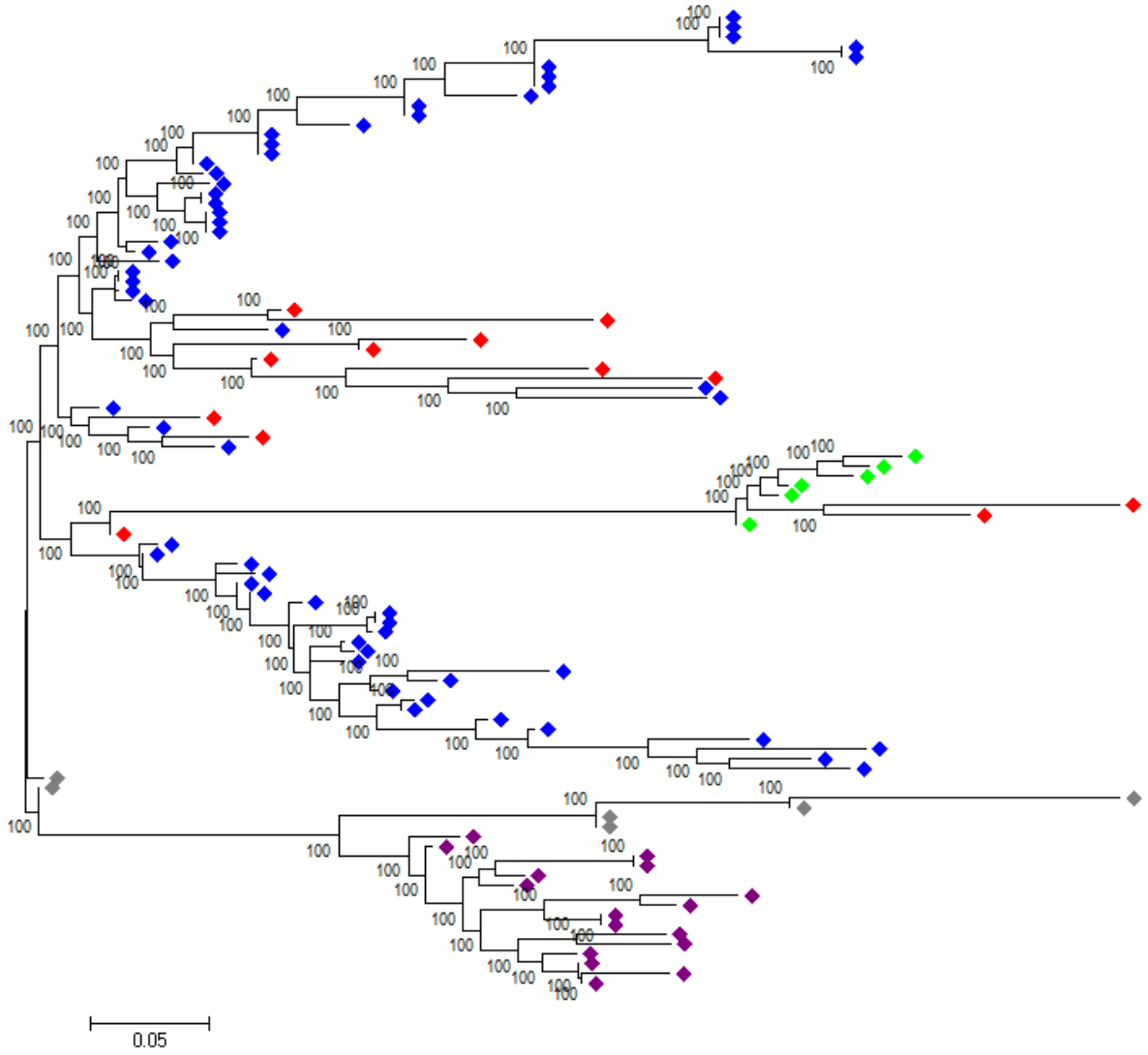


Figure S12.14. A 25-kb candidate region of introgression with fragment size that extends from 5420001 to 5445001 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The haplotype analysis suggests region of ambiguity

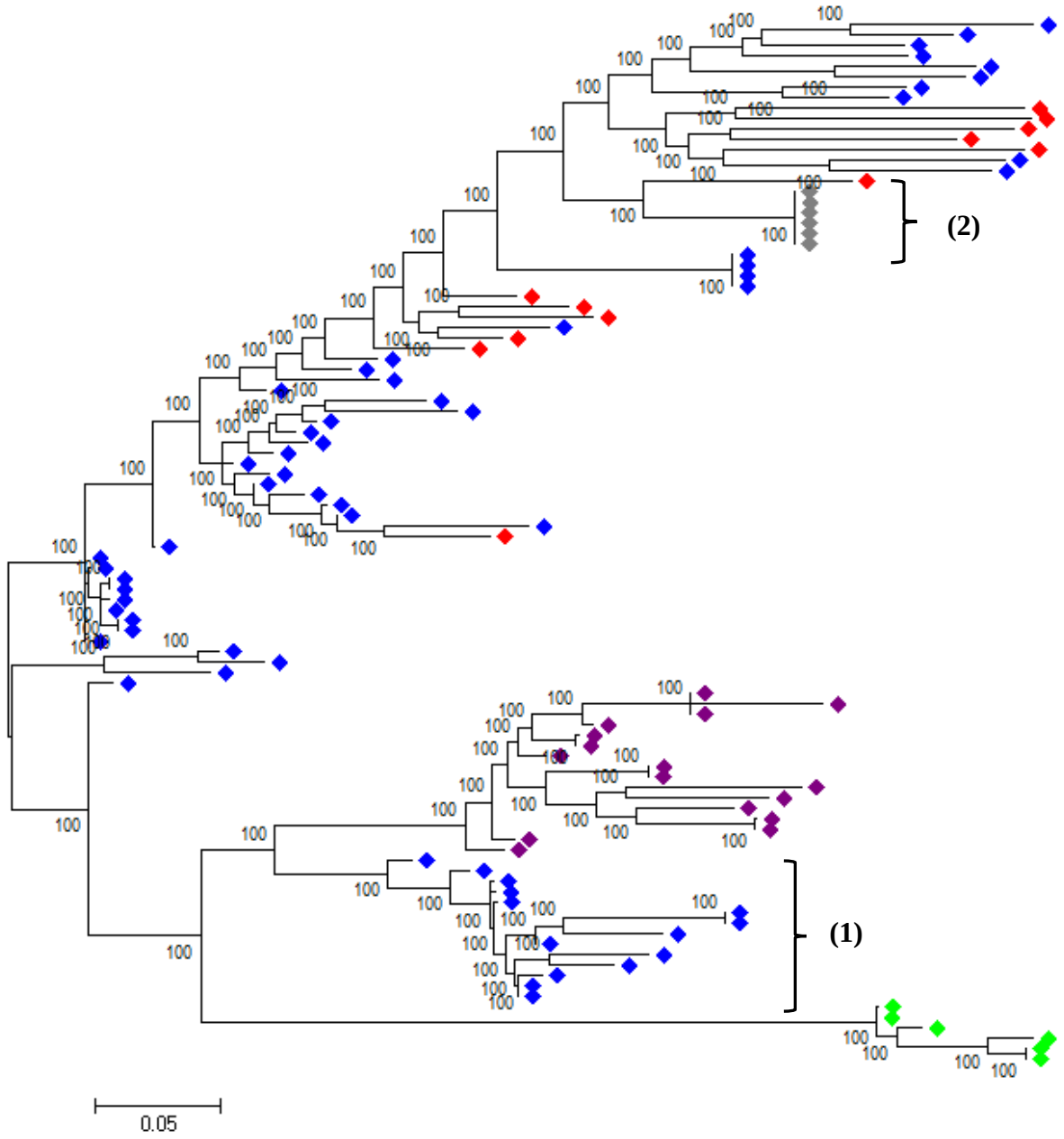


Figure S12.15. Haplotype based maximum likelihood tree for two patterns of introgression based. 55-kb fragment size that extends from 2680001 to 2735001 bp on chromosome 19. ◆: Domestic chicken, ◆: red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of Ceylon junglefowl or grey junglefowl haplotypes into chicken from Ethiopia. **(2)** Introgression of domestic chicken into grey junglefowl. In absence of wild grey junglefowl haplotypes I cannot be certain about the species introgression origin.

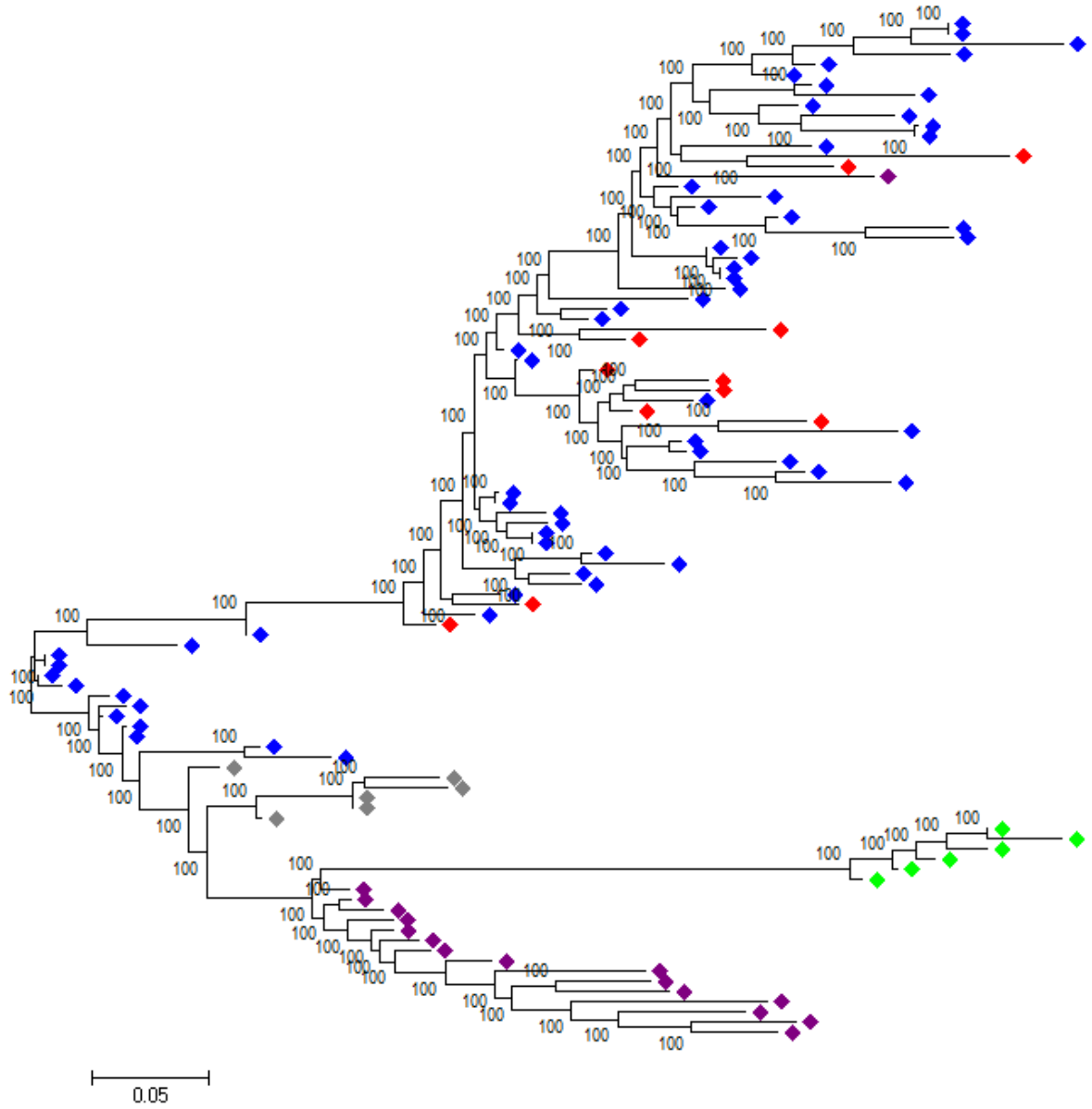


Figure S12.16. A 30-kb candidate region of introgression with fragment size that extends from 4005001 to 4035001 bp on chromosome 23. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The haplotype analysis suggests region of ambiguity

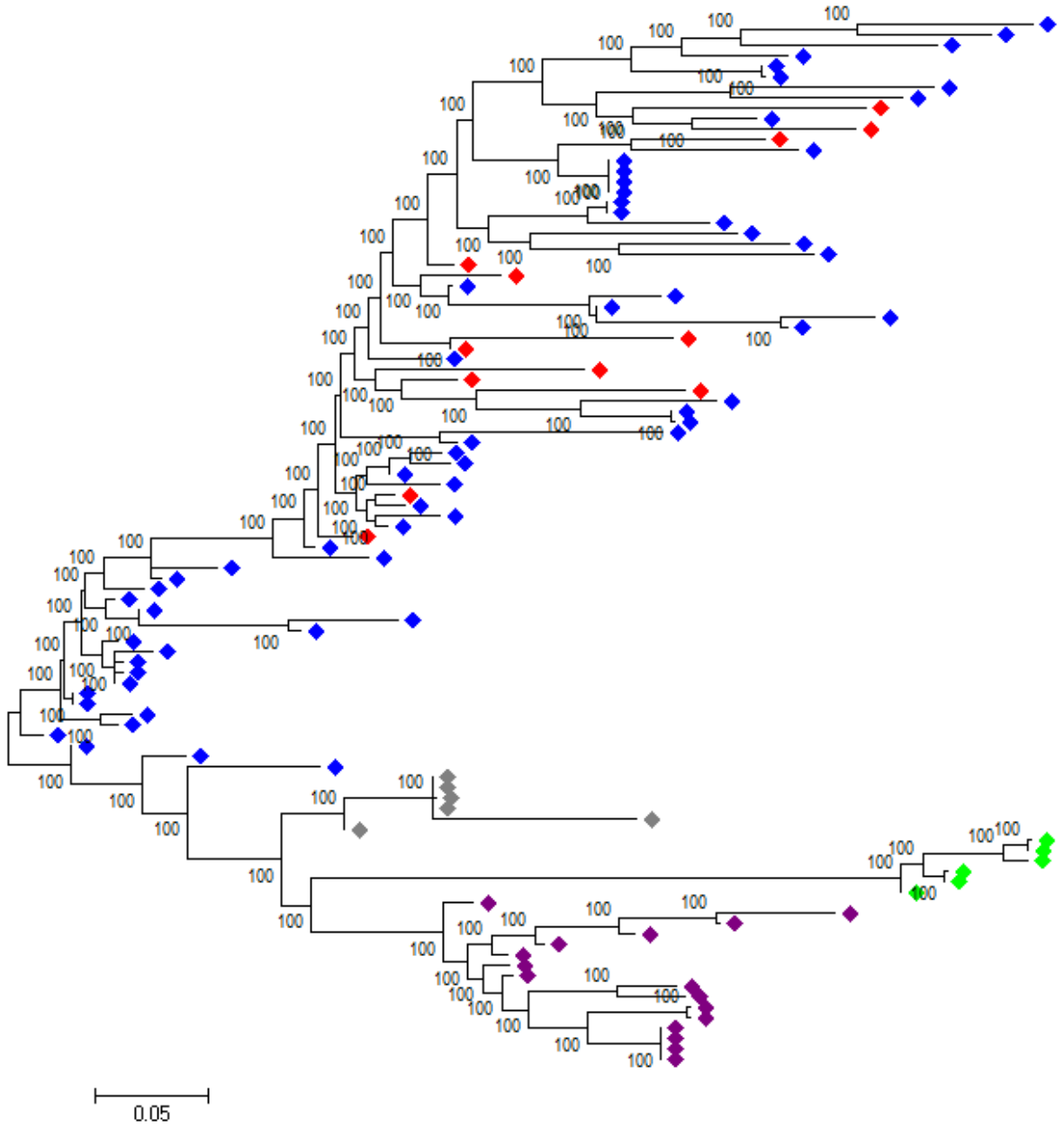


Figure S12.17. A 25-kb candidate region of introgression with fragment size that extends from 325001 to 350001 bp on chromosome 25. ◆: Domestic chicken, ◆: red junglefowl, ◆: Grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. The haplotype analysis suggests region of ambiguity

Chapter 5

Introgression of Javan red junglefowl *Gallus g. bankiva* to domestic chicken: Insights from autosomal whole-genome sequences

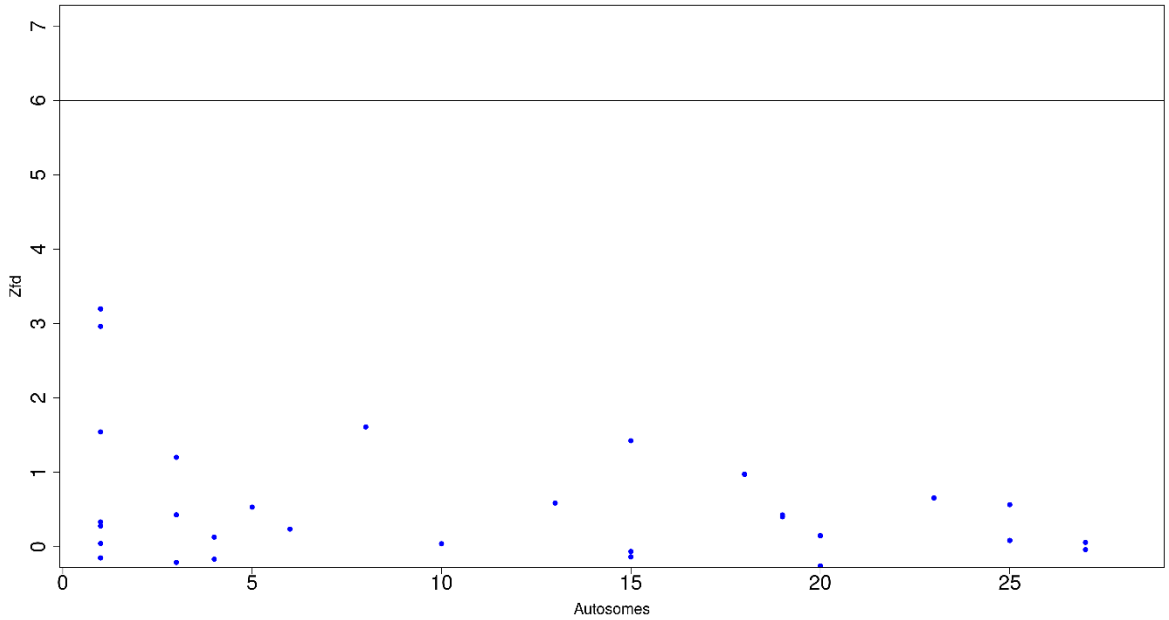


Figure S1. Chromosomal distribution of 20 kb windows having less than 100 SNPs for the Javan red junglefowl – domestic chicken ABBA/BABA plot (n = 170 windows). The Y-axis correspond to the values obtained from *Zfd* score (horizontal line represents the significant threshold). X–axis includes chromosomes 1 – 28.

Figures S2. Haplotypes relationships analysis for candidate regions of introgression from Javan red junglefowl into domestic chicken. ABBA – BABA test Javan red junglefowl - domestic chicken.

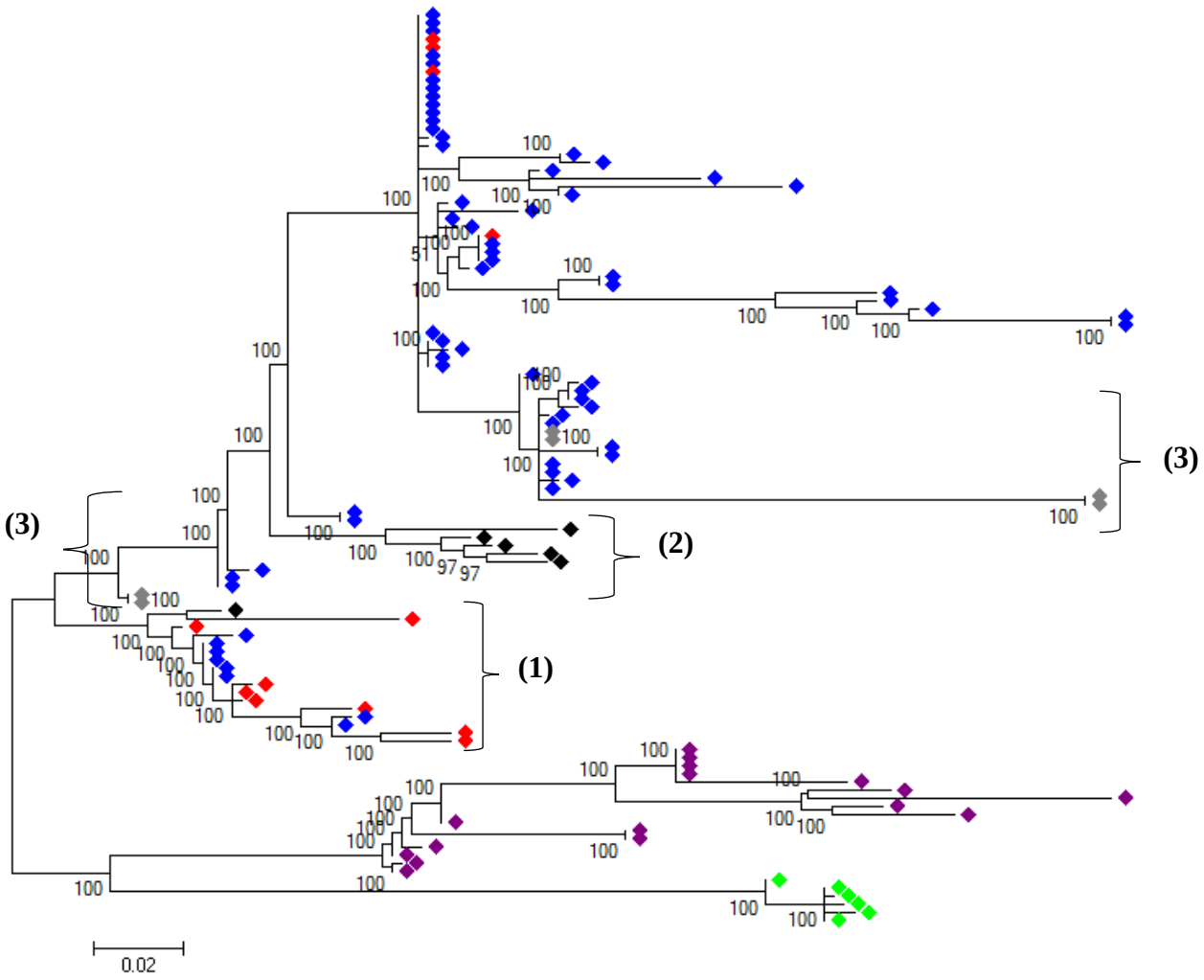


Figure S2.1. Haplotype based maximum likelihood tree for three patterns of introgression based on the 20-kb fragment size that extends from 16520000-16540000 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of a single haplotype of Javan red junglefowl into some domestic chicken and red junglefowl samples, **(2)** introgression of some domestic chicken haplotypes into Javan red junglefowl, **(3)** introgression of domestic chicken haplotypes into grey junglefowl.

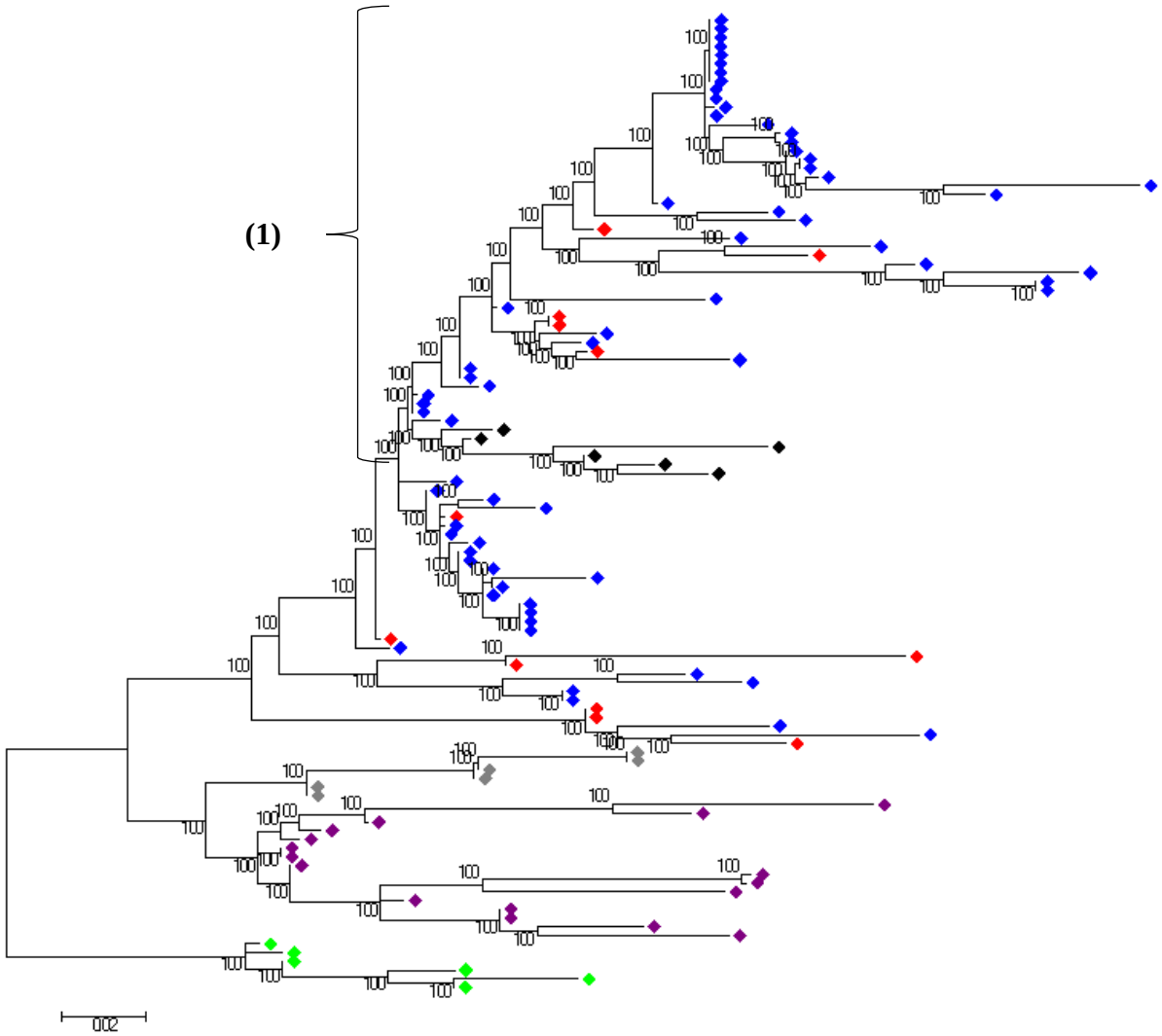


Figure S2.2. Haplotype based maximum likelihood tree for region of introgression from **(1)** Javan red junglefowl into domestic chicken/red junglefowl haplotypes. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 30-kb fragment size that extends from 23745000 to 23775000 bp on chromosome 1.

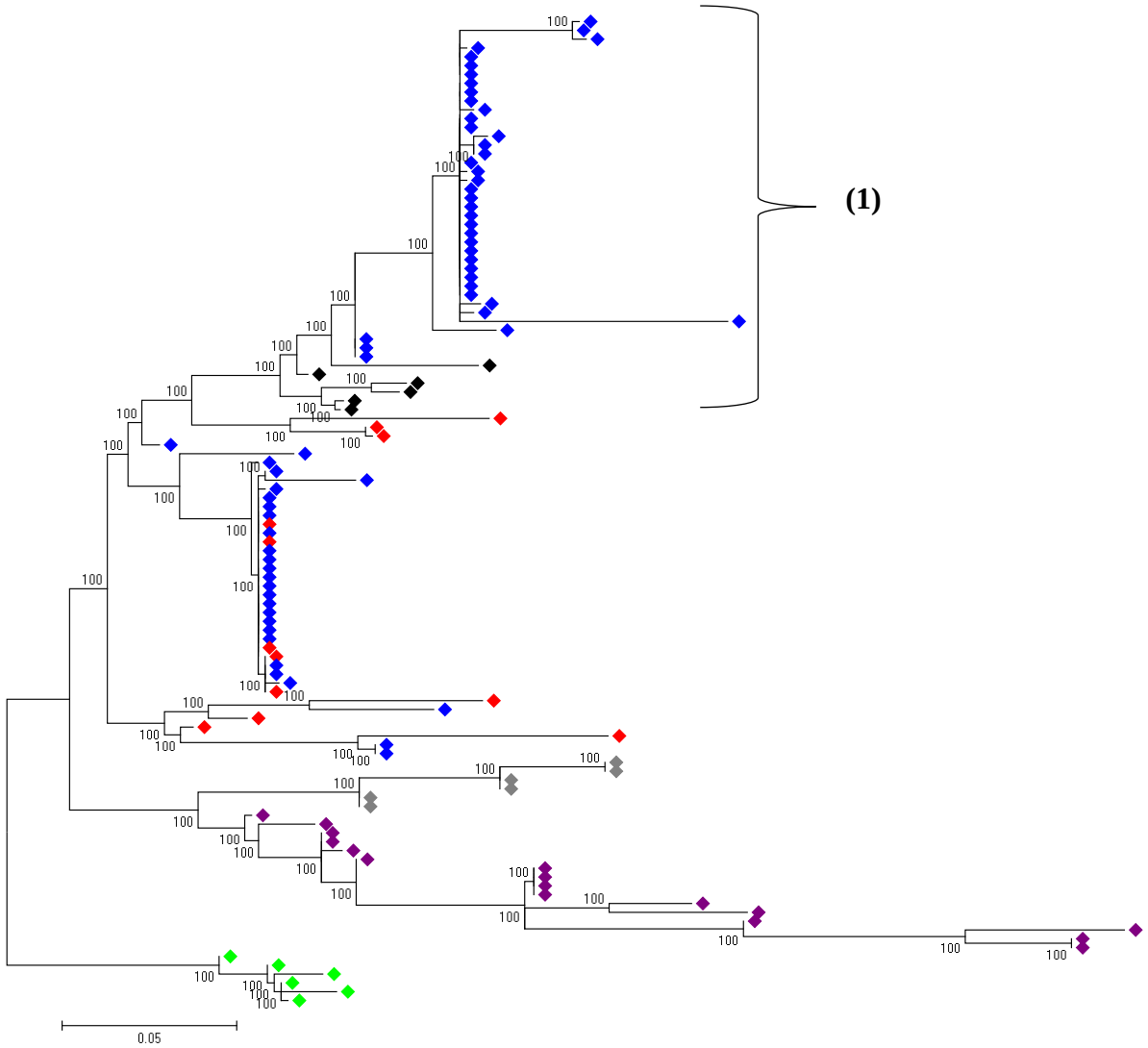


Figure S2.3. Haplotype based maximum likelihood tree for region of introgression from **(1)** Javan red junglefowl into domestic chicken/red junglefowl haplotypes. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 25-kb fragment size that extends from 32250000 to 32275000 bp on chromosome 1.

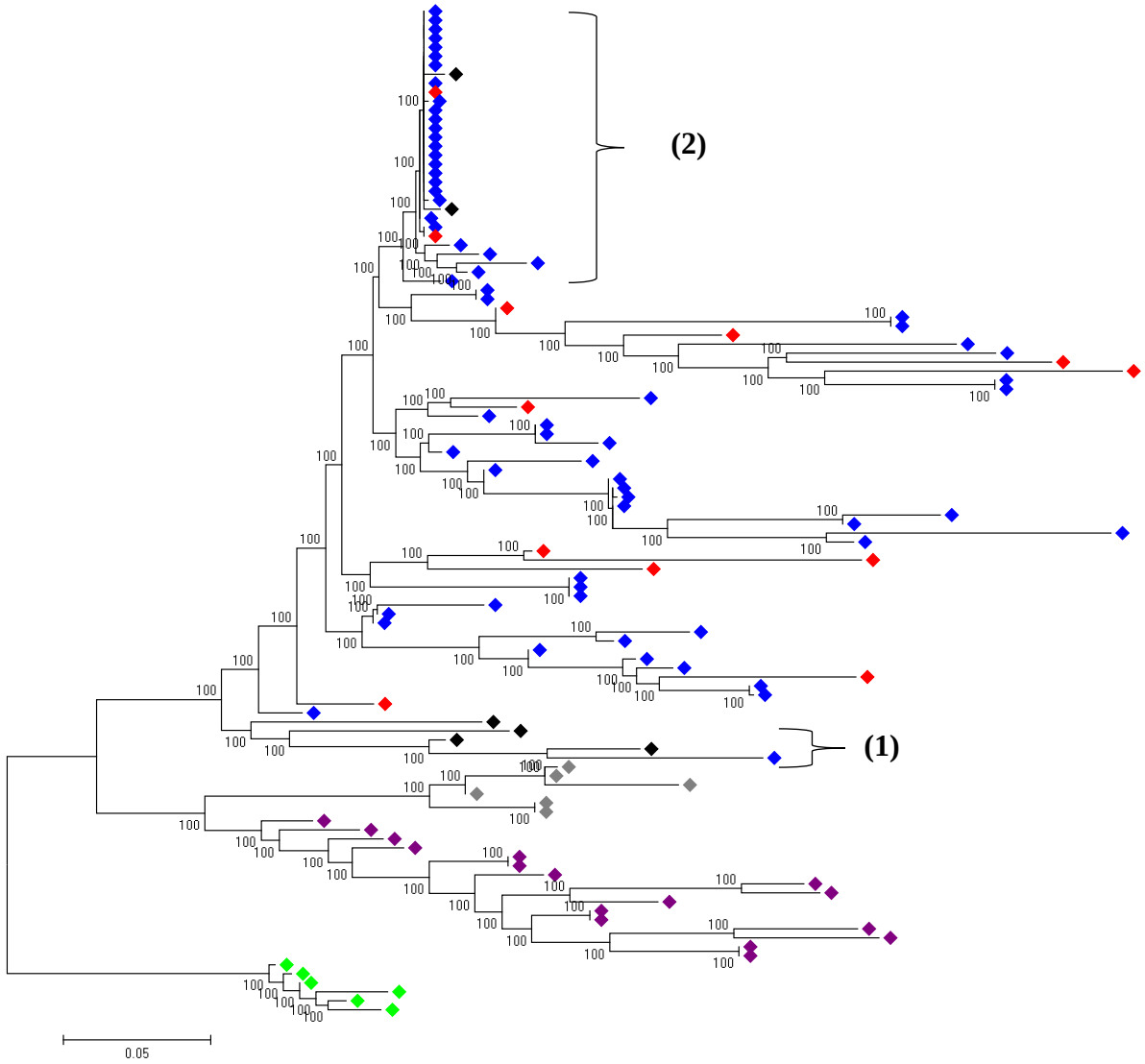


Figure S2.4. Haplotype based maximum likelihood tree for two patterns of introgression based on the 20-kb fragment size that extends from 176660000 to 176680000 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of Javan red junglefowl haplotypes into KeduHitam763 sample **(2)** introgression of domestic chicken/red junglefowl haplotypes into Javan red junglefowl.

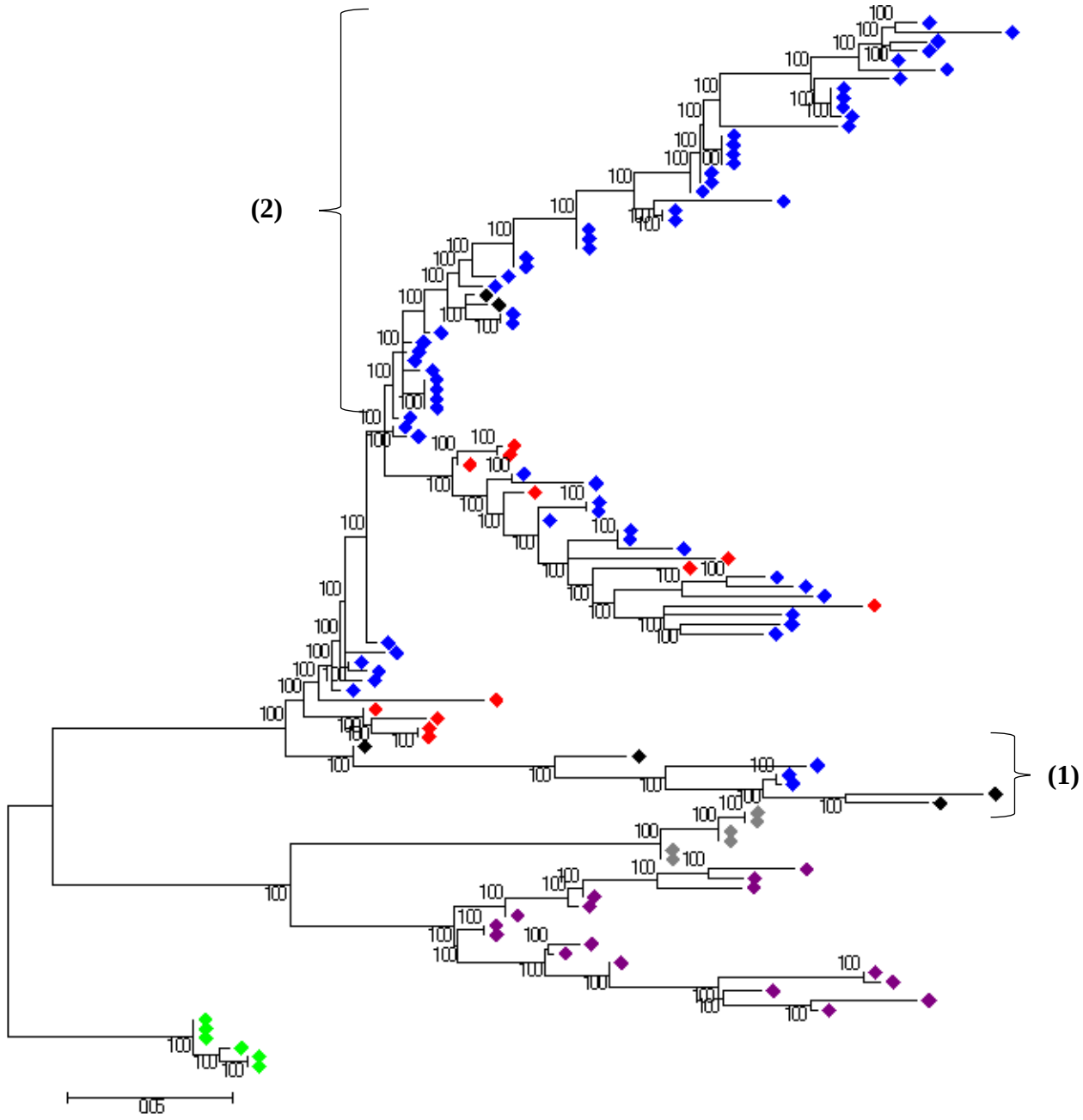


Figure S2.5. Haplotype based maximum likelihood tree for two patterns of introgression based on the 35-kb fragment size that extends from 70260000 to 70295000 bp on chromosome 2. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of Javan red junglefowl haplotypes into Sri Lanka10, 13 and 15 samples. **(2)** Introgression of domestic chicken haplotypes into Javan red junglefowl.

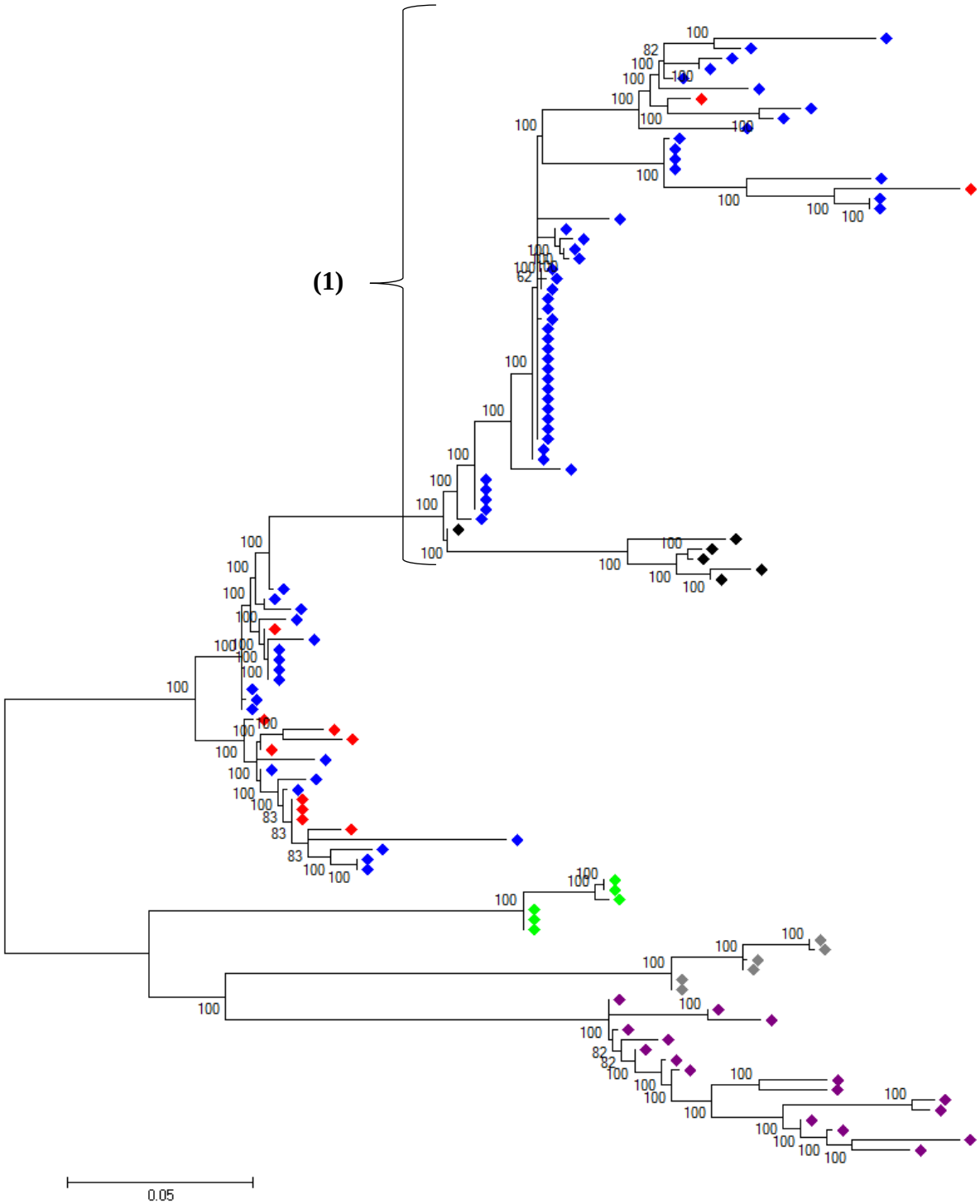


Figure S2.6. Haplotype based maximum likelihood tree for region of introgression from **(1)** Javan red junglefowl into domestic chicken. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 45-kb fragment size that extends from 70675000 to 70720000 bp on chromosome 2.

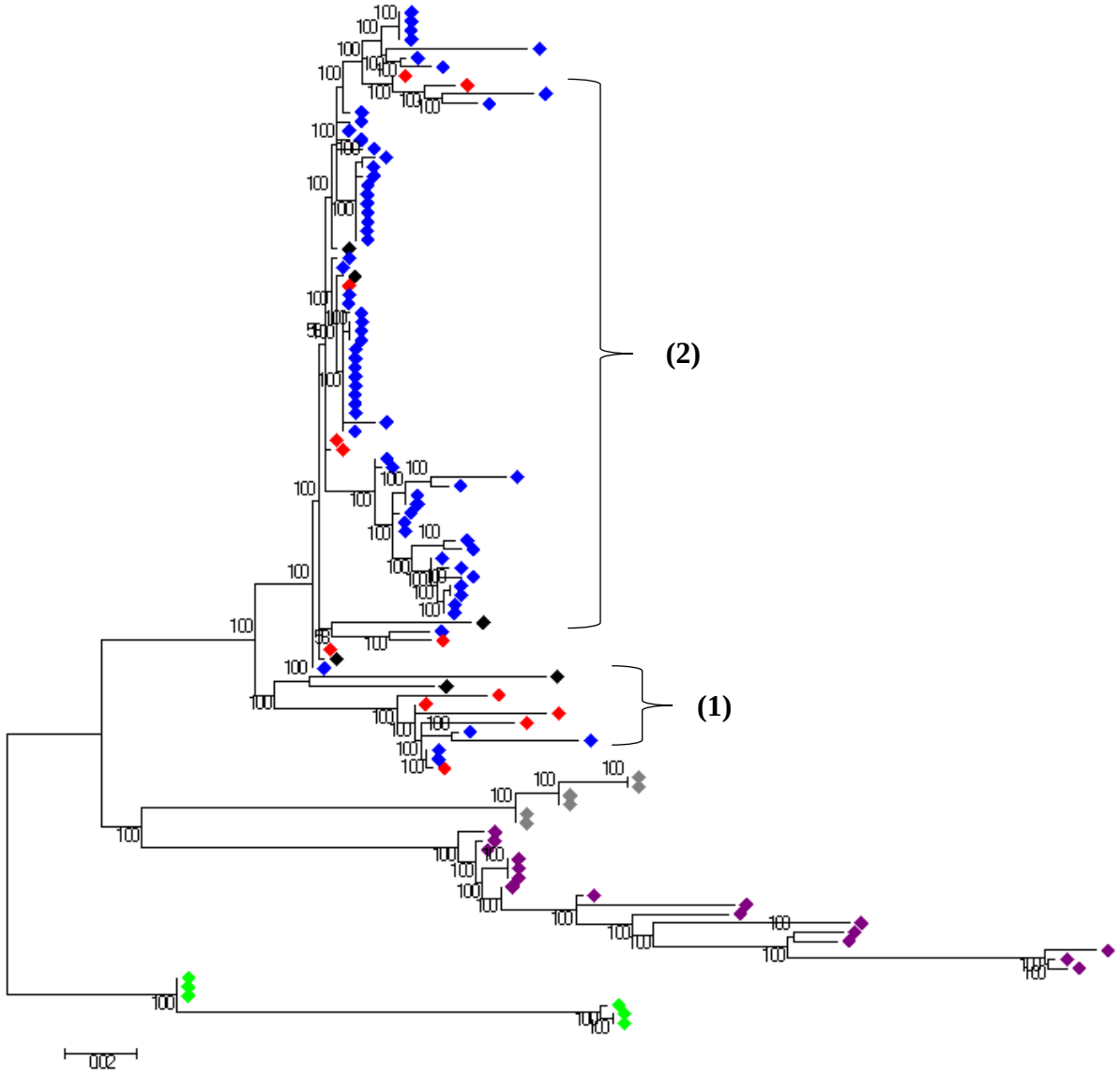


Figure S2.7. Haplotype based maximum likelihood tree for two patterns of introgression based on the 35-kb fragment size that extends from 71755000 to 71790000 bp on chromosome 2. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of Javan red junglefowl haplotypes into Sri Lanka9 and 12, KeduHitam766, red junglefowl3, 5 and Koen. **(2)** Introgression of domestic chicken/red junglefowl haplotypes into Javan red junglefowl.

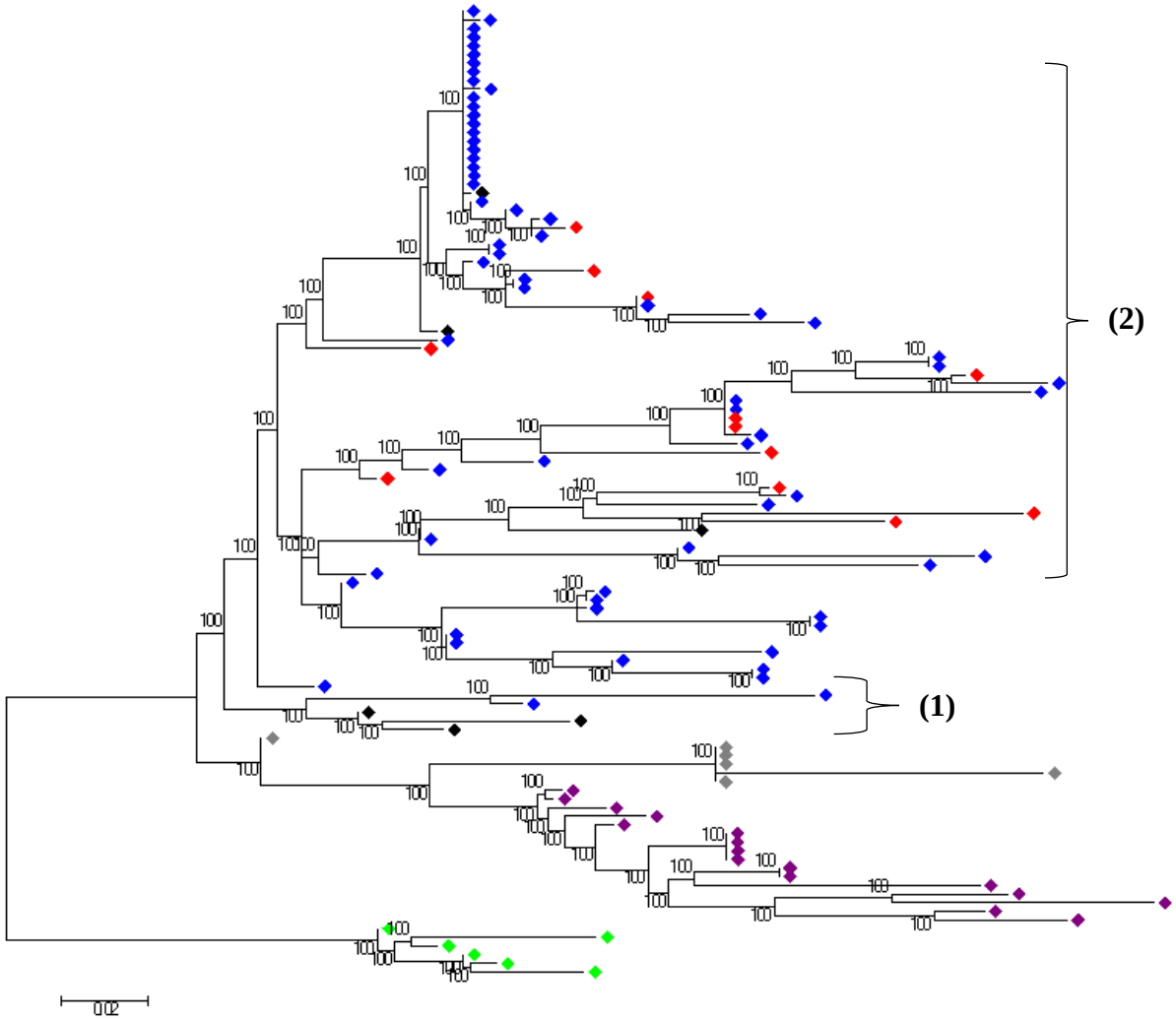


Figure S2.8. Haplotype based maximum likelihood tree for two patterns of introgression based on the 20-kb fragment size that extends from 63290000 to 63310000 bp on chromosome 3. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of Javan red junglefowl haplotypes into KeduHitam761 and 763. **(2)** Introgression of domestic chicken/red junglefowl haplotypes into Javan red junglefowl.

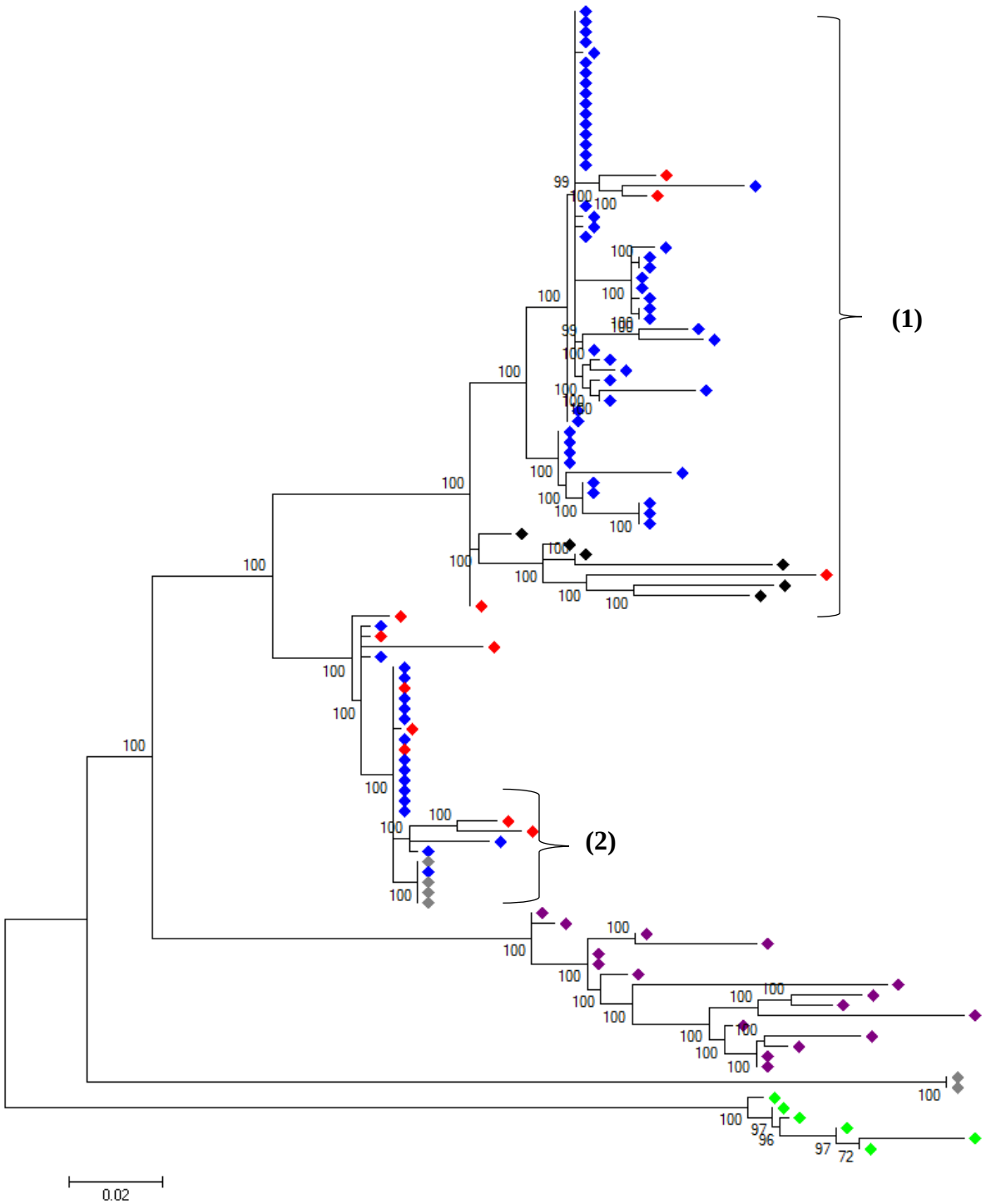


Figure S2.9. Haplotype based maximum likelihood tree for two patterns of introgression based on the 30-kb fragment size that extends from 27440000 to 27470000 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of Javan red junglefowl haplotypes into domestic chicken. **(2)** Introgression of domestic chicken/red junglefowl haplotypes into grey junglefowls.

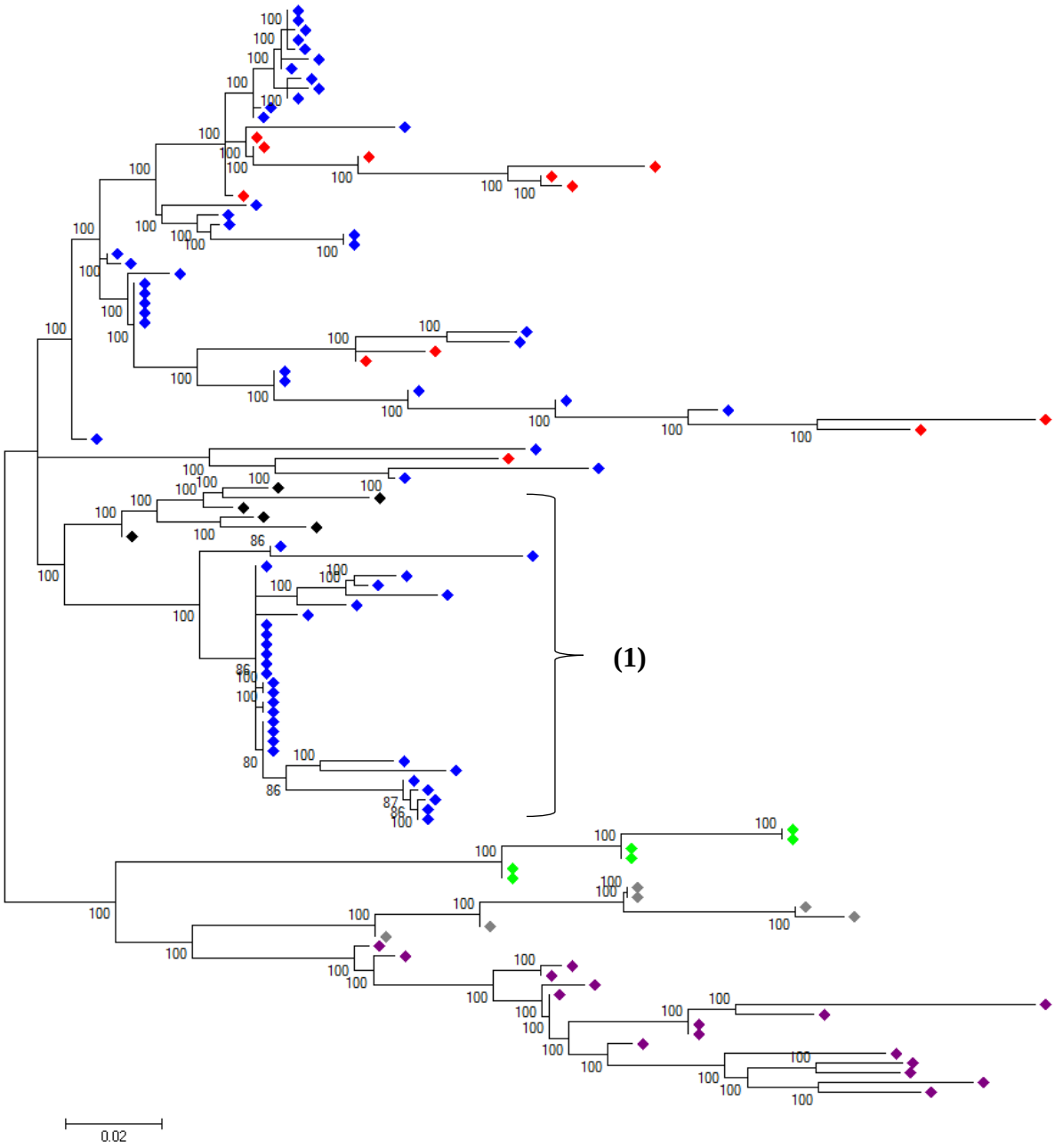


Figure S2.10. Haplotype based maximum likelihood tree for region of introgression from **(1)** Javan red junglefowl into domestic chicken haplotypes from Ethiopia, Saudi Arabia, Sri Lanka, hybrid Mechelse Koekoek and parent Poulet de Bresse. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 30-kb fragment size that extends from 67000000 to 67030000 bp on chromosome 4.

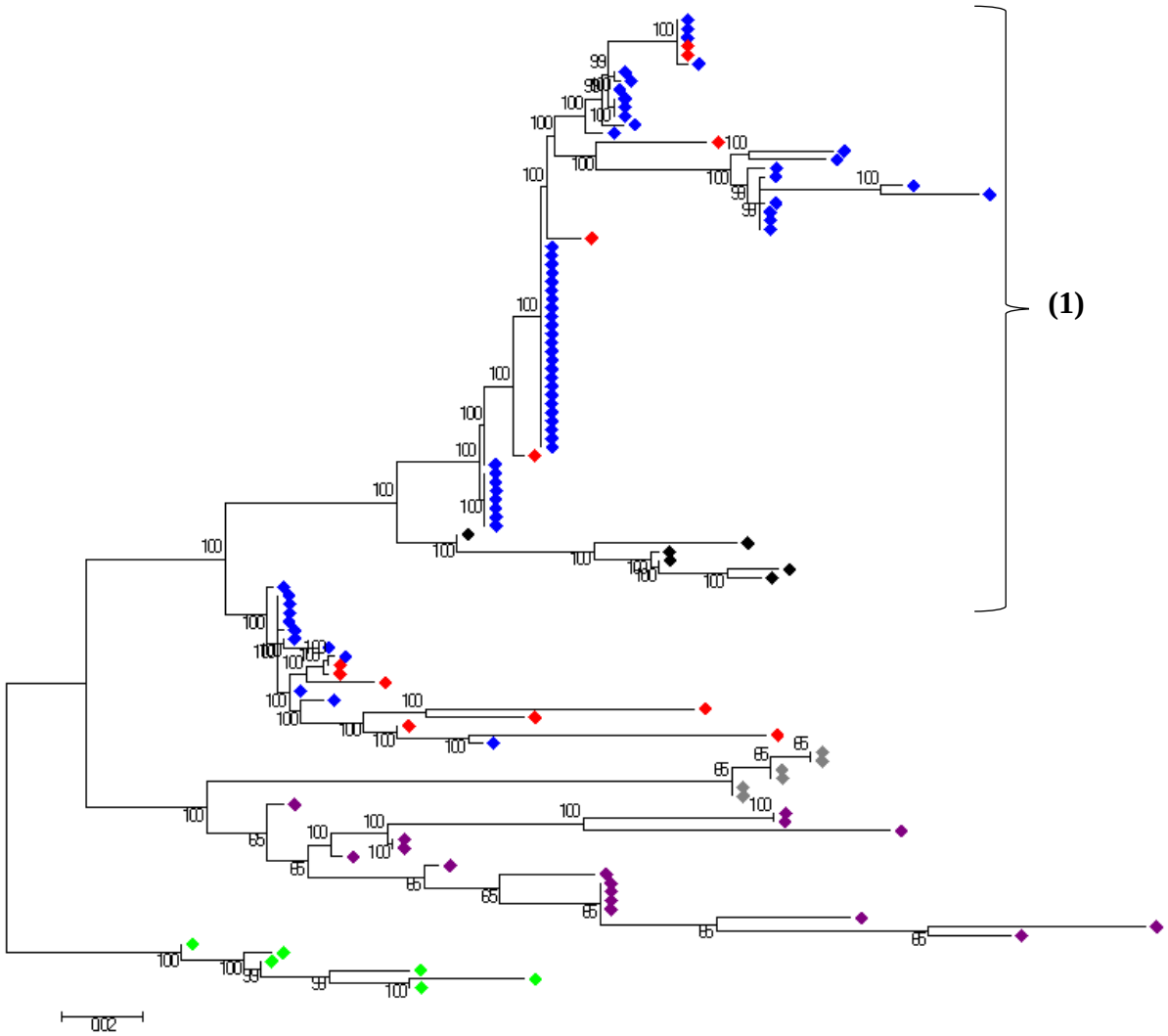


Figure S2.11. Haplotype based maximum likelihood tree for region of introgression from **(1)** Javan red junglefowl into domestic chicken/red junglefowl haplotypes with representatives from all populations. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 35-kb fragment size that extends from 40880000 to 40915000 bp on chromosome 5.

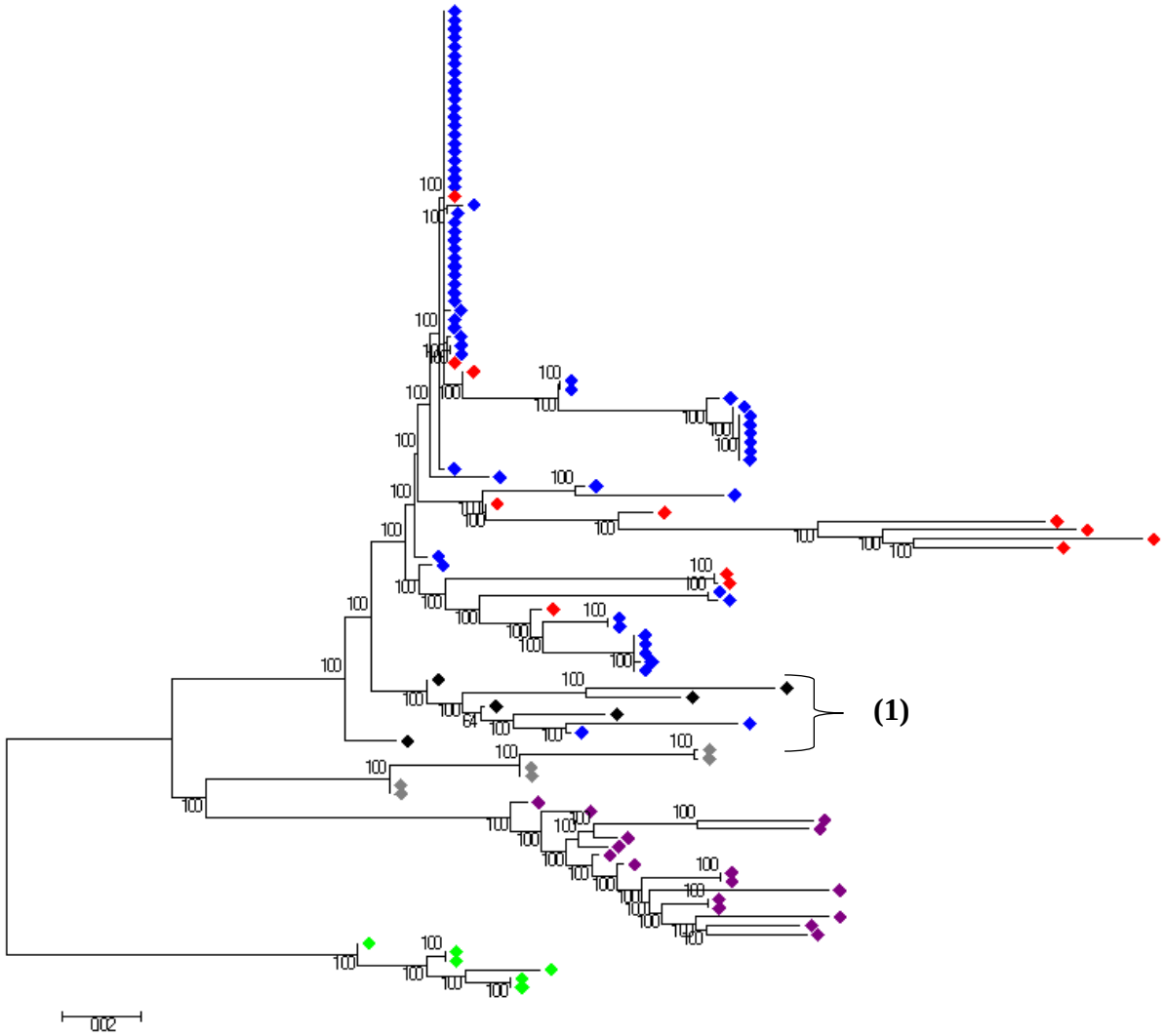


Figure S2.12. Haplotype based maximum likelihood tree for region of introgression from **(1)** Javan red junglefowl into domestic chicken haplotypes from Mechelse Koekoek and Poulet de Bresse. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 30-kb fragment size that extends from 17915000 to 17945000 bp on chromosome 6.

Figure S3. Haplotypes relationships analysis for candidate regions of introgression from grey junglefowl (*Gallus sonneratii*) into domestic chicken. ABBA – BABA test was Javan red junglefowl - domestic chicken.

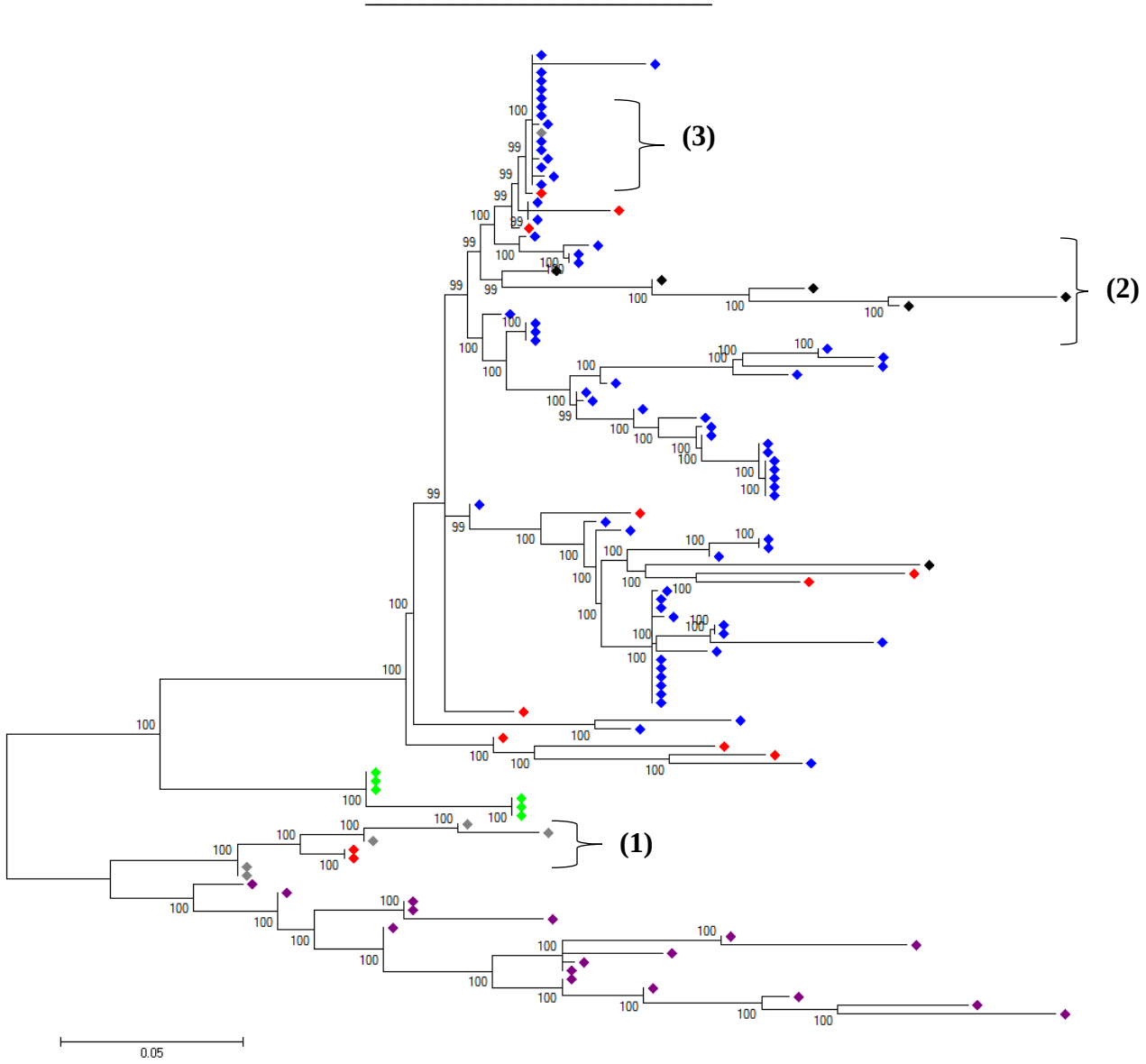


Figure S3.1 Haplotype based maximum likelihood tree for region of introgression from (**1**) grey junglefowl into red junglefowl2 sample **(2)** domestic chicken into Javan red junglefowl and **(3)** domestic chicken into grey junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 1485000 to 1505000 bp on chromosome 21.

Figures S4. Haplotypes relationships analysis for candidate regions of introgression from domestic chicken into Javan red junglefowl. ABBA – BABA test Javan red junglefowl - domestic chicken.

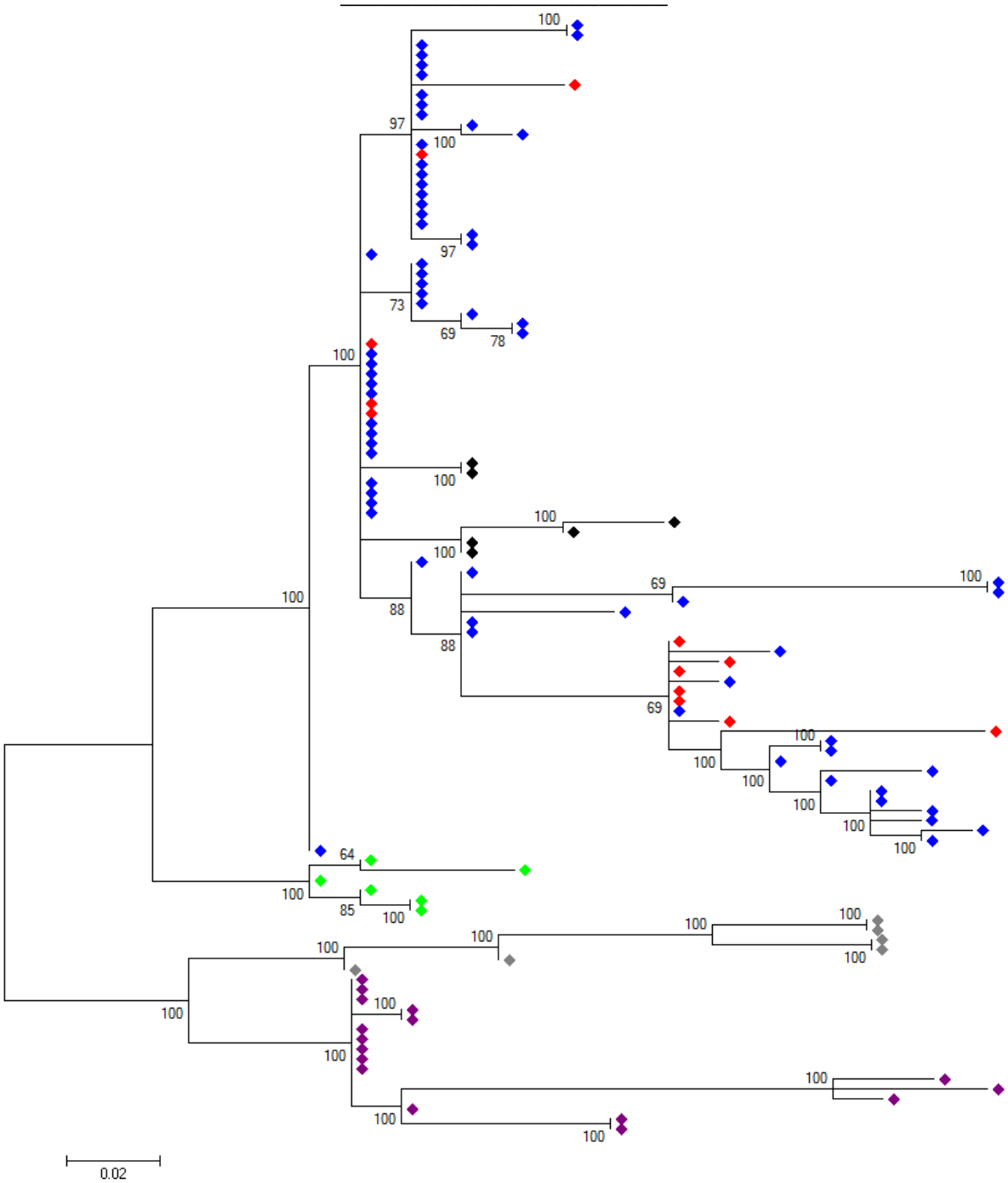


Figure S4.1. Haplotype based maximum likelihood tree for region of introgression from domestic chicken/red junglefowl haplotypes into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 31510000 to 31530000 bp on chromosome 1.

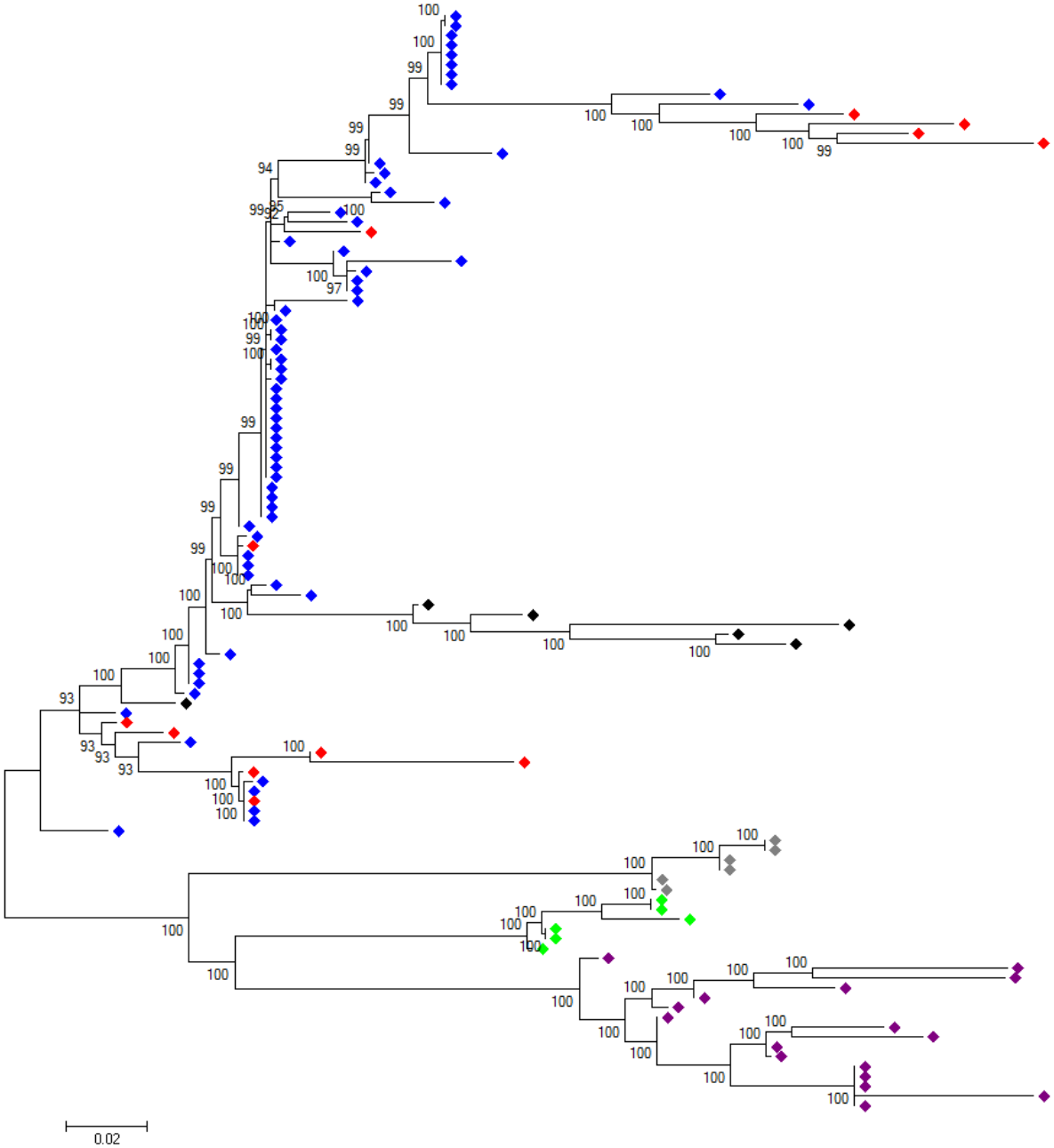


Figure S4.2. Haplotype based maximum likelihood tree for region of introgression from domestic chicken/red junglefowl haplotypes into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 30-kb fragment size that extends from 58890000 to 58920000 bp on chromosome 1.

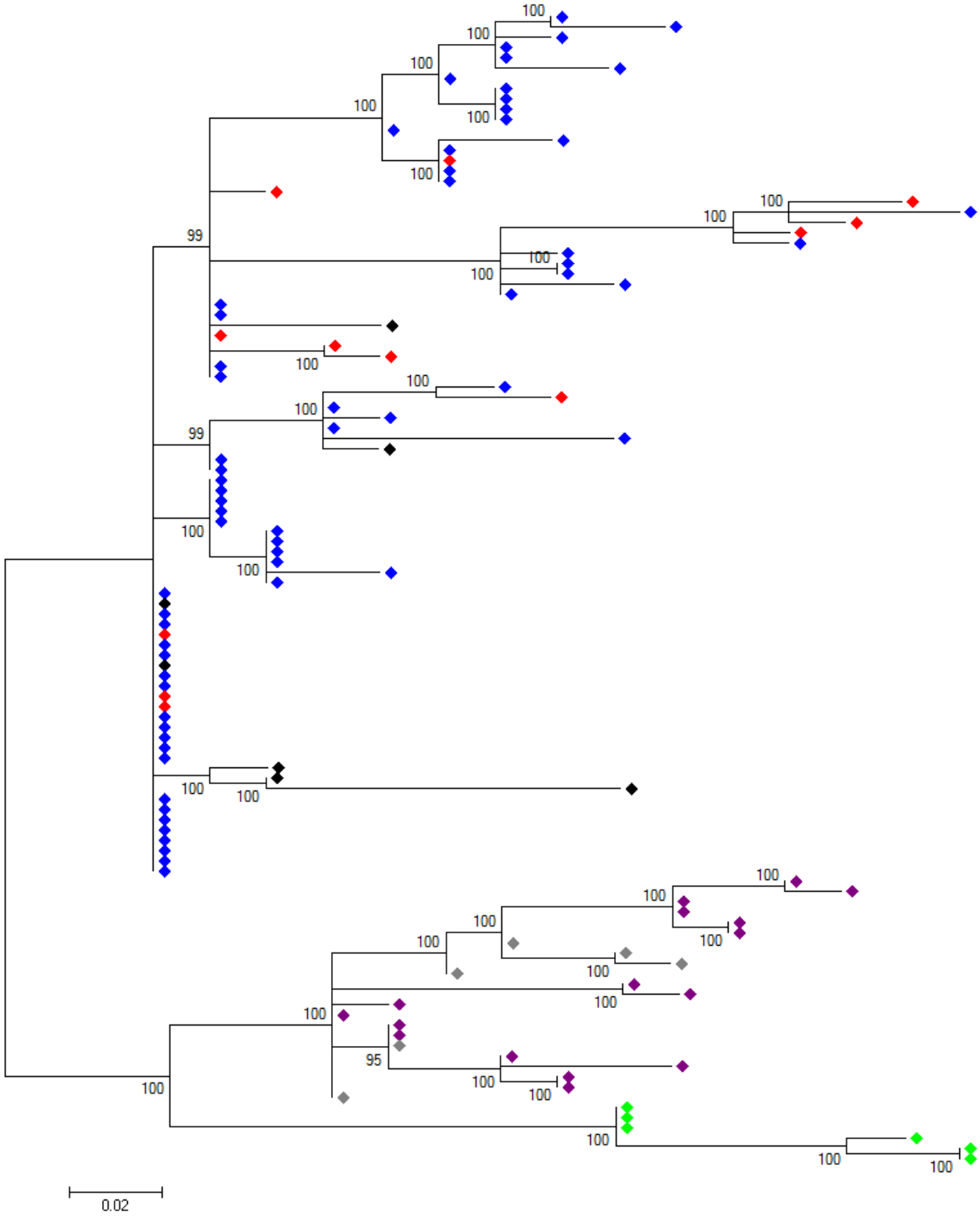


Figure S4.3. Haplotype based maximum likelihood tree for region of introgression from domestic chicken/red junglefowl haplotypes into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 62605000 to 62625000 bp on chromosome 1.

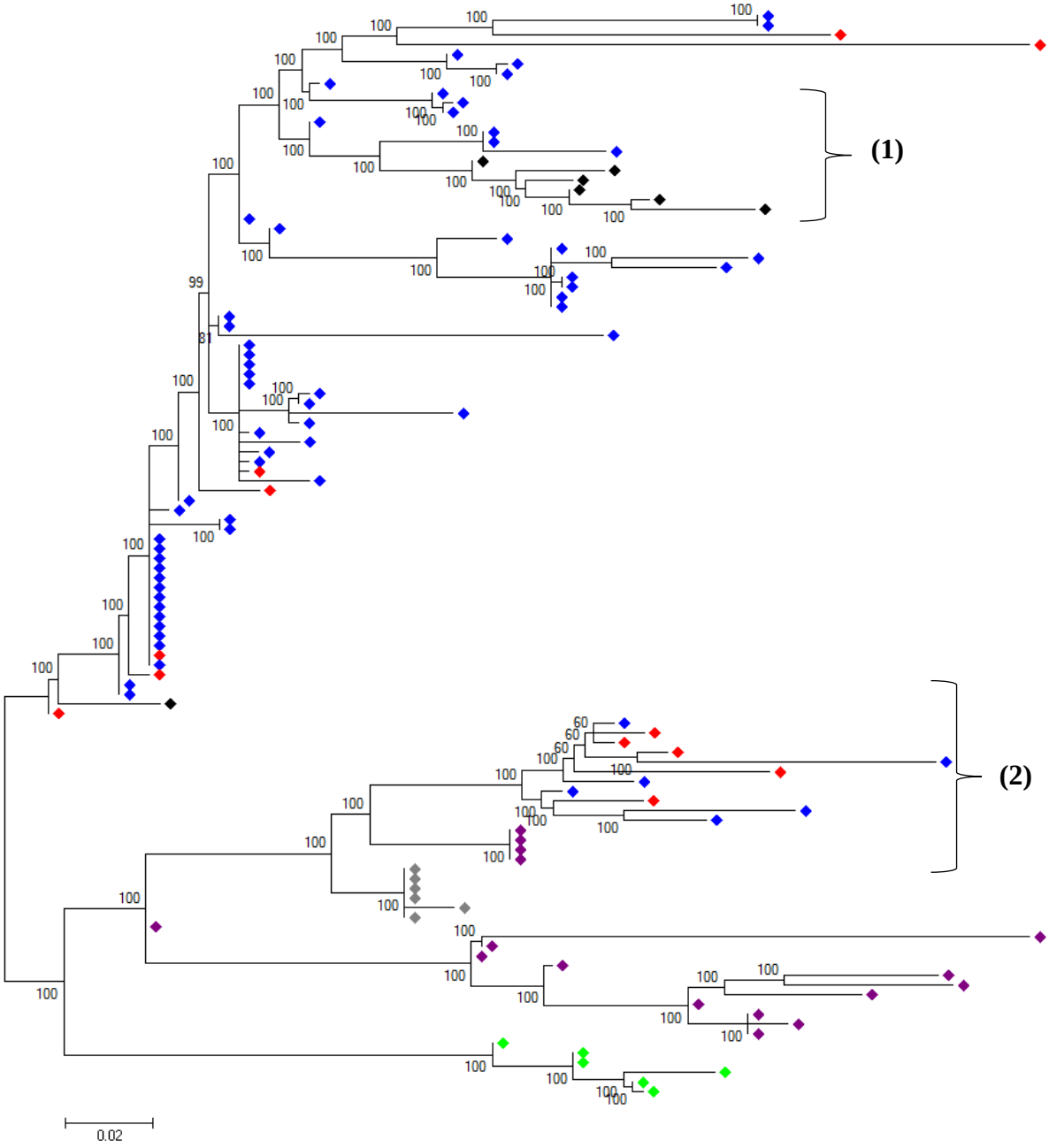


Figure S4.4. Haplotype based maximum likelihood tree for region of introgression **(1)** from domestic chicken/red junglefowl haplotypes into Javan red junglefowl and **(2)** a likely shared ancestry or an ambiguous relationship between red junglefowl/domestic chickens and Ceylon junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 25-kb fragment size that extends from 100590000 to 100615000 bp on chromosome 1.

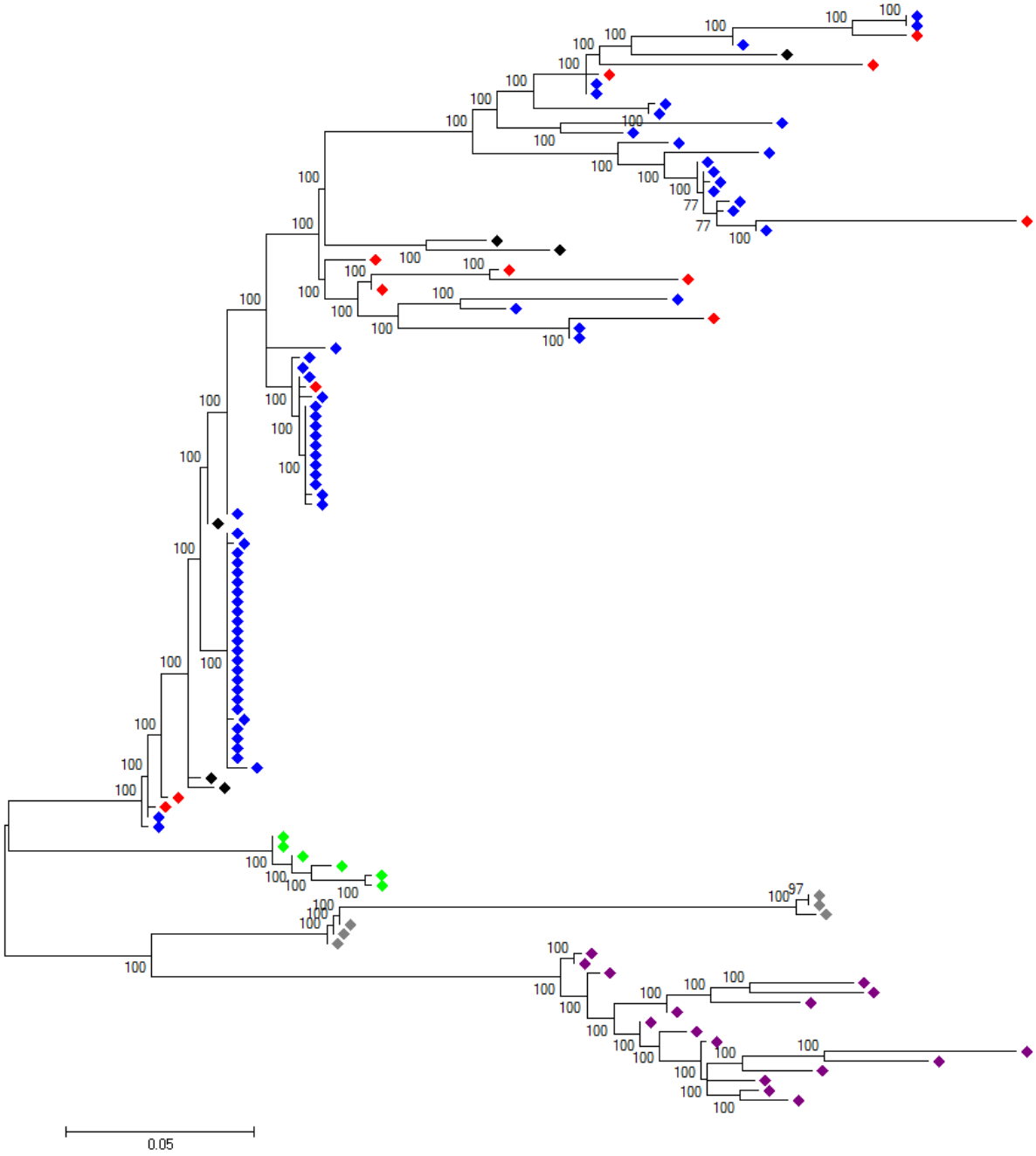


Figure S4.5. Haplotype based maximum likelihood tree for region of introgression from domestic chicken/red junglefowl haplotypes into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 102680000 to 102700000 bp on chromosome 1.

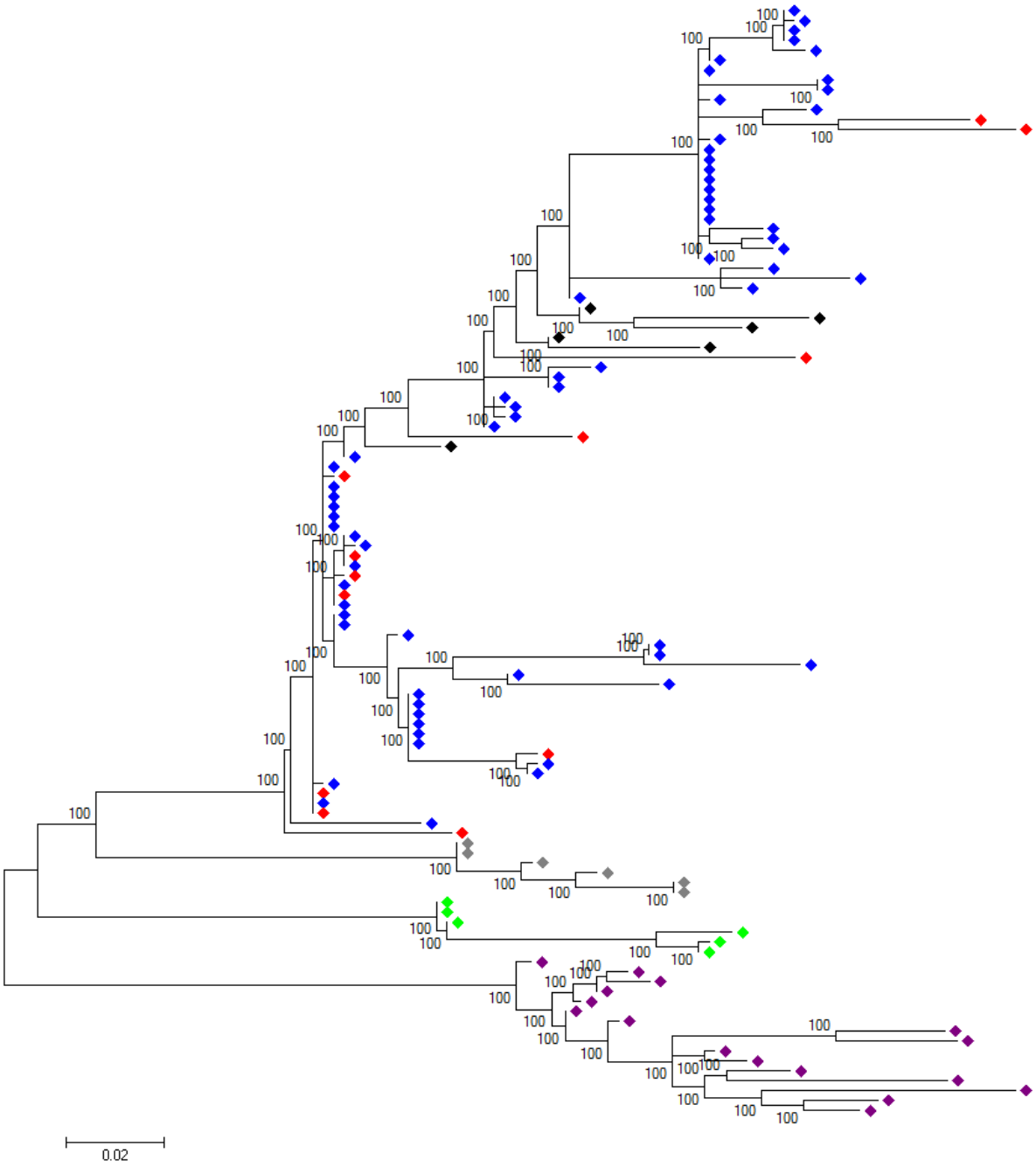


Figure S4.6. Haplotype based maximum likelihood tree for region of introgression from domestic chicken/red junglefowl haplotypes into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 102870000 to 102890000 bp on chromosome 1.

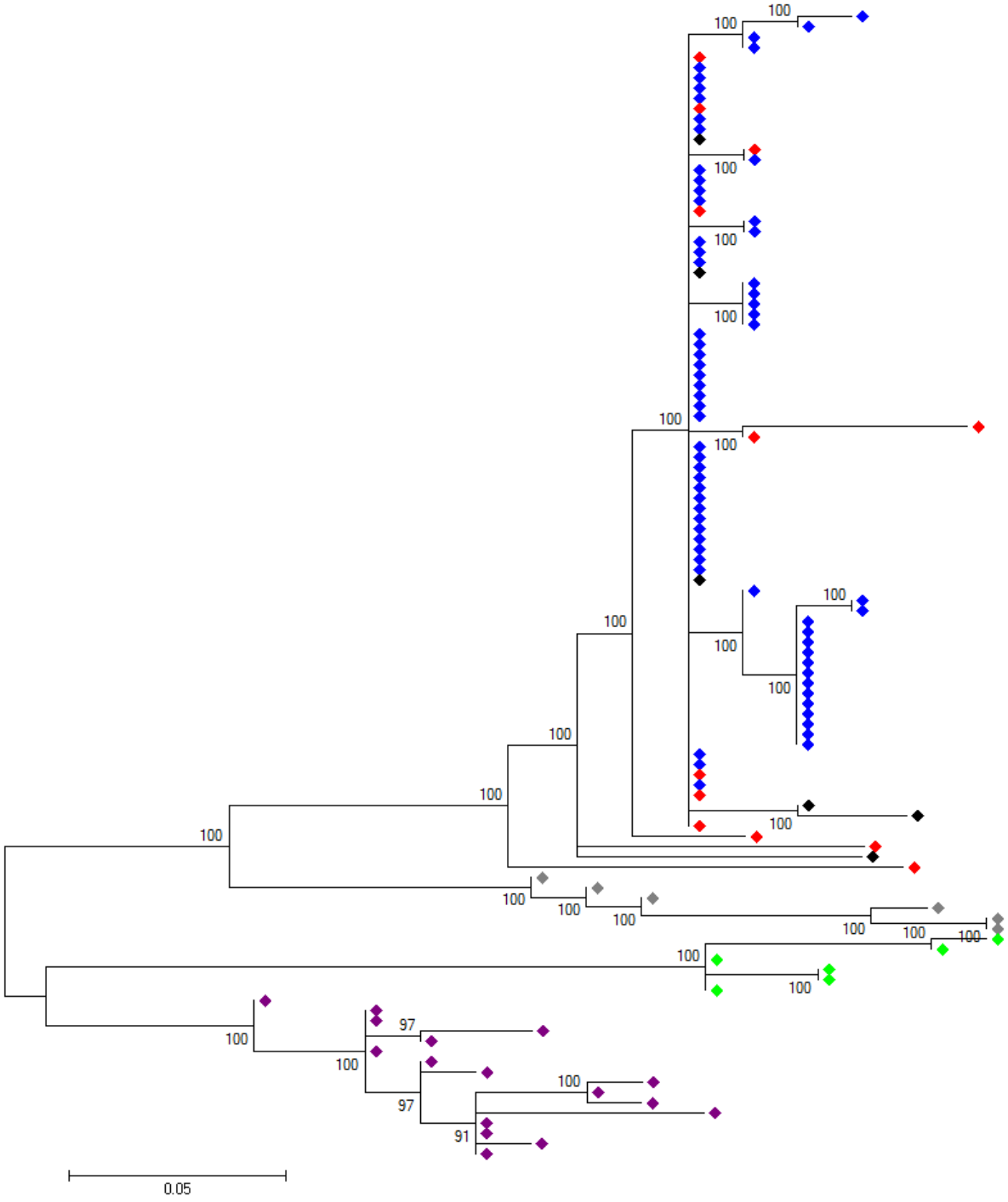


Figure S4.7. Haplotype based maximum likelihood tree for region of introgression from domestic chicken/red junglefowl haplotypes into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 122100000 to 122120000 bp on chromosome 1.

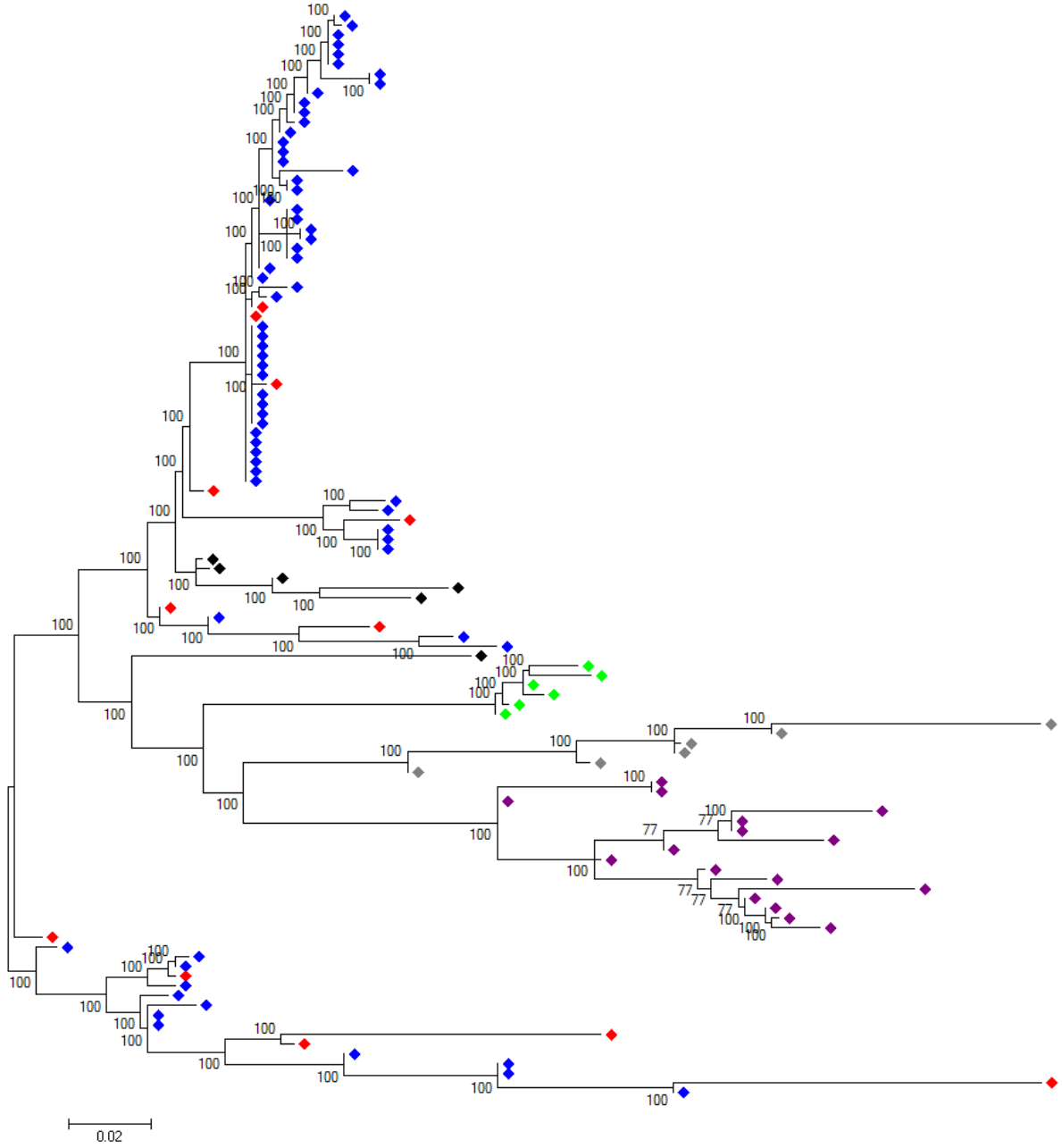


Figure S4.8. Haplotype based maximum likelihood tree for region of introgression from domestic chicken/red junglefowl haplotypes into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 25-kb fragment size that extends from 125700000 to 125725000 bp on chromosome 1.

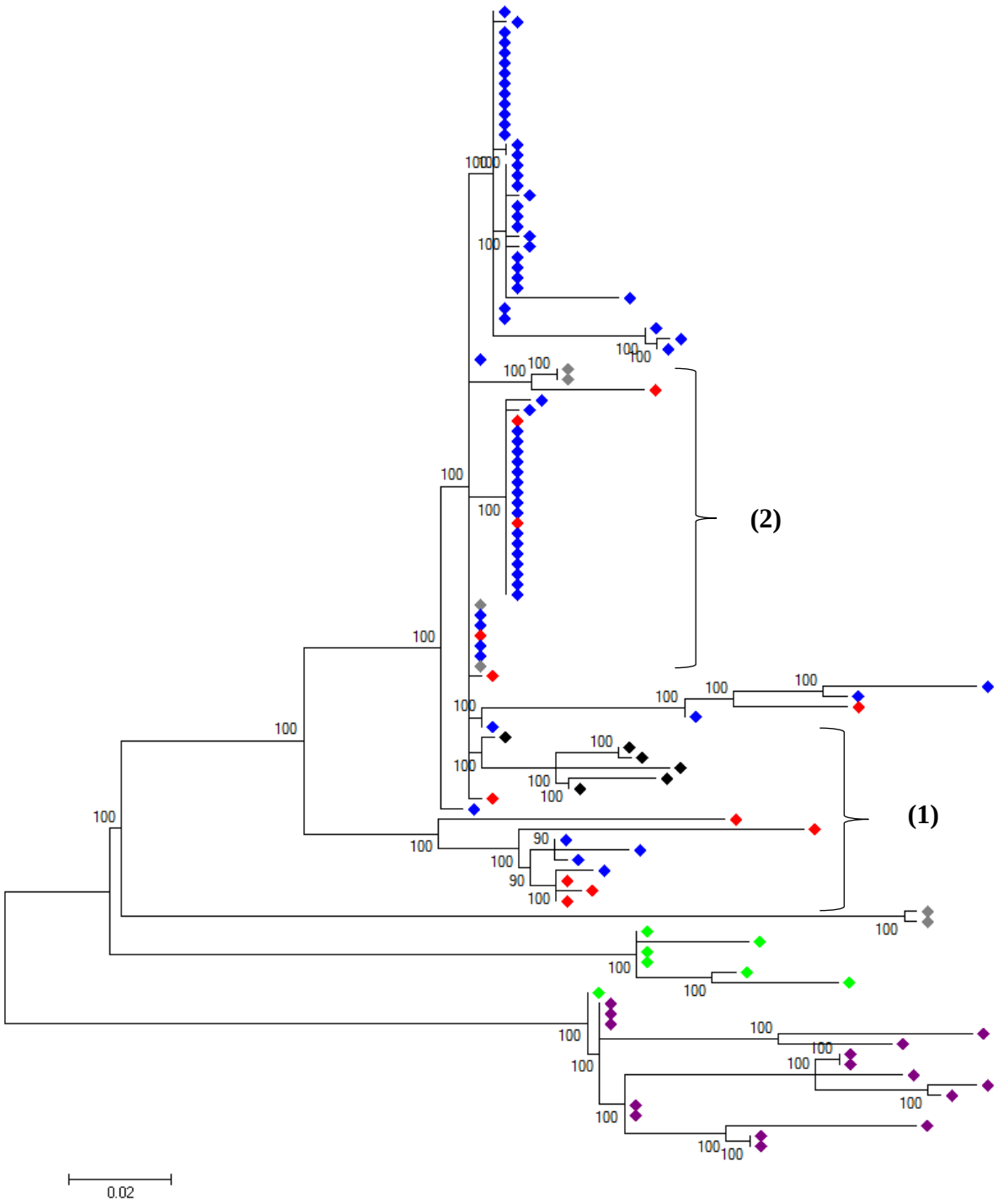


Figure S4.9. Haplotype based maximum likelihood tree for two patterns of introgression based on the 25-kb fragment size that extends from 141305000 to 141330000 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of domestic chicken/red junglefowl haplotypes into Javan red junglefowl **(2)** introgression of domestic chicken/red junglefowl haplotypes into grey junglefowl

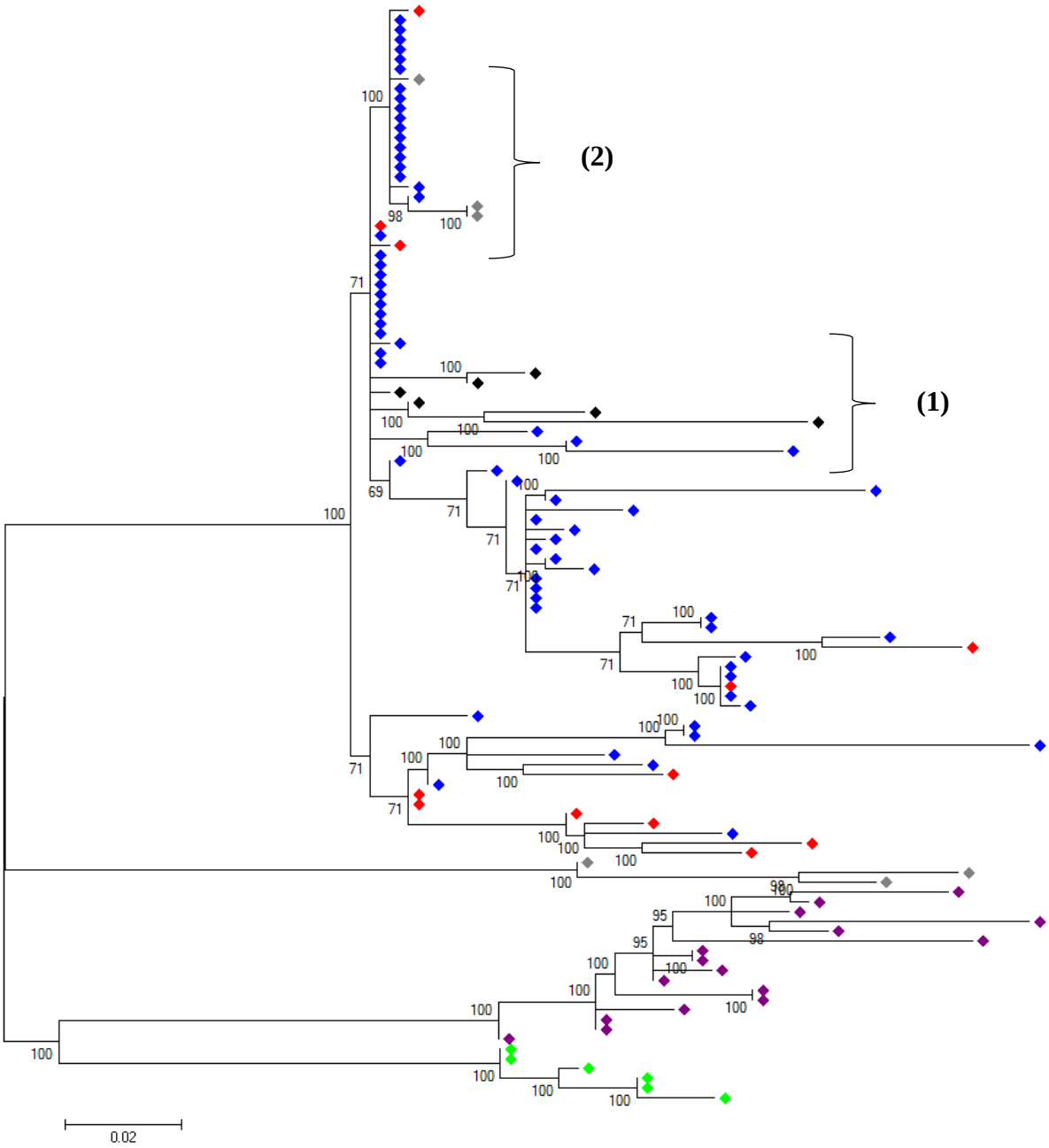


Figure S4.10. Haplotype based maximum likelihood tree for two patterns of introgression based on the 20-kb fragment size that extends from 149755000 to 149775000 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of domestic chicken/red junglefowl haplotypes into Javan red junglefowl **(2)** introgression of domestic chicken/red junglefowl haplotypes into grey junglefowl

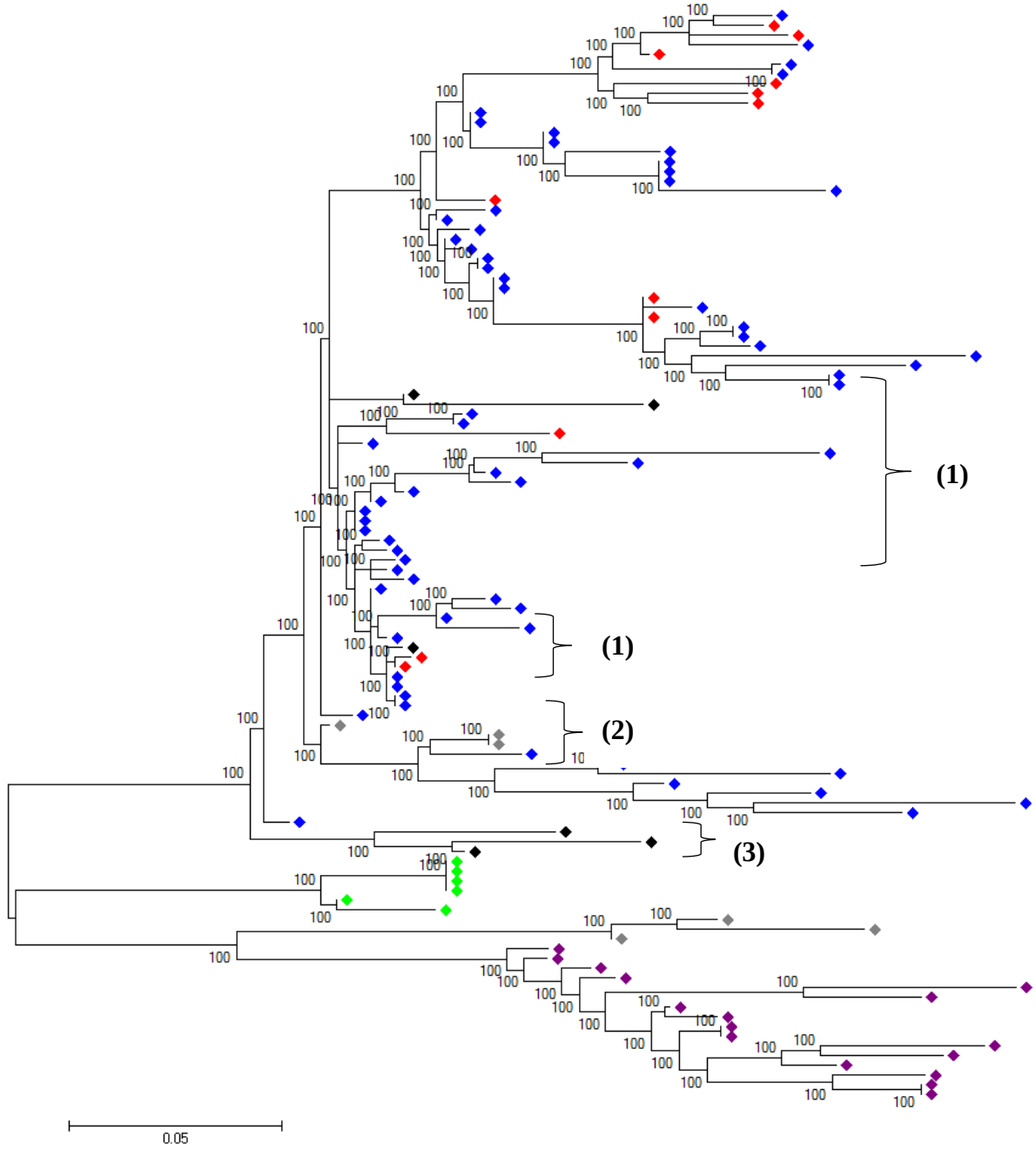


Figure S4.11. Haplotype based maximum likelihood tree for two patterns of introgression based on the 25-kb fragment size that extends from 152125000 to 152150000 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of domestic chicken/red junglefowl haplotypes into Javan red junglefowl **(2)** Introgression of domestic chicken haplotypes into grey junglefowl **(3)** three Javan red junglefowl haplotypes showing likely shared ancestry with domestic chicken/red junglefowl haplotypes

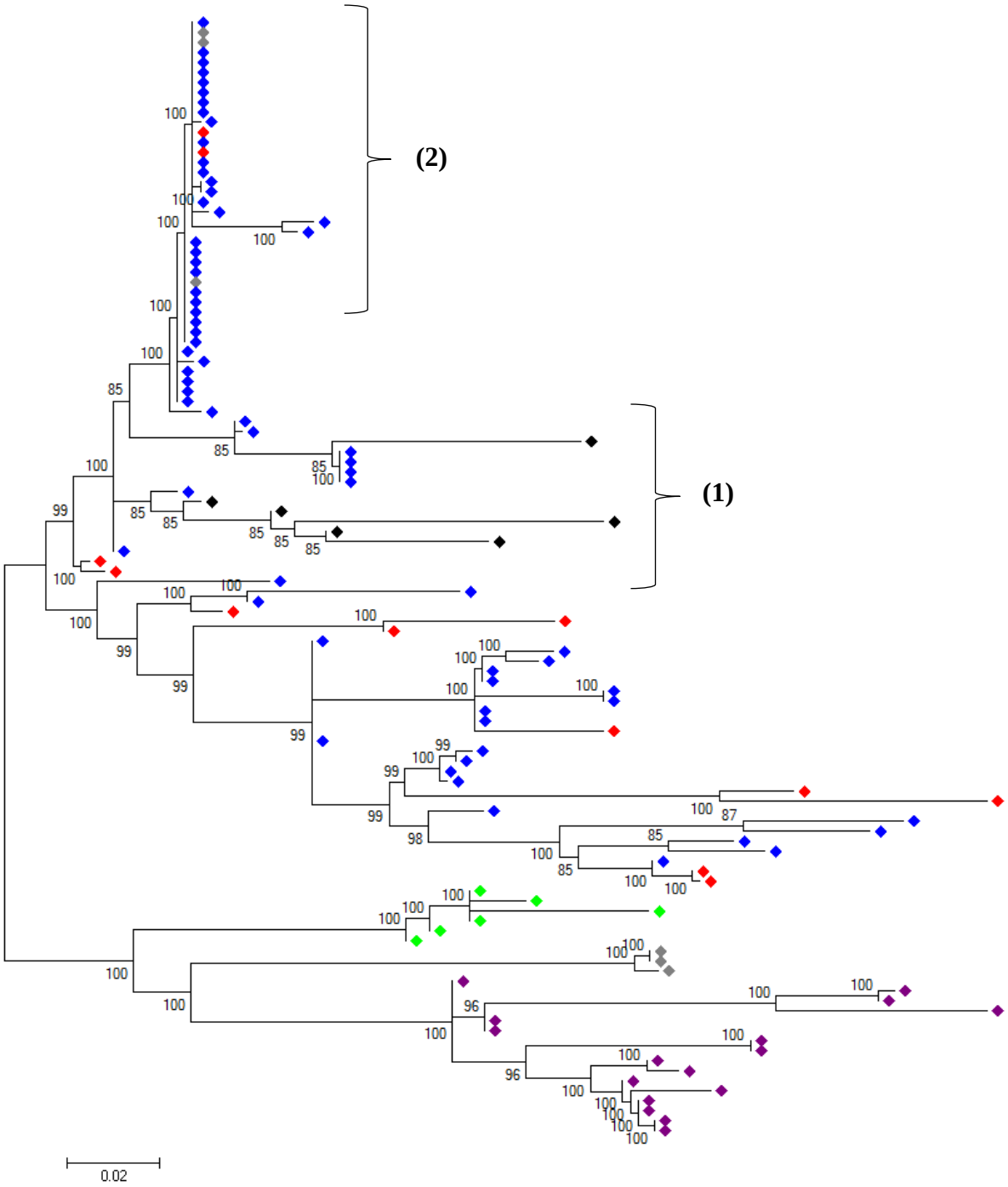


Figure S4.12. Haplotype based maximum likelihood tree for two patterns of introgression based on the 30-kb fragment size that extends from 156380000 to 156410000 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of domestic chicken/red junglefowl haplotypes into Javan red junglefowl **(2)** introgression of domestic chicken/red junglefowl haplotypes into grey junglefowl.

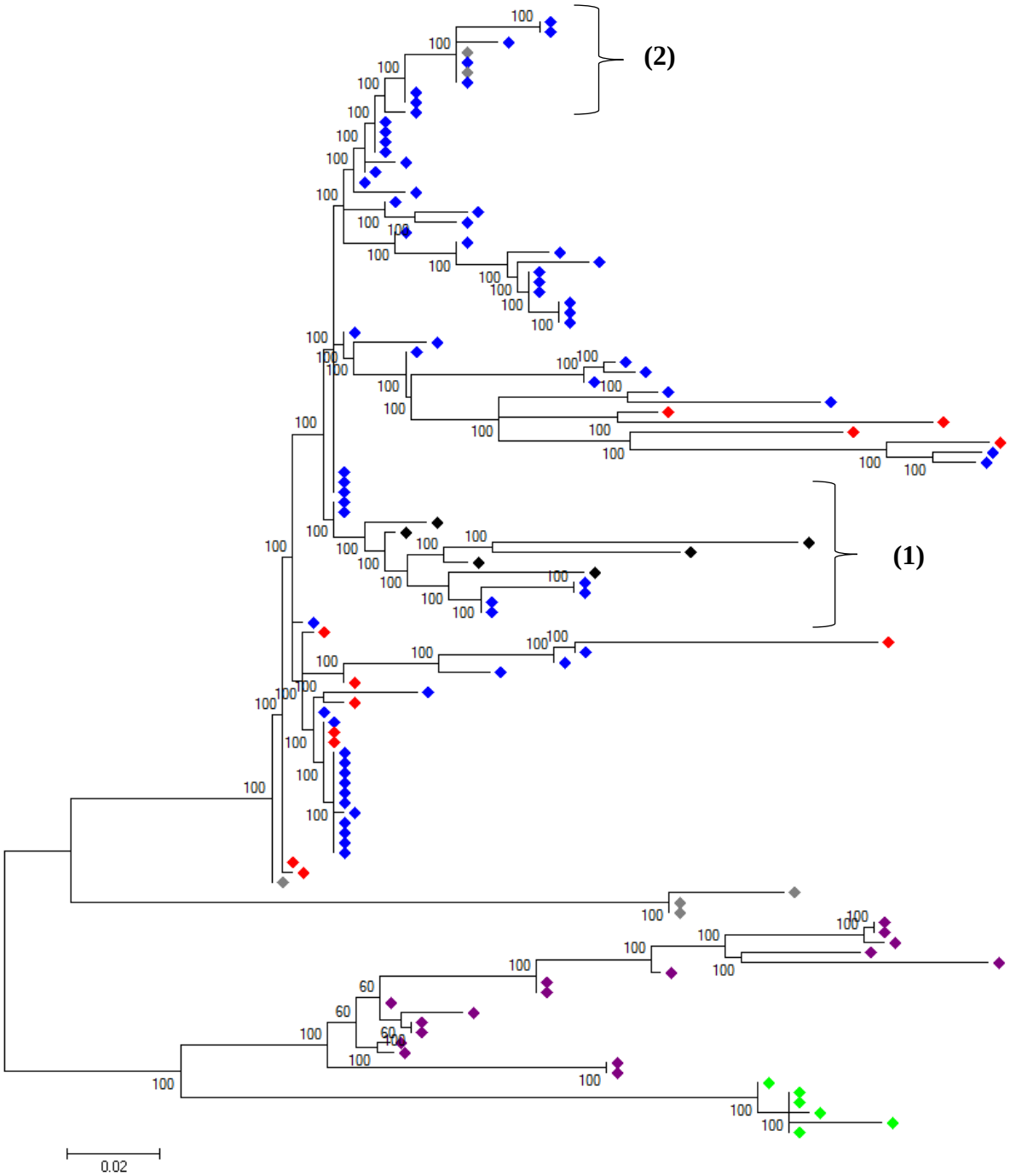


Figure S4.13. Haplotype based maximum likelihood tree for two patterns of introgression based on the 25-kb fragment size that extends from 160905000 to 160930000 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of domestic chicken/red junglefowl haplotypes into Javan red junglefowl **(2)** introgression of domestic chicken/red junglefowl haplotypes into grey junglefowl

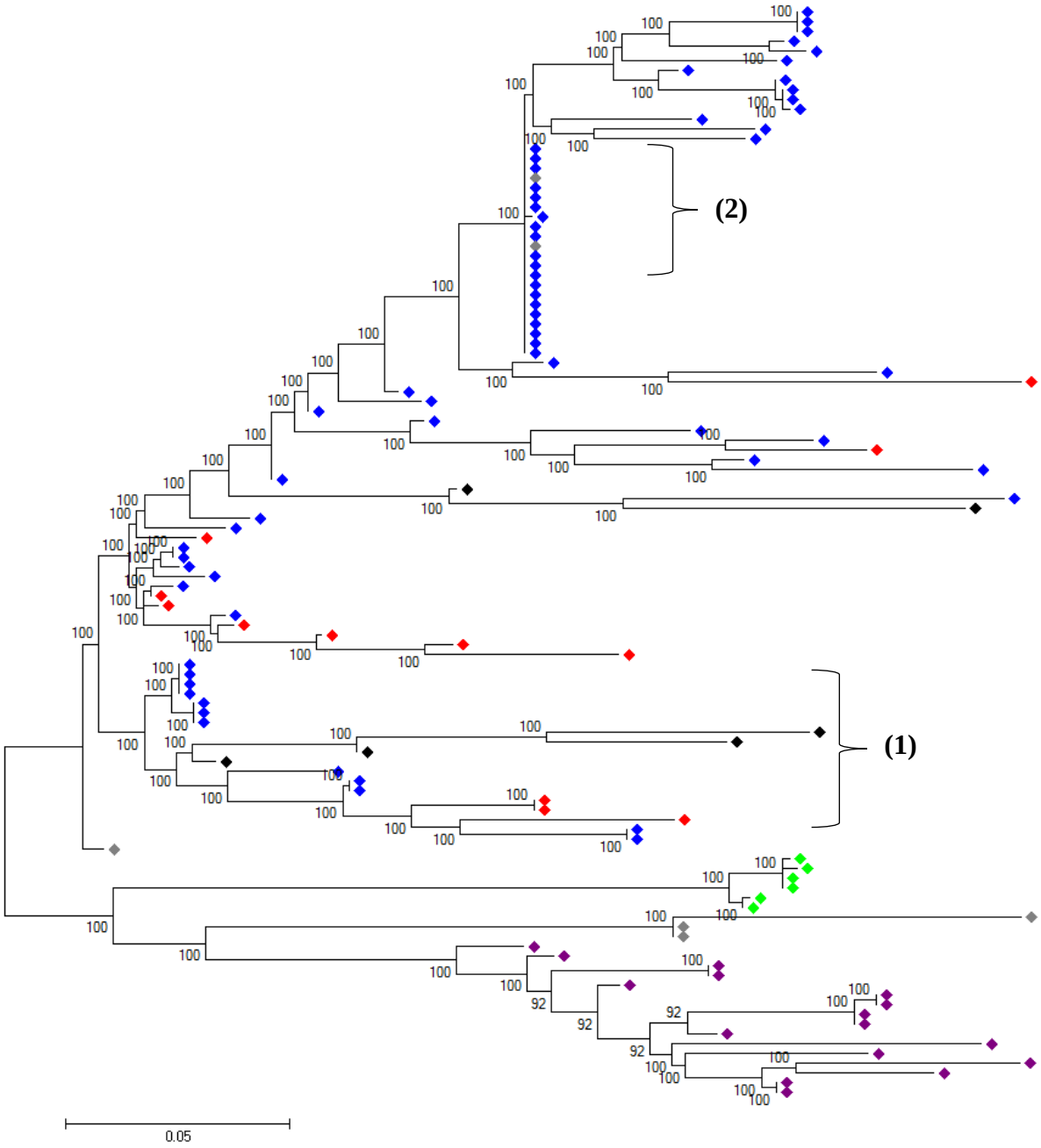


Figure S4.14. Haplotype based maximum likelihood tree for two patterns of introgression based on the 25-kb fragment size that extends from 161815000 to 161840000 bp on chromosome 1. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of domestic chicken haplotypes into Javan red junglefowl **(2)** introgression of domestic chicken/red junglefowl haplotypes into grey junglefowl.

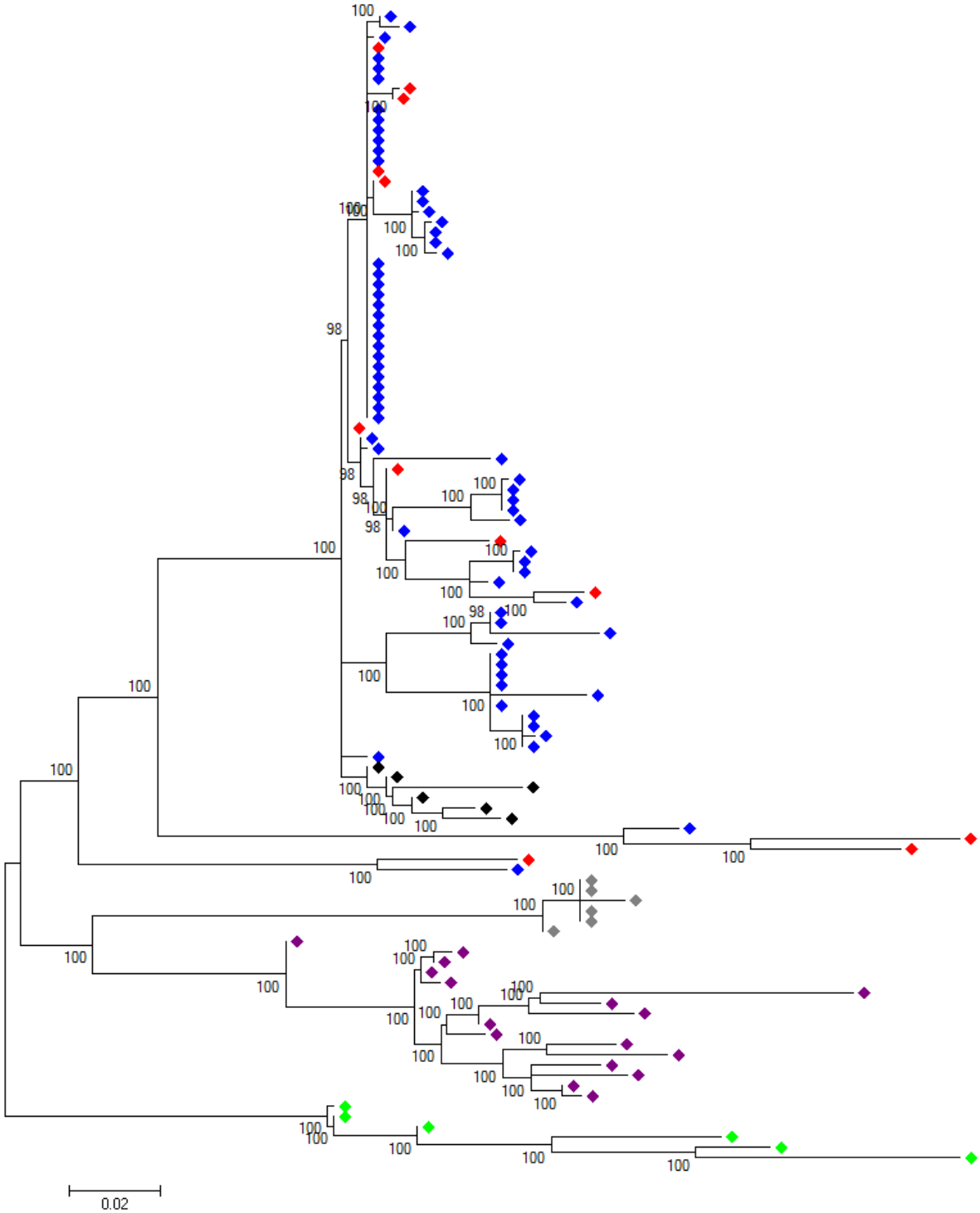


Figure S4.15. Haplotype based maximum likelihood tree for region of introgression from domestic chicken into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 30-kb fragment size that extends from 54905000 to 54935000 bp on chromosome 2.

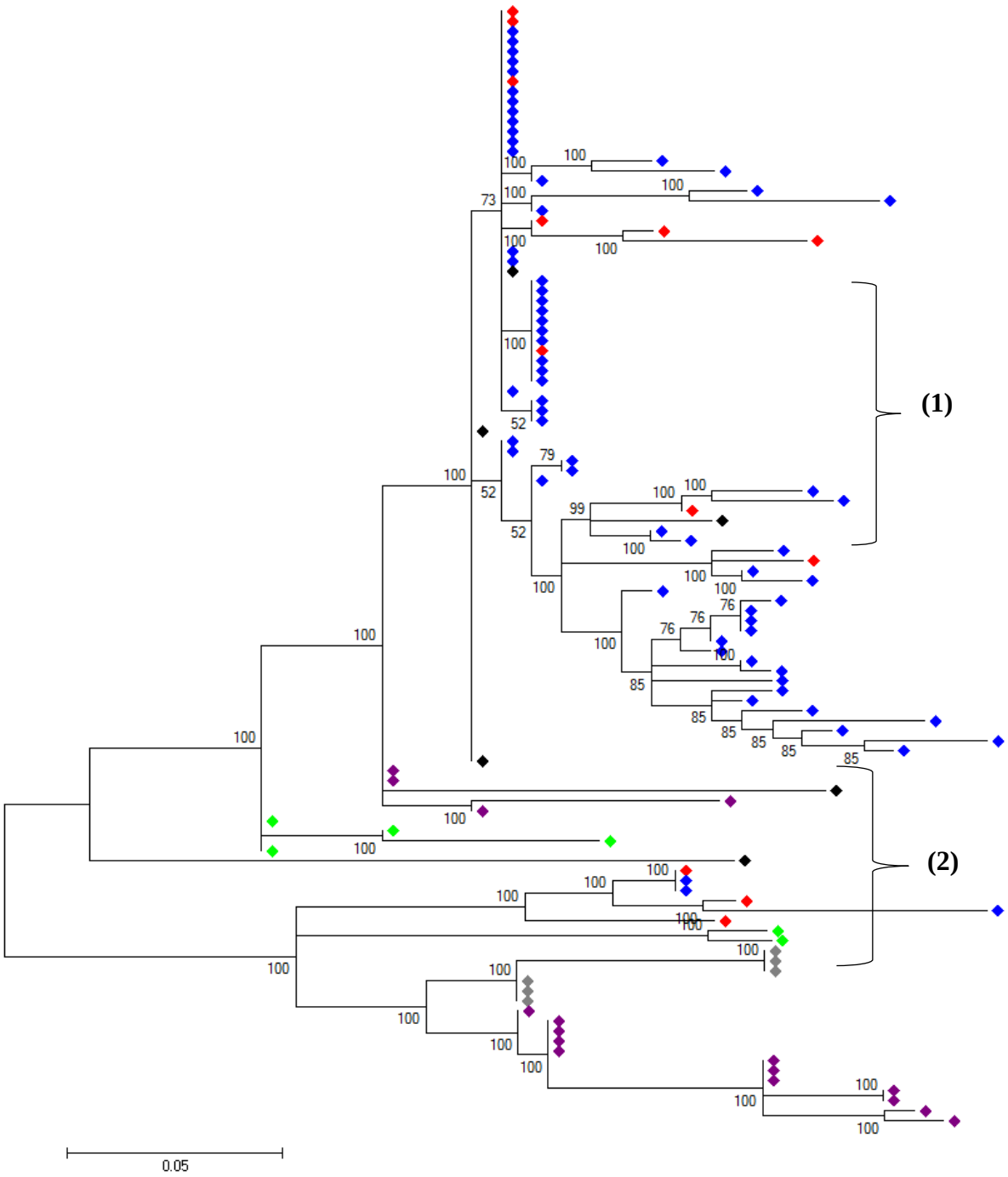


Figure S4.16. Haplotype based maximum likelihood tree for region of introgression **(1)** from domestic chicken into Javan red junglefowl **(2)** showing a likely ambiguous relationship between red junglefowl/domestic chickens, Javan red junglefowl, Ceylon and green junglefowls. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 59610000 to 59630000 bp on chromosome 2.

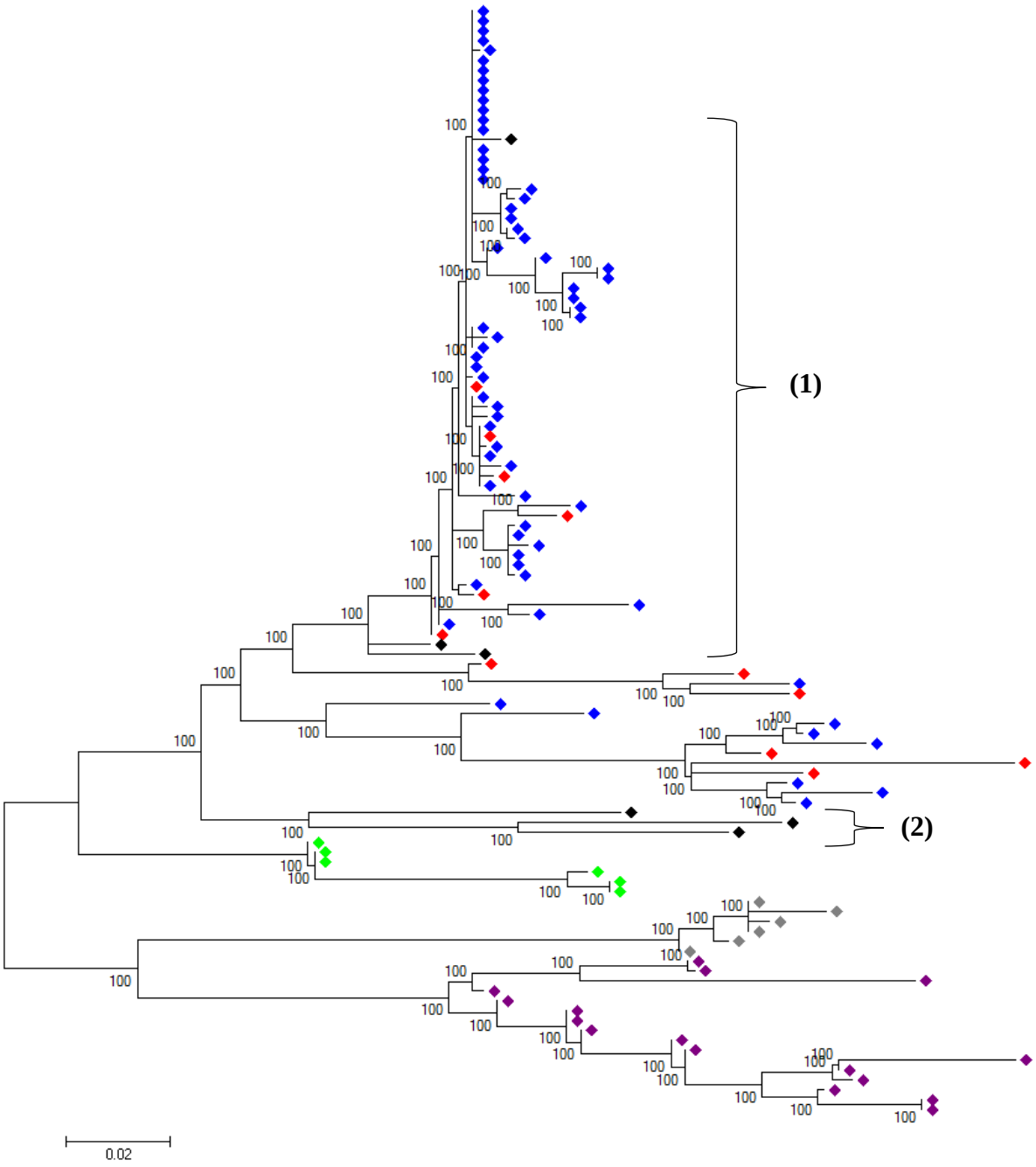


Figure S4.17. Haplotype based maximum likelihood tree for region of introgression **(1)** from domestic chicken into Javan red junglefowl **(2)** showing a likely shared ancestry or an ambiguous relationship between red junglefowl/domestic chickens. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 30-kb fragment size that extends from 95680000 to 95710000 bp on chromosome 2.

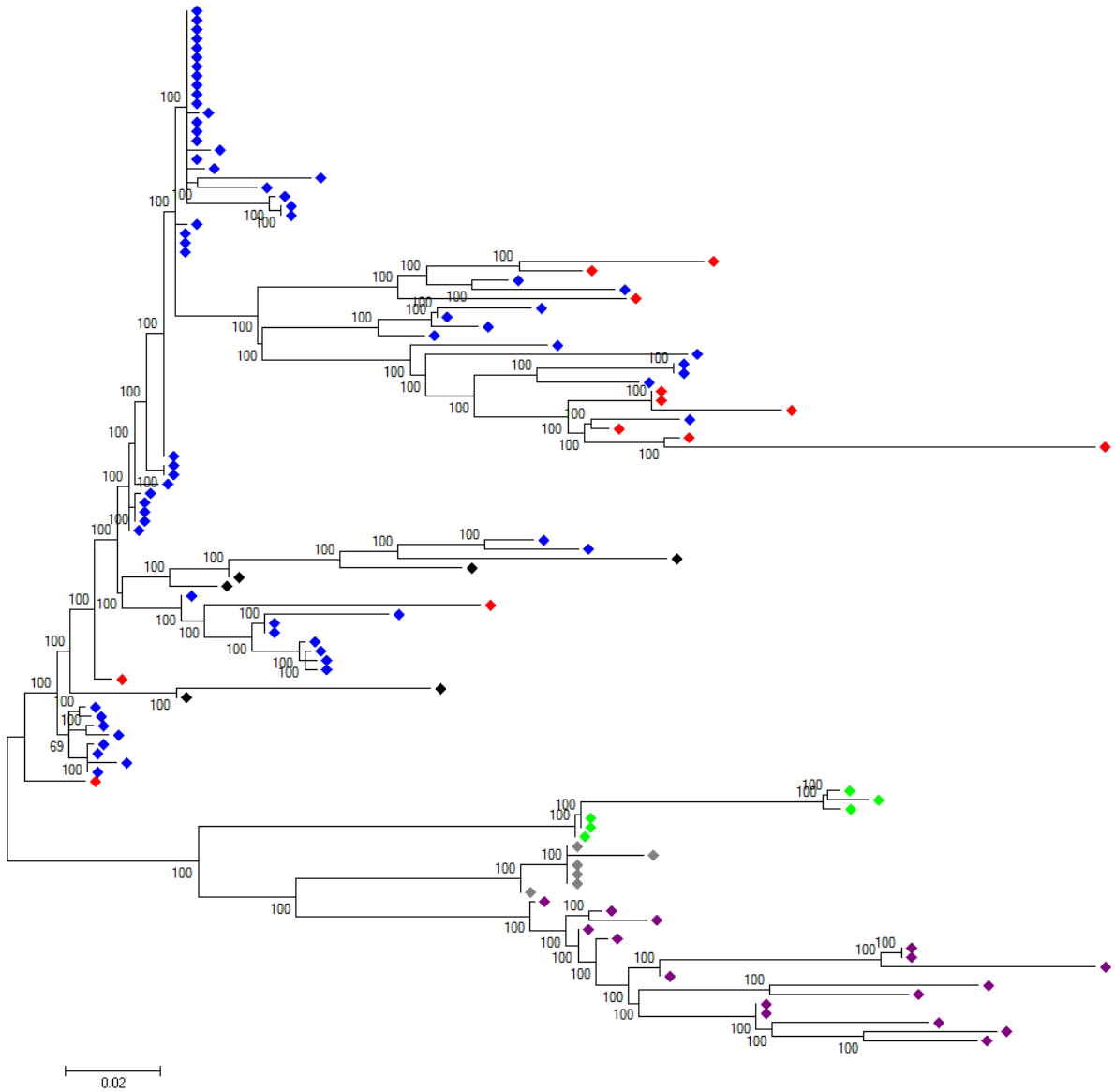


Figure S4.18. Haplotype based maximum likelihood tree for region of introgression from domestic chicken into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 35-kb fragment size that extends from 121930000 to 121965000 bp on chromosome 2.

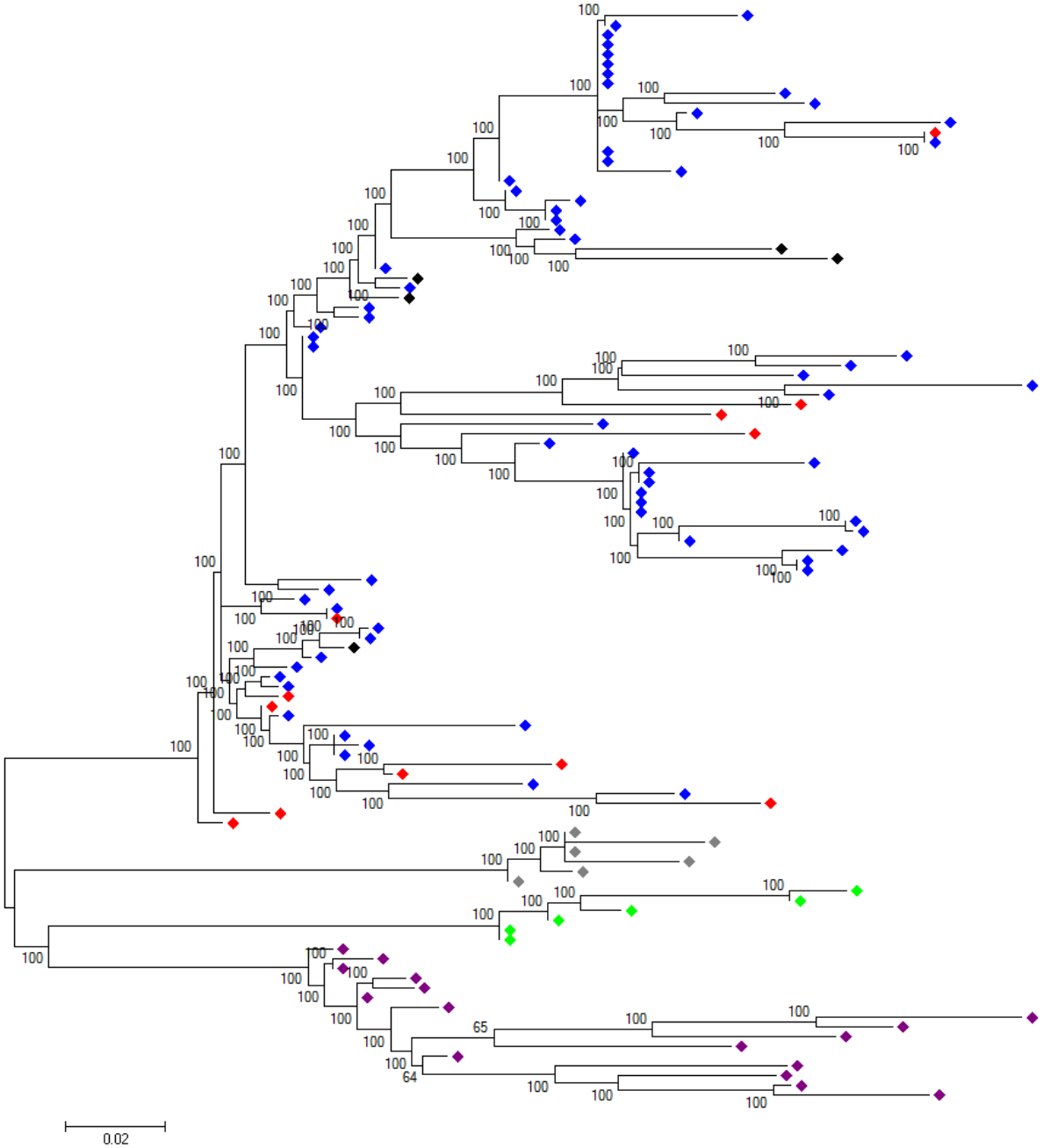


Figure S4.19. Haplotype based maximum likelihood tree for region of introgression from domestic chicken into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 32160000 to 32180000 bp on chromosome 3.

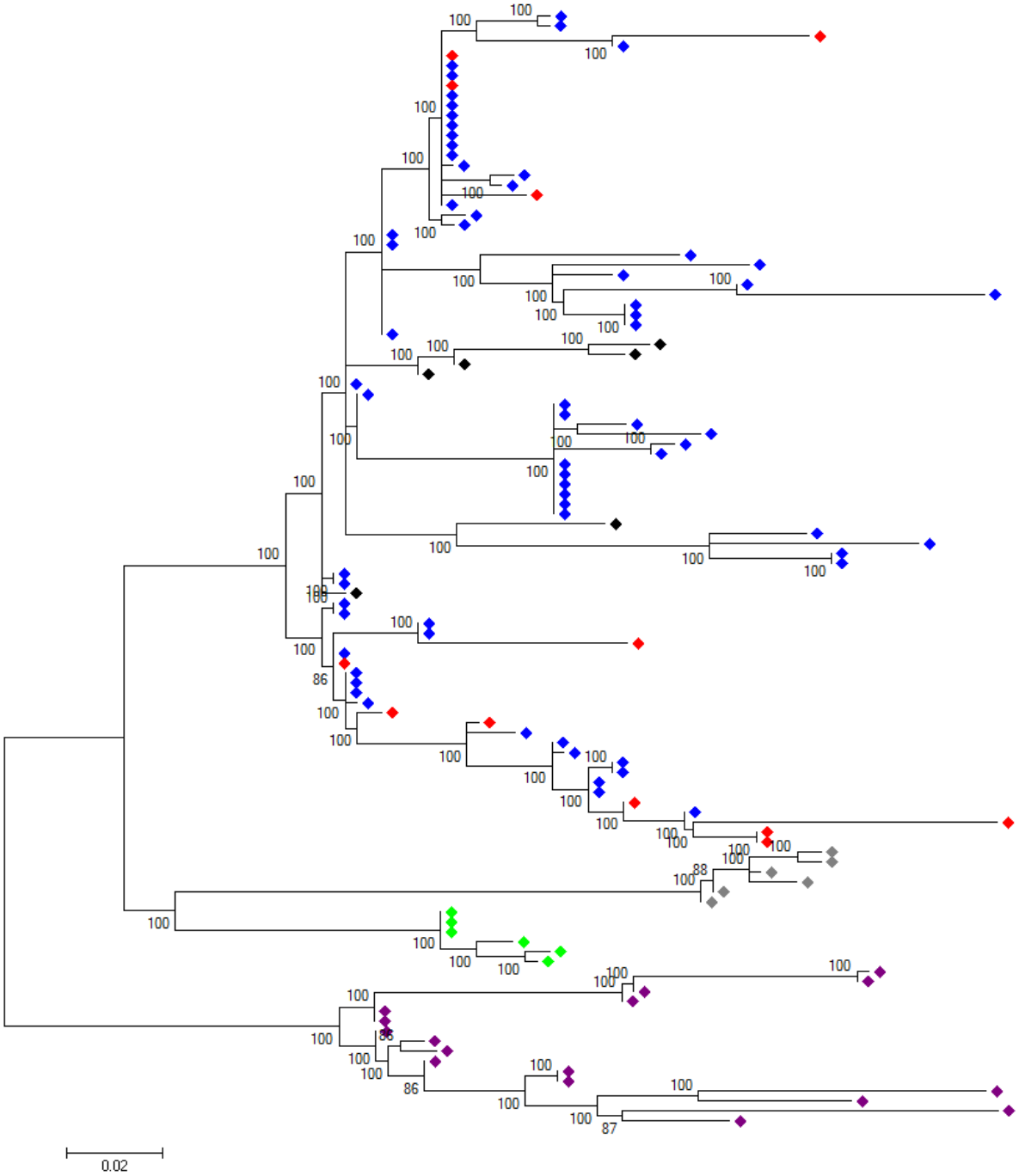


Figure S4.20. Haplotype based maximum likelihood tree for region of introgression from domestic chicken into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 89105000 to 89125000 bp on chromosome 3.

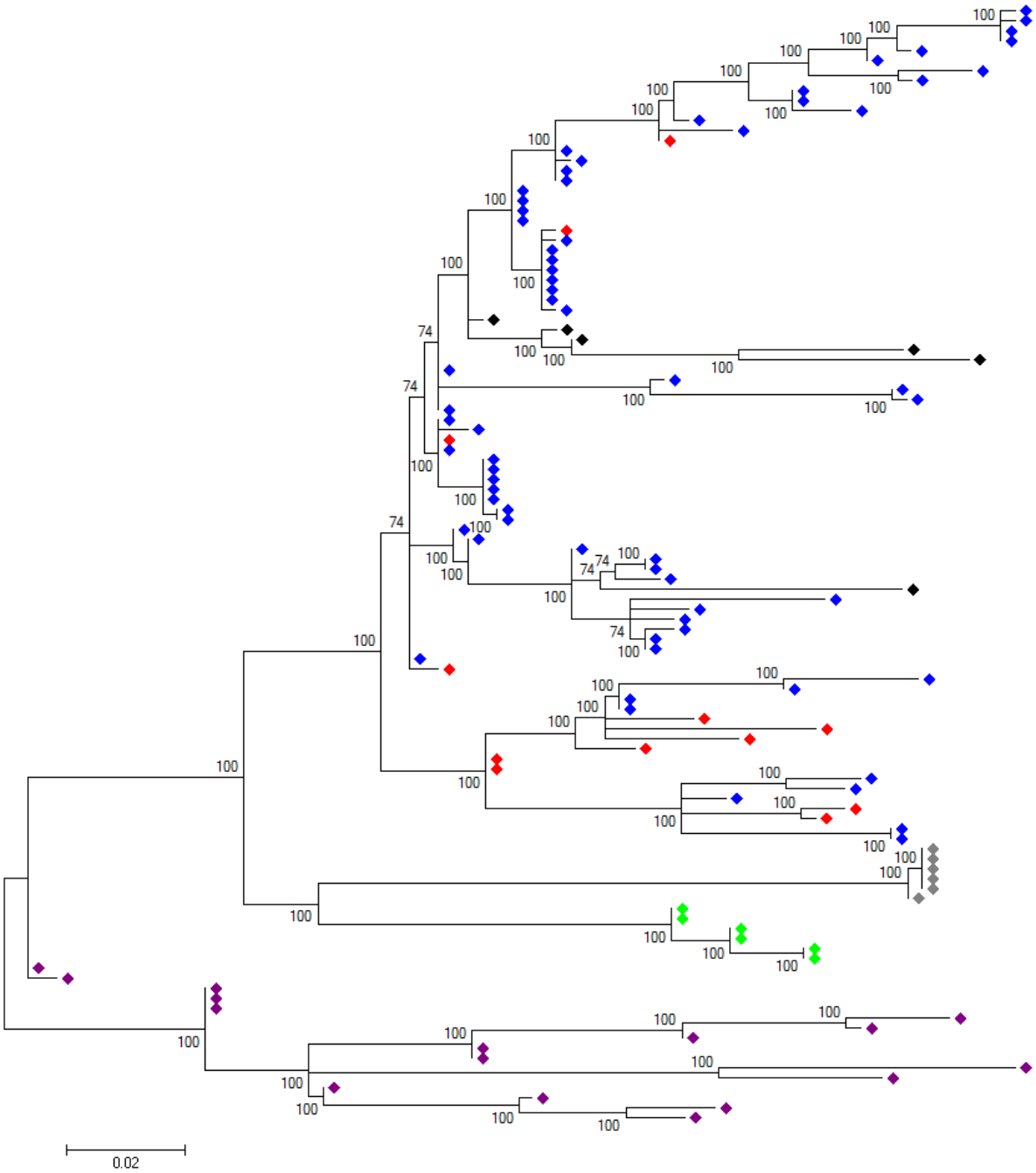


Figure S4.21. Haplotype based maximum likelihood tree for region of introgression from domestic chicken into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 97745000 to 97765000 bp on chromosome 3.

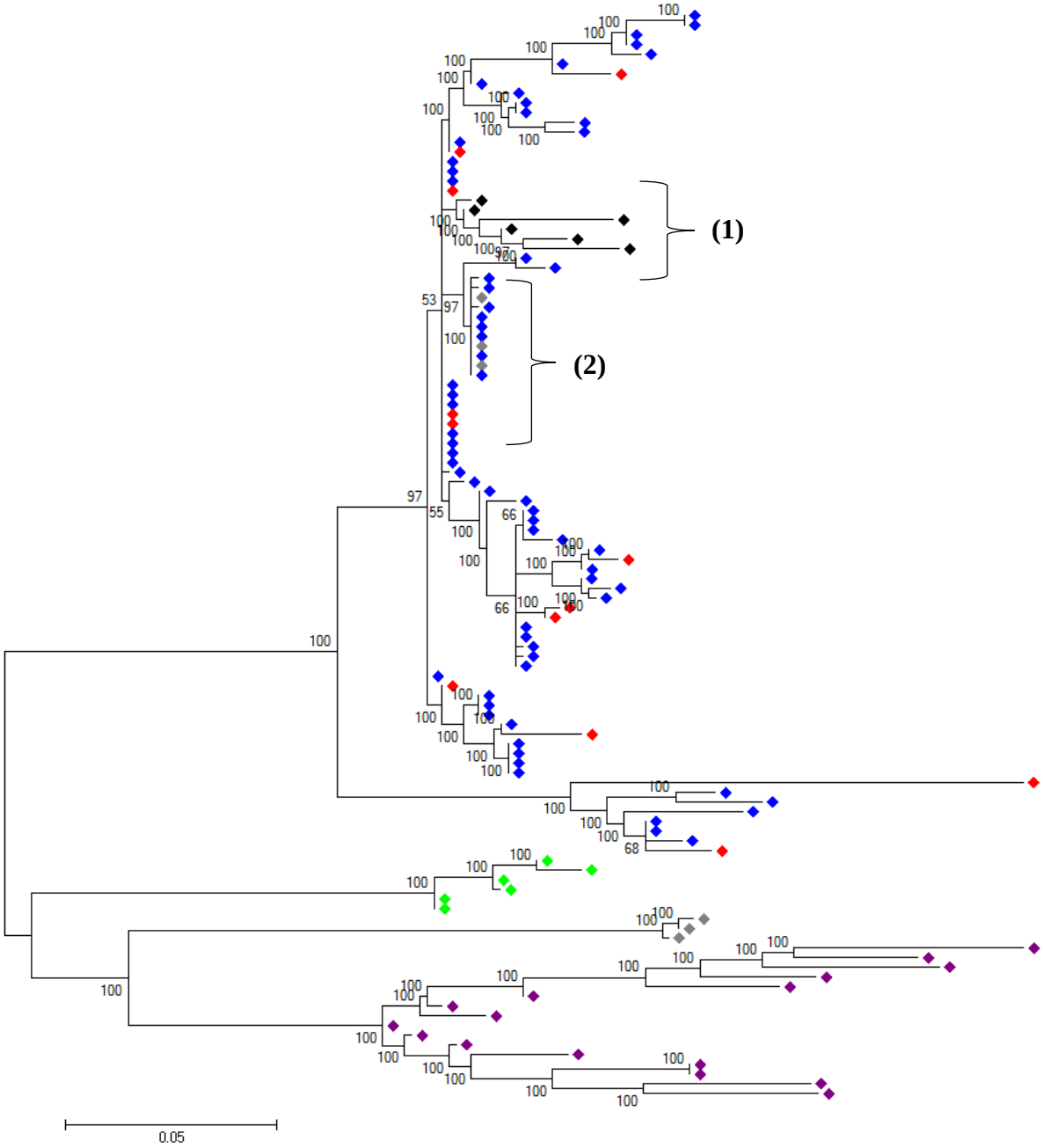


Figure S4.22. Haplotype based maximum likelihood tree for two patterns of introgression based on the 25-kb fragment size that extends from 20160000 to 20185000 bp on chromosome 4. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl, ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. **(1)** Introgression of domestic chicken/red junglefowl haplotypes into Javan red junglefowl **(2)** introgression of domestic chicken/red junglefowl haplotypes into grey junglefowl.

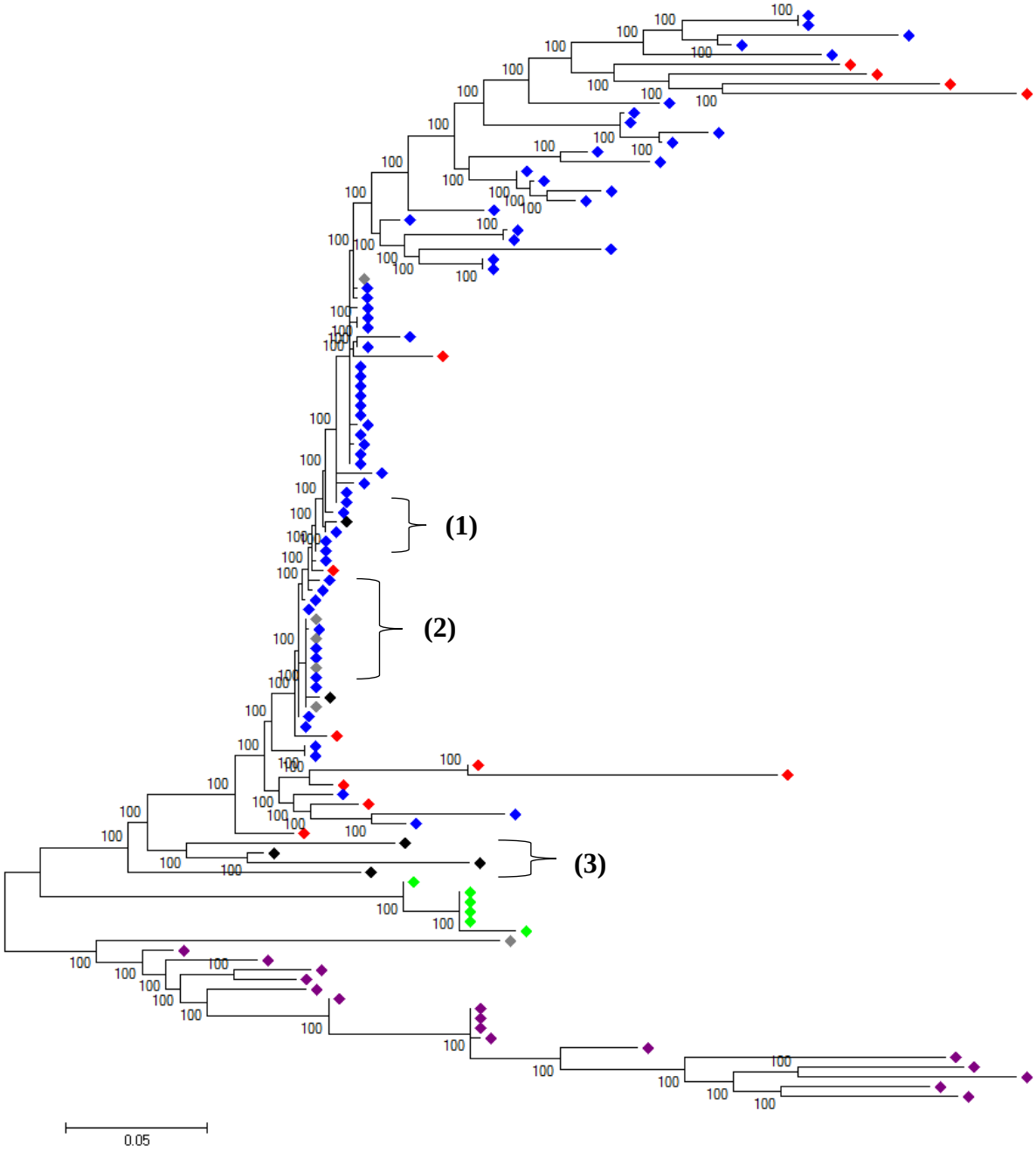


Figure S4.23. Haplotype based maximum likelihood tree for region of introgression from **(1)** domestic chicken/red junglefowl haplotypes into Javan red junglefowl **(2)** domestic chicken into grey junglefowl **(3)** a likely shared ancestry or an ambiguous relationship between red junglefowl/domestic chickens and Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 25-kb fragment size that extends from 39245000 to 39270000 bp on chromosome 4.

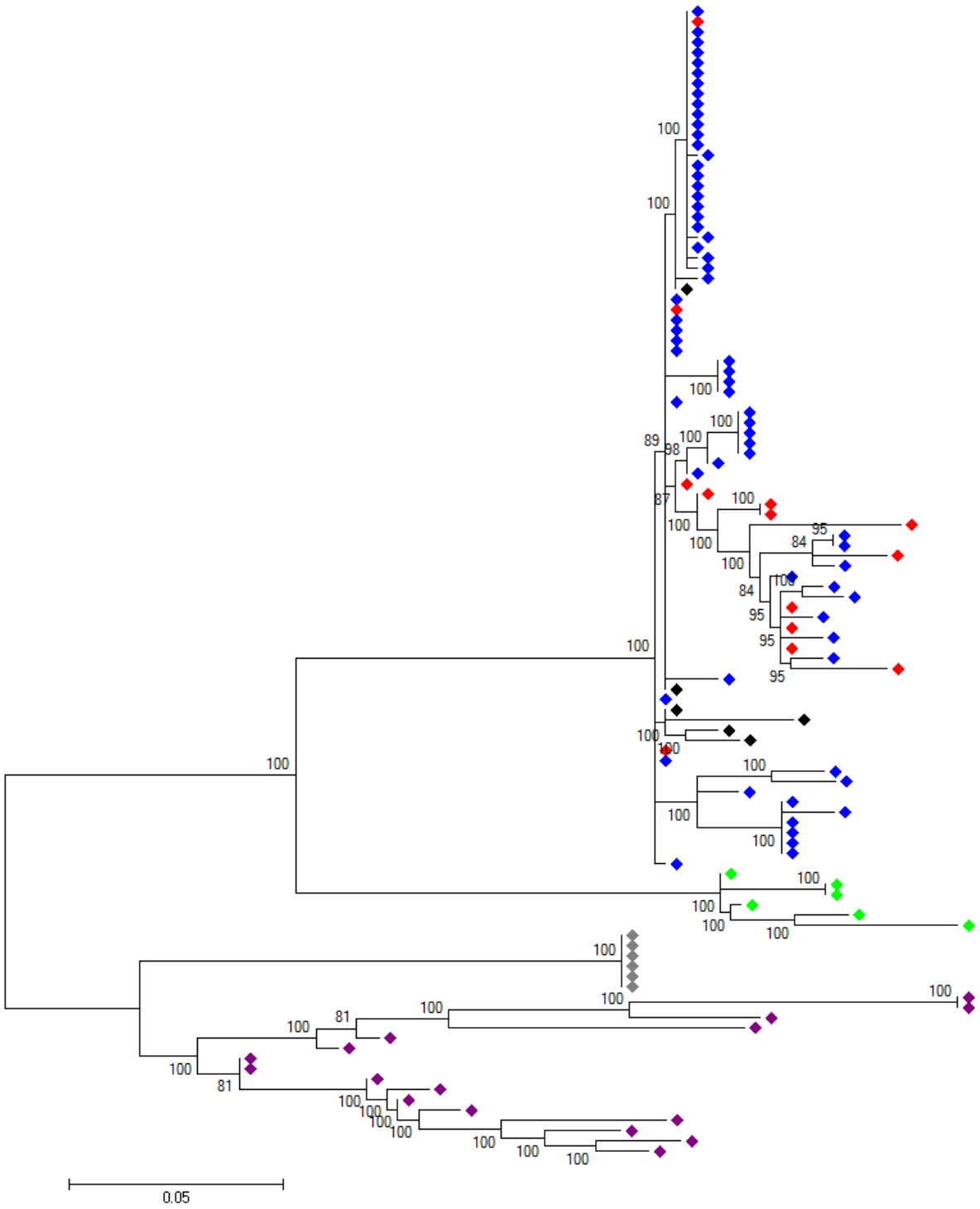


Figure S4.24. Haplotype based maximum likelihood tree for region of introgression from domestic chicken/red junglefowl haplotypes into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 30970000 to 30990000 bp on chromosome 5.

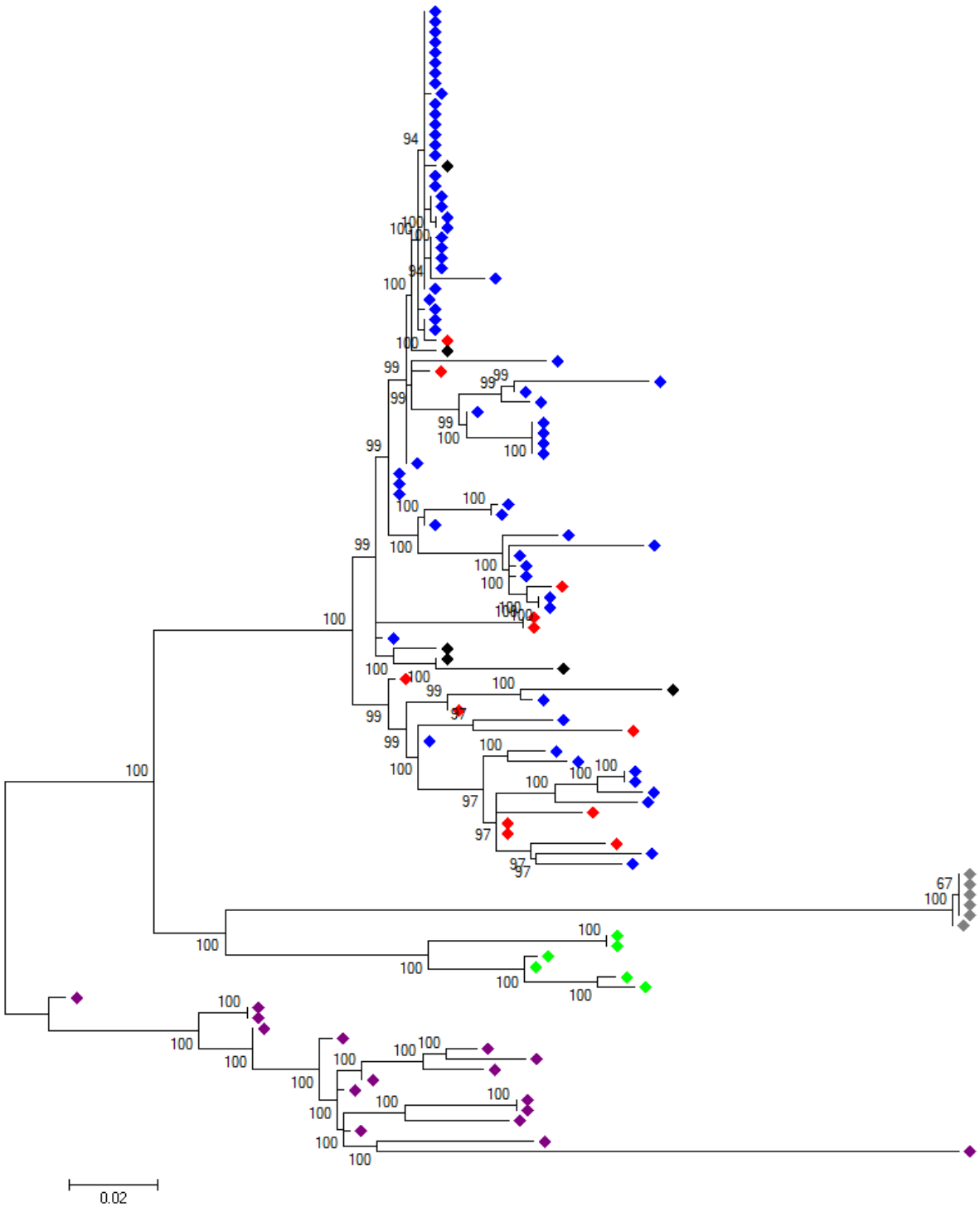


Figure S4.25. Haplotype based maximum likelihood tree for region of introgression from domestic chicken/red junglefowl haplotypes into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 40-kb fragment size that extends from 31115000 to 31155000 bp on chromosome 5.

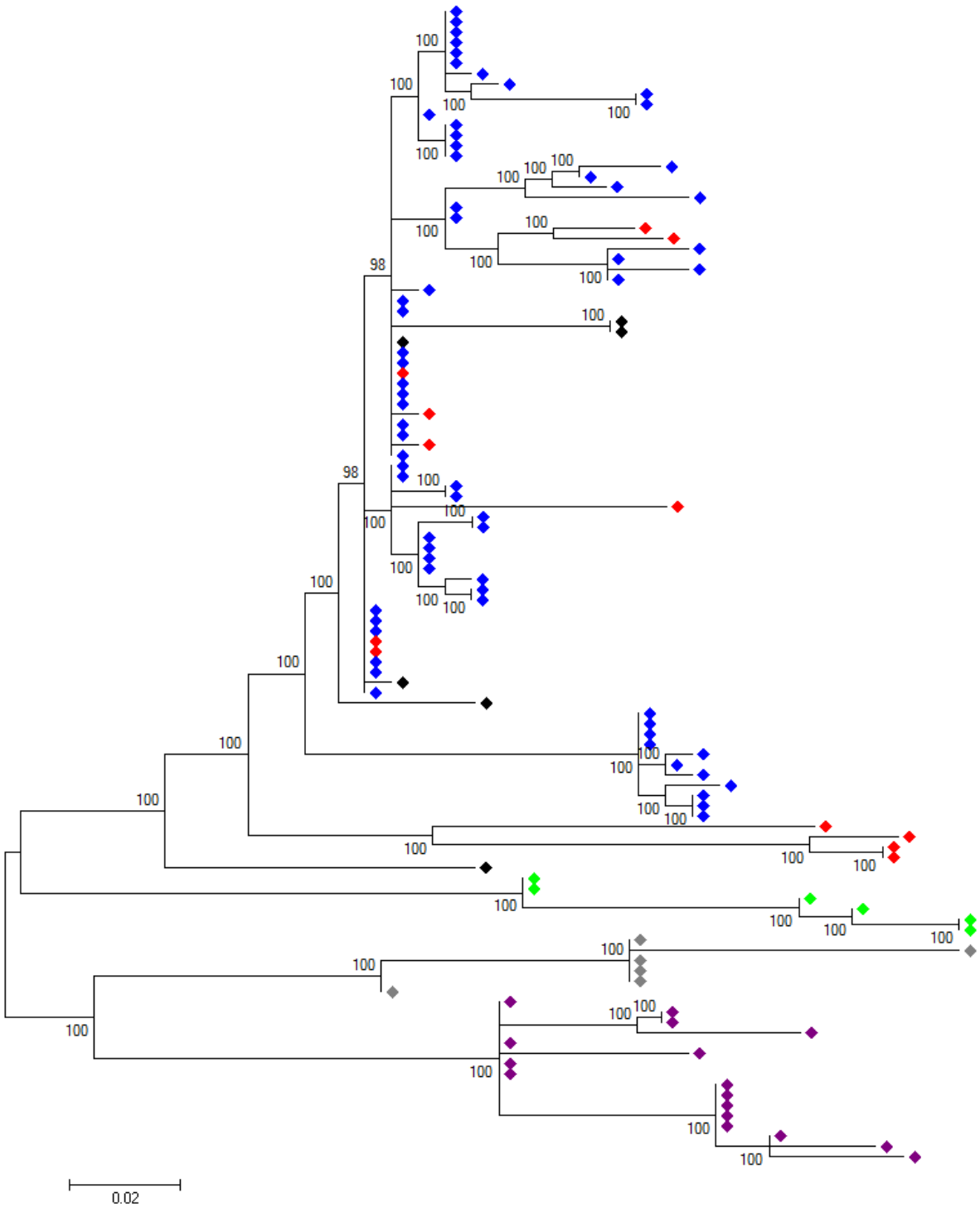


Figure S4.26. Haplotype based maximum likelihood tree for region of introgression from domestic chicken/red junglefowl into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 59380000 to 59400000 bp on chromosome 5.

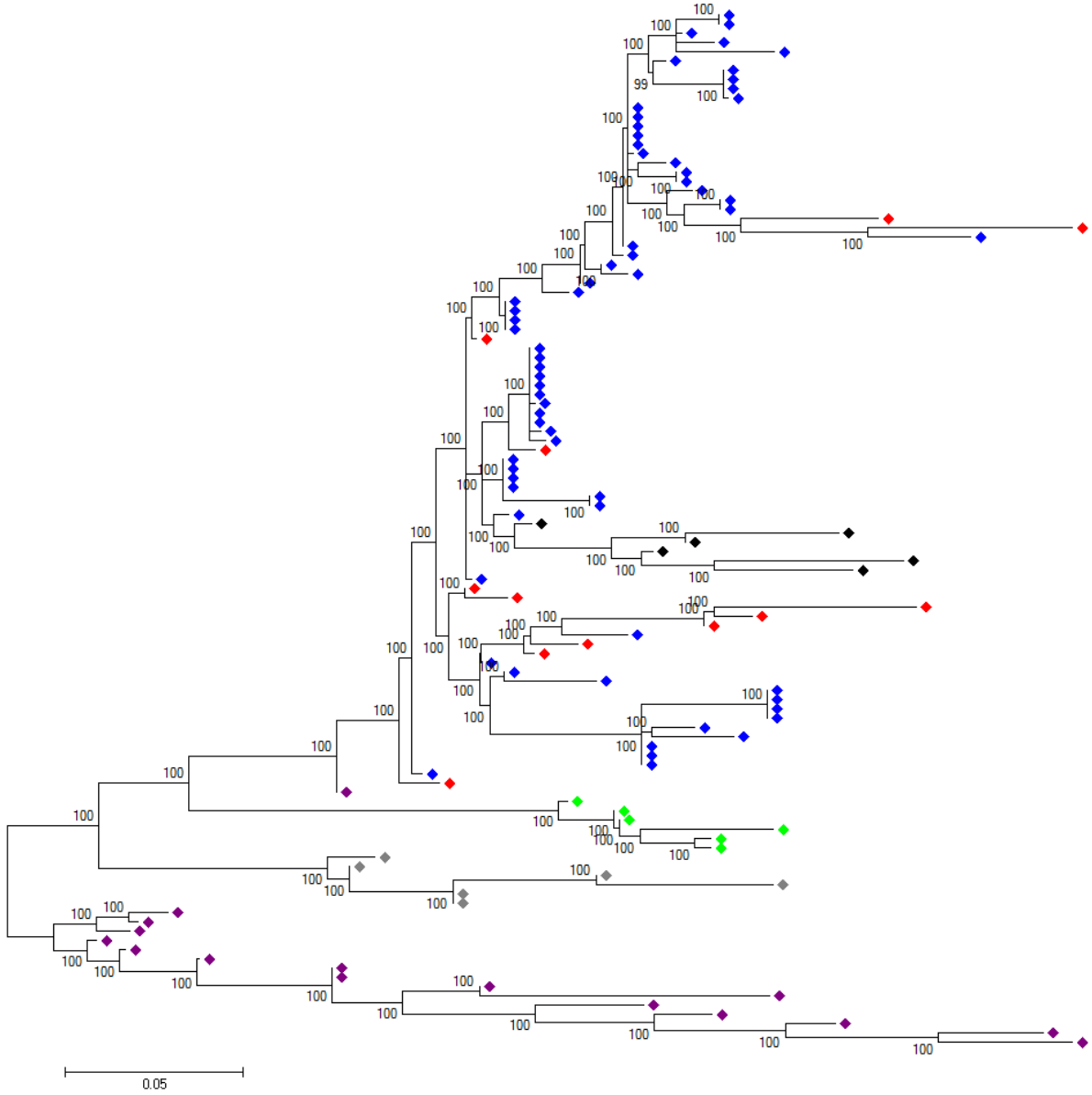


Figure S4.27. Haplotype based maximum likelihood tree for region of introgression from domestic chicken/red junglefowl into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 30-kb fragment size that extends from 6085000 to 6115000 bp on chromosome 7.

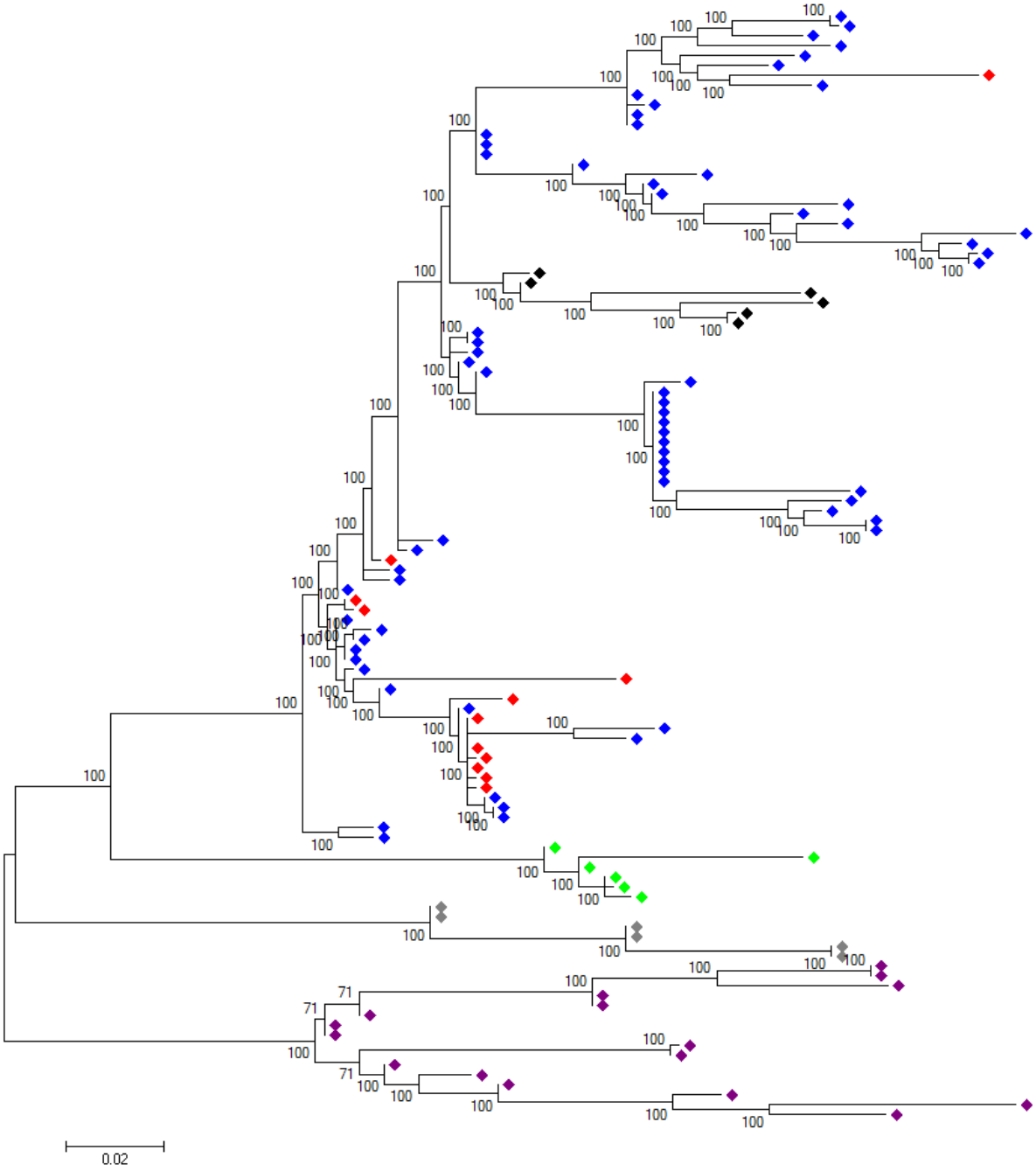


Figure S4.28. Haplotype based maximum likelihood tree for region of introgression from domestic chicken/red junglefowl into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 30-kb fragment size that extends from 19020000 to 19050000 bp on chromosome 7.

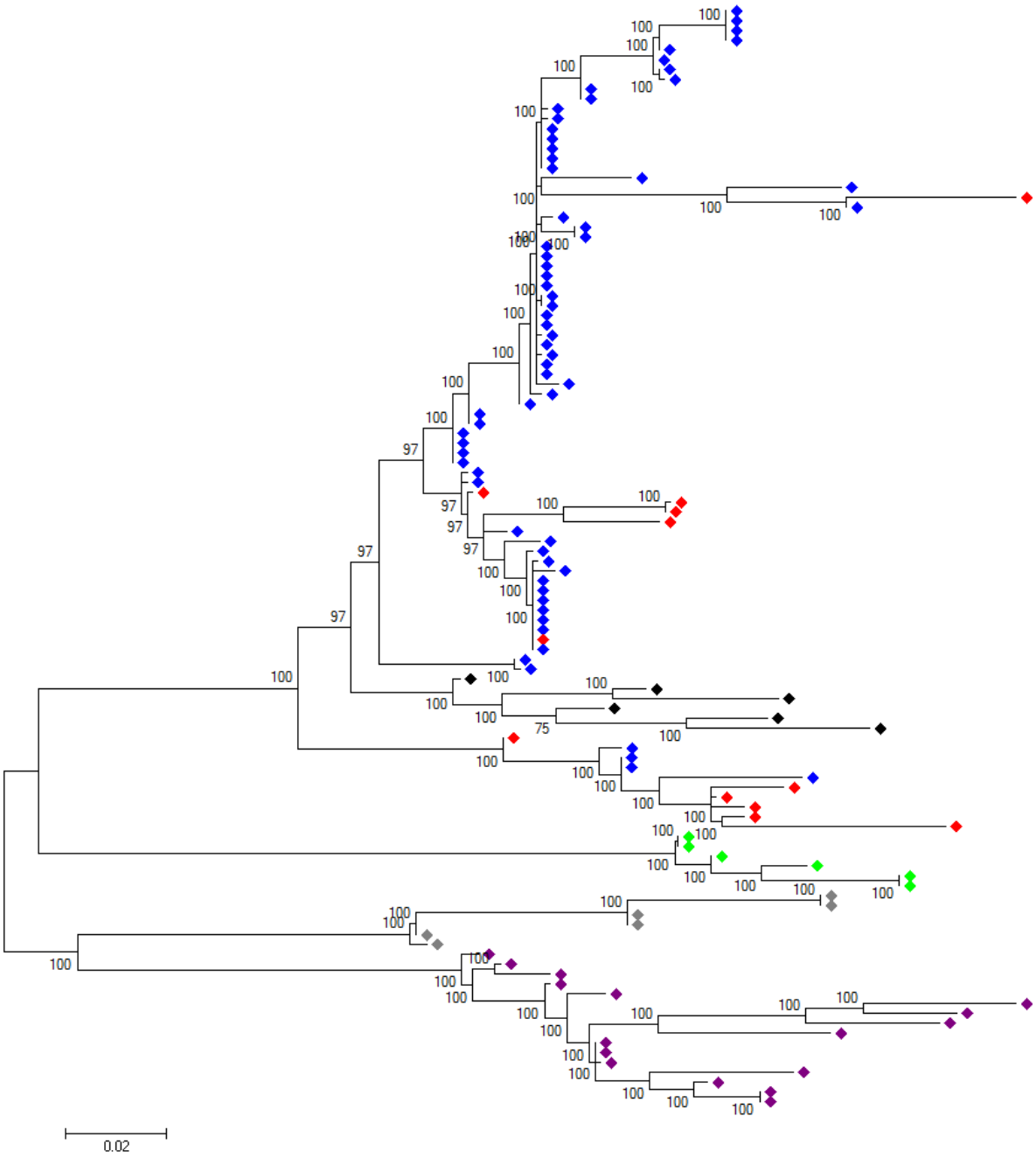


Figure S4.29. Haplotype based maximum likelihood tree for region of introgression from domestic chicken/red junglefowl into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 30-kb fragment size that extends from 695000 to 725000 bp on chromosome 8.

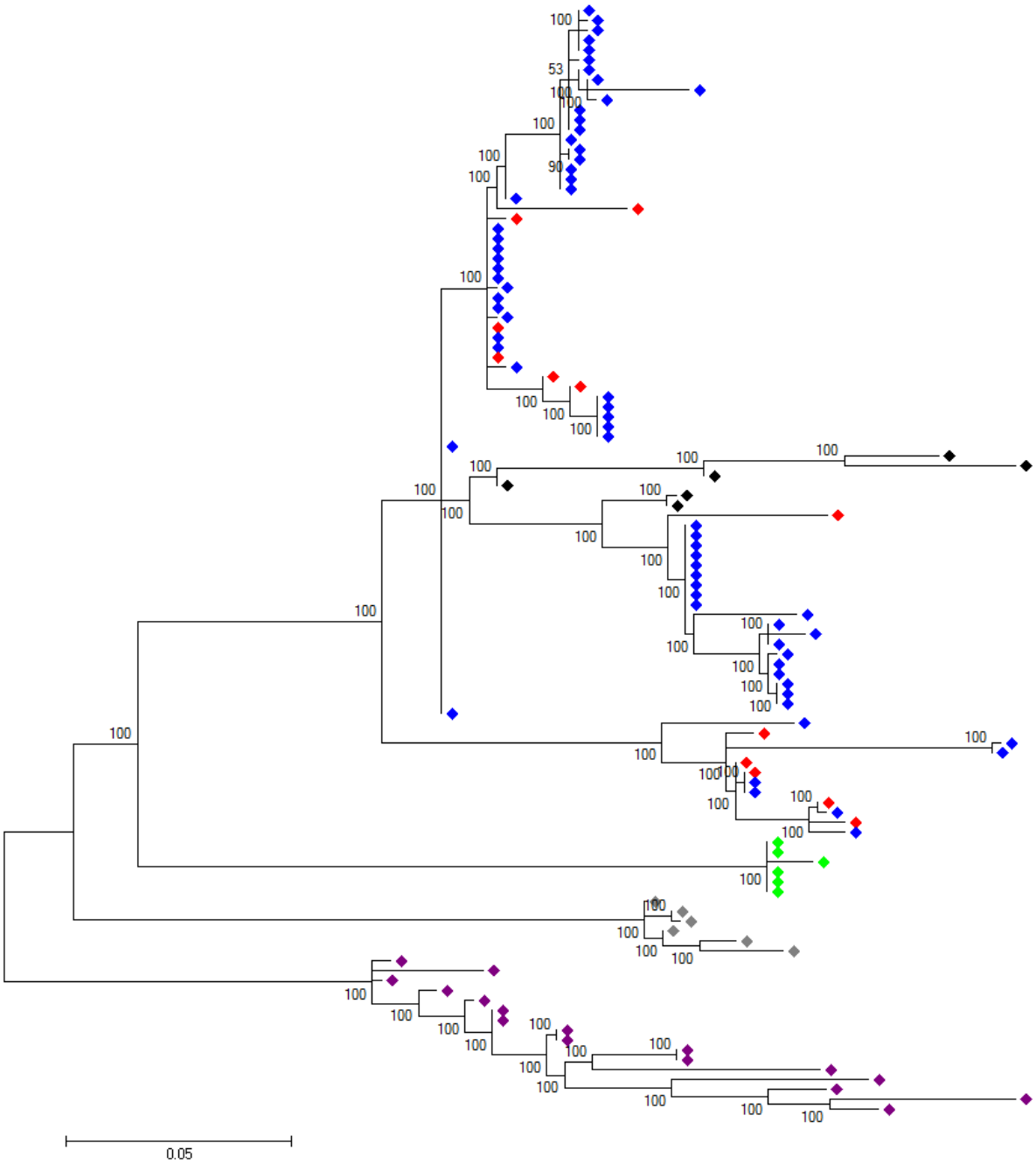


Figure S4.30. Haplotype based maximum likelihood tree for region of introgression from domestic chicken into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 35-kb fragment size that extends from 10810000 to 10845000 bp on chromosome 8.

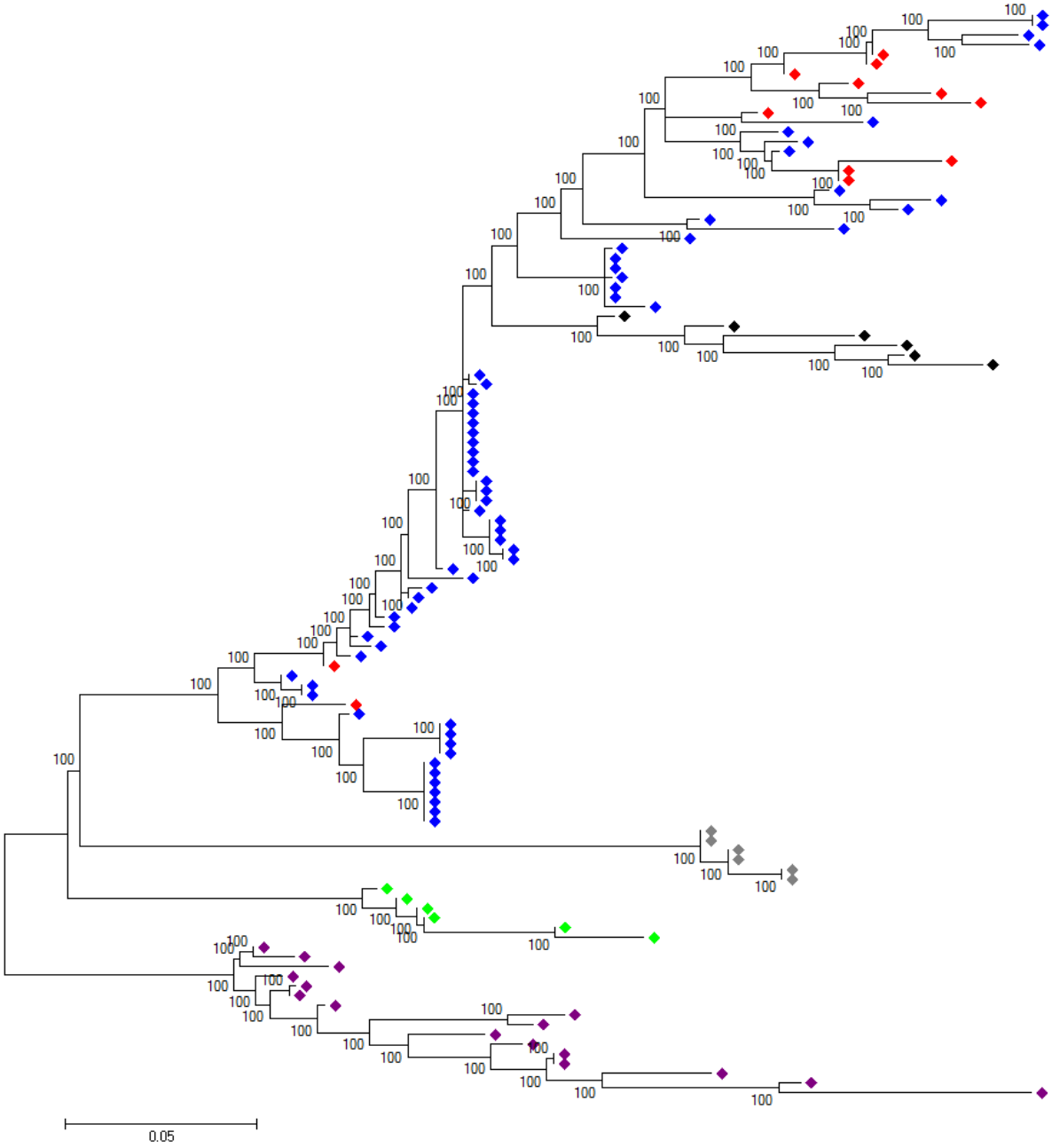


Figure S4.31. Haplotype based maximum likelihood tree for region of introgression from domestic chicken/red junglefowl haplotypes into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 30-kb fragment size that extends from 10450000 to 10480000 bp on chromosome 14.

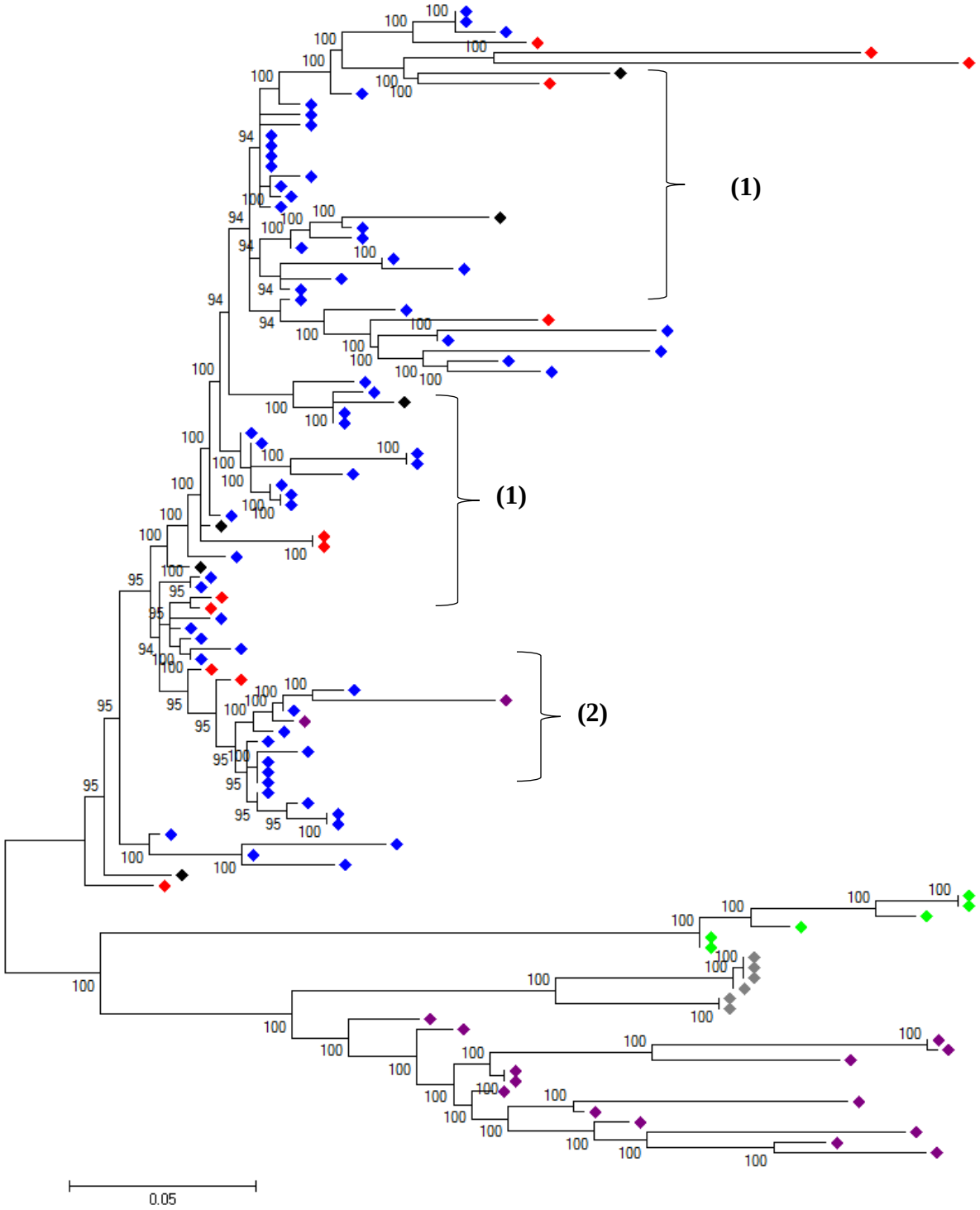


Figure S4.32. Haplotype based maximum likelihood tree for region of introgression **(1)** domestic chicken/red junglefowl haplotypes into Javan red junglefowl (**2**) domestic chicken/red junglefowl haplotypes into Lafayetii2k sample. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 30-kb fragment size that extends from 5020000 to 5050000 bp on chromosome 23.

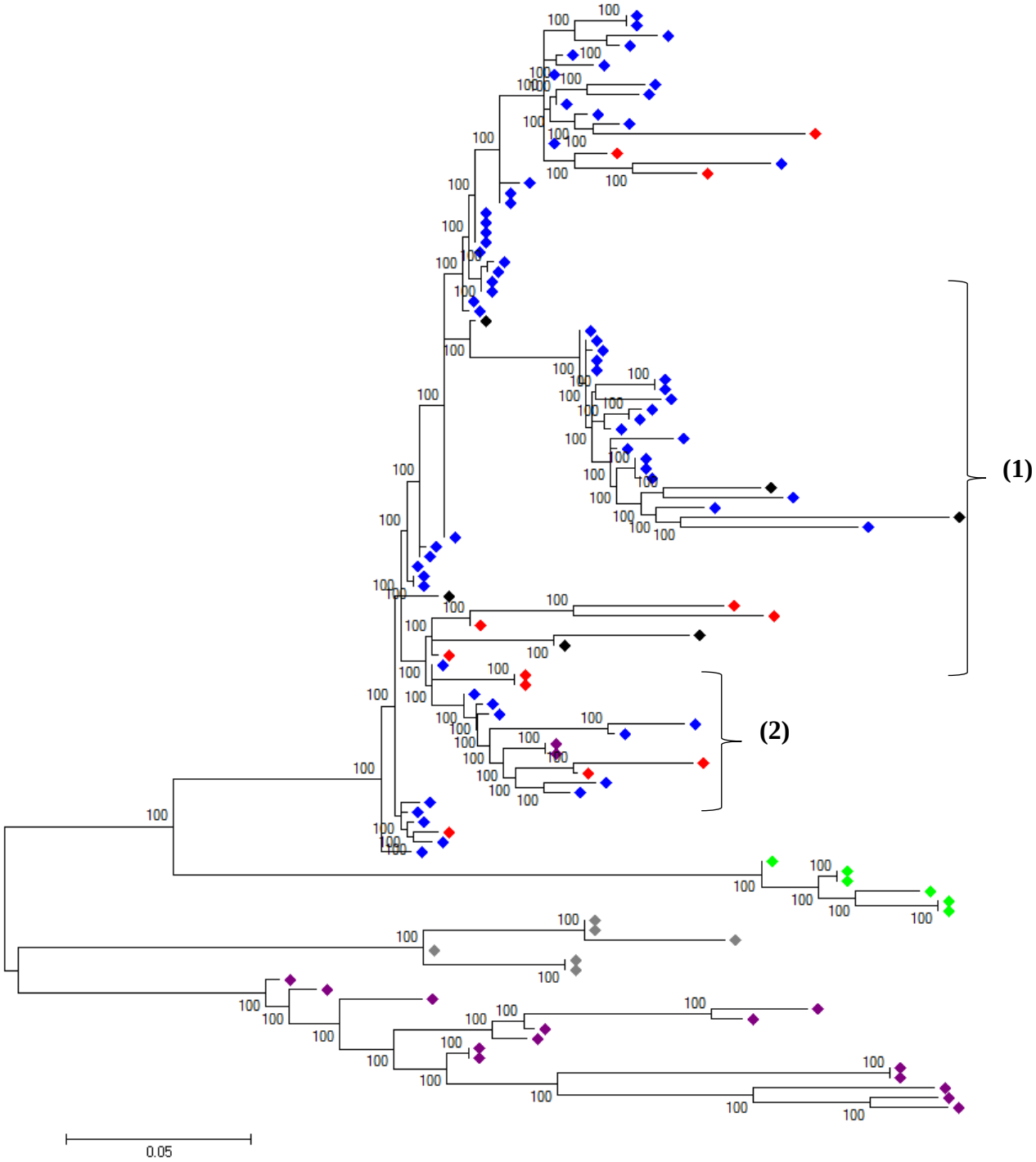


Figure S4.33. Haplotype based maximum likelihood tree for region of introgression **(1)** domestic chicken/red junglefowl haplotypes into Javan red junglefowl (**2**) domestic chicken/red junglefowl haplotypes into Lafayetii2k sample. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is 85 kb fragment size that extends from 5225000 to 5310000 bp on chromosome 23.

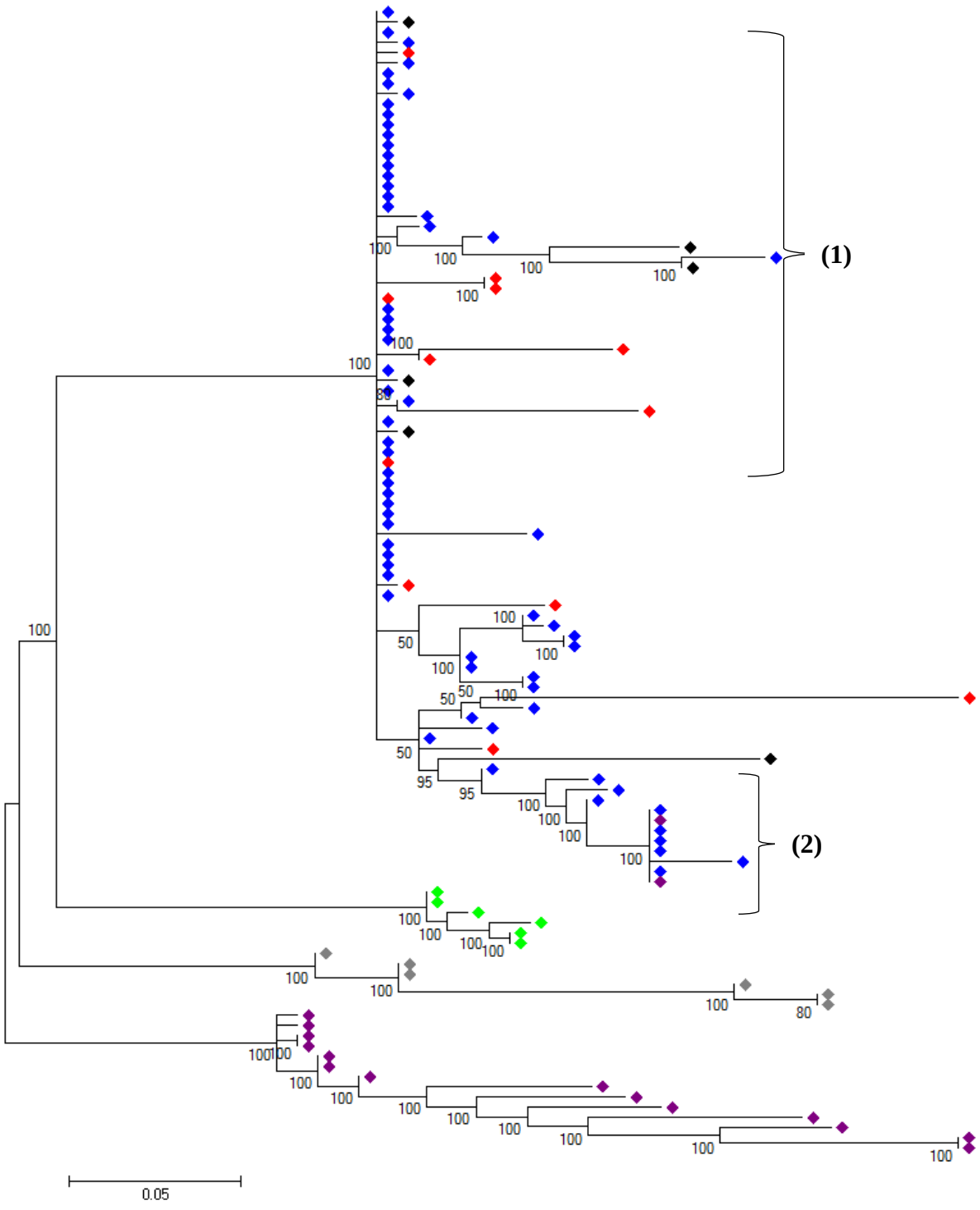


Figure S4.34. Haplotype based maximum likelihood tree for region of introgression **(1)** domestic chicken/red junglefowl haplotypes into Javan red junglefowl (**2**) domestic chicken haplotypes into Lafayetii2k sample. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 35-kb fragment size that extends from 5330000 to 5365000 bp on chromosome 23.

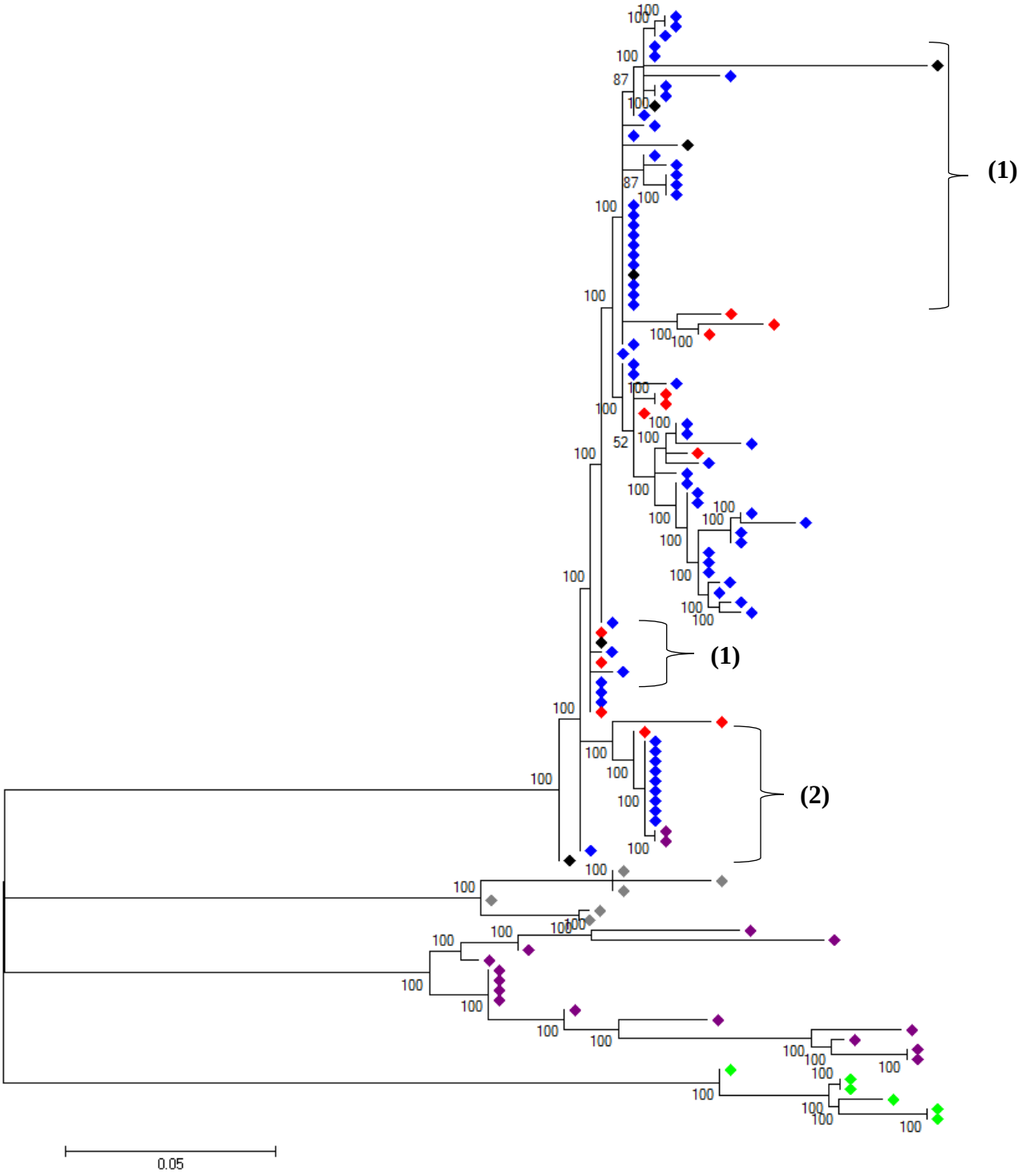


Figure S4.35. Haplotype based maximum likelihood tree for region of introgression (**1**) from domestic chicken/red junglefowl haplotypes into Javan red junglefowl **(2)** domestic chicken haplotypes into Lafayetii2k sample. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 45-kb fragment size that extends from 5510000 to 5555000 bp on chromosome 23.

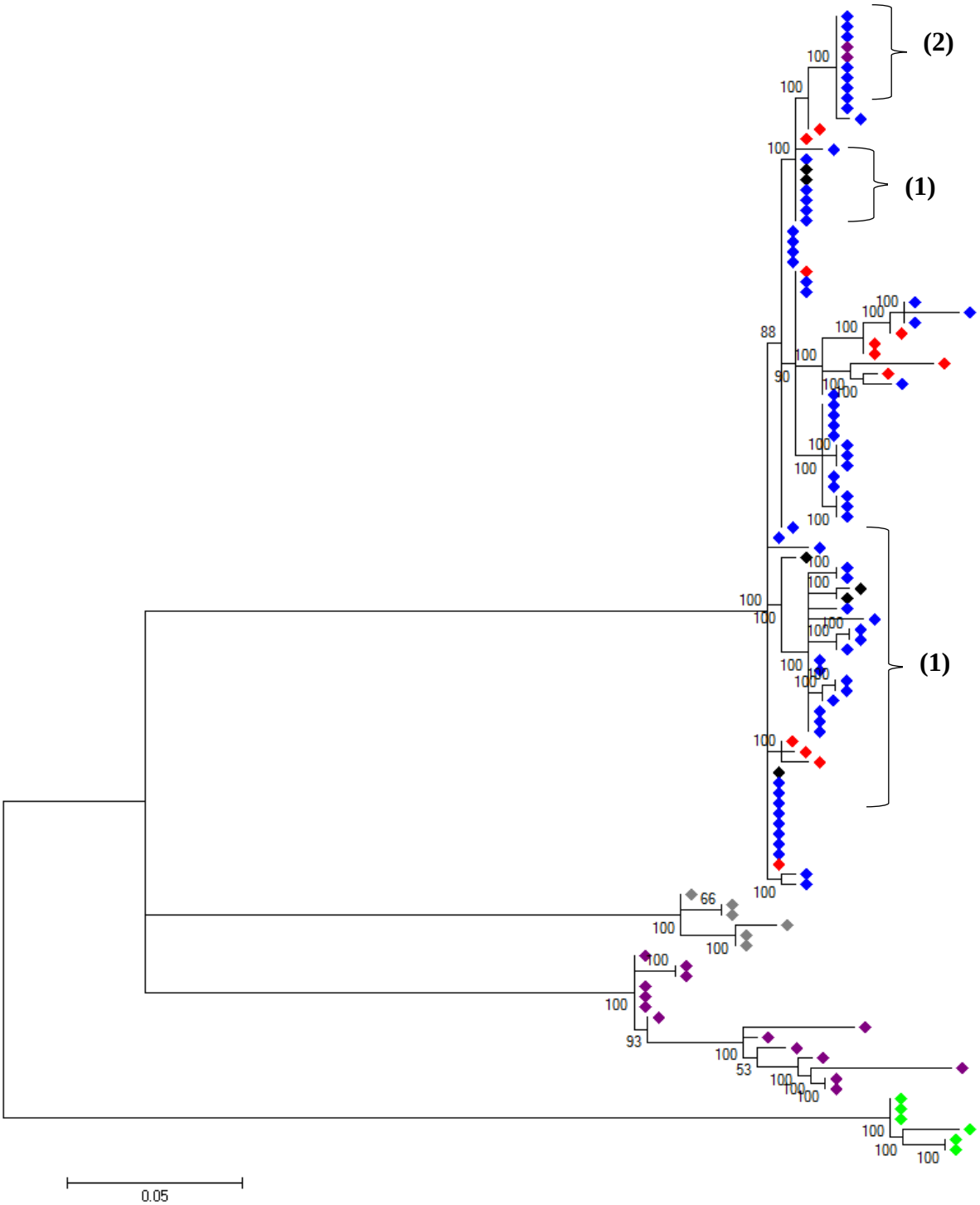


Figure S4.36. Haplotype based maximum likelihood tree for region of introgression (**1**) domestic chicken/red junglefowl haplotypes into Javan red junglefowl **(2)** domestic chicken haplotypes into Lafayetii2k sample. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 5625000 to 5645000 bp on chromosome 23.

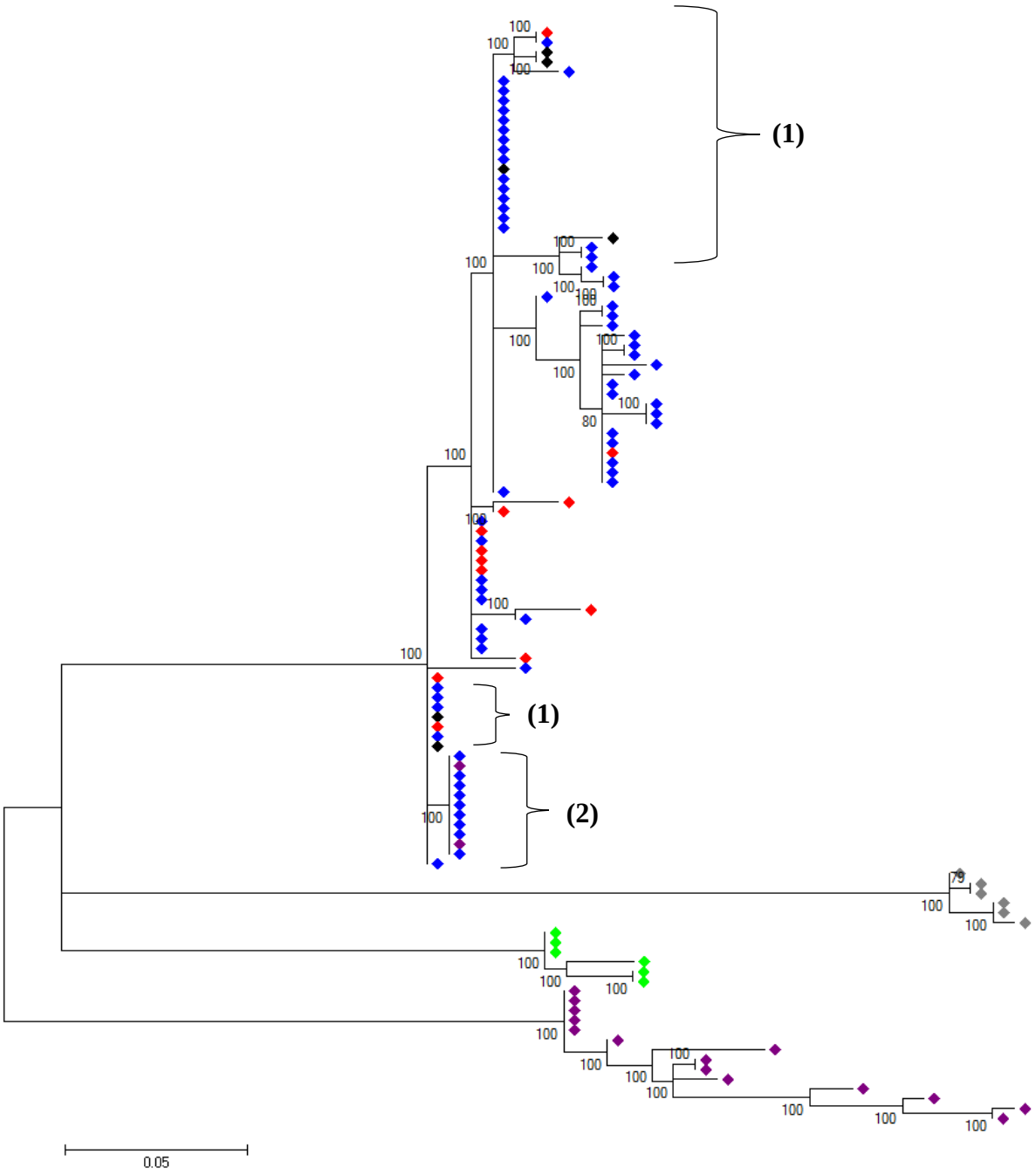


Figure S4.37. Haplotype based maximum likelihood tree for region of introgression (**1**) domestic chicken/red junglefowl haplotypes into Javan red junglefowl **(2)** domestic chicken haplotypes into Lafayetii2k sample. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 5700000 to 5720000 bp on chromosome 23.

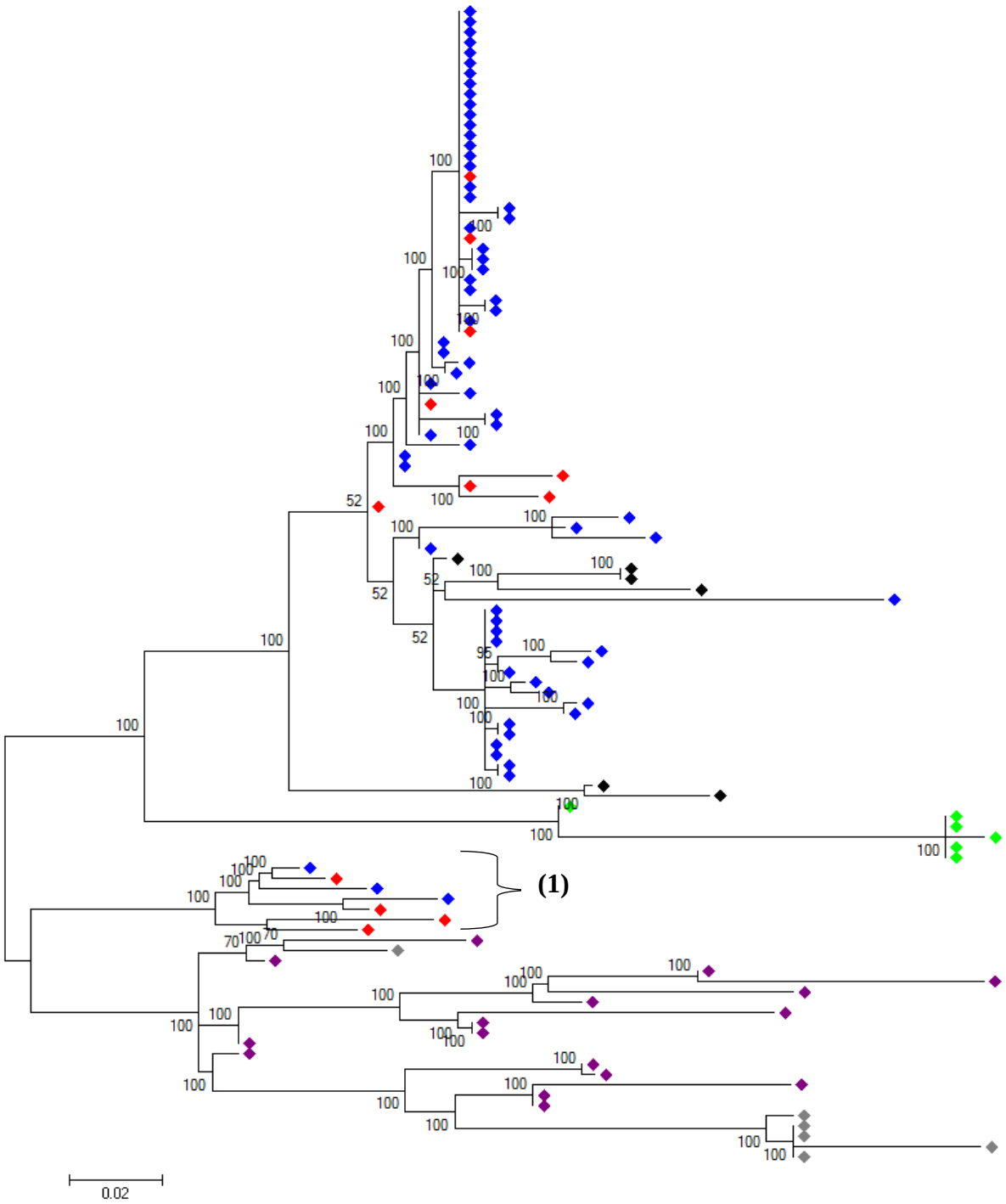


Figure S4.38. Haplotype based maximum likelihood tree for region of introgression from domestic chicken into Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 40-kb fragment size that extends from 160000 to 200000 bp on chromosome 27. **(1)** An unlikely introgressed haplotypes rather, an ambiguous relationship.

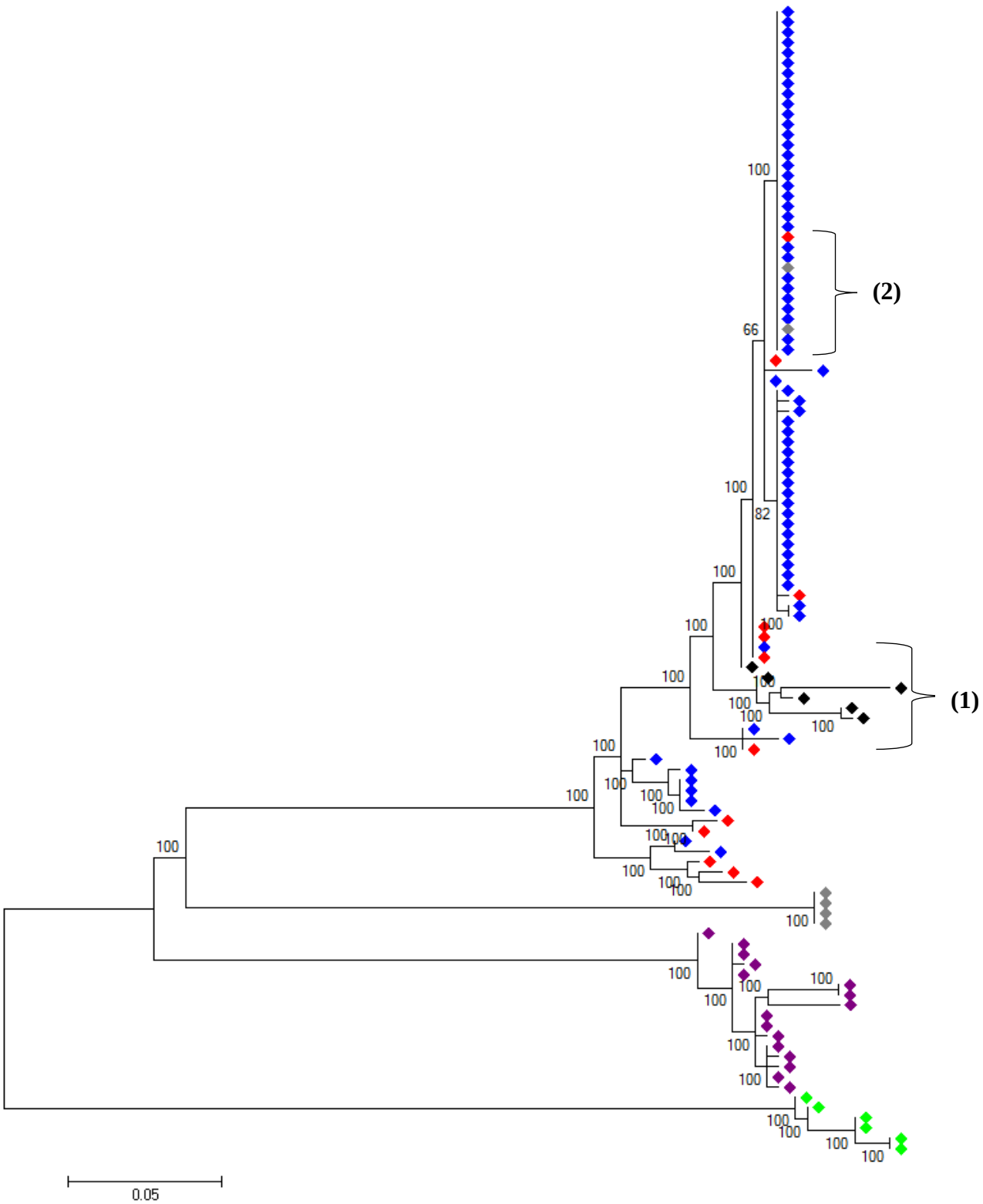


Figure S4.39. Haplotype based maximum likelihood tree for region of introgression from **(1)** domestic chicken/red junglefowl haplotypes into Javan red junglefowl and **(2)** domestic chicken/red junglefowl haplotypes into grey junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 25-kb fragment size that extends from 5110000 to 5135000 bp on chromosome 27.

Figures S5. The regions reported in this section are peaks above the genome threshold in the Javan red junglefowl - domestic chicken ABBA/BABA analysis, but the haplotype trees show inconclusive relationships or shared ancestry.

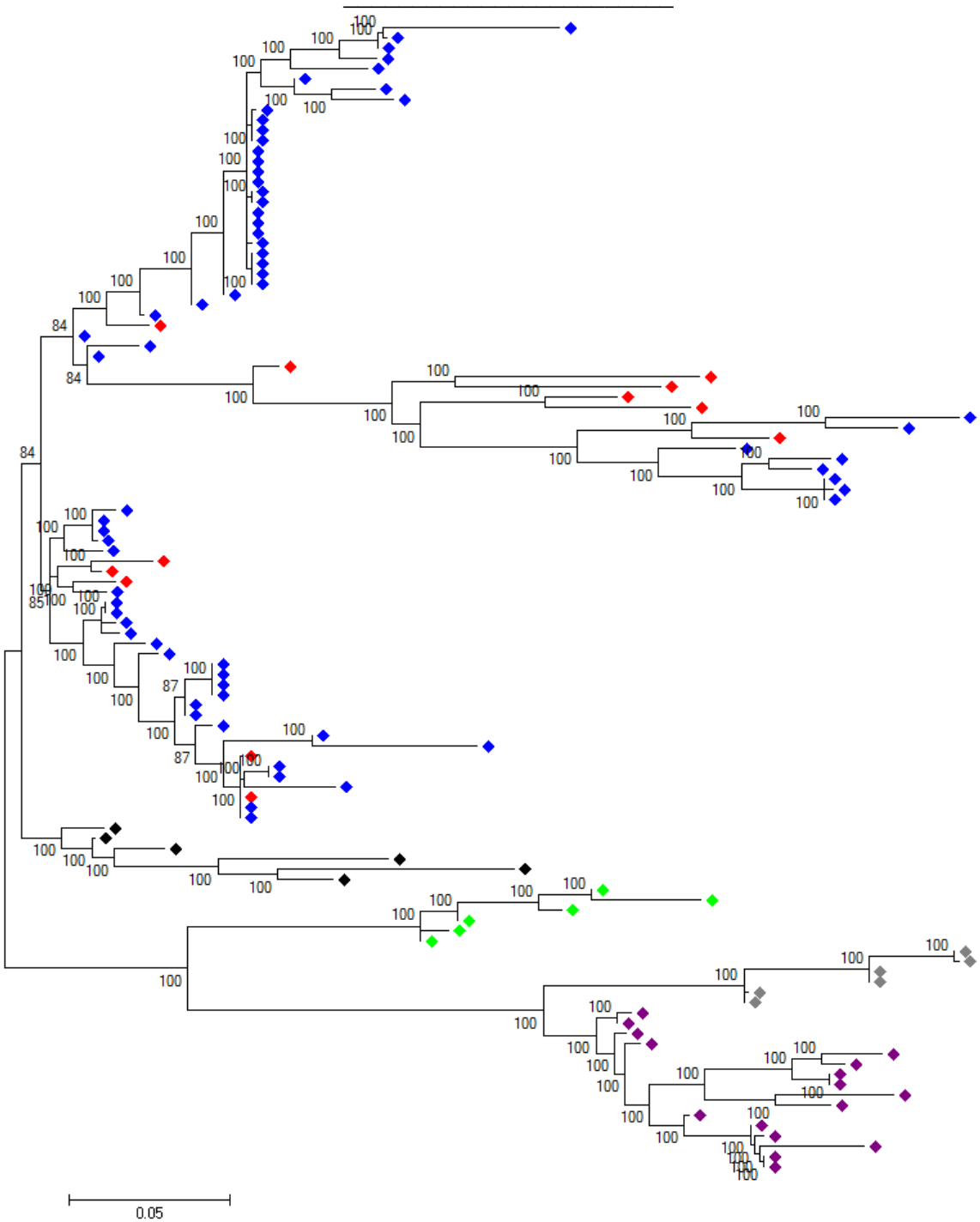


Figure S5.1 Haplotype based maximum likelihood tree for region of shared ancestry or ambiguous relationships between Javan red junglefowl and domestic chicken/red junglefowl haplotypes. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 25-kb fragment size that extends from 83180000 to 83205000 bp on chromosome 2.

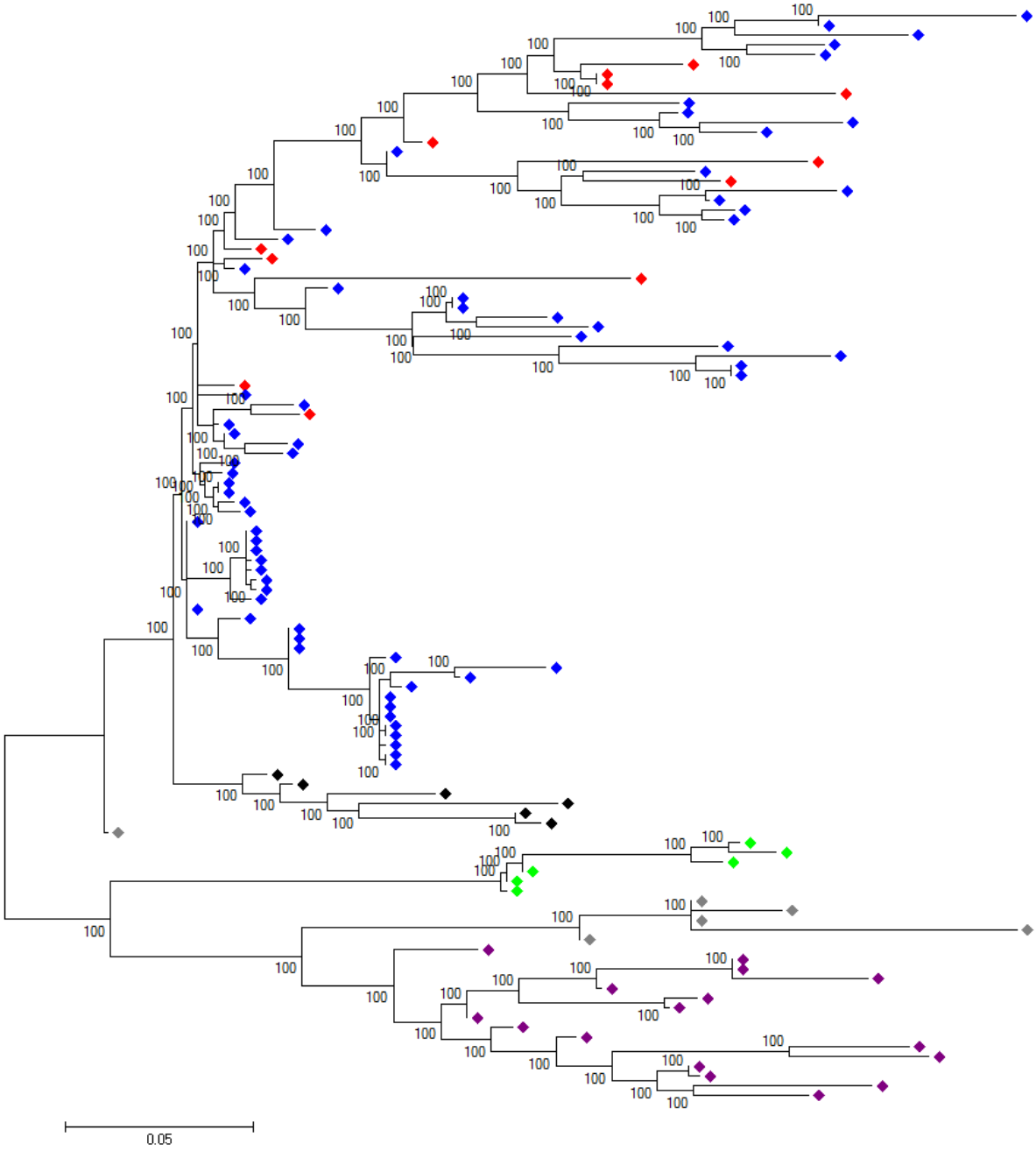


Figure S5.2 Haplotype based maximum likelihood tree for region of shared ancestry or ambiguous relationships between Javan red junglefowl and domestic chicken/red junglefowl haplotypes. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 25-kb fragment size that extends from 12160000 to 12185000 bp on chromosome 3.

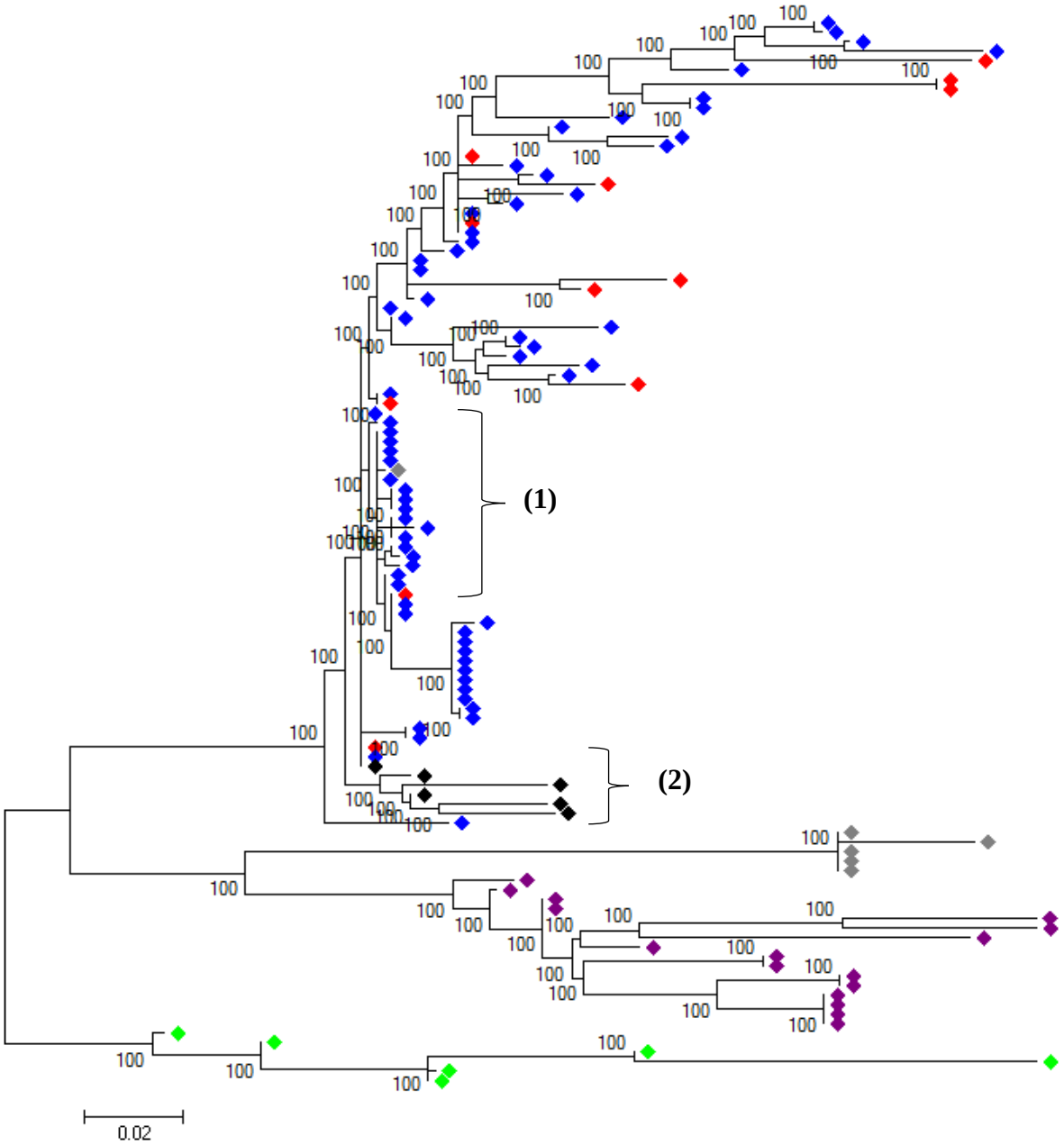


Figure S5.3 Haplotype based maximum likelihood tree for region of (**1**) introgression from domestic chicken/red junglefowl haplotypes to grey junglefowl (**2**) shared ancestry or ambiguous relationships between Javan red junglefowl and domestic chicken/red junglefowl haplotypes. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 25-kb fragment size that extends from 52495000 to 52520000 bp on chromosome 3.

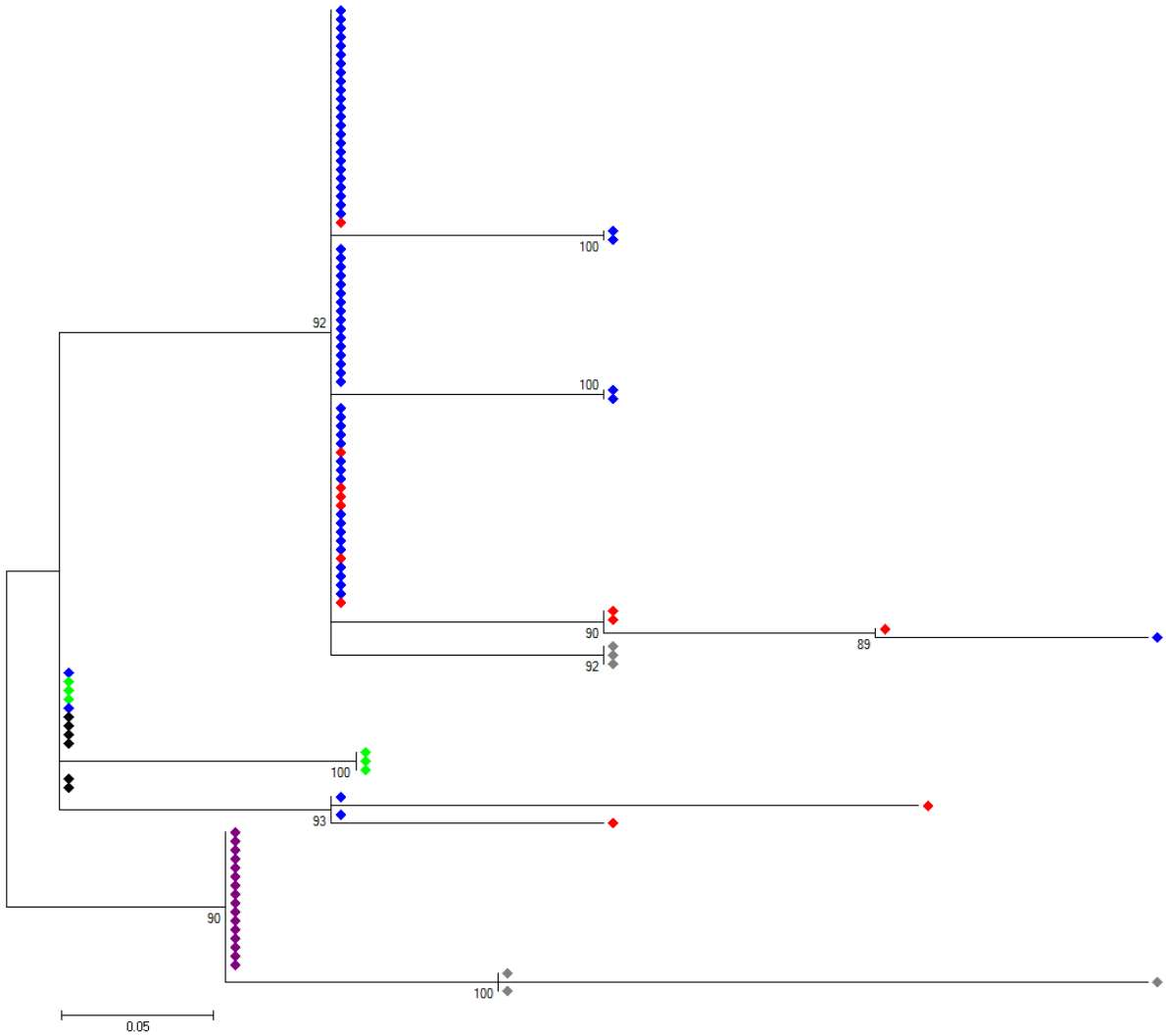


Figure S5.4 Haplotype based maximum likelihood tree for region showing ambiguity across all haplotypes. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 58800000 to 58820000 bp on chromosome 5.

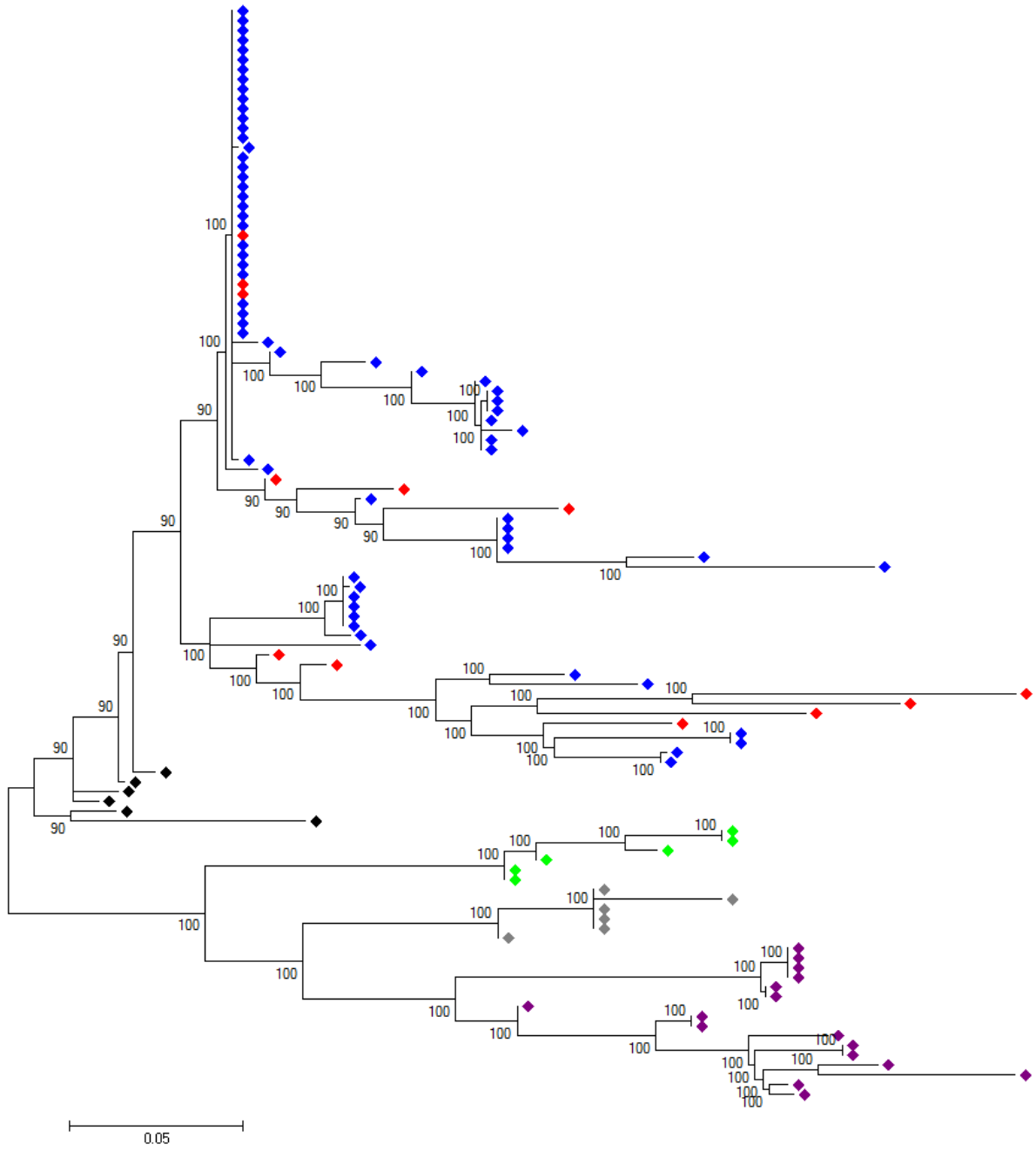


Figure S5.5 Haplotype based maximum likelihood tree for region showing a likely shared ancestry or an ambiguous relationship between red junglefowl/domestic chickens and Javan red junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 11645000 to 11665000 bp on chromosome 6.

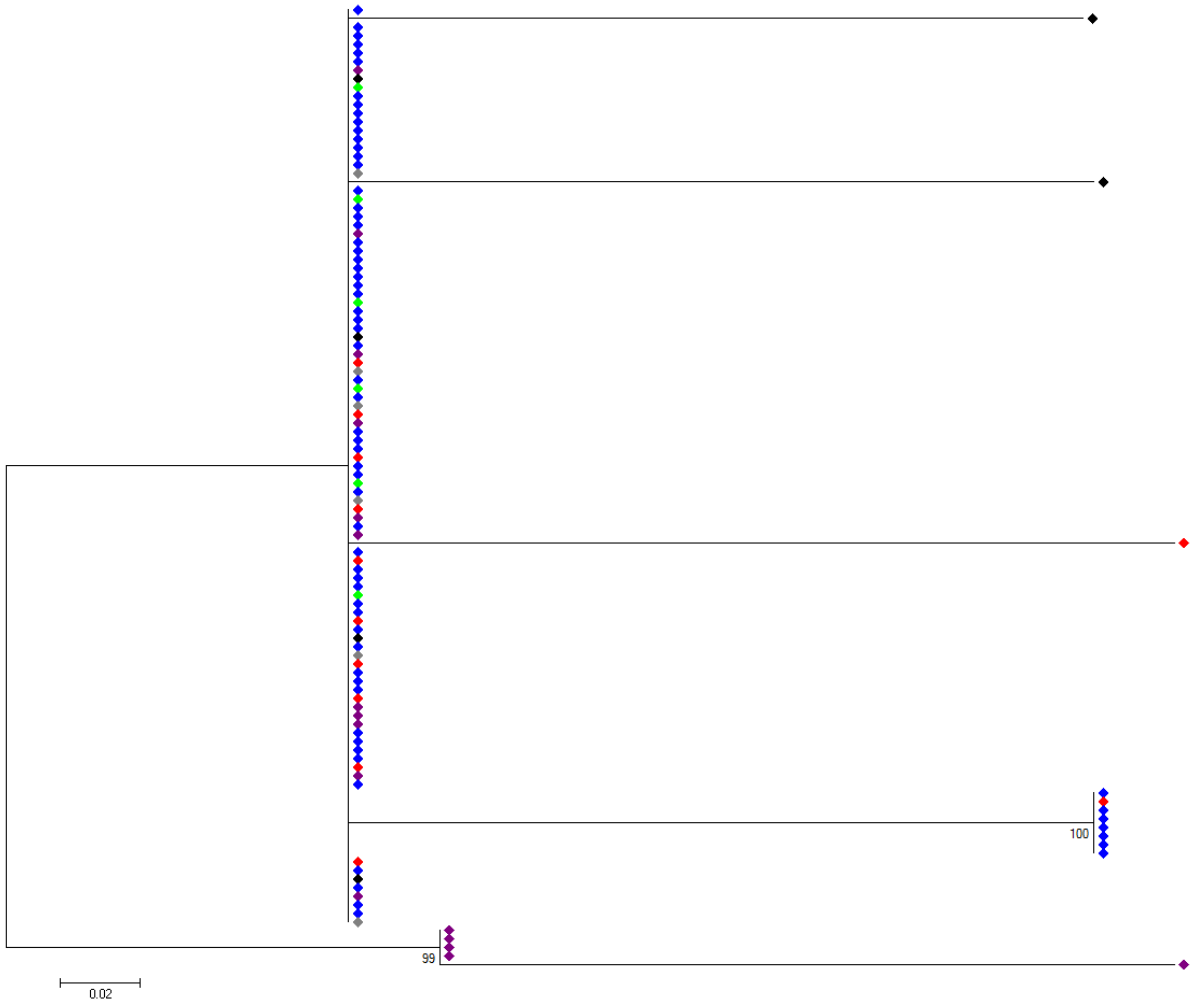


Figure S5.6 Haplotype based maximum likelihood tree for region showing ambiguous relationships across all haplotypes. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 1310000 to 1330000 bp on chromosome 11.

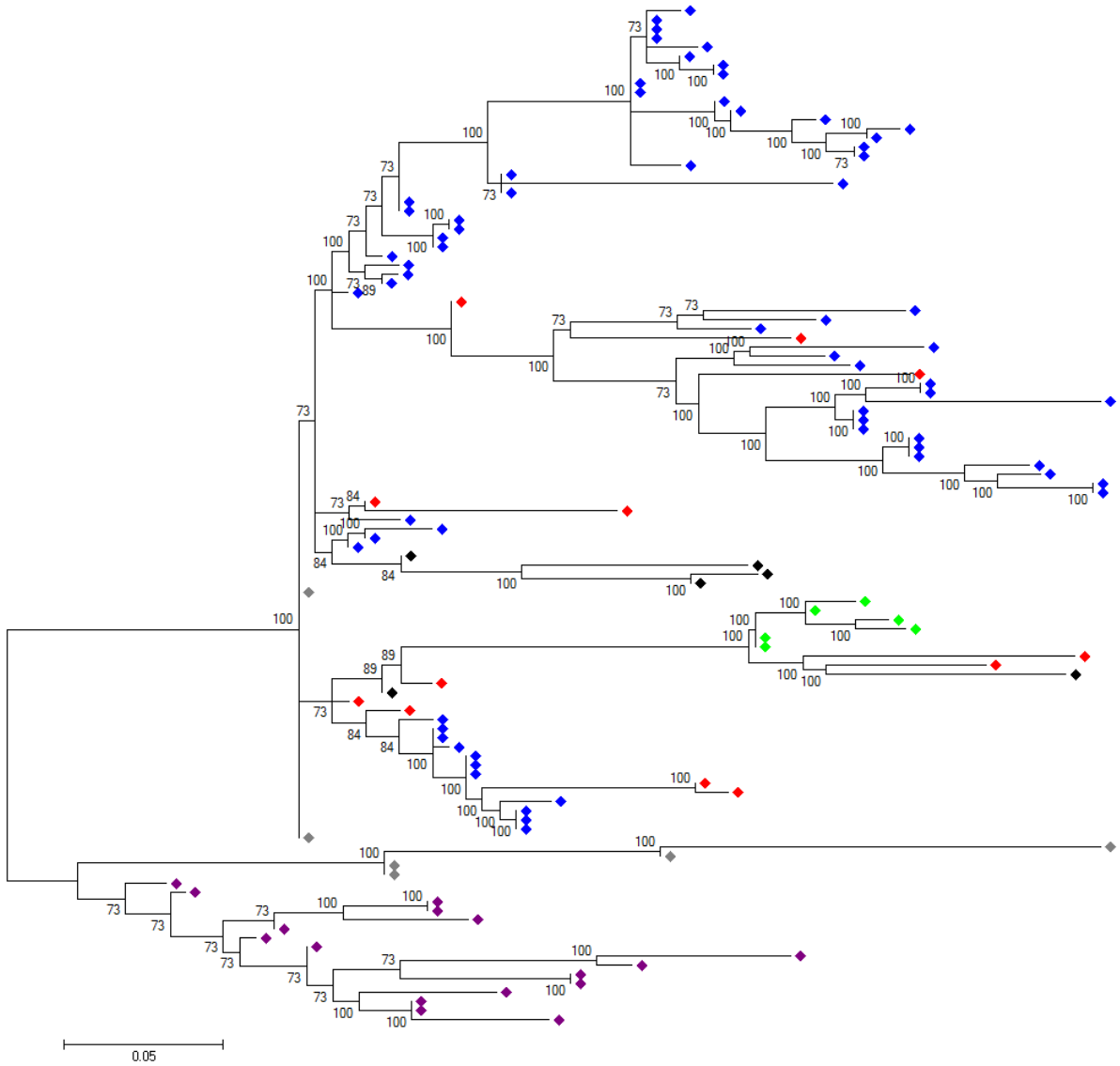


Figure S5.7 Haplotype based maximum likelihood tree for region showing ambiguous relationships among red junglefowl/domestic chickens, Javan red junglefowl, grey junglefowl and green junglefowl. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 25-kb fragment size that extends from 5425000 to 5450000 bp on chromosome 17.

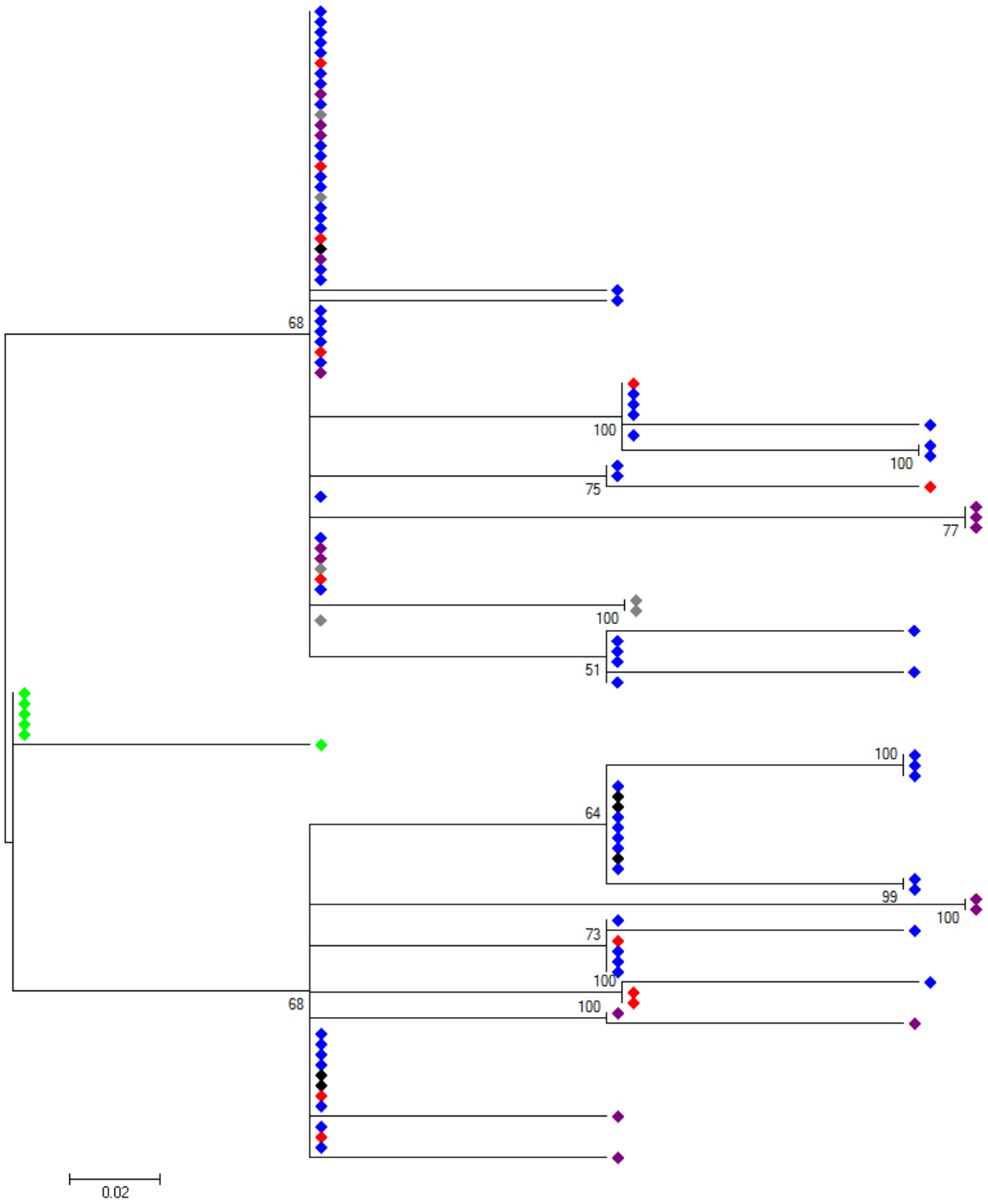


Figure S5.8 Haplotype based maximum likelihood tree for region showing ambiguous relationships across all haplotypes. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 30-kb fragment size that extends from 590000 to 620000 bp on chromosome 19.

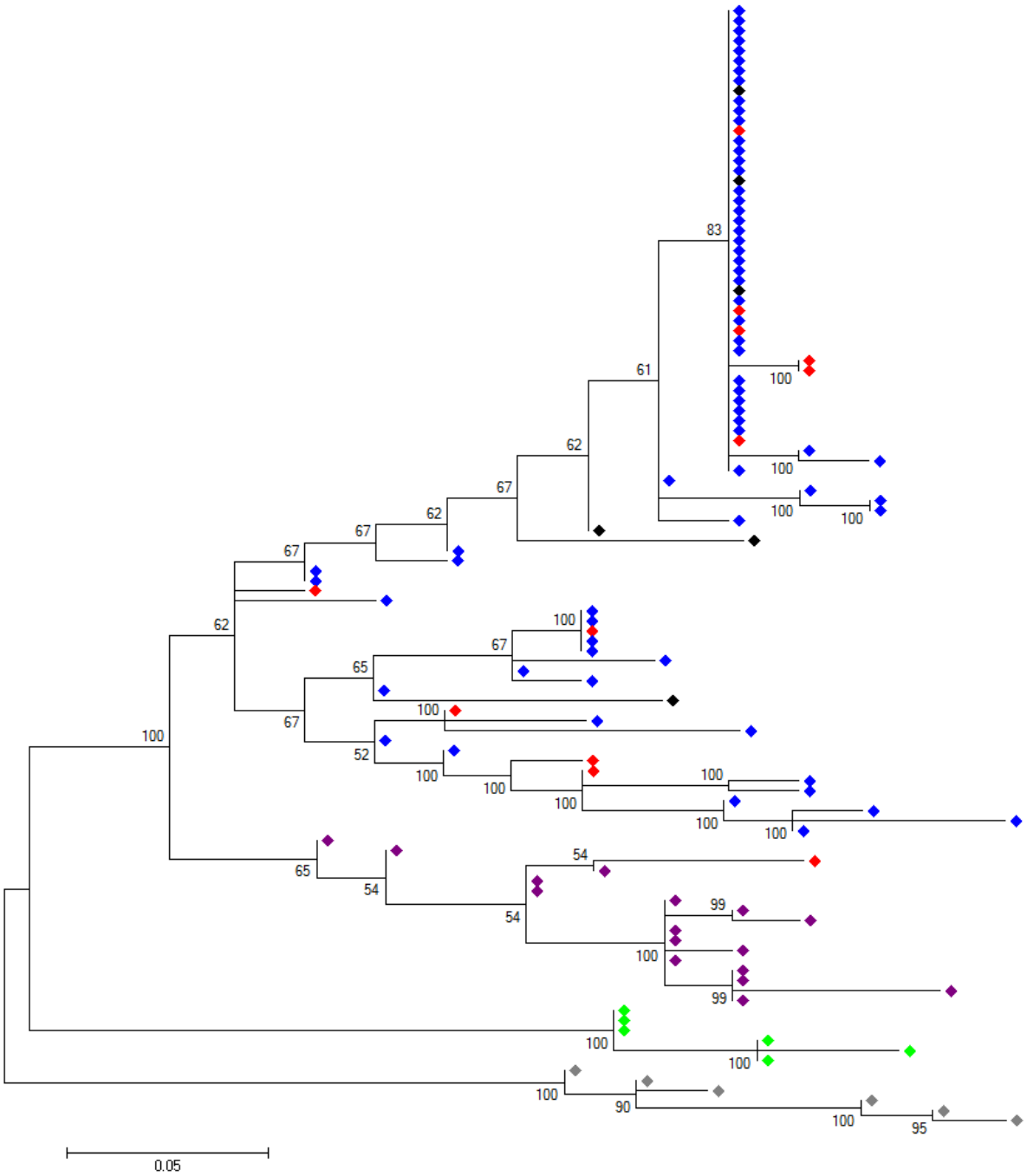


Figure S5.9 Haplotype based maximum likelihood tree for region showing ambiguous relationships across all haplotypes. ◆: Domestic chicken, ◆: red junglefowl, ◆: Javan red junglefowl ◆: grey junglefowl, ◆: Ceylon junglefowl, ◆: green junglefowl. This region is a 20-kb fragment size that extends from 2820000 to 2840000 bp on chromosome 20.