

University of Nottingham
UK | CHINA | MALAYSIA

Essential or Redundant? Disentangling the (GR)ESAG Families of *Trypanosoma brucei*

James William Chamberlain

MSc, PGCert, BSc (Hons)

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

School of Life Sciences

September 2017

Contents

I. Abstract **4**

II. Acknowledgements **5**

III. Declaration **6**

IV. Abbreviations **7**

Chapter 1. Introduction **9**

1.1. Contingency genes of pathogens **9**

1.2. Antigenic variation in *Trypanosoma brucei* **10**

1.2.1. Variant surface glycoprotein (VSG) 10

1.2.2. VSG expression sites 12

1.2.3. VSG switching 14

1.3. Functions of ESAGs and related genes **16**

1.3.1. The *T. brucei* transferrin receptor 17

1.3.2. Other ESAGs with less well defined function 20

1.3.3. A number of ESAGs localise to the host-parasite interface 22

1.3.4. Potential roles for (GR)ESAG families in parasite transmission 24

1.3.5. A role for ESAGs in active BES transcription 25

1.3.6. Adaptation to different hosts 27

1.4. Evolution of *(GR)ESAG* families **31**

1.4.1. *(GR)ESAG* families of *T. brucei* 31

1.4.2. Evolutionary origins of *(GR)ESAG* families and their conservation between African trypanosomes 38

1.5. Antigenic variation and linked genes in other parasites **41**

1.5.1. Apicomplexan parasites 41

1.5.2. *Giardia lamblia* 44

1.6. Aims **45**

Chapter 2. Methods **46**

2.1. Bioinformatics **46**

2.1.1. PSI-BLAST 46

2.2. Molecular biology **46**

2.2.1. Production of tagging constructs 47

2.2.2. Production of RNAi constructs 47
2.2.3. Production of KO fragments 48
2.2.4. Production of CreKO constructs 48
2.2.5. Diagnostic PCR 49
2.2.6. RNA extraction, cDNA synthesis and amplification of the actively transcribed *VSG* 49
2.3. Cell lines and culture 49
2.3.1. Axenic culture of *T. b. brucei* 50
2.3.2. Transfection 50
2.3.3. Growth rate analysis 51
2.4. Immunoblotting 51
2.5. Microscopy 52
2.6. Southern blotting 53
Chapter 3. A repertoire of GFP-tagged BES ESAGs 55
3.1. Introduction 55
3.2. Completing a repertoire of GFP-tagged BES1 ESAGs 56
3.3. Obtaining an RNAi competent BES3-active line 66
3.4. Generation of cell lines expressing GFP-tagged ESAG5 and ESAG10 71
3.5. Summary 78
Chapter 4. Reverse genetics approach to study ESAGs and (GR)ESAG families 81
4.1. Introduction 81
4.2. Generation of pan-family RNAi constructs 82
4.3. RNAi mediated knockdown of GFP-tagged ESAGs and (GR)ESAG families 86
4.4. Ablation of (GR)ESAG families, but not individual BES ESAGs, induces growth defects *in vitro* 88
4.5. Investigation of PAD1 and TfR levels in RNAi lines 90
4.6. Summary 99
Chapter 5. Generation of a BES with reduced *ESAG* repertoire ... 104
5.1. Introduction 104
5.2. Homologous recombination-based modification of BES1 . 106

5.3. Cre recombinase-based modification of BES1 114
5.4. CreKO lines do not have growth defects 122
5.5. Summary 129

Chapter 6. Discussion 131
6.1. RNAi as a technique to ablate (GR)ESAG families 133
6.2. CRISPR/Cas9 to study (GR)ESAG families 134
6.3. The roles of (GR)ESAG families *in vivo* 136
6.4. Producing an active BES with reduced *ESAG* repertoire ... 137
6.5. Virulence factors or contingency genes? 140

V. References 142

VI. Appendix 164

I. Abstract

The protozoan parasite *Trypanosoma brucei* escapes its mammalian host's immune response by switching expression of its major surface antigen, variant surface glycoprotein (VSG), which forms a protective coat around the cell. The actively expressed VSG is transcribed from one of 14 bloodstream expression sites (BESs), which also encode variable repertoires of expression site associated genes (ESAGs), most of which localise to the parasite surface. These genes form distinct families, are co-transcribed with *VSG*, and the repertoire expressed changes upon *in situ* switch. Specific ESAGs are known to protect the parasite from human serum lytic effects, modulate the host's innate immune response, mediate uptake of essential nutrients, and may adapt the parasite to the serum of different mammalian hosts. In addition to BES-linked *ESAGs*, these families also encompass non-BES encoded genes related to *ESAGs* (*GRESAGs*), some of which, unlike the *ESAGs*, are expressed in both mammalian and insect lifecycle stages.

To investigate the functional relationship between BES and non-BES (GR)ESAGs, I established a system to specifically ablate only BES-derived transcript or multiple *(GR)ESAG* family members. The former was carried out by tagging BES-linked ESAGs with GFP, and then targeting the *GFP* ORF by RNAi. The latter involved pan-family RNAi constructs, designed using regions with high identity across individual *(GR)ESAG* families. Pan-family, but not BES-specific, knockdown induced growth defects, showing that transcript from the active BES itself is not essential for parasite survival *in vitro*. To investigate this further, a strategy based on Cre recombinase genome engineering was developed to specifically reduce the *ESAG* repertoire contained within the active BES. Stable cell lines ready for recombination were obtained, and will be used for analysis of modified BESs.

Together, the results achieved show that (GR)ESAG families, as opposed to the actively-expressed ESAG copy, are important for parasite survival *in vitro*, and there is redundancy between BES-linked and non-BES transcripts. Therefore, the often-overlooked GRESAGs may play more significant roles than originally thought.

II. Acknowledgements

Firstly, I am immensely grateful to Dr Catarina Gadelha, for allowing me the opportunity to undertake this project, and supporting and believing in me (even through music festivals, lost posters and a hospitalisation). Thank you to Dr Bill “The Oracle” Wickstead, for always having the answers. Guidance provided by you both has been invaluable.

Thanks to all the members of the Allers, Friel, Gadelha, Huett, Nieduszynski and Wickstead labs (Sarah, Jen and Simon in particular) for making the last four years so enjoyable, and always being there for tea, coffee or a pint when necessary, even through the grumpiness. Thank you also for the shared equipment, expertise and help along the way.

I am grateful to my family for their constant support, especially Catriona, who has kept me fed throughout the duration of this degree.

This work was supported by a BBSRC DTP programme stipend. Work in the Gadelha lab is funded by the Royal Society.

III. Declaration

This thesis is the result of my own work, except where included data is explicitly mentioned, which has been undertaken during my period of registration for this degree at The University of Nottingham.

Portions of the work presented in Chapter 3 have been previously published as Gadelha C, Zhang W, Chamberlain JW, Chait BT, Wickstead B, Field MC. 2015. Architecture of a host–parasite interface: complex targeting mechanisms revealed through proteomics. *Mol Cell Proteomics* **14**: 1911–1926. A copy can be found in Appendix 1.

IV. Abbreviations

BAC	bacterial artificial chromosome
BES	bloodstream expression Site
BLAST	blasticidin S deaminase
BLASTn	nucleotide basic local alignment search tool
BLE	bleomycin phosphotransferase
bp	base pair
BSF	bloodstream form
cAMP	cyclic adenosine monophosphate
cDNA	complementary DNA
CLH	clathrin heavy chain
ConA	concanavalin A
DAPI	4,6-diamidino-2-phenylindole
DNA	deoxyribonucleic acid
DSB	double strand break
dsRNA	double stranded RNA
eGFP	enhanced GFP
ESAG	expression site associated gene
GFP	green fluorescent protein
GPI	glycosylphosphatidylinositol
GRESAG	gene related to ESAG
HSV-TK	herpes simplex virus thymidine kinase
HYG	hygromycin phosphotransferase
IC	intermediate chromosome
Ig	immunoglobulin
ISG	invariant surface glycoprotein
kb	kilobase
kDa	kilodalton
LRR	leucine rich repeat
Mb	megabase
MBC	megabase chromosome
MC	minichromosome
MES	metacyclic expression site

mRNA	messenger RNA
NEO	neomycin phosphotransferase
ORF	open reading frame
PAD	proteins associated with differentiation
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCF	procyclic form
PCR	polymerase chain reaction
PFGE	pulsed-field gel electrophoresis
PSI-BLAST	position-specific iterated basic local alignment search tool
PUR	puromycin N-acetyltransferase
rDNA	ribosomal DNA
RES	resistance-associated expression site
RING	really interesting new gene
RNA	ribonucleic acid
RNAi	RNA interference
SDS	sodium dodecyl sulphate
sfGFP	superfolder GFP
SIF	stumpy induction factor
siRNA	small interfering RNA
SRA	serum resistance-associated
ssRNA	single stranded RNA
TAR	transformation-associated recombination
TfR	transferrin receptor
UTR	untranslated region
VSG	variant surface glycoprotein

Chapter 1. Introduction

1.1. Contingency genes of pathogens

Survival and persistence of any pathogen in their host is paramount to ensuring forward transmission and spread. Pathogens must therefore adapt to an environment that has evolved specifically to repel this challenge, and control or eradicate the threat, initiating a molecular 'arms race' between the host and invader. Infectious protozoa, fungi, bacteria and viruses have all developed mechanisms that facilitate host colonisation and overcome its defences. In some cases this requires rapid adaptation of the microorganism to different environments and challenges, which is often mediated by 'contingency' genes (Moxon *et al.* 1994; Barry *et al.* 2003). These enable rapid phenotypic changes through either switching expression of specific genes 'on' or 'off', a common feature of bacteria called phase variation; or switching expression within a gene family to an immunologically distinct variant, called antigenic variation (Moxon *et al.* 2006; Deitsch *et al.* 2009). Phase variation can be considered a simplified version of antigenic variation, with phase control often mediated through genetic changes (changing the DNA sequence of the organism); whereas true antigenic variation commonly involves large heterogenic gene families, whose often mutually exclusive expression is under epigenetic control (Deitsch *et al.* 2009; Duraisingh and Horn 2016). Both of these strategies are used to facilitate immune evasion, host cell invasion and adherence, and virulence factor production in numerous bacterial, fungal and protozoan pathogens (Deitsch *et al.* 1997; Moxon *et al.* 2006; Deitsch *et al.* 2009). What is striking are the common themes shared by these microorganisms to enable antigenic variation. Pathogen contingency genes are often located in subtelomeric regions of their genome (in both eukaryotes and eubacteria such as *Borrelia* spp.), and have become disproportionately large in these organisms to encode contingency gene families and specialised transcription units that facilitate their expression (Barry *et al.* 2003; Li 2012; Obado *et al.* 2016). These dynamic genomic regions undergo high rates of recombination, which has enabled expansion and diversification of contingency gene families, and plays important roles in controlling antigenic variation (Barry *et al.* 2003;

Palmer and Brayton 2007; Vink *et al.* 2012). A well-studied and extreme example of a pathogen that relies on antigenic variation to evade its host's immune response is the kinetoplastid parasite *Trypanosoma brucei*, which will form the focus of this thesis.

1.2. Antigenic variation in *Trypanosoma brucei*

T. brucei are unicellular flagellate protozoan parasites of mammals. Spread by the bite of an insect vector (tsetse flies of the genus *Glossina*), they cause disease of man (human African trypanosomiasis, or sleeping sickness) and animals (animal African trypanosomiasis, or nagana) in sub-Saharan Africa (Barrett *et al.* 2003). Human disease is caused by two specific subspecies of *T. brucei*: *T. b. gambiense* and *T. b. rhodesiense*; while disease of animals is caused by all *T. brucei* trypanosomes (including subspecies *T. b. brucei*) and related parasites *Trypanosoma congolense* and *Trypanosoma vivax*. Other *T. brucei* subspecies, *T. b. evansi* and *T. b. equiperdum*, have escaped Africa by adapting to mechanical and sexual transmission, respectively (Claes *et al.* 2005; Lai *et al.* 2008; Desquesnes *et al.* 2013; Cuypers *et al.* 2017). *T. b. evansi* is spread by biting insects and infects a range of animals (including camels, equines, and dogs) causing a disease known as surra; while *T. b. equiperdum* causes dourine in horses (Desquesnes *et al.* 2013; Gizaw *et al.* 2017). All of these parasites are known to undergo antigenic variation in their respective mammalian hosts; *T. b. brucei* is the most widely researched and best understood, with the features of this system thought to be shared by related subspecies. Mechanisms underpinning antigenic variation in the more distantly related *T. congolense* and *T. vivax* remain to be elucidated (Barbet and McGuire 1978; Barry 1986).

1.2.1. Variant surface glycoprotein (VSG)

A catalogue of >3000 *VSG* genes are encoded by the *T. b. brucei* genome, around 80% of which consists of partial genes or pseudogenes (Cross *et al.* 2014). These are located in subtelomeric arrays on the 11

‘megabase’ chromosomes (MBC, 1-6 Mb, analogous to chromosomes in other diploid organisms), and subtelomerically in ~1-5 ‘intermediate’ chromosomes (IC, 200-900 kb) and ~100 ‘minichromosomes’ (MC, 50-150 kb) whose cores are made of 177-bp repeat elements (Ersfeld *et al.* 1999). The encoded ~50-60 kDa protein is highly glycosylated and attached to membranes by a C-terminal glycosylphosphatidylinositol (GPI)-anchor (Ferguson *et al.* 1988; Mehlert *et al.* 1998). Despite low primary sequence similarity between variants, the tertiary structure of these proteins is predicted to be conserved, and consists of two long N-terminal α-helices that sit perpendicular to the cell membrane, producing an elongated shape (Figure 1; Blum *et al.* 1993; Carrington and Boothroyd 1996). Approximately 5×10^6 VSG dimers form a monolayer of around 12-15 nm thick on the cell surface (Vickerman 1969; Auffret and Turner 1981; Jackson *et al.* 1985). This dense intercalated coat is thought to shield invariant surface proteins from view of the host’s immune system, and inhibit complement activation by the alternative pathway (through deposition of complement component 3b on the parasite surface; Ferrante and Allison 1983), while allowing access to nutrients (Mehlert *et al.* 2012). However, structural modelling of invariant proteins with large extracellular domains (ESAG4 and ISG, introduced later) indicated that they may extend out of the VSG coat (Schwede *et al.* 2015). As antibodies specific to one of these invariant proteins are generated in infected hosts, they must be exposed to the immune system, although it is unknown whether this occurs at the parasite surface or upon cell lysis (Jackson *et al.* 1993). VSG elicits a humoral response from the host that results in clearance of parasites expressing this variant (Dempsey and Mansfield 1983; Diffley 1985). Therefore the antigenicity of VSG may act to ‘distract’ the immune system from invariant surface proteins. This may be analogous to the virulence factor Protein A employed by *Staphylococcus aureus*, where this ‘superantigen’ dominates the humoral immune response, limiting the production of protective antibodies against other pathogen antigens, thus enabling immune evasion (Pauli *et al.* 2014). VSG is known to be released into the host serum, which may enhance this feature (Diffley *et al.* 1980). Continuous recycling of the VSG coat occurs through the flagellar pocket, the sole site of *T. brucei* endo- and exocytosis

(Engstler *et al.* 2004). This facilitates removal of immunoglobulin-VSG complexes from the parasite surface, preventing activation of the classical complement pathway and phagocytosis by macrophages, therefore promoting survival (Engstler *et al.* 2007; Cheung *et al.* 2016).

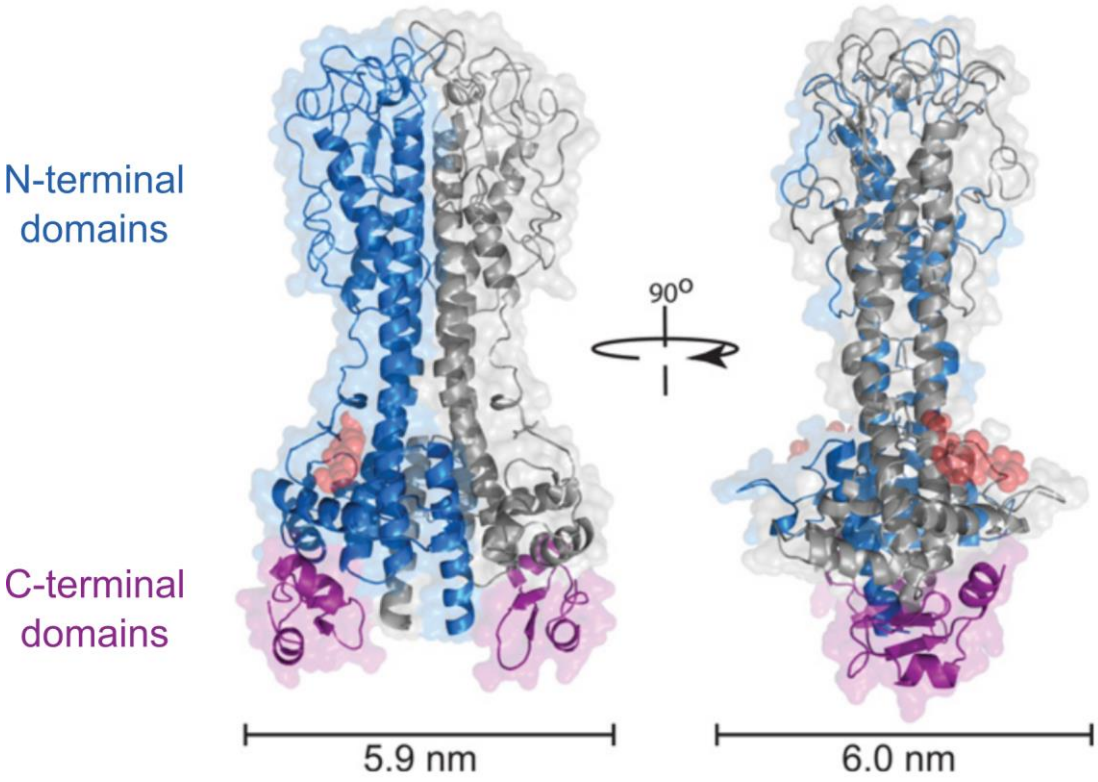


Figure 1. Structure of VSG221 homodimer. Space-filling model of VSG dimer, the N-terminal domain of one monomer is shown in blue, the other in grey (Freymann *et al.* 1990). C-terminal domains are coloured in purple (Chattopadhyay *et al.* 2005); N-linked oligosaccharide in the N-terminal domain is shown in red. The two long α-helices in each monomer that contribute to the elongated tertiary structure can be seen. N- and C-terminal domain structures were solved independently, so their relative positions are estimated in this model. Adapted from Schwede *et al.* 2015.

1.2.2. VSG expression sites

The actively expressed VSG is transcribed from one of 14 bloodstream expression sites (BESs) located subtelomerically on MBC and IC (Hertz-Fowler *et al.* 2008). The repertoire of BESs was realised through sequencing projects, first by sequencing *T. b. brucei* bacterial artificial chromosome (BAC) library clones, and later completed by transformation-associated recombination (TAR) cloning (Berriman *et al.* 2002; Becker *et al.* 2004). Despite evidence of extensive recombination between BESs, they share a conserved architecture and a number of unrelated genes separate the

telomere proximal *VSG* gene from the BES promoter (Figure 2; Hertz-Fowler *et al.* 2008). Unusually, co-transcription of these expression site associated genes (*ESAGs*) with *VSG* is mediated by RNA Polymerase I (which exclusively synthesises rDNA in other eukaryotes) from the active BES when in the bloodstream stage of the parasite lifecycle (Günzl *et al.* 2003). *ESAG* repertoires vary dramatically between BESs, but follow a conserved order. *ESAG6* and *ESAG7* are located directly downstream of the BES promoter, closely followed by *ESAG5* and an *ESAG3* pseudogene. *ESAG4*, *ESAG8* and *ESAG3* follow this in the majority of BESs, with *ESAG12* occasionally intervening between *ESAG4* and *ESAG8*. A unit of *ESAG2*, pseudogenic *ESAG11* and *ESAG1* commonly complete the set of *ESAGs* and are found just upstream of a region consisting of variable numbers of 70-bp repeats. This region separates *VSG* from *ESAG* coding genes and occasionally contains *ESAG3* and *VSG* pseudogenes. Around half of all BESs also contain an upstream region with an additional BES promoter and encode a single gene, *ESAG10*. Why these genes are located, and maintained, at these privileged and unusual genomic loci is unknown, and remains a topic of debate. The use of RNA Polymerase I, as opposed to RNA Polymerase II, to transcribe the active BES may be linked to the higher processivity (around 10-fold) of the former (Pays *et al.* 1990); thus enabling the production of greater amounts of *VSG* mRNA, and maintain levels of this abundant protein, a theory which may also be applicable to the ESAGs (Pays *et al.* 2001).

Specialised metacyclic expression sites (MESs) are the sole loci of *VSG* transcription for the quiescent, mammalian infective lifecycle stage found in the tsetse fly salivary gland lumen. Six MESs have been identified, and the *VSG* encoded at five of these sites is known to be expressed *in vitro* (Kolev *et al.* 2012; Cross *et al.* 2014). Again, telomere adjacent *VSG* is transcribed from these sites by RNA Polymerase I but, unlike BESs, these expression sites do not encode intervening *ESAGs* or 70-bp repeat regions (Figure 2), and are the only example of monocistronic transcription in trypanosomes (Alarcon *et al.* 1994; Ginger *et al.* 2002; Cross *et al.* 2014). Despite the active MES lacking *ESAGs*, these genes are transcribed in the metacyclic stage (Graham and Barry 1991; Helm *et al.* 2009; Christiano *et al.* 2017). This may

actually represent transcription of non-expression site genes related to *ESAGs* (*GRESAGs*), which will be discussed in more detail later. Reduction in the length and number of coding genes of MESs, compared to BESs, is hypothesised to require fewer resources, for example RNA Polymerase I complexes and trans-splicing of mRNA, therefore reducing the burden for MES activation in these non-dividing cells (Christiano *et al.* 2017).

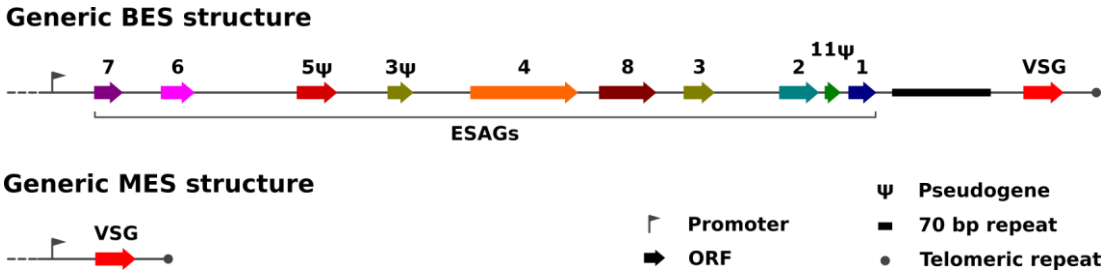


Figure 2. Generic bloodstream and metacyclic VSG expression site structures. Diagrammatic representation of the sites of *VSG* transcription in the bloodstream and metacyclic lifecycle stages. BESs are heterogenic but share a similar architecture, with varying repertoires of *ESAGs* separating the promoter and telomere proximal *VSG*, whereas MESs are more compact and have a much simpler structure, lacking *ESAGs* and 70-bp repeats. Not drawn to scale.

1.2.3. VSG switching

In order to evade the host humoral response, the VSG coat must be periodically switched for an antigenically distinct variant, which can occur via a number of different mechanisms. Frequency of VSG switching is a delicate balancing act: a sub-population of cells with distinct VSG coats must be maintained to replace those cleared by the humoral response against the previous coat, but switching must not occur too often to prevent exposure of a vast repertoire of variants to the host immune system, and immunity generated. This topic has been researched and reviewed extensively (Horn 2014; Hovel-Miner *et al.* 2015) and will only be covered briefly here. Transcription is initiated at all BESs, but quickly silenced in all but one, the active BES, resulting in expression of a single *VSG* gene (Vanhamme *et al.* 2000; Kassem *et al.* 2014). Switching of the actively transcribed *VSG* occurs by four known mechanisms: *in situ* switch, telomere exchange, duplicative gene conversion and mosaic gene formation (Figure 3). Transcription, or *in situ*, switching relies on the silencing of transcription of the active BES, and

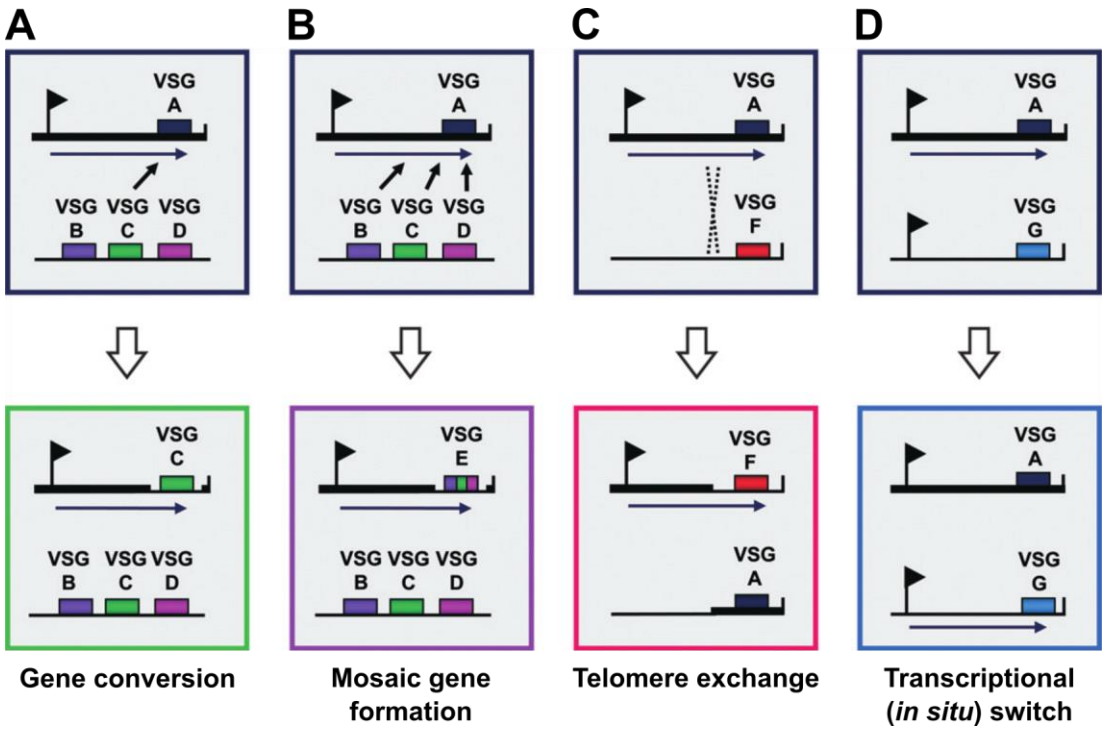


Figure 3. *T. brucei* variant antigen switching mechanisms. Flags represent BES promoters and arrows indicate active transcription. Coloured box borders correspond to the expressed VSG. **A**, Gene conversion is the most common switching mechanism *in vivo* and relies on DNA recombination of an intact *VSG* gene into the active expression site from the silent repertoire. The previously active gene is concurrently lost. **B**, Mosaic gene formation involves recombination of multiple *VSG* gene fragments into the active BES to create a novel gene, and also results in deletion of the previously active gene. **C**, In telomere exchange, or crossover, introduction of a new *VSG* into the active BES is accompanied by the telomere associated with this gene, while the previously active *VSG* is moved to a silent telomere. **D**, Transcription from the active BES is terminated, and is initiated at a previously silent BES. Known as an *in situ* switch. Adapted from Vink *et al.* 2012.

activation of another, resulting in expression of a different VSG and ESAG repertoire (Bernards *et al.* 1984; Myler *et al.* 1984). Telomere exchange involves DNA recombination between the active BES and an inactive BES or chromosome end, but appears to be a rare event (Pays *et al.* 1985; Rudenko *et al.* 1996; Hertz-Fowler *et al.* 2008). Finally, gene conversion and mosaic gene formation rely on recombination of whole *VSG* genes, or multiple fragments, into the active BES respectively, from the vast silent *VSG* catalogue, resulting in loss of the previously active *VSG* gene (Bernards *et al.* 1981; Pays *et al.* 1983; Roth *et al.* 1989; Kamper and Barbet 1992). Of these two, the former appears to be the predominant mechanism of switching, while the latter plays a role in generating *VSG* diversity and prolonging infection (Robinson *et al.* 1999; Mugnier *et al.* 2015). Antigenic variation has

been observed to occur in a hierarchy, with complete subtelomeric *VSGs* dominating early infection (Morrison *et al.* 2005). It is unknown if this represents transcriptional (*in situ*) switching, or gene conversion between the active BES and telomere-adjacent functional *VSG* genes on silent BESs or other loci, such as MCs, which appear to be a reservoir for functionally competent *VSG* genes (Cross *et al.* 2014). Activation of intact *VSG* genes in silent subtelomeric arrays follows, and is superseded by mosaic gene formation from pseudogene fragments later in infection (Marcello and Barry 2007). Control of monoallelic VSG expression and switching mechanisms is tightly regulated and has been linked to DNA replication, nuclear organisation and epigenetic factors (Glover *et al.* 2013b; Horn 2014; Li 2015; Hovel-Miner *et al.* 2015; Duraisingh and Horn 2016; Devlin *et al.* 2017).

1.3. Functions of ESAGs and related genes

ESAGs are characterised by their genomic localisation and co-transcription with *VSG* from the active BES in bloodstream form (BSF) parasites exclusively (Cully *et al.* 1986). These genes form 12 distinct families that are encoded in various combinations on the highly pleomorphic BESs. Each family also encompasses non-BES (chromosome-internal and subtelomeric) genes homologous to the BES *ESAGs*. These are known as genes related to *ESAGs* (*GRESAGs*) and, like all other protein-coding genes in eukaryotes, are transcribed by RNA Polymerase II (Alexandre *et al.* 1990). Unlike ESAGs, many GRESAG family members are thought to be constitutively expressed in both the BSF, and the procyclic form (PCF) insect lifecycle stage (Pays *et al.* 1989; Alexandre *et al.* 1990; Graham and Barry 1991). BES-encoded *ESAGs* and non-BES *GRESAGs* represent the two arms of *(GR)ESAG* families, which vary dramatically in size and diversity (Berriman *et al.* 2005; Hertz-Fowler *et al.* 2008; discussed in more depth later). Intra-family *ESAG* sequences are highly conserved, while *GRESAG* family members are often more divergent. The functions of the majority of (GR)ESAG families remain a mystery despite a great deal of interest in the field. The privileged genomic location of *ESAGs*, their co-transcription and co-switching (in the case of *in situ* switching) with *VSG*, as well as their BSF

specific expression, lead to theories that these genes may encode parasite virulence factors, or enable adaptation to the bloodstreams of different mammalian hosts. The latter fuelled by the evidence that *in situ* switching occurs early in infection, and a hypothesis that a BES encoding host specific ESAGs would provide an advantage and be selected at this stage (Marcello and Barry 2007). The roles of GRESAG family members are less well understood; research of these families is complicated by lack of adherence to the correct nomenclature, with GRESAG family members often referred to as ESAGs. Furthermore, although these genes represent 12 distinct families, they are often grouped as a single entity, making contributions of individual families difficult to decipher. Intra-family sequence similarity also introduces errors when mapping transcriptomic data. *ESAG* coding sequences were generated by sequencing the BESs of the common laboratory Lister 427 strain, while the reference genome sequence is that of the TREU 927 strain, in which the BESs have not been sequenced to completion (Berriman *et al.* 2005; Hertz-Fowler *et al.* 2008). So *ESAG* transcripts from 427 cells may erroneously map to *GRESAGs* of the 927 genome, potentially leading to artefactual interpretations. Recently a group has shown that transcriptomic data can be accurately mapped even between highly conserved *ESAG* sequences, and has used this technique to re-analyse published datasets (Hutchinson *et al.* 2016). Application of this mapping method to transcriptomic research, and consistency in nomenclature, will help clarify literature and aid in the elucidation of the roles of these families. In this section I will discuss what we know about each family and present evidence that these families may be involved in a number of aspects of parasite biology.

1.3.1. The *T. brucei* transferrin receptor

Iron is essential for parasite survival, and is scavenged from the host bloodstream as a complex with transferrin (Schell *et al.* 1991; Steverding 1998). The *T. brucei* transferrin receptor (TfR), a heterodimer formed of ESAG6 and ESAG7, is the most well studied *ESAG* product (Steverding *et al.* 1994; Chaudhri *et al.* 1994; Ligtenberg *et al.* 1994; Salmon *et al.* 1994;

Steverding *et al.* 1995). These homologous proteins share similar N-termini encoding a signal peptide, but divergent C-termini, with ESAG6 encoding a GPI-anchor addition sequence used to attach the TfR to the plasma membrane (Pays *et al.* 1989; Schell *et al.* 1991). Both subunits are heterologously glycosylated and, while this does not appear necessary for dimerisation or transferrin binding, modelling of the interaction between TfR and VSG suggests that it maintains the accessibility of transferrin without dispersing VSG molecules and exposing invariant antigens (Figure 4; Maier and Steverding 2008; Mehlert *et al.* 2012). Amino acids 205-215 and 223-238 of ESAG6 and ESAG7, respectively, are responsible for transferrin interaction (Salmon *et al.* 1997). Based on homology of both proteins with the N-terminal domain of VSG, these residues form surface exposed loops to generate a ligand-binding domain at the receptor's most membrane distal point (Blum *et al.* 1993; Salmon *et al.* 1997; Steverding 2000). One transferrin molecule interacts with the heterodimeric complex, but ESAG6 and ESAG7 homodimers are incapable of binding transferrin (Chaudhri *et al.* 1994; Ligtenberg *et al.* 1994; Salmon *et al.* 1994; Steverding *et al.* 1995). Immunoelectron microscopy has localised the TfR to the parasite flagellar pocket, where binding to host transferrin can occur (Ligtenberg *et al.* 1994; Salmon *et al.* 1994). Endocytosis traffics the receptor and ligand to endosomes, where the low pH releases transferrin to be degraded in the lysosome, while TfR is recycled back to the pocket to continue the uptake cycle (Steverding 2000; Kabiri and Steverding 2000).

Each *(GR)ESAG6* and *(GR)ESAG7* family encode a single non-BES copy. However, TfR appears to be expressed exclusively from the BESs, with 20% of all *ESAG6* transcript originating from inactive BESs (Pays *et al.* 1989; Ansorge *et al.* 1999). Starving the parasite of iron through the use of chelators, or growth in sub-optimal serum (discussed later), results in up-regulation of the TfR and its escape from the flagellar pocket onto the cell body (Fast *et al.* 1999; Mußmann *et al.* 2003; Mußmann *et al.* 2004). However, recent work has identified that cell body-localised receptor is not able to bind transferrin, and may represent ESAG6 homodimers (Tiengwe *et al.* 2017). Either *ESAG6* or *ESAG7* can be successfully knocked out of the

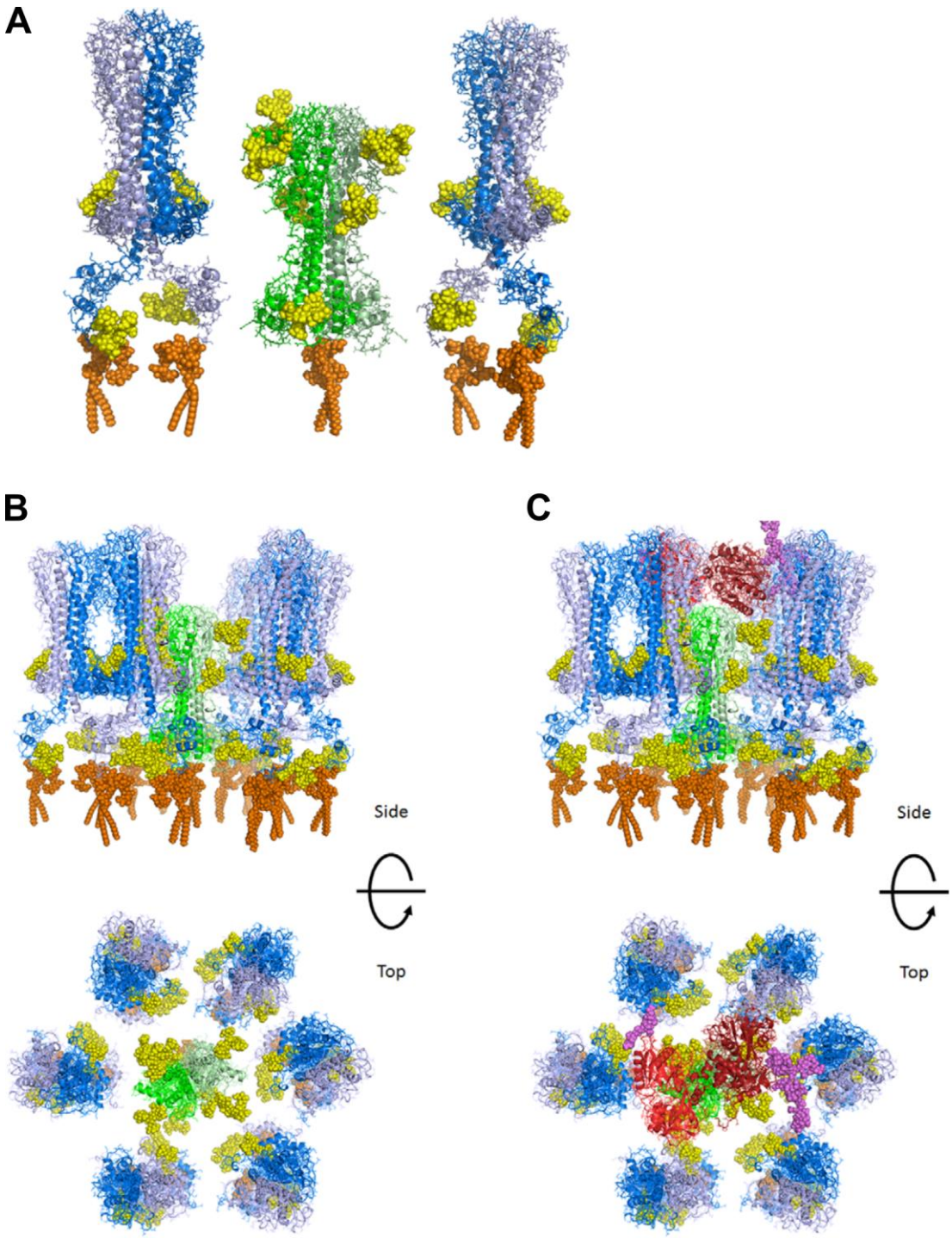


Figure 4. Modelling of TfR and its interaction with transferrin and the VSG coat. A, *T. brucei* heterodimeric TfR flanked by two VSG MITat1.2 homodimers. Homology between ESAG6/7 and the VSG N-terminal domain was used to produce this model structure. TfR and VSG are decorated with N-glycans and C-terminal GPI-anchors, one on each VSG monomer and ESAG6. **B**, Arrangement of TfR and VSG in the parasite coat. This model indicates that the receptor may be recessed into the coat, potentially to shield it from view of the host immune system. **C**, Interaction between transferrin and TfR can occur within the VSG coat. Key: VSG, blue; TfR, green; VSG and TfR N-glycans, yellow; GPI-anchor, orange; transferrin, red; transferrin N- and O-glycans, purple. Adapted from Mehlert *et al.* 2012.

active BES, but removal of both simultaneously has not been observed, suggesting that it is essential for parasite survival. This is in line with reports of RNAi-mediated TfR knockdown resulting in growth defects (Batram *et al.* 2014; Tiengwe *et al.* 2016).

1.3.2. Other ESAGs with less well defined function

Definitive roles for the other ESAGs are not as well understood as that for the TfR, and have generally received less attention. *ESAG4* encodes a receptor-type adenylate cyclase, with a large extracellular N-terminal region, and shorter intracellular C-terminal catalytic domain separated by a membrane-spanning helix (Pays *et al.* 1989). Enzymatic activity of ESAG4 was confirmed by functional complementation of adenylate cyclase-deficient yeast mutants (Ross *et al.* 1991; Paindavoine *et al.* 1992). The (GR)ESAG4 family encompasses a number of non-BES encoded receptor-type adenylate cyclases that are expected to share a similar structure to ESAG4 (Alexandre *et al.* 1990; Naula and Seebeck 2000; Salmon *et al.* 2012a). *GRESAG4s* are known to be transcribed in both BSF and PCF (Pays *et al.* 1989; Alexandre *et al.* 1990, 1996), and some family members have been shown to have lifecycle stage-specific expression (Saada *et al.* 2014; Shimogawa *et al.* 2015). ESAG4 and GRESAG4s localise to the flagellar membrane, shown by immunoelectron microscopy and immunofluorescence microscopy, respectively (Paindavoine *et al.* 1992; Saada *et al.* 2014), while native fluorescence microscopy of green fluorescent protein (GFP)-tagged ESAG4 shows a flagellar membrane and flagellar pocket localisation (Gadelha *et al.* 2015). In fact, multiple GRESAG4s are present at the BSF cell surface at any one time (Bridges *et al.* 2008), and PCF-specific GRESAG4s appear to localise to distinct flagellar membrane subdomains (Saada *et al.* 2014). These apparent receptors are hypothesised to have a role in cAMP signalling, and require ligand binding for dimerisation and catalytic activity (Naula *et al.* 2001; Bieger and Essen 2001), with ligand specificity thought to be conferred by their divergent N-terminal domains (Paindavoine *et al.* 1992; Alexandre *et al.* 1996; Seebeck *et al.* 2001).

As mentioned, the definitive role of this (GR)ESAG family is unknown, but some activities have been elucidated. ESAG4 is capable of modulating the host's innate immune system by inhibiting tumour necrosis factor alpha production, suggesting a role as a parasite virulence factor (Salmon *et al.* 2012b). This feature was thought to occur upon phagocytosis and lysis of *T. b. brucei* in the host's liver (Salmon *et al.* 2012b), but the discovery of a GRESAG4 in highly fuseogenic vesicles released by the parasite may provide another route of adenylate cyclase delivery to host cells (Szempruch *et al.* 2016). Knockdown of ESAG4 and the two most closely related GRESAG4s (Tb927.11.17040 and Tb927.10.16190, one of which is up-regulated in the PCF; Saada *et al.* 2014) impaired cytokinesis, a phenotype rescued by introduction of re-coded *ESAG4* (resistant to knockdown; Salmon *et al.* 2012a). While removal of *ESAG4* from the active BES (of the AnTat 1.1E strain) did not affect the growth or infectivity of cells, it was accompanied by up-regulation of the same two *GRESAG4s* (Salmon *et al.* 2012a). These results demonstrate the essentiality of, and a redundancy between, this (GR)ESAG4 sub-family. Interestingly, knockdown of these ESAG4-like GRESAG4s did not induce the same defect in PCF cells (Saada *et al.* 2014). Given that ESAG4 is exclusive to BSF, this implies a different function or requirement between lifecycle stages. In fact, knockdown of the two ESAG4-like GRESAG4s together, or another PCF-specific GRESAG4, appears to impact social motility of PCF cells, generating a 'hypersocial' phenotype; while ablation of another PCF-specific GRESAG4 reduced cell motility (Saada *et al.* 2014). Distinct expression profiles, localisations, and potential ligands of GRESAG4s suggests that this family are directly, or indirectly, involved in a number of different functions which include cytokinesis, social motility and modulating the host immune response.

The *(GR)ESAG10* family encode putative folate and biopterin transporters. Trypanosoma are auxotrophic for both molecules, and must therefore obtain them from the environment (Ouellette *et al.* 2002). ESAG10 is homologous to a *Leishmania donovani* biopterin/folate transporter BT1, and is predicted to encode 12 transmembrane domains, but its ligand has not been experimentally tested (Lemley *et al.* 1999). A role as a folate transporter

appears likely, as this activity was identified for the three most ESAG10-like GRESAG10s (Dewar *et al.* 2016). A cell body and flagellar pocket localisation has been shown for one of these GRESAG10s when tagged with GFP (Gadelha *et al.* 2015). Unusually, *ESAG10* appears to be transcribed from BESs during the insect lifecycle stage (Hutchinson *et al.* 2016), suggesting that silencing of the upstream BES promoter may not be as strict in PCF as BSF cells; the implications of this are unknown.

1.3.3. A number of ESAGs localise to the host-parasite interface

Another factor contributing to the hypothesis that the ESAGs may be parasite virulence factors is that the majority are predicted to encode membrane-associated domains, including signal peptides, GPI-anchor addition sequences and membrane-spanning helices (Figure 5A). This localisation has been confirmed for a number of ESAGs of unknown function, as well as those discussed above. GFP-tagged GRESAG5 has been shown to localise to the flagellar pocket and endosomal system (Gadelha *et al.* 2015). Members of the *(GR)ESAG5* family encode two BPI (bactericidal/permeability-increasing protein)/LBP (lipopolysaccharide-binding protein)/PLUNC (palate, lung and nasal epithelium clone)-like domains (Barker *et al.* 2008), which have diverse functions in other eukaryotes, but their role in trypanosomes is unknown. RNAi targeting *ESAG5* did not induce a phenotype (Batram *et al.* 2014). However, this experiment was conducted in cells with an active BES encoding an *ESAG5* pseudogene.

Along with families discussed so far, ESAG2, GRESAG11 and ESAG12 were found to be significantly represented in a *T. b. brucei* cell surface proteome, and were shown to have a range of cellular localisations when visualised by endogenous locus tagging with GFP (Figure 5B; Gadelha *et al.* 2015). Cell body localisations were identified for ESAG2 and GRESAG11, with the latter also decorating the flagellar membrane, which is consistent with a potential role at the host parasite interface (Gadelha *et al.* 2015). While ESAG12 localised to the endosomal system, and may be present at the surface at low levels (Gadelha *et al.* 2015). ESAG1 and ESAG3, which were not significantly enriched in the proteome, showed signal consistent with an

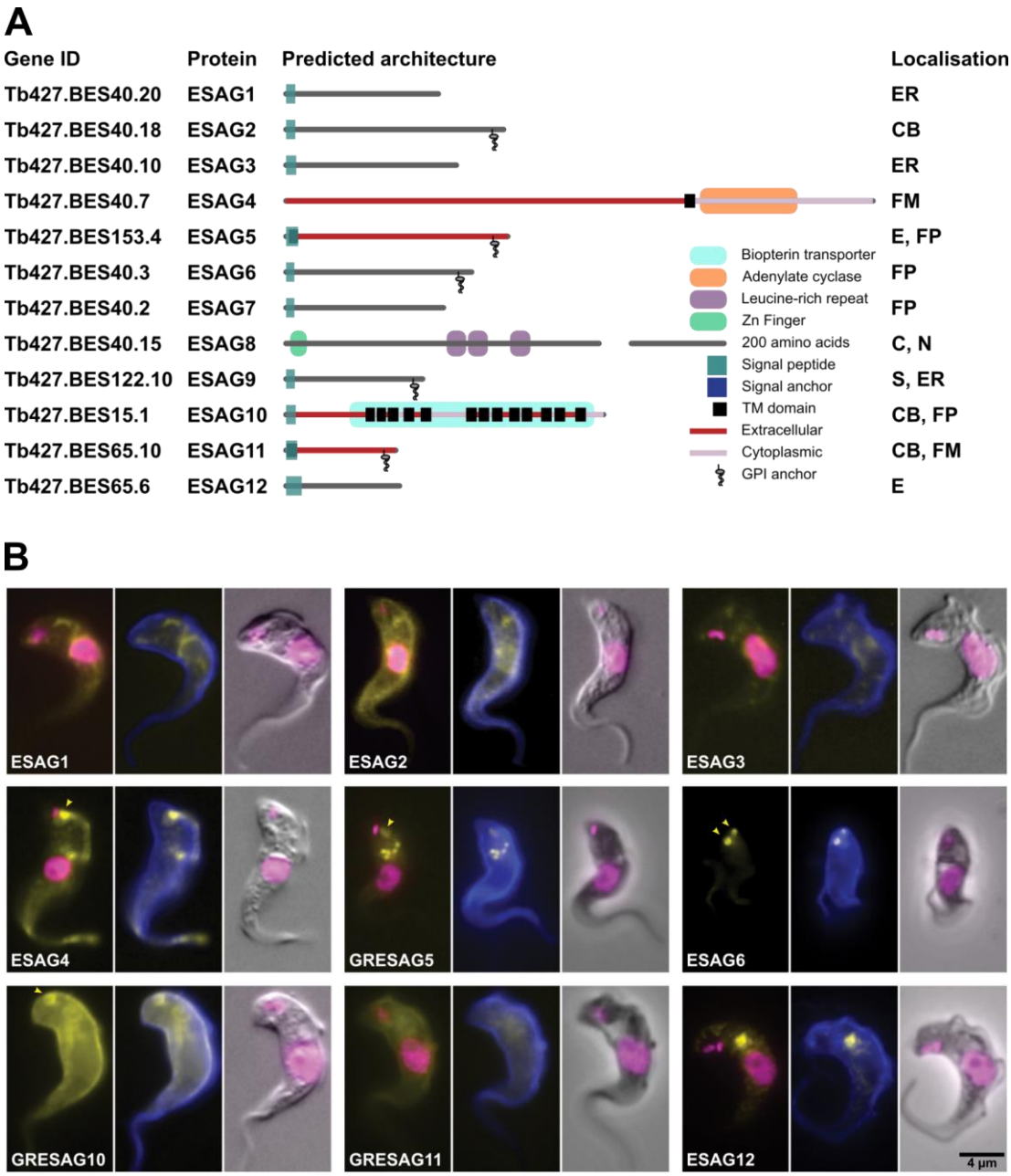


Figure 5. The majority of ESAGs encode membrane-associated domains and have a range of localisations. A, Diagrammatic representations of ESAG protein architecture. Predicted features and Pfam domains are annotated. Cellular localisations are indicated on the right; code: C, cytoplasm; CB, cell body; E, endosomes; ER, endoplasmic reticulum; FM, flagellar membrane; FP, flagellar pocket; N, nuclear; S, secreted. Protein feature prediction performed with SignalP, PredGPI and TMHMM. Specific ESAGs were modelled as a representative of their family (gene IDs shown). References within text. **B**, Native fluorescence microscopy of ESAGs and GRESAGs tagged at the endogenous locus with GFP shows that most (GR)ESAG families encode surface-associated proteins. Signal from GFP is shown in yellow. Cells have been counterstained with concanavalin A (ConA, blue) and DAPI (magenta). The flagellar pocket is indicated by yellow arrowhead. Adapted from Gadelha *et al.* 2015.

endoplasmic reticulum localisation when GFP-tagged (Figure 5B; Gadelha *et al.* 2015). The functions of all of these proteins remain to be elucidated, but knockdown of (GR)ESAG1, (GR)ESAG2, (GR)ESAG3 and (GR)ESAG12 family members is known to induced growth defects (Batram *et al.* 2014).

1.3.4. Potential roles for (GR)ESAG families in parasite transmission

BSF cells undergo differentiation into a transmission-adapted quiescent 'stumpy' form upon reaching a critical cell density (Reuner *et al.* 1997). This is driven by a quorum sensing mechanism whereby 'slender' proliferative cells respond to accumulation of an unknown small, soluble parasite-derived molecule known as stumpy induction factor (SIF) (Vassella *et al.* 1997). Accumulation of stumpy form cells limits the levels of parasitaemia while maintaining a transmission-competent cell reservoir. This prevents excessive damage to the host, prolonging the infection and therefore the opportunity for transmission (reviewed by MacGregor *et al.* 2012; Silvester *et al.* 2017). Differentiation to the stumpy form is accompanied by loss of monoallelic control of BES transcription, and a reduction in BES-associated transcripts, although transcription still initiates and is eventually silenced at all BESs (Amiguet-Vercher *et al.* 2004). Consistent with this finding, transcriptomic analysis of slender and stumpy form cells indicated a reduction in *ESAG* transcripts in the latter (Kabani *et al.* 2009), while transcription of some *(GR)ESAG* family members, specifically *GRESAG9* and *GRESAG4* genes, are up-regulated in the transmission-adapted stage (Jensen *et al.* 2009; MacGregor *et al.* 2014).

The *(GR)ESAG9* family is unusual in containing only one *ESAG9* gene encoded on BES11, but a number of non-BES family members (Berriman *et al.* 2005; Hertz-Fowler *et al.* 2008). Some of these *(GR)ESAG9s* of unknown function are specifically transcribed in stumpy form cells, and upon translation are secreted (Barnwell *et al.* 2010). (GR)ESAG9 therefore represents another potential family with a role at the host-pathogen interface. Differential expression of one GRESAG9 appears to be controlled by a short element in the gene's 3' UTR, which was also found to be responsive to cAMP analogues (Monk *et al.* 2013). A genome-wide RNAi screen recently

identified a negative regulator of *GRESAG9* transcription, REG9.1 (regulator of ESAG9), which, when ablated, resulted in spontaneous differentiation to stumpy form, and up-regulated transcription of *GRESAG9* and two other cell surface-associated gene families (Rico *et al.* 2017).

Transition from slender to stumpy is associated with a peak in adenylate cyclase activity (Mancini and Patton 1981; Vassella *et al.* 1997). Intracellular cAMP increases in slender form cells between log phase growth and peak parasitaemia, then returns to normal upon differentiation (Mancini and Patton 1981). cAMP analogues were also able to stimulate transition from slender to stumpy form (Vassella *et al.* 1997). However, conflicting literature indicates that these effects are in fact mediated by cAMP hydrolysis products, and cAMP may not play a direct role in differentiation (Laxman *et al.* 2006; Laxman and Beavo 2007). Eight of the nine BESs shown to be active *in vitro* encode an *ESAG4* (Hertz-Fowler *et al.* 2008), so a tempting hypothesis would be that silencing of BES transcription, associated with transition to stumpy form, is responsible for the observed drop in adenylate cyclase activity. ESAG4 protein levels reduce by 92% upon transition to stumpy form, yet ablation of ESAG4 and the two ESAG4-like GRESAG4s did not affect differentiation (Salmon *et al.* 2012a), so their role in this process may be indirect. Up-regulated transcription of a *GRESAG4* may represent a replacement for housekeeping adenylate cyclase activity of ESAG4, given that some members of this family have redundant functions (Salmon *et al.* 2012a). Expression of stumpy associated transcripts, *GRESAG9* and *PAD* (proteins associated with differentiation, discussed later) are also modulated by cAMP (MacGregor and Matthews 2012; Monk *et al.* 2013). Further dissection of the cAMP signalling pathway, and its role in BSF differentiation, is required to confirm a role for the (GR)ESAG4 family in this process.

1.3.5. A role for ESAGs in active BES transcription

Expression of VSG from a site other than the active BES has been shown to result in gradual attenuation of BES transcription and transition to a stumpy-like phenotype (Batram *et al.* 2014). These cells express PAD1 (a stumpy-form marker and transducer of the *cis*-aconitate differentiation signal

on their cell surface membrane; Dean *et al.* 2009), have a reduced growth rate, and become sensitive to *cis*-aconitate, all of which are reversible (Batram *et al.* 2014), which is not the case for true stumpy form cells (Dean *et al.* 2009). It has been suggested that this phenotype represents a safeguard capable of enabling cell survival or differentiation upon *in situ* switch to a non-functional BES, encoding either an ineffectual ESAG repertoire or defective VSG (Figure 6; Zimmermann *et al.* 2017). Three ESAGs have been proposed to be involved in BES attenuation signalling, and when their transcripts or encoded protein become limiting, the safeguard mechanism kicks in (Batram *et al.* 2014; Zimmermann *et al.* 2017). This is based on activation of a *PAD1* transcription reporter assay upon (GR)ESAG1, (GR)ESAG2 or ESAG8 knockdown (Batram *et al.* 2014). The assay utilises *GFP* flanked by *PAD1* UTRs, which are known to be involved in stage-specific transcription of this gene (MacGregor and Matthews 2012), so GFP expression is taken as a proxy for differentiation to stumpy-form (Batram *et al.* 2014). Therefore, ablation of these (GR)ESAGs appears to induce differentiation, and is accompanied by a growth defect (Batram *et al.* 2014). However, *ESAG1* and *ESAG8* can be knocked out of the active BES without consequence (Carruthers *et al.* 1996; Hoek and Cross 2001), so their role in sensing BES attenuation is unclear.

ESAG8 is the only protein of this group not to encode membrane-associated domains (Figure 5A). Instead, nuclear localisation signal, RING (really interesting new gene) finger and leucine-rich repeat (LRR) domains are encoded by this cytoplasmic and nucleolar protein (Hoek *et al.* 2000). Given its cellular localisation and protein-protein interaction-mediating domains, a role in controlling BES transcription was speculated for these proteins. However, while no such role has been confirmed, an interaction between ESAG8 and essential cytoplasmic protein PUF1 (of the Puf family of RNA regulatory proteins) has been shown (Hoek *et al.* 2002). Overexpression of PUF1 made parasites avirulent, but this does not appear to be dependent on interaction with ESAG8 (Hoek *et al.* 2002). Another potential role for this ESAG is one in mitochondrial activation, suggested after identification of ESAG8 interaction with PIE8 (protein interacting with ESAG8)

in a yeast two-hybrid screen (Pérez-Morga and Pays 1999). This protein's localisation to the mitochondrion coincided with its reactivation upon stumpy-to-procyclic form differentiation (Pérez-Morga and Pays 1999). However, interaction between these two proteins has not been shown in *T. brucei*, and ESAG8's specific role in mitochondrial activation has not been confirmed.

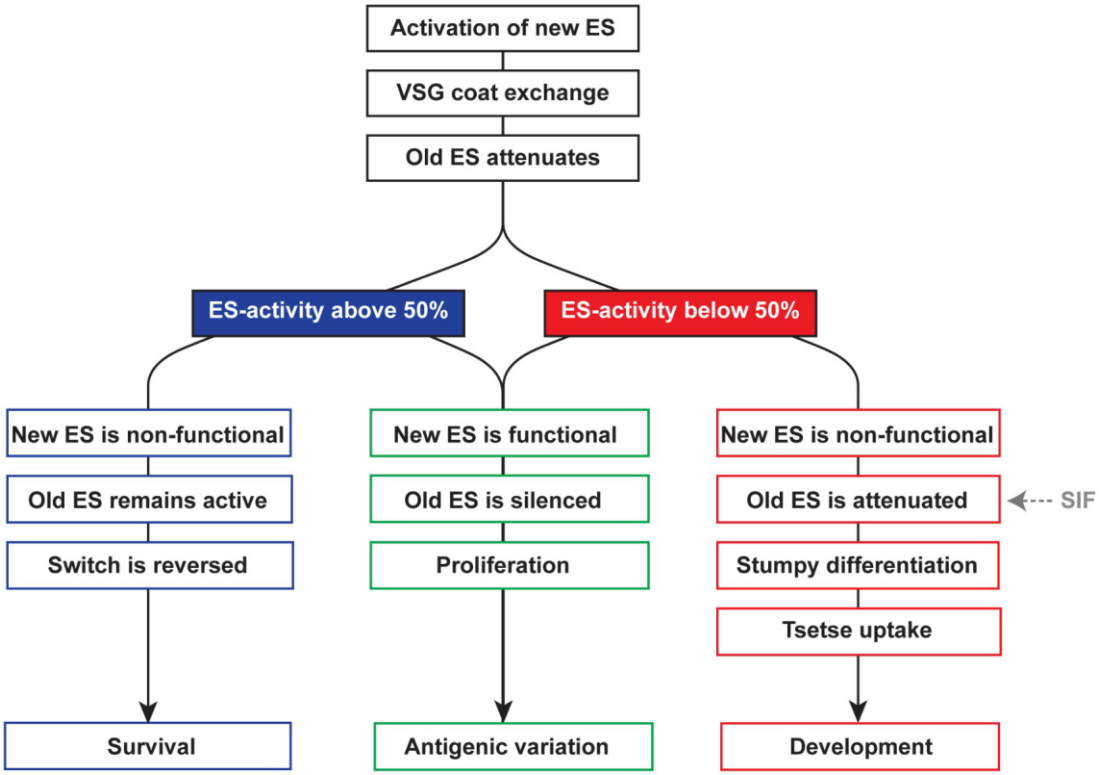


Figure 6. BES switching safeguard model. If *in situ* switching to a non-functional BES (or ES) occurs, this model proposes two mechanisms of parasite adaptation. When activation of a new non-functional BES occurs, but transcription from the old BES remains above 50%, then the switch can be reversed, enabling parasite survival (blue). Alternatively, if transcription from the previously active BES drops below 50%, and the new BES cannot sustain parasite survival, SIF-independent differentiation to stumpy form occurs, facilitating survival through transmission (red). Successful *in situ* switch to a functional BES is shown in green. ESAG1, ESAG2 and ESAG8 are proposed to have a role in sensing expression site activity at the step highlighted in blue and red. Adapted from Zimmermann *et al.* 2017.

1.3.6. Adaptation to different hosts

One common theory is that ESAGs play a role in adapting the parasite to the bloodstreams of different mammalian hosts, with specific ESAG repertoires providing a survival advantage in certain hosts (Pays *et al.* 2001; Pays 2006). This hypothesis fits neatly with the hierarchical switching of VSG: with *in situ* switching occurring early in infection, and then recombination-based switching mechanisms taking over, thus maintaining

expression of the same ESAG repertoire. If specific repertoires do provide an advantage in certain hosts, cells that have activated this 'host-specific' BES may outcompete those that have not, ensuring that this BES becomes dominant throughout the course of infection. This section will discuss the evidence for a role of ESAGs in host specificity, which is founded mostly on the *T. brucei* TfR and *T. b. rhodesiense*-specific protein, SRA (serum resistance-associated).

The finding that TfRs encoded on different BESs had variable affinities for transferrin from different mammals (Bitter *et al.* 1998) sparked great interest in the possibility that this could indicate an adaptation to life in different host bloodstreams, and a great deal of conflicting evidence has been produced since. An *ESAG6/ESAG7* pair are found adjacent to the promoter in every BES shown to be active *in vitro* (Hertz-Fowler *et al.* 2008), and relatively small changes in the primary sequence of these highly conserved proteins results in drastic changes in TfR ligand-binding affinity (Bitter *et al.* 1998; Gerrits *et al.* 2002). However, the high abundance of parasite-accessible host transferrin argues against the need for a high-affinity receptor, as even low-affinity receptors should be able to mediate uptake of sufficient iron for survival (Steverding 2003). Combined with the fact that around 20% of *ESAG6* transcript is derived from inactive BESs, with various affinities for transferrin, then high-affinity receptors may still be present in the flagellar pocket (Ansorge *et al.* 1999). Another hypothesis was that expression of a high-affinity TfR would maintain transferrin uptake even in the presence of a humoral response against the receptor (Gerrits *et al.* 2002), but titres of anti-TfR antibody generated during chronic murine infection were insufficient to deprive cells of iron (Steverding 2006). Despite these findings, cells expressing TfR with low affinity for canine transferrin displayed a number of different methods for adapting to this serum; these included switching to different BESs encoding TfRs with higher affinity for this transferrin variant, overexpression of TfR, recombination of TfR components from other BESs into the active BES, and increased transcription of *ESAG7* from silent BESs (Gerrits *et al.* 2002; Mußmann *et al.* 2004; Van Luenen *et al.* 2005). These findings strongly indicate that expression of high-affinity TfR is important for

parasite survival. However, no correlation between TfR affinity and host could be identified for two BESs preferentially active in human and mouse infection (Salmon *et al.* 2005). When reviewing this literature, it is worth noting that other ESAGs with unknown functions and ligands are expressed from the same loci; therefore, affinity for host transferrin may not be the only pressure exerted on selection of the active BES. For example, investigation of ESAG6, ESAG7 and ESAG3 expression during growth of *T. b. gambiense* in a range of sera identified that specific genes were preferentially transcribed when grown in pig and human sera (Cordon-Obras *et al.* 2013). These cells had a faster growth rate and greater transferrin uptake than those grown in goat sera, which transcribed a broader range of these three genes (Cordon-Obras *et al.* 2013). If host specific BESs do exist, this evidence suggests that *T. b. gambiense* lacks a 'goat-adapted' expression site.

A *T. b. rhodesiense*-specific ESAG, SRA, is encoded on a single BES, and provides resistance to the trypanolytic activity of human apolipoprotein L1 (APOL1; reviewed by Pays *et al.* 2014). Briefly, APOL1 in human serum is complexed with haemoglobin, among other components, and is endocytosed by the parasite upon interaction with the *T. brucei* haptoglobin-haemoglobin receptor in the flagellar pocket (Vanhollebeke *et al.* 2008). APOL1 forms anion channels in the lysosome membrane, causing osmotic swelling due to influx of chlorine ions, eventually resulting in cell lysis (Pérez-Morga *et al.* 2005). SRA localises to endosomes (Bart *et al.* 2015) where it interacts with APOL1, prevents pore formation and inhibits lysis (Vanhamme *et al.* 2003). Introduction of SRA alone into *T. b. brucei* was sufficient to induce resistance to human serum-mediated lysis (Xong *et al.* 1998). SRA is used as an example of a role for ESAGs in host adaptation as it is transcribed exclusively from a single expression site, the resistance-associated expression site (RES; Figure 7), which must be active for cells to survive in human serum (Xong *et al.* 1998). Removal of human serum resulted in spontaneous reversion of resistance, suggesting that the 'human-adapted' BES is selected against when *T. b. rhodesiense* is exposed to mouse or rat serum (in which these cells were grown; Xong *et al.* 1998). *T. b. gambiense*, the other human infective subspecies, does not encode SRA and has

developed constitutive resistance to APOL1-mediated lysis through distinct adaptations. These include reduced uptake of APOL1 due to down-regulated expression of (Kieft *et al.* 2010), and a single amino acid substitution in (DeJesus *et al.* 2013), the *T. b. gambiense* haptoglobin-haemoglobin receptor, as well as stiffening of the parasite endosomal membrane by a subspecies-specific protein, TgsGP (Uzureau *et al.* 2013; Capewell *et al.* 2013).

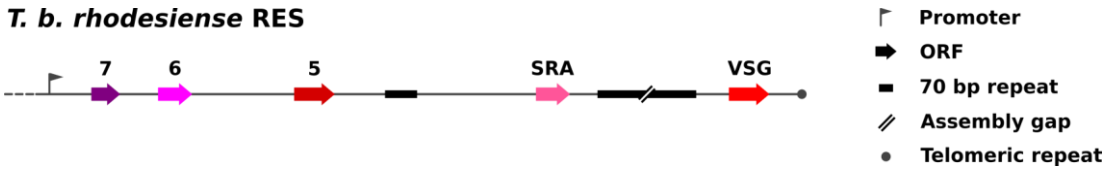


Figure 7. Structure of the *T. b. rhodesiense* resistance-associated expression site (RES) that confers resistance to human serum. The RES is a relatively small BES and has a reduced *ESAG* repertoire, encoding only *ESAG7*, *ESAG6*, *ESAG5* and subspecies specific-*SRA*. Not drawn to scale.

If the ESAGs do indeed have roles in parasite adaptation to different hosts, then a reasonable assumption is that species with wider host ranges would have evolved more diverse ESAGs. Relatively few studies have addressed this potentially interesting question. Examination of *ESAG6* sequences in *T. b. equiperdum* (narrow host range), *T. brucei* and *T. b. evansi* (both with broad host ranges) showed a correlation between host range and *ESAG* diversity, with the latter two species encoding a greater number of more diverse *ESAG6* variants than the former (Isobe *et al.* 2003; Witola *et al.* 2005). However, analysis of *ESAG6*, *ESAG5*, *ESAG2* coding sequences of *T. b. brucei*, *T. b. gambiense*, and *T. b. equiperdum* BESs, obtained through TAR cloning, contradicted these findings (Young *et al.* 2008). *T. b. gambiense*, with an intermediate host range, encodes fewer and more homogenous BESs than *T. b. equiperdum* (narrow host range), with only *ESAG2* displaying greater diversity in *T. b. gambiense* out of the three loci investigated (Young *et al.* 2008). *T. b. brucei* with the broadest host range has the most, and most diverse, BESs of the three parasites (Young *et al.* 2008). Recent evidence suggests that the genetic distance between *T. b. gambiense* and *T. b. brucei* strains is larger than that between *T. b. brucei* and *T. b. equiperdum* (Cuypers *et al.* 2017). Given that the

evolutionary dynamics of BESs are not fully understood, the higher degree of BES heterogeneity seen in *T. b. equiperdum* compared to *T. b. gambiense* may be due to a more recent divergence from *T. b. brucei*. With enough time, the BES diversity of *T. b. equiperdum* may reduce to reflect its narrow host range. However, while all three BES loci (*ESAG6*, *ESAG5* and *ESAG2*) show evidence of adaptive evolution in *T. b. brucei* and *T. equiperdum*, only *ESAG2* in *T. b. gambiense* appears to be under diversifying selection (Young *et al.* 2008). Complete sequencing of these TAR clones (and the BESs of other related parasites), would help clarify if BES number or *ESAG* diversity correlate with host range. This would also identify any novel ESAGs encoded by different species (such as SRA in *T. b. rhodesiense*), and certainly paint a clearer picture about how BESs have evolved since these parasites diverged from their common ancestor.

1.4. Evolution of *(GR)ESAG* families

Each *(GR)ESAG* family is comprised of BES-encoded *ESAGs* and homologous chromosomal-internal or subtelomeric *GRESAGs*. Just as sequencing of *T. b. brucei* BESs identified the repertoire of *ESAG* family members, sequencing of the genome revealed the true extent of these families (Berriman *et al.* 2005; Hertz-Fowler *et al.* 2008). *ESAGs* appear to be a *T. brucei*-specific adaptation, as they are not found in more distantly related parasites, while *GRESAG* sequences are (Jackson *et al.* 2013). This section will discuss what we know about the *(GR)ESAG* families of *T. brucei* and their evolutionary origins.

1.4.1. *(GR)ESAG* families of *T. brucei*

(GR)ESAG families are highly variable in terms of the number of genes they encompass and their diversity. While closely related *ESAG* family members are contained to the *T. brucei* BESs, *GRESAGs* are found throughout the genome in chromosomal-internal and subtelomeric loci, and can vary from just one family member to hundreds, including partial genes and pseudogenes (Table 1; Figure 8; Berriman *et al.* 2005). Albeit

Table 1. Size and distribution of *(GR)ESAG* families in *T. b. brucei*.

(GR)ESAG family	*ESAG* copies	*GRESAG* copies	Reciprocal monophyly	Distribution		
				T. congolense	*T. vivax*	*T. cruzi*
1	11	21	Yes			
2	11	18	Yes	+	+	
3	26	112	No	+	+	+
4	12	53	Yes	+	+	+
5	13	7	Yes	+	+	+
6	14	1	Yes	+	+	
7	16	1	Yes	+	+	
8	14	2	Yes	+	+	+
9	1	17	Yes			+
10	7	7	Yes	+	+	+
11	11	12	No	+	+	
12	5	1	Yes			

Copy numbers are provided for BES (*ESAG*), and non-BES (*GRESAG*), linked family members, including pseudogenes. Reciprocal monophyly is achieved where mutually exclusive clades of *ESAGs* and *GRESAGs* occur. The presence of *(GR)ESAG* family homologues in other trypanosomes (*T. congolense* IL3000, *T. vivax* Y486 and *T. cruzi* CL Brener) is marked by a plus sign. Data presented here is adapted from Hertz-Fowler *et al.* 2008; and Jackson *et al.* 2013.

maintaining a similar structure, BES *ESAG* repertoires are highly heterogeneous (Figure 9). These unusual loci undergo extensive recombination throughout their length, with increasing frequency towards the telomere (Young *et al.* 2008; Hertz-Fowler *et al.* 2008). Variable *ESAG* repertoires have been shaped by numerous duplications and deletions, making each BES a mosaic (Hertz-Fowler *et al.* 2008). Regardless of this, the majority of *ESAGs* are monophyletic, suggesting that they have a common origin and have evolved exclusively after transposition to BESs (Jackson *et al.* 2013). Some *(GR)ESAG* families encode *ESAG*-like *GRESAGs* (genes that are more closely related to *ESAG* family members than *GRESAGs* of the same family), representing either the *ESAG* ancestor, or secondary transposition from BESs (Jackson *et al.* 2013).

Only a single non-BES family member is encoded by the *(GR)ESAG6*, *(GR)ESAG7* and *(GR)ESAG12* families. *GRESAG6/7* are found as a tandem pair at a strand switch region of chromosome 7 and may have originated through secondary transposition from a BES. These genes do not appear to be transcribed (Pays *et al.* 1989), and it is unknown if their products would

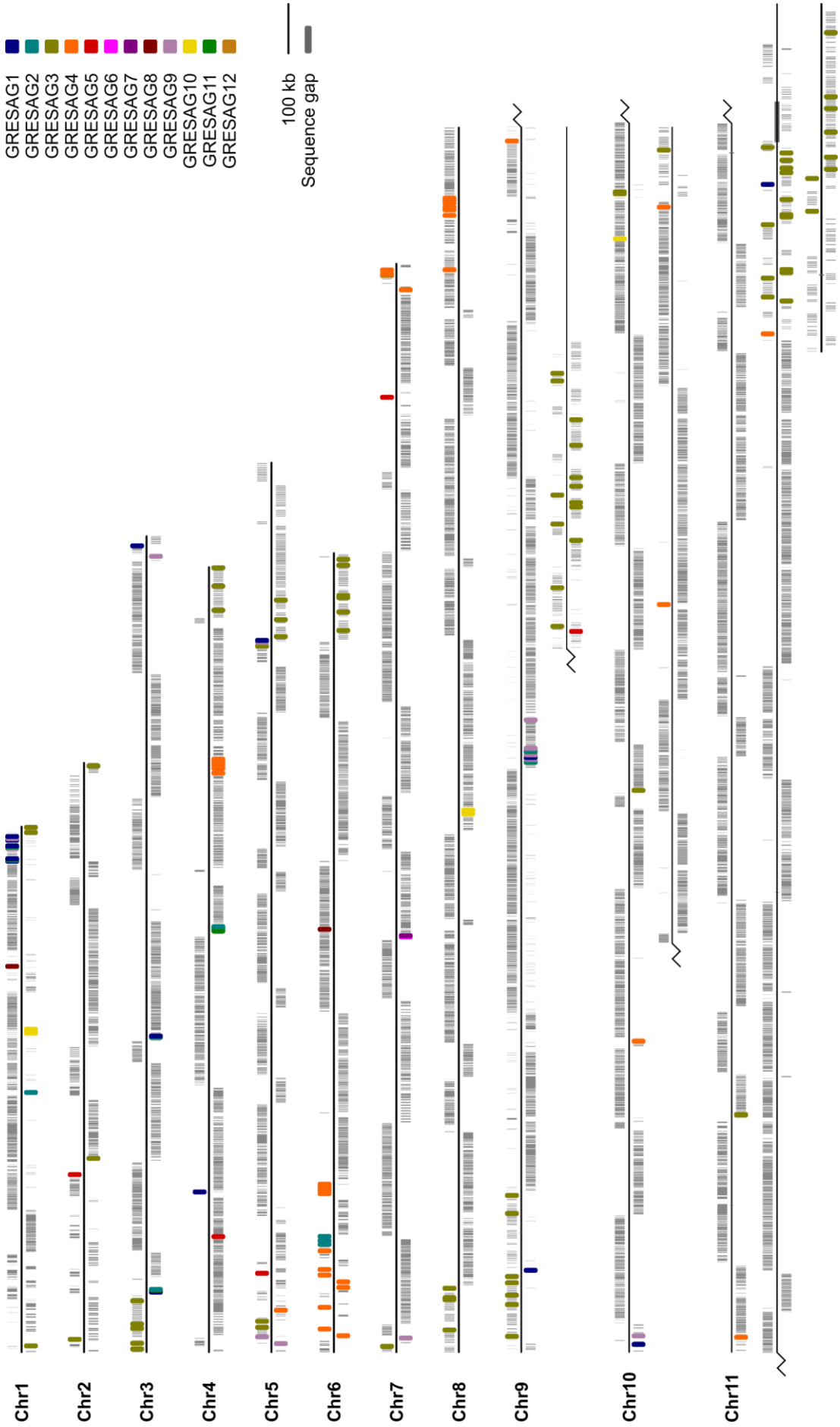

GRESAG1
GRESAG2
GRESAG3
GRESAG4
GRESAG5
GRESAG6
GRESAG7
GRESAG8
GRESAG9
GRESAG10
GRESAG11
GRESAG12
100 kb
Sequence gap
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11

Figure 8. *GRESAG* genes are often distributed in subtelomeric regions. These families are found throughout chromosomal-internal and subtelomeric loci on the 11 *T. brucei* MBC, and are commonly associated with other *GRESAGs* and *VSG* arrays. Grey and coloured bars represent open reading frames of protein coding genes. Bars above black lines indicate transcription to the right, bars below the lines shows transcription to the left. Chromosomes 9-11 have been split over multiple lines.

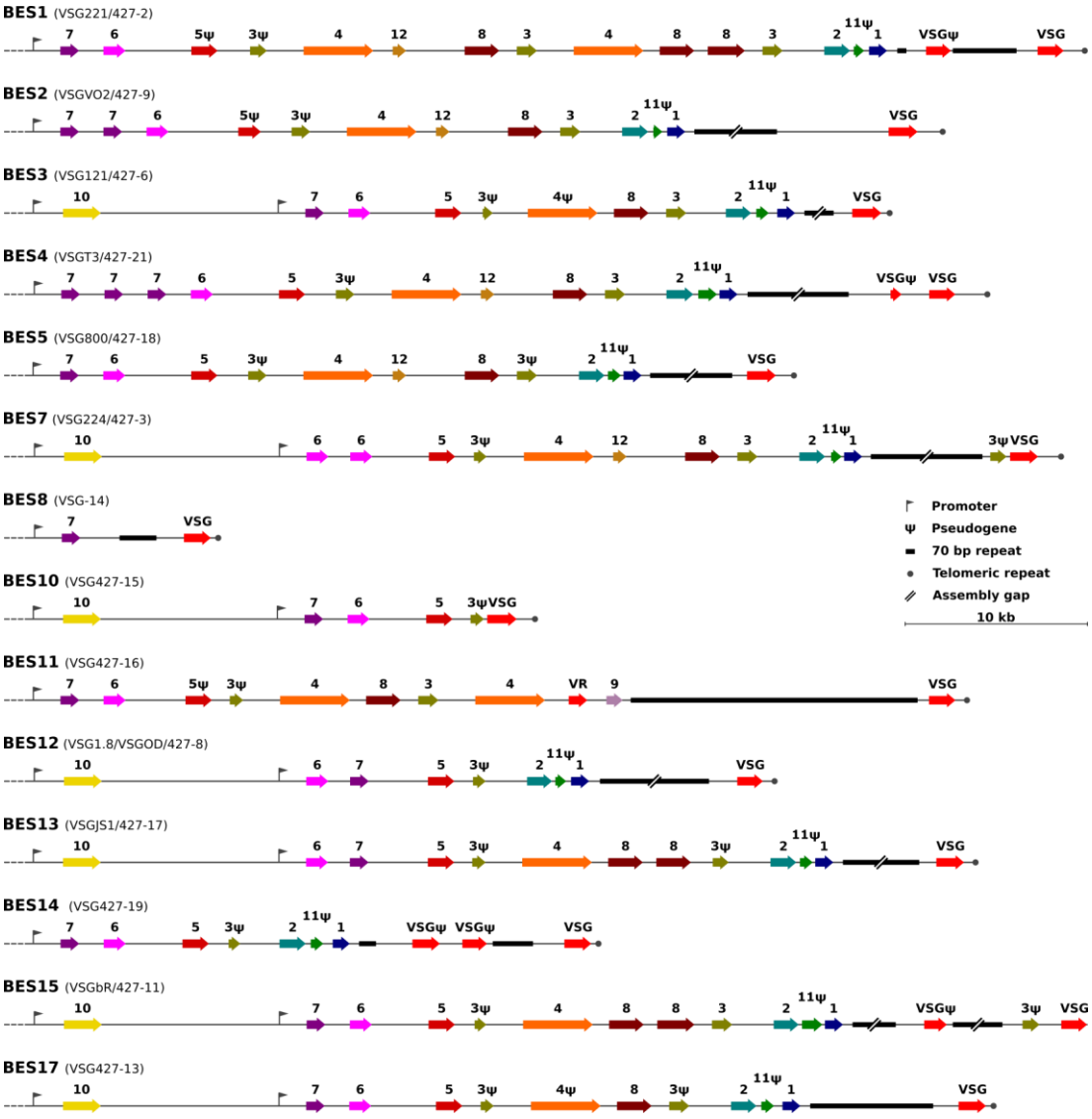


Figure 9. BESs are highly heterogeneous. Diagrammatic representation of the 14 unique *T. b. brucei* BESs. Size and *ESAG* repertoires vary dramatically between BESs, but an overall structure is conserved. Half of the BESs include a 13 kb upstream region containing an additional promoter. Some BESs, such as BES8 and BES10, have lost the majority of *ESAGs* and have not been successfully activated *in vitro* (Hertz-Fowler *et al.* 2008). BES-specific telomere proximal *VSG* gene names are shown in BES labels. Drawn to scale.

form a functional TfR. Given the essentiality of iron for parasite survival, this may explain the presence of *ESAG6/7* in every BES that has been shown to be active, including the *T. b. rhodesiense* RES (Schell *et al.* 1991; Xong *et al.* 1998; Hertz-Fowler *et al.* 2008). Homologues of *(GR)ESAG6* and *(GR)ESAG7* are encoded adjacent to procyclin loci, named *PAG1-2* and *4-5* (procyclin-associated genes; Koenig-Martin *et al.* 1992). Like the components of the TfR, PAG1 and PAG4 are predicted to encode a GPI-anchor addition site, while PAG2 and PAG5 do not. Unlike the BES ubiquitous *ESAG6* and *ESAG7*, *ESAG12* is found in only five BESs and was first identified when sequencing of the *T. b. brucei* BES repertoire was completed (Hertz-Fowler *et al.* 2008). The single *GRESAG12* is nearly identical to the *ESAG12s* and may have a redundant function given that BESs lacking *ESAG12* can be activated *in vitro* (Hertz-Fowler *et al.* 2008), and knockdown of the entire family induces a growth defect (Batram *et al.* 2014).

(GR)ESAG9 and *(GR)ESAG11* families are unusual in that only one BES encodes an *ESAG9*, while the majority of BESs encode *ESAG11*, but they are all pseudogenic (Hertz-Fowler *et al.* 2008). A number of full-length *GRESAG11* genes are found in the *T. b. brucei* genome, often associated with *GRESAG2* and *GRESAG1*, and in this way are similar to their BES-located relatives (Redpath *et al.* 2000). *ESAG11* genes are not monophyletic, which may be due to secondary transpositions from the BESs, making it difficult to determine which family members evolved first (Jackson *et al.* 2013). The conserved *ESAG2*/*ESAG11* pseudogene/*ESAG1* unit is replaced by *ESAG9* and *VR* (*VSG* related) genes on BES11 (Hertz-Fowler *et al.* 2008). The *(GR)ESAG9* family includes a number of non-BES members at subtelomeric loci, the products of which were discussed earlier. Infrequent identification of *ESAG9* on BESs may be linked with the stumpy-specific expression, and presumably function, of GRESAG family members. If the (GR)ESAG9 family do have a role in parasite transmission, then activation of the BES encoding ESAG9 may be beneficial in infection niches such as adipose tissue or skin where the parasite does not face the same pressures experienced during life in the bloodstream. Elucidation of the functions of these families may shed light on their unusual distribution amongst the BESs.

GRESAG1 and *GRESAG2* genes, some of which are pseudogenic, are often found in subtelomeres and strand-switch regions with *GRESAG11*. *GRESAG1s* are known to be transcribed in BSF cells (Morgan *et al.* 1996) and, again, may have a redundant function to *ESAG1*, given that it can be disrupted without consequence in the active BES (Carruthers *et al.* 1996), but knockdown of (GR)ESAG1 members induces growth defects (Batram *et al.* 2014). *GRESAG2* is known to be transcribed from a procyclin gene locus in PCF cells (Berberof *et al.* 1991). Knockdown of (GR)ESAG2 family members has been shown to induce a growth defect and significant loss-of-fitness in two independent investigations in BSF (Batram *et al.* 2014), but not PCF (Hutchinson *et al.* 2016); so the significance of GRESAG2 expression in this lifecycle stage is unknown. It is also unclear if the induced phenotype was caused by knockdown of ESAG2 or GRESAG2 proteins (Batram *et al.* 2014; Hutchinson *et al.* 2016).

The largest family, consisting mainly of gene fragments and pseudogenes, is *(GR)ESAG3*, which has a huge repertoire of subtelomeric members. This genomic distribution has drawn comparisons between the *(GR)ESAG3* family and subtelomeric *VSG* arrays (that also contain *GRESAG3* pseudogenes), in which the majority are pseudogenes, and a suggestion that this family undergoes antigenic variation (Jackson *et al.* 2013). Given the endoplasmic reticulum localisation of GFP-labelled ESAG3 (Gadelha *et al.* 2015), this hypothesis seems unlikely, but still begs the question as to why this family has undergone such a major expansion. Receptor-like adenylate cyclases of the (GR)ESAG4 family are also abundant and, unlike the (GR)ESAG3 family, encode mainly putatively functional members. *GRESAG4* genes are commonly found in tandem, and some family members have been characterised and divided into four sub-families, *GRESAG4.1-4* (Alexandre *et al.* 1990, 1996; Naula *et al.* 2001). The *ESAG4* expansion appears to have originated from a *GRESAG4* located in the subtelomeric boundary of chromosome 11, and the largest sub-family, *GRESAG4.4*, was expanded through segmental duplication of a tandem array of adenylate cyclase genes between chromosomes 4 and 8 (Jackson *et al.* 2013).

ESAG10 is encoded by around half of *T. b. brucei* BESs on an upstream region containing an additional promoter (Hertz-Fowler *et al.* 2008). It is likely then that there is a functional redundancy between (GR)ESAG10 family members. Three *(GR)ESAG10* genes on chromosome 8, closely related to *ESAG10*, encode folate transporters, and probably compensate for lack of *ESAG10* when a BES without this upstream region is active (Dewar *et al.* 2016). Four other *GRESAG10* genes (three on chromosome 1, the other on chromosome 10) encode putative biopterin transporters. Two chromosomal-internal *GRESAG8* genes encoding RING and LRR domains are annotated in the genome. These are more closely related to other conserved LRR-encoding genes, suggesting that they may represent more distant ancestors of *ESAG8* (Jackson *et al.* 2013). This implies that ESAG8 is expressed exclusively from the BES, which is consistent with the finding that *ESAG8* can be knocked out of this locus without consequence (Hoek and Cross 2001), and BESs lacking *ESAG8* can be activated *in vitro* (Hertz-Fowler *et al.* 2008). However, RNAi targeting all *ESAG8* sequences induced a growth defect (Batram *et al.* 2014), suggesting that ESAG8 expression is important for cell survival and that the two GRESAG8s cannot compensate for loss of ESAG8. These contradictory results indicate that further investigation is needed to understand the roles of this (GR)ESAG family.

(GR)ESAG5 is a relatively small but diverse family (Barker *et al.* 2008). The seven *GRESAG5* genes are located mainly at chromosomal-internal loci and do not appear to encode pseudogenes, while *ESAG5* is pseudogenic on two BESs (both of which can be activated *in vitro*; Hertz-Fowler *et al.* 2008). Intra-family redundancy is probable given that RNAi targeting ESAG5s and a closely related GRESAG5 (that probably arose through transposition from a BES; Jackson *et al.* 2013) did not affect cell growth (Batram *et al.* 2014). The *T. b. rhodesiense* RES encodes an *ESAG5* that clusters separately from those encoded by *T. b. brucei* BESs (Barker *et al.* 2008). If a role in adaptation is to be believed for the ESAGs, then perhaps this ESAG5 has evolved to function in human serum, given that it is preferentially expressed during infection of man (Xong *et al.* 1998), and has diverged from its (non-human infective) *T. b. brucei*-encoded relatives.

1.4.2. Evolutionary origins of *(GR)ESAG* families and their conservation between African trypanosomes

A number of *(GR)ESAG* families are conserved in other African trypanosomes, so may have a common origin. *ESAGs* appear to be a *T. brucei*-specific adaptation, as they are only distantly related to homologues in other species; while *GRESAG* family members are more well conserved (Jackson *et al.* 2013). *(GR)ESAG1* and *(GR)ESAG12* families show no homology outside of *T. brucei*, and may thus have evolved *de novo* in this parasite (Jackson *et al.* 2013). Some families, such as *(GR)ESAG3-5*, *8* and *10*, are conserved between the African trypanosomes *T. congolense* and *T. vivax*, and the more distantly related American trypanosome *T. cruzi* (Alexandre *et al.* 1996; Barker *et al.* 2008; Jackson *et al.* 2013). In all but (*GR)ESAG3*, these families encode common domains that are shared between larger protein families, suggesting a more ancient origin. *T. congolense* and *T. vivax* encode one and two *GRESAG3* genes, respectively, highlighting the scale of family expansion that has taken place in *T. brucei*. One *GRESAG5* locus on chromosome 4 is conserved between these African trypanosomes and *T. cruzi*, and may represent the ancestral *(GR)ESAG5* family progenitor. Whilst other *T. brucei* and *T. congolense GRESAG5s* cluster together, *T. vivax* genes are monophyletic, suggesting that expansion of this family happened after the split of *T. vivax* from the *T. brucei*/*T. congolense* common ancestor, but prior to divergence of these two species. A similar distribution is observed for *GRESAG4*, where *T. brucei* and *T. congolense* sequences are paraphyletic, while these families are monophyletic in both *T. vivax* and *T. cruzi* (Jackson *et al.* 2013). All *GRESAG10* genes are conserved between *T. brucei* and *T. congolense*, and the chromosome 8 *GRESAG10* locus is conserved between all four species (Jackson *et al.* 2013).

ESAG11 was first identified by screening a BSF cDNA expression library with antiserum raised against *T. brucei* ISG (invariant surface glycoprotein) (Redpath *et al.* 2000). Like *GRESAG11*, *ISG* is preferentially transcribed by BSF cells and localises to the whole cell surface membrane (Ziegelbauer *et al.* 1992; Redpath *et al.* 2000; Gadelha *et al.* 2015). The functions of both of

these families remain a mystery. *ISG* and *(GR)ESAG11* families are proposed to be related based on shared predicted secondary structure and distribution of conserved cysteine residues (Jackson *et al.* 2013). Homologues of the *ISG/(GR)ESAG11* family are also found in *T. vivax* and *T. congolense*. Although homologues of *GRESAG9* cannot be identified in other African trypanosomes, it has been suggested that this family is distantly related to the large and highly polymorphic *T. cruzi* mucin-associated surface protein (MASP) family (Barnwell *et al.* 2010). Like *GRESAG9*, these surface-associated proteins are predicted to encode a N-terminal signal peptide and C-terminal GPI-anchor addition site (Figure 5A); they are thought to participate in immune evasion and invasion of host cells, but a definitive function is unknown (De Pablos and Osuna 2012).

TfR components, (GR)ESAG6 and (GR)ESAG7, and their PAG relatives, are homologous to the N-terminal domain of VSG; specifically the a-type VSG lineage (a-VSG; Hobbs and Boothroyd 1990; Salmon *et al.* 1997; Jackson *et al.* 2012). This is one of two that generate antigenic diversity in *T. brucei*, and are grouped by the number and position of conserved cysteine residues (Carrington *et al.* 1991; Marcello and Barry 2007). Both the TfR and PAGs have homologues in *T. congolense*, which lacks a-VSG. The opposite is true for *T. vivax*, which encodes *a-VSG*-like genes, but not *TfR* genes or *PAGs*, suggesting that these families diverged from *a-VSG* in the *T. brucei*/*T. congolense* common ancestor, after the speciation of *T. vivax* (Jackson *et al.* 2012). *T. congolense TfR* genes are scattered throughout the subtelomeres, while all but one pair are found on *T. brucei* BESs. The encoded proteins are expected to perform the same function due to conservation of amino acids in the transferrin-binding domain. Another ESAG is thought to have evolved from VSG: the (GR)ESAG2 family are homologues of b-VSG, the second lineage responsible for antigenic diversity in *T. brucei*. Two *b-VSG*-like families are found in *T. congolense*; phylogenetic analysis confirms that *(GR)ESAG2* sequences form a clade within one of these families (*Fam13*), and are more closely related to *Fam13* than *T. brucei b-VSG* (Jackson *et al.* 2012). This suggests that, like the *TfR*, *(GR)ESAG2/Fam13* diverged from *b-VSG* in the common

T. brucei/*T. congolense* ancestor, as opposed to independent evolution post-speciation. While *(GR)ESAG2* genes are not thought to be involved in antigenic variation, other *Fam13* members encode published *T. congolense* VSG, and are variant antigens in this species (Jackson *et al.* 2012). *T. vivax* does encode a *b-VSG*-like family, but it is only distantly related to *b-VSG* subfamilies of *T. brucei* and *T. congolense* (Jackson *et al.* 2012). Finally, SRA is another example of a VSG-related ESAG with a novel function (De Greef and Hamers 1994; Campillo and Carrington 2003). Both TfR components, PAG, (GR)ESAG2 and SRA are thought to share a similar N-terminal fold to VSG, highlighting the functional plasticity of the VSG structure (Carrington and Boothroyd 1996; Salmon *et al.* 1997; Campillo and Carrington 2003; Jackson *et al.* 2013).

Generally speaking, the largest expansion of *(GR)ESAG* families has occurred in *T. brucei*, with the majority of families containing more members than those of *T. congolense* (*TfR* and (*GR)ESAG2*/*Fam13* are notable exceptions), and *T. vivax* encoding the smallest *(GR)ESAG* families. While the BES-linked *ESAG* family members represent a *T. brucei*-specific expansion, other *GRESAG* family members are found in chromosomal-internal and subtelomeric loci in all species. As discussed at the start of this chapter, subtelomeres provide a means to generate genetic diversity due to high rates of recombination, and these genomic regions are utilised by pathogens to evolve large and diverse gene families. The expansion of *(GR)ESAG* families may therefore be a result of their subtelomeric location. The differential expansion observed between the African trypanosomes could be linked to how these parasites generate diversity in their respective variant antigen repertoires, which appear to have higher recombination frequency in *T. brucei* and *T. congolense* than *T. vivax* (Jackson *et al.* 2012). Given the similar genomic location of *GRESAGs* and genes involved in antigenic variation in these species, it stands to reason that recombination-based diversification of antigen repertoires may also apply to the *GRESAGs*. Complete sequencing of the *T. congolense* and *T. vivax* subtelomeric regions, and identification of their VSG expression sites, could lead to the definition of more *(GR)ESAG* family members, or species-specific *ESAGs*, as

well as shedding light on the mechanisms underlying antigenic variation in these parasites.

1.5. Antigenic variation and linked genes in other parasites

Nine out of the 12 *T. b. brucei* ESAGs have cell surface localisations or are secreted. Although specific roles for the majority of these proteins are unknown, it is clear that the parasite has linked genes involved in antigenic variation with those encoding proteins that function at the host-parasite interface, through conservation of *ESAGs* and *VSG* at BESs. Numerous other parasites are known to undergo antigenic variation, but do they also link other genes with those involved in this process? Three examples of other parasites that may share this common theme with *T. brucei* are discussed here: the apicomplexa *Plasmodium falciparum* and *Babesia* spp., and a fellow *T. brucei* excavate, *Giardia lamblia*.

1.5.1. Apicomplexan parasites

A number of parasitic protozoa of the phylum apicomplexa have undergone expansion of subtelomeric gene families thought to be involved in mediating host-parasite interactions (Reid 2015). *P. falciparum* and *Babesia* spp. both live intracellularly in erythrocytes of their respective hosts, causing malaria and babesiosis, respectively. These parasites encode large gene families with roles in antigenic variation, and appear to link these genes with those encoding proteins that function at the host-pathogen interface. *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) is a cytoadhesin encoded by the highly pleomorphic *var* family encompassing around 60 genes (Gardner *et al.* 2002). This extremely diverse family is found in groups of 1-3 genes in most of the *P. falciparum* subtelomeres, and in tandem clusters at chromosomal-internal loci (Gardner *et al.* 2002). Mutually exclusive *var* gene expression, and transport of the encoded PfEMP1 variant to the erythrocyte surface, results in a specific adhesive phenotype; antigenic variation occurs through switching *var* gene expression to an immunologically distinct variant, and changes tissue tropism due to interaction of PfEMP1 with

specific host receptors (Scherf *et al.* 1998). Similar to *T. brucei*, control of *var* gene expression appears to be regulated by epigenetic factors (Duraisingh and Horn 2016), and recombination shapes diversity of the family (Kraemer *et al.* 2007). Two other variant gene families are found interspersed at subtelomeres, associated with *var* genes. These are the repetitive interspersed (*rif*), and the subtelomeric variant open reading frame (*stevor*) gene families, which encode ~150 RIFIN and ~30 STEVOR proteins, respectively (Cheng *et al.* 1998; Gardner *et al.* 2002). Both families are known be exported to the erythrocyte membrane, and undergo expression switching (Kyes *et al.* 1999; Fernandez *et al.* 1999; Kaviratne *et al.* 2002; Lavazec *et al.* 2006, 2007; Blythe *et al.* 2008). Also, both are thought to encode single-pass transmembrane proteins, and expose semi-conserved N-terminal regions to the external milieu (Bachmann *et al.* 2015). Along with PfEMP1, RIFIN and STEVOR mediate rosetting, whereby infected cells bind to uninfected erythrocytes forming rosettes (Rowe *et al.* 1997; Niang *et al.* 2009, 2014; Goel *et al.* 2015). This is thought to contribute to parasite virulence and immune evasion. Like *T. brucei*, *P. falciparum* maintains expanded gene families encoding variant antigens and proteins that mediate host-pathogen interactions at subtelomeric loci. In contrast to *T. brucei*, the families described here appear to combine antigenic variation with host-pathogen interaction functions, a need that may have been lost in the African trypanosome due to the shielding nature of the VSG coat.

The protein mediating antigenic variation in *Babesia bovis* is known as variant erythrocyte surface antigen (VESA), a heterodimer encoded by multi-copy *ves1α* and *ves1β* gene families and, as its name suggests, is transported to the surface of infected host erythrocytes (Allred *et al.* 2000; Xiao *et al.* 2010). 72 *ves1α* and 43 *ves1β* genes are found throughout the *B. bovis* chromosomes, the majority of which are arranged in a head-to-head orientation (Brayton *et al.* 2007). These units are known as loci of active transcription (LAT), with expression driven by bidirectional promoters located in the *ves1α-ves1β* intergenic region (Al-Khedery and Allred 2006; Wang *et al.* 2012). Transcription occurs from a single LAT at a time, and appears to involve chromatin remodelling (Zupańska *et al.* 2009; Huang *et al.* 2013).

B. bovis is the most studied species of the genus and thus represents the prototype for antigenic variation across *Babesia* spp.. Sequencing of *B. bigemina* and *B. divergens* genomes has revealed homologues of *ves1* genes that have diverged between species (Jackson *et al.* 2014). *B. bovis* and *B. bigemina* encode similar sized *ves1* repertoires, while the family is smaller in *B. divergens* and are typically found in subtelomeric arrays, as opposed to oriented in LATs (Jackson *et al.* 2014). The second largest *B. bovis* gene family, named *SmORF* (small open reading frame), is found closely associated with *ves1* genes throughout the genome (Brayton *et al.* 2007). This novel gene family is not represented in the other *Babesia* spp. discussed here, but diverse families with similar genomic distributions to *SmORF* genes, named *ves2*, are found in these species and are homologous to the N-terminal encoding sequence of *ves1* (Jackson *et al.* 2014). Phylogenetic analysis shows that, while all *ves1* genes share a common ancestor, *ves2* in *B. bigemina* and *B. divergens* has diverged independently from *ves1* genes in these species (Jackson *et al.* 2014). Similar to *VSG* of *T. brucei* and *var* of *P. falciparum*, diversity of *ves1* genes is driven by recombination that appears to occur less frequently in *SmORF* and *ves2* genes (although *SmORF* repertoires are known to vary between *B. bovis* strains; Ferreri *et al.* 2012; Jackson *et al.* 2014). Proteomic analysis of *B. bigemina* shows that multiple VESA2 (encoded by *ves2*) variants are expressed at the same time (or switch very rapidly) and that they may be more abundant than VESA1 in this species (Jackson *et al.* 2014). Multiple *SmORF* genes are also transcribed simultaneously (Ferreri *et al.* 2012). These results suggest that *ves1* and *SmORF*/*ves2* gene expression is not controlled by the same mechanism in spite of their genomic proximity. Functions have not been defined for SMORF or VESA2, but they are predicted to be secreted, and may therefore function at the host-parasite interface (Brayton *et al.* 2007; Jackson *et al.* 2014). It is also thought that SMORF may have evolved independently in *B. bovis* to perform the same function as VESA2 in *B. bigemina* and *B. divergens* (Jackson *et al.* 2014). Although these *Babesia* gene families are generally not subtelomerically located, it does appear again that this parasite links gene families encoding proteins involved in antigenic variation with those that function at the host-

parasite interface. The co-option of a variant antigen gene, in *ves1*, to perform a (presumably) different function in *ves2*, is analogous to the evolution of proteins with novel functions from VSG in *T. brucei*.

1.5.2. *Giardia lamblia*

G. lamblia is a parasite of mammals and colonises the small intestine lumen, where it lives exclusively extracellularly (Adam 2001). Here it persists and causes disease despite generation of host humoral response (Nash 1997). Immune evasion is achieved through antigenic variation, driven by switching the expression of variant-specific surface proteins (VSPs) that cover the parasite surface (Nash 2002). A single VSP is expressed at one time, controlled post-transcriptionally through an RNAi-like mechanism (Morrison *et al.* 2007; Prucca *et al.* 2008), and involves histone modification (Carranza *et al.* 2016). Controversy surrounds the *G. lamblia vsp* repertoire size, which varies from ~70-300 depending on the analysis, but inversely to *T. brucei* and *P. falciparum*, they are rarely found in subtelomeres (Morrison *et al.* 2007; Adam *et al.* 2010; Li *et al.* 2013). Instead, they are present in linear arrays of up to 10 genes, and in tandem pairs, in either head-to-head or tail-to-tail configuration (Adam *et al.* 2010). Diversity of the family appears to have occurred via duplication and recombination of *vsp* genes (similar to generation of variant antigen diversity in the other parasites discussed here) despite their mainly chromosomal-internal locations (Adam *et al.* 2010). Interestingly, the *vsp* family expansion appears to have occurred at a similar time to expansion of mammal species, and this co-evolution may represent the parasitic origin of *G. lamblia* (Sun *et al.* 2010). While the function of VSP outside of antigenic variation is not well understood, there is evidence of expression of host-specific variants independent of pressure exerted by the immune response, suggesting that VSP may have a role in mediating host-parasite interactions and adapting *G. lamblia* to different mammals (Singer *et al.* 2001). Despite the lack of genes linked with the *vsp* family, *G. lamblia* may have combined antigenic variation and host-adaptation functions in its variant antigen repertoire, forgoing the need to link these genes as previously discussed here.

1.6. Aims

The majority of (GR)ESAG families (9 out of 12) function at the host-pathogen interface. Specific family members, the ESAGs, are co-transcribed with the *T. brucei* variant antigen from privileged genomic loci, and *in situ* switching results in expression of a different VSG and ESAG repertoire. As discussed above, while definitive functions for the majority of these proteins are unknown, current evidence implies roles in parasite transmission, adaptation to different hosts and BES transcription attenuation, suggesting that these families are of central importance for parasite survival. However, some *ESAGs* can be removed from the active BES, supposedly functionally compensated by GRESAG family members, which complies with the activation of BESs lacking various *ESAGs*. The questions remain then, whether (GR)ESAG families are essential for parasite survival, and if intra-family redundancy exists. By answering this, and elucidation of the roles of these families, we may be able to better understand the linking of these families with the parasite variant antigen, which appears to also occur in other parasitic organisms. Research presented in this thesis seeks to investigate the essentiality of (GR)ESAG families, and disentangle the roles played by ESAGs and their GRESAG counterparts.

Chapter 2. Methods

2.1. Bioinformatics

Signal peptide and GPI-anchor addition sequence were predicted using SignalP v3.0b (Nielsen *et al.* 1997; Nielsen and Krogh 1998) and PredGPI (Pierleoni *et al.* 2008), respectively. Transmembrane domains were predicted using TMHMM v2.0c (Sonnhammer *et al.* 1998; Krogh *et al.* 2001). Multiple DNA sequence alignments were produced using MAFFT v7 (Katoh and Standley 2013), and edited with Jalview v2.0 (Waterhouse *et al.* 2009). Neighbour-joining clustering was generated from MAFFT alignments by Phylo.io 1.0.0 (Robinson *et al.* 2016).

2.1.1. PSI-BLAST

T. brucei TREU 927 and Lister 427 predicted proteome sequences were downloaded from TriTrypDB release 9.0 (http://tritrypdb.org), and Lister 427 BES sequences combined with the TREU 927 proteome to create a searchable database. To identify (GR)ESAG5 and (GR)ESAG10 family members, iterative searching of this custom made database was carried out to completion using Position-Specific Iterated (PSI)-BLAST (NCBI; Altschul *et al.* 1997) from the BLAST+ suite (Camacho *et al.* 2009) with an e-value threshold of 2×10^{-3}, using ESAG5 and ESAG10 proteins as input sequences (Tb427.BES15.4 and Tb427.BES15.1, respectively). BLASTn searches were performed against a transcript version of this database or the VSGnome (Cross *et al.* 2014) DNA database.

2.2. Molecular biology

Unless otherwise stated, *T. b. brucei* Lister 427 wild type genomic DNA was used as PCR template. Restriction enzymes were used at 20- and 10-fold excess to digest inserts and plasmids, respectively, and at 1.5-fold excess in plasmid linearisation for parasite transfection. Generally, plasmid and insert were used at 1:3 molar ratio in ligation reactions carried out with T4 DNA ligase at room temperature for 1 h, 12°C for 4 h or overnight at 4°C

(depending on the reaction complexity), cloned into XL1-Blue *Escherichia coli* by heat shock, and grown at 37°C for 16 h. Silica-adsorption-based DNA purification was routinely carried out on PCR products, digestion reactions and plasmid isolated from *E. coli*. Constructs were Sanger sequenced over the ligation site to ensure the expected *in silico* sequence had been obtained. Prior to transfection into *T. b. brucei*, plasmids were linearised with NotI (unless otherwise stated), purified by anion-exchange chromatography and precipitated in ethanol. Oligopeptides and synthetic DNA (in the form of gBlocks Gene Fragments) were purchased from Integrated DNA Technologies and are shown in Appendix 2; construct maps are shown in Appendix 3.

2.2.1. Production of tagging constructs

Plasmids to endogenous-locus tag ESAG5 and ESAG6 at the N-terminus, and ESAG10 at the C-terminus, with superfolder GFP (sfGFP) were derived from pSiS-HhsfG (Gadelha *et al.* 2015). For N-terminus tagging, ~200 bp fragments of the target gene 5’ UTR and 5’ ORF-end (downstream of sequence encoding predicted signal peptide and in frame with plasmid coding sequence) were cloned into XbaI-BamHI sites of pSiS-HhsfG. C-terminal tagging required fragments of the target gene 3’ ORF-end (upstream of the stop codon and in frame with plasmid coding sequence) and 3’ UTR to be cloned into HindIII-AvrII sites of pSiS-HhsfG.

2.2.2. Production of RNAi constructs

To knockdown chimeric proteins and (GR)ESAG families, ~450-1000 bp sections of target gene ORF were cloned into BamHI-ClaI/HindIII sites of the tetracycline-inducible RNAi construct p2T7-177 (Wickstead *et al.* 2002) that drives T7 RNA polymerase transcription of the cloned fragment from the MC. To target RNAi to divergent members of the (GR)ESAG5 and (GR)ESAG10 families, multiple *GRESAG5* or *GRESAG10* ORF targeting fragments (pan-family RNAi fragments 2-5) of ~150-250 bp were combined by fusion PCR and then cloned into a version of p2T7-177 where the selection marker *BLE* had been exchanged for *BLAST*, so it could be used in conjunction with

p2T7-177 containing *ESAG5* or *ESAG10* ORF fragments (pan-family RNAi fragment 1). This required amplification of the p2T7-177 rDNA promoter region with 5’ SphI and 3’ XbaI sites, and *BLAST* ORF (from the vector pEnG0B obtained from Bill Wickstead, University of Nottingham) with 5’ XbaI and 3’ RsrII sites, which were cloned into SphI/RsrII digested p2T7-177, creating p2T7-B177.

2.2.3. Production of KO fragments

DNA fragments designed to knockout sections of BES1 were produced by amplifying 5’ and 3’ targeting fragments and *BLAST* ORF (from pEnG0B) individually, then combined by fusion PCR. KO fragments containing *HYG* were produced the same way, with *HYG* flanked by actin UTRs amplified from pEnG0 (obtained from Bill Wickstead), then cloned into XhoI-BamHI sites of pBlueScript II SK (+) (Stratagene). Prior to transfection, these KO fragments were released from the plasmid through digestion with XhoI and BamHI.

2.2.4. Production of CreKO constructs

To produce CreKO_5’ and CreKO_3’ constructs to facilitate the Cre-recombinase based reduction of BES1, herpes simplex virus thymidine kinase gene fusion with *BLAST* (*TK-BLAST*) or *PURO*, both with actin 3’ UTR, were amplified from pMC-T7-HSVTK-B (D’Archivio and Wickstead 2017) and pEnG0P (obtained from Bill Wickstead), respectively. Along with synthetic DNA encoding *loxP* and aldolase 5’ UTR sequence, these PCR products were cloned into EcoRI-SpeI sites of pBlueScript II SK (+) to produce pBS+lox-aldolase5'-TK-BLAST-actin3'UTR and pBS+lox-aldolase5'-PURO-actin3'UTR.

DNA sequence to target homologous recombination into BES1 was cloned into KpnI-EcoRI sites of pBS+lox-aldolase5'-TK-BLAST-actin3'UTR and SpeI-NotI sites of pBS+lox-aldolase5'-PURO-actin3'UTR to produce various CreKO_5’ and CreKO_3’ constructs. These constructs were linearised with NdeI prior to transfection.

2.2.5. Diagnostic PCR

Diagnostic PCR to investigate integration of tagging constructs into the trypanosome genome, or identify the presence of *VSG* genes, was performed on genomic DNA extracted from specific cell lines indicated in the text. 2×10^5 mid-log phase parasites were resuspended in 120 µl of 50 mM NaOH, 2 mM EDTA and heated at 95°C for 10 min, then neutralised with 24 µl of 450 mM Tris-HCl pH 6.8 and used as PCR template.

2.2.6. RNA extraction, cDNA synthesis and amplification of the actively transcribed *VSG*

To identify the actively transcribed *VSG*, total RNA was isolated from 3×10^7 cells through silica-adsorption, and contaminating DNA removed by treatment with DNase I. cDNA was produced from 0.8 µg of total RNA incubated with 5 µM oligo(dT) primer, 10 mM DTT, 200U reverse transcriptase for 1 h at 42°C and enzyme inactivated through heating at 80°C for 5 min. cDNA was stored at -20°C.

The actively transcribed *VSG* was amplified from cDNA using primers that bind in the 5' miniexon (CAGGATCCGTGTTAAAATATATCA) and conserved *VSG* 3' UTR (CAGGATCCGTGTTAAAATATATCA), then Sanger sequenced using the same primers.

2.3. Cell lines and culture

All cell lines were derived from the *T. b. brucei* Lister 427 single marker line (known as S16) expressing transgenic T7 RNA polymerase and Tet-repressor protein from the tubulin locus (Wirtz *et al.* 1999), or wild type *T. b. brucei* Lister 427 cells (obtained from Mark Carrington, University of Cambridge). Appendix 4 outlines derivatives of these cell lines, constructs used to modify them, and the selection pressure used to maintain stable transfection. *T. b. brucei* EATRO 1125 clone AnTat 1.1 also used in this study were obtained from Luisa Figueiredo (University of Lisbon). BSF stage cells were used exclusively.

2.3.1. Axenic culture of *T. b. brucei*

Trypanosomes were routinely cultured in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 1 mM cysteine, 50 µM bathocuproine, 192 µM 2-mercaptoethanol, 90 µM cytosine, 1 mM hypoxanthine, 90 µM uracil and 15% foetal calf serum (Gibco), at 37°C and 5% CO_2. Medium was 0.22 µm filter sterilised prior to use. Cells were maintained at mid-log phase by regular dilution, such that population density remained between approximately 1×10^3 and 1×10^6 cells/ml. Transfected trypanosomes were cultured in medium supplemented with drugs to maintain selection. Markers and drug concentrations used in this study are found in Table 2. RNAi was induced by the addition 1 µg/ml of tetracycline to medium.

Table 2. Selection markers and drug concentrations used in this study

Selection marker	Drug	Concentration (µg/ml)
NEO (neomycin phosphotransferase)	G418	2.5
HYG (hygromycin phosphotransferase)	Hygromycin	5
BLE (bleomycin phosphotransferase)	Phleomycin	1
PUR (puromycin N-acetyltransferase)	Puromycin	0.1
BLAST (blasticidin S deaminase)	Blasticidin	5
HSV-TK (herpes simplex virus thymidine kinase)	Ganciclovir	100

2.3.2. Transfection

For stable transfection of *T. b. brucei*, 2.5×10^7 cells harvested from mid-log phase culture were resuspended with 10 µg linearised plasmid (or KO fragment) in 120 µl Tb-BSF buffer (90 mM $NaHPO_4$, 5 mM KCl, 0.15 mM $CaCl_2$, 50 mM HEPES pH 7.3; Burkard *et al.* 2011) and electroporation performed using Amaxa Nucleofector (Lonza) program Z-001. Cells were recovered for 8 h prior to addition of drug and aliquoted in 96 well plates at limiting dilutions to select for stable clonal transfectants. Transient transfection using circular plasmids followed the same procedure but did not receive drug, or plating to limiting dilution.

2.3.3. Growth rate analysis

An ATP-based luciferase assay was used to monitor cell growth in culture over 5-day periods. Cell lines of interest were diluted to 1×10^3 cells/ml in 1.5 ml wells on a 96 deep-well plate and were counted at 24 h intervals including t=0 h. Prior to reading, wells were resuspended and 100 µl transferred to black 96 well plates. 50 µl of CellTiter-Glo Reagent 2.0 (Promega) was added to cells, incubated for 10 min, and luminescence detected using a Glomax 96 Microplate Luminometer (Promega). A set of 8 ATP standards in √10 dilution series, from 316 nM to 100 pM, were included on each plate to normalise reads between time points. A cell standard √10 dilution series, from 3.16×10^6 to 1×10^2 cells/ml, was included on the t=0 h plate to convert arbitrary luminescence units obtained from the ATP standards to cell density, by calculating the average amount of ATP per cell. This relies on the linear relationship between luminescent signal and cell density in this range, setting the limit of detection. Cultures were subbed at t=72 h or t=96 h depending on cell density, to maintain log-phase growth. Short-term growth monitoring was conducted using an Improved Neubauer haemocytometer, with cell counts obtained every 8 h over a 24 h period.

2.4. Immunoblotting

Whole cell lysates were produced by incubating 2.5×10^5 cells/µl washed parasites in Laemmli buffer modified to pH 7.2 (2% SDS, 0.4 M 2-mercaptoethanol, 10% glycerol, 50 mM Tris-HCl) at 95°C for 5 min, 50°C for 10 min (when resolving for ESAG10-GFP), or room temperature for 30 min (when probing for PAD1), and stored at -80°C until needed. 2.5×10^6 cell equivalents per lane were separated by SDS-polyacrylamide gel electrophoresis (PAGE) on 4-20% Tris-glycine gels in 25 mM Tris, 250 mM glycine, 0.1% SDS buffer.

Proteins were electrophoretically transferred to 0.45 µm pore-size nitrocellulose membrane for 14 h at 0.04 A in 25 mM Tris, 192 mM glycine, 0.02% SDS and 10% methanol. Immobilised proteins were visualised by staining the nitrocellulose membrane with 0.1% Ponceau-S in 5% acetic acid

prior to blocking with 5% milk in buffer outlined in Table 3 for 30 min. Buffers used were Tris-buffered saline (TBS; 150 mM NaCl, 20 mM Tris-HCl pH 7.5), TBS-Tween (TBS-T; TBS with 0.1% Tween-20), high salt TBS-T (TBS-T with 0.5M NaCl) and phosphate-buffered saline with Tween (PBS-T; 140 mM NaCl, 10 mM $NaHPO_4$, 3 mM KCl, 0.1% Tween-20). Chimeric GFP-labelled proteins were detected with a mixture of two mouse anti-GFP monoclonal antibodies (7.1 and 13.1; Roche); clathrin heavy chain was detected using rabbit anti-TbCLH serum (obtained from Mark Field, University of Dundee; Morgan *et al.* 2001); PAD1 was detected using rabbit anti-PAD1 anti-peptide polyclonal antibody (obtained from Keith Matthews, University of Edinburgh; Dean *et al.* 2009); and TfR was detected using rabbit anti-TfR polyclonal antibodies (obtained from Piet Borst; Netherlands Cancer Institute; Bitter *et al.* 1998). Either horseradish peroxidase-conjugated goat anti-mouse IgG polyclonal or goat anti-rabbit IgG polyclonal antibodies (Sigma) were used as secondary antibody, and detected with chemiluminescence. Antibodies and concentrations used are shown in Table 3, along with buffers used to dilute them. Generally, membranes were incubated in antibody solution for 1 h, or with primary antibody for 16 h at 4°C to increase binding. Four washes of 5 min with rocking were performed prior to incubation of secondary antibody and detection. For reprobing, membranes were stripped with 62.5 mM Tris-HCl pH 6.8, 2% SDS and 100 mM 2-mercaptoethanol at 60°C.

Table 3. Antibody concentrations and buffers used in immunoblotting.

Primary antibody	Secondary antibody	Buffer
0.8 µg/ml mouse anti-GFP	1:5000 goat anti-mouse IgG in 1% milk	TBS-T
1:2000 rabbit anti-TbCLH	1:20000 goat anti-rabbit IgG in 1% milk	TBS
1:1000 rabbit anti-PAD1 in 5% milk	1:20000 goat anti-rabbit IgG in 1% milk	PBS-T
1:5000 rabbit anti-TfR in 1% milk	1:20000 goat anti-rabbit IgG in 1% milk	High salt TBS-T

2.5. Microscopy

For analysis of GFP-tagged protein localisation by native fluorescence microscopy, cells were harvested from mid-log phase cultures, washed twice in PBS plus 20mM glucose, allowed to adhere onto glutaraldehyde-

derivatised silanised glass slides for 2 min (at density of 2×10^7 cells/ml), fixed for 10 min in 2.5% formaldehyde, and mounted in a solution containing a DNA stain and photostabilising agent (1% 1,4-diazabicyclo(2.2.2)octane, 0.25 mg/ml 4',6-diamidino-2-phenylindole, 90% glycerol, 50 mM sodium phosphate pH 8).

For analysis of the actively expressed VSG, cells were prepared as above and after fixation, slides were incubated for 1 h with 1:500 dilutions of polyclonal rabbit anti-VSG221 (Gadelha *et al.* 2015) or monoclonal mouse anti-VSG121 (obtained from Miguel Navarro, The Institute of Parasitology and Biomedicine, Spain; Landeira *et al.* 2009) antibodies. Slides were then washed three times in PBS and incubated for 1 h with either goat anti-rabbit or anti-mouse IgG (depending on primary antibody used) polyclonal antibody conjugated to Alexa Fluor 568 (Invitrogen) at 2 µg/ml. Slides were washed as before and counter-stained with 2 µg/ml concanavalin A (ConA) conjugated to fluorescein isothiocyanate (FITC; Sigma) for 20 min prior to mounting as above.

Images were captured on a BX51 microscope equipped with a 100× UPlanApo objective (NA 1.35; Olympus) and CoolSnap-HQ (Photometrics) 6.45 µm/pixel CCD camera, and processed in ImageJ (National Institute of Health, USA; rsb.info.nih.gov).

2.6. Southern blotting

Cell plugs were produced by embedding 5×10^7 washed parasites in 0.8% low-melt agarose in TES buffer (120 mM NaCl, 20 mM EDTA, 25 mM Tris-HCl pH 8). Plugs were incubated in NDS buffer (500 mM EDTA, 1% N-lauroyl sarcosine, 10 mM Tris-HCl pH 9) with 1 mg/ml proteinase K for 24 h at 50°C, followed by NDS (pH 8) with 1 mg/ml proteinase K under the same conditions. Plugs were stored at 4°C in NDS (pH 8) until needed, and dialysed with 10 mM Tris-HCl pH 8 before use.

Agarose-embedded DNA was digested with SwaI, PvuI, BsiWI, or ApaI endonuclease and subjected to pulsed-field gel electrophoresis (PFGE) in a

contour-clamped homogeneous electric field electrophoresis apparatus (CHEF-DR III; Bio-Rad), loading DNA from 1.7×10^7 cells/lane. DNA separation was performed in 1% agarose (SeaKem Gold) in TB(0.1)E buffer (90 mM Tris borate, 0.2 mM EDTA, pH 8.2) held at 14°C for 20 h at 5.2 V/cm, with switching time ramped linearly 2–10 s and an included angle of 120°. DNA gels were stained in ethidium bromide and prepared for transfer by UV nicking (80 mJ, 250 nm UV) followed by equilibration in 0.4 M NaOH, 1.5 M NaCl and then transferred to positively-charged nylon membrane by capillary transfer in the same solution. After transfer, membranes were neutralised with 0.5 M Tris-HCl pH 7 and cross-linked (120 mJ, 250 nm UV). Fluorescein-labelled probes were generated by random priming from unlabelled *eGFP*, *sfGFP*, *HYG, BLAST, PURO, VSG121, VSG224* and *VSG221* coding sequences (amplified using primers in Appendix 2). Denatured template DNA (100 ng) were incubated for 5 h at 37°C with 0.1 mM dATP, 0.1 mM dCTP, 0.1 mM dGTP, 0.67 mM dTTP, 0.33 mM Fluorescein-dUTP, 2 µM random heptamers and 5 U Klenow fragment. Hybridisation was performed overnight in 1% SDS, 5% dextran sulphate, 10% blocking solution (Roche), 750 mM NaCl, 75 mM sodium citrate pH 7 at 60°C. Blots were washed to a stringency of 0.1% SDS, 30 mM NaCl, 3 mM sodium citrate pH 7 at 62°C, then blocked for 1 h in 10% blocking solution in BufferA (0.3 M NaCl, 0.1 M Tris-HCl pH 9.5). Blocked membrane was incubated with anti-fluorescein alkaline phosphatase-conjugated Fab fragments (Roche) at 1:5000 in BufferA with 5% bovine serum albumin in for 1 h, then washed in BufferA with 0.3% Tween20 prior to detection with chemiluminescence. For reprobing, membranes were stripped with just boiled 0.3% SDS, 0.3 M NaOH.

Chapter 3. A repertoire of GFP-tagged BES ESAGs

3.1. Introduction

Despite discovery of the first ESAGs occurring over 20 years ago, definitive functions of most of these gene families remain elusive. Along with their functional roles, another important question concerning the *ESAGs* is why the parasite has moved and maintains specific members of these families at a privileged genomic location: the VSG expression sites. As discussed previously, the ESAGs are hypothesised to adapt the parasite to different mammalian bloodstreams; however, non-BES encoded family members, the GRESAGs, are rarely considered when discussing these families. Little research has gone into dissecting these two arms of each (GR)ESAG family, comparing the roles of BES encoded ESAGs to their GRESAG counterparts. The work presented here aims to clarify this.

The study of protein function through ablation is well established. In trypanosomes, RNAi mediated gene silencing has been used to investigate numerous individual proteins and protein families (reviewed by Kolev *et al.* 2011; discussed in Chapter 4). In order to disentangle the roles of ESAGs and GRESAGs of the same family, we aimed to investigate the phenotype induced upon RNAi-mediated knockdown of the actively expressed BES-linked ESAG family member, and compare this to the phenotype produced when multiple (GR)ESAG family members are ablated. However, the high degree of sequence identity shared between *ESAGs* and *GRESAGs* makes it challenging to design an RNAi strategy to knockdown just one of these sets of genes. In fact, RNAi constructs that target (GR)ESAG families, or 'pan-family' constructs, can be produced easily and have been employed in other studies (Salmon *et al.* 2012a; Batram *et al.* 2014; Tiengwe *et al.* 2016). RNAi mediated knockdown of the actively expressed BES ESAG, without targeting GRESAGs of the same family, requires unique sequence of this gene to be targeted. To facilitate this, we planned to tag the active BES-linked ESAG at the endogenous locus with GFP, then knockdown chimeric protein using RNAi targeting the *GFP* ORF. In order to achieve this, a set of cell lines expressing GFP-tagged ESAGs from the active BES was required.

The ability to genetically modify bloodstream form *T. brucei* through integration of an exogenous construct via homologous recombination was first achieved in 1993 (Carruthers *et al.* 1993). Homologous recombination-mediated transfection of procyclic cells had been described earlier, but its application to BSF was limited by challenges in culturing this stage of the parasite lifecycle. The technique relies on the innate machinery used by cells to repair double strand breaks that often occur during DNA replication. Introduction of a linear construct flanked by endogenous sequence into the *T. brucei* cells via electroporation induces integration into genome at the targeted site. It has since been shown that efficiency of homologous recombination is dependent on the length and homology of the targeting sequence (Barnes and McCulloch 2007). This feature has made manipulation of the *T. brucei* genome a relatively simple task, and has facilitated a range of reverse genetics experiments. These include expression of transgenes (ten Asbroek *et al.* 1990), gene knockout (Sommer *et al.* 1996) and gene tagging, to enable downstream purification of protein (Estévez *et al.* 2001), or producing chimeras with fluorescent protein (Kelly *et al.* 2007). Endogenous locus tagging of genes enables the addition of protein coding sequence at the N- or C-terminus, while maintaining close to endogenous levels of expression. Tagging genes to encode a fluorescent protein (for example GFP) can be used to identify the cellular localisation of the resulting chimera, which can inform the function of the protein. Here I will use the additional sequence added by the GFP-tag to specifically target RNAi mediated knockdown of actively expressed BES ESAGs.

3.2. Completing a repertoire of GFP-tagged BES1 ESAGs

The lab recently developed a suite of vectors designed to tag proteins with predicted cell surface localisations (Gadelha *et al.* 2015). These vectors, pSiS and pSiG, tag genes at their endogenous locus, the resulting chimeric protein containing three HA (haemagglutinin) epitope tags and either eGFP (enhanced GFP) or sfGFP (superfolder GFP; Figure 10). The vectors also replace endogenous signal peptide or GPI anchor addition sequences to

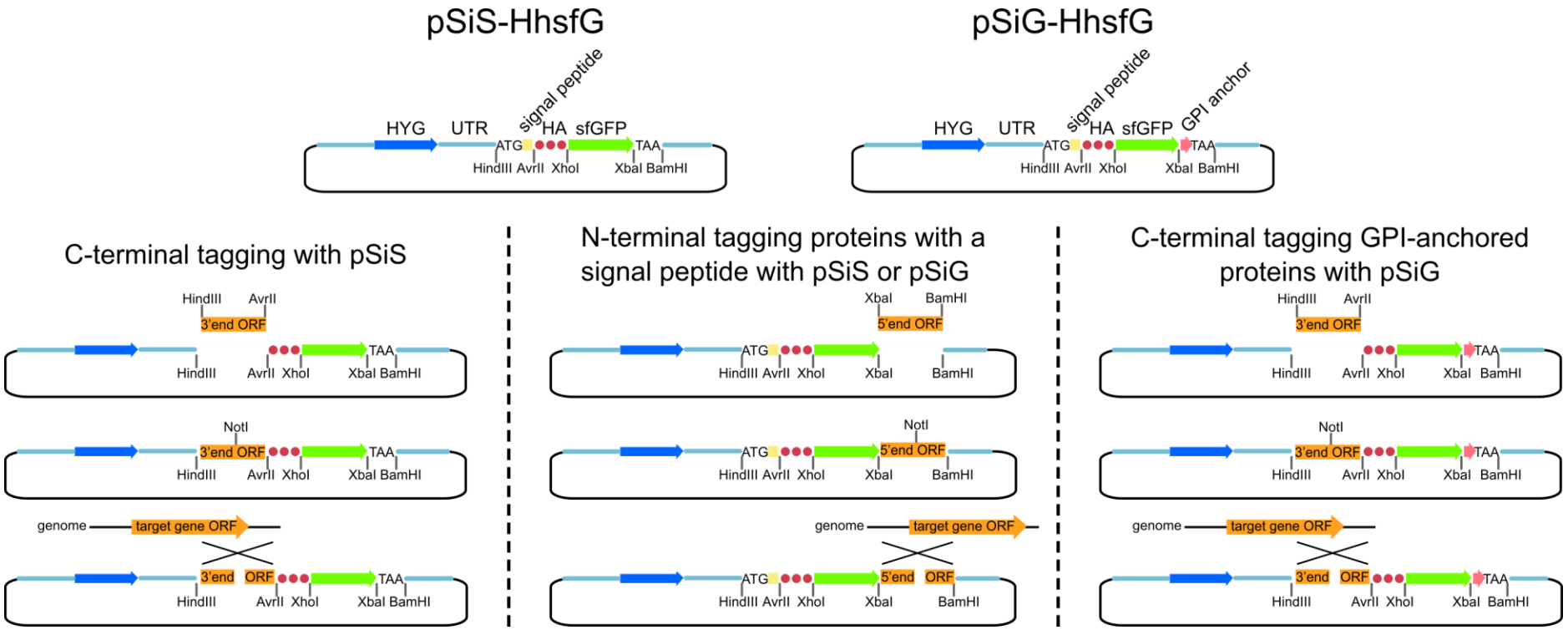


Figure 10. A genetic toolkit for membrane protein localisation. This diagram outlines the molecular biology strategy employed to tag proteins at the N- or C-terminus with three HA epitope tags and sfGFP. pSiS is used to tag membrane-associated proteins that do not encode a GPI-anchor addition sequence at the C-terminus. Proteins that do encode a GPI-anchor addition sequence can be tagged at the C-terminus with pSiG, where the endogenous GPI-anchor addition sequence is replaced by one encoded in the tagging construct. Both pSiS and pSiG can be used to tag proteins with a signal peptide at the N-terminus, replacing the native signal peptide. These vectors also introduce *HYG* into the genome as a selectable marker. Targeting sequences (orange) are amplified from genomic DNA by PCR. Primers used to amplify these fragments include tags encoding restriction endonuclease recognition sites to enable ligation into the constructs and to facilitate linearisation of the vector prior to transfection.

ensure that processing of the resulting polypeptide is not disrupted, and the native cellular localisation of the protein is maintained. These vectors were employed to validate the localisations of a set of proteins identified in the *T. brucei* cell surface proteome (surfeome; Gadelha *et al.* 2015). Integration into the parasite genome is facilitated by homologous recombination between ~200 bp targeting sequences. N-terminal tagging requires sequence homologous to the 5' UTR and N-terminus-encoding ORF (just downstream of any predicted signal peptide cleavage site) of the target gene; while C-terminal tagging utilises sequence homologous to the 3' UTR and C-terminus-encoding ORF (upstream of any GPI anchor addition sequence cleavage site, or upstream of the STOP codon in ORFs encoding predicted transmembrane proteins) of the target gene. Strategies for N- and C-terminal tagging with pSiS and pSiG are shown in Figure 10.

The BSF surfeome identified and validated by Gadelha *et al.* included a number of ESAGs and GRESAGs, which prompted the tagging and identification of cellular localisation of a representative of each family. (GR)ESAG families that were not identified, or not enriched, in this dataset were also included to ensure that the tagging constructs do not force non-surface proteins to this localisation, and provide a method of investigating the surfeome for false-positives. In (GR)ESAG families with a member encoded on BES1, this protein was tagged; while the GRESAG most similar to its family's ESAGs was tagged for other families (Table 4). As a rule, predicted transmembrane proteins were tagged at the C-terminus, while those encoding GPI-anchor addition sequences were tagged at the N-terminus. To complete this set, an N-terminally tagged ESAG6-expressing line was required.

Tagging construct pSiS was modified to N-terminally tag ESAG6 encoded on BES1, replacing the endogenous signal peptide coding sequence with that encoded on pSiS, three HA epitopes and sfGFP. The high degree of DNA sequence identity shared between *ESAG6* genes on different BESs means that the tagging construct can potentially integrate into a silent expression site. To ensure that the *ESAG6* gene on BES1 specifically is tagged, the construct needed to be transfected into BES1-active (expressing

the BES1 specific VSG: VSG221) S16 cells. Integration into the active BES ensures transcription of the construct's selectable marker (in this case hygromycin phosphotransferase (*HYG*)), and promotes survival of BES1-active transfectants. Should the construct integrate into an inactive BES, HYG would not be expressed, or would rely on leaky transcription, as in the case of *ESAG6* mRNA derived from "silent" BESs (Ansorge *et al.* 1999). The active BES of two S16 cultures was assessed by immunofluorescence microscopy using anti-VSG221 antibody (Figure 11). ESAG6 tagging construct was used to transfect the S16 line shown to have VSG221 present on the cell surface (herein referred to as VSG221-positive), and therefore BES1-active.

Table 4. The set of GFP-tagged (GR)ESAG cell lines generated by Gadelha *et al.*

Family member	Gene targeted	Termini-tagged	Fluorescent protein	Designation
ESAG1	Tb427.BES40.20	N	eGFP	GFP-ESAG1
ESAG2	Tb427.BES40.18	N	eGFP	GFP-ESAG2
ESAG3	Tb427.BES40.16	N	eGFP	GFP-ESAG3
ESAG4	Tb427.BES40.13	C	eGFP	ESAG4-GFP
GRESAG5	Tb427.7.6860	N	sfGFP	GFP-GRESAG5
ESAG8	Tb427.BES40.14	C	eGFP	ESAG8-GFP
GRESAG10	Tb427.08.3630	C	eGFP	GRESAG10-GFP
GRESAG11	Tb427.01.5110	N	eGFP	GFP-GRESAG11
ESAG12	Tb427.BES40.8	C	sfGFP	ESAG12-GFP

ESAG6 and *ESAG7* are related and therefore share a high degree of sequence identity, in both ORF and UTRs (Appendix 5). The targeting sequences used to tag *ESAG6* on BES1 are homologous with *ESAG7* sequence, so the tagging construct can integrate into either of these genes. To identify which had been tagged, positive transfectants were screened by diagnostic PCR. Primers that anneal in *sfGFP* (encoded in the tagging construct) and a divergent region of *ESAG6* or *ESAG7* ORF were used to amplify a product of ~800 bp if integration into either gene had taken place (Figure 12A). Genomic DNA purified from transfectants and parental cells was used as template in this diagnostic PCR reaction. Figure 12B shows that

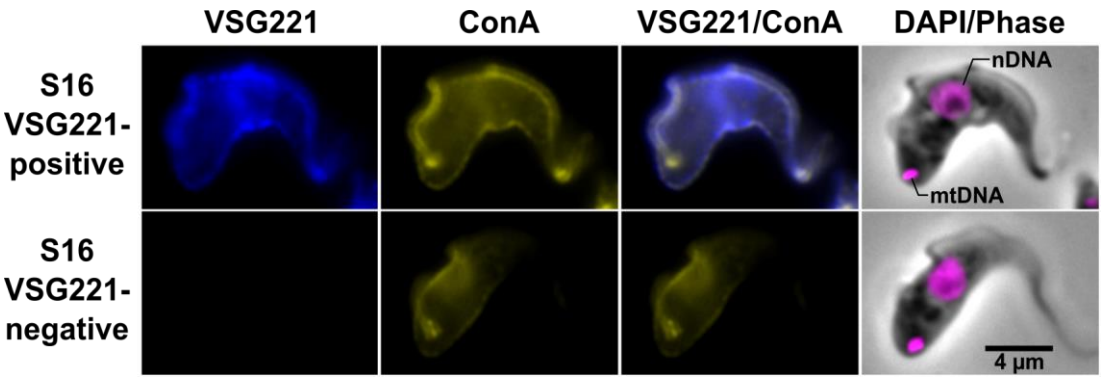


Figure 11. S16 cultures express different *VSG* genes. Immunofluorescence microscopy on two S16 cell cultures shows that one expresses VSG221 (VSG221-positive) and is BES1 active, the other is VSG221-negative with an unknown active BES. VSG221 detected using polyclonal antibodies against VSG221, blue; Concanavalin A (ConA; interacts with sugar-containing moieties that decorate VSG) used as a cell surface counterstain, yellow; DAPI labelled DNA, magenta. Nuclear (nDNA) and mitochondrial (mtDNA) DNA are indicated. Images are representative of the signal distribution observed for each cell line, all images processed to the same levels.

the construct has integrated into either, or both, *ESAG6* and *ESAG7* in 18 of the 29 clones tested.

Whole cell lysate of clones, identified as tagged on either ESAG6 or ESAG7 by diagnostic PCR (Figure 12), was investigated for expression of chimeric protein by immunoblotting with anti-GFP antibody (Figure 13A). GFP-ESAG6 (following the nomenclature outlined in Table 4) has a predicted molecular weight of 72 kDa, while GFP-ESAG7 is 68 kDa. A number of clones show signal of approximately the predicted molecular weight, with GFP-ESAG7 being consistently smaller than GFP-ESAG6 as expected. ESAG6 and ESAG7 are both predicted to be glycosylated, and ESAG6 modified by the addition of a GPI-anchor, which will add to their molecular weight and affect how they migrate in SDS-PAGE. Lysate of a line encoding C-terminally tagged ESAG6 (ESAG6-GFP, produced by the lab previously) was included as a control, and migrates to a similar size as GFP-ESAG6. GFP-ESAG6 clones 9 and 19, and GFP-ESAG7 clone 1 appear to be tagged on BES1 based on levels of chimeric protein. A number of clones screened do show reduced chimeric protein levels, in some cases this is below the limit of detection of the technique. Given that approximately 20% of *ESAG6* mRNA results from transcription of silent BESs (Ansorge *et al.* 1999), these

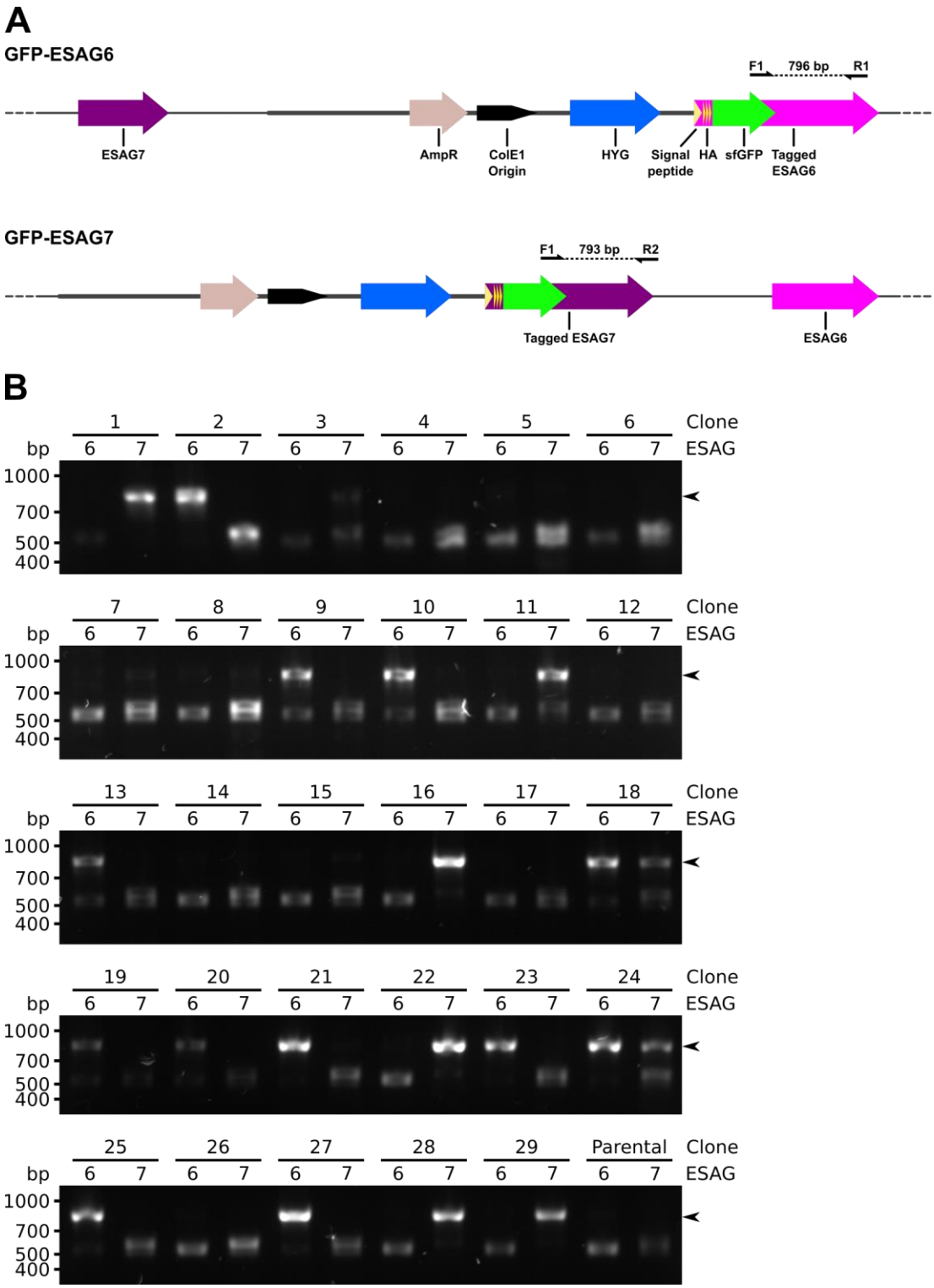


Figure 12. ESAG6 tagging construct can integrate into *ESAG6* or *ESAG7*. A, Schematic illustrating the diagnostic PCR strategy to identify integration of the ESAG6 tagging construct into *ESAG6* or *ESAG7*. Primer F1 anneals in *sfGFP* and is used in both reactions; primers R1 and R2 anneal specifically in *ESAG6* and *ESAG7*, respectively. DNA will only be amplified if the construct has integrated into either of these genes with the corresponding R primer. Genomic DNA represented by a grey line, tagging construct sequence represented by thicker grey line. Diagram drawn to scale except primers, which anneal at arrowhead tips. **B**, Diagnostic PCR on clones resulting from transfection with ESAG6 tagging construct shows integration into *ESAG7* in clones 1, 11, 16, 22, 28 and 29; into *ESAG6* in clones 2, 9, 10, 13, 19, 20, 21, 23, 25 and 27; and into both *ESAG6* and *ESAG7* in clones 18 and 24.

Arrowheads indicate the expected product if integration of tagging construct into this gene was successful. Non-specific binding of primers has resulted in amplification of a ~540 bp product in both reactions, from transfectant and parental DNA.

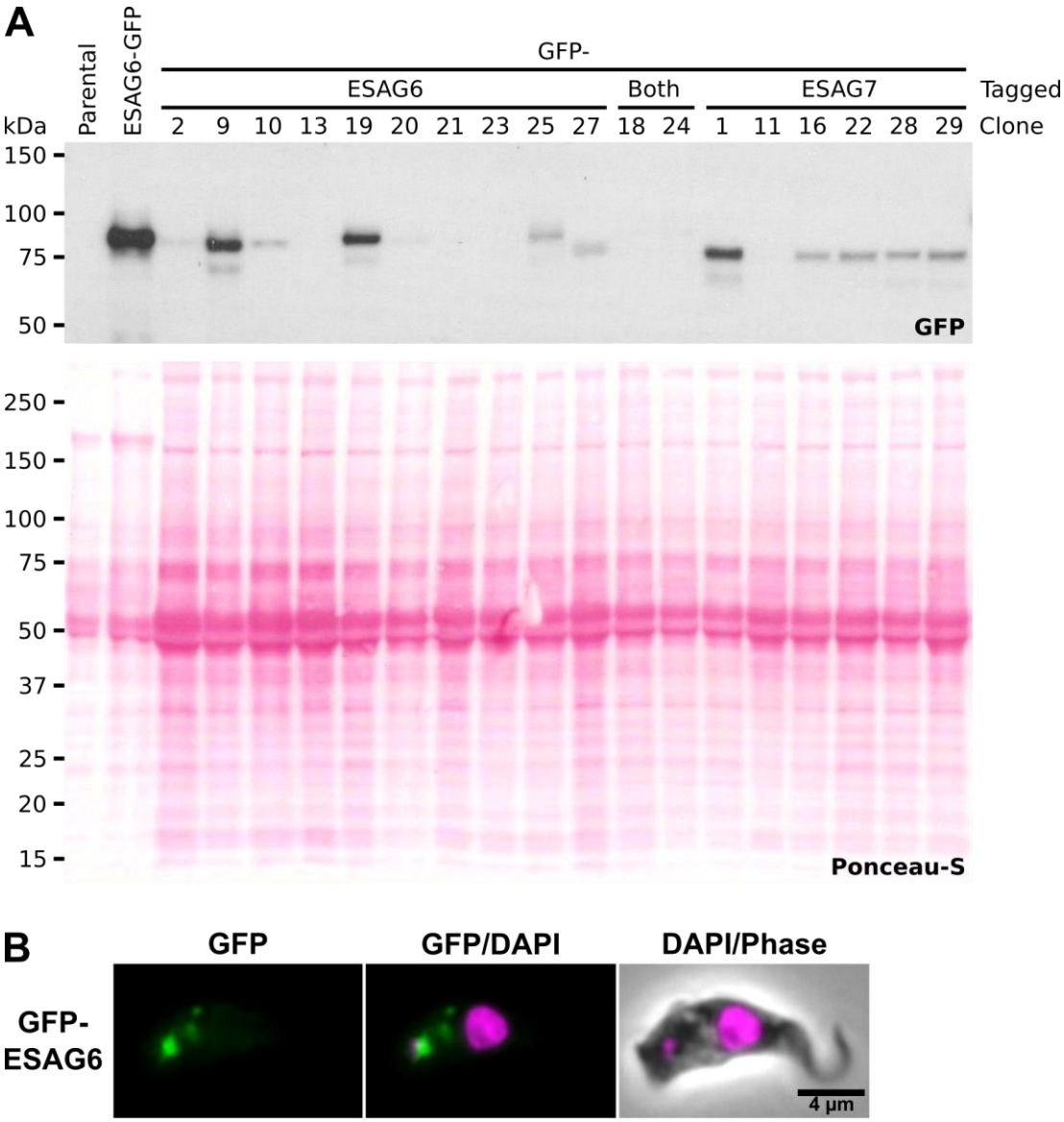


Figure 13. Chimeric protein is expressed by cells transfected with ESAG6 tagging construct. A, Immunoblot of lines with ESAG6 tagging construct integrated into *ESAG6*, *ESAG7* or both according to diagnostic PCR show expression of chimeric protein. GFP-fusion proteins detected with monoclonal antibodies against GFP. Nitrocellulose membrane stained with Ponceau-S provided as a loading control. **B**, Native fluorescence microscopy of GFP-ESAG6 clone 9 shows chimeric protein localisation to the flagellar pocket and endosomal system. Signal from GFP is shown in green. Cells were counterstained with DAPI, magenta.

cell lines may be tagged on a silent BES, with enough leaky transcription of selection marker to provide resistance to hygromycin.

In order to determine the cellular localisation of GFP-ESAG6, cells were investigated by native fluorescence microscopy (Figure 13B). GFP signal from clone 9 showed a flagellar pocket and endosomal localisation, which is in line with previous findings (Salmon *et al.* 1994), and the localisation of ESAG6-GFP (data not shown).

As mentioned previously, high intra-family sequence identity can result in integration of tagging vectors into genes other than the active BES *ESAG*. A Southern blot was designed to determine if the ESAG6 tagging construct – as well as those listed in Table 4 – had integrated into BES1. When the Tb427 genome is digested with SwaI, the resulting restriction fragment containing BES1 is approximately 58 kb; integration of the tagging construct into BES1 results in an increase of ~5 kb in this restriction fragment (Figure 14A). SwaI-digested genomic DNA of parental and tagged lines was separated by PFGE, transferred to nylon membrane, probed with *VSG221* ORF (sequence unique to BES1) probe, then stripped and re-probed with a mixture of *eGFP*, *sfGFP* and *HYG* ORF (sequence unique to the tagging constructs) probes. Figure 14B shows that restriction fragments containing *VSG221* and tagging construct features have migrated slower than that of the parental line, indicating that all tagged lines targeting ESAGs on BES1 from Table 4, and the GFP-ESAG6 line, had all integrated into BES1. Differences in the actual and published BES1 restriction fragment (between Figure 14A and 14B), and variability in size of this fragment between tagged lines, are attributed to variation in the number of repeats in the 70-bp repeat region.

ESAG3, *ESAG4*, and *ESAG8* are encoded as multiple copies on BES1 (Hertz-Fowler *et al.* 2008). Figure 14 indicates that one copy of each of these families is tagged on BES1, but does not define which copy is modified. A diagnostic PCR was designed to identify which copy of each family had been tagged (Figure 15A). This shows that *ESAG3.2* (Tb427.BES40.16), *ESAG4.2* (Tb427.BES40.13) and *ESAG8.2* (Tb427.BES40.14) are tagged on BES1 (numbered from 5’ to 3’; Figure 15B).

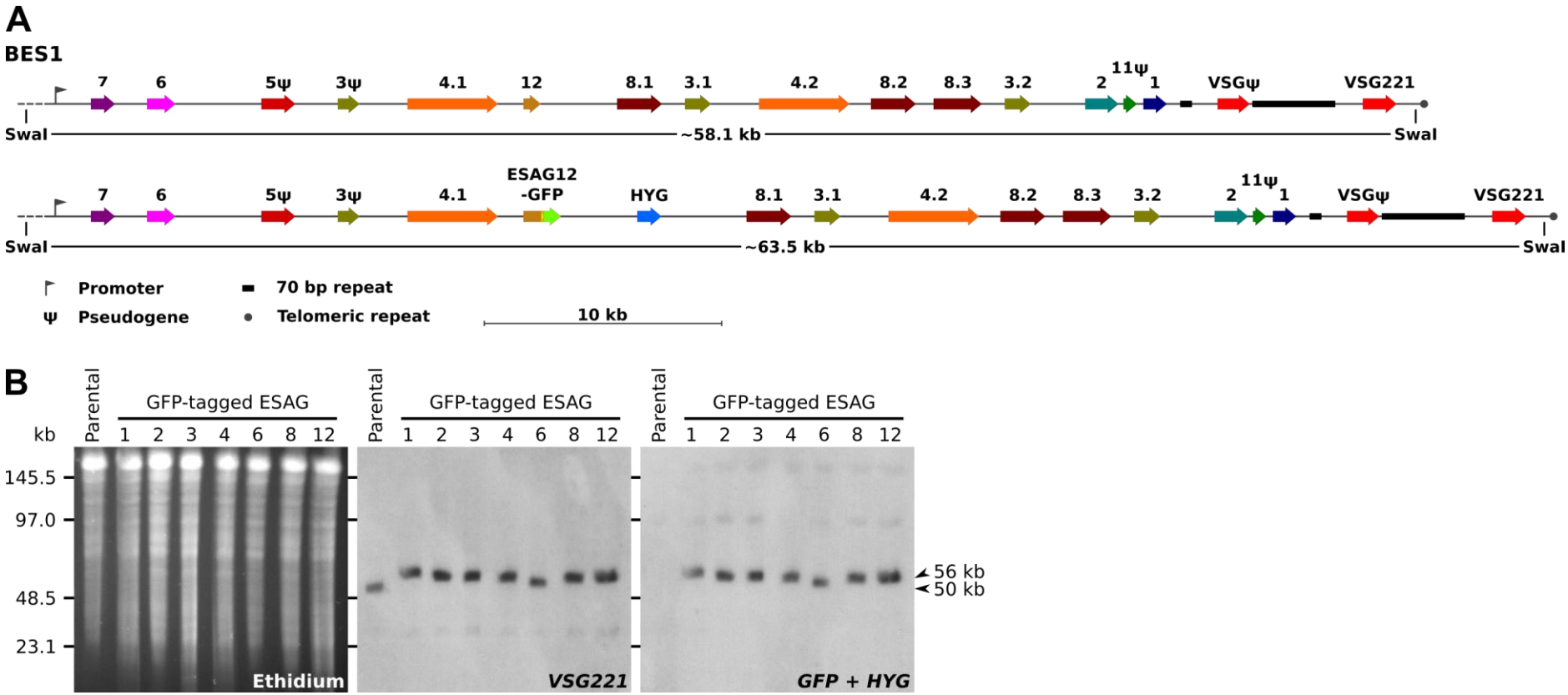
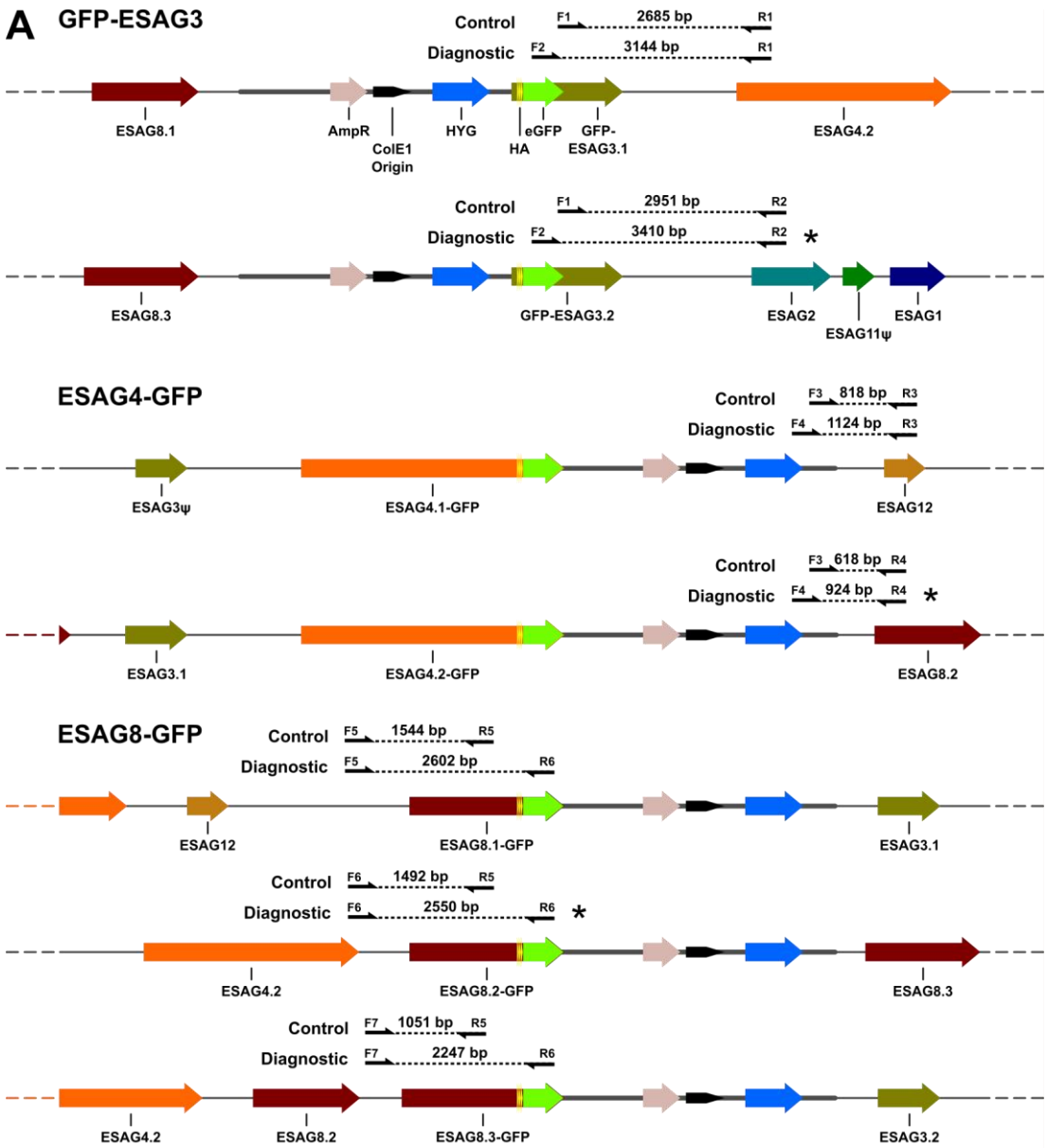


Figure 14. GFP-tagged ESAGs are encoded on BES1. A, Schematic outlining the size of the BES1-containing restriction fragment when genomic DNA is digested with SwaI endonuclease. Integration of tagging constructs pSiS or pSiG into BES1 (in this example to tag ESAG12) increases the fragment size by ~5 kb. Diagram drawn to scale. **B**, DNA from ESAG-tagged (Table 4), GFP-ESAG6 clone 9, and parental cell lines were digested with SwaI endonuclease, and separated by PFGE. Hybridisation of *VSG221* or a mix of *eGFP*, *sfGFP* and *HYG* probes with restriction fragments shows that tagging constructs have integrated into BES1 in all lines tested. Ethidium bromide stained agarose gel is provided as a loading control.


A
GFP-ESAG3
Control
Diagnostic
F1
2685 bp
R1
F2
3144 bp
R1
ESAG8.1
AmpR
ColE1 Origin
HYG
HA
eGFP
GFP-ESAG3.1
ESAG4.2
Control
Diagnostic
F1
2951 bp
R2
F2
3410 bp
R2
*
ESAG8.3
GFP-ESAG3.2
ESAG2
ESAG11ψ
ESAG1
ESAG4-GFP
Control
Diagnostic
F3
818 bp
R3
F4
1124 bp
R3
ESAG3ψ
ESAG4.1-GFP
ESAG12
Control
Diagnostic
F3
618 bp
R4
F4
924 bp
R4
*
ESAG3.1
ESAG4.2-GFP
ESAG8.2
ESAG8-GFP
Control
Diagnostic
F5
1544 bp
R5
F5
2602 bp
R6
ESAG12
ESAG8.1-GFP
ESAG3.1
Control
Diagnostic
F6
1492 bp
R5
F6
2550 bp
R6
*
ESAG4.2
ESAG8.2-GFP
ESAG8.3
Control
Diagnostic
F7
1051 bp
R5
F7
2247 bp
R6
ESAG4.2
ESAG8.2
ESAG8.3-GFP
ESAG3.2

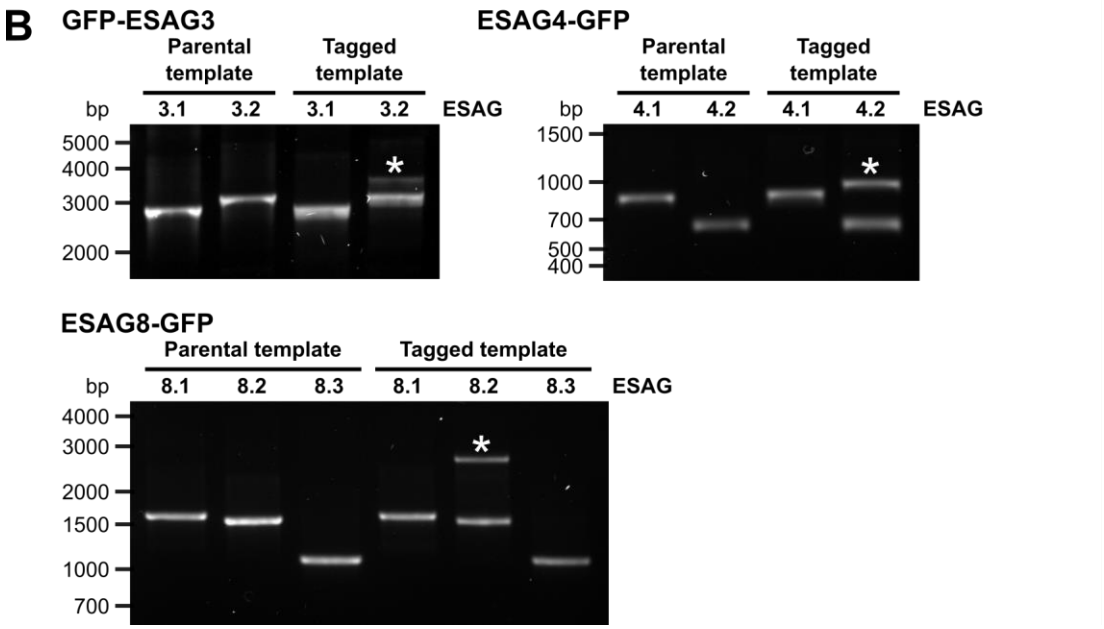

B
GFP-ESAG3
Parental template
Tagged template
bp
3.1
3.2
3.1
3.2
ESAG
5000
4000
3000
2000
*
ESAG4-GFP
Parental template
Tagged template
bp
4.1
4.2
4.1
4.2
ESAG
1500
1000
700
500
400
*
ESAG8-GFP
Parental template
Tagged template
bp
8.1
8.2
8.3
8.1
8.2
8.3
ESAG
4000
3000
2000
1500
1000
700
*

Figure 15. *ESAG3.2*, *ESAG4.2* and *ESAG8.2* are tagged on BES1. A, Diagnostic PCR strategy schematic to identify which BES1 copy of *ESAG3*, *ESAG4* and *ESAG8* was tagged. For each potentially tagged gene a "control" product is amplified (primers anneal outside of tagging construct sequence); integration of tagging construct enables amplification of a "diagnostic" product (one primer binds within *eGFP*). Genomic DNA represented by a grey line, tagging construct sequence represented by thicker grey line. Diagram drawn to scale except primers, which anneal at arrowhead tips. Asterisks in the schematic correspond to experimentally amplified diagnostic PCR products. **B**, Amplification of these products from genomic DNA isolated from parental and GFP-tagged cell lines. The diagnostic product is amplified from copy number two in each family, indicating that this is the tagged gene.

3.3. Obtaining an RNAi competent BES3-active line

Section 3.2 describes completion of a set of GFP-tagged ESAGs on BES1. However, BES1 does not encode a member of every (GR)ESAG family. To complete a set of GFP-tagged active BES ESAGs, tagged ESAG5 and ESAG10 lines were required. ESAG9 and ESAG11 are excluded from this set as they are not canonically represented in BESs: ESAG9 is only encoded on a single BES, and ESAG11 is only present on BESs as pseudogenes (Redpath *et al.* 2000; Hertz-Fowler *et al.* 2008). ESAG5 on BES1 is a pseudogene, and BES1 does not have the upstream section that encodes ESAG10 (Hertz-Fowler *et al.* 2008). BES3 is another commonly active expression site *in vitro*, and encodes putatively functional ESAG5 and ESAG10 (Hertz-Fowler *et al.* 2008).

In order to obtain GFP-tagged ESAG5 and ESAG10 lines, a BES3-active line was required. Initially we aimed to obtain cells that had undergone *in situ* switching from our BES1-active RNAi competent S16 culture (Figure 11). To select for a population that had switched away from BES1, cells were incubated with a range of concentrations of anti-VSG221 native serum (Figure 16A; produced in Gadelha *et al.* 2015). We hypothesised that cells with VSG221 on their surface would be opsonised by anti-VSG221 antibodies and lysed by complement activation. VSG221-positive S16 cells were also incubated with naive native serum, and VSG221-negative cells (Figure 11) were incubated with anti-VSG221 native serum as controls. As expected, VSG221-positive cells in anti-VSG221 serum formed clumps and died, which occurred at a faster rate as serum concentration was increased, while both control samples were unaffected (data not shown). Presence of VSG221 on the surface of cells that grew out of this experiment was

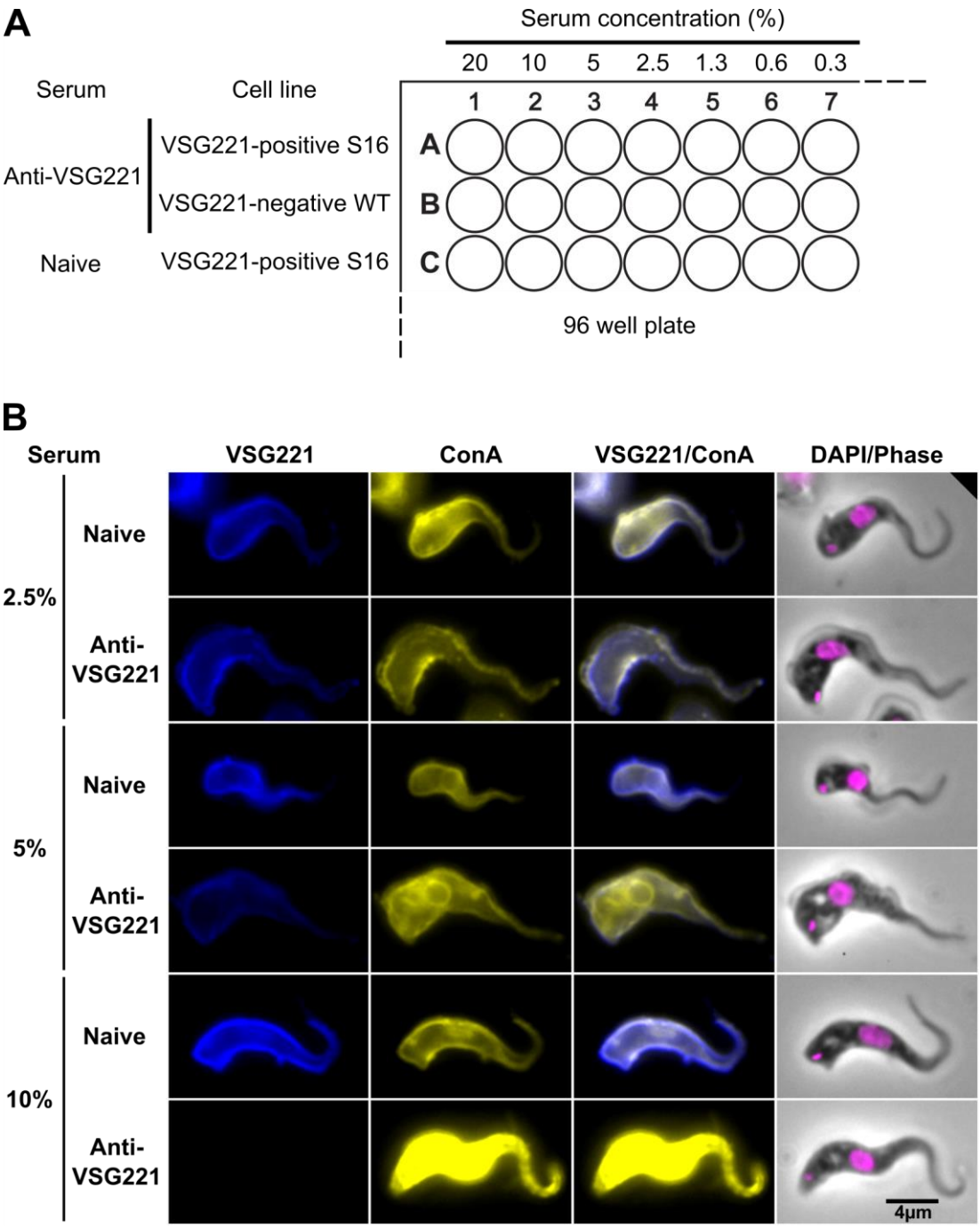


Figure 16. VSG221-negative cells can be selected from VSG221 expressing populations. A, Diagram of the experimental set-up to obtain cells that had switched away from BES1. 2.5x10^5 cells per well were incubated with serum ranging from 20% to 0.25% in a two-fold dilution series in a 96 well plate. **B**, Immunofluorescence microscopy on cells outgrowing from VSG221-positive S16 culture, post-challenge with anti-VSG221 serum. 10% anti-VSG221 serum was required to remove all VSG221-positive cells from the population. VSG221 signal displayed in blue; ConA, yellow; DAPI, magenta. Images are representative of the signal distribution observed for each cell line, and were processed to the same levels.

investigated by immunofluorescence microscopy (Figure 16B). Cells that grew out of 2.5 and 5% anti-VSG221 serum remained VSG221-positive,

while 10% serum was sufficient to remove all VSG221 expressing cells. These VSG221-negative cells show a greatly increased signal when stained with ConA. This suggests that the new VSG is decorated with more mannose or glucose residues that interact with the lectin, or they are more accessible for binding.

Clones were obtained from the population of VSG221-negative S16 cells (that had grown out of challenge with 10% anti-VSG221 serum). These were investigated for VSG121 expression using anti-VSG121 monoclonal antibody, a kind gift of Miguel Navarro (Figure 17A). Wild type cells expressing VSG121, a kind gift of Mark Carrington, were included in this experiment as a positive control. This showed that two of the six clones screened were VSG121-positive and potentially BES3-active.

To answer if this change in VSG expression was the result of gene conversion or *in situ* switching, the presence of *VSG221* and *VSG121* in the genome of these switched clonal populations was investigated by PCR. If the expressed VSG gene is switched by gene conversion, the previous VSG gene will be deleted from the genome. Whereas if transcription from the active BES is terminated and transcription from a previously silent BES activated – an *in situ* switch – DNA encoding the previous *VSG* gene will still be present in the genome. Figure 17B shows amplification of a *VSG121* fragment from VSG221-positive S16 cells, VSG121-positive WT and switched cells. However, *VSG221* cannot be amplified from the VSG121-positive switched clones. This indicates that these cells have switched VSG through gene conversion rather than *in situ* switching, and therefore are still BES1-active.

BES3-active cells could not be obtained through switching VSG221-positive S16 cells, so VSG121-positive wild type cells (Figure 17A, 17B) were made RNAi competent by introducing T7 RNA polymerase and tetracycline repressor (TetR) genes into the genome through transfection with the Single Marker Oxford (SmOx) construct (Poon *et al.* 2012). Clones obtained from this transfection (herein referred to as SmOx), and their derivatives, remained VSG121-positive (Figure 18A). In order to validate VSG121-

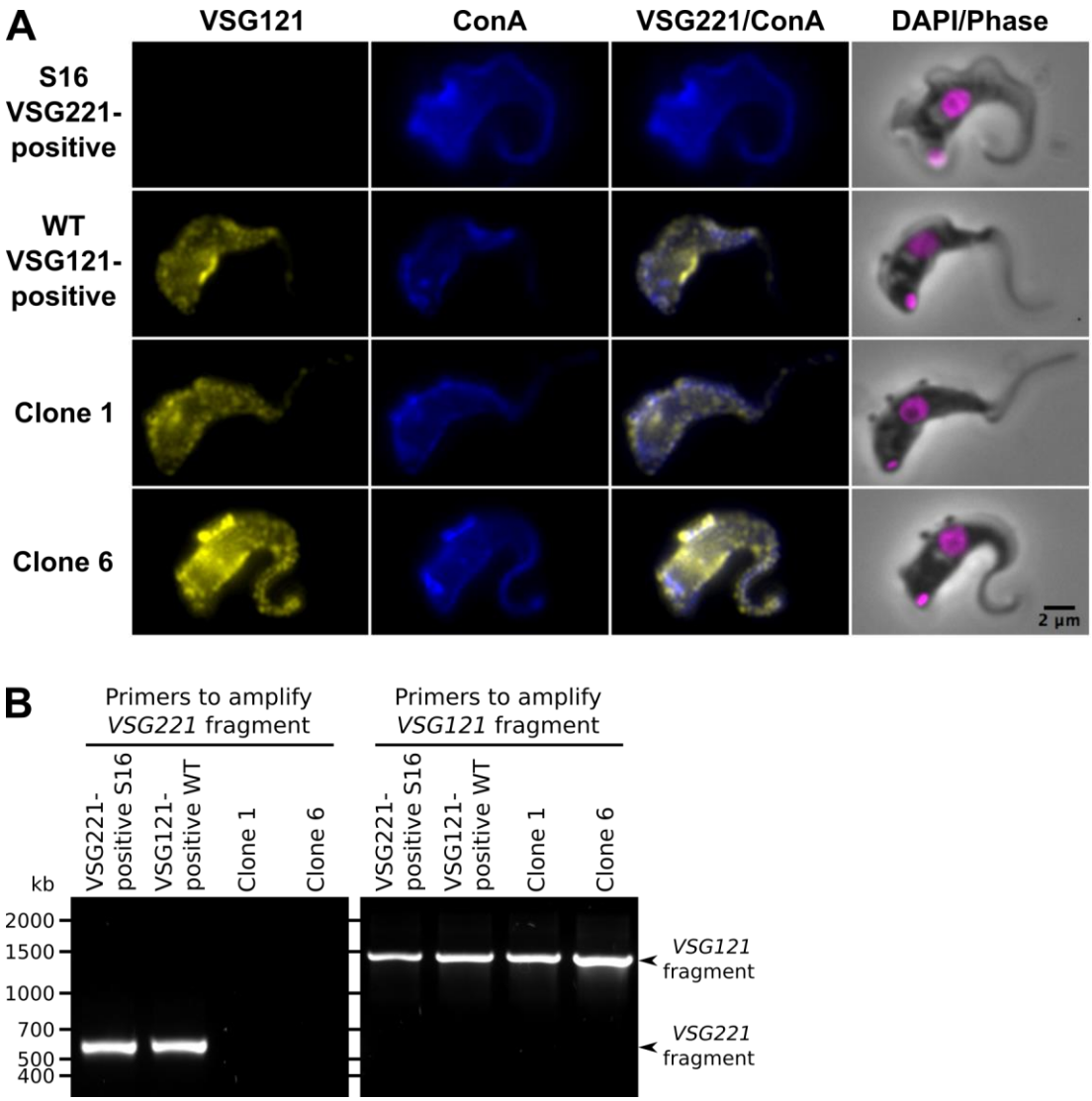


Figure 17. VSG121-positive cells obtained from the VSG221-positive S16 population have switched by gene conversion. A, Immunofluorescence microscopy on clones growing out from VSG221-positive S16 culture, after incubation with 10% anti-VSG221, shows that clones 1 and 6 are VSG121-positive. VSG221-positive S16 and VSG121-positive WT lines are included as controls. VSG121 signal displayed in yellow; ConA, blue; DAPI, magenta. Images are representative of the signal distribution observed for each cell line, all images processed to the same levels. **B**, Amplification of *VSG121*, but not *VSG221*, ORF fragment by PCR from clone 1 and 6 genomic DNA shows that the active VSG has switched by gene conversion in these lines. Fragments of both genes can be amplified from VSG221-positive S16 and VSG121-positive WT cell lines. Expected products indicated by arrowheads.

positive SmOx cells as RNAi competent, knockdown of a protein with a well-characterised phenotype, clathrin heavy chain (CLH), was performed. An inducible RNAi construct (p2T7-177, see Chapter 4 for more details) containing a fragment of the *CLH* ORF (p2T7-177-CHC, a kind gift of Sarah Whipple, University of Nottingham) was introduced into these cells, RNAi

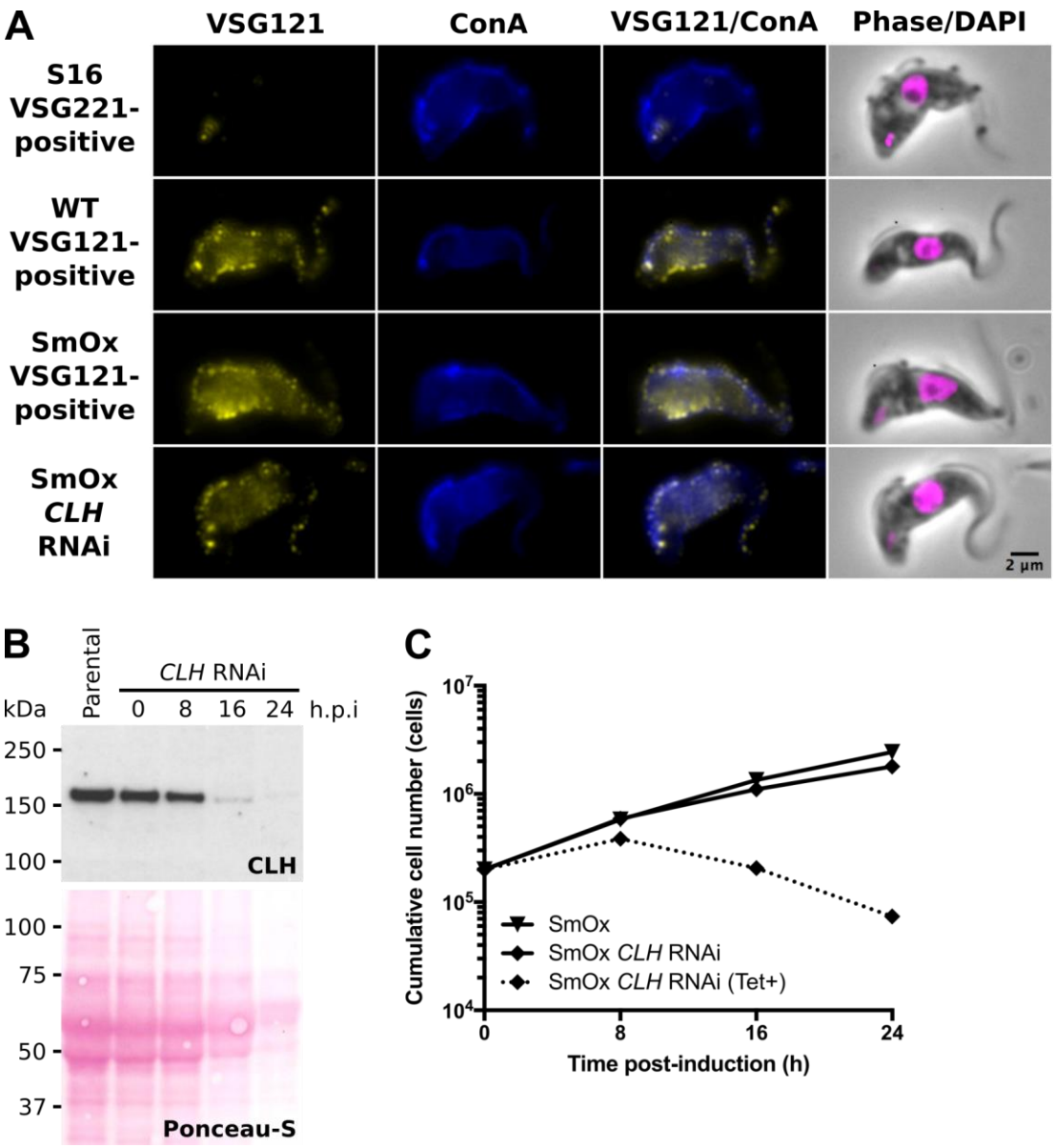


Figure 18. BES3-active cells can be made RNAi competent. A, Immunofluorescence microscopy shows that VSG121-positive WT cells maintain the same active VSG when transfected with SmOx and RNAi constructs. VSG221-positive S16 cells included as a control. VSG121 signal displayed in yellow; ConA, blue; DAPI, magenta. Images are representative of the signal distribution observed for each cell line, all images processed to the same levels. **B**, Immunoblot of VSG121-positive SmOx cells show depletion of CLH when RNAi is induced. A control lysate from the parental lane is included. CLH was detected with anti-TbCLH antiserum. Nitrocellulose membrane stained with Ponceau-S provided as a loading control. Hours post-induction (h.p.i). **C**, Growth of VSG121-positive SmOx *CLH* RNAi cells in culture. RNAi-induced ablation of CLH results in a severe growth defect. Parental and non-induced cells are included as controls.

induced, and protein levels monitored by immunoblot using antiserum raised against TbCLH (a kind gift of Mark Field; Morgan *et al.* 2001; Figure 18B). In induced cells CLH levels were reduced, resulting in a severe growth defect (Figure 18C) and characteristic “BigEye” phenotype (observed by light

microscopy, data not shown). This result recapitulated the findings of Allen *et al.* 2003, and validated the VSG121-positive SmOx line as RNAi competent.

3.4. Generation of cell lines expressing GFP-tagged ESAG5 and ESAG10

Now that a BES3-active RNAi competent cell line was available, tagging of ESAG5 and ESAG10 could take place. pSiS was modified to tag ESAG5 at the N-terminus with sfGFP and replace the native signal peptide. This construct was transfected into VSG121-positive SmOx cells and chimeric protein expression monitored by immunoblot using anti-GFP antibody (Figure 19). GFP-ESAG5 has a predicted molecular weight of approximately 82 kDa (after signal peptide cleavage). However, every clone screened exhibited signal at around 100 kDa. This increased size is likely due to N-glycosylation, which is a predicted feature of these proteins, and GFP-GRESAG5 migrated a similar position when assessed by the same method (Gadelha *et al.* 2015).

As with the ESAGs tagged on BES1 (Section 3.2), integration of the ESAG5 tagging construct into the active expression site (BES3) must be investigated due to promiscuity in targeting of the construct to other *ESAG5* genes. Genomic DNA from GFP-ESAG5 cells, and its parental line, were fragmented using PvuI restriction endonuclease, which cuts just downstream of the BES promoter and in the *HYG* resistance gene. Therefore, integration of the ESAG5 tagging construct results in a shorter restriction fragment containing the BES3 *VSG* gene, *VSG121* (Figure 20A). Visualisation of *VSG121, sfGFP* and *HYG* containing restriction fragments using complementary DNA probes shows that the tagging construct had integrated into BES3 (Figure 20B). When probing for *VSG121*, multiple restriction fragments were labelled. BLASTn of the *VSG121* probe sequence (used in Figure 20B) against the *T. brucei* Lister 427 VSGnome (Cross *et al.* 2014) identified three *VSG* genes with a high degree of sequence identity (Appendix 6). Cross-hybridisation of the *VSG121* probe with these additional two closely related genes is likely to be the cause of labelling multiple restriction fragments.

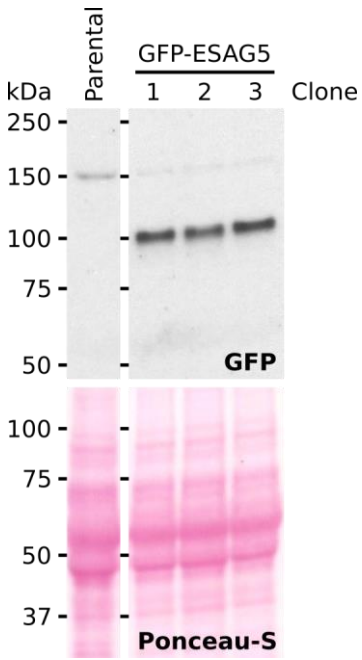
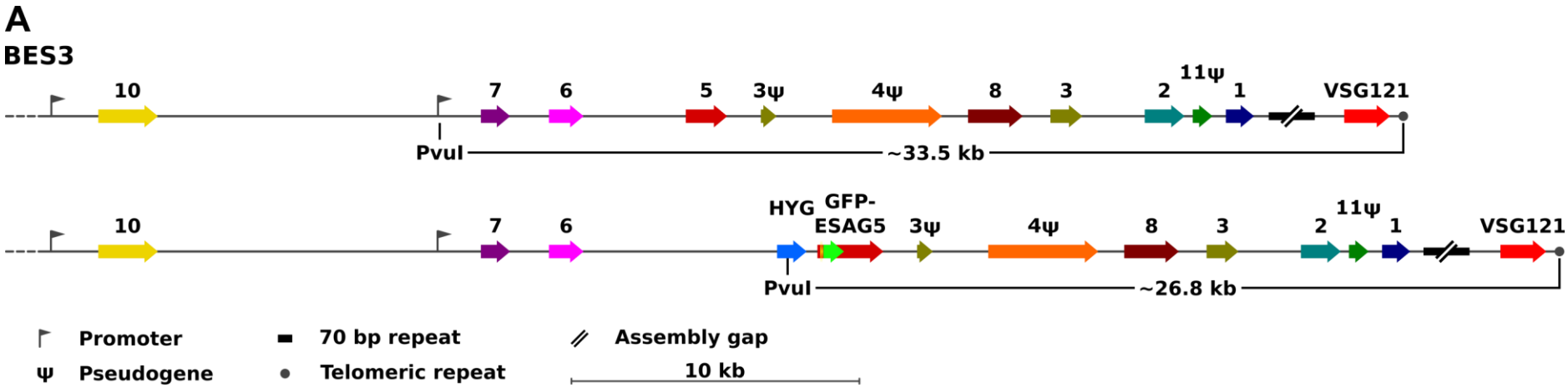


Figure 19. GFP-tagged ESAG5 is expressed in transfected lines. Immunoblot of whole cell lysate from lines transfected with ESAG5 tagging construct show expression of chimeric protein, detected with monoclonal antibodies against GFP. Nitrocellulose membrane stained with Ponceau-S provided as a loading control.

Native fluorescence microscopy shows an endoplasmic reticulum and endocytic vesicle localisation for GFP-ESAG5 (Figure 21). This cellular localisation of GFP-ESAG5 is similar to that of GFP-GRESAG5, albeit without the additional flagellar pocket signal shown by the GRESAG (Gadelha *et al.* 2015). This slight difference in localisation could be due to an inherent difference in native localisation of the BES3 ESAG5 compared to the GRESAG5 tested, or due to different levels at which the proteins are expressed (discussed in Section 3.5).

The last GFP-tagged ESAG needed to complete the set was ESAG10. As before, pSiS was modified to target *ESAG10* on BS3, but this time to integrate into the 3' end of the ORF, removing the native STOP codon and inserting the *HA* tags and *sfGFP* coding sequence in frame. Transfection of the ESAG10 tagging construct into BES3-active SmOx cells produced very few outgrowing cells resistant to hygromycin: only three clones grew out of two independent transfections. ESAG10-GFP has an expected size of around 105 kDa. When assessed by anti-GFP immunoblot each clone displayed signal at approximately 85-90 kDa (Figure 22). This variation from



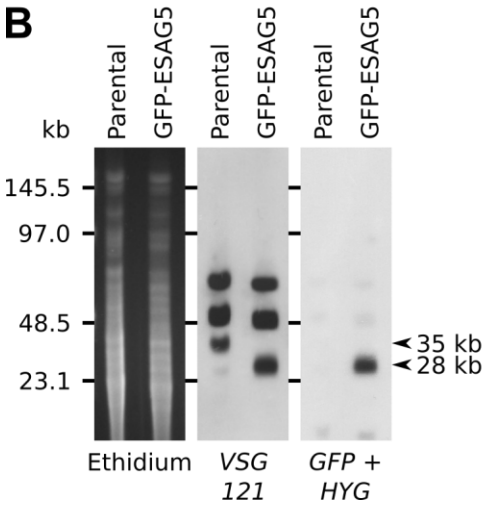


Figure 20. GFP-ESAG5 is encoded on BES3. A, Schematic outlining the size of the restriction fragment containing a section of BES3 when genomic DNA is digested with PvuI. Integration of ESAG5 tagging construct into BES3 introduces a PvuI site, reducing the fragment size by ~7 kb. Diagram drawn to scale. **B**, DNA from GFP-ESAG5 clone 1 and parental cell lines were digested with PvuI endonuclease, and separated by PFGE. Hybridisation of separated DNA to either *VSG121* or a mix of *sfGFP* and *HYG* probes shows that the ESAG5 tagging construct has integrated in to BES3. Ethidium bromide stained agarose gel is provided as a loading control.

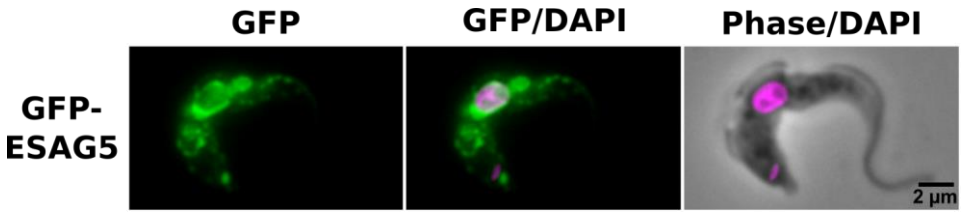


Figure 21. GFP-ESAG5 localises to the endoplasmic reticulum and endosomal system. Native fluorescence microscopy of GFP-ESAG5 clone 1, signal from GFP is shown in green. Cells were counterstained with DAPI, magenta.

the expected molecular weight may result from aberrant migration of the chimeric protein by SDS-PAGE due to the high degree of hydrophobicity inherent to ESAG10 because of its 13 predicted transmembrane domains. Clones 1 and 3 show similar chimeric protein levels, while clone 2 expressed far less ESAG10-GFP, and grew at a slower rate than the other clones and the parental line. This suggested that the tagging construct in clone 2 had integrated into an inactive BES and the small amount of transcription of this BES produced enough HYG to provide a degree of resistance to the drug pressure, albeit with a growth defect. To explore this hypothesis, these lines were investigated by anti-VSG121 immunofluorescence microscopy (data not shown). Clone 2 and parental lines remained VSG121-positive, but the other clones no longer had VSG121 present on the cell surface, and had presumably switched BES. These results prompted speculation that the tagging construct had integrated into an expression site other than BES3 in all three instances.

To determine which BES had become active in the ESAG10-GFP clone 1 line, RNA was extracted from these cells (as well as a number of control lines with known VSG expression through immunofluorescence microscopy) and converted to cDNA using reverse transcriptase. PCR using primers that anneal in the 5' miniexon (spliced onto trypanosome mRNA) and in a conserved region of the *VSG* 3' UTR were employed to amplify the actively expressed *VSG* gene (as described by Hertz-Fowler *et al.* 2008). Sequencing this PCR product revealed that *VSG224* was actively transcribed by the ESAG10-tagged line, which is specific to BES7 – a BES that is known to include the upstream expression site section encoding ESAG10 (Hertz-Fowler *et al.* 2008). Confirmation of integration of the tagging construct into

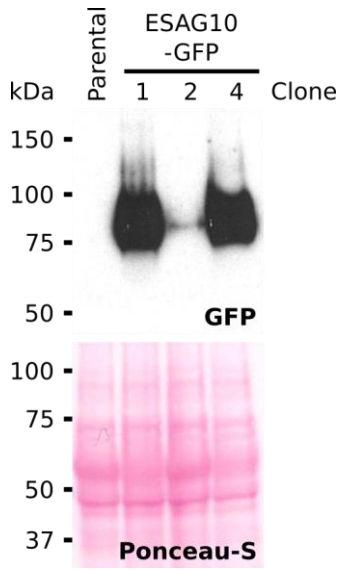


Figure 22. ESG10-GFP is expressed in transfected lines. Immunoblot of whole cell lysate from lines transfected with ESAG10 tagging construct show expression of chimeric protein, detected with monoclonal antibodies against GFP. Nitrocellulose membrane stained with Ponceau-S provided as a loading control.

BES7 was conducted again by Southern blot, designed so that integration of the construct would increase the size of the BES7-containing restriction fragment (Figure 23A). This time genomic DNA was digested with BsiWI restriction endonuclease, separated by PFGE, and probes against *VSG224*, *sfGFP* and *HYG* ORFs used. A shift in size in the tagged cell line, compared to the parental BES7-containing restriction fragment, can be seen in Figure 23B, indicating integration into this BES. The difference in expected restriction fragment size in Figure 23A, and that obtained in Figure 23B is attributed to variation in the 70-bp, and telomeric repeats between the published BES7 sequence and the line examined.

We next asked if BES3 in our VSG121-positive SmOx line encoded the upstream ESAG10 gene, as per the published sequence (Hertz-Fowler *et al.* 2008). To answer this, genomic DNA from the GFP-ESAG5 and parental cell lines were fragmented with ApaI restriction endonuclease. This cuts in the *ESAG10* ORF, a restriction site that is conserved between all *ESAG10* genes and *GRESAG10* genes on chromosome 8. Therefore, compared to Figure 20A where these genomes were digested with PvuI, restriction fragments containing BES3 should be 11 and 23 kb larger between parental and tagged lines, respectively, when digested with ApaI (Figure 23A). When these

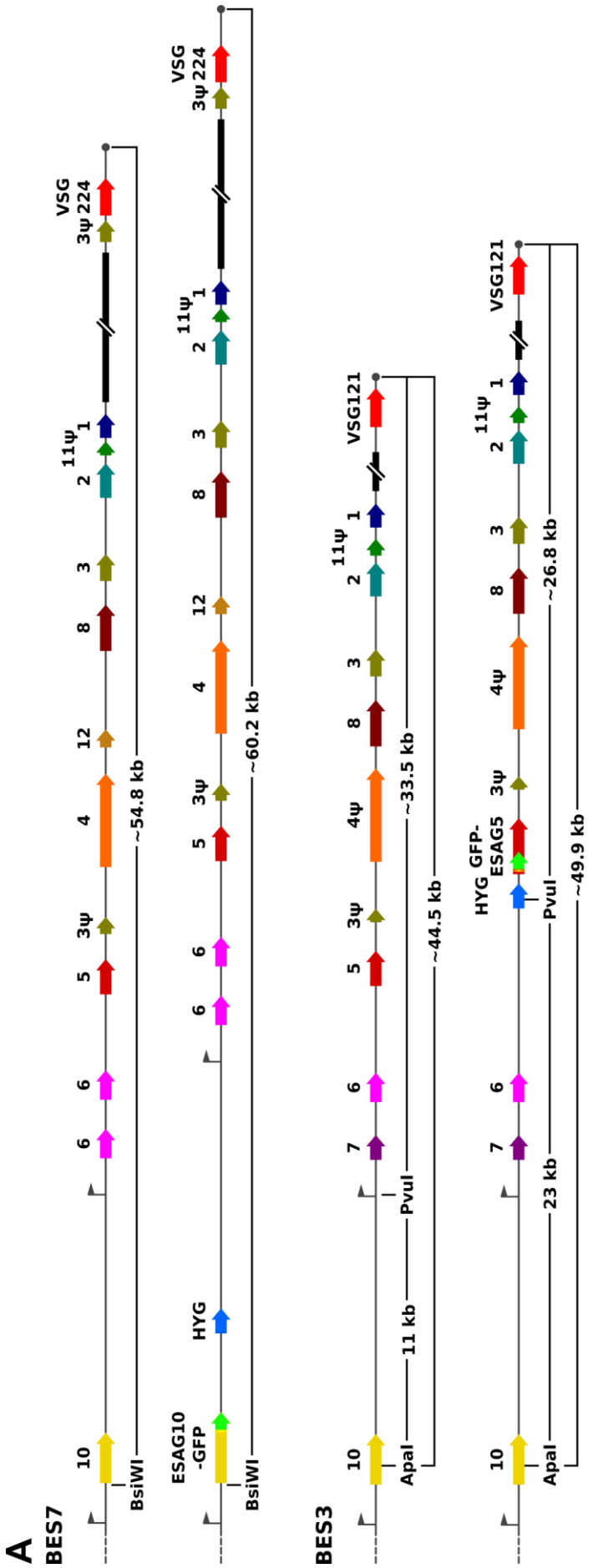


Figure 23. ESAG10-GFP is encoded on BES7. BES3 does not encode ESAG10. A, Schematic outlining the size of restriction fragment containing a section of BES7 when genomic DNA is digested with BsiWI. Integration of ESAG10 tagging construct into BES7 increases the fragment size by ~5 kb. *VSG121*-containing restriction fragments resulting from digestion of genomic DNA with ApaI and PvuI are also shown. *VSG121*-containing ApaI restriction fragments should be 11 or 23 kb larger than those produced by PvuI digestion of parental or GFP-ESAG5 genomic DNA, respectively. Diagrams drawn to scale.

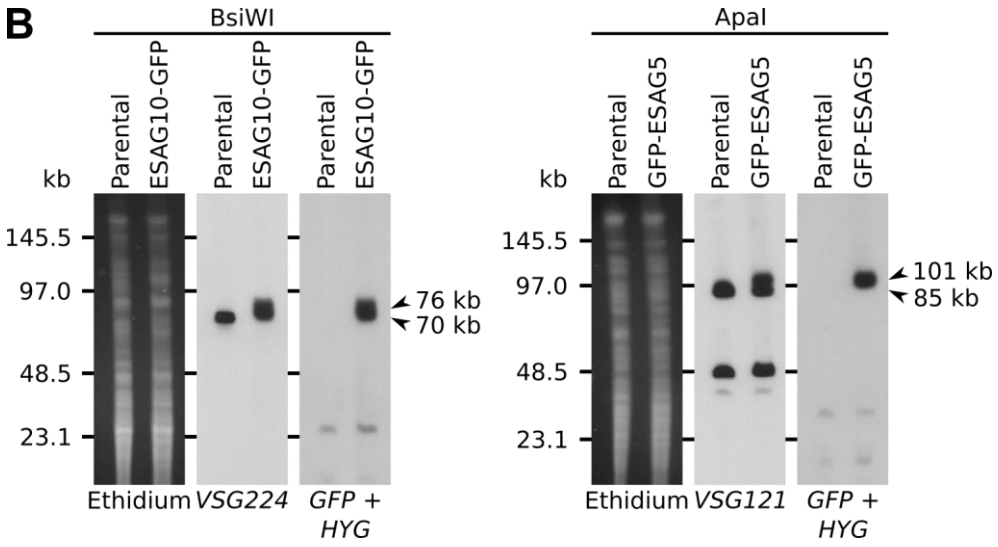


Figure 23. ESAG10-GFP is encoded on BES7. BES3 does not encode ESAG10. B, DNA from ESAG10-GFP clone 1, and parental, cell lines were digested with BsiWI endonuclease, and separated by PFGE. Hybridisation of separated DNA to either *VSG224* or a mix of *sfGFP* and *HYG* probes shows that ESAG10 is tagged on BES7. DNA from GFP-ESAG5 clone 1 and parental cell lines were digested with ApaI endonuclease, and separated by PFGE. Hybridisation of separated DNA to either *VSG121* or a mix of *sfGFP* and *HYG* probes is shown. *VSG121*-containing restriction fragments are much larger than expected, suggesting that the ApaI site in ESAG10 is not present. Ethidium bromide stained agarose gel is provided as a loading control.

fragmented genomes were probed for *VSG121*, restriction fragments containing BES3 were much larger than expected. Given the results shown in Figure 20B, the BES3-containing restriction fragments should be around 46 and 51 kb in parental and tagged lines, respectively. However, fragments of approximately 85 and 101 kb were obtained (Figure 23B). This suggests that, barring the unlikely mutation of the ApaI site in *ESAG10*, BES3 in this line and its derivatives is lacking an *ESAG10* gene, and probably the whole upstream BES section.

Native fluorescence microscopy of ESAG10-GFP showed signal on the cell body, excluded from the flagellar membrane (Figure 24). This localisation is similar to GRESAG10-GFP (Gadelha *et al.* 2015), but displaying greater intracellular signal from the endoplasmic reticulum and endosomal system, and lacking a flagellar pocket localisation. As with GFP-ESAG5, this slight variation between the GFP-tagged ESAG and a GRESAG of the same family may be inherent to the protein.

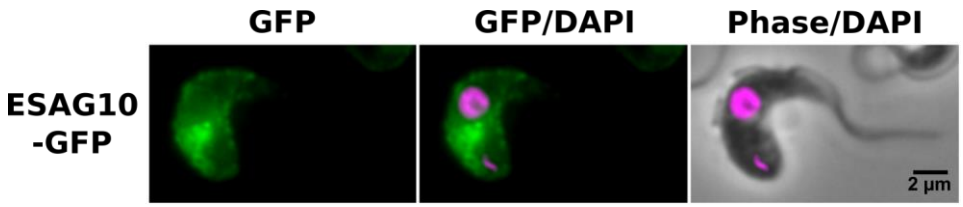


Figure 24. ESAG10-GFP localises to the cell body. Native fluorescence microscopy of ESAG10-GFP clone 1, signal from GFP is shown in green. Cells were counterstained with DAPI, magenta.

3.5. Summary

Data presented in this chapter describe the completion of a set of cell lines expressing a GFP-tagged ESAG family member from the active BES. GFP-tagged ESAG1, 2, 3.2, 4.2, 6, 8.2 and 12 are expressed from BES1, GFP-ESAG5 is expressed from BES3, and ESAG10-GFP is expressed from BES7. ESAG9 and ESAG11 are excluded from this set, as they are not typically represented across BESs. I have shown that parasites that have switched their VSG coat can be selected from a population of VSG221-positive cells when incubated in native anti-VSG221 serum. VSG121 is commonly activated post-switch from VSG221, being expressed by two out of 6 clones tested, and this switch had occurred by gene conversion rather than *in situ* BES switch. This foiled plans to obtain an RNAi competent BES3-active S16 line. However, a wild type BES3-active line was made RNAi competent through introduction of T7 RNA polymerase and TetR into the genome, and validated by testing knockdown of CLH, which reproduced the published phenotype.

Tagging ESAG6 at the N- or C-termini with GFP does not appear to impact the cellular localisation of chimeric protein, with both GFP-ESAG6 (this study) and ESAG6-GFP (Catarina Gadelha, University of Nottingham, unpublished data) targeted to the flagellar pocket. ESAG5 and ESAG10 show a similar localisation to one of their GRESAG counterparts, but both appear to lack the flagellar pocket signal observed in the GRESAGs. This difference in localisation may be native to each tagged protein: ESAG5 and the GRESAG5 tagged share 77% protein similarity, while ESAG10 and the GRESAG10 tested share 96%. It is not inconceivable that these differences

in protein primary sequence could result in a different cellular localisation of the folded protein. Levels of each chimeric protein may also influence their localisation. *ESAGs* are transcribed by RNA polymerase I, so transcript is presumed be more abundant than that of *GRESAGs* transcribed by RNA polymerase II. However, this does not strictly relate to protein levels in trypanosomes. Two *T. brucei* membrane-associated proteins have displayed level-dependent cellular localisations. Overexpression of the TfR or a histidine acid phosphatase (TbMBAP1) results in the protein's redistribution over the parasite surface, from the flagellar pocket and endosomes, respectively (Mußmann *et al.* 2003; Engstler *et al.* 2005). This may be a feature shared by (GR)ESAG5 and (GR)ESAG10 families, with protein level-dependent accumulation at the flagellar pocket.

Attempting to tag ESAG10 on BES3 resulted in selection of clones that had switched away from this BES, and a clone that grew at a reduced rate under hygromycin pressure (clone 2). This lead to the discovery that BES3 in the VSG121-positive SmOx line does not encode *ESAG10*. Upstream and downstream BES promoter regions present in this expression site architecture are identical, and can therefore recombine to remove the upstream BES section, a phenomenon that has been observed previously (Gottesdiener *et al.* 1991, 1992; Navarro and Cross 1996). Removal, or addition (through duplication), of this section of BES may represent an additional mechanism for adapting the parasite to the bloodstream of different mammalian hosts. This may be beneficial in an environment where the transporter substrate availability is limited, and increased uptake provides an advantage to the parasite. Low transfection efficiency of the ESAG10 tagging construct can be attributed to the need for transfectants to have switched to a BES that encoded this protein. In the case of the clone taken forward, this was BES7. As it had been observed when tagging ESAG6 and in ESAG10-GFP clone 2, integration of tagging construct into a silent BES is possible. The fact that more clones were not obtained displaying the same phenotype suggests that transcription from the upstream BES promoter does not behave in the same way as the downstream promoter, which is responsible for transcription of *ESAG6* and *ESAG7* from silent BESs

(Ansorge *et al.* 1999). In the case of ESAG10-GFP clone 2, this cell line may be resistant to the selection pressure due to a mutation in the machinery that maintains monoallelic transcription from the active BES, allowing transcription initiation from inactive BESs, and enough expression of ESAG10-GFP and HYG to survive.

Chapter 4. Reverse genetics approach to study ESAGs and (GR)ESAG families

4.1. Introduction

Protein ablation resulting from RNAi-mediated transcript degradation has been established in trypanosomes for many years, and has proved a powerful tool for study of protein function. In fact, *T. brucei* was the first protozoan parasite where RNAi was shown to be viable, through transfection of dsRNA homologous to the α-tubulin 5' UTR (Ngô *et al.* 1998). The full potential of this feature was unlocked through the development of tetracycline inducible expression, via introduction of *tetR* from *E. coli* into the *T. brucei* genome (Wirtz and Clayton 1995), and further improved by the production of lines co-expressing TetR and T7 RNA polymerase (Wirtz *et al.* 1999; Poon *et al.* 2012). While originally designed for inducible protein expression, these systems were also able to produce dsRNA either via stem-loop formation (Shi *et al.* 2000; Bastin *et al.* 2000), or through the use of head-to-head T7 promoters (Wang *et al.* 2000; LaCount *et al.* 2000), under the control of a Tet operator sequence. Initially, these inducible RNAi constructs were targeted to the transcriptionally silent rDNA spacer, to prevent leaky transcription of dsRNA in the absence of tetracycline. However, the work presented here makes use of a tetracycline-regulated vector modified from that described by LaCount *et al.* that integrates into the *T. brucei* minichromosomes, and provides improved transcriptional regulation (p2T7-177; Wickstead *et al.* 2002).

Gene silencing by dsRNA was first observed in *Caenorhabditis elegans* almost 20 years ago (Fire *et al.* 1998). The machinery required for this mechanism was described shortly after, and homologues of the proteins involved discovered in *T. brucei.* These include a member of the Argonaute family of proteins (Durand-Dubief and Bastin 2003; Shi *et al.* 2004), and two Dicer-like proteins (Shi *et al.* 2006; Patrick *et al.* 2009), whose native role is the production of regulatory small interfering RNA (siRNA, reviewed by Atayde *et al.* 2011). Briefly, Dicer cleaves dsRNA into approximately 22 bp long siRNA fragments (Bernstein *et al.* 2001), single strands of which are

loaded into Argonaute to form the RNA-induced silencing complex (RISC), and direct cleavage of complementary mRNA, inhibiting its translation (Liu *et al.* 2004).

As mentioned previously, RNAi has become a popular method of protein function investigation by ablation, evidenced by use in numerous studies including genome wide screens (reviewed by Kolev *et al.* 2011). This rivals the traditional route of protein ablation through gene knockout, mainly due to its ease of use, and ability to silence multiple genes with a single target sequence. Producing gene knockouts in trypanosomes requires removal of both alleles from the diploid genome and, unless all genes of interest are found as a cluster, the limited number of available selection markers makes targeting multiple genes difficult. Other drawbacks include the difficulties associated with knocking out essential genes, which is mitigated by the use of inducible systems, and the risk of obtaining cells that have activated compensatory networks, which potentially masks phenotypes (Rossi *et al.* 2015).

Here I describe the use of tetracycline-inducible RNAi constructs to knockdown GFP-tagged ESAGs (produced in Chapter 3) and (GR)ESAG families (Figure 25). Silencing of transcript encoding GFP-fusion proteins can be achieved by targeting the *GFP* ORF, which in these cell lines will result in ablation of only the actively expressed, GFP-tagged, ESAG family member. But by targeting RNAi to a conserved sequence of a minimum 22 bp in length (Bernstein *et al.* 2001), knockdown of multiple (GR)ESAG family members can be achieved. The latter approach has been employed to study these families (Salmon *et al.* 2012a; Batram *et al.* 2014; Tiengwe *et al.* 2016). Investigating the phenotypes produced in these cell lines will aid in the elucidation of protein function, and understand the roles played by ESAGs and GRESAGs of the same family.

4.2. Generation of pan-family RNAi constructs

As mentioned, RNAi targeted to multiple (GR)ESAG family members has been used to investigate the roles of these proteins. Constructs designed to

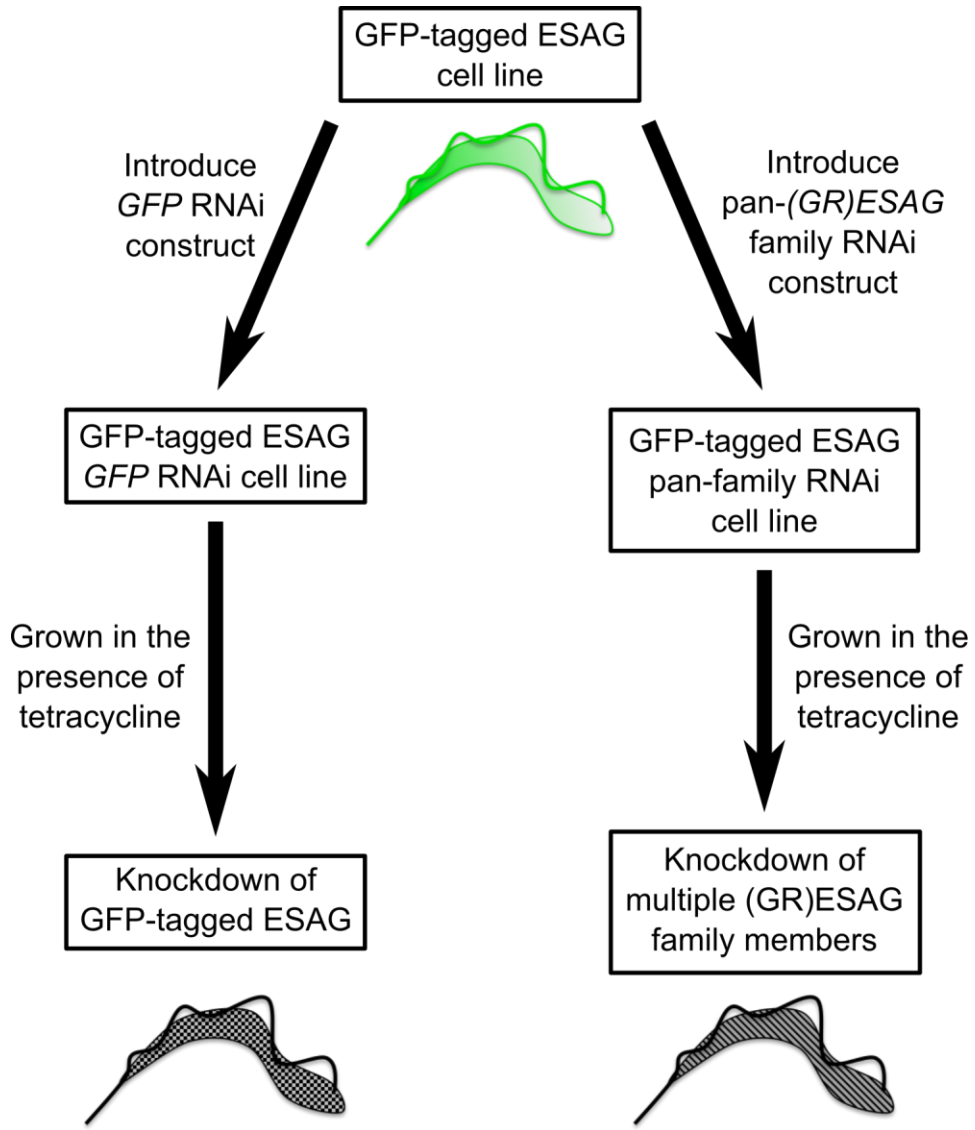


Figure 25. An RNAi strategy that enables the independent investigation of ESAGs and multiple (GR)ESAG family members. *GFP* and pan-*(GR)ESAG* family RNAi constructs were used to transfect GFP-tagged ESAG lines generated in Chapter 1. When grown in the presence of tetracycline this will result in knockdown of just the GFP-tagged ESAG or (GR)ESAG family members, respectively.

knockdown (GR)ESAG1-3, 6/7-8 and 12 (Batram *et al.* 2014) and (GR)ESAG4 (Salmon *et al.* 2012a) family members use *ESAG* ORF fragments to produce dsRNA that will induce RNAi. High DNA sequence identity between *ESAGs* means that these constructs will target RNAi to all *ESAGs* of their respective families. Some *GRESAGs* also share conserved sequence (of more than 22 consecutive bases) with *ESAGs* of the same family, so are also targeted for ablation. All of these published RNAi constructs have been shown to induce a phenotype, so overlapping targeting

fragments were used here. These fragments were cloned into inducible RNAi construct p2T7-177 and used to transfect the corresponding GFP-tagged ESAG cell line produced in Chapter 3. BLASTn was used to find *T. b. brucei* transcripts that had 22 bp of consecutive sequence identity with the published targeting fragments and those used in this study. This confirmed that the same genes would be targeted by both strategies, IDs of which are provided in Appendix 7. Each of these 'pan-family' RNAi constructs will knockdown all ESAG family members and at least one GRESAG, except in the case of (GR)ESAG8, where the non-BES genes share low sequence identity with their BES-linked relatives, thus only ESAG8s are targeted. Homology between *(GR)ESAG6* and *(GR)ESAG7* families enables both to be targeted with the same DNA fragment, facilitating knockdown of the parasite TfR. Only pan-family RNAi constructs targeting (GR)ESAG5 and (GR)ESAG10 families needed to be designed and produced. Knockdown of three (of seven) GRESAG10s was recently shown to confirm their putative function by reducing folate uptake, and was not accompanied by growth defects (Dewar *et al.* 2016), whilst RNAi targeting (GR)ESAG5 family members did not induce a measurable phenotype (Batram *et al.* 2014). After consideration of these results we aimed to investigate the effects of simultaneously knocking down the complete (GR)ESAG5 and (GR)ESAG10 families by producing pan-family RNAi constructs that target all family members.

In order to design RNAi constructs that were capable of knocking down entire (GR)ESAG families, these families first needed to be defined. Continued sequencing projects (in particular the assembly of chromosomes 9 and 11) and updates to TriTrypDB allow for the discovery of previously unknown *GRESAGs*, which may not be annotated. PSI-BLAST (Position-Specific Iterated Basic Local Alignment Search Tool) was employed to identify newly sequenced potential *GRESAGs* and update *(GR)ESAG5* and *(GR)ESAG10* families. This tool utilises a position-specific score matrix generated from sequence alignments, to iteratively search protein databases for related sequences, making it more sensitive at identifying distantly related proteins than the protein BLAST program (Altschul *et al.* 1997).

Isolation and sequencing of the *T. brucei* BESs was made possible through the use of transformation-associated recombination (TAR) to clone these telomeric sections into yeast artificial chromosomes (Becker *et al.* 2004; Hertz-Fowler *et al.* 2008). This in turn revealed the BES-linked *ESAG* repertoire, and the discovery of *ESAG12* (Hertz-Fowler *et al.* 2008). Protein sequences derived from Lister 427 BESs were combined with TREU 927 protein sequences from TriTrypDB release 9.0 to produce a comprehensive *T. brucei brucei* protein database. PSI-BLAST analysis of this database was initiated with BES3 encoded ESAG5 or ESAG10 protein sequence as a query, and allowed to run to convergence with an e-value threshold of 2×10^{-3}. This showed that the seven previously identified *GRESAG10* sequences make up the non-BES complement of this family, while three putatively functional chromosome 11 *GRESAG5* genes were revealed.

Multiple DNA sequence alignment enabled selection of potential fragments that would direct RNAi-mediated knockdown of multiple family members (multiple sequence alignments of *(GR)ESAG5* and *(GR)ESAG10* families provided, Appendices 8 and 9). *ESAG5* or *ESAG10* ORF fragments were sufficient to target RNAi to all *ESAG* family members, and the most closely related *GRESAG(s)* (Appendices 7 and 10). This *ESAG5* or *ESAG10* fragment, named pan-family RNAi fragment 1, was cloned into p2T7-177 and used to transfect GFP-ESAG5 or ESAG10-GFP cell lines produced in Chapter 3, respectively.

ESAGs and *GRESAGs* of *(GR)ESAG5* and *(GR)ESAG10* families share between 40-83 and 52-94% sequence identity at the DNA level, respectively, so additional RNAi fragments were required to target all family members. The use of multiple targeting fragments in tandem in a RNAi construct has been shown to facilitate knockdown of dissimilar proteins (Wang *et al.* 2000). An additional two RNAi fragments were required to knockdown the remaining GRESAG10s (not targeted by *(GR)ESAG10* pan-family RNAi fragment 1), while four additional fragments were required to target the remaining GRESAG5s (Appendices 7 and 10). In each case, these additional fragments were amplified from genomic DNA and combined into a single fragment by fusion PCR. This was then cloned into p2T7-177 encoding *BLAST* (p2T7-

B177), as opposed to *BLEO*, so it could be used in conjunction with the construct containing RNAi fragment 1. These were then introduced into the corresponding tagged line containing p2T7-177 with (GR)ESAG5 or (GR)ESAG10 RNAi fragment 1, to produce *(GR)ESAG5* and *(GR)ESAG10* pan-family RNAi lines, respectively. These RNAi fragments are highlighted on *(GR)ESAG5* and *(GR)ESAG10* family multiple DNA sequence alignments (Appendices 8 and 9), and the specific family members targeted by each RNAi fragment shown in Appendices 7 and 10.

RNAi constructs containing fragments of *eGFP* or *sfGFP* ORF were also produced, and used to transfect lines with eGFP- or sfGFP-tagged ESAGs, respectively. When induced, only the GFP-tagged ESAG will be knocked down in these lines, whereas multiple (GR)ESAG family members will be knocked down in pan-family RNAi lines (Figure 25).

4.3. RNAi mediated knockdown of GFP-tagged ESAGs and (GR)ESAG families

Clones resulting from transfection with RNAi constructs described in Section 4.2 were grown in the presence and absence of tetracycline for 24 hours, and lines showing the highest degree of chimeric protein ablation were taken forward (data not shown). RNAi was induced in these *GFP* and pan-family RNAi lines, and levels of GFP-tagged ESAGs monitored by immunoblot (Figure 26). All chimeric proteins migrated to approximately the predicted molecular weight, or correlated with those observed previously (Chapter 3; Gadelha *et al.* 2015). GFP-tagged ESAG1 and ESAG2 appear larger than their predicted sizes. These proteins are known to be glycosylated, which will increase their molecular weight (Cully *et al.* 1986; Nolan *et al.* 1999). The apparent size of GFP-ESAG3 is larger than expected, and ESAG3 is predicted to be modified by N-glycosylation, but does not appear to contain poly-N-acetyllactosamine based on its lack of interaction with tomato lectin (Nolan *et al.* 1999). ESAG8-GFP appears slightly smaller than expected, but its size is similar to that of published GFP-ESAG8 (Hoek *et al.* 2000).

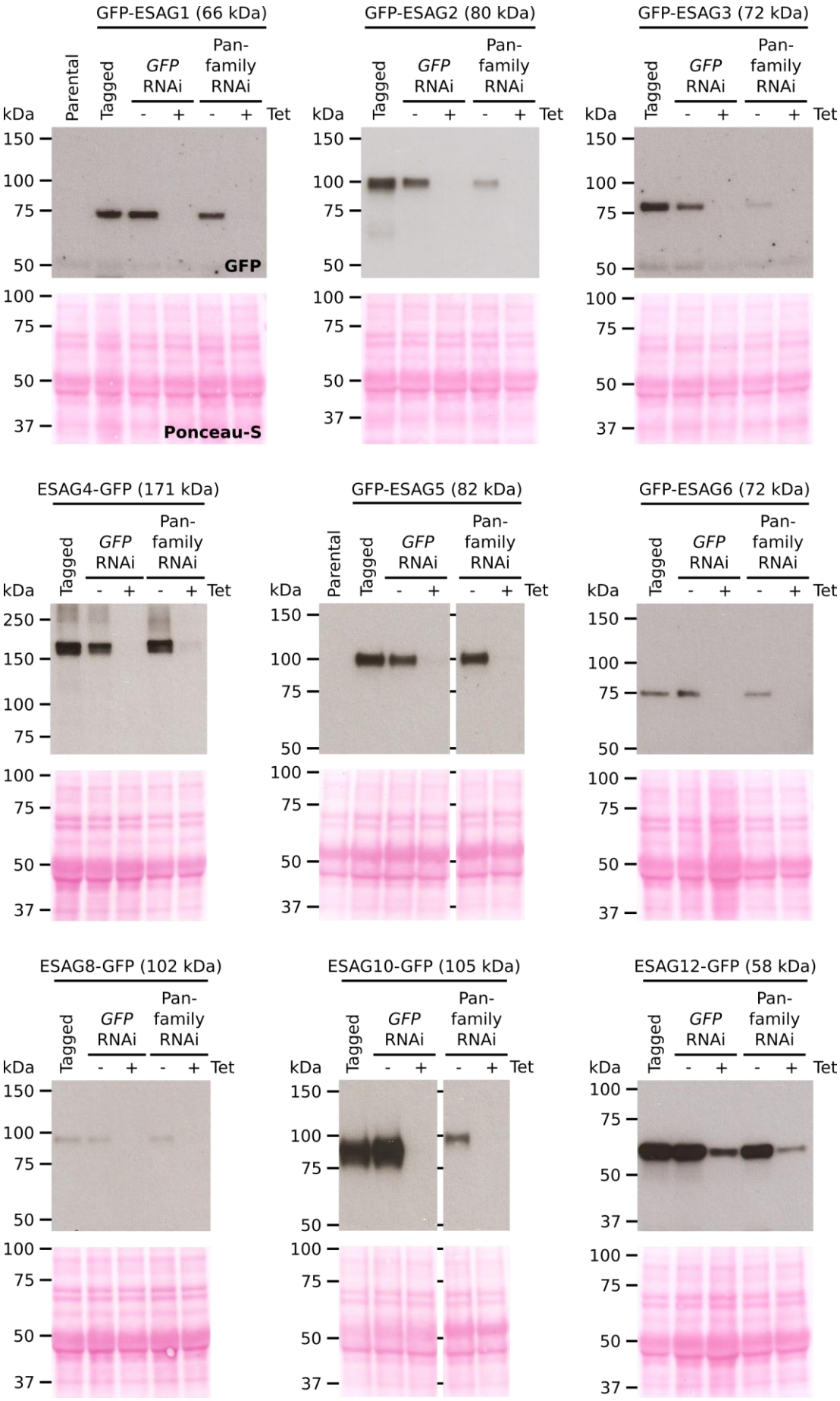

GFP-ESAG1 (66 kDa)
Parental
Tagged
GFP RNAi
Pan-family RNAi
kDa
- + - + Tet
150
100
75
50
GFP
100
75
50
37
Ponceau-S
GFP-ESAG2 (80 kDa)
Tagged
GFP RNAi
Pan-family RNAi
kDa
- + - + Tet
150
100
75
50
100
75
50
37
GFP-ESAG3 (72 kDa)
Tagged
GFP RNAi
Pan-family RNAi
kDa
- + - + Tet
150
100
75
50
100
75
50
37
ESAG4-GFP (171 kDa)
Tagged
GFP RNAi
Pan-family RNAi
kDa
- + - + Tet
250
150
100
75
100
75
50
37
GFP-ESAG5 (82 kDa)
Parental
Tagged
GFP RNAi
Pan-family RNAi
kDa
- + - + Tet
150
100
75
50
100
75
50
37
GFP-ESAG6 (72 kDa)
Tagged
GFP RNAi
Pan-family RNAi
kDa
- + - + Tet
150
100
75
50
100
75
50
37
ESAG8-GFP (102 kDa)
Tagged
GFP RNAi
Pan-family RNAi
kDa
- + - + Tet
150
100
75
50
100
75
50
37
ESAG10-GFP (105 kDa)
Tagged
GFP RNAi
Pan-family RNAi
kDa
- + - + Tet
150
100
75
50
100
75
50
37
ESAG12-GFP (58 kDa)
Tagged
GFP RNAi
Pan-family RNAi
kDa
- + - + Tet
100
75
50
37
100
75
50
37

Figure 26. GFP-tagged ESAGs can be specifically ablated by *GFP* or pan-*(GR)ESAG* family RNAi. Immunoblot of whole cell lysate from GFP-tagged ESAG lines (Chapter 3) and derivative *GFP* and pan-family RNAi lines shows that both RNAi constructs can knockdown chimeric protein. RNAi lines were grown in the absence (-) or presence (+) of tetracycline for 24 hours to induce dsRNA production. Parental VSG221-positive S16 and VSG121-positive SmOx lines are included with GFP-ESAG1 and GFP-ESAG5 samples, respectively, as negative controls. Chimeric protein was detected with monoclonal anti-GFP antibodies. Expected approximate chimeric protein sizes (post signal peptide and GPI-anchor addition sequence cleavage) are indicated. Nitrocellulose membrane stained with Ponceau-S provided as a loading control.

Both *GFP* and pan-family RNAi reduced levels of the GFP-tagged ESAG in all lines. Knockdown of the chimeric protein here is used as a proxy for evidence of targeting multiple *(GR)ESAG* family members in pan-family RNAi lines. Ablation below the limit of detection was achieved in most cases. ESAG12-GFP remained detectable after 24 hours of both *GFP* and pan-family RNAi induction. *(GR)ESAG4* pan-family RNAi also failed to completely remove chimeric protein from cells. Pan-family constructs targeting *(GR)ESAG2*, *(GR)ESAG3* and *(GR)ESAG10* appear to show leaky transcription from the RNAi construct, with chimeric protein levels reduced in non-induced samples compared to parental cells. Consistent with previous findings, ESAG8-GFP is only present at very low levels in the cell, and was nearing the limit of detection by immunoblot (Hoek *et al.* 2000; Gadelha *et al.* 2015). The *(GR)ESAG5* and *(GR)ESAG10* pan-family RNAi fragment 1 constructs (described in Section 4.2), were sufficient to ablate GFP-tagged ESAG5 and ESAG10, respectively, when assessed by immunoblot (Appendix 11A).

4.4. Ablation of (GR)ESAG families, but not individual BES ESAGs, induces growth defects *in vitro*

To investigate the essentiality of ESAGs compared to (GR)ESAG families, the growth of these RNAi lines was monitored for five days in the presence and absence of tetracycline (Figure 27). In every instance, ablation of the tagged-ESAG via *GFP* RNAi fails to affect the growth rate of cells. However, knockdown of multiple (GR)ESAG family members induces growth defects in all families except (GR)ESAG4 and (GR)ESAG12. Pan-family

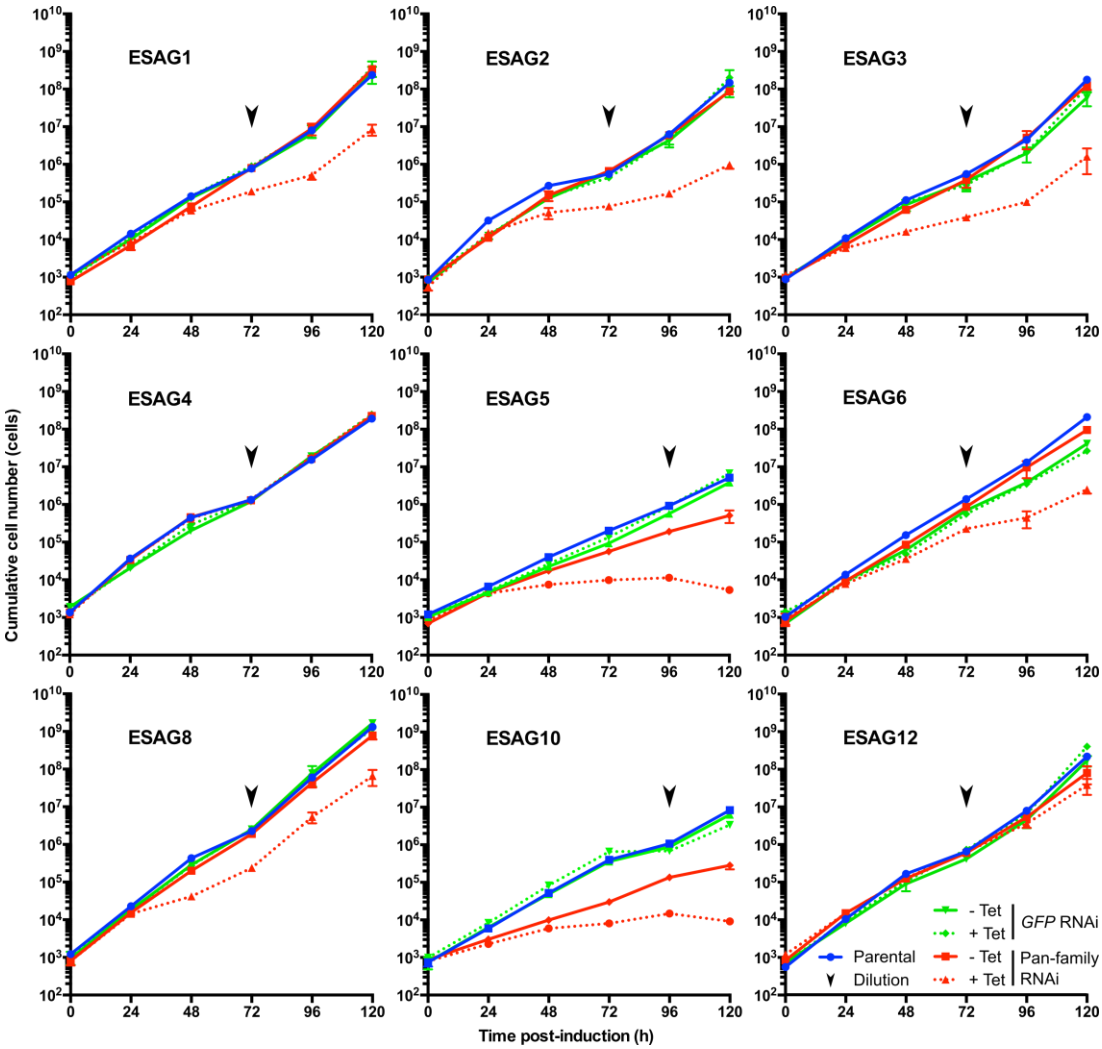


Figure 27. Ablation of (GR)ESAG families, but not GFP-tagged ESAGs, induces growth defects. Growth of GFP-tagged and derivative RNAi lines was monitored for five days. Induced *GFP* RNAi lines grow at the same rate as parental lines, while ablation of (GR)ESAG families by pan-family RNAi results in growth defects. Parental, *GFP* RNAi and pan-family RNAi cultures are shown in blue, green and red, respectively. Cultures containing tetracycline are shown with a dashed line, absence of tetracycline is represented by a solid line. Arrowheads indicate which day cultures were diluted into fresh media to maintain log-phase growth.

RNAi in these lines failed to fully ablate chimeric protein levels (Figure 26), so the remaining protein may be sufficient to supplement the role of the family and maintain normal growth. Knockdown of GFP-tagged ESAG5 and ESAG10 using the pan-family RNAi fragment 1 (which targets *ESAGs* and most *ESAG*-like *GRESAGs*, in each family) was insufficient to induce the growth defects observed in Figure 27 (Appendix 11B). These results indicate that most (GR)ESAG families are required for cell growth *in vitro*, while the actively expressed BES-linked ESAG of each family is not. This also

suggests that there is intra-family functional redundancy, and that silencing of the ESAG can be complemented by GRESAG family members. The severity of growth defects observed in pan-family RNAi lines decreases, and growth rate returns to normal in a number of cases between 96 and 120 hours of induction. This effect has been previously observed when performing RNAi in trypanosomes (Allen *et al.* 2003; Batram *et al.* 2014).

Both GFP-tagged ESAG5 and ESAG10 lines show a reduced growth rate compared to BES1-active cells. This trait is seen consistently between their parental lines, VSG121-positive SmOx cells compared to the VSG221-positive S16 cells (data not shown). This phenotype may result from the different ESAG repertoire or different biochemical properties of the VSG encoded on BESs 1, 3 and 7. *(GR)ESAG5* and *(GR)ESAG10* pan-family RNAi lines also exhibit a reduced growth rate compared to the tagged line. In the case of *(GR)ESAG10*, this could be caused by leaky transcription of the RNAi construct (shown in Figure 26). However, this was not a feature of the *(GR)ESAG5* pan-family RNAi line.

4.5. Investigation of PAD1 and TfR levels in RNAi lines

To further investigate the phenotype induced by (GR)ESAG family and individual ESAG knockdown, levels of PAD1 and TfR were examined. Part of the PAD family of proteins responsible for transduction of the differentiation signal, PAD1 is a multipass transmembrane transporter present on the surface of transmission-adapted stumpy form cells (Dean *et al.* 2009), and has been used as a marker for this lifecycle stage on numerous occasions. Batram *et al.* have previously used a reporter cell line with the *GFP* ORF under control of *PAD1* UTRs, with GFP levels are monitored by flow cytometry as a proxy for PAD1 expression. They have shown that RNAi targeting *(GR)ESAG1*, *(GR)ESAG2* and *ESAG8* transcripts activates this reporter, and hypothesise that deregulation of ESAG homeostasis acts as a signal for stumpy form differentiation (Batram *et al.* 2014).

In order to disentangle the role played by ablation of ESAGs or (GR)ESAG families in inducing PAD1 expression, whole cell lysates

(produced at room temperature to prevent aggregation of PAD1) of *GFP* and pan-family RNAi lines were probed with anti-PAD1 antibody (a kind gift of Keith Matthews, Figure 28). Non-induced, 24- and 72-hour RNAi induction whole cell lysates were produced, as Batram *et al.* observed increasing reporter signal during this period, peaking at the 72-hour time point. Overgrown (allowed to reach stationary phase) pleomorphic EATRO 1125 (clone AnTat 1.1, a kind gift of Luisa Figueiredo) whole cell lysate was included as a positive control and produced a signal of just below 50 kDa. The predicted molecular weight of PAD1 is 66 kDa but, like ESAG10, encodes a number of hydrophobic transmembrane domains, which may affect its migration on SDS-PAGE and cause the deviation from the expected size. No other samples probed with anti-PAD1 antibody exhibit this signal, indicating that PAD1 expression was not induced by ESAG or (GR)ESAG family knockdown. Lack of signal produced by overgrown VSG221-positive S16 parental cells is consistent with previous findings, as this laboratory-adapted “monomorphic” cell line does not express PAD1 at detectable levels (Dean *et al.* 2009). However, this line has been shown to express PAD1 when a second VSG is expressed, and does not explain the activation of a PAD1-reporter upon (GR)ESAG family knockdown (Batram *et al.* 2014).

Knockdown of GFP-tagged ESAGs was also monitored in this experiment (Figure 28). RNAi lines show a similar degree of chimeric protein knockdown compared to Figure 26. As before, *(GR)ESAG4* and *(GR)ESAG12* pan-family RNAi failed to fully ablate their respective chimera, as does *GFP* RNAi in the ESAG12-GFP line. ESAG12-GFP also shows a number of degradation products, and potential dimer formation (evidenced by a band of approximately 120 kDa, around twice the size of ESAG12-GFP), compared to Figure 26. This may result from producing these lysates at room temperature: it is possible that ESAG12-GFP is sensitive to proteolysis by enzymes that are still active when cells are lysed at room temperature, and this more gentle sample production method may enable ESAG12-GFP dimers to remain associated. ESAG8-GFP was barely visible in parental and non-induced samples, but no longer detectable in induced cells. After 24 hours of *GFP* RNAi induction, ESAG10-GFP levels were equal to those in

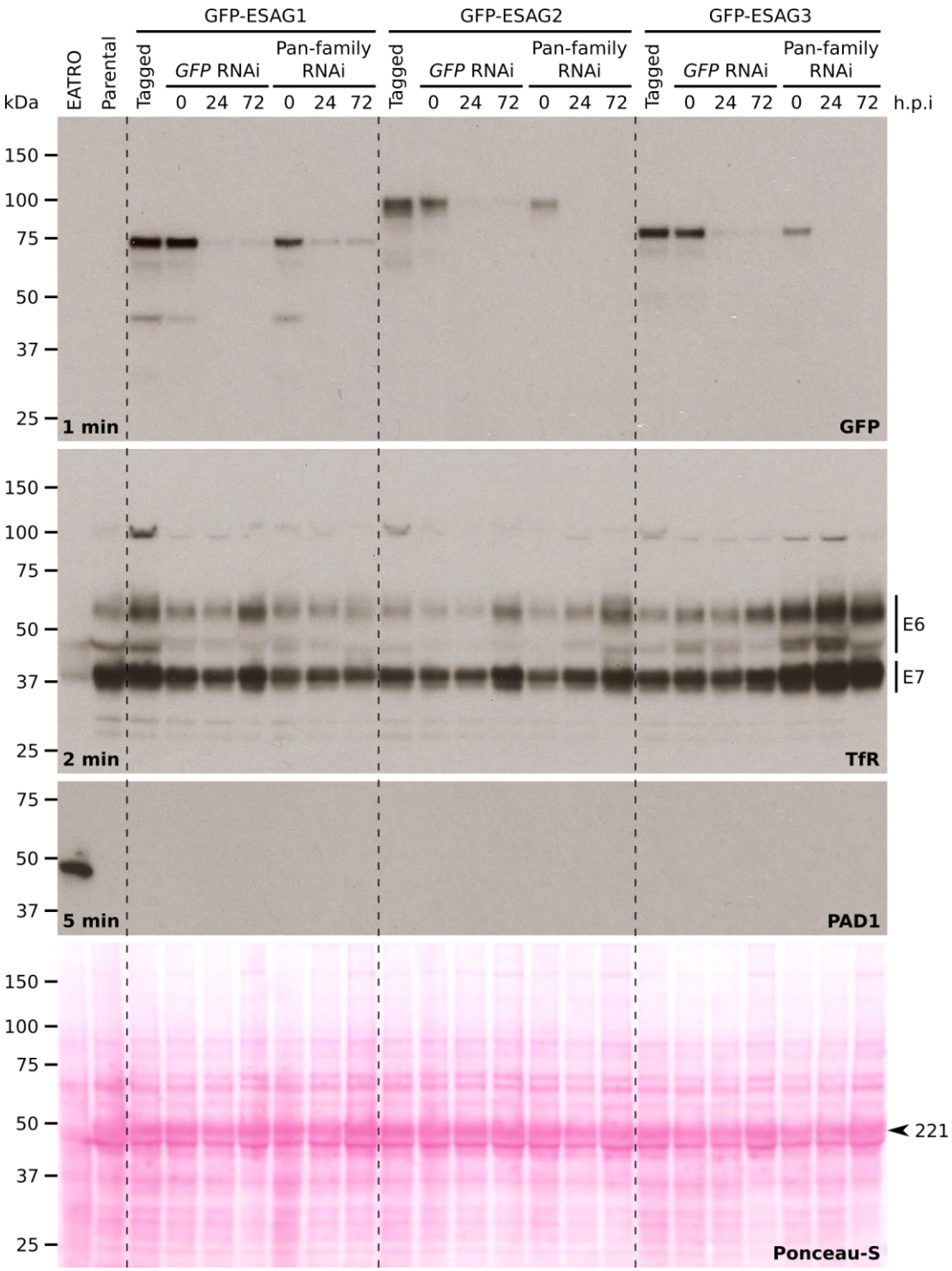
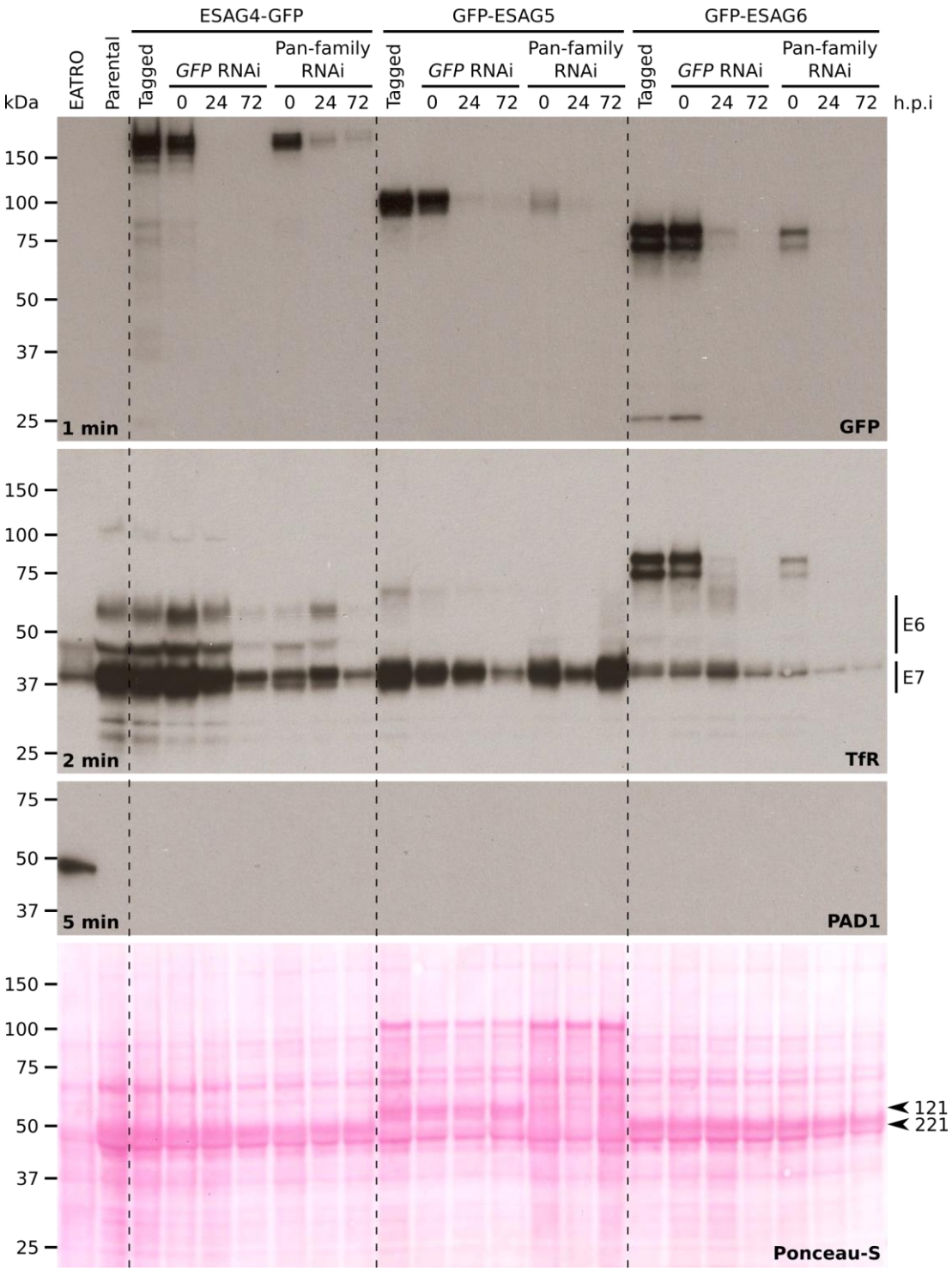
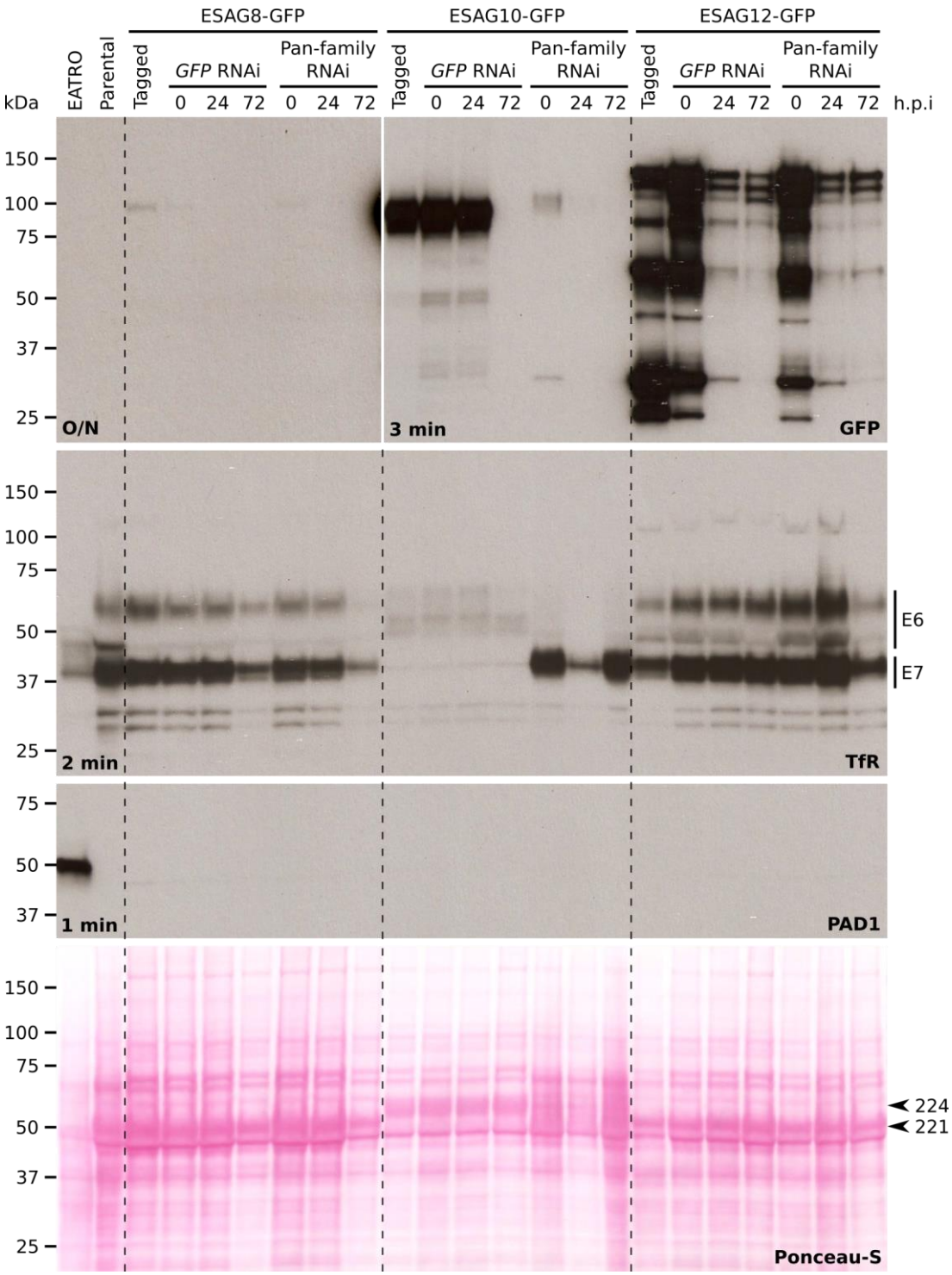


Figure 28. Knockdown of GFP-tagged ESAGs and (GR)ESAG families does not induce PAD1 expression, but does impact TfR levels. Immunoblot of whole cell lysate from GFP-tagged ESAG, and derivative *GFP* and pan-family RNAi lines. Ablation of ESAGs and (GR)ESAG families fails to induce expression of the stumpy-form associated PAD1, but showed variable effects on TfR levels. RNAi lines were grown in the absence or presence of



tetracycline for 24 and 72 hours (h.p.i = hours post induction). Overgrown pleomorphic EATRO 1125 cells included as a positive control for PAD1 expression; overgrown parental VSG221-positive S16 cells also included (two left-hand-most lanes). Chimeric protein detected with monoclonal antibodies against GFP; TfR with TfR specific polyclonal antibody,



and PAD1 with a PAD1 anti-peptide antibody. Exposure times are shown on their respective blot. The position of ESAG6 and ESAG7 bands is indicated. Nitrocellulose membrane stained with Ponceau-S provided as a loading control. Arrowheads indicate Ponceau-S stained VSG221, VSG121 or VSG224 bands.

the non-induced sample. This is inconsistent with results presented in Figure 26, where ESAG10-GFP was fully ablated when induced for the same period of time. Tetracycline may have been omitted from the culture used to produce this sample, an issue that will be addressed later. GFP-ESAG5 and ESAG10-GFP pan-family RNAi cell lines appear to have switched VSG (based on visualisation of the VSG band in these samples by Ponceau-S; Figure 28). As discussed in Chapter 3, these lines may have switched away from the previously active BES and remain resistant to hygromycin due to leaky expression from the now silent BES. The reason for this switch is unknown.

ESAG6 and ESAG7 are the component parts of the *T. brucei* TfR. Given that approximately 20% of *ESAG6* mRNA is derived from silent BESs (Ansorge *et al.* 1999), we decided to investigate if expression of TfR from other BESs was sufficient to prevent a growth defect upon ablation of GFP-ESAG6 expressed from the active BES. Whole cell lysates previously probed with anti-PAD1 antibody were probed with anti-TfR antibody (a kind gift of Piet Borst, Figure 28). ESAG6 and ESAG7 encoded on BES1 have expected molecular weights (after cleavage of predicted signal peptides and GPI-anchor addition sequence) of 39 and 36 kDa, respectively. Of the three major bands labelled by anti-TfR antibody, ESAG7 is presumed to be the smallest, migrating to around 40 kDa. The larger two bands are expected to be ESAG6, migrating to approximately 60 and 45 kDa. These distinct migration patterns may be due to heterogeneity in the α-galactose side chains of the GPI anchor of ESAG6 (Mehlert and Ferguson 2007). These two bands appear to be ~25 kDa larger in GFP-ESAG6 lines. This corresponds with the size of GFP fused to ESAG6, providing further evidence for the identity of these bands. Both of these TfR subunits are known to be glycosylated, which may be the cause of variation between predicted and experimental molecular weights (Mehlert *et al.* 2012). A band of ~100 kDa apparent molecular weight is also visualised in some samples, and may represent the two monomers in complex as the BES1 TfR. As observed for ESAG12-GFP, production of whole cell lysates at room temperature may not be harsh enough to cause dissociation of these complexes. This antibody was produced through

immunisation with TfR derived from BES1 (Bitter *et al.* 1998). ESAG6 and ESAG7 encoded on different BESs share a high degree of identity at the amino acid level (~80-100%), but the cross-reactivity of this polyclonal antibody is unknown. However, signal observed in BES1-active S16, BES3- and BES7-active SmOx and EATRO 1125 cell line lysates suggests that this anti-TfR antibody exhibits a high-degree of cross-reactivity with different *T. brucei* TfRs. It is unclear if the lower TfR levels observed in EATRO 1125, BES3- and BES7-active lines compared to BES1-active lines is due to reduced binding of anti-TfR antibody, or expression of less TfR from these BESs. Variation in the size of TfR subunits can also be observed between different active BESs; this is presumably due to different glycosylation patterns between TfRs.

Knockdown of GFP-ESAG6 by both *GFP* and pan-family RNAi is visualised when lysates are probed with anti-TfR and anti-GFP antibodies. ESAG7 levels are also reduced upon pan-family RNAi induction, as expected. Contrary to our hypothesis that TfR from other BESs would supplement GFP-ESAG6 ablated by *GFP* RNAi, no additional ESAG6 or ESAG7 signal is detectable. In fact, signal derived from ESAG7 is also reduced. Either up-regulated TfR from silent BESs is not recognised by this antibody, or the amount of BES1 TfR that remains when *GFP* RNAi is induced is sufficient for cells to grow normally. Knockdown of other ESAGs and (GR)ESAG families also appears to affect TfR levels; however in some cases small variations may be due to inconsistent loading of lysate sample (for example, 72-hour induced ESAG8-GFP RNAi samples show less Ponceau-S stained total protein and TfR). TfR levels are consistently increased in *(GR)ESAG3* pan-family RNAi lines, even in the absence of tetracycline. TfR levels are reduced in all *(GR)ESAG4* pan-family and 72 hour induced *GFP* RNAi samples compared to the parental and non-induced *GFP* RNAi samples. Ablation of GFP-ESAG5 after 72 hours via *GFP* RNAi results in reduced TfR levels, and in both *(GR)ESAG5* and *(GR)ESAG10* pan-family RNAi lines TfR levels are reduced after 24 hours of induction. However, the VSG band visible by Ponceau-S staining of *(GR)ESAG5* and *(GR)ESAG10* pan-family RNAi cell lines indicates that a switch may have

taken place (as mentioned previously), which could contribute to these results. Further evidence for a VSG switch in *(GR)ESAG10* pan-family RNAi cell lines is the TfR signal, which looks more similar to the BES3-active GFP-ESAG5 lines than its own BES7-active parental line. It is unknown how these various RNAi lines are affecting TfR levels in these cells; further work will be required to elucidate the mechanisms behind this.

To investigate the effect of *GFP* and pan-family RNAi on TfR levels in GFP-tagged ESAG5 and ESAG10 lines (as these lines showed the most significant TfR level variation upon RNAi induction in Figure 28), and to repeat the 24-hour *GFP* RNAi induction in ESAG10-GFP cells, these lysate samples were reproduced. In order to eliminate potential artefacts introduced by producing whole cell lysates at room temperature (as they were in Figure 28), samples prepared at 95°C (or 50°C for the ESAG10-GFP cell line, and its derivatives) were also probed for the presence of TfR and GFP by immunoblot (Figure 29). Loading of these samples was more consistent than Figure 28, making results easier to interpret. TfR in BES7-active ESAG10-GFP and its derivative *GFP* RNAi lines was still not detectable. After 72 hours of *GFP* RNAi the ESAG6 signal appears, suggesting that a BES switch has occurred in this sample. All other RNAi lines show a similar trend: ESAG7 levels reduce slightly the longer RNAi is induced. This was reproduced between room temperature and heated lysates, and seen in both *GFP* and pan-family RNAi lines in GFP-ESAG5 lines and *(GR)ESAG10* pan-family RNAi samples.

The *(GR)ESAG10* pan-family RNAi line shows a different TfR signal and VSG band than the parental tagged line (Figure 29), which was also observed in Figure 28. The VSG and TfR signal look similar to that present in GFP-ESAG5 lines, suggesting that these cells had reverted to BES3. All GFP-ESAG5 lines appear to express the same VSG, so the possible switch observed in Figure 28 may in fact be an artefact of producing cell lysates at room temperature; as the VSG band appears much clearer when stained by Ponceau-S in heated lysates compared to those prepared at room temperature (Figure 29). To confirm if a VSG switch had taken place in these lines, total RNA was isolated, converted to cDNA, and the actively

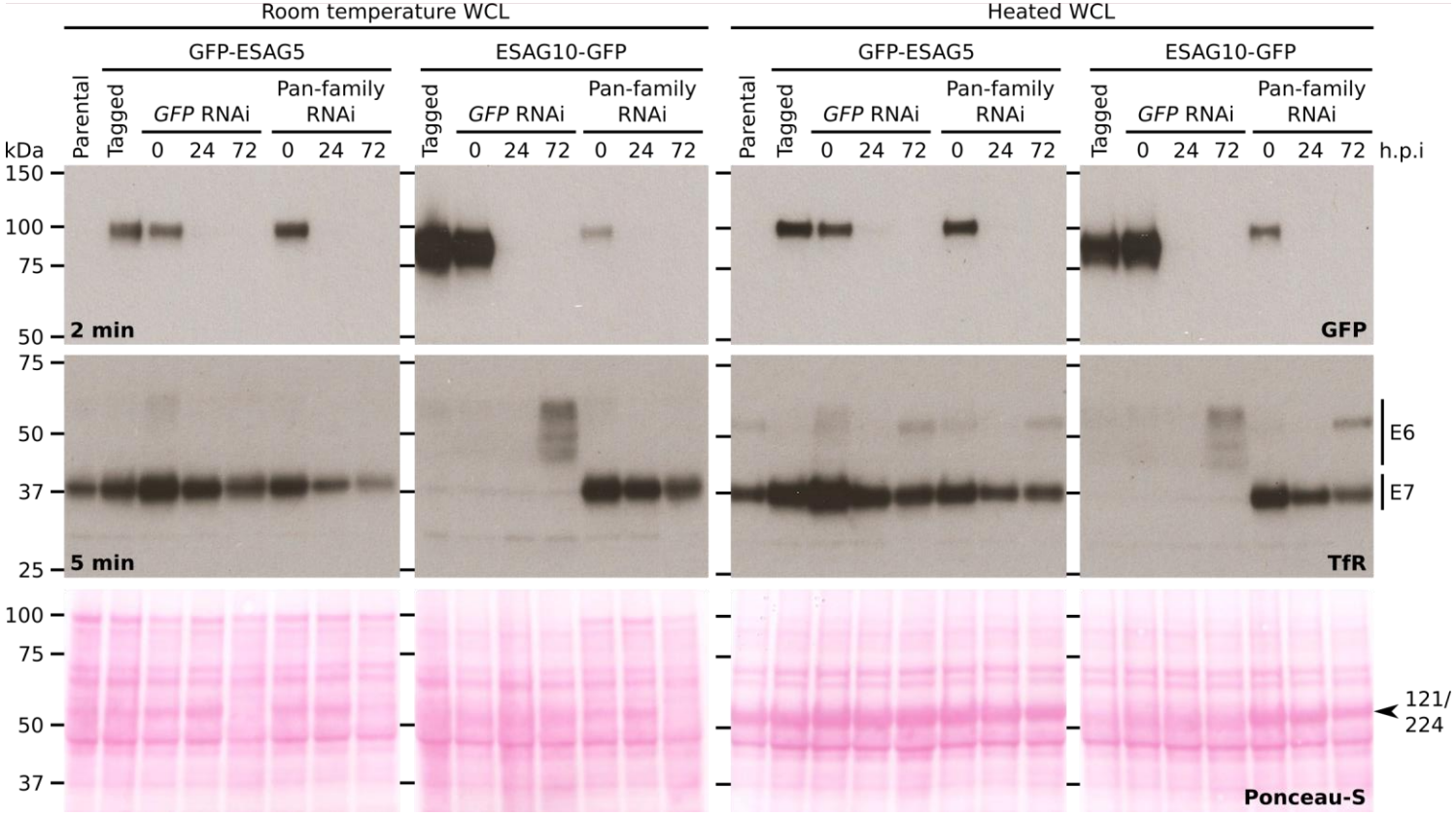


Figure 29. Levels of TfR are reduced upon ablation of (GR)ESAG5 and (GR)ESAG10 families and GFP-ESAG5. Immunoblot of whole cell lysate of GFP-tagged ESAG5 and ESAG10 lines, and derivative *GFP* and pan-family RNAi lines. Samples were prepared at room temperature or by heating at 95°C (or 50°C in ESAG10-GPF lines). RNAi lines were grown in the absence or presence of tetracycline for 24 and 72 hours (h.p.i = hours post induction). Chimeric protein detected with monoclonal antibodies against GFP, and TfR with anti-TfR polyclonal antibody. Exposures of 2 and 5 minutes were used for GFP and TfR blots, respectively. The position of ESAG6 and ESAG7 bands is indicated. Nitrocellulose membrane stained with Ponceau-S provided as a loading control. Ponceau-S stained VSG121 and VSG224 bands are indicated by an arrowhead.

transcribed *VSG* amplified and sequenced (as conducted in Section 3.4). Cryostocks of these lines, frozen immediately after transfection clone selection, were defrosted and also investigated to confirm a possible VSG switch. This showed that all GFP-ESAG5 RNAi line derivatives express VSG121, so are BES3-active (data not shown). ESAG10-GFP lines had indeed switched: *GFP* RNAi lines and the cryostock pan-family RNAi line were all BES7-active, while the pan-family RNAi line used to produce lysates in Figures 28 and 29 had reverted to BES3 (data not shown). This line appears to be stably BES3-active, and survive hygromycin challenge through leaky expression from BES7, which accounts for the reduced levels of ESAG10-GFP observed. This may be the cause of the slower growth rate shown by this line (Figure 27), and represents a similar phenotype to ESAG10-GFP clone 2 discussed in Chapter 3.

4.6. Summary

Here I describe the use of PSI-BLAST to identify (GR)ESAG family members present in TriTrypDB release 9.0. Pan-family RNAi constructs capable of knocking down ESAG and GRESAG family members were then designed and introduced into GFP-tagged ESAG lines. *(GR)ESAG5* and *(GR)ESAG10* families are so divergent at the DNA level that additional RNAi constructs were required for full coverage of pan-family RNAi. *GFP* RNAi constructs were also introduced into GFP-tagged ESAG lines to enable specific knockdown of the actively expressed BES-linked ESAG family member. Varying degrees of chimeric protein ablation was achieved with these RNAi constructs, but (GR)ESAG family knockdown induced growth defects in the majority of cases, while knockdown of the ESAG family member alone did not reproduce this phenotype. These data provide evidence for intra-family functional redundancy, and that (GR)ESAG families, but not the actively expressed ESAG family member, are essential for parasite survival *in vitro*. For (GR)ESAG families with multiple copies encoded on BES1 (i.e. ESAG3, ESAG4 and ESAG8) the *GFP* RNAi in these lines will only knockdown the GFP-tagged copy, so a putatively functional

copy will also be transcribed from the BES in these lines. This limitation will be addressed in Chapter 5.

Knockdown of ESAG4s and two closely related GRESAG4s is known to induce a severe cytokinesis defect (Salmon *et al.* 2012a), which was not reproduced by my *(GR)ESAG4* pan-family RNAi line using a very similar targeting fragment. Only pan-*(GR)ESAG4* and *(GR)ESAG12* family RNAi failed to induce growth defects, but also failed to fully ablate chimeric protein, the remains of which may be sufficient to fulfil the roles of these families and enable cells to survive. The use of different, or additional, RNAi fragments may improve the degree of knockdown and could be pursued in the future. Generally, the phenotype induced by my pan-family RNAi constructs is less severe than previously published examples (Salmon *et al.* 2012a; Batram *et al.* 2014; Tiengwe *et al.* 2016). There are numerous potential sources for this variability, including the use of different cell lines and culture methods, and also the penetrance of RNAi achieved through use of different RNAi constructs, their site of integration and RNAi targeting fragment. The parental line used here (for all except GFP-tagged ESAG5 and ESAG10, and derivative RNAi lines) is also used in these other studies (Wirtz *et al.* 1999), so is unlikely to be the source of variability; however, different RNAi constructs may be. Batram *et al.* and Salmon *et al.* utilise earlier derivatives (Shi *et al.* 2000; LaCount *et al.* 2000) of the RNAi construct used here, that integrate into the rDNA locus as opposed to minichromosomes. Tiengwe *et al.* employs the pLEW100v5X:Pex11 construct to produce dsRNA via stem-loop formation (Silverman *et al.* 2011). These differences may explain the disparity between our results.

The *(GR)ESAG8* pan-family RNAi construct used here only targets knockdown of ESAG8s due to low sequence similarity shared between *ESAGs* and the two *GRESAGs* of this family. Despite this, knockdown of ESAG8s induces a growth defect, suggesting that the distantly related GRESAG8s cannot complement ESAG8 function. This is an odd discovery given that BES12, which does not encode an ESAG8, can be activated *in vitro* (Hertz-Fowler *et al.* 2008), and ESAG8 can be conditionally knocked-out of the active BES (Hoek and Cross 2001). *ESAG8* transcript was reduced to

4% in this knockout line and encoded protein could not be detected, but cells grew at a similar rate to the wild type line (Hoek and Cross 2001). As with all RNAi experiments the results observed here, and previously (Batram *et al.* 2014), could be due to off-target effects.

BES1-active cells consistently grow at a faster rate than BES3- and BES7-active cells *in vitro*. This feature could be inherent to the active BES, a result of the repertoire of ESAGs encoded, or a biochemical property of the active VSG. TfRs derived from different BESs are known to have variable affinities for different mammalian transferrin (Bitter *et al.* 1998; Gerrits *et al.* 2002). If the TfR encoded on BES3 or BES7 has a lower affinity for bovine transferrin (in which these cells are grown) than BES1 TfR, this could limit the rate of growth of these cells. Equally, any of the less well-characterised ESAGs encoded by these BESs may be responsible for this limited growth rate via a similar mechanism. If VSGs encoded on these BESs fold at a slower rate, or have a less common codon usage than VSG221, this could also affect growth. Alternatively, this difference could be due to mutations in the BES1-active S16 line that make it better adapted to life in culture than the BES3-active WT line that was subsequently made competent for RNAi and used in the tagging of ESAG5 and ESAG10 (Chapter 3). The further reduced growth rate of the *(GR)ESAG10* pan-family RNAi line appears to be a result of reverting to BES3 from BES7, and reliance on leaky transcription of *HYG* from an inactive BES to survive this selection pressure. However, this does not explain the slower growth rate of the *(GR)ESAG5* pan-family RNAi line. It is unclear why the *(GR)ESAG10* pan-family RNAi line has switched back to BES3, or what advantage this provides to cells.

Ablation of ESAGs and (GR)ESAG families does not induce PAD1 expression. Batram *et al.* show that RNAi using fragments of *ESAG1*, *ESAG2* and *ESAG8* ORFs was sufficient to activate a PAD1 expression reporter line, but PAD1 levels were never directly investigated (Batram *et al.* 2014). However, endogenous PAD1 was visualised by immunofluorescence microscopy with anti-PAD1 antibody upon BES transcription attenuation induced by expression of a second VSG from the rDNA locus (Batram *et al.* 2014). The same RNAi-competent parental cell line was used in my study,

but does not appear able to produce PAD1 when ESAGs or (GR)ESAG families are ablated. These cells, despite having the same origin, have been cultured independently for many years and therefore may have distinct genotypes. It is possible that the cell line used here is incapable of producing PAD1 *in vitro*, while others apparently are. To investigate these differing results, the PAD1 reporter utilised by Batram *et al.* could be introduced into the RNAi lines described here, and GFP levels monitored upon ESAG and (GR)ESAG family ablation. Also, the presence of different stumpy form markers could be probed in these RNAi lines to answer if disruption of ESAG or (GR)ESAG family homeostasis does induce differentiation. Alternatively, the DNA content of these cells could be monitored by flow cytometry of propidium iodide-stained samples to see if RNAi induces accumulation of G1 cells, or arrest in other stages of the cell cycle.

Levels of TfR were affected upon knockdown of some ESAGs and (GR)ESAG families. Interestingly, none of the three ESAGs claimed to be involved in BES attenuation by Batram *et al.* (ESAG1, ESAG2 and ESAG8) had an effect on TfR levels when ablated, which one would expect if transcription from the active BES is being silenced. Further investigation is required to understand the mechanisms behind this response. Despite ~20% of *ESAG6* mRNA originating from silent BESs (Ansorge *et al.* 1999), TfR produced from this transcript could not be visualised when probing with anti-TfR antibody by immunoblot. However, this may be due to lack of antibody cross-reactivity with these TfRs. The leaky transcription of *ESAG6* from silent BESs was measured in cells where *ESAG6* had been knocked out of the active BES. *T. brucei* is known to adapt to growth in serum containing transferrin for which the actively expressed TfR has a low affinity, through a number of mechanisms (Gerrits *et al.* 2002; Mußmann *et al.* 2004; Van Luenen *et al.* 2005), which may have occurred in this *ESAG6* knockout line to compensate for the loss of this TfR component from the active BES. Therefore, lines with functional TfR expressed from the active BES may not transcribe *ESAG6* from silent BESs, hence its absence here. Perhaps surprisingly, ablation of ESAG6 derived from the active BES did not affect cell growth, given that both ESAG6 and ESAG7 are required to form

functional heterodimeric TfR (Steverding *et al.* 1994; Salmon *et al.* 1994; Chaudhri *et al.* 1994). Other functional TfR may be present in these cells that is not detectable with this antibody. Alternatively, these cells may be able to survive with any remaining ESAG6 not removed by knockdown, as only a small amount of iron is required for parasite growth (Steverding 1998), or through depletion of intracellular iron stores (Taylor and Kelly 2010); if *GFP* RNAi induced cells were monitored over a longer period a defect may become apparent.

Further investigation of the phenotypes induced in pan-family and *GFP* RNAi lines may shed light on the functions of these families and disentangle the roles played by ESAG and GRESAG family members. This, in turn, could aid in understanding the linking of ESAGs and *T. brucei* BESs. The work presented in this chapter highlights the importance of GRESAGs when investigating these families, and that these generally overlooked family members may in fact be as, if not more, important to parasite survival *in vitro* than ESAGs. Given that these families (the ESAG members in particular) are hypothesised to play roles in adapting the parasite to life in different mammalian serum, investigating phenotypes of these RNAi lines *in vivo* represents an important next step. In this environment, with additional pressures from the host and different availability of nutrients, the importance of ESAGs may be revealed.

Chapter 5. Generation of a BES with reduced *ESAG* repertoire

5.1. Introduction

Data presented in Chapter 4 indicates that expression of individual ESAGs from the active BES is not essential *in vitro*, as assessed by RNAi-mediated knockdown of these proteins. Therefore we hypothesised that the repertoire of ESAGs expressed from the active BES could be ablated, without detriment to the parasite. RNAi is not a viable technique to achieve this, but the repertoire of *ESAGs* encoded on BES1 could be reduced through knockout, achieving full ablation of these proteins. The drawbacks of traditional gene knockout in trypanosomes outlined in Section 4.1 are alleviated somewhat by the presence of BES sequences as single copy in the parasite genome; therefore requiring knockout of only one allele, and use of a single selection marker.

We theorised that removal of the multicopy *ESAG* genes on BES1 was likely to be possible, given that a number of the other BESs do not encode these additional copies and have been shown to be active *in vitro* (Hertz-Fowler *et al.* 2008). Once these genes have been knocked out, the remaining ESAG of each family can be tagged with GFP, and investigated with *GFP* RNAi as in Chapter 4. This would remedy the limitation that *GFP* RNAi in GFP-tagged ESAG3, ESAG4 and ESAG8 lines only ablates one of the copies expressed from BES1.

We also aimed to investigate if the entire *ESAG* complement encoded by BES1 could be removed. However, this appeared to be a more improbable outcome, given that expression of a functional TfR from the active BES appears essential to the cell. This is based on the inability to activate BES8 *in vitro*, which has a greatly reduced *ESAG* repertoire – only encoding *ESAG7* (Hertz-Fowler *et al.* 2008). Also, transfection of knockout constructs targeting *ESAG6/7* 5’ and 3’ UTRs removed either gene with a similar frequency, but never both simultaneously, suggesting that this result is detrimental to the cell (Ansorge *et al.* 1999).

Based on this reasoning, three distinct knockout schemes were devised: “Multicopy ESAG knockout”, where additional copies of *ESAG3*, *ESAG4* and *ESAG8* are removed from BES1, leaving just one putatively functional copy of each; “Partial ESAG knockout”, where all *ESAGs* except those encoding the TfR (*ESAG6* and *ESAG7*) are knocked out of BES1; and “Complete ESAG knockout”, where the entire BES1 *ESAG* repertoire is excised.

Two techniques were employed to achieve this: traditional knockout via homologous recombination, and Cre recombinase-mediated excision of sequence. The Cre/*loxP* system has been adapted for use in a plethora of organisms for numerous applications due to the reliability and predictability of genomic modification it enables. Briefly, Cre recombinase catalyses the excision of sequence flanked by 34-bp *loxP* sites in direct orientation to an episome, leaving a single *loxP* “scar”. Cre can also invert this intervening sequence when *loxP* sites are in opposing orientation. Cre recombinase was first used in PCF *T. brucei* by Barrett *et al.*, and further developed for use in BSF by the Cross lab (Barrett *et al.* 2004; Scahill *et al.* 2008; Kim *et al.* 2013). This was driven by the need to preserve the few selection markers available for use in *T. brucei* (requiring one per allele when knocking out diploid genes by traditional homologous recombination, but can be deleted and re-used when flanked by *loxP*), which makes knockout of multiple genes difficult, and the need for a conditional knockout strategy to study essential genes. These make use of the negative selection marker herpes simplex virus thymidine kinase (HSV-TK) which, when expressed, induces sensitivity to the nucleoside analogue ganciclovir. This drug is phosphorylated by HSV-TK, and further phosphorylated by cellular kinases to the triphosphate form, which can be incorporated into DNA and inhibit elongation. *HSV-TK* is deleted upon Cre-mediated recombination in these strategies, inducing a reversion to ganciclovir resistance.

This chapter details a number of attempted strategies to achieve reduction of the BES1 *ESAG* repertoire, using both homologous recombination and Cre/*loxP* techniques.

5.2. Homologous recombination-based modification of BES1

In order to reduce the *ESAG* repertoire of BES1, as described in Section 5.1, the traditional homologous-recombination knockout strategy was first employed. To direct homologous recombination, and thus deletion of intervening sequence, targeting fragments of ~500 bp were amplified from BES1 and ligated to a selectable marker ORF (in this case *BLAST*) by fusion PCR, to produce KO fragments (Figure 30). These fragments are capable of knocking out the multicopy *ESAGs* (KO Multi), all but *ESAG6* and *ESAG7* (KO Par) and completely removing *ESAGs* from BES1 (KO All, as outlined in Section 5.1). Targeting fragments originated from *ESAG* UTRs in each instance. Therefore, upon homologous recombination of the KO fragment into BES1, *BLAST* would be completed with both 5' and 3' UTRs, to ensure correct processing and translation of *BLAST* mRNA. For example, to knockout all *ESAGs* on BES1, the 5' targeting fragment used *ESAG7* 5' UTR sequence ending at the last base upstream of the ORF start codon, and 3' UTR from the first base downstream of the *ESAG1* stop codon as 3' targeting sequence. The same 3' targeting fragment can be utilised by both KO All and KO Par constructs, so it is labelled as the KO All/Par 3' targeting fragment (Figure 30). Amplification of KO fragments by PCR provided sufficient material to introduce into cells, omitting the need for an intermediate cloning step.

KO fragments were used to transfect VSG221-positive (BES1-active) S16 and VSG121-positive (BES3-active) SmOx cell lines. BES1-active cells were transfected as the fragments are designed to integrate into this BES, facilitating transcription of the selection marker and survival of positive transfectants. KO fragments were also introduced into a BES3-active (BES1-silent) line in case integration of the fragments into BES1 was lethal to cells, which should not affect a BES3-active line. This control requires leaky transcription of *BLAST* from BES1 in these BES3-active cells to survive the selection pressure (given the survival of lines with selection markers encoded on silent BESs presented in Chapters 3 and 4, this is not implausible). Efficiency of transfection with KO fragments was poor, with only two clones

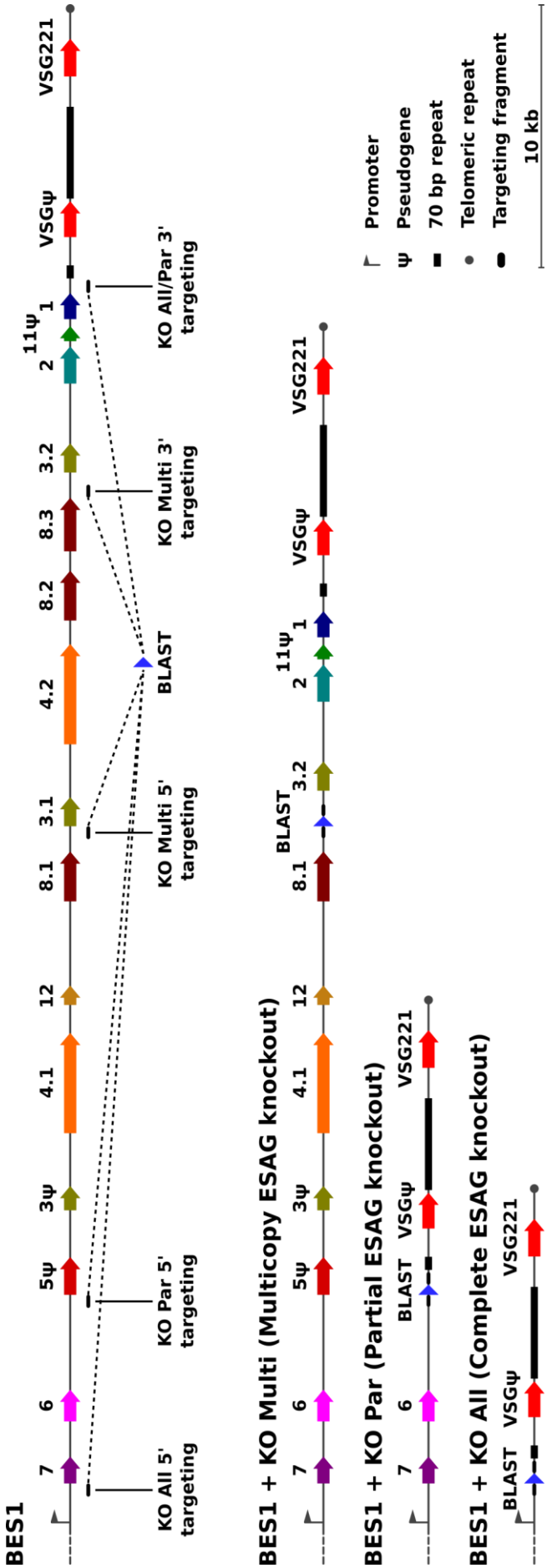


Figure 30. A homologous recombination-based strategy to reduce the *ESAG* complement of BES1. Schematic to illustrate the strategy to reduce the repertoire of *ESAGs* encoded by BES1. *BLAST* ORF is flanked by targeting sequence homologous to BES1 in order to induce excision of sections of the BES upon recombination. Three knockout schemes were designed: removing *ESAG3*, *ESAG4* and *ESAG8* genes to leave only one copy of each (Multicopy ESAG knockout facilitated by the KO Multi fragment); removal of all *ESAGs* except *ESAG6* and *ESAG7* (Partial ESAG knockout by KO Par knockout fragment); and excision of the entire BES1 *ESAG* repertoire (Complete ESAG knockout achieved using the KO All fragment). Diagrams drawn to scale.

growing out of selection with blasticidin from six independent transfections (three KO fragments with S16 or SmOx cell lines). These two clones resulted from transfection of S16 (BES1-active) cells with the KO Multi fragment, and SmOx (BES3-active) cells with the KO All fragment. Success of knockout was investigated by Southern blot. As shown previously, digestion of the *T. brucei* genome by SwaI endonuclease *in silico* produces a ~58 kb restriction fragment containing BES1. Knockout of *ESAG* genes will reduce the size of this fragment to ~46 and ~13 kb in the case of Multicopy and Complete ESAG KO strategies, respectively (Figure 31A). When genomic DNA of these two clones and their parental lines were digested by SwaI, separated by PFGE and probed for presence of *VSG221* and *BLAST* (introduced by the KO fragment), the expected shifts in fragment sizes were not detected (Figure 31B). In fact, *BLAST* appears to have been knocked-in to BES1 in the case of S16 + KO Multi, as evidenced by the small increase in the BES1 restriction fragment size. Introduction of the KO All fragment into SmOx cells resulted in integration into an unknown location. Both of these mis-integration events have produced cells resistant to blasticidin, so able to express BLAST, but have not successfully knocked out BES1 *ESAGs*.

A number of factors may have influenced the low efficiency of clone selection in this experiment, and inability to knockout sections of BES1. Firstly, the targeting fragments used were not unique to BES1, as the UTR sequences are conserved across *ESAGs* on other BESs, and potentially *GRESAG* family members, providing additional sites for homologous recombination to occur. Producing KO fragments by fusion PCR may not have provided enough DNA for transfection, and if additional clones had been obtained through use of greater DNA mass, cell lines with correctly integrated fragments may have been obtained. The only blasticidin resistant clones from transfection with these KO fragments had arisen due to mis-integration of the construct. This suggests that *BLAST* may not be transcribed properly upon correct integration into BES1. Also, BLAST had not been used previously in the lab as a selection marker transcribed from a BES, and the consequence of this unknown.

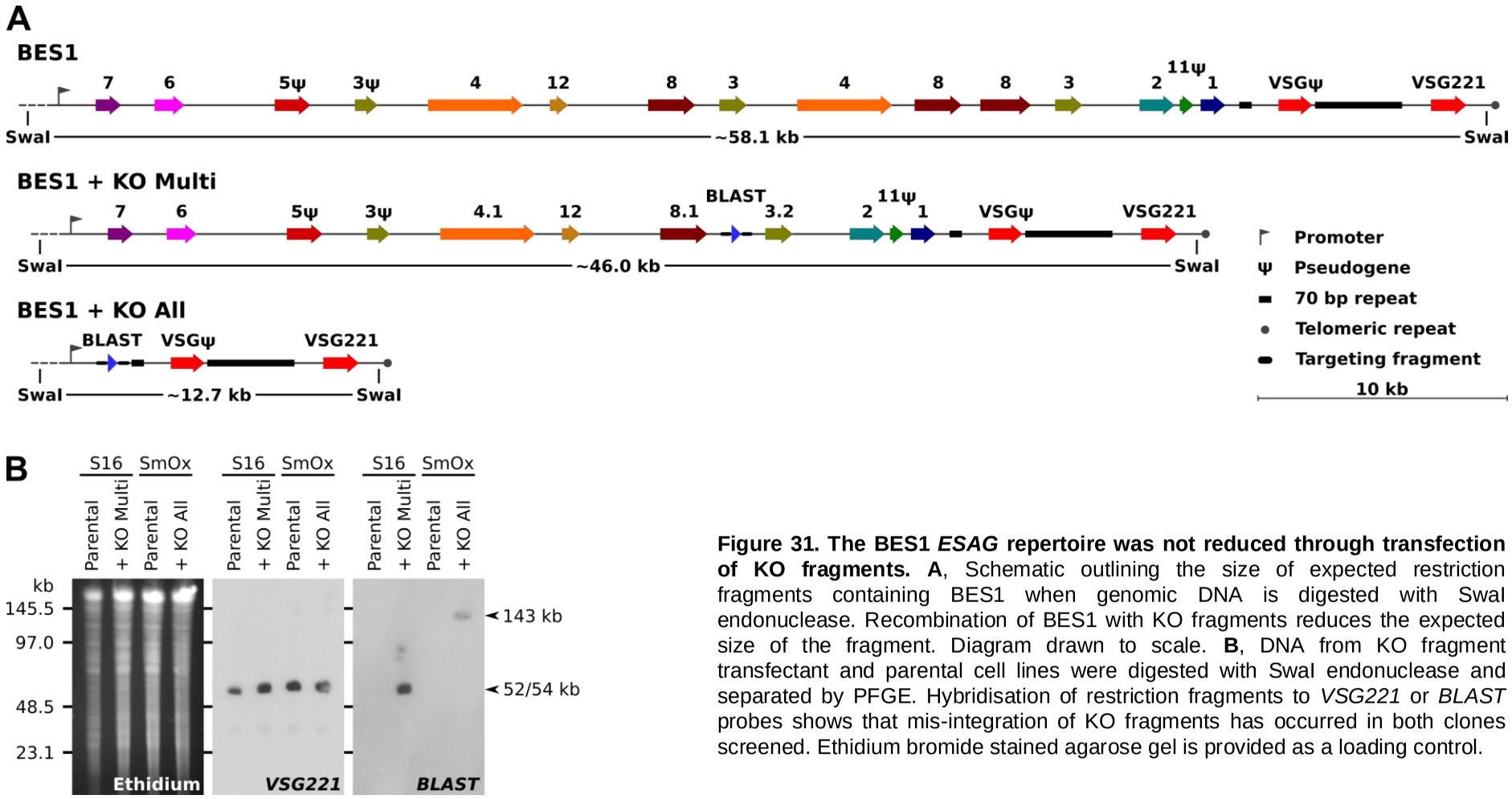


Figure 31. The BES1 *ESAG* repertoire was not reduced through transfection of KO fragments. **A**, Schematic outlining the size of expected restriction fragments containing BES1 when genomic DNA is digested with SwaI endonuclease. Recombination of BES1 with KO fragments reduces the expected size of the fragment. Diagram drawn to scale. **B**, DNA from KO fragment transfectant and parental cell lines were digested with SwaI endonuclease and separated by PFGE. Hybridisation of restriction fragments to *VSG221* or *BLAST* probes shows that mis-integration of KO fragments has occurred in both clones screened. Ethidium bromide stained agarose gel is provided as a loading control.

To remedy these points, a number of modifications were implemented. *BLAST* was replaced by *HYG* as this marker had previously been introduced to expression sites, and was able to select for transfectants (Chapter 3). Also leaky expression of *HYG* from silent BESs was sufficient for cells to survive challenge with hygromycin. Actin UTRs flanking the selection marker were incorporated to promote its expression. This allowed use of different targeting fragments, as the new KO fragments need not utilise *ESAG* UTRs (Figure 32). To prevent knock-in of *HYG*, new targeting fragments for KO Multi were selected. This construct will now integrate between *ESAG12* and *ESAG8.3*, leaving *ESAG8.3* as the single-copy representative of the (GR)ESAG8 family (Figure 32), as opposed to *ESAG8.1* as previously. Also, a fragment of the BES1 *VSG* pseudogene will be used as KO All/Par 3' targeting fragment (Figure 32). This sequence is unique to BES1, promoting integration into this expression site, and preventing the ability of these KO fragments integrating elsewhere in the genome due to sequence shared by *(GR)ESAG1* family members. Disruption of this pseudogene is not detrimental to the parasite (Davies *et al.* 1997). These new KO fragments were cloned into pBlueScript II SK (+) (a kind gift of Bill Wickstead) to obtain sufficient DNA mass for transfection, and could be released from the vector prior to transfection by digestion with BamHI and XhoI endonucleases.

As before, these new KO fragments were used to transfect BES1-active S16, and BES3-active SmOx cell lines. More clones were obtained than with the previous strategy, and again screened for integration by Southern blot. Expected sizes of the BES1-containing SwaI restriction fragment were approximately 46, 19 and 12 kb upon integration of KO Multi, KO Par and KO All, respectively (Figure 33A). To maximise the number of cell lines screened, surviving transfection recovery populations were plated to limiting dilution and clones obtained from these populations are preceded with "P" (Figure 33B). Investigation of PFGE-separated SwaI-digested genomic clone DNA identified only a single transfectant with a smaller than parental BES1 restriction fragment (~23 kb) that hybridised with both *VSG221* and *HYG* probes (Figure 33B). This line, SmOx + KO Par clone P1, shows additional fragments that hybridise with *VSG221* and *HYG* probes, suggesting that

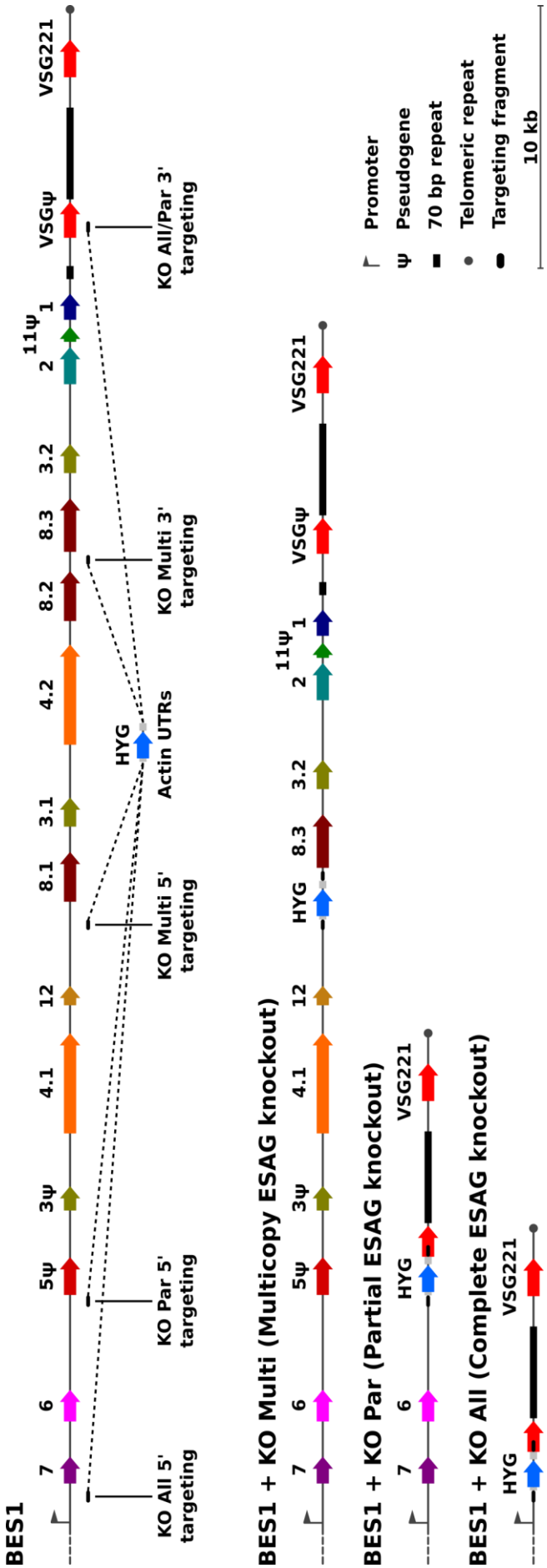


Figure 32. An updated homologous recombination-based strategy to reduce the *ESAG* complement of BES1. Schematic to illustrate the updated strategy to reduce the repertoire of *ESAG*s encoded by BES1. *HYG* ORF is flanked by actin UTRs and targeting sequences homologous to BES1 in order to induce excision of sections of the BES upon recombination. The three knockout schemes have the same outcome as outlined previously. Diagram drawn to scale.

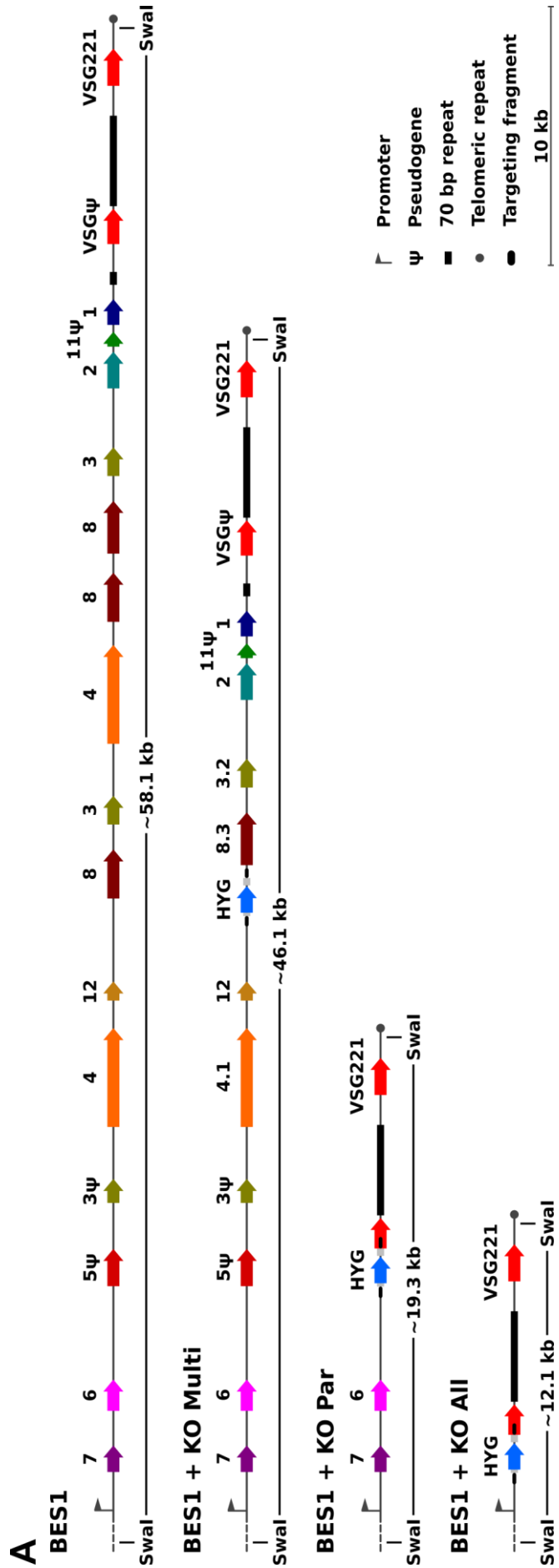


Figure 33. The BES1 *ESAG* repertoire was not reduced through transfection of updated KO fragments. A, Schematic outlining the size of expected restriction fragments containing BES1 when genomic DNA is digested with SwaI endonuclease. Integration of updated KO fragments into BES1 reduces the expected size of the fragment. Diagram drawn to scale.

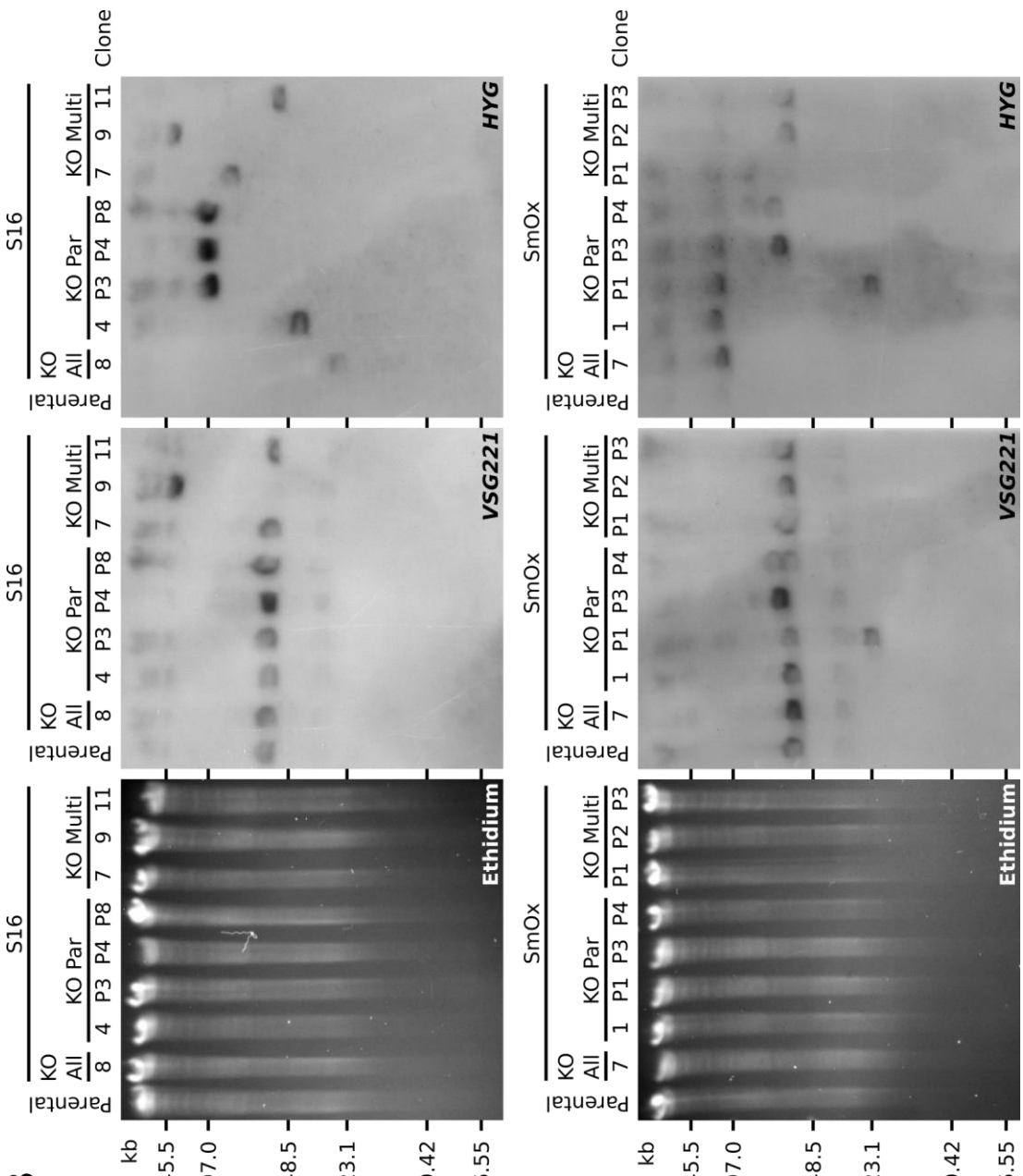


Figure 33. The BES1 *ESAG* repertoire was not reduced through transfection of updated KO fragments. B, DNA from parental cell lines, and clones resulting from their transfection with updated KO constructs, were digested with SwaI endonuclease and separated by PFGE. Hybridisation of restriction fragments to *VSG221* or *HYG* probes shows that transfection of these KO fragments has failed to reduce the *ESAG* repertoire of BES1. Ethidium bromide stained agarose gel is provided as a loading control.

BES1 may have been duplicated, and *HYG* copied elsewhere in the genome. Alternatively, this line may not be clonal, but a population of two clones with different KO Par integrations: one into BES1, one elsewhere in the genome. SmOx + KO Par clone P4 also shows multiple *VSG221* containing fragments, and again may actually contain multiple lines. In other clones, including S16 + KO Multi 11, SmOx + KO Par P3 and P4, and SmOx + KO Multi P2 and P3, *HYG* appears to have integrated into BES1 without knocking out the desired fragments. The BES1 fragment appears to be slightly larger, or a similar size to the parental in these lines. *HYG* appears to have integrated elsewhere in the genome in all other clones screened. It is unknown why integration of these KO fragments deviates from their design, but given the failure of a number of attempts to reduce the BES1 *ESAG* repertoire by homologous recombination, a new approach was required.

5.3. Cre recombinase-based modification of BES1

Given the difficulties encountered when attempting to obtain the desired knockouts via homologous recombination, this strategy was abandoned in favour of Cre recombinase-mediated excision of BES1 fragments. As described in Section 5.1, the DNA fragment to be excised must be flanked by *loxP* sites in direct orientation. Building upon the work of Cross *et al.*, *loxP* sites would be introduced into BES1 with selectable markers, Cre expression induced from a transiently transfected vector, and BES1 with a reduced *ESAG* complement obtained. Given the distance between *loxP* sites (>20 kb), two separate integrations of this DNA element into BES1 were required. This was facilitated by introduction of *loxP* encoding vectors, CreKO_5' and CreKO_3', into BES1 at the upstream and downstream recombination sites, respectively (Figures 34A and 34B). These vectors encode *loxP* sites adjacent to selectable markers. CreKO_5' encodes HSV-TK as a chimera with BLAST (TK-BLAST, coding sequence was a kind gift of Simon D'Archivio) under aldolase 5' and actin 3' UTRs to facilitate transcription (Figure 34A). This enables the use of positive (to ensure integration) and negative (to ensure loss upon recombination) selection, and

is used as a fusion-protein to prevent loss of either marker individually. The CreKO_3' construct utilises PURO as a positive selection marker, under the same UTRs as *TK-BLAST* (Figure 34A). Targeting fragments were added to these constructs to integrate into sites that, upon Cre mediated-recombination, will produce the three knockout schemes detailed in Section 5.1 (Figure 34B). Excision of DNA between *loxP* sites will remove the desired *ESAGs*, as well as *TK-BLAST*, leaving PURO as the only selection marker (Figure 34C).

BES1-active S16 cells were transfected with CreKO_3' constructs (insert sites 4 and 5 in Figure 34B), with transfection of corresponding CreKO_5' construct to follow in correctly integrated lines. Only the BES1-active line was used in this strategy, as integration of these constructs should not have a deleterious effect on cells. Southern blot was again used to screen integration, and ApaI endonuclease used to fragment cell genomes. An ApaI site is present upstream of the BES1 promoter, further than the SwaI site utilised previously, so the BES1-containing restriction fragment is larger than that produced through fragmentation with SwaI, at 77 kb (Figure 35A). Integration of CreKO_3'Multi or CreKO_3'All/Par introduces another ApaI site, so the *VSG221*-containing restriction fragment will be reduced to around 22 or 12 kb, respectively (Figure 35A). When parental genomic DNA was probed for *VSG221*, the labelled restriction fragment appeared much larger than expected, at between 97 and 146 kb (Figure 35B). This suggests that the ApaI site is much further upstream than expected (from the BES1-containing BAC sequence; Berriman *et al.* 2002). CreKO_3'All/Par has not integrated into BES1 in any of the clones screened; while a number of clones have been successfully transfected with CreKO_3'Multi. Variability between labelled restriction fragments is attributed to the remaining telomeric repeats (Figure 35B), which were removed in previous screens where SwaI was used (Figures 31B and 33B). Clones 1, 3, 4, 6, 7, 8 and 10 have integrated CreKO_3'Multi into BES1 based on *VSG221* probe signal (Figure 35B). However, *PURO* probe labelling of the same ~23 kb restriction fragment can only be claimed confidently for clones 6, 7, 8 and 10, due to the high background in this reprobed blot (which is attributed to a reduction in the UV

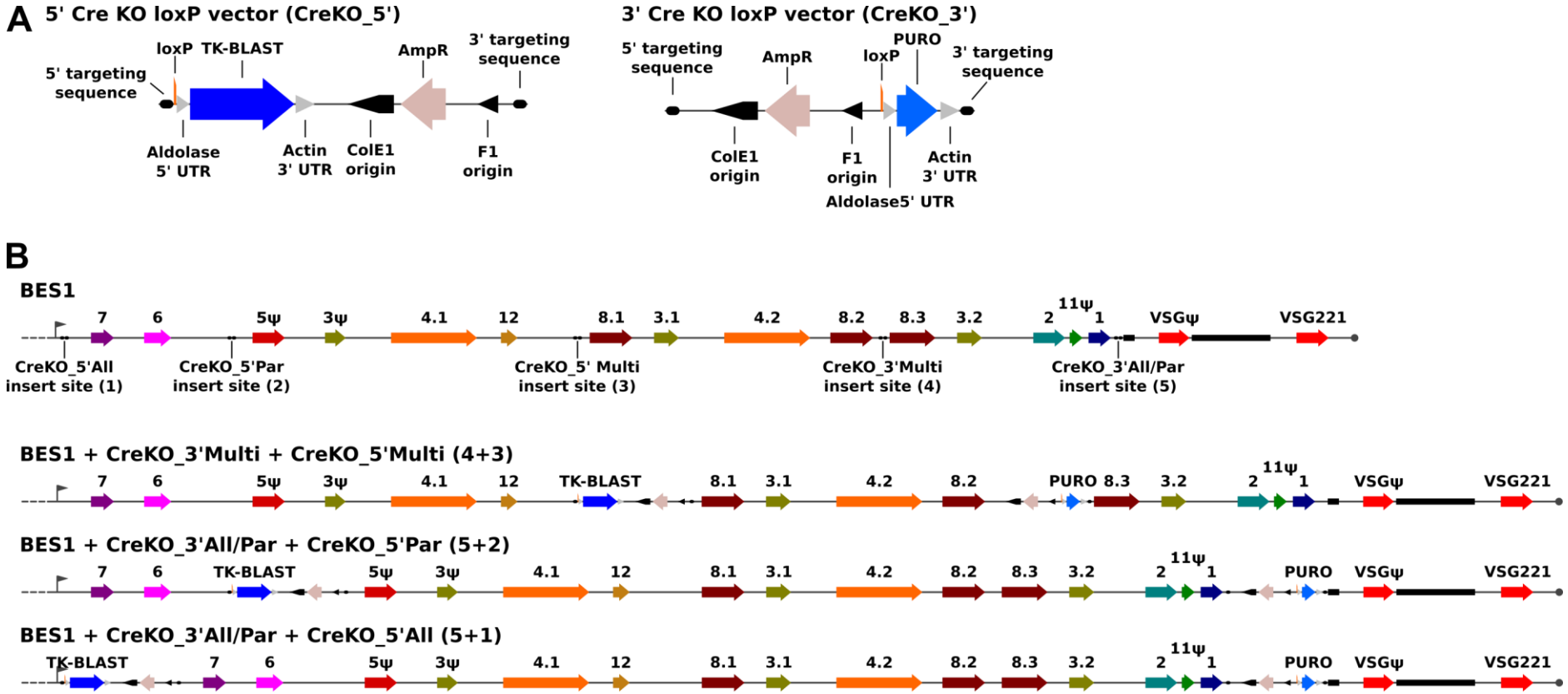


Figure 34. A Cre recombinase-based strategy to reduce the *ESAG* complement of BES1. A, Vectors encoding selectable markers and *loxP* sites, named CreKO_5' and CreKO_3', contain sequence homologous to sections of BES1 outlined in part B. Each knockout scheme (Multicopy, Partial and Complete ESAG knockout) requires one of each of these constructs integrated into BES1, prior to Cre recombinase-mediated excision of sequence surrounded by *loxP* sites. **B**, Schematic to illustrate the Cre recombinase-based strategy to reduce the repertoire of *ESAGs* encoded by BES1. Integration sites of CreKO constructs for each of the three knockout schemes is indicated. Incorporation of these vectors into BES1 is shown.

C

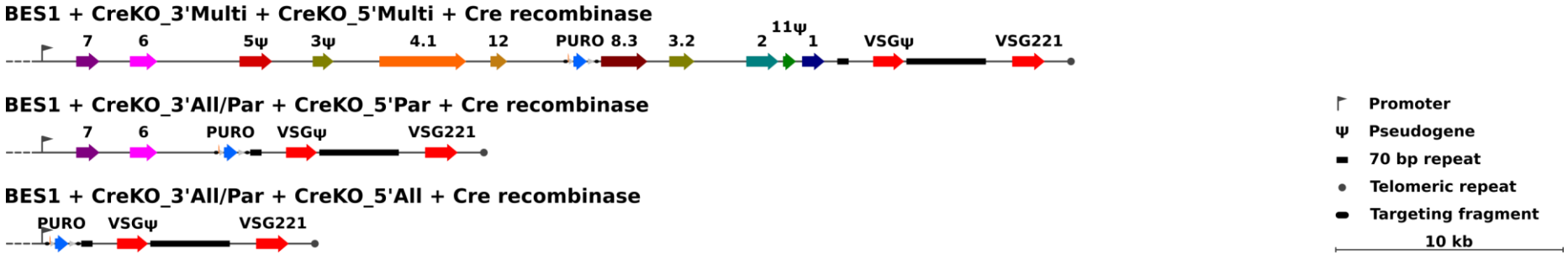


Figure 34. A Cre recombinase-based strategy to reduce the *ESAG* complement of BES1. C, Resulting BES1 sequence in the three knockout schemes following Cre-mediated excision of the *loxP* flanked sequence shown in part B. Diagrams in part B and C are drawn to scale.

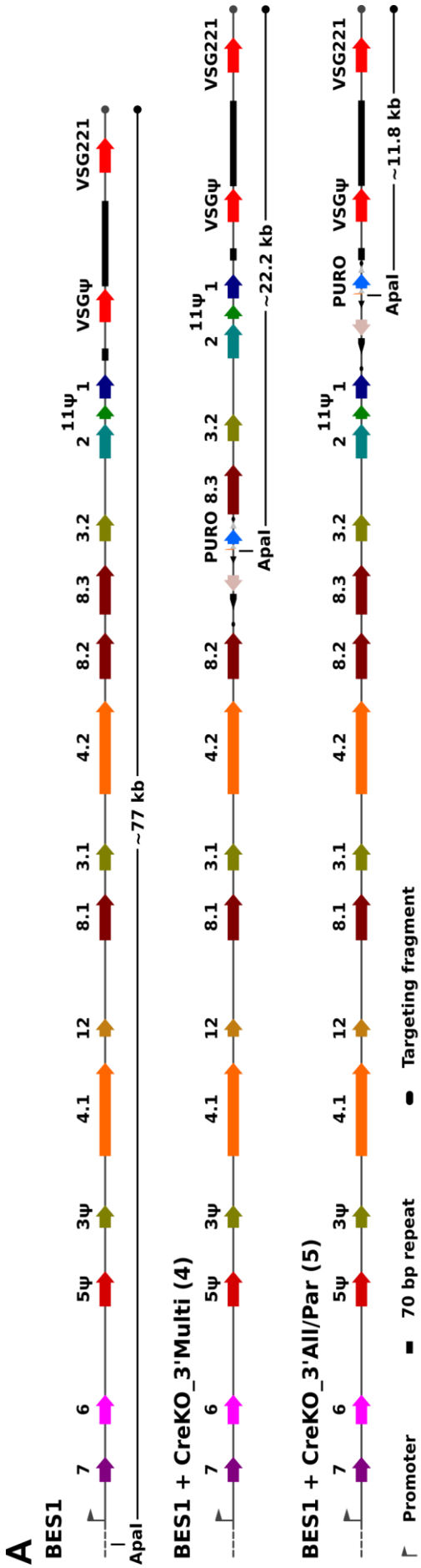


Figure 35. CreKO_3'Multi, but not CreKO_3'All/Par, was integrated into BES1. A, Schematic outlining the size of expected *VSG221*-containing restriction fragments when genomic DNA is digested with ApaI endonuclease. Integration of CreKO_3' constructs into BES1 reduces the expected size of the fragment due to the ApaI site present in these constructs. Insert sites used in labels relate to Figure 34B. Diagram drawn to scale.

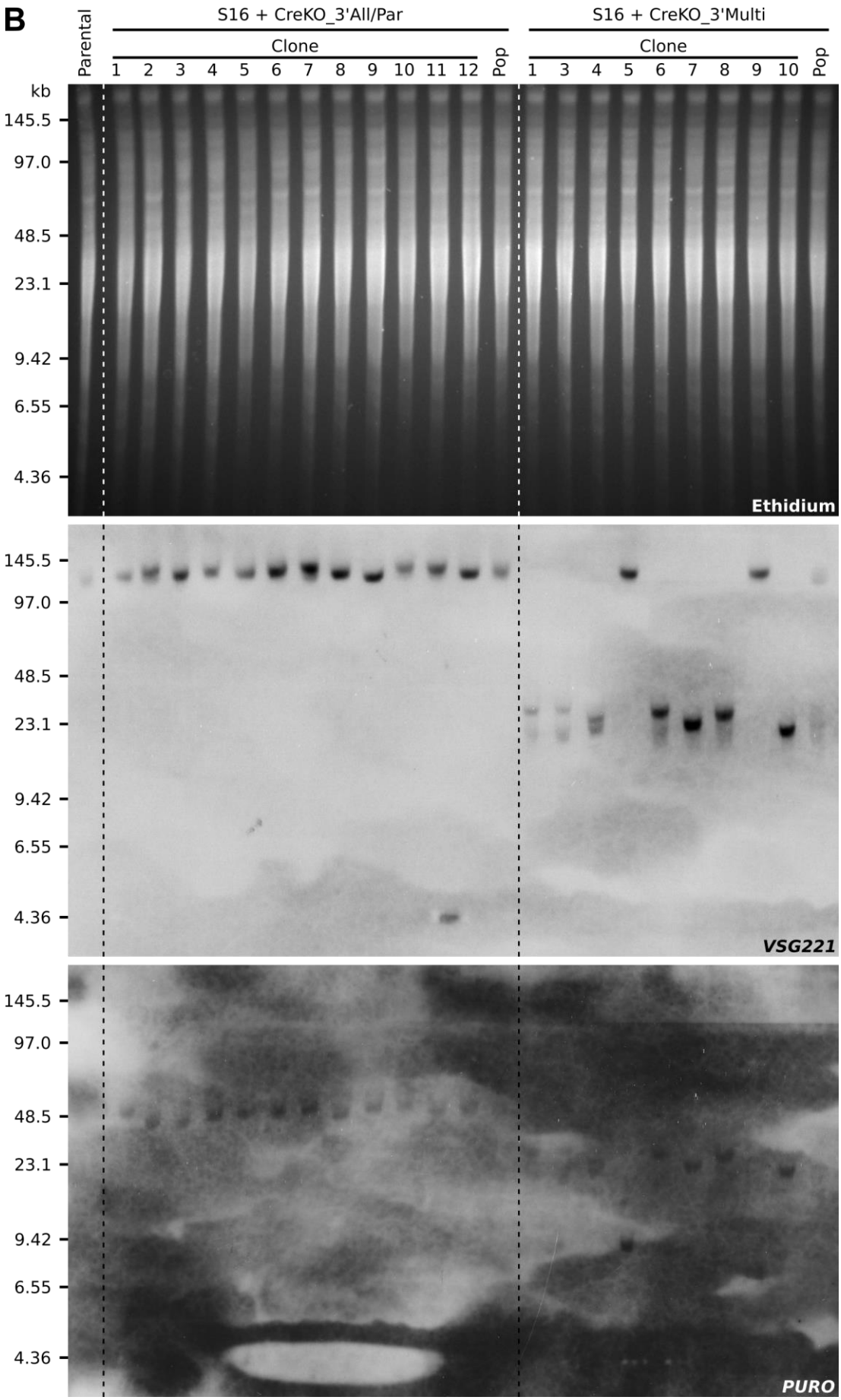

B
Parental
S16 + CreKO_3'All/Par
Clone
1 2 3 4 5 6 7 8 9 10 11 12
Pop
S16 + CreKO_3'Multi
Clone
1 3 4 5 6 7 8 9 10
Pop
kb
145.5
97.0
48.5
23.1
9.42
6.55
4.36
Ethidium
VSG221
PURO

Figure 35. CreKO_3'Multi, but not CreKO_3'All/Par, was integrated into BES1. B, DNA from clones and recovery populations (Pop) transfected with CreKO_3' constructs, and the parental line, were digested with ApaI endonuclease and separated by PFGE. Hybridisation of restriction fragments to *VSG221* or *PURO* probes shows that CreKO_3'All/Par has not integrated into BES1 in any of the clones screened. CreKO_3'Multi appears to have integrated into BES1 in 7 of the 9 clones screened. Ethidium bromide stained agarose gel is provided as a loading control.

exposure used to crosslink the DNA to nylon membrane). Clone 10 was taken forward.

The ability of the CreKO_3'Multi construct to integrate successfully into BES1, but not CreKO_3'All/Par, suggests that the problem arose from the targeting fragments used in the latter. Two new targeting strategies were designed for this construct specifically. One would integrate into the BES1 *VSG* pseudogene (CreKO_3'All/Par-VSGψ), and a second from the *ESAG1* 3' UTR to the *VSG* pseudogene, knocking out the upstream 70-bp repeat region (CreKO_3'All/Par-E1VSGψ), and both utilising the same 3' targeting fragment (insert sites 5b and 5a, respectively, in Figure 36A). As mentioned, disruption of this pseudogene and the upstream 70-bp repeat region is not detrimental to the parasite (Davies *et al.* 1997). Due to the use of the same 3' targeting fragment in these new CreKO_3'All/Par constructs, the same sequence will be produced upon recombination for Complete or Partial ESAG knockout schemes, independent of which version (CreKO_3'All/Par-VSGψ or CreKO_3'All/Par-E1VSGψ) was used (Figure 36B). Integration of CreKO_3'All/Par-VSGψ increases the SwaI-produced BES1 restriction fragment to ~62 kb, while CreKO_3'All/Par-E1VSGψ only increases it to ~60 kb due to removal of the 70-bp repeat region (Figure 37A). These expected shifts in restriction fragment size were observed in every clone screened compared to the parental line when SwaI-digested genomic DNA was probed for *VSG221* and *PURO* (Figure 37B), indicating successful integration into BES1 in each case. This change of targeting sequence has dramatically improved the rate of integration, suggesting that this may have been the issue with the previous strategy (Figure 34B). S16 + CreKO_3'All/Par-VSGψ clone 5 and S16 + CreKO_3'All/Par-E1VSGψ clone 1 were taken forward.

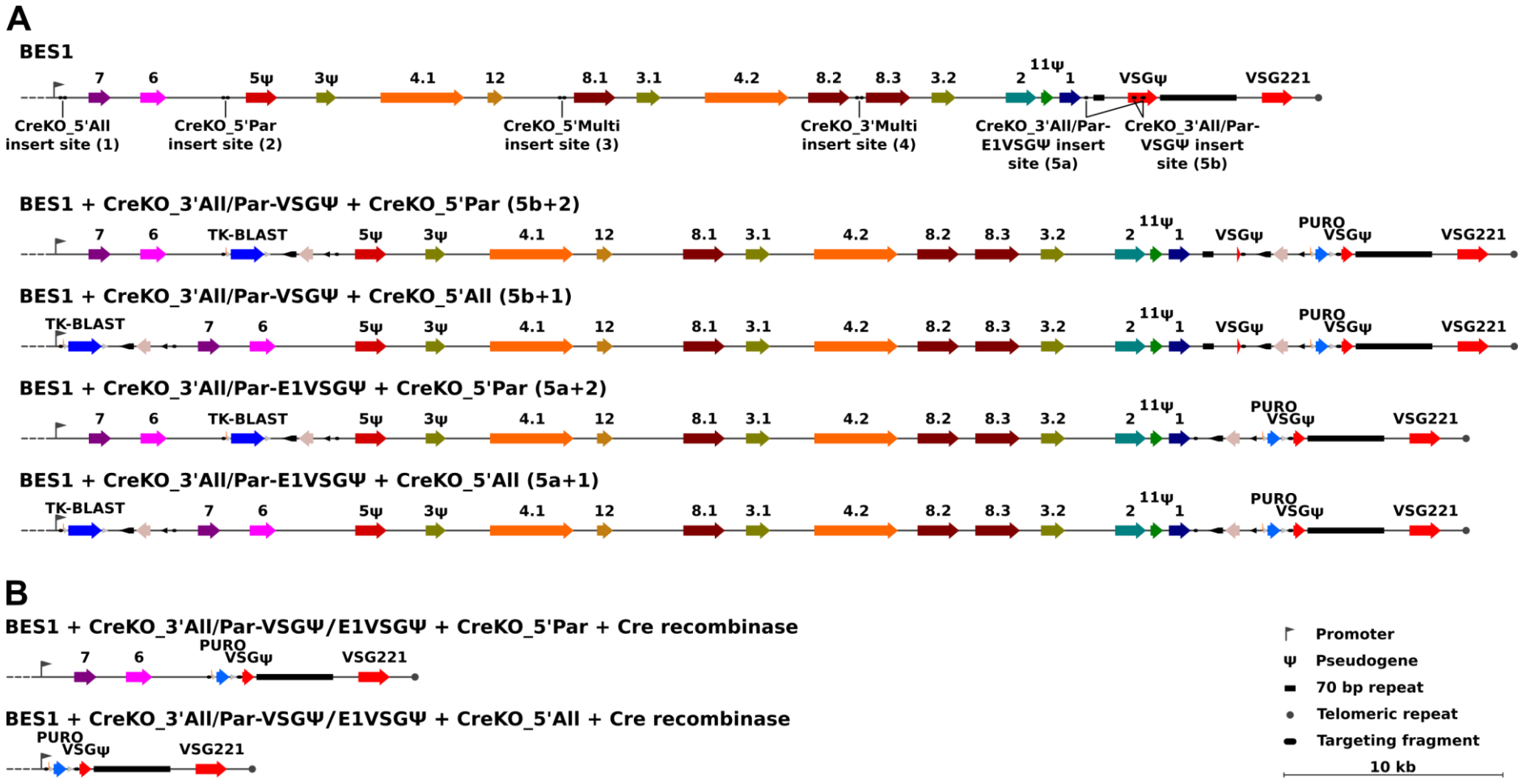


Figure 36. An updated Cre recombinase-based strategy to reduce the *ESAG* complement of BES1. A, Schematic to illustrate the updated Cre recombinase-based strategy to reduce the repertoire of *ESAGs* encoded by BES1. Integration sites of CreKO constructs for each of the three knockout schemes are indicated. Two updated 3' integration sites for Complete and Partial ESAG knockout schemes are shown; the new constructs targeted to these

sites are known as CreKO_3'All/Par-VSGψ (insert site 5A) and CreKO_3'All/Par-E1VSGψ (insert site 5B). Incorporation of these vectors into BES1 is shown. **B**, Resulting BES1 sequence in Complete and Partial ESAG knockout schemes following Cre-mediated excision of the *loxP* flanked sequence shown in part A. Use of the same 3' targeting sequence in both CreKO_3'All/Par-VSGψ and CreKO_3'All/Par-E1VSGψ results in production of the same reduced BES1 sequence in Complete or Partial ESAG knockout schemes, independent of the CreKO_3'All/Par construct version used. Diagrams in part A and B are drawn to scale.

The next step in producing the knockout lines was to introduce CreKO_5' constructs into the successfully integrated CreKO_3' lines. CreKO_5'All and CreKO_5'Par were used to transfect S16 + CreKO_3'All/Par-VSGψ clone 5 and S16 + CreKO_3'All/Par-E1VSGψ clone 1, and CreKO_5'Multi in S16 + CreKO_3'Multi clone 10 (insert sites 1-3 in Figure 36A). Integration of these constructs in S16 + CreKO_3'All/Par-VSGψ and S16 + CreKO_3'All/Par-E1VSGψ lines will produce BES1-containing SwaI restriction fragments of ~67 and ~65 kb, respectively, while integration of CreKO_5'Multi will also produce a fragment of ~67 kb (Figure 38A). Transfectants were again screened by Southern blot for integration into BES1 (Figure 38B). The majority of clones investigated by *VSG221* and *BLAST* probes showed integration of CreKO_5' construct into BES1 (Figure 38B). Only S16 + CreKO_3'All/Par-VSGψ + CreKO_5'All clone 3, and S16 + CreKO_3'All/Par-E1VSGψ + CreKO_5'Par clones 2 and 3 failed to show the expected shifts in restriction fragment size outlined in Figure 38A. The presence of *PURO* was also investigated to ensure that the CreKO_3' constructs remained in BES1, which was the case in every cell line screened. Clones with both CreKO constructs correctly integrated now have sections of BES1 flanked by *loxP* sequence, ready to be excised by Cre recombinase. Clone 1 of each of the dual CreKO_3' and CreKO_5' integrated lines was taken forward.

5.4. CreKO lines do not have growth defects

Prior to inducing Cre expression in dual integrated CreKO lines, their growth was monitored for five days in culture to investigate if these modifications to the active BES were tolerated by the parasite (Figure 39). All of the lines tested exhibited similar growth rates to parental S16 cells. In

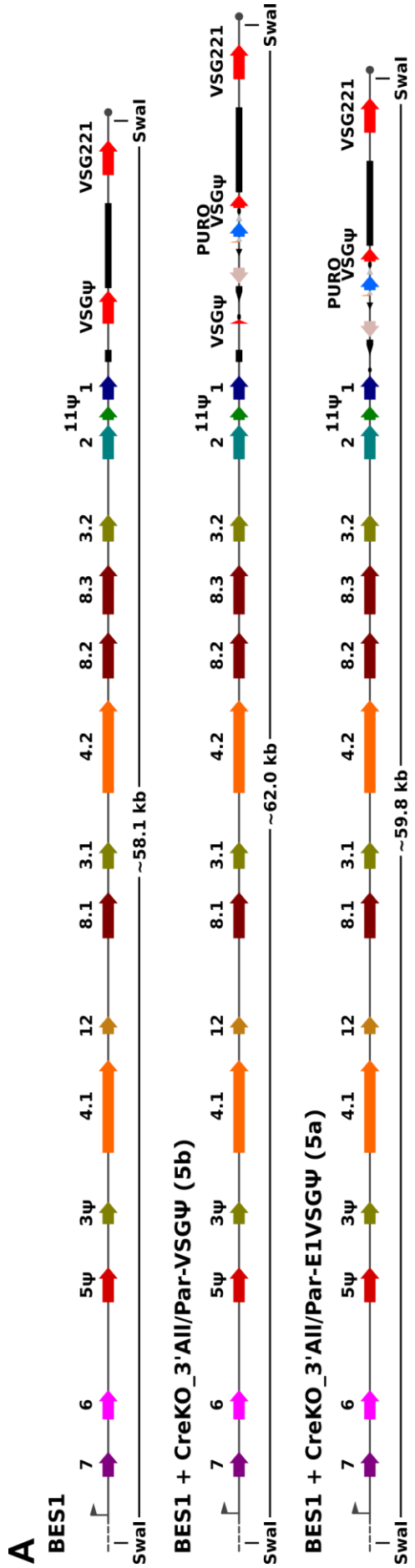


Figure 37. Both CreKO_3'All/Par-VSGψ and CreKO_3'All/Par-E1VSGψ can be successfully integrated into BES1. A, Schematic outlining the size of expected BES1-containing restriction fragments when genomic DNA is digested with SwaI endonuclease. Integration of CreKO_3'All/Par constructs into BES1 increases the expected size of the fragment. Insert sites used in labels relate to Figure 36. Diagram drawn to scale.

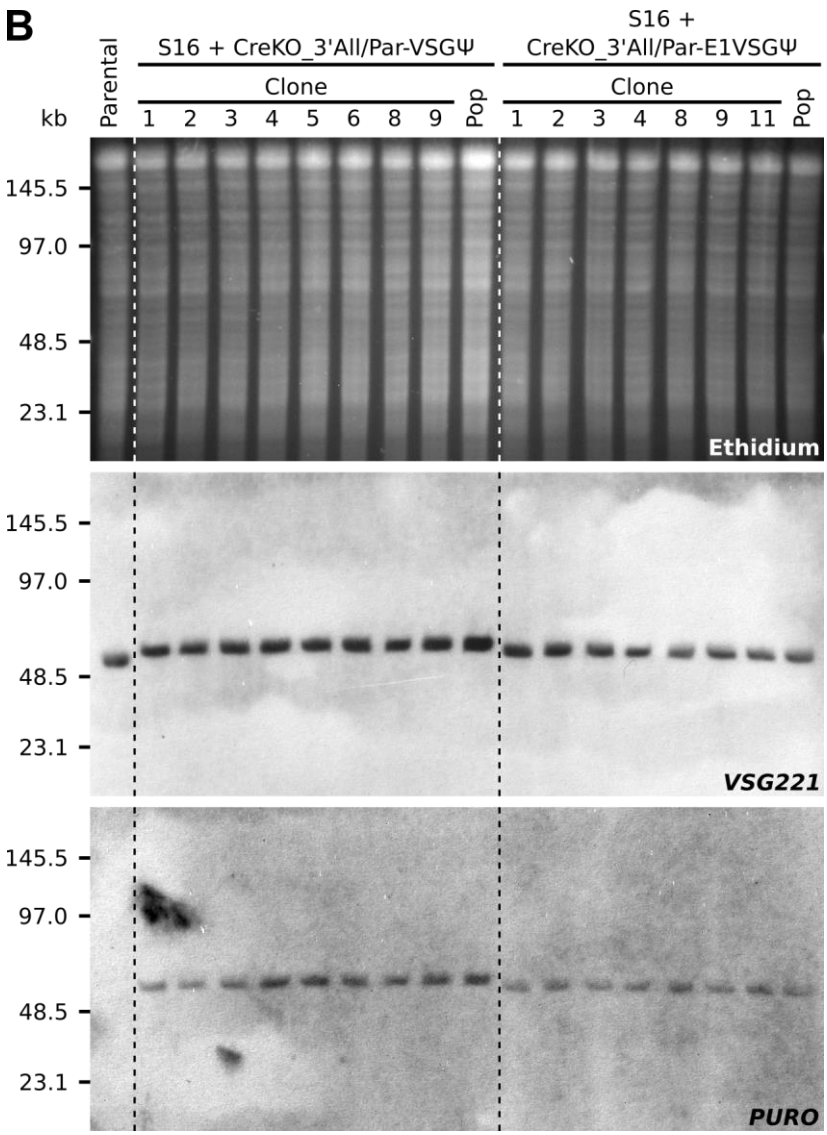


Figure 37. Both CreKO_3'All/Par-VSGψ and CreKO_3'All/Par-E1VSGψ can be successfully integrated into BES1. B, DNA from clones and recovery populations (Pop) transfected with CreKO_3'All/Par constructs, and the parental line, were digested with SwaI endonuclease and separated by PFGE. Hybridisation of restriction fragments to *VSG221* or *PURO* probes shows that CreKO_3'All/Par constructs have integrated into BES1 in all clones screened. Ethidium bromide stained agarose gel is provided as a loading control.

order to enable Cre-mediated excision of the *loxP* flanked sections of BES1 in these lines, they first needed to be transfected with an inducible Cre expression vector (pLEW100-Cre-EP1, Addgene plasmid 33345, Scahill *et al.* 2008). This construct integrates into the rDNA locus and enables tetracycline inducible expression of Cre recombinase from a T7 promoter. The strategy for this experiment is outlined in Figure 40A. Dual CreKO integration lines were transiently transfected with circular pLEW100-Cre-EP1, and recovered in medium containing tetracycline to induce Cre recombinase

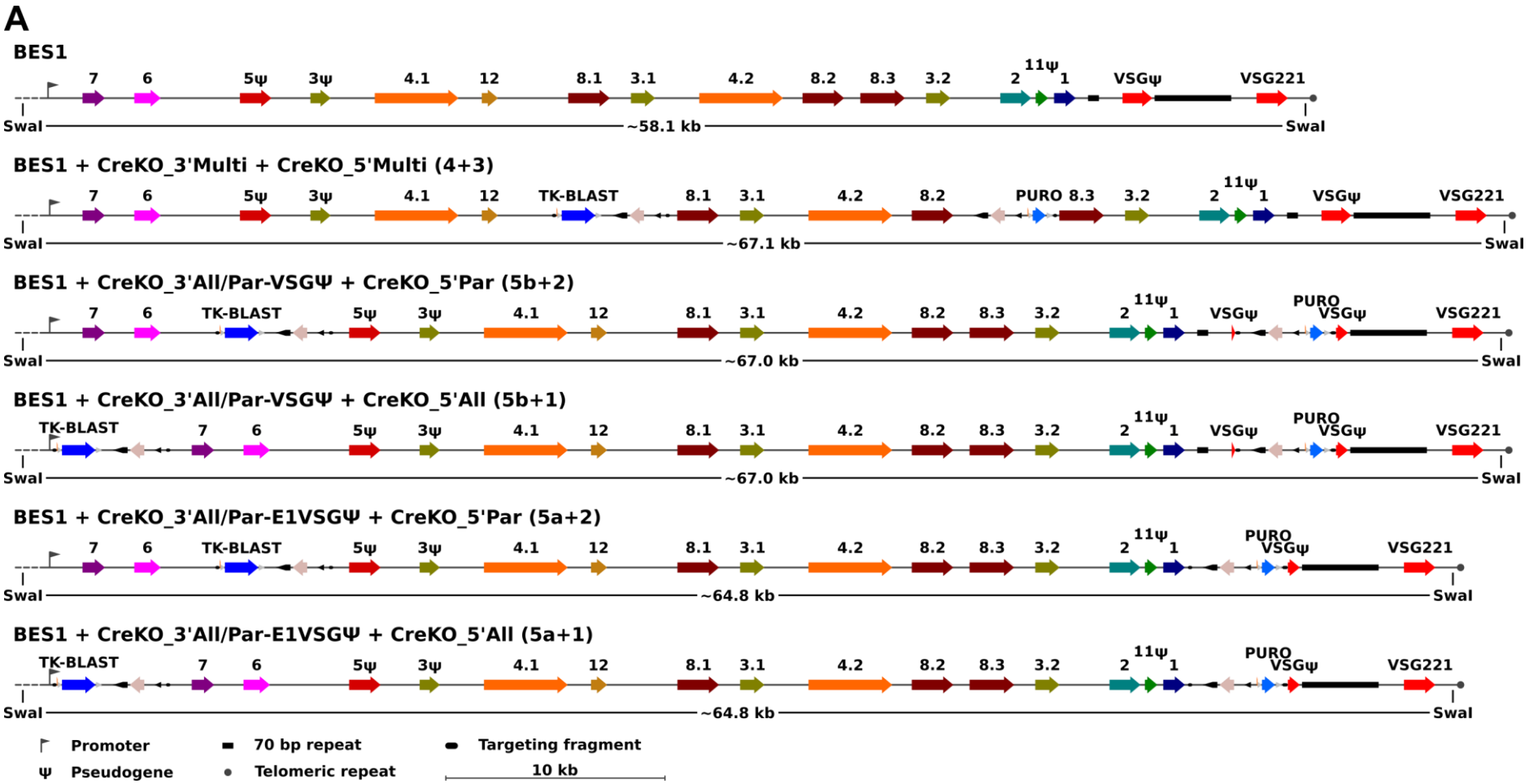


Figure 38. Integration of CreKO_5' constructs into cell lines previously transfected with corresponding CreKO_3' construct. A, Schematic outlining the size of expected BES1-containing restriction fragments when genomic DNA is digested with SwaI endonuclease. Integration of CreKO_5' constructs into BES1 encoding CreKO_3' sequence increases the expected size of the fragment. Insert sites used in labels relate to Figure 36. Diagram drawn to scale.

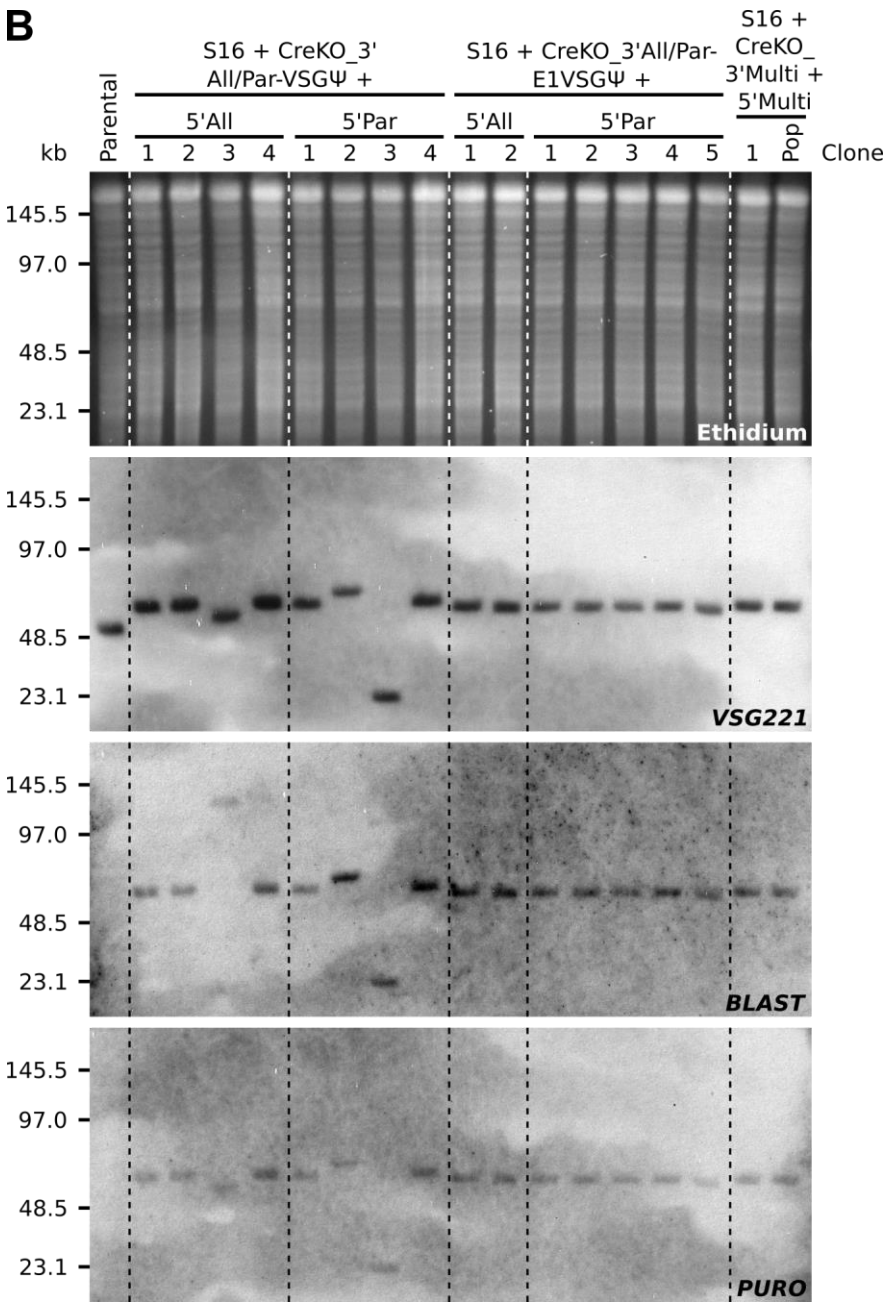


Figure 38. Integration of CreKO_5' constructs into cell lines previously transfected with corresponding CreKO_3' construct. B, DNA from clones and recovery populations (Pop) transfected with CreKO_5' constructs, and the parental S16 line, were digested with SwaI endonuclease and separated by PFGE. Hybridisation of restriction fragments to *VSG221*, *BLAST* or *PURO* probes is shown. Ethidium bromide stained agarose gel is provided as a loading control.

expression. Cells were incubated for eight hours to allow time for accumulation of Cre, and excision of the BES1 fragments. Cells were then washed to remove tetracycline and incubated for a further 16 hours prior to addition of ganciclovir. This incubation period was included to provide time for turnover of TK-BLAST in cells where recombination had taken place,

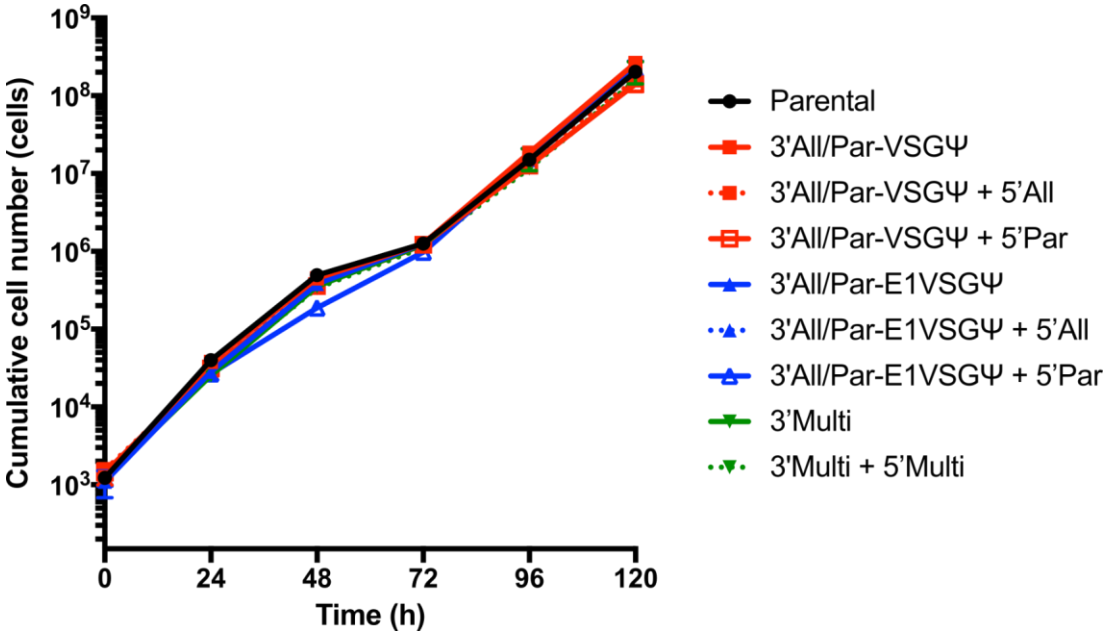


Figure 39. Lines transfected with CreKO constructs do not exhibit growth defects. Growth of parental S16, CreKO_3' integrated, and dual CreKO_3' and CreKO_5' integrated cell lines monitored over five days in culture. All lines grow at a rate similar to their parental. Cultures were diluted at 72 hours to maintain log phase growth.

removing the negative selection marker from these cells. Therefore, addition of ganciclovir would only kill cells where Cre-mediated excision of the TK-BLAST encoding fragment had not occurred. Puromycin pressure was maintained throughout to select for BES1-active cells. Ganciclovir was used at 100 µg/ml, as this concentration has previously been used with the TK-BLAST fusion protein (D'Archivio and Wickstead 2017). Susceptibility of dual CreKO integration cell lines to ganciclovir was monitored to ensure this negative selection marker was functional. Within 72 hours of addition, cell numbers were reduced below the limit of linearity of the luciferase-based cell count assay used here (Figure 40B). Ganciclovir addition to lines transfected with pLEW100-Cre-EP1, and treated as outlined above, also resulted in population decline. However, these cells did not recover as expected, with no successful reduced-BES1 survivors growing out from this pressure. Further work is required to understand why Cre-mediated recombination did not take place in these cells, so this experiment can be repeated and produce lines with reduced BES1 *ESAG* repertoires (discussed in Section 5.5).

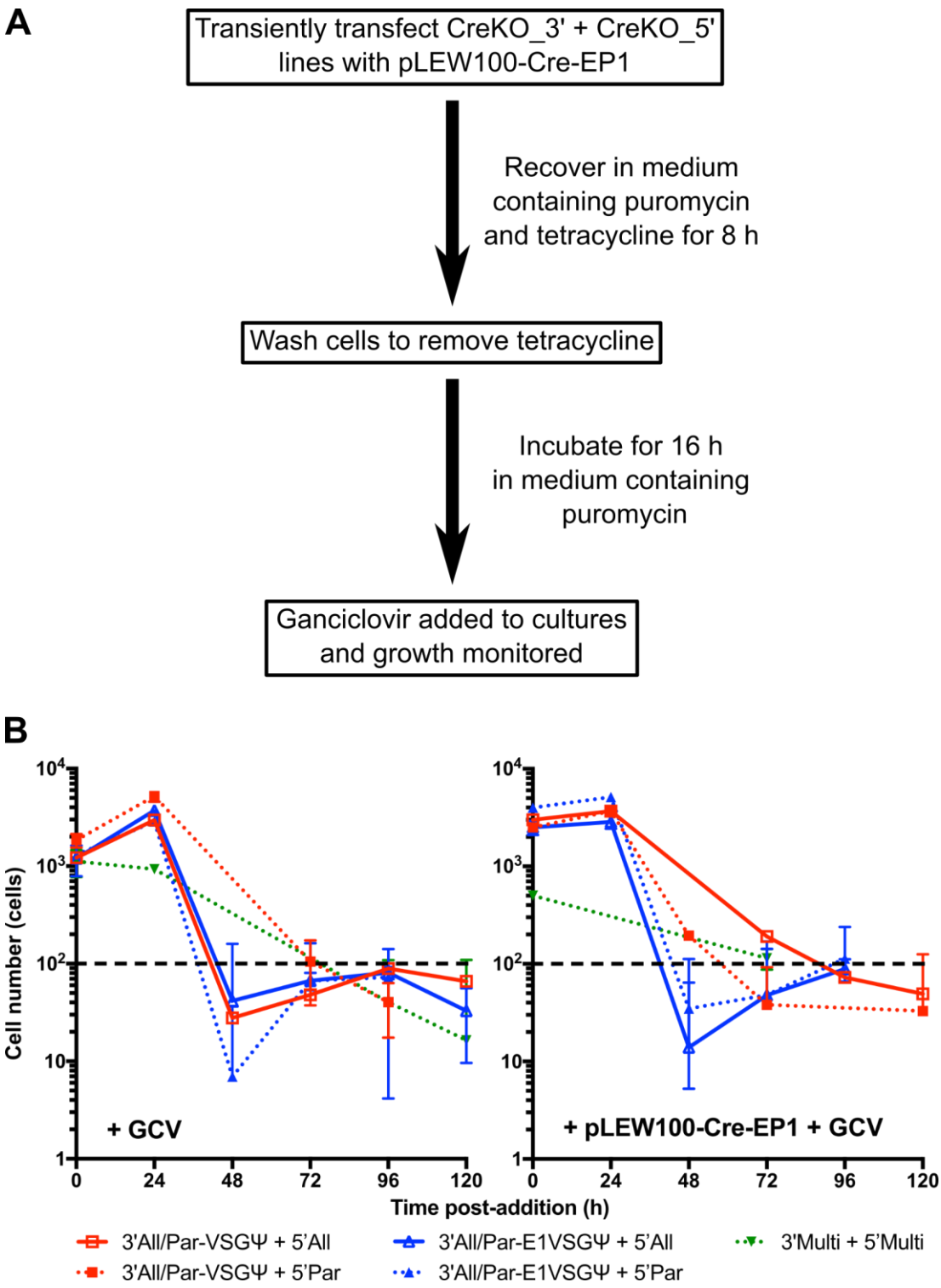


Figure 40. Expression of Cre recombinase in dual CreKO integrated lines does not make cells resistant to ganciclovir. A, Experimental design to obtain cells where *loxP* flanked BES1 sequence had been excised by Cre recombinase in dual CreKO integrated lines. **B**, Growth of dual CreKO_3' and CreKO_5' integrated lines monitored over five days in the presence of 100 µg/ml ganciclovir (GCV), with or without transient transfection and induction of Cre recombinase expression vector (pLEW100-Cre-EP1, as outlined in part A). All lines are susceptible to ganciclovir and die within 4 days of its addition. The limit of linearity is indicated by the black dashed line at 100 cells.

5.5. Summary

Despite employing two different techniques, and multiple strategies and modifications, cells with a reduced BES1 *ESAG* repertoire could not be obtained in this study. Lines containing CreKO_5' and CreKO_3' constructs integrated into BES1 were successfully produced, did not display impaired growth, and were sensitive to negative selection. However, recombination between *loxP* sites could not be induced upon introduction of Cre expression vectors into these lines. A possible explanation for the failure of both homologous-recombination and Cre recombinase-based methods for knocking out sections of BES1 is that it is lethal to the cell. If this is due to the *ESAG* complement of the BES, one would expect the KO Multi strategy to succeed given that BES3 and BES7, shown to be active *in vitro* in Chapter 3, do not encode multiple copies of *ESAG3*, *ESAG4* and *ESAG8*. Hertz-Fowler *et al.* were able to activate BESs with a reduced *ESAG* repertoire *in vitro*. In fact, one of these, BES12, lacks putatively functional *ESAG3*, *ESAG4* and *ESAG8* coding sequence completely (Hertz-Fowler *et al.* 2008). Alternatively, modification of the active BES in this way may induce cells to switch to a different BES; in which case, the reduced transcription of a selection marker from BES1 may result in the death of switched cells. However, deletion of sections of BES1 in BES3 active cells should be possible if this were the case, but this was also unachievable though homologous recombination-based methods.

There are a number of potential reasons why induction of Cre recombinase expression in dual integrated CreKO lines failed to result in out-growers that had excised *loxP* flanked sequence: for example, 1) puromycin pressure was maintained throughout the experiment, so if recombination between *loxP* sites resulted in disruption of *PURO* transcription, cells may have become sensitive to this drug (i.e. successful recombinants would be killed). 2) Levels of TK-BLAST protein and transcript may not have sufficiently depleted after excision of the BES1 fragment for cells to fully revert to ganciclovir resistance. 3) The genomic location of this marker may also play a role. Previous use of HSV-TK for this purpose has relied on its transcription from chromosomal-internal loci (Scahill *et al.* 2008). If greater

amounts of transcript are produced from the BES, then encoded protein may be present in the cell for a longer period, maintaining sensitivity to ganciclovir.

If I were to repeat this experiment, a number of factors would be changed. The inducible Cre expression vector was transiently transfected to avoid use of an additional selection marker, and potential toxicity observed in PCF (Barrett *et al.* 2004). However, this does not seem to represent as severe an issue in BSF (Scahill *et al.* 2008). Thus, I would stably transfect pLEW100-Cre-EP1 into dual integrated CreKO lines to allow for further troubleshooting. This would include induction of Cre expression in the absence of puromycin, and titrating the period of time left post-induction before the addition of ganciclovir. Puromycin pressure could then be applied once ganciclovir resistant clones have grown out. This would have the added benefit of identifying if cells where recombination has taken place have also switched BES. Should these modifications yield lines with successfully recombined BES1 (monitored by Southern blot), the experiment would be repeated with transient pLEW100-Cre-EP1 transfection and an updated protocol. Once these lines are obtained, investigating the phenotype associated with reducing the BES1 repertoire will aid in understanding their essentiality of these genes *in vitro*, and if their removal from the active BES induces any adaptive response by the parasite, such as switching the active expression site.

Chapter 6. Discussion

Despite the presence of *(GR)ESAG* family members at important genomic loci – the VSG expression sites – little is known about the function of the majority of these families. The study of multi-gene families is challenging due to their high DNA sequence similarity, which can cause issues when attempting to target both homologous recombination and RNA interference to specific genes, and also due to the potential for functional redundancy between family members. Here I present a method to specifically target an individual family member for RNAi-mediated ablation, through tagging at the endogenous locus with GFP, then using *GFP* ORF sequence as template for dsRNA production.

To obtain a complete set of GFP-tagged, active BES-encoded ESAGs, a BES3 active cell line was required to tag ESAG5 and ESAG10. Switched cells were selected for from a population of BES1-active cells using anti-serum raised against the BES1-resident VSG – VSG221. We expected to obtain cells that had undergone an *in situ* switch to BES3, as this is the preferred form of switching *in vitro* (Aitcheson *et al.* 2005), and BES3 is a commonly active *in vitro* BES. Instead, VSG221-specific serum pressure selected for cells that had undergone gene duplication of the BES3 *VSG121* into BES1, evidenced by the loss of *VSG221*. As a substitute, a BES3-active wild type cell line was made RNAi competent, and used to produce GFP-tagged ESAG5 and ESAG10 cell lines. BES3 in this line appears to lack the upstream expression site section that normally encodes ESAG10, and by tagging this gene with a resistance marker on a different BES, a BES7-active line was acquired. Cells derived from this wild type line consistently grew at a slower rate compared to the S16 line, which was employed to tag other ESAGs. This feature may be inherent to the wild type line, or result from transcripts derived from the active BESs (BES3 or BES7). Loss of the BES3 upstream region has been reported previously, but the biological significance of this rearrangement is unknown (Gottesdiener *et al.* 1991, 1992; Navarro and Cross 1996).

Pan-family RNAi constructs, capable of targeting knockdown of multiple genes, were produced by taking advantage of the sequence similarity between (GR)ESAG family members. Bioinformatic tools were used to identify previously unknown *GRESAG5* genes, and three were found present on *T. b. brucei* chromosome 11. RNAi constructs housing multiple targeting fragments were produced to target all members of the divergent *(GR)ESAG5* and *(GR)ESAG10* families. These, along with the *(GR)ESAG6/7* and *(GR)ESAG12* pan-family RNAi constructs, target all known members of their respective families. This combinatorial RNAi strategy could be applied elsewhere to target knockdown of two or more unrelated proteins, which may be useful when investigating multi-component pathways.

Both *GFP* and pan-family RNAi constructs were able to ablate chimeric BES-encoded protein, but only pan-family constructs were capable of inducing growth defects in parasites. These results show that (GR)ESAG families as a whole, and not necessarily BES-linked family members, are important for parasite survival *in vitro*. The lack of growth phenotype produced upon knockdown of the actively expressed ESAG indicates that there may be functional redundancy between BES-linked and non-BES encoded family members. This result is not surprising given the high similarity at the DNA and protein level between these two (GR)ESAG family arms. In addition, BESs can be activated *in vitro* that do not encode representatives of all families (Hertz-Fowler *et al.* 2008), and some *ESAGs* can be removed from the active BES without detriment to the cell (Carruthers *et al.* 1996; Hoek and Cross 2001). Finally, at least for one family, GRESAGs have been shown to functionally complement loss of the BES-linked gene (Salmon *et al.* 2012a). The RNAi cell lines produced here did not induce PAD1 expression, as has been described in similar lines elsewhere, but levels of TfR were affected. The mechanism behind this phenotype is unknown, but occurred in both *GFP* and pan-family RNAi lines, suggesting that it was not linked to growth defects induced in the latter.

The apparently dispensable nature of ESAGs led me to attempt to reduce the repertoire encoded by the active BES. Strategies to remove sections of BES by homologous recombination were unsuccessful. Cell lines were

produced with *loxP* sites and selectable markers introduced into the active BES, and Cre-mediated recombination between these sites should result in three versions of a reduced BES, with sequentially smaller *ESAG* repertoires. Unfortunately, this was not completed due to time constraints, but should be achievable with troubleshooting, and answer an important question: what is the minimal BES *ESAG* complement? Given that specific results and experimental limitations were discussed at the end of each results chapter, more broad topics and future avenues of investigation will be explored here.

6.1. RNAi as a technique to ablate (GR)ESAG families

Pan-family RNAi constructs have been used here, and elsewhere (Salmon *et al.* 2012a; Batram *et al.* 2014; Tiengwe *et al.* 2016), that are capable of targeting multiple (GR)ESAG family members at once. This relies on conserved sequence identity of around 22 consecutive base pairs (Bernstein *et al.* 2001). However, sites of Dicer-mediated cleavage of dsRNA are unknown, so siRNA containing these specific 22 bp stretches may not be produced, and thus may not target degradation of related transcripts. Measuring the degree of pan-family knockdown, based on ablation of GFP-tagged ESAG, serves as a surrogate, and does not inform if other family members are effectively targeted. This limitation could be remedied through the use of a number of techniques, including northern blot, quantitative reverse transcription PCR (RT-qPCR) or RNA sequencing (RNA-seq) to investigate mRNA levels of *(GR)ESAG* family members. These experiments would confirm if specific family members were targeted by pan-family RNAi, however, as gene expression is post-transcriptionally controlled in trypanosomes (Clayton and Shapira 2007), transcript and protein levels are not strictly relative, so these techniques would still only provide an indirect read-out.

As well as confirming RNAi-mediated degradation of specific *GRESAG* transcripts, RNA-seq represents a powerful tool for the study of these RNAi lines if reads are correctly mapped, taking account of intra-family sequence identity. It would show if any *GRESAGs* are up-regulated to compensate for

the loss of BES-derived transcript in *GFP* RNAi lines, as has been shown when *ESAG4* is knocked out of the active BES (Salmon *et al.* 2012a). It may also shed light on the function of these families, by identifying associated proteins or pathways, whose transcript may fluctuate when RNAi is induced.

Phenotypes produced by pan-*(GR)ESAG* family RNAi constructs, described elsewhere, are more severe than those presented in this study. Use of stem-loop dsRNA, as opposed to the head-to-head promoter employed here to produce complementary ssRNA, may improve RNAi penetrance. This has been shown to be the case in yeast, where stronger RNAi phenotypes were induced by dsRNA hairpins, compared to dsRNA derived from head-to-head promoters, as the former produced more siRNA (Drinnenberg *et al.* 2009). This feature may also be shared by *T. brucei*, where hairpin dsRNA-mediated knockdown of two kinesins induced significantly greater growth defects than dsRNA from head-to-head promoters using identical targeting sequence (Stretton and Wickstead, personal communication). However, increased levels of siRNA have been shown to result in off-target effects (Caffrey *et al.* 2011), and thus may be the source of greater phenotypes presented in these examples. The reduced siRNA produced by head-to-head promoters may arise due to the need for complementary single RNA strands to encounter in the nucleus, while complementary bases encoded on the single strand used to produce hairpins are fundamentally spatially ‘close’.

6.2. CRISPR/Cas9 to study (GR)ESAG families

Genome engineering using CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and Cas9 (CRISPR-associated protein-9) was in its relative infancy at the beginning of this project. This transformative technology, based on the bacterial adaptive immune system (to protect from viruses and plasmids), has been adapted for use in numerous eukaryotic organisms, and represents a powerful new genome-editing tool (reviewed by Doudna and Charpentier 2014; Salsman and Dellaire 2016). Briefly, Cas9 in complex with ‘single guide’ RNA (sgRNA), targets the endonuclease activity

of this CRISPR-associated protein to complementary DNA sequence, where it introduces a site-specific double-strand break (DSB). This stimulates DNA repair by non-homologous end-joining (NHEJ), which often results in insertions and deletions (indels), or homologous recombination. The latter can make use of exogenous DNA sequence as a template to initiate homology-directed repair, and thus has applications in gene knockout, tagging, and site-directed mutagenesis, to name a few. Error-prone repair by NHEJ can be used to silence genes by introducing mutations that result in production of non-functional protein, and even deletions or chromosome translocations when DSBs are targeted to multiple genomic locations.

This system has recently been introduced into apicomplexa and trypanosomatids (reviewed by Lander *et al.* 2016a). CRISPR-Cas9-mediated genome engineering has been reported in *Leishmania* spp. (Sollelis *et al.* 2015; Zhang and Matlashewski 2015; Zhang *et al.* 2017), *T. cruzi* (Peng *et al.* 2014; Lander *et al.* 2015, 2016b, 2017), and more recently *T. brucei* (Beneke *et al.* 2017). Repair of DSBs in these parasites is dominated by microhomology-mediated end joining (MMEJ), as opposed to NHEJ, and results in deletion of sequence between regions of microhomology that surround the DSB (Glover *et al.* 2008, 2011). Significantly, knockdown of the β-galactofuranosyl glycosyltransferase family, consisting of 65 genes, was achieved through multiplexing of three sgRNA to target DSBs to all family members in *T. cruzi* (Peng *et al.* 2014). Therefore, CRISPR/Cas9 represents an alternative tool to RNAi for the study of the (GR)ESAG families in *T. brucei*, or in fact any multigene family. Importantly, it will allow study of contingency gene families in organisms that lack the capacity for RNAi, such as *T. cruzi*, most *Leishmania* spp. and *Plasmodium* spp. (Kolev *et al.* 2011).

CRISPR/Cas9 also represents an additional mechanism to produce reduced-*ESAG*-repertoire BESs. By targeting DSBs to two intra-chromosomal loci using two sgRNA, large-scale deletions (from a few kb to Mb) have been achieved in mammalian genomes, relying either on NHEJ or homology-directed repair (Salsman and Dellaire 2017). This scale of genome editing has not yet been attempted in *T. brucei* with this system, but targeting

of DSBs, as opposed to Cre-mediated recombination, to the active BES, may facilitate its modification with greater efficiency.

6.3. The roles of (GR)ESAG families *in vivo*

ESAGs are hypothesised to have roles in adapting the parasite to different mammalian bloodstreams, based mainly on variable TfR affinities for different transferrin molecules, and SRA-mediated resistance to human serum. Thus, investigating the behaviour of *GFP* RNAi lines *in vivo* may reveal the importance of ESAGs during infection, as their ligands may be more scarce than when grown in culture, or functions specific to this complex environment. Given the redundancy between ESAGs and GRESAGs highlighted by this study, and the adaptable nature of these families, evidenced by up-regulation and rearrangement of family members in response to disruption of the actively expressed BES-linked gene (Gerrits *et al.* 2002; Mußmann *et al.* 2004; Van Luenen *et al.* 2005; Salmon *et al.* 2012a), it is unlikely that severe defects would be produced by ablation of chimeric protein in these lines. However, more subtle defects, or a failure to thrive, may occur if the specific ESAG did provide an advantage to parasite growth. This phenotype could be revealed *in vivo* through dual inoculation with RNAi and parental lines, and monitoring the relative proportions of each of these populations through the course of infection, which could be achieved via qPCR or next-generation sequencing techniques. For this to work, the BES encoding a GFP-tagged ESAG would need to be selected for by the presence of hygromycin, to prevent switching and therefore transcription of functional *ESAG* from a different BES. The DR4 mouse strain constitutively expresses HYG, and is resistant to this drug pressure, so could be used in these experiments (Uzureau *et al.* 2007). In the absence of selection pressure, an increased rate of switching away from this BES in induced *GFP* RNAi cells could also indicate that there is a benefit to activating a BES, from which this *ESAG* transcript is not degraded.

If ESAGs do adapt the parasite to different mammalian hosts, and specific variants have higher affinity for host ligand, or differential activity in different

hosts, the high rate of recombination that occurs at BESs suggests that 'host-specific' BESs (encoding an optimised set of ESAGs that provide an advantage in a certain host) would not be maintained over extended time periods. It is tempting to propose a mechanism whereby *in situ* switching occurring early in an infection, provides an advantage to parasites that activate a BES that is better suited to the current host environment, potentially allowing these cells to outcompete those that have not activated this specific BES. Then during infection, recombination of BES sequence may enable accumulation of host-specific *ESAGs*, and formation of a host-specialised BES. Expression of specific ESAG6, ESAG7 and ESAG3 subsets in *T. b. gambiense* grown in pig or human serum provides some evidence for this potential mechanism, as these cells also grew at a faster rate and took up more transferrin than cells grown in calf or goat serum, which expressed a more diverse range of these proteins (Cordon-Obras *et al.* 2013).

To this end, investigation of BES evolution through infection could prove interesting, and generate evidence for this adaptation mechanism. This could be achieved through isolating genomic DNA from parasites at certain points during infection, sequencing the VSGs expressed by the population (through targeted *VSG* RNA sequencing, or VSG-seq, described by Mugnier *et al.* 2015), and performing TAR cloning to isolate BESs from the population. TAR clones encoding BESs which were active during infection could then be identified through the ability to amplify *VSGs* shown to be transcribed at that time point. Sequencing of the 'infection-active' BESs would identify if, and to what extent, recombination occurs at these loci during infection, and if *ESAG* repertoires change. Equally, if recombination is so high that specific BESs cannot be tracked throughout the course of infection, it suggests that *ESAG* repertoires are highly dynamic and, if necessary, the cell can adapt to their presence in, or absence from, the active BES.

6.4. Producing an active BES with reduced *ESAG* repertoire

Removal of specific sets of *ESAGs* from the active BES during this study has proved challenging. I have produced cell lines that, upon expression of

Cre recombinase, should contain an active BES with reduced *ESAG* repertoire. If this were made possible by the troubleshooting methods outlined in Section 5.5, for all three reduction schemes, then this would suggest that there is no minimum *ESAG* repertoire for parasite survival *in vitro*. With all *ESAGs* removed, BES1 would have an architecture more similar to a MES, that completely lack *ESAGs*, and are known to be active in the quiescent metacyclic stage. It is likely that the active BES must encode a functional TfR, suggesting that the removal of all but *ESAG6* and *ESAG7* should be possible, and that these genes represent the minimal repertoire for survival *in vitro*. Investigation of reduced-BES1 lines by RNA-seq, as described above, would identify if up-regulation of GRESAG family members has occurred to compensate for the loss of BES-derived transcripts, and may in turn shed light on why the parasite maintains these specific genes at a privileged genomic location. Successful recombination of BES1 in these cell lines will represent the first instance of Cre recombinase-mediated modification of a BES, and the largest genomic modification using this system in *T. brucei*.

If Cre-mediated removal of BES sequence is not possible, it suggests that either the disrupted BES-derived transcripts are essential, or the scale or site of modification is not conducive to cell survival. Modification of BESs in the PCF stage should not affect the cell, as these regions are all transcriptionally silent. If recombined BES1 could not be obtained in BSF cells, differentiation to the insect lifecycle stage could be achieved through incubation of BSF cells at 27°C in PCF media containing *cis*-aconitate in excess of 3 mM (Czichos *et al.* 1986). Cre recombinase expression vector could then be transfected (stably or transiently) into these differentiated cells, and recombination induced, with its success monitored by diagnostic PCR. Recent literature outlines a procedure to complete the parasite differentiation of PCF to BSF *in vitro* through transient expression of RNA binding protein 10 (RBP10; Mugo *et al.* 2017). This could be employed to convert these PCF cells with reduced BES1, back into BSF, followed by addition of puromycin to select for cells with an active BES1. If this BES can be activated in these lines, it implies that Cre-mediated modification of the active BES in BSF cells

is not viable. But if BES1 cannot be activated, then it could suggest that these various sets of *ESAGs* are essential for cell survival.

The *T. b. rhodesiense* RES represents the smallest-*ESAG*-repertoire containing expression site known to be active *in vivo*, with only *TfR*, *ESAG5* and *SRA* protein coding genes transcribed along with *VSG* (Xong *et al.* 1998). Investigation of cells with reduced BES1 *in vivo* may reveal that the minimum *ESAG* repertoire is different between life in culture and the mammalian host. SRA is only required for survival in human serum, and BES1 does not encode functional ESAG5 and is known to be active *in vitro* and *in vivo*, suggesting again that the TfR components may be the only essential non-*VSG* transcripts arising from the active BES. Loss of *ESAG* sequences from the active BES may impair the ability for cells to adapt fully to the host environment. Thus, monitoring the capacity of reduced-BES1 cells to establish infection, parasitaemia levels, and chronicity will reveal if these genes are important for parasite virulence. The *in vivo* dynamics of these cells could be investigated with a competition assay (described above) to see if a reduced BES impairs the ability of the parasite to thrive in this environment. Also, monitoring if the parasite attempts to repair the active BES by replacing lost *ESAG* sequences could provide evidence that these genes do provide an advantage to cells during infection, and could identify how BESs that have undergone drastic recombination can be made functionally competent.

Subtelomeric BESs are fragile, with DSBs occurring naturally at a higher rate than chromosomal-internal loci, a feature with central importance to antigenic variation (Boothroyd *et al.* 2009; Glover *et al.* 2013a). DSB formation in the 70-bp repeat region induces *VSG* switching via gene duplication, and thus recombination of donor sequence from the silent catalogue into the active BES (Hovel-Miner *et al.* 2016). Telomeric exchange is thought to occur at the 50-bp repeat region upstream of BESs (Pays *et al.* 1985; Navarro and Cross 1996), but also occurs within BES sequence (Rudenko *et al.* 1996), although precise sites of recombination are unknown. Removal of *ESAGs* from the active BES may therefore impact switching frequency of cells via telomere exchange, as sequence homologous to other

BESs is no longer present, and can therefore not act as donor template to repair DSBs that occur in this region. This hypothesis could be tested by observing the switch frequency, and mechanism, between parental and reduced-BES cell lines. This could be achieved by the selection of VSG switchers through incubation of cell populations in anti-VSG221 anti-serum as described in Chapter 3, followed by cloning of survivors by limiting dilution. The type of switch could then be diagnosed. Cells that have undergone gene conversion will no longer encode *VSG221* (which can be assessed by PCR, also described in Chapter 3). Transcriptional switching, or telomere exchange upstream of the selection marker (*PURO* in this case), will revert cells to puromycin sensitivity, while telomere exchange downstream of *PURO* will remain resistant (but still encode *VSG221*). Southern blotting could be used to diagnose between *in situ* switching and telomere exchange occurring upstream of *PURO*, as the genomic location of this gene will have changed in the latter. However, switching in these monomorphic cell lines is infrequent, so this experiment may have to be repeated in pleomorphic lines, or DSBs experimentally induced in regions of active native and reduced BESs using the yeast mitochondrial endonuclease I-SceI, which has previously been used to introduce DNA lesions into *T. brucei* BESs (Glover *et al.* 2007, 2013a; Boothroyd *et al.* 2009; Hovel-Miner *et al.* 2016).

6.5. Virulence factors or contingency genes?

(GR)ESAG family members share genomic loci with the major *T. brucei* contingency gene family, *VSG*, and undergo variant switching, which may induce a change in phenotype. However, they are not themselves typical contingency genes, as they do not appear to spontaneously switch at a set frequency, and expression is instead dependent on lifecycle stage and genomic location. As molecules that uptake nutrients from the host (ESAG6/7, and probably ESAG10), and modulate the host immune system (ESAG4 and SRA), these represent parasite virulence factors. Given that other less-defined families are also present at the host-parasite interface, it is tempting to speculate that they have similar roles, but further investigation is required to confirm this. Investigation of two *T. b. rhodesiense* clones that

differ dramatically in virulence phenotype showed that the same BES was active in both lines, suggesting that BES derived-transcript did not affect the clonal virulence (Inverso *et al.* 2010). However, the *ESAG* complement of this expression site is unknown, and represents a very small sample size.

Transcription of *ESAGs* from the BES provides a lifecycle-specific control of expression, as this only occurs in BSF cells. But in some cases, GRESAGs are also expressed in this lifecycle stage, and appear to functionally complement the BES-derived protein. So why are these genes positioned at the expression site? When SRA is transfected into the *T. brucei* rDNA locus, mRNA levels are only 5% of that from the *T. b. rhodesiense* RES, but still provides resistance to human serum (Xong *et al.* 1998). This suggests that increased transcript levels mediated by RNA polymerase I are not important to the resistance mechanism. SRA need only be produced by BSF cells, but *T. b. gambiense* constitutively expresses the protein that confers resistance in this subspecies (TgsGP), thus the parasite need not rely on activation of a specific BES to survive this environment. The production of numerous other proteins is lifecycle-stage specific, controlled through mechanisms other than their genomic location, so the benefit of ESAG expression from the active BES remains unknown.

Linkage of contingency gene families, and those with roles at the host parasite interface, appears to be conserved across protozoan parasites, and may also be a feature of other pathogens. This study has identified redundancy between *(GR)ESAG* family members at distinct genomic locations, and that transcription of BES-linked family members is not essential for parasite survival *in vitro*. These data do not explain why members of these specific *T. brucei* gene families have moved, and are maintained, at the VSG expression site. This feature, and the functions of a number of these families, remain enigmatic.

V. References

Adam RD. 2001. Biology of *Giardia lamblia*. *Clin Microbiol Rev* **14**: 447–475.

Adam RD, Nigam A, Seshadri V, Martens CA, Farneth GA, Morrison HG, Nash TE, Porcella SF, Patel R. 2010. The *Giardia lamblia vsp* gene repertoire: characteristics, genomic organization, and evolution. *BMC Genomics* **11**: 424.

Aitcheson N, Talbot S, Shapiro J, Hughes K, Adkin C, Butt T, Sheader K, Rudenko G. 2005. VSG switching in *Trypanosoma brucei*: antigenic variation analysed using RNAi in the absence of immune selection. *Mol Microbiol* **57**: 1608–1622.

Al-Khedery B, Allred DR. 2006. Antigenic variation in *Babesia bovis* occurs through segmental gene conversion of the *ves* multigene family, within a bidirectional locus of active transcription. *Mol Microbiol* **59**: 402–414.

Alarcon CM, Son HJ, Hall T, Donelson JE. 1994. A monocistronic transcript for a trypanosome variant surface glycoprotein. *Mol Cell Biol* **14**: 5579–5591.

Alexandre S, Paindavoine P, Hanocq-Quertier J, Paturiaux-Hanocq F, Tebabi P, Pays E. 1996. Families of adenylate cyclase genes in *Trypanosoma brucei*. *Mol Biochem Parasitol* **77**: 173–182.

Alexandre S, Paindavoine P, Tebabi P, Pays A, Halleux S, Steinert M, Pays E. 1990. Differential expression of a family of putative adenylate/guanylate cyclase genes in *Trypanosoma brucei*. *Mol Biochem Parasitol* **43**: 279–288.

Allen CL, Goulding D, Field MC. 2003. Clathrin-mediated endocytosis is essential in *Trypanosoma brucei*. *EMBO J* **22**: 4991–5002.

Allred DR, Carlton JM-R, Satcher RL, Long JA, Brown WC, Patterson PE, O'Connor RM, Stroup SE. 2000. The *ves* multigene family of *B. bovis* encodes components of rapid antigenic variation at the infected erythrocyte surface. *Mol Cell* **5**: 153–162.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Amiguet-Vercher A, Pérez-Morga D, Pays A, Poelvoorde P, Xong HV, Tebabi P, Vanhamme L, Pays E. 2004. Loss of the mono-allelic control of the *VSG* expression sites during the development of *Trypanosoma brucei* in the bloodstream. *Mol Microbiol* **51**: 1577–1588.

Ansorge I, Steverding D, Melville S, Hartmann C, Clayton C. 1999. Transcription of "inactive" expression sites in African trypanosomes leads to expression of multiple transferrin receptor RNAs in bloodstream forms. *Mol Biochem Parasitol* **101**: 81–94.

Atayde VD, Tschudi C, Ullu E. 2011. The emerging world of small silencing RNAs in protozoan parasites. *Trends Parasitol* **27**: 321–327.

Auffret CA, Turner MJ. 1981. Variant specific antigens of *Trypanosoma brucei* exist in solution as glycoprotein dimers. *Biochem J* **193**: 647–650.

Bachmann A, Scholz JAM, Janßen M, Klinkert M-Q, Tannich E, Bruchhaus I, Petter M. 2015. A comparative study of the localization and membrane topology of members of the RIFIN, STEVOR and *Pf*MC-2TM protein families in *Plasmodium falciparum*-infected erythrocytes. *Malar J* **14**: 274.

Barbet AF, McGuire TC. 1978. Crossreacting determinants in variant-specific surface antigens of African trypanosomes. *Proc Natl Acad Sci U S A* **75**: 1989–1993.

Barker AR, Wickstead B, Gluenz E, Gull K. 2008. Bioinformatic insights to the *ESAG5* and *GRESAG5* gene families in kinetoplastid parasites. *Mol Biochem Parasitol* **162**: 112–122.

Barnes RL, McCulloch R. 2007. *Trypanosoma brucei* homologous recombination is dependent on substrate length and homology, though displays a differential dependence on mismatch repair as substrate length decreases. *Nucleic Acids Res* **35**: 3478–3493.

Barnwell EM, van Deursen FJ, Jeacock L, Smith KA, Maizels RM, Acosta-Serrano A, Matthews K. 2010. Developmental regulation and extracellular release of a *VSG* expression-site-associated gene product from *Trypanosoma brucei* bloodstream forms. *J Cell Sci* **123**: 3401–3411.

Barrett B, LaCount DJ, Donelson JE. 2004. *Trypanosoma brucei*: A first-generation CRE-*loxP* site-specific recombination system. *Exp Parasitol* **106**: 37–44.

Barrett MP, Burchmore RJS, Stich A, Lazzari JO, Frasch AC, Cazzulo JJ, Krishna S. 2003. The trypanosomiases. *Lancet* **362**: 1469–1480.

Barry JD. 1986. Antigenic variation during *Trypanosoma vivax* infections of different host species. *Parasitology* **92**: 51–65.

Barry JD, Ginger ML, Burton P, McCulloch R. 2003. Why are parasite contingency genes often associated with telomeres? *Int J Parasitol* **33**: 29–45.

Bart JM, Cordon-Obras C, Vidal I, Reed J, Perez-Pastrana E, Cuevas L, Field MC, Carrington M, Navarro M. 2015. Localization of serum resistance-associated protein in *Trypanosoma brucei rhodesiense* and transgenic *Trypanosoma brucei brucei*. *Cell Microbiol* **17**: 1523–1535.

Bastin P, Ellis K, Kohl L, Gull K. 2000. Flagellum ontogeny in trypanosomes studied via an inherited and regulated RNA interference system. *J Cell Sci* **113**: 3321–3328.

Batram C, Jones NG, Janzen CJ, Markert SM, Engstler M. 2014. Expression site attenuation mechanistically links antigenic variation and development in *Trypanosoma brucei*. *Elife* **3**: e02324.

Becker M, Aitcheson N, Byles E, Wickstead B, Louis E, Rudenko G. 2004. Isolation of the repertoire of *VSG* expression site containing telomeres of *Trypanosoma brucei* 427 using transformation-associated recombination in yeast. *Genome Res* **14**: 2319–2329.

Beneke T, Madden R, Makin L, Valli J, Sunter J, Gluenz E. 2017. A CRISPR Cas9 high-throughput genome editing toolkit for kinetoplastids. *R Soc Open Sci* **4**: 170095.

Berberof M, Pays A, Pays E. 1991. A similar gene is shared by both the variant surface glycoprotein and procyclin gene transcription units of *Trypanosoma brucei*. *Mol Cell Biol* **11**: 1473–1479.

Bernards A, Lex HT, Van der Ploeg A, Frasch CC, Borst P, Boothroyd JC, Coleman S, Cross GAM. 1981. Activation of trypanosome surface glycoprotein genes involves a duplication-transposition leading to an altered 3′ end. *Cell* **27**: 497–505.

Bernards A, Titia DL, Michels PAM, Liu AYC, Huisman MJ, Borst P. 1984. Two modes of activation of a single surface antigen gene of *Trypanosoma brucei*. *Cell* **36**: 163–170.

Bernstein E, Caudy AA, Hammond SM, Hannon GJ. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363–366.

Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, *et al.* 2005. The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309**: 416–422.

Berriman M, Hall N, Sheader K, Bringaud F, Tiwari B, Isobe T, Bowman S, Corton C, Clark L, Cross GAM, *et al.* 2002. The architecture of variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Mol Biochem Parasitol* **122**: 131–140.

Bieger B, Essen LO. 2001. Structural analysis of adenylate cyclases from *Trypanosoma brucei* in their monomeric state. *EMBO J* **20**: 433–445.

Bitter W, Gerrits H, Kieft R, Borst P. 1998. The role of transferrin-receptor variation in the host range of *Trypanosoma brucei*. *Nature* **391**: 499–502.

Blum ML, Down JA, Gurnett AM, Carrington M, Turner MJ, Wiley DC. 1993. A structural motif in the variant surface glycoproteins of *Trypanosoma brucei*. *Nature* **362**: 603–609.

Blythe JE, Xue YY, Kuss C, Bozdech Z, Holder AA, Marsh K, Langhorne J, Preiser PR. 2008. *Plasmodium falciparum* STEVOR proteins are highly expressed in patient isolates and located in the surface membranes of infected red blood cells and the apical tips of merozoites. *Infect Immun* **76**: 3329–3336.

Boothroyd CE, Dreesen O, Leonova T, Ly KI, Figueiredo LM, Cross GAM, Papavasiliou FN. 2009. A yeast-endonuclease-generated DNA break induces antigenic switching in *Trypanosoma brucei*. *Nature* **459**: 278–281.

Brayton KA, Lau AOT, Herndon DR, Hannick L, Kappmeyer LS, Berens SJ, Bidwell SL, Brown WC, Crabtree J, Fadrosh D, *et al.* 2007. Genome sequence of *Babesia bovis* and comparative analysis of apicomplexan hemoprotozoa. *PLoS Pathog* **3**: 1401–1413.

Bridges DJ, Pitt AR, Hanrahan O, Brennan K, Voorheis HP, Herzyk P, de Koning HP, Burchmore RJS. 2008. Characterisation of the plasma membrane subproteome of bloodstream form *Trypanosoma brucei*. *Proteomics* **8**: 83–99.

Caffrey DR, Zhao J, Song Z, Schaffer ME, Haney SA, Subramanian RR, Seymour AB, Hughes JD. 2011. siRNA off-target effects can be reduced at concentrations that match their individual potency. *PLoS One* **6**: e21503.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Campillo N, Carrington M. 2003. The origin of the serum resistance associated (*SRA*) gene and a model of the structure of the SRA polypeptide from *Trypanosoma brucei rhodesiense*. *Mol Biochem Parasitol* **127**: 79–84.

Capewell P, Clucas C, DeJesus E, Kieft R, Hajduk S, Veitch N, Steketee PC, Cooper A, Weir W, MacLeod A. 2013. The TgsGP gene is essential for resistance to human serum in *Trypanosoma brucei gambiense*. *PLoS Pathog* **9**: e1003686.

Carranza PG, Gargantini PR, Prucca CG, Torri A, Saura A, Svärd S, Lujan HD. 2016. Specific histone modifications play critical roles in the control of encystation and antigenic variation in the early-branching eukaryote *Giardia lamblia*. *Int J Biochem Cell Biol* **81**: 32–43.

Carrington M, Boothroyd J. 1996. Implications of conserved structural motifs in disparate trypanosome surface proteins. *Mol Biochem Parasitol* **81**: 119–126.

Carrington M, Miller N, Blum M, Roditi I, Wiley D, Turner M. 1991. Variant specific glycoprotein of *Trypanosoma brucei* consists of two domains each having an independently conserved pattern of cysteine residues. *J Mol Biol* **221**: 823–835.

Carruthers VB, Navarro M, Cross GAM. 1996. Targeted disruption of expression site-associated gene-1 in bloodstream-form *Trypanosoma brucei*. *Mol Biochem Parasitol* **81**: 65–79.

Carruthers VB, Van der Ploeg LHT, Cross GAM. 1993. DNA-mediated transformation of bloodstream-form *Trypanosoma brucei*. *Nucleic Acids Res* **21**: 2537–2538.

Chattopadhyay A, Jones NG, Nietlispach D, Nielsen PR, Voorheis HP, Mott HR, Carrington M. 2005. Structure of the C-terminal domain from *Trypanosoma brucei* variant surface glycoprotein MITat1.2. *J Biol Chem* **280**: 7228–7235.

Chaudhri M, Steverding D, Kittelberger D, Tjia S, Overath P. 1994. Expression of a glycosylphosphatidylinositol-anchored *Trypanosoma brucei* transferrin-binding protein complex in insect cells. *Proc Natl Acad Sci U S A* **91**: 6443–6447.

Cheng Q, Cloonan N, Fischer K, Thompson J, Waine G, Lanzer M, Saul A. 1998. *stevor* and *rif* are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. *Mol Biochem Parasitol* **97**: 161–176.

Cheung JLY, Wand N V, Ooi C-P, Ridewood S, Wheeler RJ, Rudenko G. 2016. Blocking synthesis of the variant surface glycoprotein coat in *Trypanosoma brucei* leads to an increase in macrophage phagocytosis due to reduced clearance of surface coat antibodies. *PLoS Pathog* **12**: e1006023.

Christiano R, Kolev NG, Shi H, Ullu E, Walther TC, Tschudi C. 2017. The proteome and transcriptome of the infectious metacyclic form of *Trypanosoma brucei* define quiescent cells primed for mammalian invasion. *Mol Microbiol* **106**: 74–92.

Claes F, Büscher P, Touratier L, Goddeeris BM. 2005. *Trypanosoma equiperdum*: master of disguise or historical mistake? *Trends Parasitol* **21**: 316–321.

Clayton C, Shapira M. 2007. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. *Mol Biochem Parasitol* **156**: 93–101.

Cordon-Obras C, Cano J, González-Pacanowska D, Benito A, Navarro M, Bart J-M. 2013. *Trypanosoma brucei gambiense* adaptation to different mammalian sera is associated with *VSG* expression site plasticity. *PLoS One* **8**: e85072.

Cross GAM, Kim H-S, Wickstead B. 2014. Capturing the variant surface glycoprotein repertoire (the VSGnome) of *Trypanosoma brucei* Lister 427. *Mol Biochem Parasitol* **195**: 59–73.

Cully DF, Gibbs CP, Cross GAM. 1986. Identification of proteins encoded by variant surface glycoprotein expression site-associated genes in *Trypanosoma brucei*. *Mol Biochem Parasitol* **21**: 189–197.

Cuypers B, Van den Broeck F, Van Reet N, Meehan CJ, Cauchard J, Wilkes JM, Claes F, Goddeeris B, Birhanu H, Dujardin J-C, *et al.* 2017. Genome-Wide SNP Analysis Reveals Distinct Origins of *Trypanosoma evansi* and *Trypanosoma equiperdum*. *Genome Biol Evol* **9**: 1990–1997.

Czichos J, Nonnengaesser C, Overath P. 1986. *Trypanosoma brucei*: *cis*-aconitate and temperature reduction as triggers of synchronous transformation of bloodstream to procyclic trypomastigotes *in vitro*. *Exp Parasitol* **62**: 283–291.

D'Archivio S, Wickstead B. 2017. Trypanosome outer kinetochore proteins suggest conservation of chromosome segregation machinery across eukaryotes. *J Cell Biol* **216**: 379–391.

Davies KP, Carruthers VB, Cross GA. 1997. Manipulation of the *vsg* co-transposed region increases expression-site switching in *Trypanosoma brucei*. *Mol Biochem Parasitol* **86**: 163–177.

De Greef C, Hamers R. 1994. The serum resistance-associated (SRA) gene of *Trypanosoma brucei rhodesiense* encodes a variant surface glycoprotein-like protein. *Mol Biochem Parasitol* **68**: 277–284.

Dean S, Marchetti R, Kirk K, Matthews KR. 2009. A surface transporter family conveys the trypanosome differentiation signal. *Nature* **459**: 213–217.

De Pablos LM, Osuna A. 2012. Multigene families in *Trypanosoma cruzi* and their role in infectivity. *Infect Immun* **80**: 2258–2264.

Deitsch KW, Lukehart SA, Stringer JR. 2009. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat Rev Microbiol* **7**: 493–503.

Deitsch KW, Moxon ER, Wellems TE. 1997. Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal Infections. *Microbiol Mol Biol Rev* **61**: 281–293.

DeJesus E, Kieft R, Albright B, Stephens NA, Hajduk SL. 2013. A single amino acid substitution in the group 1 *Trypanosoma brucei gambiense* haptoglobin-hemoglobin receptor abolishes TLF-1 binding. *PLoS Pathog* **9**: e1003317.

Dempsey WL, Mansfield JM. 1983. Lymphocyte function in experimental African trypanosomiasis. V. Role of antibody and the mononuclear phagocyte system in variant-specific immunity. *J Immunol* **130**: 405–411.

Desquesnes M, Holzmuller P, Lai D-H, Dargantes A, Lun Z-R, Jittaplapong S. 2013. *Trypanosoma evansi* and surra: a review and perspectives on origin, history, distribution, taxonomy, morphology, hosts, and pathogenic effects. *Biomed Res Int* **2013**: 194176.

Devlin R, Marques CA, McCulloch R. 2017. Does DNA replication direct locus-specific recombination during host immune evasion by antigenic variation in the African trypanosome? *Curr Genet* **63**: 441–449.

Dewar S, Sienkiewicz N, Ong HB, Wall RJ, Horn D, Fairlamb AH. 2016. The role of folate transport in antifolate drug action in *Trypanosoma brucei*. *J Biol Chem* **291**: 24768–24778.

Diffley P. 1985. *Trypanosoma brucei*: immunogenicity of the variant surface coat glycoprotein of virulent and avirulent subspecies. *Exp Parasitol* **59**: 98–107.

Diffley P, Strickler JE, Patton CL, Waksman BH. 1980. Detection and quantification of variant specific antigen in the plasma of rats and mice infected with *Trypanosoma brucei brucei*. *J Parasitol* **66**: 185–191.

Doudna JA, Charpentier E. 2014. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**: 1258096.

Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR, Bartel DP. 2009. RNAi in budding yeast. *Science* **326**: 544–550.

Duraisingh MT, Horn D. 2016. Epigenetic regulation of virulence gene expression in parasitic protozoa. *Cell Host Microbe* **19**: 629–640.

Durand-Dubief M, Bastin P. 2003. TbAGO1, an argonaute protein required for RNA interference, is involved in mitosis and chromosome segregation in *Trypanosoma brucei*. *BMC Biol* **1**: 2.

Engstler M, Pfohl T, Herminghaus S, Boshart M, Wiegertjes G, Heddergott N, Overath P. 2007. Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. *Cell* **131**: 505–515.

Engstler M, Thilo L, Weise F, Grünfelder CG, Schwarz H, Boshart M, Overath P. 2004. Kinetics of endocytosis and recycling of the GPI-anchored variant surface glycoprotein in *Trypanosoma brucei*. *J Cell Sci* **117**: 1105–1115.

Engstler M, Weise F, Bopp K, Grünfelder CG, Günzel M, Heddergott N, Overath P. 2005. The membrane-bound histidine acid phosphatase *Tb*MBAP1 is essential for endocytosis and membrane recycling in *Trypanosoma brucei*. *J Cell Sci* **118**: 2105–2118.

Ersfeld K, Melville SE, Gull K. 1999. Nuclear and genome organization of *Trypanosoma brucei*. *Parasitol today* **15**: 58–63.

Estévez AM, Kempf T, Clayton C. 2001. The exosome of *Trypanosoma brucei*. *EMBO J* **20**: 3831–3839.

Fast B, Kremp K, Boshart M, Steverding D. 1999. Iron-dependent regulation of transferrin receptor expression in *Trypanosoma brucei*. *Biochem J* **342**: 691–696.

Ferguson MAJ, Homans SW, Dwek RA, Rademacher TW. 1988. Glycosyl-phosphatidylinositol moiety that anchors *Trypanosoma brucei* variant surface glycoprotein to the membrane. *Science* **239**: 753–759.

Fernandez V, Hommel M, Chen Q, Hagblom P, Wahlgren M. 1999. Small, clonally variant antigens expressed on the surface of the *Plasmodium falciparum*-infected erythrocyte are encoded by the *rif* gene family and are the target of human immune responses. *J Exp Med* **190**: 1393–1404.

Ferrante A, Allison AC. 1983. Alternative pathway activation of complement by African trypanosomes lacking a glycoprotein coat. *Parasite Immunol* **5**: 491–498.

Ferreri LM, Brayton KA, Sondgeroth KS, Lau AOT, Suarez CE, McElwain TF. 2012. Expression and strain variation of the novel "*small open reading frame*" (*smorf*) multigene family in *Babesia bovis*. *Int J Parasitol* **42**: 131–138.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**: 806–811.

Freymann D, Down J, Carrington M, Roditi I, Turner M, Wiley D. 1990. 2.9 Å resolution structure of the N-terminal domain of a variant surface glycoprotein from *Trypanosoma brucei*. *J Mol Biol* **216**: 141–160.

Gadelha C, Zhang W, Chamberlain JW, Chait BT, Wickstead B, Field MC. 2015. Architecture of a host-parasite interface: complex targeting mechanisms revealed through proteomics. *Mol Cell Proteomics* **14**: 1911–1926.

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, *et al.* 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**: 498–511.

Gerrits H, Mußmann R, Bitter W, Kieft R, Borst P. 2002. The physiological significance of transferrin receptor variations in *Trypanosoma brucei*. *Mol Biochem Parasitol* **119**: 237–247.

Ginger ML, Blundell PA, Lewis AM, Browitt A, Günzl A, Barry JD. 2002. *Ex vivo* and *in vitro* identification of a consensus promoter for *VSG* genes expressed by metacyclic-stage trypanosomes in the tsetse fly. *Eukaryot Cell* **1**: 1000–1009.

Gizaw Y, Megersa M, Fayera T. 2017. Dourine: a neglected disease of equids. *Trop Anim Health Prod* **49**: 887–897.

Glover L, Alsford S, Beattie C, Horn D. 2007. Deletion of a trypanosome telomere leads to loss of silencing and progressive loss of terminal DNA in the absence of cell cycle arrest. *Nucleic Acids Res* **35**: 872–880.

Glover L, Alsford S, Horn D. 2013a. DNA break site at fragile subtelomeres determines probability and mechanism of antigenic variation in African trypanosomes. *PLoS Pathog* **9**: e1003260.

Glover L, Hutchinson S, Alsford S, Mcculloch R, Field MC, Horn D. 2013b. Antigenic variation in African trypanosomes: the importance of chromosomal and nuclear context in *VSG* expression control. *Cell Microbiol* **15**: 1984–1993.

Glover L, Jun J, Horn D. 2011. Microhomology-mediated deletion and gene conversion in African trypanosomes. *Nucleic Acids Res* **39**: 1372–1380.

Glover L, McCulloch R, Horn D. 2008. Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. *Nucleic Acids Res* **36**: 2608–2618.

Goel S, Palmkvist M, Moll K, Joannin N, Lara P, R Akhouri R, Moradi N, Öjemalm K, Westman M, Angeletti D, *et al.* 2015. RIFINs are adhesins implicated in severe *Plasmodium falciparum* malaria. *Nat Med* **21**: 314–317.

Gottesdiener K, Chung HM, Brown SD, Lee MG, Van der Ploeg LHT. 1991. Characterization of VSG gene expression site promoters and promoter-associated DNA rearrangement events. *Mol Cell Biol* **11**: 2467–2480.

Gottesdiener KM, Goriparthi L, Masucci JP, Van der Ploeg LHT. 1992. A proposed mechanism for promoter-associated DNA rearrangement events at a variant surface glycoprotein gene expression site. *Mol Cell Biol* **12**: 4784–4795.

Graham S V, Barry JD. 1991. Expression site-associated genes transcribed independently of variant surface glycoprotein genes in *Trypanosoma brucei*. *Mol Biochem Parasitol* **47**: 31–41.

Günzl A, Bruderer T, Laufer G, Schimanski B, Tu L-C, Chung H-M, Lee P-T, Lee MG-S. 2003. RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. *Eukaryot Cell* **2**: 542–551.

Helm JR, Hertz-Fowler C, Aslett M, Berriman M, Sanders M, Quail MA, Soares MB, Bonaldo MF, Sakurai T, Inoue N, *et al.* 2009. Analysis of expressed sequence tags from the four main developmental stages of *Trypanosoma congolense*. *Mol Biochem Parasitol* **168**: 34–42.

Hertz-Fowler C, Figueiredo LM, Quail MA, Becker M, Jackson A, Bason N, Brooks K, Churcher C, Fahkro S, Goodhead I, *et al.* 2008. Telomeric expression sites are highly conserved in *Trypanosoma brucei*. *PLoS One* **3**: e3527.

Hobbs MR, Boothroyd JC. 1990. An expression-site-associated gene family of trypanosomes is expressed *in vivo* and shows homology to a variant surface glycoprotein gene. *Mol Biochem Parasitol* **43**: 1–16.

Hoek M, Cross GAM. 2001. Expression-site-associated-gene-8 (*ESAG8*) is not required for regulation of the *VSG* expression site in *Trypanosoma brucei*. *Mol Biochem Parasitol* **117**: 211–215.

Hoek M, Engstler M, Cross GAM. 2000. Expression-site-associated gene 8 (*ESAG8*) of *Trypanosoma brucei* is apparently essential and accumulates in the nucleolus. *J Cell Sci* **113**: 3959–3968.

Hoek M, Zanders T, Cross GAM. 2002. *Trypanosoma brucei* expression-site-associated-gene-8 protein interacts with a *Pumilio* family protein. *Mol Biochem Parasitol* **120**: 269–283.

Horn D. 2014. Antigenic variation in African trypanosomes. *Mol Biochem Parasitol* **195**: 123–129.

Hovel-Miner G, Mugnier M, Papavasiliou FN, Pinger J, Schulz D. 2015. A host-pathogen interaction reduced to first principles: antigenic variation in *T. brucei*. *Results Probl Cell Differ* **57**: 23–46.

Hovel-Miner G, Mugnier MR, Goldwater B, Cross GAM, Papavasiliou FN. 2016. A conserved DNA repeat promotes selection of a diverse repertoire of *Trypanosoma brucei* surface antigens from the genomic archive. *PLoS Genet* **12**: e1005994.

Huang Y, Xiao Y-P, Allred DR. 2013. Unusual chromatin structure associated with monoparalogous transcription of the *Babesia bovis ves* multigene family. *Int J Parasitol* **43**: 163–172.

Hutchinson S, Glover L, Horn D. 2016. High-resolution analysis of multi-copy variant surface glycoprotein gene expression sites in African trypanosomes. *BMC Genomics* **17**: 806.

Inverso JA, Uphoff TS, Johnson SC, Paulnock DM, Mansfield JM. 2010. Biological variation among African trypanosomes: I. Clonal expression of virulence is not linked to the variant surface glycoprotein or the variant surface glycoprotein gene telomeric expression site. *DNA Cell Biol* **29**: 215–227.

Isobe T, Holmes EC, Rudenko G. 2003. The transferrin receptor genes of *Trypanosoma equiperdum* are less diverse in their transferrin binding site than those of the broad-host range *Trypanosoma brucei*. *J Mol Evol* **56**: 377–386.

Jackson AP, Allison HC, Barry JD, Field MC, Hertz-Fowler C, Berriman M. 2013. A cell-surface phylome for African trypanosomes. *PLoS Negl Trop Dis* **7**: e2121.

Jackson AP, Berry A, Aslett M, Allison HC, Burton P, Vavrova-Anderson J, Brown R, Browne H, Corton N, Hauser H, *et al.* 2012. Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. *Proc Natl Acad Sci* **109**: 3416–3421.

Jackson AP, Otto TD, Darby A, Ramaprasad A, Xia D, Echaide IE, Farber M, Gahlot S, Gamble J, Gupta D, *et al.* 2014. The evolutionary dynamics of variant antigen genes in *Babesia* reveal a history of genomic innovation underlying host-parasite interaction. *Nucleic Acids Res* **42**: 7113–7131.

Jackson DG, Owen MJ, Voorheis HP. 1985. A new method for the rapid purification of both the membrane-bound and released forms of the variant surface glycoprotein from *Trypanosoma brucei*. *Biochem J* **230**: 195–202.

Jackson DG, Windle HJ, Voorheis HP. 1993. The identification, purification, and characterization of two invariant surface glycoproteins located beneath the surface coat barrier of bloodstream forms of *Trypanosoma brucei*. *J Biol Chem* **268**: 8085–8095.

Jensen BC, Sivam D, Kifer CT, Myler PJ, Parsons M. 2009. Widespread variation in transcript abundance within and across developmental stages of *Trypanosoma brucei*. *BMC Genomics* **10**: 482.

Kabani S, Fenn K, Ross A, Ivens A, Smith TK, Ghazal P, Matthews K. 2009. Genome-wide expression profiling of *in vivo*-derived bloodstream parasite stages and dynamic analysis of mRNA alterations during synchronous differentiation in *Trypanosoma brucei*. *BMC Genomics* **10**: 427.

Kabiri M, Steverding D. 2000. Studies on the recycling of the transferrin receptor in *Trypanosoma brucei* using an inducible gene expression system. *Eur J Biochem* **267**: 3309–3314.

Kamper SM, Barbet AF. 1992. Surface epitope variation via mosaic gene formation is potential key to long-term survival of *Trypanosoma brucei*. *Mol Biochem Parasitol* **53**: 33–44.

Kassem A, Pays E, Vanhamme L. 2014. Transcription is initiated on silent variant surface glycoprotein expression sites despite monoallelic expression in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A* **111**: 8943–8948.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772–780.

Kaviratne M, Khan SM, Jarra W, Preiser PR. 2002. Small variant STEVOR antigen is uniquely located within Maurer's clefts in *Plasmodium falciparum*-infected red blood cells. *Eukaryot Cell* **1**: 926–935.

Kelly S, Reed J, Kramer S, Ellis L, Webb H, Sunter J, Salje J, Marinsek N, Gull K, Wickstead B, *et al.* 2007. Functional genomics in *Trypanosoma brucei*: a collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci. *Mol Biochem Parasitol* **154**: 103–109.

Kieft R, Capewell P, Turner CMR, Veitch NJ, MacLeod A, Hajduk S. 2010. Mechanism of *Trypanosoma brucei gambiense* (group 1) resistance to human trypanosome lytic factor. *Proc Natl Acad Sci U S A* **107**: 16137–16141.

Kim HS, Li Z, Boothroyd C, Cross GAM. 2013. Strategies to construct null and conditional null *Trypanosoma brucei* mutants using Cre-recombinase and loxP. *Mol Biochem Parasitol* **191**: 16–19.

Koenig-Martin E, Yamage M, Roditi I. 1992. A procyclin-associated gene in *Trypanosoma brucei* encodes a polypeptide related to ESAG 6 and 7 proteins. *Mol Biochem Parasitol* **55**: 135–145.

Kolev NG, Ramey-Butler K, Cross GAM, Ullu E, Tschudi C. 2012. Developmental progression to infectivity in *Trypanosoma brucei* triggered by an RNA-binding protein. *Science* **338**: 1352–1353.

Kolev NG, Tschudi C, Ullu E. 2011. RNA interference in protozoan parasites: achievements and challenges. *Eukaryot Cell* **10**: 1156–1163.

Kraemer SM, Kyes SA, Aggarwal G, Springer AL, Nelson SO, Christodoulou Z, Smith LM, Wang W, Levin E, Newbold CI, *et al.* 2007. Patterns of gene recombination shape *var* gene repertoires in *Plasmodium falciparum*: comparisons of geographically diverse isolates. *BMC Genomics* **8**: 45.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567–580.

Kyes SA, Rowe JA, Kriek N, Newbold CI. 1999. Rifins: a second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum*. *Med Sci* **96**: 9333–9338.

LaCount DJ, Bruse S, Hill KL, Donelson JE. 2000. Double-stranded RNA interference in *Trypanosoma brucei* using head-to-head promoters. *Mol Biochem Parasitol* **111**: 67–76.

Lai D-H, Hashimi H, Lun Z-R, Ayala FJ, Lukes J. 2008. Adaptations of *Trypanosoma brucei* to gradual loss of kinetoplast DNA: *Trypanosoma equiperdum* and *Trypanosoma evansi* are petite mutants of *T. brucei*. *Proc Natl Acad Sci U S A* **105**: 1999–2004.

Landeira D, Bart J-M, Van Tyne D, Navarro M. 2009. Cohesin regulates *VSG* monoallelic expression in trypanosomes. *J Cell Biol* **186**: 243–254.

Lander N, Chiurillo MA, Docampo R. 2016a. Genome editing by CRISPR/Cas9: a game change in the genetic manipulation of protists. *J Eukaryot Microbiol* **63**: 679–690.

Lander N, Chiurillo MA, Storey M, Vercesi AE, Docampo R. 2016b. CRISPR/Cas9-mediated endogenous C-terminal tagging of *Trypanosoma cruzi* genes reveals the acidocalcisome localization of the inositol-1,4,5-trisphosphate receptor. *J Biol Chem* **291**: 25505–25515.

Lander N, Chiurillo MA, Vercesi AE, Docampo R. 2017. Endogenous C-terminal tagging by CRISPR/Cas9 in *Trypanosoma cruzi*. *Bio-protocol* **7**: e2299.

Lander N, Li Z-H, Niyogi S, Docampo R. 2015. CRISPR/Cas9-induced disruption of paraflagellar rod protein 1 and 2 genes in *Trypanosoma cruzi* reveals their role in flagellar attachment. *MBio* **6**: e01012-15.

Lavazec C, Sanyal S, Templeton TJ. 2007. Expression switching in the *stevor* and *Pfmc-2TM* superfamilies in *Plasmodium falciparum*. *Mol Microbiol* **64**: 1621–1634.

Lavazec C, Sanyal S, Templeton TJ. 2006. Hypervariability within the Rifin, Stevor and *Pf*mc-2TM superfamilies in *Plasmodium falciparum*. *Nucleic Acids Res* **34**: 6696–6707.

Laxman S, Beavo JA. 2007. Cyclic nucleotide signaling mechanisms in trypanosomes: possible targets for therapeutic agents. *Mol Interv* **7**: 203–215.

Laxman S, Riechers A, Sadilek M, Schwede F, Beavo JA. 2006. Hydrolysis products of cAMP analogs cause transformation of *Trypanosoma brucei* from slender to stumpy-like forms. *Proc Natl Acad Sci U S A* **103**: 19194–19199.

Lemley C, Yan S, Dole VS, Madhubala R, Cunningham ML, Beverly SM, Myler PJ, Stuart KD. 1999. The *Leishmania donovani* LD1 locus gene *ORFG* encodes a biopterin transporter (BT1). *Mol Biochem Parasitol* **104**: 93–105.

Li B. 2015. DNA double-strand breaks and telomeres play important roles in *Trypanosoma brucei* antigenic variation. *Eukaryot Cell* **14**: 196–205.

Li B. 2012. Telomere components as potential therapeutic targets for treating microbial pathogen infections. *Front Oncol* **2**: 156.

Li W, Saraiya AA, Wang CC. 2013. Experimental verification of the identity of variant-specific surface proteins in *Giardia lamblia* trophozoites. *MBio* **4**: e00321-13.

Ligtenberg MJ, Bitter W, Kieft R, Steverding D, Janssen H, Calafat J, Borst P. 1994. Reconstitution of a surface transferrin binding complex in insect form *Trypanosoma brucei*. *EMBO J* **13**: 2565–2573.

Liu J, Carmell MA, Rivas F V, Marsden CG, Thomson JM, Song J-J, Hammond SM, Joshua-Tor L, Hannon GJ. 2004. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**: 1437–1441.

MacGregor P, Ivens A, Shave S, Collie I, Gray D, Auer M, Matthews KR. 2014. High-throughput chemical screening for antivirulence developmental phenotypes in *Trypanosoma brucei*. *Eukaryot Cell* **13**: 412–426.

MacGregor P, Matthews KR. 2012. Identification of the regulatory elements controlling the transmission stage-specific gene expression of PAD1 in *Trypanosoma brucei*. *Nucleic Acids Res* **40**: 7705–7717.

MacGregor P, Szöőr B, Savill NJ, Matthews KR. 2012. Trypanosomal immune evasion, chronicity and transmission: an elegant balancing act. *Nat Rev Microbiol* **10**: 431–438.

Maier A, Steverding D. 2008. Expression and purification of non-glycosylated *Trypanosoma brucei* transferrin receptor in insect cells. *Exp Parasitol* **120**: 205–207.

Mancini PE, Patton CL. 1981. Cyclic 3',5'-adenosine monophosphate levels during the developmental cycle of *Trypanosoma brucei brucei* in the rat. *Mol Biochem Parasitol* **3**: 19–31.

Marcello L, Barry JD. 2007. Analysis of the *VSG* gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. *Genome Res* **17**: 1344–1352.

Mehlert A, Ferguson MAJ. 2007. Structure of the glycosylphosphatidylinositol anchor of the *Trypanosoma brucei* transferrin receptor. *Mol Biochem Parasitol* **151**: 220–223.

Mehlert A, Wormald MR, Ferguson MAJ. 2012. Modeling of the N-glycosylated transferrin receptor suggests how transferrin binding can occur within the surface coat of *Trypanosoma brucei*. *PLoS Pathog* **8**: e1002618.

Mehlert A, Zitzmann N, Richardson JM, Treumann A, Ferguson MAJ. 1998. The glycosylation of the variant surface glycoproteins and procyclic acidic repetitive proteins of *Trypanosoma brucei*. *Mol Biochem Parasitol* **91**: 145–152.

Monk SL, Simmonds P, Matthews KR. 2013. A short bifunctional element operates to positively or negatively regulate *ESAG9* expression in different developmental forms of *Trypanosoma brucei*. *J Cell Sci* **126**: 2294–2304.

Morgan GW, Allen CL, Jeffries TR, Hollinshead M, Field MC. 2001. Developmental and morphological regulation of clathrin-mediated endocytosis in *Trypanosoma brucei*. *J Cell Sci* **114**: 2605–2615.

Morgan RW, El-Sayed NMA, Kepa JK, Pedram M, Donelson JE. 1996. Differential expression of the expression site-associated gene I family in African trypanosomes. *J Biol Chem* **271**: 9771–9777.

Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, *et al.* 2007. Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. *Science* **317**: 1921–1926.

Morrison LJ, Majiwa P, Read AF, Barry JD. 2005. Probabilistic order in antigenic variation of *Trypanosoma brucei*. *Int J Parasitol* **35**: 961–972.

Moxon ER, Rainey PB, Nowak MA, Lenski RE. 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol* **4**: 24–33.

Moxon R, Bayliss C, Hood D. 2006. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet* **40**: 307–333.

Mugnier MR, Cross GAM, Papavasiliou FN. 2015. The *in vivo* dynamics of antigenic variation in *Trypanosoma brucei*. *Science* **347**: 1470–1473.

Mugo E, Egler F, Clayton C. 2017. Conversion of procyclic-form *Trypanosoma brucei* to the bloodstream form by transient expression of RBP10. *Mol Biochem Parasitol* **216**: 49–51.

Mußmann R, Engstler M, Gerrits H, Kieft R, Toaldo CB, Onderwater J, Koerten H, van Luenen HGAM, Borst P. 2004. Factors affecting the level and localization of the transferrin receptor in *Trypanosoma brucei*. *J Biol Chem* **279**: 40690–40698.

Mußmann R, Janssen H, Calafat J, Engstler M, Ansorge I, Clayton C, Borst P. 2003. The expression level determines the surface distribution of the transferrin receptor in *Trypanosoma brucei*. *Mol Microbiol* **47**: 23–35.

Myler P, Nelson RG, Agabian N, Stuart K. 1984. Two mechanisms of expression of a predominant variant antigen gene of *Trypanosoma brucei*. *Nature* **309**: 282–284.

Nash TE. 1997. Antigenic variation in *Giardia lamblia* and the host's immune response. *Philos Trans R Soc Lond B Biol Sci* **352**: 1369–1375.

Nash TE. 2002. Surface antigenic variation in *Giardia lamblia*. *Mol Microbiol* **45**: 585–590.

Naula C, Schaub R, Leech V, Melville S, Seebeck T. 2001. Spontaneous dimerization and leucine-zipper induced activation of the recombinant catalytic domain of a new adenylyl cyclase of *Trypanosoma brucei*, GRESAG4.4B. *Mol Biochem Parasitol* **112**: 19–28.

Naula C, Seebeck T. 2000. Cyclic AMP signaling in trypanosomatids. *Parasitol today* **16**: 35–38.

Navarro M, Cross GAM. 1996. DNA rearrangements associated with multiple consecutive directed antigenic switches in *Trypanosoma brucei*. *Mol Cell Biol* **16**: 3615–3625.

Ngô H, Tschudi C, Gull K, Ullu E. 1998. Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A* **95**: 14687–14692.

Niang M, Bei AK, Madnani KG, Pelly S, Dankwa S, Kanjee U, Gunalan K, Amaladoss A, Yeo KP, Bob NS, *et al.* 2014. STEVOR is a *Plasmodium falciparum* erythrocyte binding protein that mediates merozoite invasion and rosetting. *Cell Host Microbe* **16**: 81–93.

Niang M, Yan Yam X, Preiser PR. 2009. The *Plasmodium falciparum* STEVOR multigene family mediates antigenic variation of the infected erythrocyte. *PLoS Pathog* **5**: e1000307.

Nielsen H, Engelbrecht J, Brunak S, von Heijne G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**: 1–6.

Nielsen H, Krogh A. 1998. Prediction of signal peptides and signal anchors by a hidden Markov model. *Proceedings Int Conf Intell Syst Mol Biol* **6**: 122–130.

Nolan DP, Geuskens M, Pays E. 1999. *N*-linked glycans containing linear poly-*N*-acetyllactosamine as sorting signals in endocytosis in *Trypanosoma brucei*. *Curr Biol* **9**: 1169–1172.

Obado SO, Glover L, Deitsch KW. 2016. The nuclear envelope and gene organization in parasitic protozoa: specializations associated with disease. *Mol Biochem Parasitol* **209**: 104–113.

Ouellette M, Drummelsmith J, El Fadili A, Kündig C, Richard D, Roy G. 2002. Pterin transport and metabolism in *Leishmania* and related trypanosomatid parasites. *Int J Parasitol* **32**: 385–398.

Paindavoine P, Rolin S, Van Assel S, Geuskens M, Jauniaux JC, Dinsart C, Huet G, Pays E. 1992. A gene from the variant surface glycoprotein expression site encodes one of several transmembrane adenylate cyclases located on the flagellum of *Trypanosoma brucei*. *Mol Cell Biol* **12**: 1218–1225.

Palmer GH, Brayton KA. 2007. Gene conversion is a convergent strategy for pathogen antigenic variation. *Trends Parasitol* **23**: 408–413.

Patrick KL, Shi H, Kolev NG, Ersfeld K, Tschudi C, Ullu E. 2009. Distinct and overlapping roles for two Dicer-like proteins in the RNA interference pathways of the ancient eukaryote *Trypanosoma brucei*. *Proc Natl Acad Sci U S A* **106**: 17933–17938.

Pauli NT, Kim HK, Falugi F, Huang M, Dulac J, Henry Dunand C, Zheng N-Y, Kaur K, Andrews SF, Huang Y, *et al.* 2014. *Staphylococcus aureus* infection induces protein A–mediated immune evasion in humans. *J Exp Med* **211**: 2331–2339.

Pays E. 2006. The variant surface glycoprotein as a tool for adaptation in African trypanosomes. *Microbes Infect* **8**: 930–937.

Pays E, Coquelet H, Tebabi P, Pays A, Jefferies D, Steinert M, Koenig E, Williams RO, Roditi I. 1990. *Trypanosoma brucei*: constitutive activity of the VSG and procyclin gene promoters. *EMBO J* **9**: 3145–3151.

Pays E, Guyaux M, Aerts D, Van Meirvenne N, Steinert M. 1985. Telomeric reciprocal recombination as a possible mechanism for antigenic variation in trypanosomes. *Nature* **316**: 562–564.

Pays E, Lips S, Nolan D, Vanhamme L, Pérez-Morga D. 2001. The *VSG* expression sites of *Trypanosoma brucei*: multipurpose tools for the adaptation of the parasite to mammalian hosts. *Mol Biochem Parasitol* **114**: 1–16.

Pays E, Tebabi P, Pays A, Coquelet H, Revelard P, Salmon D, Steinert M. 1989. The genes and transcripts of an antigen gene expression site from *T. brucei*. *Cell* **57**: 835–845.

Pays E, Van Assel S, Laurent M, Darville M, Vervoort T, Van Meirvenne N, Steinert M. 1983. Gene conversion as a mechanism for antigenic variation in trypanosomes. *Cell* **34**: 371–381.

Pays E, Vanhollebeke B, Uzureau P, Lecordier L, Pérez-Morga D. 2014. The molecular arms race between African trypanosomes and humans. *Nat Rev Microbiol* **12**: 575–584.

Peng D, Kurup SP, Yao PY, Minning TA, Tarleton RL. 2014. CRISPR-Cas9-mediated single-gene and gene family disruption in *Trypanosoma cruzi*. *MBio* **6**: e02097-14.

Pérez-Morga D, Pays E. 1999. A protein linked to mitochondrion development in *Trypanosoma brucei*. *Mol Biochem Parasitol* **101**: 161–172.

Pérez-Morga D, Vanhollebeke B, Paturiaux-Hanocq F, Nolan DP, Lins L, Homblé F, Vanhamme L, Tebabi P, Pays A, Poelvoorde P, *et al.* 2005. Apolipoprotein L-I promotes trypanosome lysis by forming pores in lysosomal membranes. *Science* **309**: 469–472.

Pierleoni A, Martelli P, Casadio R. 2008. PredGPI: a GPI-anchor predictor. *BMC Bioinformatics* **9**: 392.

Poon S, Peacock L, Gibson W, Gull K, Kelly S. 2012. A modular and optimized single marker system for generating *Trypanosoma brucei* cell lines expressing T7 RNA polymerase and the tetracycline repressor. *Open Biol* **2**: 110037.

Prucca CG, Slavin I, Quiroga R, Elías E V, Rivero FD, Saura A, Carranza PG, Luján HD. 2008. Antigenic variation in *Giardia lamblia* is regulated by RNA interference. *Nature* **456**: 750–754.

Redpath MB, Windle H, Nolan D, Pays E, Voorheis HP, Carrington M. 2000. ESAG11, a new VSG expression site-associated gene from *Trypanosoma brucei*. *Mol Biochem Parasitol* **111**: 223–228.

Reid AJ. 2015. Large, rapidly evolving gene families are at the forefront of host–parasite interactions in *Apicomplexa*. *Parasitology* **142**: S57–S70.

Reuner B, Vassella E, Yutzy B, Boshart M. 1997. Cell density triggers slender to stumpy differentiation of *Trypanosoma brucei* bloodstream forms in culture. *Mol Biochem Parasitol* **90**: 269–280.

Rico E, Ivens A, Glover L, Horn D, Matthews KR. 2017. Genome-wide RNAi selection identifies a regulator of transmission stage-enriched gene families and cell-type differentiation in *Trypanosoma brucei*. *PLoS Pathog* **13**: e1006279.

Robinson NP, Burman N, Melville SE, Barry JD. 1999. Predominance of duplicative *VSG* gene conversion in antigenic variation in African trypanosomes. *Mol Cell Biol* **19**: 5839–5846.

Robinson O, Dylus D, Dessimoz C. 2016. Phylo.io: interactive viewing and comparison of large phylogenetic trees on the web. *Mol Biol Evol* **33**: 2163–2166.

Ross D, Raibaud A, Florent I, Sather S, Gross M, Storm D, Eisen H. 1991. The trypanosome VSG expression site encodes adenylate cyclase and a leucine-rich putative regulatory gene. *EMBO J* **10**: 2047–2053.

Rossi A, Kontarakis Z, Gerri C, Nolte H, Hölper S, Krüger M, Stainier DYR. 2015. Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* **524**: 230–233.

Roth C, Bringaud F, Layden RE, Baltz T, Eisen H. 1989. Active late-appearing variable surface antigen genes in *Trypanosoma equiperdum* are constructed entirely from pseudogenes. *Proc Natl Acad Sci U S A* **86**: 9375–9379.

Rowe JA, Moulds JM, Newbold CI, Miller LH. 1997. *P. falciparum* rosetting mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. *Nature* **388**: 292–295.

Rudenko G, McCulloch R, Dirks-Mulder A, Borst P. 1996. Telomere exchange can be an important mechanism of variant surface glycoprotein gene switching in *Trypanosoma brucei*. *Mol Biochem Parasitol* **80**: 65–75.

Saada EA, Kabututu ZP, Lopez M, Shimogawa MM, Langousis G, Oberholzer M, Riestra A, Jonsson ZO, Wohlschlegel JA, Hill KL. 2014. Insect stage-specific receptor adenylate cyclases are localized to distinct subdomains of the *Trypanosoma brucei* flagellar membrane. *Eukaryot Cell* **13**: 1064–1076.

Salmon D, Bachmaier S, Krumbholz C, Kador M, Gossmann JA, Uzureau P, Pays E, Boshart M. 2012a. Cytokinesis of *Trypanosoma brucei* bloodstream forms depends on expression of adenylyl cyclases of the ESAG4 or ESAG4- like subfamily. *Mol Microbiol* **84**: 225–242.

Salmon D, Geuskens M, Hanocq F, Hanocq-Quertier J, Nolan D, Ruben L, Pays E. 1994. A novel heterodimeric transferrin receptor encoded by a pair of VSG expression site-associated genes in *T. brucei*. *Cell* **78**: 75–86.

Salmon D, Hanocq-Quertier J, Paturiaux-Hanocq F, Pays A, Tebabi P, Nolan DP, Michel A, Pays E. 1997. Characterization of the ligand-binding site of the transferrin receptor in *Trypanosoma brucei* demonstrates a structural relationship with the N-terminal domain of the variant surface glycoprotein. *EMBO J* **16**: 7272–7278.

Salmon D, Paturiaux-Hanocq F, Poelvoorde P, Vanhamme L, Pays E. 2005. *Trypanosoma brucei*: growth differences in different mammalian sera are not due to the species-specificity of transferrin. *Exp Parasitol* **109**: 188–194.

Salmon D, Vanwalleghem G, Morias Y, Denoeud J, Krumbholz C, Lhommé F, Bachmaier S, Kador M, Gossmann J, Dias FBS, *et al.* 2012b. Adenylate cyclases of *Trypanosoma brucei* inhibit the innate immune response of the host. *Science* **337**: 463–466.

Salsman J, Dellaire G. 2017. Precision genome editing in the CRISPR era. *Biochem Cell Biol* **95**: 187–201.

Scahill MD, Pastar I, Cross GAM. 2008. CRE recombinase-based positive-negative selection systems for genetic manipulation in *Trypanosoma brucei*. *Mol Biochem Parasitol* **157**: 73–82.

Schell D, Evers R, Preis D, Ziegelbauer K, Kiefer H, Lottspeich F, Cornelissen AW, Overath P. 1991. A transferrin-binding protein of *Trypanosoma brucei* is encoded by one of the genes in the variant surface glycoprotein gene expression site. *EMBO J* **10**: 1061–1066.

Scherf A, Hernandez-Rivas R, Buffet P, Bottius E, Benatar C, Pouvelle B, Gysin J, Lanzer M. 1998. Antigenic variation in malaria: *in situ* switching, relaxed and mutually exclusive transcription of *var* genes during intra-erythrocytic development in *Plasmodium falciparum*. *EMBO J* **17**: 5418–5426.

Schumann Burkard G, Jutzi P, Roditi I. 2011. Genome-wide RNAi screens in bloodstream form trypanosomes identify drug transporters. *Mol Biochem Parasitol* **175**: 91–94.

Schwede A, Macleod OJS, MacGregor P, Carrington M. 2015. How does the VSG coat of bloodstream form African trypanosomes interact with external proteins? *PLOS Pathog* **11**: e1005259.

Seebeck T, Gong K, Kunz S, Schaub R, Shalaby T, Zoraghi R. 2001. cAMP signalling in *Trypanosoma brucei*. *Int J Parasitol* **31**: 491–498.

Shi H, Djikeng A, Mark T, Wirtz E, Tschudi C, Ullu E. 2000. Genetic interference in *Trypanosoma brucei* by heritable and inducible double-stranded RNA. *RNA* **6**: 1069–1076.

Shi H, Djikeng A, Tschudi C, Ullu E. 2004. Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retroposon transcript abundance. *Mol Cell Biol* **24**: 420–427.

Shi H, Tschudi C, Ullu E. 2006. An unusual Dicer-like1 protein fuels the RNA interference pathway in *Trypanosoma brucei*. *RNA* **12**: 2063–2072.

Shimogawa MM, Saada EA, Vashisht AA, Barshop WD, Wohlschlegel JA, Hill KL. 2015. Cell surface proteomics provides insight into stage-specific remodeling of the host-parasite interface in *Trypanosoma brucei*. *Mol Cell Proteomics* **14**: 1977–1988.

Silverman JS, Schwartz KJ, Hajduk SL, Bangs JD. 2011. Late endosomal Rab7 regulates lysosomal trafficking of endocytic but not biosynthetic cargo in *Trypanosoma brucei*. *Mol Microbiol* **82**: 664–678.

Silvester E, McWilliam K, Matthews K. 2017. The cytological events and molecular control of life cycle development of *Trypanosoma brucei* in the mammalian bloodstream. *Pathogens* **6**: 29.

Singer SM, Elmendorf HG, Conrad JT, Nash TE. 2001. Biological selection of variant-specific surface proteins in *Giardia lamblia*. *J Infect Dis* **183**: 119–124.

Sollelis L, Ghorbal M, Macpherson CR, Martins RM, Kuk N, Crobu L, Bastien P, Scherf A, Lopez-Rubio JJ, Sterkers Y. 2015. First efficient CRISPR-Cas9-mediated genome editing in *Leishmania* parasites. *Cell Microbiol* **17**: 1405–1412.

Sommer JM, Hua S, Li F, Gottesdiener KM, Wang CC. 1996. Cloning by functional complementation in *Trypanosoma brucei*. *Mol Biochem Parasitol* **76**: 83–89.

Sonnhammer ELL, von Heijne G, Krogh A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings Int Conf Intell Syst Mol Biol* **6**: 175–182.

Steverding D. 1998. Bloodstream forms of *Trypanosoma brucei* require only small amounts of iron for growth. *Parasitol Res* **84**: 59–62.

Steverding D. 2006. On the significance of host antibody response to the *Trypanosoma brucei* transferrin receptor during chronic infection. *Microbes Infect* **8**: 2777–2782.

Steverding D. 2003. The significance of transferrin receptor variation in *Trypanosoma brucei*. *Trends Parasitol* **19**: 125–127.

Steverding D. 2000. The transferrin receptor of *Trypanosoma brucei*. *Parasitol Int* **48**: 191–198.

Steverding D, Stierhof YD, Chaudhri M, Ligtenberg M, Schell D, Beck-Sickinger AG, Overath P. 1994. ESAG 6 and 7 products of *Trypanosoma brucei* form a transferrin binding protein complex. *Eur J Cell Biol* **64**: 78–87.

Steverding D, Stierhof YD, Fuchs H, Tauber R, Overath P. 1995. Transferrin-binding protein complex is the receptor for transferrin uptake in *Trypanosoma brucei*. *J Cell Biol* **131**: 1173–1182.

Sun J, Jiang H, Flores R, Wen J. 2010. Gene duplication in the genome of parasitic *Giardia lamblia*. *BMC Evol Biol* **10**: 49.

Szempruch AJ, Sykes SE, Kieft R, Dennison L, Becker AC, Gartrell A, Martin WJ, Nakayasu ES, Almeida IC, Hajduk SL, *et al.* 2016. Extracellular vesicles from *Trypanosoma brucei* mediate virulence factor transfer and cause host anemia. *Cell* **164**: 246–257.

Taylor MC, Kelly JM. 2010. Iron metabolism in trypanosomatids, and its crucial role in infection. *Parasitology* **137**: 899–917.

ten Asbroek AL, Ouellette M, Borst P. 1990. Targeted insertion of the neomycin phosphotransferase gene into the tubulin gene cluster of *Trypanosoma brucei*. *Nature* **348**: 174–175.

Tiengwe C, Bush PJ, Bangs JD. 2017. Controlling transferrin receptor trafficking with GPI-valence in bloodstream stage African trypanosomes. *PLoS Pathog* **13**: e1006366.

Tiengwe C, Muratore KA, Bangs JD. 2016. Surface proteins, ERAD and antigenic variation in *Trypanosoma brucei*. *Cell Microbiol* **18**: 1673–1688.

Uzureau P, Felu C, De Muylder G, Raldine, Pays E, Vanhamme L. 2007. G418, phleomycin and hygromycin selection of recombinant *Trypanosoma brucei* parasites refractory to long-term *in vitro* culture. *Mol Biochem Parasitol* **154**: 90–94.

Uzureau P, Uzureau S, Lecordier L, Fontaine F, Tebabi P, Homblé F, Grélard A, Zhendre V, Nolan DP, Lins L, *et al.* 2013. Mechanism of *Trypanosoma brucei gambiense* resistance to human serum. *Nature* **501**: 430–434.

Van Luenen HGAM, Kieft R, Mußmann R, Engstler M, ter Riet B, Borst P. 2005. Trypanosomes change their transferrin receptor expression to allow effective uptake of host transferrin. *Mol Microbiol* **58**: 151–165.

Vanhamme L, Paturiaux-Hanocq F, Poelvoorde P, Nolan DP, Lins L, Abbeele J Van Den, Pays A, Tebabi P, Xong HV, Jacquet A, *et al.* 2003. Apolipoprotein L-I is the trypanosome lytic factor of human serum. *Nature* **422**: 83–87.

Vanhamme L, Poelvoorde P, Pays A, Tebabi P, Xong HV, Pays E. 2000. Differential RNA elongation controls the variant surface glycoprotein gene expression sites of *Trypanosoma brucei*. *Mol Microbiol* **36**: 328–340.

Vanhollebeke B, De Muylder G, Nielsen MJ, Pays A, Tebabi P, Dieu M, Raes M, Moestrup SK, Pays E. 2008. A haptoglobin-hemoglobin receptor conveys innate immunity to *Trypanosoma brucei* in humans. *Science* **320**: 677–681.

Vassella E, Reuner B, Yutzy B, Boshart M. 1997. Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. *J Cell Sci* **110**: 2661–2671.

Vickerman K. 1969. On the surface coat and flagellar adhesion in trypanosomes. *J Cell Sci* **5**: 163–193.

Vink C, Rudenko G, Seifert HS. 2012. Microbial antigenic variation mediated by homologous DNA recombination. *FEMS Microbiol Rev* **36**: 917–948.

Wang X, Xiao YP, Bouchut A, Al-Khedery B, Wang H, Allreda DR. 2012. Characterization of the unusual bidirectional *ves* promoters driving VESA1 expression and associated with antigenic variation in *Babesia bovis*. *Eukaryot Cell* **11**: 260–269.

Wang Z, Morris JC, Drew ME, Englund PT. 2000. Inhibition of *Trypanosoma brucei* gene expression by RNA interference using an integratable vector with opposing T7 promoters. *J Biol Chem* **275**: 40174–40179.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191.

Wickstead B, Ersfeld K, Gull K. 2002. Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of *Trypanosoma brucei*. *Mol Biochem Parasitol* **125**: 211–216.

Wirtz E, Clayton C. 1995. Inducible gene expression in trypanosomes mediated by a prokaryotic repressor. *Science* **268**: 1179–1183.

Wirtz E, Leal S, Ochatt C, Cross GAM. 1999. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol Biochem Parasitol* **99**: 89–101.

Witola WH, Sarataphan N, Inoue N, Ohashi K, Onuma M. 2005. Genetic variability in ESAG6 genes among *Trypanosoma evansi* isolates and in comparison to other Trypanozoon members. *Acta Trop* **93**: 63–73.

Xiao Y-P, Al-Khedery B, Allred DR. 2010. The *Babesia bovis* VESA1 virulence factor subunit 1b is encoded by the 1β branch of the *ves* multigene family. *Mol Biochem Parasitol* **171**: 81–88.

Xong HV, Vanhamme L, Chamekh M, Chimfwembe CE, Van Den Abbeele J, Pays A, Van Meirvenne N, Hamers R, De Baetselier P, Pays E. 1998. A *VSG* expression site–associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. *Cell* **95**: 839–846.

Young R, Taylor JE, Kurioka A, Becker M, Louis EJ, Rudenko G. 2008. Isolation and analysis of the genetic diversity of repertoires of *VSG* expression site containing telomeres from *Trypanosoma brucei gambiense*, *T. b. brucei* and *T. equiperdum*. *BMC Genomics* **9**: 385.

Zhang W-W, Lypaczewski P, Matlashewski G. 2017. Optimized CRISPR-Cas9 genome editing for *Leishmania* and its use to target a multigene family, induce chromosomal translocation, and study DNA break repair mechanisms. *mSphere* **2**: e00340-16.

Zhang W-W, Matlashewski G. 2015. CRISPR-Cas9-mediated genome editing in *Leishmania donovani*. *MBio* **6**: e00861.

Ziegelbauer K, Multhaup G, Overath P. 1992. Molecular characterization of two invariant surface glycoproteins specific for the bloodstream stage of *Trypanosoma brucei*. *J Biol Chem* **267**: 10797–10803.

Zimmermann H, Subota I, Batram C, Kramer S, Janzen CJ, Jones NG, Engstler M. 2017. A quorum sensing-independent path to stumpy development in *Trypanosoma brucei*. *PLoS Pathog* **13**: e1006324.

Zupańska AK, Drummond PB, Swetnam DM, Al-Khedery B, Allred DR. 2009. Universal primers suitable to assess population dynamics reveal apparent mutually exclusive transcription of the *Babesia bovis ves1α* gene. *Mol Biochem Parasitol* **166**: 47–53.

VI. Appendix

Appendix 1.

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

⌘ *Author's Choice*

© 2015 by The American Society for Biochemistry and Molecular Biology, Inc.
This paper is available on line at http://www.mcponline.org

Architecture of a Host–Parasite Interface: Complex Targeting Mechanisms Revealed Through Proteomics*

Catarina Gadelha‡§, Wenzhu Zhang¶, James W. Chamberlain‡, Brian T. Chait¶, Bill Wickstead‡, and Mark C. Field‖**

Surface membrane organization and composition is key to cellular function, and membrane proteins serve many essential roles in endocytosis, secretion, and cell recognition. The surface of parasitic organisms, however, is a double-edged sword; this is the primary interface between parasites and their hosts, and those crucial cellular processes must be carried out while avoiding elimination by the host immune defenses. For extracellular African trypanosomes, the surface is partitioned such that all endo- and exocytosis is directed through a specific membrane region, the flagellar pocket, in which it is thought the majority of invariant surface proteins reside. However, very few of these proteins have been identified, severely limiting functional studies, and hampering the development of potential treatments. Here we used an integrated biochemical, proteomic and bioinformatic strategy to identify surface components of the human parasite *Trypanosoma brucei*. This surface proteome contains previously known flagellar pocket proteins as well as multiple novel components, and is significantly enriched in proteins that are essential for parasite survival. Molecules with receptor-like properties are almost exclusively parasite-specific, whereas transporter-like proteins are conserved in model organisms. Validation shows that the majority of surface proteome constituents are *bona fide* surface-associated proteins and, as expected, most present at the flagellar pocket. Moreover, the largest systematic analysis of trypanosome surface molecules to date provides evidence that the cell surface is compartmentalized into three distinct domains with free diffusion of molecules in each, but selective, asymmetric traffic between. This work provides a paradigm for the compartmentalization of a cell surface and a resource for its analysis. ***Molecular & Cellular Proteomics 14: 10.1074/mcp.M114.047647, 1911–1926, 2015.***

The cell surface is the major point of interaction between unicellular parasites and their surroundings, and is the site for many essential functions such as nutrient uptake, host recognition, and environment sensing. This interface, however, also represents the primary target for host immune attack. To evade adaptive immune defenses, many pathogens (including the causative agents of malaria, Lyme disease and AIDS) use some form of antigenic variation - the expression of a series of immunologically distinct surface proteins (1, 2)). As an exclusively extracellular parasite of the blood, African trypanosomes have made a huge investment in this strategy. In the human-infective species *Trypanosoma brucei*, around ten million copies of a single variant surface glycoprotein (VSG)[1] form a dense surface coat that protects the parasite against complement-mediated lysis. Periodic switching of the single expressed VSG gene from a vast silent library enables trypanosomes to avoid clearance by the host's adaptive immune response, prolonging infection and increasing the chances of transmission. The monoallelic expression of *VSG* is achieved through tight regulation from telomeric expression sites (ES), with only one from about 20 ES being transcriptionally active at any one time.

For the strategy of antigenic variation to work, the African trypanosome surface coat must be kept free of many essential invariant antigens that might otherwise elicit an immune response. Most of these are thought to be sequestered within a specialized region of the surface membrane at the base of the flagellum called the flagellar pocket (FP). This relatively small membrane domain is the sole site for all endocytosis and exocytosis performed by African trypanosomes, and has

From the ‡School of Life Sciences, University of Nottingham, Nottingham, UK, NG2 7UH; §Department of Pathology, University of Cambridge, Cambridge, UK, CB2 1QP; ¶Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, 10021; ‖Division of Biological Chemistry and Drug Discovery, University of Dundee, Dundee, UK, DD1 5EH

Received December 18, 2014, and in revised form, April 22, 2015
Published, MCP Papers in Press, April 30, 2015, DOI 10.1074/mcp.M114.047647

⌘ *Author's Choice*—Final version free via Creative Commons CC-BY license.

Author contributions: C.G. conceived the study; C.G., B.T.C., and M.C.F. designed research; C.G., W.Z., and J.W.C. performed research; B.T.C. and B.W. contributed new reagents or analytic tools; C.G., W.Z., J.W.C., B.T.C., B.W., and M.C.F. analyzed data; C.G. and M.C.F. wrote the paper; B.T.C. and B.W. revised manuscript.

[1] The abbreviations used are: VSG, variant surface glycoprotein; FP, flagellar pocket; ESP, enriched surface-labeled protein; UTR, untranslated region; TM, transmembrane; GPI, glycosylphosphatidylinositol.

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

the highest rate of endocytosis for any system thus far observed (3, 4). Thus, the FP is a crucial interface between the parasite and host. Unsurprising, disruption of FP function by loss of the associated cytoskeleton or endocytic vesicular traffic is lethal (5–7), highlighting the potential of this host–parasite interface as a therapeutic target.

Our current understanding of FP function and its possible exploitation for therapeutic gain have been significantly inhibited by the paucity of data on its molecular composition. Mining the parasite genome for genes encoding simple characteristics of membrane association is of limited predictive power, as a large proportion of predicted membrane proteins unlikely to be on the cell surface, and in-silico generated datasets are often not amenable to validation studies (for example, the parasite genome is predicted to encode 257 GPI-anchored proteins, 1963 transmembrane proteins, and over 7000 potentially glycosylated proteins, from a pool of only 9202 predicted proteins (genedb.org, v4)). Attempts to purify specific FP components, however, have been hampered by technical difficulties in isolation and thus far none have succeeded in providing validated, high-confidence datasets (8–11).

Only a few validated FP constituents are known to date. The first reported was the heterodimeric transferrin receptor (12) encoded by the ES associated genes 6 and 7 (*ESAG6* and *ESAG7*). Since then, only four more proteins have been specifically localized to the FP of bloodstream-form stages: the haptoglobin–hemoglobin receptor (13), an aquaporin (14), a hypothetical protein identified by proteomics of flagellar fractions (15), and a protein associated with differentiation (16). These components likely represent only a tiny subset of the FP proteome. Here, we address this knowledge gap using a comparative, semi-quantitative approach for the high-confidence identification of cell surface proteins in bloodstream-form *Trypanosoma brucei*. By creating a new genetic toolkit for endogenous locus tagging of membrane proteins, we validate our proteomic set by localization of 25 putative surface molecules of unknown function. As well as demonstrating the location for many novel FP components, we show that individual proteins access different combinations of cell surface membrane domains, and present a bioinformatics analysis of sorting signals. From these data, we propose a new model for the domain organization of the *T. brucei* surface.

EXPERIMENTAL PROCEDURES

Isolation of Surface Membrane Proteins—We used bloodstream-form *Trypanosoma brucei* Lister 427 expressing VSG221 (BES1/MI-Tat 1.2/VSG427–2/TAR 40), as monitored by immunofluorescence microscopy using an affinity-purified polyclonal antibody anti-VSG221. 5×10^8 mid-log phase cells were harvested by centrifugation and resuspended at 2×10^8 cells ml^{-1} in PBS (10 mM PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.5) plus 20 mM glucose. Cells were held on ice while pulsed with 500 μM fluorescein-hexanoate-NHS (referred hereafter to as fluorescein) dissolved in DMSO and HPG buffer (20 mM HEPES pH 7.5, 140 mM NaCl, 20 mM glucose). Pulse duration was 15 min on ice, during which time cells remained actively motile and morphologically normal (as assessed by light microscopy). Fluorescence microscopy showed fluorescein to be exclusively associated to the parasite cell surface (Fig. 1*B*). At the end of this period, unreacted fluorescein was blocked by the addition of TBS (25 mM Tris-HCl pH 7.5, 150 mM NaCl) plus 0.25% w/v glycine, and removed by washing cells in TBS plus 20 mM glucose. Fluorescein-labeled cells were lysed with 2% v/v Igepal CA-630 and 2% w/v CHAPS in the presence of protease inhibitors (5 μM E-64d, 2 mM 1,10-phenanthroline, 50 μM leupeptin, 7.5 μM pepstatin A, 500 μM phenylmethylsulfonyl fluoride, 1 mM EDTA, 1 mM DTT) and 200 μg ml^{-1} DNase I, and centrifuged at $20{,}000 \times g$ for 30 min to separate soluble labeled proteins from the insoluble fraction. To increase identification sensitivity toward less abundant surface membrane proteins, we included a VSG-depletion step by affinity chromatography, for which a polyclonal antibody anti-VSG221 was generated (please see below). The soluble fraction was allowed to bind to 8 mg polyclonal antibody anti-VSG221 conjugated to protein G-Sepharose 4 fast flow (GE Healthcare). Then the soluble fraction partially depleted of VSG was allowed to bind to 30 mg protein G-Dynabeads (Invitrogen) cross-linked to 400 μg polyclonal antibody anti-fluorescein for 1 h, after which period unbound material was collected as flow-through and beads were washed several times in the presence of high salt and detergent (500 mM NaCl, 0.02% v/v Tween-20). Bound proteins were deglycosylated native *on column* with 1000U of PNGase F for 1 h before acid then basic elutions in 0.2 M glycine pH 2.5 and 0.2 M triethanolamine pH 11 respectively. To control for nonspecific binding to anti-fluorescein column, a parallel isolation was carried out with unlabeled cells. To account for possible cell lysis during the surface labeling step, 1×10^8 cells were subjected to hypotonic lysis by resuspension in 20 mM Hepes pH 7.5 in the presence of the protease inhibitors aforementioned for 30 min at room temperature, and then pulsed with fluorescein as above.

Mass Spectrometry—Proteins in the final eluate were precipitated with cold acetone, solubilized in Laemmli buffer, and treated with 1 M iodoacetamide to alkylate reduced cysteines. Proteins were resolved by SDS-PAGE using pre-cast gels and standard techniques. Post-electrophoresis gels were stained with SyproRuby (Life Technologies) for imaging, or Coomassie blue for band excision. Mass spectrometry analysis of proteins that were digested in-gel was performed on an LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Scientific). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD002180.

Label-free quantitation of mass spectrometry results—mzXML data files were uploaded onto the Central Proteomics Facilities Pipeline (release version 2.1.1; www.proteomics.ox.ac.uk), which uses Mascot, X!Tandem and OMSSA search engines. The data were searched with the following peptide modifications: fluorescein (K), acetylation (protein N terminus), carbamidomethylation (C), oxidation (M), and deamidation (N/Q). Note that we expect the majority of peptides, even those derived from fluoresceinated proteins, not to contain the fluorescein modification, as only a few lysines in any given surface protein would be accessible. Precursor mass tolerance was set at 20 ppm, MS/MS fragment ion tolerance at 0.5 Da, and number of missed cleavages permitted at 2. Searches were performed against a custom, non-redundant trypanosome protein sequence database combining predicted protein sequences from TREU927 and Lister 427 genomic data (tritrypdb.org), with the inclusion of ES and VSG sequences (17, 18), and containing in total 20,195 entries. The resulting peptide identifications from each search engine were validated with PEPTIDEPROPHET and PROTEINPROPHET and lists compiled at the peptide and protein level. IPROPHET was used to combine the iden-

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

tifications from three search engines and further refine identifications and probabilities. Normalized spectral index quantitation (SINQ) was applied to the grouped meta-searches to give protein-level quantitation between labeled samples and controls (19). All lists of peptide and protein identifications were generated with a probability cut-off corresponding to 1% false discovery rate (FDR) relative to a target decoy database. Only proteins identified with 2 or more spectra were considered for further analysis.

Bioinformatics—Signal peptide and anchor sequences were predicted from the first 70 aa of each coding sequence by a stand-alone implementation of SignalP v3.0b (20, 21) using the hidden Markov model methodology, "eukaryotic" settings and thresholds of $p \geq 0.9$. For GPI-anchor prediction, to reduce false positives, proteins were considered only if they were a PredGPI hit (22) with false-positive rate ≤0.1 and also had SignalP peptide prediction with $p \geq 0.7$ (because only proteins directed to the endoplasmic reticulum are processed for anchor addition). Transmembrane domains were predicted using TMHMM v2.0c (23, 24).

Generation of a Genetic Toolkit for Membrane Protein Localization—A vector for specific tagging of GPI-anchored protein genes, named pSiG, was created by *de novo* synthesis (MrGene, Invitrogen). pSiG contains an epitope tag and fluorescent protein flanked by a signal peptide and GPI-anchor insertion sequences (derived from VSG221) up- and downstream respectively. A derivative for tagging of transmembrane protein genes, pSiS, was created by replacing the GPI-anchor insertion sequence from pSiG with a stop codon generated by annealing two primers. In these vectors, part of the targeted ORF and its UTR, at either the N- or C terminus, is cloned in frame with the epitope tag/fluorescent protein, then the plasmid is linearized for transfection and replacement of the endogenous gene fragments. Hence, the sites for targeting the specific locus are supplied by the user along with the site for linearization. The constructs contain convenient restriction sites on either side of the fluorescent protein/epitope tag for integration of short targeting sequences. Derivatives include nine different fluorescent proteins, two epitope tags and three selection markers. These vectors are available from the authors upon request, and their DNA sequences can be found on the authors' webpage (www.catarinagadelha.com/resources).

Endogenous-locus Tagging—ESPs and ESAGs predicted to encode transmembrane proteins were tagged at the C terminus, whereas those predicted to contain a GPI anchor were tagged at the N terminus (because of lack of robustness of prediction algorithms). For N-terminal tagging, PCR amplicons containing ~200bp from the 5′-end UTR (untranslated region) and ~200bp from the N-terminal end of the CDS (coding sequence) of interest were cloned together into the XbaI-BamHI sites downstream of the fluorescent protein ORF in pSiG, such that the N-terminal end of the CDS was in frame with the fluorescent protein. For C terminus tagging, PCR amplicons containing ~200bp from the 3′-end UTR and ~200bp from the C-terminal end of the CDS of interest were cloned together into the HindIII-AvrII sites upstream of the epitope tag sequence in pSiS, such that the C-terminal end of the CDS was in frame with the fluorescent protein. In the same step, a NotI linearization site was introduced between the UTR and CDS. Integration of these constructs at the targeted endogenous locus results in transgenic lines in which one allele of the CDS of interest contains fluorescent protein at its N/C terminus, but both 5′- and 3′-UTRs are identical to untagged copy. Vectors (~10 μg) were linearized by digestion with NotI restriction endonuclease and transfected into single-marker bloodstream form *T. brucei* (25) using an Amaxa Nucleofector 2b device, followed by selection of stable transformants with 5 μg/ml hygromycin. Correct integration was assessed by diagnostic PCR from genomic DNA of clonal transformants (not shown) and also immunoblotting of cell lysates separated by SDS-PAGE against a mixture of two anti-GFP monoclonals (7.1 and 13.1; Roche) at 800 ng/ml in 1% w/v skimmed milk in TBS, followed by 80 ng/ml horseradish peroxidase-conjugated goat anti-mouse immunoglobulins.

Analysis of Integration into VSG221 Expression Site—Whole-chromosome-sized DNAs were prepared as described elsewhere (26). Agarose-embedded DNA was digested with SmiI endonuclease and subjected to pulsed-field gel electrophoresis in a contour-clamped homogeneous electric field electrophoresis apparatus (CHEF-DR III; Bio-Rad), loading DNA from 1.7×10^7 cells per lane. DNA separation was performed in 1% agarose in TB(0.1)E (90 mM Tris borate, 0.2 mM EDTA, pH 8.2) held at 14 °C for 20 h at 5.2 V/cm with switching time ramped linearly 2–10 s and an included angle of 120°. DNA gels were stained in ethidium bromide and prepared for transfer by UV nicking (80 mJ, 250 nm UV) followed by equilibration in 0.4 M NaOH, 1.5 M NaCl and then transferred to positively charged nylon membrane by capillary transfer in the same solution. After transfer, membranes were neutralized with 0.5 M Tris-HCl (pH 7) and cross-linked (120 mJ, 250 nm UV). Fluorescein-labeled probes were generated by random priming from unlabeled *GFP*, *HYG* and *VSG221* coding sequences. Denatured template DNA (100 ng) were incubated for 5 h at 37 °C with 0.1 mM dATP, dCTP, dGTP, 0.67 mM dTTP, 0.33 mM Fluorescein-dUTP, 2 μM random heptamers and 5 U Klenow fragment. Hybridization was performed overnight in 1% w/v SDS, 5% w/v dextran sulfate, 10% v/v blocking solution (Roche), 750 mM NaCl, 75 mM sodium citrate (pH 7) at 60 °C. Blots were washed to a stringency of 0.1% SDS w/v, 30 mM NaCl, 3 mM sodium citrate (pH 7) at 62 °C. Hybridized probe was detected with anti-fluorescein alkaline phosphatase-conjugated antibody and chemiluminescence. For reprobing, membranes were stripped with hot 0.3% w/v SDS plus 0.3 M NaOH.

Protein localization—For analysis of localization of tagged proteins by native fluorescence, cells were harvested from mid-log phase cultures, washed twice in PBS plus 20 mM glucose, allowed to adhere onto derivatized glass slides for 2 min (at density of 2×10^7 cells/ml), fixed for 10 min in 2.5% w/v formaldehyde, counter-stained with 5 μg/ml concanavalin A (ConA; it binds to α-D-mannose and α-D-glucose moieties associated to VSG and possibly other surface proteins) conjugated to tetramethylrhodamine isothiocyanate (TRITC) for 20 min, and mounted in a solution containing DAPI and a photostabilizing agent (1% w/v 1,4-Diazabicyclo(2.2.2)octane, 90% v/v glycerol, 50 mM sodium phosphate pH 8.0, 0.25 mg/ml 4′,6-diamidino-2-phenylindole).

Generation and Purification of Polyclonal Antiserum—A fragment encoding residues 27–384 of VSG221 (Tb427.BES40.22) was amplified by PCR from *T. brucei* Lister 427 genomic DNA and cloned in frame into the bacterial expression vector pQE-30 (Qiagen) to allow expression of the coding sequence fragment fused to an N-terminal 6xHis tag. Expression of recombinant protein was induced in M15(Rep4) *Escherichia coli* (Qiagen) and protein was subsequently isolated from cleared, sonicated bacterial lysates by nickel-affinity chromatography by standard methods. 200 μg of recombinant VSG221 was used as immunogen in rabbits. Reactive antiserum was purified by binding to recombinant protein coupled to CNBr-activated Sepharose beads, washed extensively with PBS and eluted with 0.2 M glycine pH 2.5 followed by 0.2 M triethanolamine pH 11. Affinity-purified polyclonal antibodies were dialyzed against PBS and concentrated by ultrafiltration.

Immunoblotting of Surface Membrane Protein Isolation Fractions—Immunoblots to test the purification procedure (Fig. 1*C*) were performed with the following polyclonal antisera: anti-ISG65 (kind gift from Mark Carrington, University of Cambridge, UK), anti-TfR (Piet Borst, The Netherlands Cancer Institute, Netherlands), anti-p67 and anti-BiP (James Bangs, University at Buffalo (SUNY).

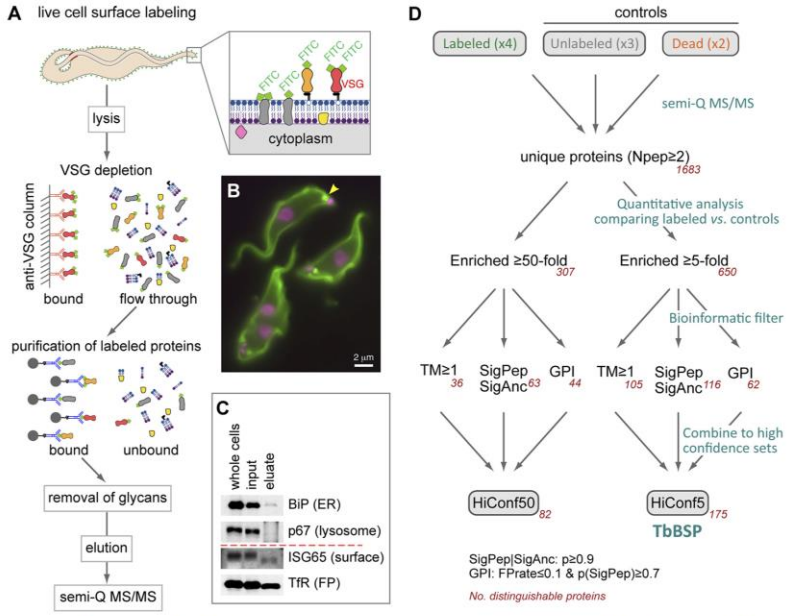


FIG. 1. **Workflow of biochemical, semi-quantitative mass spectrometry and bioinformatic methods used to identify putative cell surface proteins.** *A*, Scheme illustrating key steps in purification. *B*, Micrograph of cells following chemical modification with fluorescein (live at 0 °C). Native fluorescence at plasma membrane is predominantly derived from fluoresceinated VSG (which makes up ~90% of proteins at the parasite surface). DNA has been counter-stained with DAPI (magenta); the FP is indicated by yellow arrowhead. *C*, Immunoblots showing isolation of known surface proteins (ISG65, found on the cell surface and TfR, found in the FP) in the final purified eluate. Note faster migration of deglycosylated ISG and TfR in eluate. Common contaminants from the ER (BiP) and lysosome (p67) are highly depleted in final eluate. *D*, Schematic showing enrichment analysis (for exclusion of contaminants by comparison of labeled samples with controls) and bioinformatic filters (for prediction of membrane proteins features) applied to protein identification to produce "high-confidence" sets. The numbers of unique proteins present in each set are shown in red. The high-confidence set of 175 putative surface membrane proteins enriched 5x in labeled samples is herein referred to as the *T. brucei* bloodstream surface proteome (TbBSP). Experimental replicates of protein isolation from fluorescein-labeled live cells ("Labeled"), unlabeled cells ("Unlabeled"), and fluorescein-labeled material from lysed cells ("Dead") are indicated between brackets. See Experimental Procedures for details of protein feature prediction, and supplemental Table S2 for bioinformatics filter abbreviations.

RESULTS

Chemical Modification of the Cell Surface—A mechanistic understanding of the interface between African trypanosomes and their mammalian host requires the identification and characterization of the FP molecular composition. As the FP membrane is contiguous with the membranes of both the cell body and the flagellum, it is extremely challenging to isolate pocket proteins through classical cell fractionation procedures. To address this problem, we devised a workflow to specifically isolate cell surface proteins and generate a validated dataset of bloodstream-form cell surface constituents of *Trypanosoma brucei*. Our strategy is summarized in Fig. 1*A* and starts with the chemical modification (fluoresceination) of the surface of live cells held at low temperature (0 °C). Under these conditions, recycling of the surface coat and endocytosis are blocked, but chemical tags are still able to access proteins at both the plasma membrane (~90% of which being VSG (27, 28)) and also the FP lumen (Fig. 1*B*). Labeled cells were then solubilized and fluoresceinated surface proteins purified by affinity chromatography. The purification method was optimized by a VSG depletion step to increase sensitivity of detection of less abundant surface proteins (Fig. 1*A* and supplemental Fig. S1), and on-column enzymatic removal of N-glycans to improve mass spectrometry identification of glycosylated surface proteins (Fig. 1*A* and supplemental Fig. S1). Finally, to allow more efficient solubilization of membrane proteins and increase dynamic range, the sample was resolved by SDS-PAGE (supplemental Fig. S1), and gel regions

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

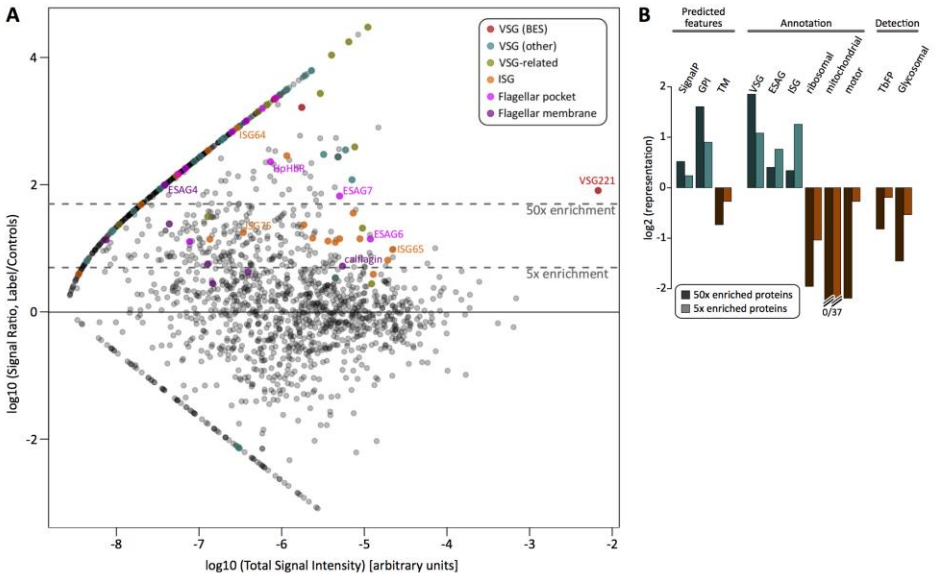


FIG. 2. **Identification of surface proteins by comparative label-free semi-quantitative mass spectrometry.** *A*, Enrichment analysis for 1683 unique proteins (integrated spectral intensity) in labeled samples *versus* unlabeled and osmotically lysed controls (see Experimental Procedures for more information). Points represent log10-transforms of total intensity (all replicates, samples and controls) against the ratio of intensity in samples *versus* summed controls. Points representing signal from VSGs (BES copies or from elsewhere in the genome), VSG-related proteins, ISGs, and proteins previously localized to the FP or flagellar membrane are highlighted. *B*, Representation of proteins with select predicted features (SignalP peptide prediction $p \geq 0.9$; PredGPI false-positive rate ≤ 0.1; ≥ 1 predicted TM domain), annotation (word match in description) or those detected in either the *T. brucei* flagellar proteome (30) or glycosomal proteome (31). Representation is the ratio of the number of hits in enriched sets *versus* all uniquely detected proteins.

subjected to tandem mass spectrometry (GeLC-MS-MS). The final eluate was enriched in an invariant surface glycoprotein (ISG65) that localizes to the cell surface, and the low-abundance transferrin receptor (ESAG6 subunit) which is found in the FP (Fig. 1*C*). High-abundance markers of internal compartments, specifically the abundant luminal ER chaperone BiP and the LAMP-like lysosomal protein p67, were either greatly reduced or undetectable (Fig. 1*C*).

Semi-quantitative Comparative Mass Spectrometry Defines a Trypanosoma brucei Surface Proteome—Our surface protein preparation is anticipated to contain many FP proteins as well as those localized more generally to the cell surface and early/recycling endosomes. Many FP components, however, are expected to be present at only tens or hundreds of copies per cell, as seen for the haptoglobin-hemoglobin receptor (13). This necessitates highly sensitive detection, but also the exclusion of inevitable contaminating proteins. To identify proteins specifically enriched in our surface protein preparation, we used a label-free semi-quantitative mass spectrometry approach against two controls: (1) to account for non-specific binding to affinity chromatography columns, we carried out parallel isolations with unlabeled cells; and (2) to account for cell lysis during the chemical modification, in which the fluorescein tag would access internal proteins as well as those at the surface, a further control was made by labeling hypotonically lysed cells. We then compared the integrated spectral intensities from mass spectrometry of material isolated from labeled *versus* control preparations, allowing for removal of contaminants through testing for signal enrichment in the labeled sample (Figs. 1*D*).

Across all preparations and replicates, we detected 1683 uniquely distinguishable proteins (each being represented by two or more detectable peptides). The full list of hits and their respective integrated spectral intensities is provided in supplemental Table S1. The most abundant protein in bloodstream-form *T. brucei* cells is VSG (27) and, as expected, VSG221 (MITat 1.2/VSG427–2) expressed from the active ES is detected in all preparations. However, its signal is highly enriched (80×) in labeled samples *versus* controls (Fig. 2*A*), despite being deliberately depleted in these preparations (Fig.

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

1*A* and supplemental Fig. S1). We also observed, ~10^3 times less abundantly, several other VSGs including those in other telomeric ESs, likely representing rare cells in the parasite population which have undergone switch events. Along with these "true" VSGs, a number of VSG-related proteins (transcribed from chromosome internal locations (29)) are also enriched in the preparation, representing the first evidence that this family of proteins is translated in bloodstream form parasites and that they are surface associated.

Our procedure enriches for VSG, several ISGs and proteins known to be localized specifically to the cell surface membrane (Fig. 2*A*). Importantly, low abundance FP proteins, such as HpHbR are also detected in these experiments and are highly enriched (250x) in the labeled preparation. Analysis of the features or annotations of enriched proteins compared with all those detected showed a substantial over-representation of those with predicted signal peptide or glycosylphosphatidylinositol (GPI) anchor, as well those with annotations for "VSG," "ESAG," or "ISG" (Fig. 2*B*). Conversely, annotations associated with ribosomes, mitochondrion or cytoskeleton motors are under-represented in the enriched cohort (Fig. 2*B*), as are proteins detected as part of the *T. brucei* flagellar proteome (30) or glycosomal proteome (31). Interestingly, we also find an under-representation of proteins with predicted transmembrane (TM) domains in labeled preparations (Fig. 2*B*). This may suggest that a significant fraction of TM proteins in trypanosomes are expected to be associated with internal membranes, although this may also reflect less efficient chemical modification of multipass proteins with few extracellular lysines (see "Discussion").

These data show that known surface, FP and flagellum membrane proteins are substantially enriched in our chemically modified preparations. Using enrichment analysis, 307 and 650 uniquely distinguishable protein hits were identified with 50- or 5-fold enrichment in the labeled sample when compared with controls (Fig. 1*D*). However, these sets are unlikely to represent only genuine membrane-associated proteins. To further improve discrimination between true surface proteins and contaminants, we applied a bioinformatic filter to create sets representing only those proteins with predicted signal peptide or signal anchor sequence, GPI-anchor addition sites or TM domains. This is equivalent to intersecting our enrichment data sets with bioinformatic prediction of membrane-association as used by Jackson *et al.* (32), and constitute "high-confidence" sets that have support from both methods. This procedure is also analogous to the approach used to analyze the trypanosome nuclear envelope and identify nuclear pore complex components (33), and here identified 82 or 175 uniquely distinguishable putative surface proteins at 50× or 5× enrichment thresholds respectively (Fig. 1*D*). The full list of these sets is given in supplemental Table S2. These sets represent hits with a high likelihood of being genuine surface proteins, and identified proteins include known FP components, VSGs and ISGs, as well as proteins with predicted function as transporters (Tb427.04.4830, Tb427tmp.02.0630, Tb427.03.4630, Tb427.08.2380, Tb427.08.3620, Tb427.04.4860, Tb427.08.650, Tb427.08.2160), permeases (Tb427.05.3390) and channels (Tb427.10.11680). We herein refer to the 5x-enriched, high-confidence set of 175 putative surface membrane proteins as the *T. brucei* bloodsteam surface proteome (TbBSP).

Most TbBSP Proteins are True Parasite Cell Surface Components—Having demonstrated an efficient enrichment of known FP proteins and related annotation in the labeled dataset, we next sought to robustly test our TbBSP by directly interrogating the cellular location of multiple protein hits of unknown localization, and looking for specific signal at the FP. A set of 25 candidates were selected from the high-confidence sets for further characterization using the following criteria: (1) they were annotated as "hypothetical" proteins for which no functional data had been previously reported for *T. brucei* at the start of this work; (2) they represented the range of general protein topologies detected, *e.g.* predicted GPI-anchored proteins, type I and type II TM proteins, and multipass TM proteins; and (3) they included proteins with enrichment ranging from 5 to >6000 times and spanning >3 orders of magnitude of mass spectrometry signal intensity. Fig. 3 shows the enrichment and architectures of these candidates, and supplemental Table S3 provides their accession numbers and predicted features.

Chimeric proteins were created by integration of tagging constructs at endogenous gene loci. Tagging cell surface proteins is potentially complicated by requirements for signaling sequences at both amino and carboxyl termini, and issues with folding of fluorescent proteins targeted through the ER. To overcome these problems, we created two new series of vectors specifically designed for the endogenous-locus tagging of genes encoding GPI-anchored and non-GPI-anchored sequences containing N-terminal signal sequences. These vectors are called the pSiG and pSiS series, respectively (supplemental Fig. S2) and include the incorporation of a "superfolder" GFP (or derivatives) with improved folding dynamics and greater resistance to the reducing environments encountered in the ER lumen or extracellular space compared with conventional GFP variants (34), plus an epitope tag (HA). The pSiG/pSiS series also include processing signals (trypanosome signal peptide or GPI-anchor addition sequences), providing a means to rapidly and accurately tag surface proteins at either N- or C terminus (supplemental Fig. S2). These vectors provide the correct FP localization of previously analyzed proteins, for example, either GPI-anchored or non-anchored subunits of the transferrin receptor (see Fig. 7). Moreover, the toolkit does not force a non-TbBSP protein (ESAG1) onto the cell surface (supplemental Fig. S2).

The 25 selected genes encoding candidate surface-associated proteins, designated as enriched in surface-labeled proteome (ESP) proteins (ESP1–25), were tagged at their endogenous loci using the vectors described above. Correct

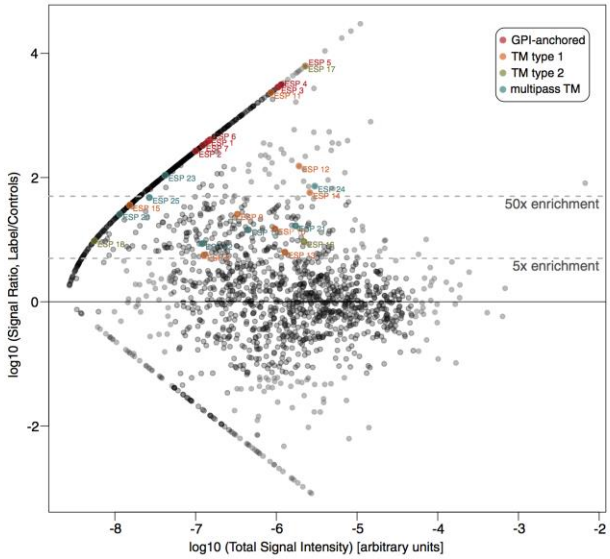


FIG. 3. **Enrichment analysis showing proteins of unknown function taken for validation by localization.** Dataset as in Fig. 2, with points representing 25 ESPs highlighted according to predicted protein architecture.

integration of the tagging construct and expression of fusion proteins was assessed with immunoblotting of whole-cell extracts (supplemental Fig. 3A). Because genes are tagged by integration at the endogenous loci, it is expected that protein expression levels will be close to those for wild-type protein and, consistent with this, different fusion proteins were expressed at different levels. Two tagged proteins (ESP4 and ESP7) did not show a detectable signal on Western blots, and were not pursued further.

For the 23 fusion proteins with detectable expression, 12 were clearly present at the FP membrane as assessed by native fluorescence (Fig. 4 and supplemental Fig. S4). These 12 proteins localized either exclusively to the FP (ESP1, 6, 10 and 11) or in addition to another surface domain - for example, five proteins localized to the FP and endosomal system (ESP12, 14, 19, 21, and 22), whereas ESP8 localized to the FP and the junction between the cell body and the flagellum membranes (the flagellum attachment zone). In addition to these 12 FP proteins, ESP13 and 24 were present across the entire cell surface (FP, flagellum and cell body) and a further four ESPs were predominantly localized to endosomes (ESP5, 9, 15, 20). This is expected, because the endosomal membrane is in constant flux with the cell surface and proteins with clear FP function, such are TfR, maintain a steady-state concentration in early/recycling endosomal compartments (35, 36). Likewise, ISGs are equally distributed between endosomes and FP/cell surface (37). Therefore, these four predominantly endosomal ESPs are likely to be transiently present at the FP, albeit at low abundance, and are thus enriched in our chemical modification procedure. ESP17 and ESP18 were found at both the cell body membrane and an intracellular compartment tentatively interpreted as the lysosome. The remaining five proteins (ESP2, 3, 16, 23, 25) localized elsewhere in the cell and may represent contaminants, although mislocalization because of tagging cannot be excluded (supplemental Fig. S4). Overall, experimental validation by cellular localization of 23 ESPs shows that we have identified 18 novel membrane proteins on the parasite cell surface, the majority of which reside at the FP (exclusively or in combination with another surface membrane domain).

Diversification of Parasite Surface Architecture—ESPs at the FP may represent promising therapeutic targets because of their exposure and potential roles in modulating essential parasitic processes, but only if those proteins are sufficiently different to host ones. To map the evolutionary distribution of ESPs, we asked if orthologs could be detected in organisms representing a wide taxonomic diversity of eukaryotes, including humans, and for which complete or near-complete genome sequences were publicly available. Phylogenetic analysis show that most ESPs are specific to African trypanosomes and closely related parasites (Fig. 5). This provides

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

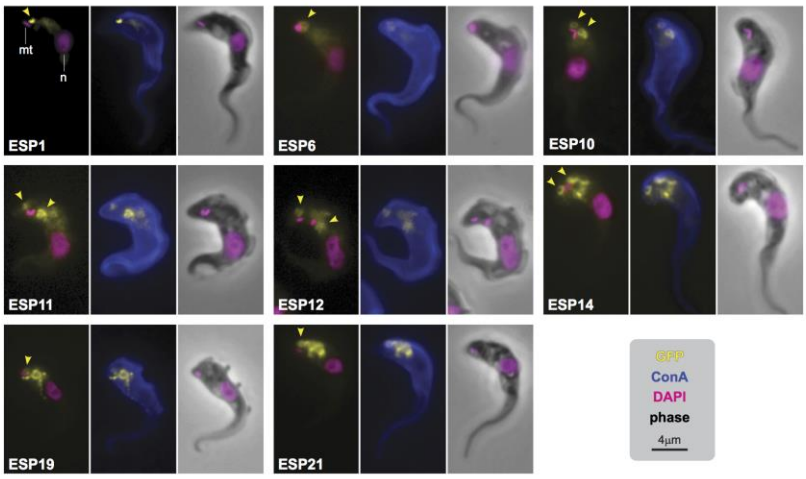


FIG. 4. **Localization of surface proteome components at the FP.** ESPs were localized by tagging the gene at the endogenous locus with an ORF encoding superfolder-GFP. Images are representative of the signal distribution observed for each cell line. Yellow: native fluorescence from superfolder-GFP; Blue: concanavalin A counterstain (ConA); Magenta: DAPI. Nuclear (n) and mitochodrial (mt) DNA contents, and FP (yellow arrowhead) are indicated.

evidence for a lineage-specific architecture for the surface membrane of kinetoplastid cells, reflecting their shared ancestry and biological similarities. Striking, however, was the finding that ESPs predicted to be GPI-anchored are often restricted to *T. brucei*, whereas type I and II TM proteins tend to be conserved in all kinetoplastids (both intra- and extracellular parasites) (Fig. 5). This distribution suggests specific protein evolution to match distinct selective pressures encountered by these parasites, such as mechanisms of survival, host immune invasion and transmission. In contrast, many of the multipass TM ESPs are from families conserved right across eukaryotes (Fig. 5) and, thus, may have arose early in eukaryotic evolution. This likely reflects the expected hierarchy of conservation, with essential transporters being more evolutionarily constrained.

Nine out of 12 ESAGs Encode Surface Associated Proteins—The first FP component identified was the TfR previously mentioned, encoded by the expression site-associated genes (*ESAG*s) 6 and 7. There are 12 distinct families of *ESAG*s (*ESAG*1 to 12) that are co-transcribed with the active VSG gene from one of ~20 telomeric expression sites (ESs). Some or all *ESAG*s may be present in a particular ES (17), and most have chromosome-internal paralogs known as genes-related to *ESAG* (*GRESAG*s). Only a few other ESAGs have been characterized in detail in *T. brucei*: ESAG8 is a protein of unknown function that has been localized to the nucleus (38, 39), whereas ESAG4 is an adenylate cyclase localized to the flagellum membrane (40), and whose activity has been associated with control of parasitemia (41). GRESAG9 is specifically expressed and secreted by the quiescent "stumpy" bloodstream-form stage (42). Finally, an ESAG specific to the subspecies *T. b. rhodesiense* - the serum resistance associated gene, or SRA - confers resistance to a trypanolytic factor associated with the heavy density lipoprotein found in normal human serum (43–46).

Given that ESAGs are co-expressed with the active VSG during infection, they are believed to play roles in parasite survival in the human host. All but two ESAGs (ESAG8 and ESAG12) are predicted to encode a signal peptide sequence, a GPI-anchor insertion site, or a TM domain, suggesting that they may be associated to the surface membrane or secreted proteins, but for the majority this has not been tested. Significantly, seven ESAGs (ESAG2, 4, 5, 6, 7, 10 and 11) are present in our high confidence datasets (Fig. 6*B*), but the remainder were not. We took this finding, and the genetic tools developed here, as an opportunity to investigate the cellular localization of all ESAGs and to test our surface proteome for false negatives (*i.e.* true surface proteins not detected in our set). We tagged every *ESAG* present in the active ES of bloodstream-form trypanosomes used in this study (BES1, Fig. 6*A*). Only pseudogenes of *ESAG5* and *ESAG11* are present in this ES, and *ESAG9* and *ESAG10* are absent (17). Because chromosome-internal copies of ESAG5, 10 and 11 were highly enriched in our surface proteome, these were also targeted for protein fusions. GRESAG9 has previously

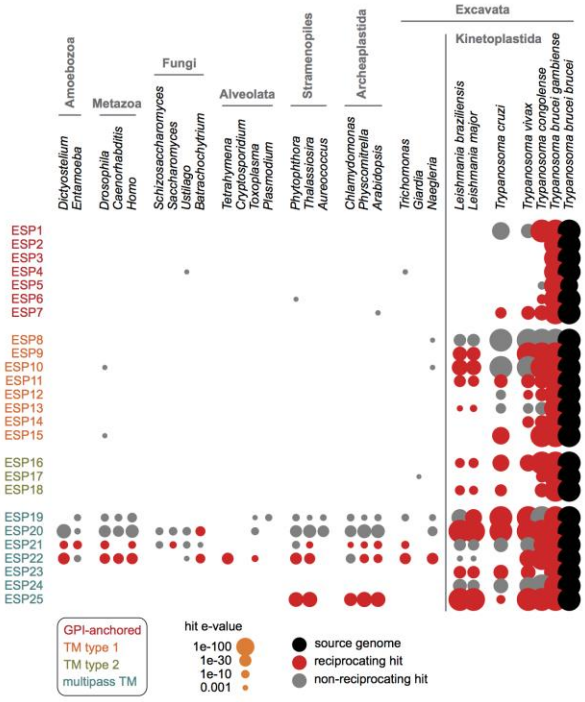


FIG. 5. **The distribution of ESPs across eukaryotes.** Conservation was investigated by analysis of BLAST hits in the predicted proteomes of model species from a wide range of eukaryotic lineages. Spot size represents the strength of BLAST hit (e-value). Red shows reciprocal best-BLAST hits between genomes; gray shows non-reciprocating hits.

been shown not to be expressed in proliferative bloodstream-form parasites (42), and was not pursued here.

Correct tagging of the active ES copy was confirmed by Southern blotting (supplemental Fig. S5) and ESAG fusion proteins were assessed for correct tagging by immunoblotting (supplemental Fig. 3*B*), and localized by native fluorescence microscopy (Fig. 7). Significantly, all ESAGs detected in our surface proteome localize to the surface membrane. ESAG6/ESAG7 localized to the FP and ESAG4 localized to the FP and flagellum membranes, as previously described (12, 41). Other surface proteome ESAGs localized to the cell body membrane (ESAG2), cell body and FP (ESAG10), cell body and flagellum (ESAG11), or FP and endosomes (ESAG5). With respect to those ESAGs not detected or not enriched in our surface proteome, ESAG12 was detected in endosomes, consistent with being also at the surface at low levels and/or recycling through the endomembrane and surface compartments. ESAG8 was expressed at levels close to the limit of detection by immunoblot when tagged at either end of the endogenous ES copy, and was undetectable in localization experiments. Importantly, ESAG1 and ESAG3 - which contain signal sequences suggestive of possible surface-association, but which were not enriched in the surface proteome—did not localize to the cell surface when tagged. These data demonstrate that (1) tagging with our vectors does not cause non-TbBSP proteins to mislocalize to the surface and (2) ESAG1 and 3 are unlikely to be surface-associated. Hence, of the 12 ESAG proteins, nine are shown to be surface-associated or secreted, five of which present at the FP membrane, clearly arguing for direct roles in host–parasite interactions by virtue of being exposed to the host environment.

Protein Localization Suggests Distinct Functional Membrane Domains Maintained by Selective Barriers—The trypanosome surface can be conceptually divided into three regions of contiguous membrane: the FP, the flagellum membrane and the cell body. Our localization data, using the same tag with 14 hypothetical proteins and 6 ESAGs that clearly

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

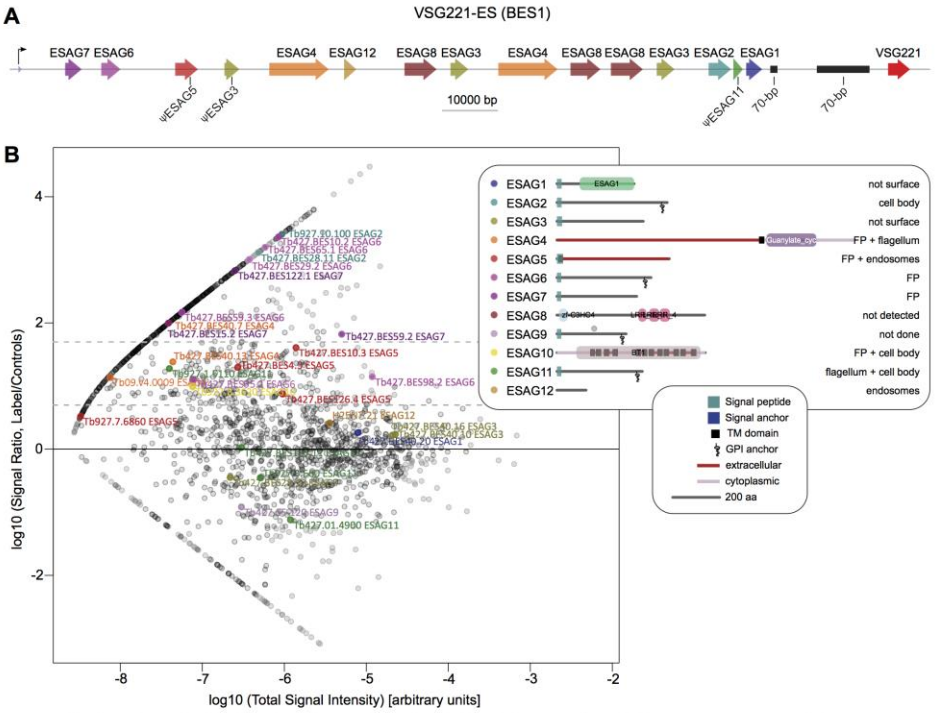


FIG. 6. **Identification and validation of ESAG proteins in the TbBSP.** *A*, Structure of the VSG221 expression site (BES1), which is active in cells used in this study. *B*, Distribution of ESAG proteins in the comparative label-free semi-quantitative mass spectrometry. Dataset as in Fig. 2, with points representing members of the 12 ESAG families (including predicted GRESAGs) highlighted. *C*, Protein architectures and localization of the ESAGs. All localization data are from ES copies, except for ESAG5, 10 and 11, which are not present in BES1) for which a detected GRESAG was used. See Experimental Procedures for details of protein feature prediction.

target the cell surface membrane, allowed us probe for the existence of these or other membrane domains with the largest set of trypanosome surface proteins systematically tested to date. Individual proteins in our sets were found to be restricted to any one of these domains or to combinations of them (Fig. 4 and supplemental Fig. 4), suggesting that the three regions indeed act as specialized domains of surface membrane, divided by selective barriers. Notably, we found only one example of sub-localization within a region (for ESP8), indicating that most proteins have free diffusion within each of the surface membrane domains.

A polarized distribution of ESPs and ESAGs implies intrinsic protein-sorting signals governing location on the cell surface. We therefore analyzed this set for the presence of common sequence motifs or structure that might regulate such sorting; However, no simple correlation between cellular localization and protein architecture emerged. For example, predicted GPI-anchored proteins were not all restricted to the FP, nor were type I TM proteins restricted to the cell body membrane (Fig. 8). Furthermore, motif elicitation analysis (MEME) detected no common motifs among ESPs and ESAGs with shared localization (data not shown). This suggests that protein topology alone may not be the primary determinant of surface domain segregation in *T. brucei*, and more complex interactions are at play.

DISCUSSION

A Surface Proteome for African Trypanosomes—Here we describe a high-confidence, validated surface proteome for the major host form of African trypanosome parasites. This was achieved through a novel biochemical preparation in

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

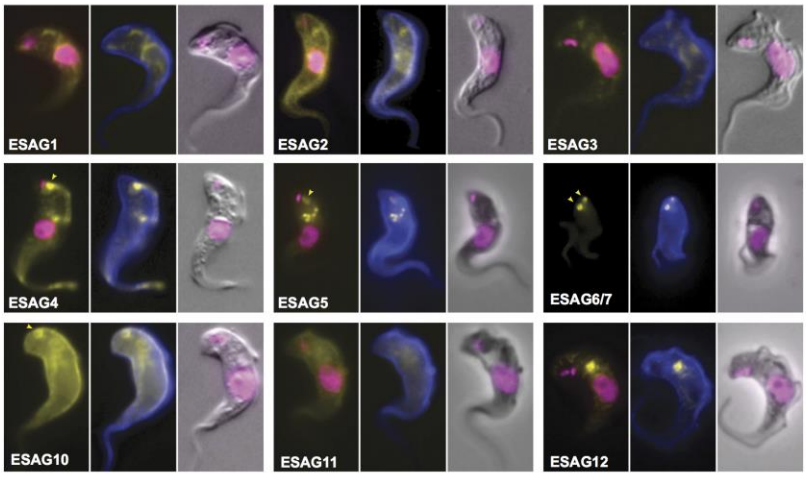


FIG. 7. **Most ESAGs encode surface-associated proteins.** 9 ESAGs were localized by tagging the respective gene at the active ES (except for ESAG5, 10 and 11, for which a surface proteome GRESAG was used) with an ORF encoding superfolder GFP and imaged using native fluorescence microscopy. Signal from superfolder-GFP is shown in yellow. Cells have been counterstained with concanavalin A (ConA, blue) and DAPI (magenta). The FP is indicated by yellow arrowhead.

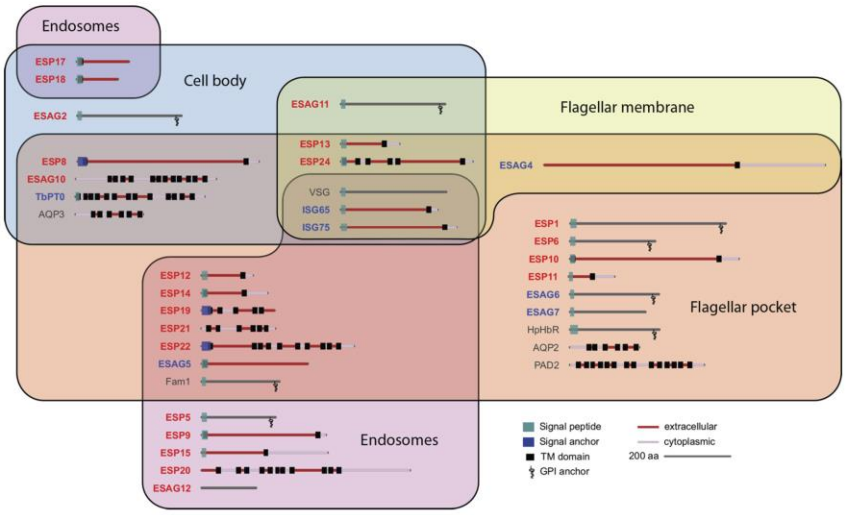


FIG. 8. **Domain architecture for surface-associated proteins in *T. brucei*.** Names of proteins which were localized as part of this study are emboldened (red: localized for the first time; blue: also localized in other studies). Data from proteins that have been previously localized by equivalent tagging methods (only) are also shown for comparison (references included in supplemental Table S3). See Experimental Procedures for details of protein feature prediction.

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

which the use of fluorescein was one of several steps optimized to increase both the specificity and sensitivity of our approach. Cell surface proteomic studies of other human pathogens, as well as mammalian cells, have frequently used the biotin-avidin based system to isolate plasma membrane proteins (47–50). In initial experiments we too used sulfo-NHS-biotin to chemically modify the surface of live trypanosomes. However, following affinity chromatography with streptavidin, we found the specificity of the approach was compromised by high background from control (unlabeled) cells, which could not be removed even on extensive washing. This may be a product of the parasite's intrinsic biochemistry: trypanosomatids (except those harboring bacterial endosymbionts) are unable to synthesize biotin (51); but this vitamin is an essential requirement for cell growth (52), and known to be incorporated into endogenous proteins (53). To avoid contamination with endogenously biotinylated parasite proteins, we abandoned biotin as a chemical tag, and moved to fluorescein labeling combined with an antigen–antibody purification system. Fluorescein is cell-impermeable, ensuring that only surface membrane proteins from intact cells are labeled by covalent modification of accessible lysine residues, and antigen-antibody columns can be washed to high stringency. Fluorescein also has an advantage that it can be followed visually or by fluorimetry during preparations. Using this approach, we have developed here a strategy for the identification of surface-exposed membrane proteins, which in trypanosomes isolates proteins that, at steady-state, reside at the FP, early/recycling endosomes, flagellum and cell body membranes.

Our surface proteome was extracted from a specific biochemical preparation coupled with comparative semi-quantitative mass spectrometry and bioinformatic filters. The bioinformatic methods used decrease the risk of contaminants in the defined TbBSP in a manner analogous to those used to describe the high-quality set of nucleoporins that compose the trypanosome nuclear pore complex (33). In that study, an initial set of 757 mass spectrometry hits was reduced by removing 448 contaminants on the basis of functionally unrelated sequence homology and gene annotations (*e.g.* ribosomal, endoplasmic reticulum and cytosolitc proteins). The remaining 309 proteins were informatically filtered for features associated with known nucleoporins (such as functional motifs, molecular weight and predicted secondary structure) (33). Here we filtered our experimental data for sequences that predict targeting to the endoplasmic reticulum and membrane anchoring (either via a GPI anchor or a transmembrane domain).

The contrasting approach of interrogating the entire genome sequence for cell surface localization on the basis of bioinformatic prediction of membrane-association is not applicable to our question because ~15% of the parasite's predicted proteome (1465 proteins) have such features. To define a predicted cell surface phylome, Jackson and colleagues (32) combined this approach with sequence clustering to look more specifically at those putative membrane-associated proteins in multigene families. Of the 50 CSP families present in *T. brucei*, 20 are detected in the surface proteome (supplemental Fig. S6). Particularly well-represented are Fam10 and Fam79 (supplemental Fig. S6), which comprise proteins of unknown function for which we present the first experimental evidence. For example, of the 7 members in Fam10, five were detected in our dataset, and we have demonstrated the surface association of one (ESP17). Significantly, however, the majority of proteins (63%) in the surface proteome are not part of multigene families (and hence not part of the CSP), yet are *bona fide* surface-associated proteins according to our validation experiments (12 out of 18 ESPs). This highlights the strength of our joint approach of sensitive, semi-quantitative detection and bioinformatic filtering.

Extent of the Surface Proteome—The surface proteome includes almost all previously characterized surface proteins for *T. brucei* (albeit rather few in number), as well as hypothetical proteins with predicted function as receptors, transporters, channels and others. These data suggest that the overall coverage of surface proteins in our high confidence set is broad, although it is to be expected that it will not be complete. A natural limitation of our approach is that it only derivatizes surface components with regions of modifiable polypeptide chain exposed to the extracellular space. This excludes proteins solely associated with the cytoplasmic side of the membrane. Hence, proteins modified by N-terminal palmitoyl- or myristoylation (such as for the flagellum calcium-binding protein calflagin (54)) are not expected to be present. Such proteins were not the focus of this work, as our primary objective was to gain knowledge of the molecular components exposed at the host–parasite interface.

A more significant cohort of proteins that may be underrepresented in our surface proteome are those with few exposed extracellular lysine residues. This may explain why aquaporins 2 and 3 (shown to localize to the FP and cell body membranes of *T. brucei* (14)) are not present in the TbBSP. Neither is a putative calcium channel protein (FS179/Tb927.10.2880) localized to the region of flagellum attachment to the cell (15)). These proteins are multipass TM proteins and are predicted to have limited sequence on the extracellular side of the membrane (*e.g.* aquaporins have only three lysines predicted to be extracellular, which may or may not be accessible to fluoresceination depending on the folding of the protein). A number of transporters and channel-like proteins are present in the surface proteome (10/175 proteins in total)—and validation showed that five detected multipass TM proteins are indeed surface-associated—but it is noteworthy that proteins with predicted TM domains were under-represented in our preparations (Fig. 2).

Confidence of Surface Prediction—We believe that a specific strength of the present work is the robust validation.

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

Alongside bioinformatic support, we also developed a genetic toolkit to test a subset of 25 candidates for FP/surface localization. The majority were true surface components (14 out of 23 detectable fusion proteins were present at the cell surface, four found predominantly in endosomal compartments that are likely to cycle to the surface in small amounts, whereas five localized elsewhere in the cell). This suggests that our surface proteome contains relatively few false-positives (~22% at the >5× threshold, and likely far fewer at greater enrichment values).

A number of ESAGs were present in our surface proteome and were localized to the cell surface when tagged, compared with only 1 out of 4 (ESAG12) not present in the TbBSP (despite containing sequence characteristics that might have suggested surface proteins). Although this is only a small set, it does indicate that the levels of false-negatives in our analysis (*i.e.* proteins that should have been detected, but were not) is also proportionally low. It is improbable that our surface proteome contains all proteins resident at the parasite surface, but results from localization of hypothetical proteins and ESAGs indicate a high confidence for the 175 proteins identified herein.

One issue with the interpretation of localization data for ESPs and ESAGs is in defining where the cell surface ends. Most of the proteins tested were detectable by light microscopy at locations in the cell consistent with being the FP, flagellum or cell body membranes. However, the plasma membrane is highly dynamic and is in constant exchange with components of the endosomal system. In trypanosomes, TfR, ISG and VSG are all present in endosomes as well as at the cell surface. It is thus possible that some of the TbBSP proteins not localized to cell surface domains are still molecules that are found transiently or in low abundance at the cell surface. In mammalian cells it is common to find many proteins cycling between the cell surface and early/recycling endosomes, but proteins as "deep" as those found in lysosomes have also been observed on the surface (55–57). African trypanosomes too have a transport route for newly synthesized lysosomal membrane glycoproteins to exit the Golgi and reach the lysosome via the FP (58), though the lysosomal marker p67 may take a direct route that bypasses the FP membrane (59). Hence, it may be biologically meaningful that proteins such as ESP15 (a type I TM protein that localized to the lysosome) is in the surface proteome, whereas p67 (also a type I TM glycoprotein) is not.

Membrane Domains and Domain Maintenance—The few surface proteins analyzed to date suggest the existence of at least three biochemically distinct domains across contiguous membranes, and emphasize the idea that individual proteins can access one or more domains on the cell surface. For example, TfR is restricted to the FP and endosomes, the adenylate cyclase encoded by ESAG4 is present at the FP and flagellum, and VSG is distributed across the entire surface membrane and endosomal system. The work here considerably expands these observations, showing that 8 ESAGs and 14 proteins of unknown function localize to one or more of three separate membrane domains: the FP, the flagellar membrane, the cell body. Moreover, these proteins do so in all possible combinations (with the exception of flagellar membrane alone, which was not observed).

Our results support a model whereby trypanosome surface organization is determined by control of access to any of three membrane domains. The finding that only one surface protein (ESP8) showed evidence of sub-domain localization suggests that diffusion within each domain is essentially free for most components. However, selective diffusion barriers or very rapid transfer systems exist between these domains. Because newly synthesized proteins are delivered to the FP, most combinations could be produced by the "opening" of symmetrical barriers at either the base of the flagellum (to access the flagellar membrane) or distal end of the FP (to access the cell body membrane). Nonetheless, the existence of proteins that are enriched in at just the cell body (ESP17, ESP18 and ESAG2) or cell body plus flagellar membrane (ESAG11) suggests that for at least some of the surface proteins the barriers or protein movement must be asymmetric.

This model raises major questions with regards to the mechanisms underlying protein sorting and retention in African trypanosomes, and elucidating such mechanisms in any cell type remains a formidable challenge. We considered that common motifs within the primary sequence might be used to target ESPs and ESAGs to their respective domains or enable them to cross specific domain boundaries, but simple common signals were not found in our analyses. It is also clear that gross protein architecture (*e.g.* GPI-anchor, type I TM, etc.) is not predictive of domain localization, suggesting that the signals are encoded by more complex or protein-specific cues.

The barriers to protein movement on the cell surface are likely to be contained in the structural features described at the boundaries between the domains—the rows of intramembrane particles seen by freeze-fracture electron microscopy forming the ciliary necklace at the junction of the flagellum and FP membranes, and the junction of the FP and neck membrane (60). The molecular identity of these particles remains unknown, but a morphologically similar configuration identified at the base of the mammalian primary cilium requires the GTPase septin for retention of receptors in that organelle (61). Alternatively, lipid composition, particularly that able to accommodate the geometric constraints of highly curved membrane sections (like that at the junction of the flagellum and the FP) could act as barriers to protein movement or as targeting signal. The distribution pattern of membrane probes and GPI-anchored YFP between the ciliary and plasma membranes are consistent with lipid composition operating in this manner (62).

For two trypanosome membrane proteins, lateral movement between surface domains appears to be dependent on protein abundance as well as identity. Over-expression of a

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

membrane-bound acid phosphatase predominantly found in endosomes causes it to re-distribute over the whole cell surface (36). In a similar manner, TfR in excess of normal levels is no longer retained in FP and endosomes, and escapes to the entire cell surface (35). The relevance of such artificial overexpression to endogenous protein targeting is uncertain, but trypanosomes grown in serum with low-affinity transferrin compensate by up-regulating the expression of TfR which, in turn, escapes the FP (35). However, it is clear that surface domain targeting in trypanosomes must be more complex than just a saturable mechanism of FP retention, as has been proposed for TfR, because we observe proteins with localizations specific to each individual domain, and combinations thereof - including proteins excluded from the FP (*e.g.* ESAG2), from most of the cell body (*e.g.* ESP8) or from the flagellum membrane (*e.g.* ESAG10).

Unraveling the Host–Parasite Interface—With a cell body entirely covered by ten million copies of a single glycoprotein, cellular functions that would normally occur at the plasma membrane of a typical eukaryotic cell are here concentrated at the FP of trypanosomes. The restriction of endocytosis and secretion to a focal point on the parasite surface allows for invariant receptors, channels and transporters, and other signaling molecules to be sequestered in an environment that is protected from the attention of host defenses, whereas the cell body membrane is mostly denuded of those proteins. Sitting at the critical interface between host and parasite, it is surprising that so few components of the FP have been described prior to this study. The essential nature of receptors such as TfR and HpHbR highlights the FP as an area of vulnerability that could be exploited in a therapeutic context. Our work has expanded this portfolio to 12 novel FP components with proven localization and identifies a total of 175 in the surface proteome, >50% of which are estimated to also be FP proteins. Importantly, 60% of surface proteome components cause a significant loss-of-fitness when knocked down individually (50/83 genes covered in a large-scale RNAi library screen (63)) compared with 42% for all genes ($p = 0.001$), showing that the TbBSP is notably enriched in genes essential for growth in the bloodstream. Because these proteins are mostly parasite specific and exposed to the extracellular space, our surface proteome is a potential source of drugable targets for disease treatment and control.

The high-confidence surface proteome described here greatly increases our knowledge of the trypanosome surface, and provides a significant resource against which hypothesis about membrane protein sorting and retention might be tested. Moreover, the methods we described are widely applicable to the study of cell membrane composition in human pathogens in general; whereas the surface compartmentalization is significant for understanding trypanosome biology and an important paradigm for surface organization in other systems.

Acknowledgments—We thank Ben Thomas (University of Oxford) for help with semi-quantitative analysis of mass spectrometry data; Jim Haseloff (University of Cambridge) for superfolder fluorescent protein constructs; and Mark Carrington (University of Cambridge) and Bob Lloyd (University of Nottingham) for access to fluorescence microscopes. We also thank Catherine Lindon (University of Cambridge), Flavia Moreira-Leite (University of Oxford), and David Horn (University of Dundee) for critical reading of the manuscript.

* This work was supported by the MRC (project grant G0900255 to CG and MCF), the Royal Society (research grant RG130195 to CG), the BBSRC (new investigator award BB/J01477X/1 to BW), the National Institutes of Health (PHS GM103314 and GM103511 grants to BTC and WZ) and the Wellcome Trust (program grant 090007/Z/09/Z to MCF). Data are available via ProteomeXchange with identifier PXD002180.

S This article contains supplemental Figs. S1 to S6 and Tables S1 to S3.

** To whom correspondence should be addressed: School of Life Sciences, University of Nottingham, Queen's Medical Centre, Nottingham NG2 7UH, United Kingdom. Tel.: 44-115-8230378; Email: catarina.gadelha@nottingham.ac.uk.

During this project, Woods and Oberholzer (and collaborators) reported on the localization of ESP8 (which the authors named FLA3) and ESP10 (therein named FS133), respectively (15, 64).

REFERENCES

1. Morrison L. J., Marcello L., and McCulloch R. (2009) Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity. *Cell Microbiol.* **11,** 1724–1734
2. Burton D. R., Poignard P., Stanfield R. L., and Wilson I. A. (2012) Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science* **337,** 183–186
3. Engstler M., Thilo L., Weise F., Grunfelder C. G., Schwarz H., Boshart M., and Overath P. (2004) Kinetics of endocytosis and recycling of the GPI-anchored variant surface glycoprotein in *Trypanosoma brucei*. *J. Cell Sci.* **117,** 1105–1115
4. Engstler M., Pfohl T., Herminghaus S., Boshart M., Wiegertjes G., Heddergott N., and Overath P. (2007) Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. *Cell* **131,** 505–515
5. Allen C. L., Goulding D., and Field M. C. (2003) Clathrin-mediated endocytosis is essential in *Trypanosoma brucei*. *EMBO J.* **22,** 4991–5002
6. Garcia-Salcedo J. A., Perez-Morga D., Gijon P., Dilbeck V., Pays E., and Nolan D. P. (2004) A differential role for actin during the life cycle of *Trypanosoma brucei*. *EMBO J.* **23,** 780–789
7. Bonhivers M., Nowacki S., Landrein N., and Robinson D. R. (2008) Biogenesis of the trypanosome endo-exocytotic organelle is cytoskeleton mediated. *PLoS Biol.* **6,** e105
8. McLaughlin J. (1987) Trypanosoma rhodesiense: antigenicity and immunogenicity of flagellar pocket membrane components. *Exp. Parasitol.* **64,** 1–11
9. Grab D. J., Webster P., Ito S., Fish W. R., Verjee Y., and Lonsdale-Eccles J. D. (1987) Subcellular localization of a variable surface glycoprotein phosphatidylinositol-specific phospholipase-C in African trypanosomes. *J. Cell Biol.* **105,** 737–746
10. Olenick J. G., Wolff R., Nauman R. K., and McLaughlin J. (1988) A flagellar pocket membrane fraction from *Trypanosoma brucei rhodesiense*: immunogold localization and nonvariant immunoprotection. *Infect. Immun.* **56,** 92–98
11. Nolan D. P., Jackson D. G., Windle H. J., Pays A., Geuskens M., Michel A., Voorheis H. P., and Pays E. (1997) Characterization of a novel, stage-specific, invariant surface protein in *Trypanosoma brucei* containing an internal, serine-rich, repetitive motif. *J. Biol. Chem.* **272,** 29212–29221
12. Salmon D., Geuskens M., Hanocq F., Hanocq-Quertier J., Nolan D., Ruben L., and Pays E. (1994) A novel heterodimeric transferrin receptor encoded by a pair of VSG expression site-associated genes in *T. brucei*. *Cell* **78,** 75–86
13. Vanhollebeke B., De Muylder G., Nielsen M. J., Pays A., Tebabi P., Dieu M., Raes M., Moestrup S. K., and Pays E. (2008) A haptoglobin-hemoglobin

receptor conveys innate immunity to *Trypanosoma brucei* in humans. *Science* **320,** 677–681
14. Baker N., Glover L., Munday J. C., Aguinaga Andres D., Barrett M. P., de Koning H. P., and Horn D. (2012) Aquaglyceroporin 2 controls susceptibility to melarsoprol and pentamidine in African trypanosomes. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 10996–11001
15. Oberholzer M., Langousis G., Nguyen H. T., Saada E. A., Shimogawa M. M., Jonsson Z. O., Nguyen S. M., Wohlschlegel J. A., and Hill K. L. (2011) Independent analysis of the flagellum surface and matrix proteomes provides insight into flagellum signaling in mammalian-infectious *Trypanosoma brucei*. *Mol. Cell. Proteomics* **10**, M111 010538
16. Dean S., Marchetti R., Kirk K., and Matthews K. R. (2009) A surface transporter family conveys the trypanosome differentiation signal. *Nature* **459,** 213–217
17. Hertz-Fowler C., Figueiredo L. M., Quail M. A., Becker M., Jackson A., Bason N., Brooks K., Churcher C., Fahkro S., Goodhead I. et al. (2008) Telomeric expression sites are highly conserved in *Trypanosoma brucei*. *PLoS One* **3,** e3527
18. Cross G. A., Kim H. S., and Wickstead B. (2014) Capturing the variant surface glycoprotein repertoire (the VSGnome) of *Trypanosoma brucei* Lister 427. *Mol. Biochem. Parasitol.* **195,** 59–73
19. Trudgian D. C., Ridlova G., Fischer R., Mackeen M. M., Ternette N., Acuto O., Kessler B. M., and Thomas B. (2011) Comparative evaluation of label-free SINQ normalized spectral index quantitation in the central proteomics facilities pipeline. *Proteomics* **11,** 2790–2797
20. Nielsen H., Engelbrecht J., Brunak S., and von Heijne G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10,** 1–6
21. Nielsen H., and Krogh A. (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6,** 122–130
22. Pierleoni A., Martelli P. L., and Casadio R. (2008) PredGPI: a GPI-anchor predictor. *BMC Bioinformatics* **9,** 392
23. Sonnhammer E. L., von Heijne G., and Krogh A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6,** 175–182
24. Krogh A., Larsson B., von Heijne G., and Sonnhammer E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305,** 567–580
25. Wirtz E., Leal S., Ochatt C., and Cross G. A. (1999) A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **99,** 89–101
26. Melville S. E., Leech V., Gerrard C. S., Tait A., and Blackwell J. M. (1998) The molecular karyotype of the megabase chromosomes of *Trypanosoma brucei* and the assignment of chromosome markers. *Mol. Biochem. Parasitol.* **94,** 155–173
27. Jackson D. G., Owen M. J., and Voorheis H. P. (1985) A new method for the rapid purification of both the membrane-bound and released forms of the variant surface glycoprotein from *Trypanosoma brucei*. *Biochem. J.* **230,** 195–202
28. Grunfelder C. G., Engstler M., Weise F., Schwarz H., Stierhof Y. D., Boshart M., and Overath P. (2002) Accumulation of a GPI-anchored protein at the cell surface requires sorting at multiple intracellular levels. *Traffic* **3,** 547–559
29. Marcello L., and Barry J. D. (2007) Analysis of the VSG gene silent archive in *Trypanosoma brucei* reveals that mosaic gene expression is prominent in antigenic variation and is favored by archive substructure. *Genome Res.* **17,** 1344–1352
30. Broadhead R., Dawe H. R., Farr H., Griffiths S., Hart S. R., Portman N., Shaw M. K., Ginger M. L., Gaskell S. J., McKean P. G. et al. (2006) Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature* **440,** 224–227
31. Colasante C., Voncken F., Manful T., Ruppert T., Tielens A. G., van Hellemond J. J., and Clayton C. (2013) Proteins and lipids of glycosomal membranes from *Leishmania tarentolae* and *Trypanosoma brucei*. *F1000Res.* **2,** 27
32. Jackson A. P., Allison H. C., Barry J. D., Field M. C., Hertz-Fowler C., and Berriman M. (2013) A cell-surface phylome for African trypanosomes. *PLoS Negl Trop Dis* **7,** e2121
33. DeGrasse J. A., DuBois K. N., Devos D., Siegel T. N., Sali A., Field M. C., Rout M. P., and Chait B. T. (2009) Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor. *Mol Cell Proteomics* **8,** 2119–2130
34. Pedelacq J. D., Cabantous S., Tran T., Terwilliger T. C., and Waldo G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79–88
35. Mussmann R., Janssen H., Calafat J., Engstler M., Ansorge I., Clayton C., and Borst P. (2003) The expression level determines the surface distribution of the transferrin receptor in *Trypanosoma brucei*. *Mol. Microbiol.* **47,** 23–35
36. Engstler M., Weise F., Bopp K., Grunfelder C. G., Gunzel M., Heddergott N., and Overath P. (2005) The membrane-bound histidine acid phosphatase TbMBAP1 is essential for endocytosis and membrane recycling in *Trypanosoma brucei*. *J. Cell Sci.* **118,** 2105–2118
37. Leung K. F., Riley F. S., Carrington M., and Field M. C. (2011) Ubiquitylation and developmental regulation of invariant surface protein expression in trypanosomes. *Eukaryot. Cell* **10,** 916–931
38. Lips S., Geuskens M., Paturiaux-Hanocq F., Hanocq-Quertier J., and Pays E. (1996) The esag 8 gene of *Trypanosoma brucei* encodes a nuclear protein. *Mol. Biochem. Parasitol.* **79,** 113–117
39. Hoek M., Engstler M., and Cross G. A. (2000) Expression-site-associated gene 8 (ESAG8) of *Trypanosoma brucei* is apparently essential and accumulates in the nucleolus. *J. Cell Sci.* **113 (Pt 22),** 3959–3968
40. Paindavoine P., Rolin S., Van Assel S., Geuskens M., Jauniaux J. C., Dinsart C., Huet G., and Pays E. (1992) A gene from the variant surface glycoprotein expression site encodes one of several transmembrane adenylate cyclases located on the flagellum of *Trypanosoma brucei*. *Mol. Cell. Biol.* **12,** 1218–1225
41. Salmon D., Vanwalleghem G., Morias Y., Denoeud J., Krumbholz C., Lhomme F., Bachmaier S., Kador M., Gossmann J., Dias F. B. et al. (2012) Adenylate cyclases of *Trypanosoma brucei* inhibit the innate immune response of the host. *Science* **337,** 463–466
42. Barnwell E. M., van Deursen F. J., Jeacock L., Smith K. A., Maizels R. M., Acosta-Serrano A., and Matthews K. (2010) Developmental regulation and extracellular release of a VSG expression-site-associated gene product from *Trypanosoma brucei* bloodstream forms. *J. Cell Sci.* **123,** 3401–3411
43. Rifkin M. R. (1978) Trypanosoma brucei: some properties of the cytotoxic reaction induced by normal human serum. *Exp. Parasitol.* **46,** 189–206
44. Hajduk S. L., Moore D. R., Vasudevacharya J., Siqueira H., Torri A. F., Tytler E. M., and Esko J. D. (1989) Lysis of *Trypanosoma brucei* by a toxic subspecies of human high density lipoprotein. *J. Biol. Chem.* **264,** 5210–5217
45. Xong H. V., Vanhamme L., Chamekh M., Chimfwembe C. E., Van Den Abbeele J., Pays A., Van Meirvenne N., Hamers R., De Baetselier P., and Pays E. (1998) A VSG expression site-associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. *Cell* **95,** 839–846
46. Vanhamme L., Paturiaux-Hanocq F., Poelvoorde P., Nolan D. P., Lins L., Van Den Abbeele J., Pays A., Tebabi P., Van Xong H., Jacquet A.et al. (2003) Apolipoprotein L-I is the trypanosome lytic factor of human serum. *Nature* **422**, 83–87
47. Roesli C., Borgia B., Schliemann C., Gunthert M., Wunderli-Allenspach H., Giavazzi R., and Neri D. (2009) Comparative analysis of the membrane proteome of closely related metastatic and nonmetastatic tumor cells. *Cancer Res.* **69,** 5406–5414
48. de Miguel N., Lustig G., Twu O., Chattopadhyay A., Wohlschlegel J. A., and Johnson P. J. (2010) Proteome analysis of the surface of *Trichomonas vaginalis* reveals novel proteins and strain-dependent differential expression. *Mol. Cell Proteomics* **9,** 1554–1566
49. Niehage C., Steenblock C., Pursche T., Bornhauser M., Corbeil D., and Hoflack B. (2011) The cell surface proteome of human mesenchymal stromal cells. *PLoS One* **6,** e20399
50. Biller L., Matthiesen J., Kuhne V., Lotter H., Handal G., Nozaki T., Saito-Nakano Y., Schumann M., Roeder T., Tannich E. et al. (2014) The cell surface proteome of *Entamoeba histolytica*. *Mol. Cell Proteomics* **13,** 132–144
51. Klein C. C., Alves J. M., Serrano M. G., Buck G. A., Vasconcelos A. T., Sagot M. F., Teixeira M. M., Camargo E. P., and Motta M. C. (2013) Biosynthesis of vitamins and cofactors in bacterium-harbouring trypanosomatids depends on the symbiotic association as revealed by genomic analyses. *PLoS One* **8,** e79786

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

52. Roitman C., Roitman I., and de Azevedo H. P. (1972) Growth of an insect trypanosomatid at 37C in a defined medium. *J. Protozool.* **19,** 346–349
53. dos Santos A. L., Abreu C. M., Batista L. M., Alviano C. S., and de Araujo Soares R. M. (2001) Cell-associated and extracellular proteinases in Blastocrithidia culicis: influence of growth conditions. *Curr. Microbiol.* **43,** 100–106
54. Emmer B. T., Souther C., Toriello K. M., Olson C. L., Epting C. L., and Engman D. M. (2009) Identification of a palmitoyl acyltransferase required for protein sorting to the flagellar membrane. *J. Cell Sci.* **122,** 867–874
55. Lippincott-Schwartz J., and Fambrough D. M. (1986) Lysosomal membrane dynamics: structure and interorganellar movement of a major lysosomal membrane glycoprotein. *J. Cell Biol.* **102,** 1593–1605
56. Harter C., and Mellman I. (1992) Transport of the lysosomal membrane glycoprotein lgp120 (lgp-A) to lysosomes does not require appearance on the plasma membrane. *J. Cell Biol.* **117,** 311–325
57. Saftig P., and Klumperman J. (2009) Lysosome biogenesis and lysosomal membrane proteins: trafficking meets function. *Nat. Rev. Mol. Cell Biol.* **10,** 623–635
58. Brickman M. J., and Balber A. E. (1994) Transport of a lysosomal membrane glycoprotein from the Golgi to endosomes and lysosomes via the cell surface in African trypanosomes. *J. Cell Sci.* **107 (Pt 11),** 3191–3200
59. Alexander D. L., Schwartz K. J., Balber A. E., and Bangs J. D. (2002) Developmentally regulated trafficking of the lysosomal membrane protein p67 in *Trypanosoma brucei*. *J. Cell Sci.* **115,** 3253–3263
60. Gadelha C., Rothery S., Morphew M., McIntosh J. R., Severs N. J., and Gull K. (2009) Membrane domains and flagellar pocket boundaries are influenced by the cytoskeleton in African trypanosomes. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 17425–17430
61. Hu Q., Milenkovic L., Jin H., Scott M. P., Nachury M. V., Spiliotis E. T., and Nelson W. J. (2010) A septin diffusion barrier at the base of the primary cilium maintains ciliary membrane protein distribution. *Science* **329,** 436–439
62. Vieira O. V., Gaus K., Verkade P., Fullekrug J., Vaz W. L., and Simons K. (2006) FAPP2, cilium formation, and compartmentalization of the apical membrane in polarized Madin-Darby canine kidney (MDCK) cells. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 18556–18561
63. Alsford S., Turner D. J., Obado S. O., Sanchez-Flores A., Glover L., Berriman M., Hertz-Fowler C., and Horn D. (2011) High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res.* **21,** 915–924
64. Woods K., Nic a'Bhaird N., Dooley C., Perez-Morga D., and Nolan D. P. (2013) Identification and characterization of a stage specific membrane protein involved in flagellar attachment in *Trypanosoma brucei*. *PLoS One* **8,** e52846

MCP MOLECULAR & CELLULAR PROTEOMICS ASBMB

Appendix 2. Primers used in this study. Uppercase indicates sequence complementary to template DNA, primer tags encoding restriction sites and sequence for use in fusion PCR are lowercase.

Endogenous-locus GFP-tagging constructs based on pSiS-HhsfG

Fragment	Forward primer (NotI/XbaI/HindIII)	Reverse primer (BamHI/NotI/XbaI)	Construct
ESAG5 UTR	tagcggccgcCCAGGGTGCATGTACG	caggatccCCTGCTGAAAATAGAGTTTGTACC	pSiS-HhsfG-ESAG5-Nt
ESAG5 ORF	catctagaCGGTCTAGTAGTCCCTTTACG	tagcggccgcCTTCCAACTGTCACGTTACGA	
ESAG6/7 UTR	tagcggccgcTTGAGTTGGTATTTGAGG	caggatccATCACTTTTTTAAATTATGCTTT	pSiS-HhsfG-ESAG6/7-Nt
ESAG6/7 ORF	catctagaTATTATGAAAATAAAAGGAATGCA	tagcggccgcCCTCAATACTGTTTTCAAGA	
ESAG10 UTR	ccaagcttAACCATATCTAAACCCTCCCA	tagcggccgcATCGCCACTGTTTCCC	pSiS-HhsfG-ESAG10-Ct
ESAG10 ORF	tagcggccgcTGCGACTTTAGTAATGTGAAGTG	catctagaATTGGCTGTGGGTTCC	

RNAi constructs based on p2T7-177

Targeting fragment	Forward primer (BamHI)	Reverse primer (ClaI/HindIII)	Construct
eGFP	caggatccCCACCTACGGCAAGCT	ccaagcttCCGATGGGGGTGTTCT	p2T7-177-eGFP
sfGFP	caggatccGCGAGGGTGAAGGTGA	ccaagcttTGATCAGCCAGCTGCA	p2T7-177-sfGFP
(GR)ESAG1 pan-family	caggatccCTTGTGTTGATGCATGGC	ccatcgatTCGGTCTTGAGTTTAGTGGTA	p2T7-177-(GR)ESAG1 pan-family
(GR)ESAG2 pan-family	caggatccTATGAAATAGTGATTGCCGTTGG	ccaagcttTCAAACTCAGCTAATGCACTTTT	p2T7-177-(GR)ESAG2 pan-family
(GR)ESAG3 pan-family	caggatccGAGAGCTCTATTCCTTCAAGAT	ccaagcttGCCTTAAACAAGTCATTCTCCT	p2T7-177-(GR)ESAG3 pan-family
(GR)ESAG4 pan-family	caggatccAATATGCTTCACTTGGGCG	ccaagcttGTCGGATAAAAGTGGACCC	p2T7-177-(GR)ESAG4 pan-family
(GR)ESAG5 pan-family 1	caggatccTGGGCAAAAGGTTAACGG	ccatcgatCACGACCAGTTCGAGTC	p2T7-177-(GR)ESAG5 pan-family RNAi fragment 1
(GR)ESAG6/7 pan-family	caggatccGTTTTGGTTTGTGCTGTTGG	ccaagcttATACTTTCCGCACCCAAG	p2T7-177-(GR)ESAG6/7 pan-family
(GR)ESAG8 pan-family	caggatccGCACTACGTGATCTGGAAG	ccaagcttCATAGAGCACCCTCAAGTG	p2T7-177-(GR)ESAG8 pan-family

(GR)ESAG10 pan-family 1	caggatccATGACCACGTCACCGA	ccatcgatGCCACCGGTACATACG	p2T7-177-(GR)ESAG10 pan-family RNAi fragment 1
(GR)ESAG12 pan-family	caggatccAGCGGTGTCAATATTCAGG	ccaagcttAGGAGGAAGGAGTTTGAAGA	p2T7-177-(GR)ESAG12 pan-family

p2T7-177 drug swap

Product	Forward primer (SphI/XbaI)	Reverse primer (XbaI/RsrII)	Construct
p2T7-177 rDNA promoter	CAGGCATGCAAGCTA	catctagaGCTTATTTTATGGCAGCAAC	p2T7-B177
BLAST	catctagaCATGCCTTTGTCTCAAGA	cgtcggaccgGTTAGCCCTCCCACA	

RNAi constructs based on p2T7-B177

Targeting fragment	Forward primer (BamHI)	Reverse primer (HindIII)	Construct
(GR)ESAG5 pan-family 2	caggatccATTCCGTGGTTTCCCC	GCAAGCGGAGTACGTT	p2T7-B177-(GR)ESAG5 pan-family RNAi fragment 2-5
(GR)ESAG5 pan-family 3	aacgtactccgcttgcGCTAATGTCCCAGAAACCA	TTGCCATGTCCTCCAC	
(GR)ESAG5 pan-family 4	gtggaggacatggcaaGCAGCTGGAGCGAATA	CGAGGATGCATTGGGA	
(GR)ESAG5 pan-family 5	tcccaatgcatcctcgGCTTCGGTGTAGTTGCT	ccaagcttAAGGGGCAATCTGGAC	
(GR)ESAG10 pan-family 2	caggatccGGTCGACAGGAAGATGG	ATCCACCAACAAGTCCG	p2T7-B177-(GR)ESAG10 pan-family RNAi fragment 2-3
(GR)ESAG10 pan-family 3	cggacttgttggtggatCTACGCGGTCACTACC	ccaagcttGCCAGAAGAGAGTCTCG	

KO fragments (*BLAST*)

Product	Forward primer	Reverse primer
BLAST	ATGCCTTTGTCTCAAGAAGAATC	TTAGCCCTCCCACACA
All 5' targeting	GACCAGCAGATGAAGCA	tcttgagacaaaggcatTATCACTTTTTTAAATTATGCTTTTTAAAACT

Par 5' targeting	TCAGTTGTGGTCACGG	tcttgagacaaaggcatCGTCGAATAGAAATAACACC
All/Par 3' targeting	tgtgggagggctaaGATATATGTGTATATACATTTTTTCTTGC	GACAATGTCGCCGAAAAC
Multi 5' targeting	ATAATAATATTATTACGATAGCAGTAGAAG	tcttgagacaaaggcatTCCCTAGCATACTAAAAAAATTTG
Multi 3' targeting	tgtgggagggctaaTGATAATTTATTTTACTTTTAACTTTGGG	CTACCATTATCCAAATATCTGC

KO fragments (*HYG* with actin UTRs cloned into pBlueScript SK II (+))

Product	Forward primer (XhoI)	Reverse primer (BamHI)
HYG with *actin* UTRs	AGGCTAGGGCACAG	GAATACTGCATAGATAACAAACGC
KO All 5' targeting	gacctcgagTCGCTTTCATAAATGCCA	tgtgccctagcctAACCTTGGTGCTAAAACA
KO Par 5' targeting	gacctcgagTCAGTTGTGGTCACG	tgtgccctagcctACACCGAATCCTGC
KO All/Par 3' targeting	cgtttgttatctatgcagtattcCGGCGACACACA	caggatccGGGTCAGTGGGTG
KO Multi 5' targeting	gacctcgagTCGTGGAATCCTTTCC	tgtgccctagcctACAAGGAGTTGCAAACT
KO Multi 3' targeting	cgtttgttatctatgcagtattcTTTCTTGTAGTGACGGAC	caggatccTTTGGTTGGCACTTGT

CreKO intermediate constructs based on pBluescript SK II (+)

Product	Forward primer (HindIII)	Reverse primer (SpeI)	Construct (cloned with loxP + aldolase 5' UTR gBlock)
TK-BLAST + *actin* 3' UTR	gataagcttATGGCTTCGTACCCC	ccactagtAATACTGCATAGATAACAAACGCA	pBS+lox-aldolase5'-TK-BLAST-actin3'UTR
PURO + *actin* 3' UTR	gataagcttATGACCGAGTACAAGCC		pBS+lox-aldolase5'-PURO-actin3'UTR

loxP + *aldolase* 5' UTR gBlock (EcoRI/HindIII)	atatcgaattcATAACTTCGTATAGCATACATTATACGAAGTTATCGGACCGTGCTCAAGCTGTGTAGCGCACGCGTTTCCTTACA TATTTCTCTAACAGGCACGGAAGCCTAACAAATACACTTGGCTTATTTTTTTGCCCCCTCATGTCTTGTACAAATATTTGCGA TAGCTTAGCTATCAGCCACATTAATCAAACAAGTATACCAACAAGCCCGAAAACAGAAACTCAACTGCAACGaagcttgac

CreKO_5' constructs based on pBS+lox-aldolase5'-TK-BLAST-actin3'UTR

Product	Forward primer (KpnI/NdeI)	Reverse primer (NdeI/EcoRI)	Construct
All 5' targeting	ctacatatgTTCGATTTATGTCGACTGTGT	tcgaattcCATGCAAAACAACACACCA	CreKO_5'All
All 3' targeting	taaggtaccCGAAGAGCAGGGGT	ctacatatgACTCTGCGTTATTACCCC	
Par 5' targeting	ctacatatgAGCCTGCCTGCA	tcgaattcTATCACAACCCTTGTACTTCTT	CreKO_5'Par
Par 3' targeting	taaggtaccGCTTATTTATATGATGGAGAAGCC	ctacatatgTAGTCCCACATTTCTGAGTATC	
Multi 5' targeting	ctacatatgTGCTTGCCACCTAAGT	tcgaattcAACGCAACTGGTAACAAC	CreKO_5'Multi
Multi 3' targeting	taaggtaccCATACATCTAAGATGTGTACGTTAAGA	ctacatatgGCACAGGTATTAGATGTTGC	

CreKO_3' constructs based on pBS+lox-aldolase5'-PURO-actin3'UTR

Product	Forward primer (SpeI/NdeI)	Reverse primer (NdeI/NotI)	Construct
All/Par 5' targeting	ctacatatgGGGAGAACTAAATTTTTGTCGG	tagcggccgcGAACCTCACAAGAAAATTTGAGAC	CreKO_3'All/Par
All/Par 3' targeting	ctactagtTGGCAAAAATATGATTTATAGGAGGA	ctacatatgTTGTGCCCGACCTTT	
Multi 5' targeting	ctacatatgATTGTTTTCTTGTAGTGACGGA	tagcggccgcTGAATCAATAACTTAAAAACCACAATAAGA	CreKO_3'Multi
Multi 3' targeting	ctactagtTTTTAAATATTTTGACTGTGAGCGG	ctacatatgCACACACTTCCACCTAAATTATTTG	
All/Par-VSGψ 5' targeting	ctacatatgCCTGCTCGCAACC	tagcggccgcCCGTCCTTGCTGTG	CreKO_3'All/Par-VSGψ
All/Par-VSGψ 3' targeting	ctactagtCCGACACACCAACC	ctacatatgGTCTGTGGGCTTGC	

CreKO_3' constructs based on CreKO_3'All/Par

Product	Forward primer (SpeI)	Reverse primer (NdeI)	Construct
3' All/Par-VSGψ 3' targeting	ctactagtCCGACACACCAACC	ctacatatgGTCTGTGGGCTTGC	CreKO_3'All/Par-E1VSGψ

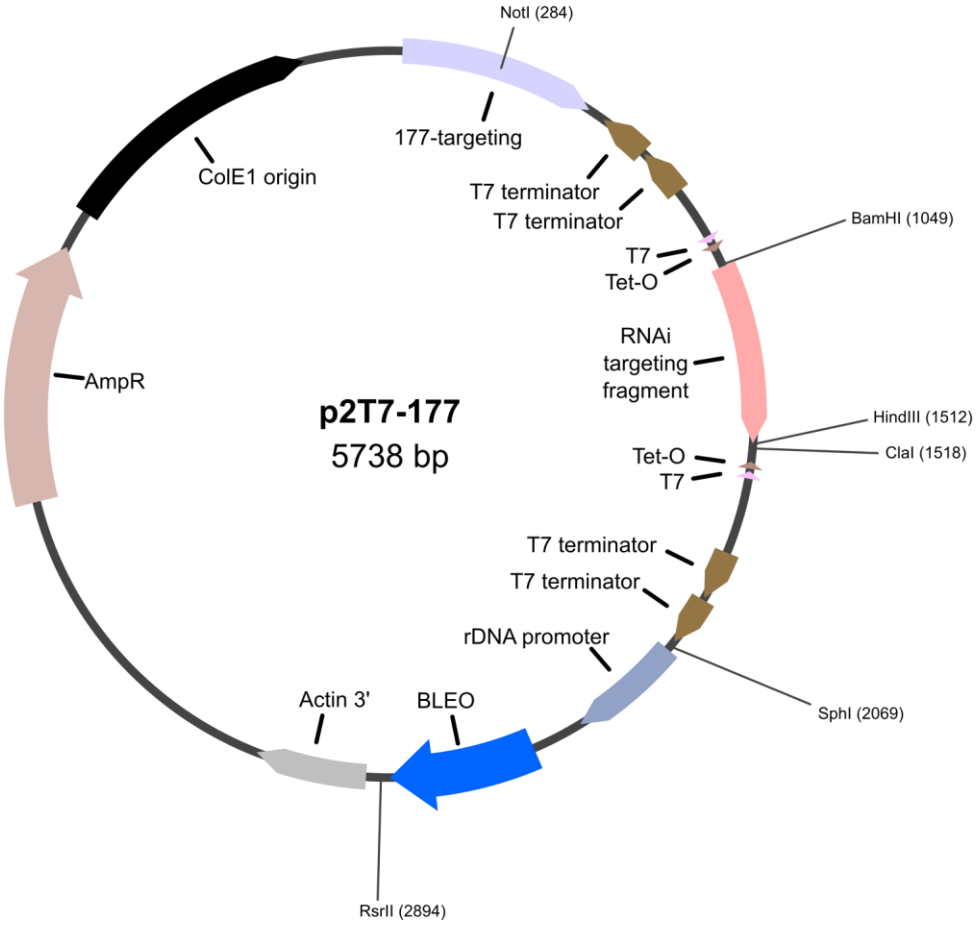
Diagnostic PCR

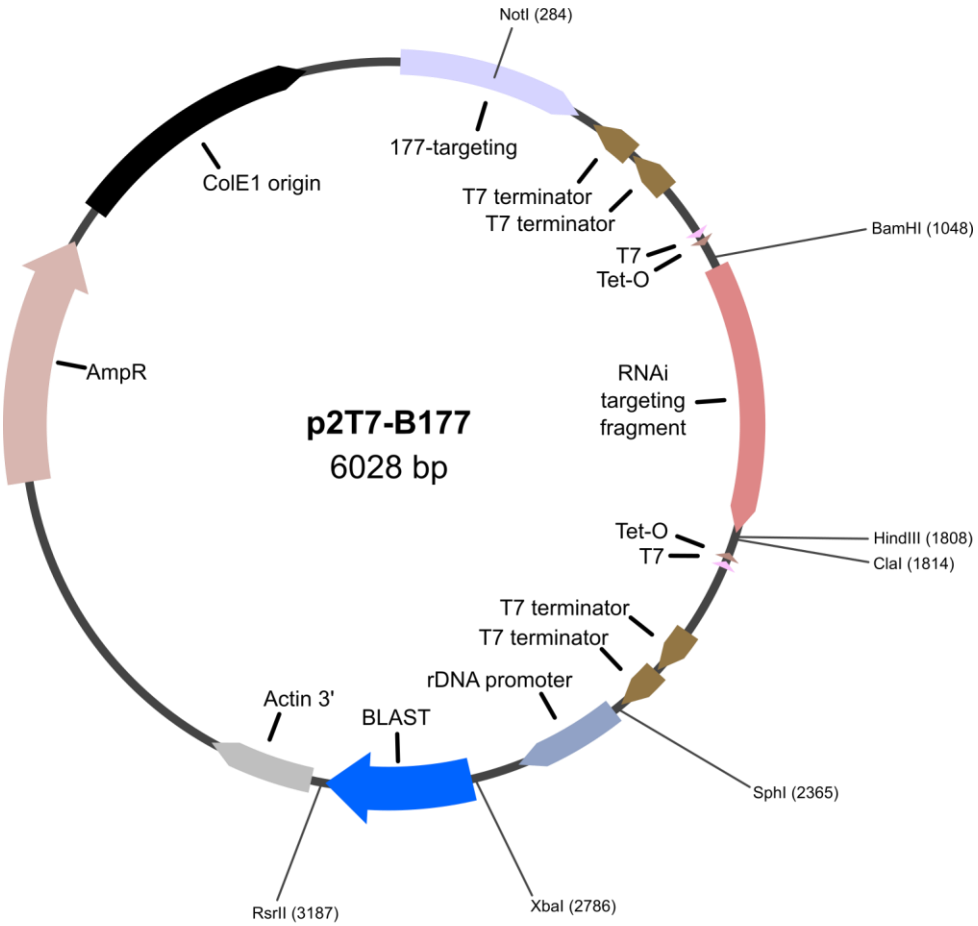
Fragment	Forward primer	Reverse primer
GFP-tagged *ESAG6* (Figure 12A. F1 + R1)	ACGCATGGTATGGATGAA	CTGCACCGTACTCTGTAG
GFP-tagged *ESAG7* (Figure 12A. F1 + R2)		CCACATCGTGAGAACTGA
GFP-tagged *ESAG3.1* control (Figure 15A. F1 + R1)	TCAGTTGGATTTGCTCATGA	CGTCGGAAGGGAATGTT
GFP-tagged *ESAG3.1* diagnostic (Figure 15A. F2 + R1)	CATGGACGAGCTGTACAA	
GFP-tagged *ESAG3.2* control (Figure 15A. F1 + R2)	TCAGTTGGATTTGCTCATGA	CACAGCCTCCTCAGAG
GFP-tagged *ESAG3.2* diagnostic (Figure 15A. F2 + R2)	CATGGACGAGCTGTACAA	
GFP-tagged *ESAG4.1* control (Figure 15A. F3 + R3)	TGTGAGTATCCTCGCC	ATGCAAATCTGCCACTTG
GFP-tagged *ESAG4.1* diagnostic (Figure 15A. F4 + R3)	GCAATAAAGCATCAGGTACC	
GFP-tagged *ESAG4.2* control (Figure 15A. F3 + R4)	TGTGAGTATCCTCGCC	CGGCCAGTCATTTCAG
GFP-tagged *ESAG4.2* diagnostic (Figure 15A. F4 + R4)	GCAATAAAGCATCAGGTACC	
GFP-tagged *ESAG8.1* control (Figure 15A. F5 + R5)	GTCTGTACGGTGGTTCTG	CACTCCCTAACACCGG
GFP-tagged *ESAG8.1* diagnostic (Figure 15A. F5 + R6)		CCCTTGCTCACACTAGTC
GFP-tagged *ESAG8.2* control (Figure 15A. F6 + R5)	AGCTTAAATTTCCGTTTGTGTGT	CACTCCCTAACACCGG
GFP-tagged *ESAG8.2* diagnostic (Figure 15A. F6 + R6)		CCCTTGCTCACACTAGTC
GFP-tagged *ESAG8.3* control (Figure 15A. F7 + R5)	GTGTGTTGTCCCTTTGGT	CACTCCCTAACACCGG
GFP-tagged *ESAG8.3* diagnostic (Figure 15A. F7 + R6)		CCCTTGCTCACACTAGTC
VSG221 (Figure 17B.)	TCAAACAAGCTTTTTGGCAACCT	TGGCTAGCCTGCATTCG
VSG121 (Figure 17B.)	GGCGTACGCGATTAGT	GGTGCCATCAGTTTCTGT

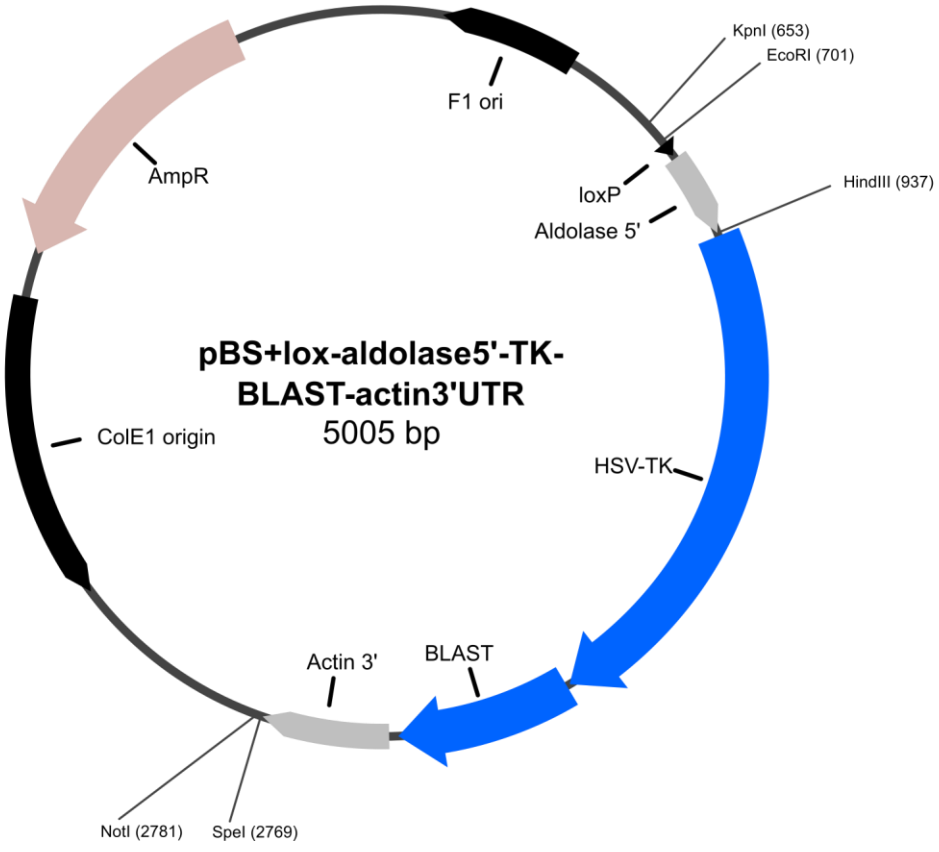
Southern blot probes

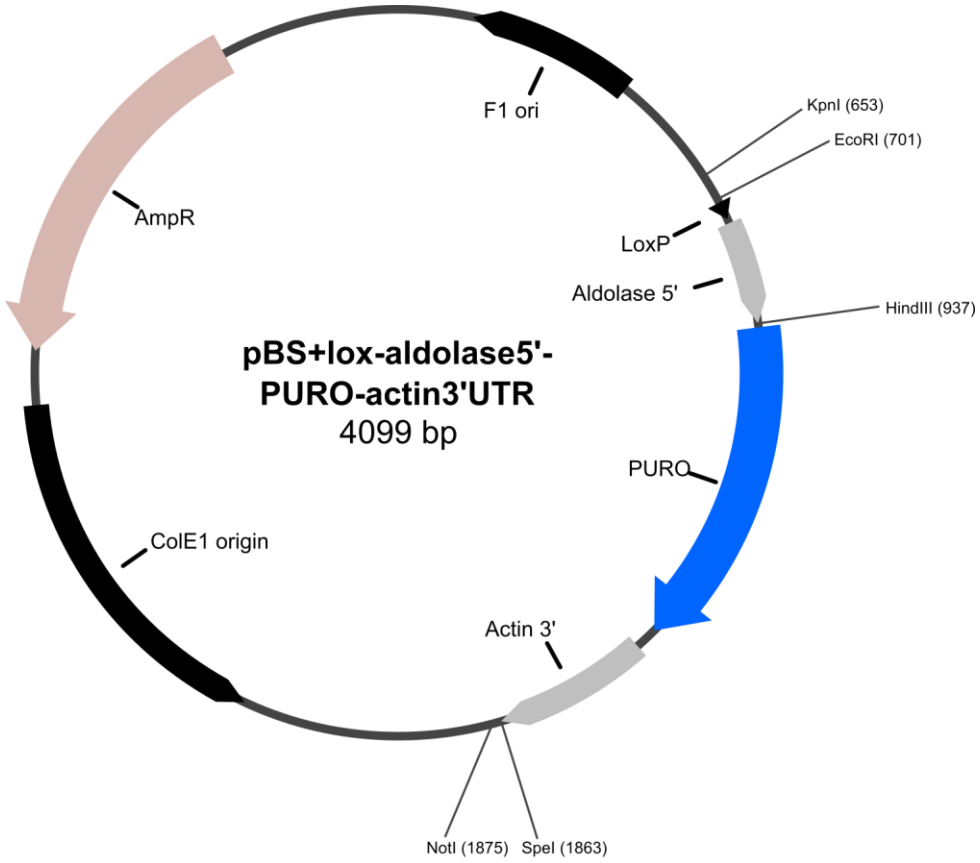
Probe	Forward primer	Reverse primer
eGFP	GGCGAGGAGCTGTTCA	TTGTACAGCTCGTCCATGC
sfGFP	GCGTAAAGGCGAAGAGCT	GTACAGTTCATCCATACCATGCG
HYG	ATGAAAAAGCCTGAACTCACCG	CTTTGCCCTCGGACGA
BLAST	ATGCCTTTGTCTCAAGAAGAATC	TTAGCCCTCCCACACA
PURO	ATGACCGAGTACAAGCC	TTAGGCACCGGGCTT
VSG221	TCAAACAAGCTTTTTGGCAACCT	GCTTGTTGTTGGCTTTCAGC
VSG121	GGCGTACGCGATTAGT	GGTGCCATCAGTTTCTGT
VSG224	ATAACCAGCATCGAAGCTG	GCAGTTTGATCCTGGCT

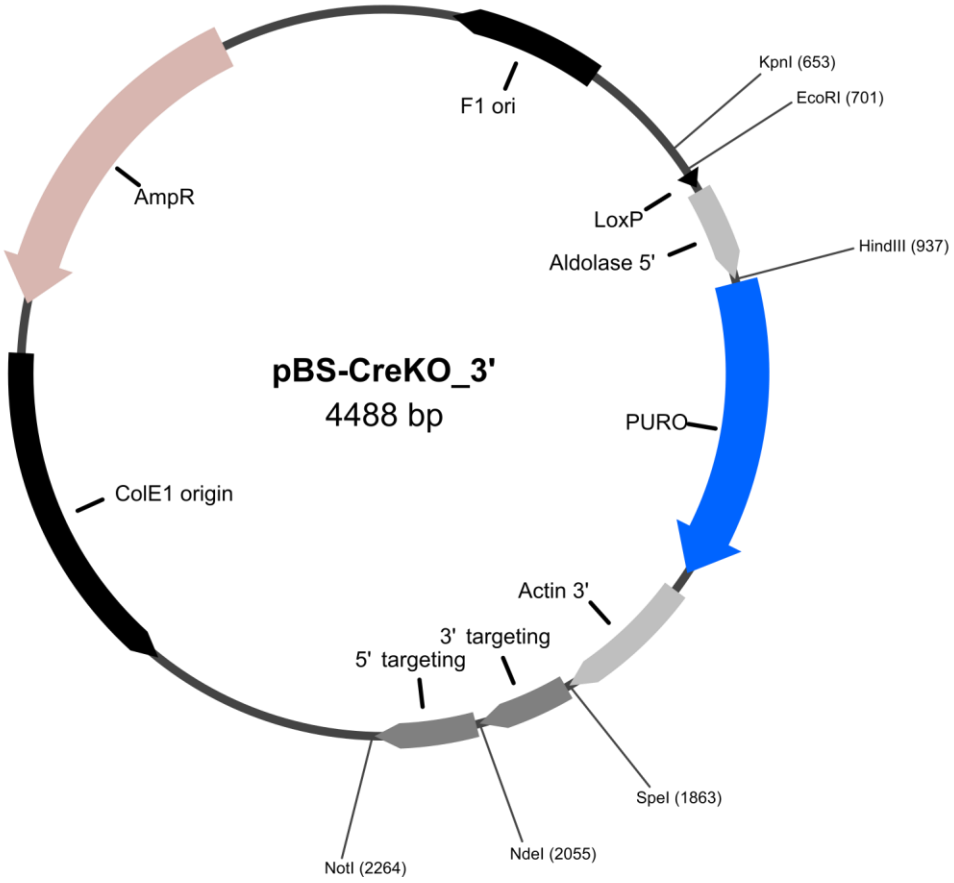

Actin 5'
ColE1 origin
HYG
AmpR
pSiS-HhsfG
5524 bp
PFR2 intergenic
Signal Peptide
HA
HindIII (1540)
AvrII (1636)
sfGFP
Aldolase 3'
XbaI (2473)
BamHI (2493)

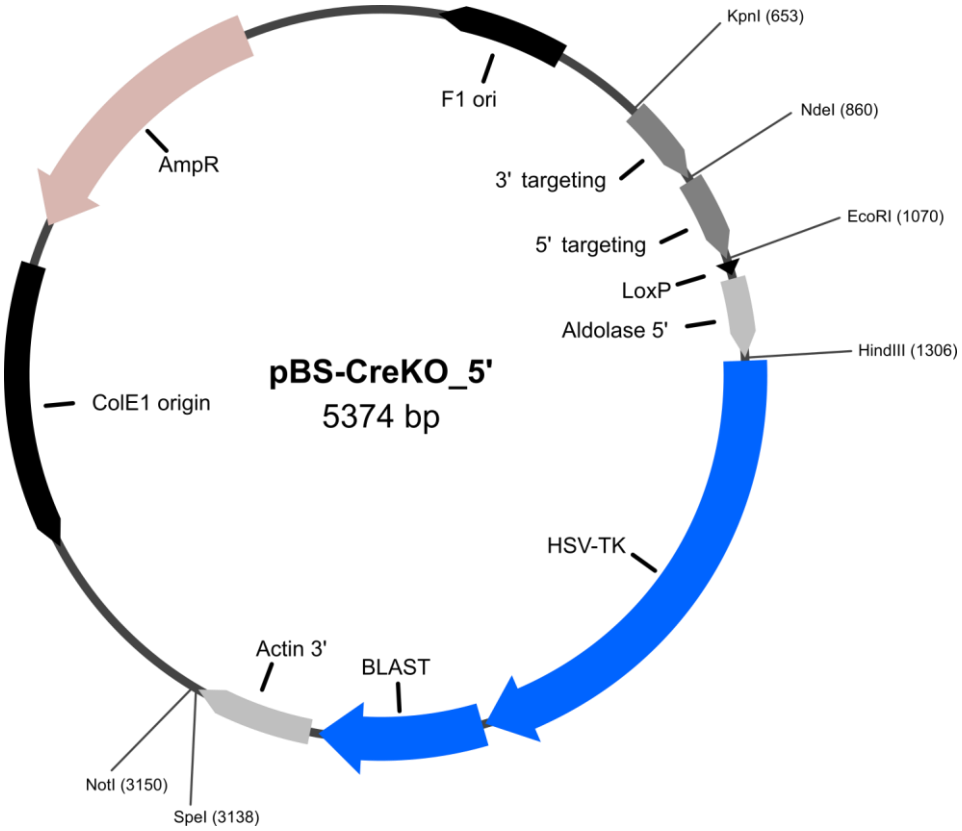

NotI (284)
177-targeting
ColE1 origin
T7 terminator
T7 terminator
BamHI (1049)
T7
Tet-O
RNAi
targeting
fragment
AmpR
p2T7-177
5738 bp
HindIII (1512)
ClaI (1518)
Tet-O
T7
T7 terminator
T7 terminator
rDNA promoter
Actin 3'
BLEO
SphI (2069)
RsrII (2894)


NotI (284)
177-targeting
ColE1 origin
T7 terminator
T7 terminator
T7
Tet-O
BamHI (1048)
AmpR
RNAi
targeting
fragment
p2T7-B177
6028 bp
Tet-O
T7
HindIII (1808)
ClaI (1814)
T7 terminator
T7 terminator
rDNA promoter
Actin 3'
BLAST
SphI (2365)
RsrII (3187)
XbaI (2786)


KpnI (653)
EcoRI (701)
F1 ori
AmpR
loxP
HindIII (937)
Aldolase 5'
pBS+lox-aldolase5'-TK-
BLAST-actin3'UTR
5005 bp
ColE1 origin
HSV-TK
Actin 3'
BLAST
NotI (2781)
SpeI (2769)


F1 ori
KpnI (653)
EcoRI (701)
AmpR
LoxP
Aldolase 5'
HindIII (937)
pBS+lox-aldolase5'-
PURO-actin3'UTR
4099 bp
PURO
ColE1 origin
Actin 3'
NotI (1875)
SpeI (1863)


F1 ori
KpnI (653)
EcoRI (701)
AmpR
LoxP
Aldolase 5'
HindIII (937)
pBS-CreKO_3'
4488 bp
PURO
ColE1 origin
Actin 3'
3' targeting
5' targeting
SpeI (1863)
NotI (2264)
NdeI (2055)

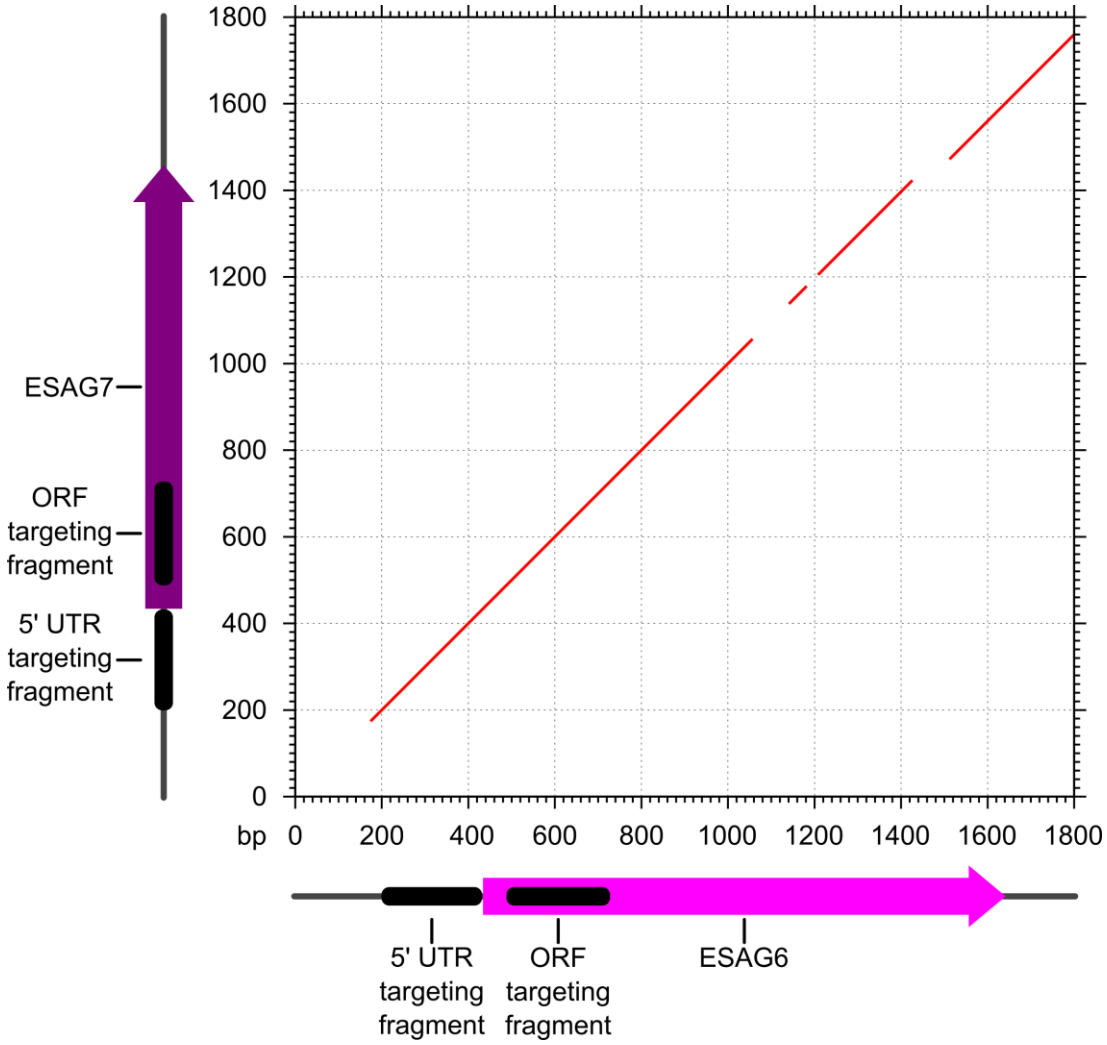


Appendix 3. Maps of constructs used in this study. Restriction enzymes used to produce constructs are shown.

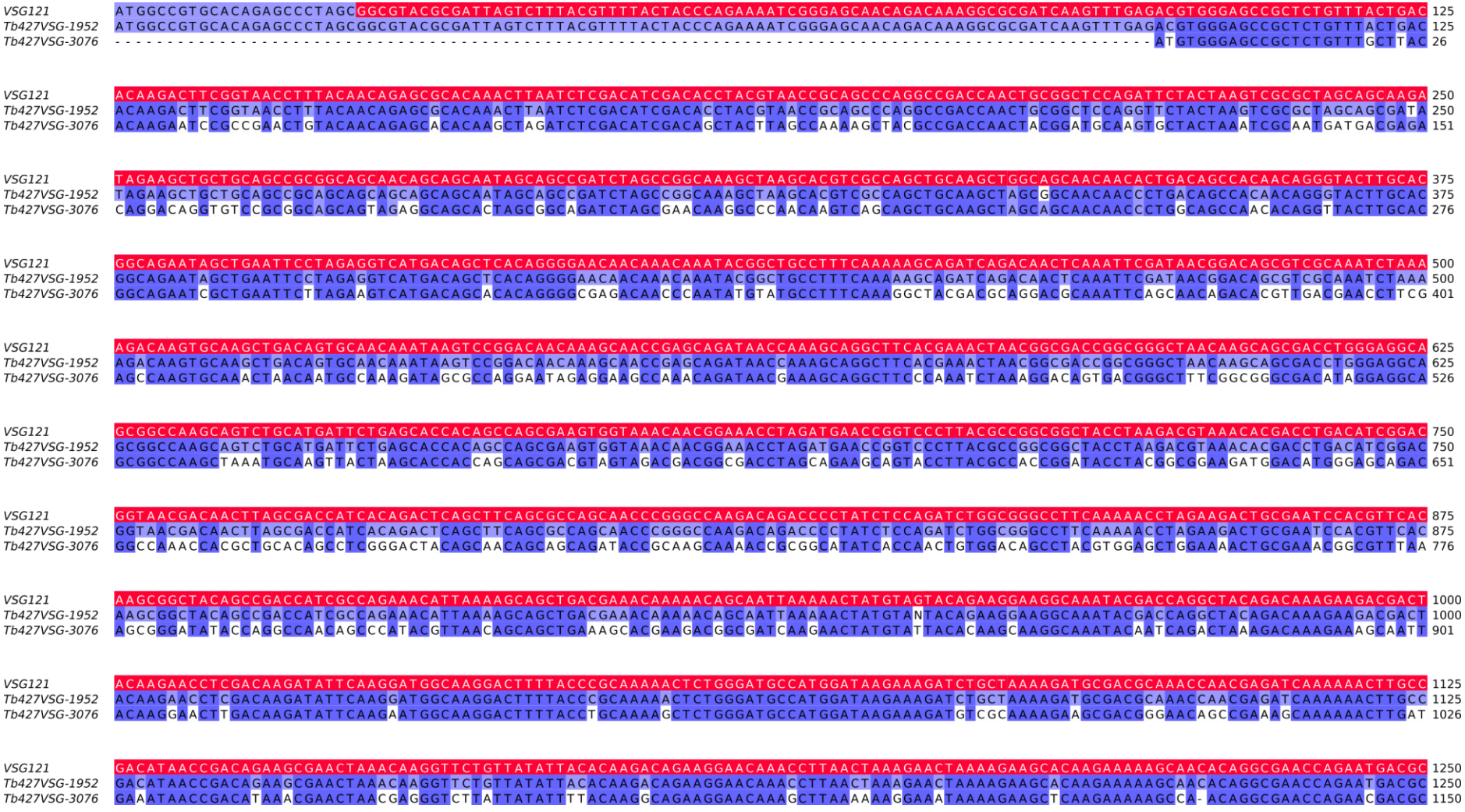
Appendix 4. Cell lines produced in this study.

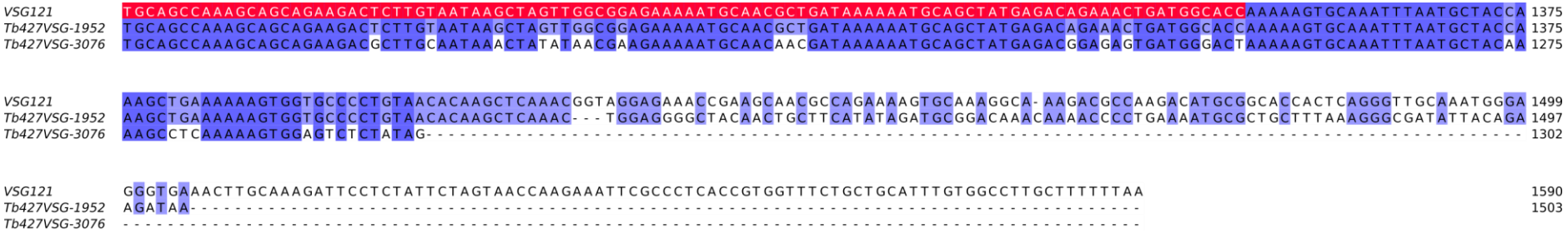
Plasmid	Background cell line	Resulting cell line	Selection marker
pSiS-HhsfG-ESAG6/7-Nt	S16	GFP-ESAG6	NEO, HYG
SmOx (Poon *et al*. 2012)	Wild type Lister 427 (obtained from Mark Carrington, University of Cambridge)	SmOx	PUR
p2T7-177-CHC (obtained from Sarah Whipple, University of Nottingham)	SmOx	SmOx *CLH* RNAi	PUR, BLE
pSiS-HhsfG-ESAG5-Nt	SmOx	GFP-ESAG5	PUR, HYG
pSiS-HhsfG-ESAG10-Ct	SmOx	ESAG10-GFP	PUR, HYG
p2T7-177-eGFP	GFP-ESAG1 (Table 4. Gadelha *et al*. 2015)	GFP-ESAG1 *GFP* RNAi	NEO, HYG, BLE
p2T7-177-eGFP	GFP-ESAG2 (Table 4. Gadelha *et al*. 2015)	GFP-ESAG2 *GFP* RNAi	NEO, HYG, BLE
p2T7-177-eGFP	GFP-ESAG3 (Table 4. Gadelha *et al*. 2015)	GFP-ESAG3 *GFP* RNAi	NEO, HYG, BLE
p2T7-177-eGFP	ESAG4-GFP (Table 4. Gadelha *et al*. 2015)	ESAG4-GFP *GFP* RNAi	NEO, HYG, BLE
p2T7-177-sfGFP	GFP-ESAG5	GFP-ESAG5 *GFP* RNAi	PUR, HYG, BLE
p2T7-177-sfGFP	GFP-ESAG6	GFP-ESAG6 *GFP* RNAi	NEO, HYG, BLE
p2T7-177-eGFP	ESAG8-GFP (Table 4. Gadelha *et al*. 2015)	ESAG8-GFP *GFP* RNAi	NEO, HYG, BLE
p2T7-177-sfGFP	ESAG10-GFP	ESAG10-GFP *GFP* RNAi	PUR, HYG, BLE
p2T7-177-sfGFP	ESAG12-GFP (Table 4. Gadelha *et al*. 2015)	ESAG12-GFP *GFP* RNAi	NEO, HYG, BLE
p2T7-177-(GR)ESAG1 pan-family	GFP-ESAG1	GFP-ESAG1 *(GR)ESAG1* pan-family RNAi	NEO, HYG, BLE
p2T7-177-(GR)ESAG2 pan-family	GFP-ESAG2	GFP-ESAG2 *(GR)ESAG2* pan-family RNAi	NEO, HYG, BLE

p2T7-177-(GR)ESAG3 pan-family	GFP-ESAG3	GFP-ESAG3 *(GR)ESAG3* pan-family RNAi	NEO, HYG, BLE
p2T7-177-(GR)ESAG4 pan-family	ESAG4-GFP	ESAG4-GFP *(GR)ESAG4* pan-family RNAi	NEO, HYG, BLE
p2T7-177-(GR)ESAG5 pan-family RNAi fragment 1	GFP-ESAG5	GFP-ESAG5 *(GR)ESAG5* pan-family RNAi fragment 1 RNAi	PUR, HYG, BLE
p2T7-177-(GR)ESAG6/7 pan-family	GFP-ESAG6	GFP-ESAG6 *(GR)ESAG6/7* pan-family RNAi	NEO, HYG, BLE
p2T7-177-(GR)ESAG8 pan-family	ESAG8-GFP	ESAG8-GFP *(GR)ESAG8* pan-family RNAi	NEO, HYG, BLE
p2T7-177-(GR)ESAG10 pan-family RNAi fragment 1	ESAG10-GFP	ESAG10-GFP *(GR)ESAG10* pan-family RNAi fragment 1 RNAi	PUR, HYG, BLE
p2T7-177-(GR)ESAG12 pan-family	ESAG12-GFP	ESAG12-GFP *(GR)ESAG12* pan-family RNAi	NEO, HYG, BLE
p2T7-B177-(GR)ESAG5 pan-family RNAi fragment 2-5	GFP-ESAG5 *(GR)ESAG5* pan-family RNAi fragment 1	GFP-ESAG5 *(GR)ESAG5* pan-family RNAi	PUR, HYG, BLE, BLAST
p2T7-B177-(GR)ESAG10 pan-family RNAi fragment 2-3	ESAG10-GFP *(GR)ESAG10* pan-family RNAi fragment 1	ESAG10-GFP *(GR)ESAG10* pan-family RNAi	PUR, HYG, BLE, BLAST
CreKO_3'Multi	S16	S16 + CreKO_3'Multi	NEO, PURO
CreKO_3'All/Par-VSGψ	S16	S16 + CreKO_3'All/Par-VSGψ	NEO, PURO
CreKO_3'All/Par-E1VSGψ	S16	S16 + CreKO_3'All/Par-E1VSGψ	NEO, PURO
CreKO_5'Multi	S16 + CreKO_3'Multi	S16 + CreKO_3'Multi + CreKO_5'Multi	NEO, PURO, TK-BLAST
CreKO_5'All	S16 + CreKO_3'All/Par-VSGψ	S16 + CreKO_3'All/Par-VSGψ + CreKO_5'All	NEO, PURO, TK-BLAST
CreKO_5'Par	S16 + CreKO_3'All/Par-VSGψ	S16 + CreKO_3'All/Par-VSGψ + CreKO_5'Par	NEO, PURO, TK-BLAST
CreKO_5'All	S16 + CreKO_3'All/Par-E1VSGψ	S16 + CreKO_3'All/Par-E1VSGψ + CreKO_5'All	NEO, PURO, TK-BLAST
CreKO_5'Par	S16 + CreKO_3'All/Par-E1VSGψ	S16 + CreKO_3'All/Par-E1VSGψ + CreKO_5'Par	NEO, PURO, TK-BLAST



Appendix 5. *ESAG6* and *ESAG7* share a high degree of sequence identity. Nucleotide sequence of *ESAG6* and *ESAG7* from BES1 was used to produce a dot plot (MAFFT). Schematics of *ESAG6* and *ESAG7* are shown adjacent to dot plot axes. Targeting fragments used in the ESAG6 N-terminal tagging construct are highlighted. Diagram drawn to scale.


VSG121 ATGGCCGTGCACAGAGCCCTAGCGGCGTACGCGATTAGTCTTTACGTTTTACTACCCAGAAAATCGGGAGCAACAGACAAAGGCGCGATCAAGTTTGAGACGTGGGAGCCGCTCTGTTTACTGAC 125
Tb427VSG-1952 ATGGCCGTGCACAGAGCCCTAGCGGCGTACGCGATTAGTCTTTACGTTTTACTACCCAGAAAATCGGGAGCAACAGACAAAGGCGCGATCAAGTTTGAGACGTGGGAGCCGCTCTGTTTACTGAC 125
Tb427VSG-3076 ATGTGGGAGCCGCTCTGTTTGCTTAC 26
VSG121 ACAAGACTTCGGTAACCTTTACAACAGAGCGCACAAACTTAATCTCGACATCGACACCTACGTAACCGCAGCCCAGGCCGACCAACTGCGGCTCCAGATTCTACTAAGTCGCGCTAGCAGCAAGA 250
Tb427VSG-1952 ACAAGACTTCGGTAACCTTTACAACAGAGCGCACAAACTTAATCTCGACATCGACACCTACGTAACCGCAGCCCAGGCCGACCAACTGCGGCTCCAGGTTCTACTAAGTCGCGCTAGCAGCGATA 250
Tb427VSG-3076 ACAAGAATCCGCCGAACTGTACAACAGAGCACACAAGCTAGATCTCGACATCGACAGCTACTTAGCCAAAAGCTACGCCGACCAACTACGGATGCAAGTGCTACTAAATCGCAATGATGACGAGA 151
VSG121 TAGAAGCTGCTGCAGCCGCGGCAGCAACAGCAGCAATAGCAGCCGATCTAGCCGGCAAAGCTAAGCACGTCGCCAGCTGCAAGCTGGCAGCAACAACACTGACAGCCACAACAGGGTACTTGCAC 375
Tb427VSG-1952 TAGAAGCTGCTGCAGCCGCAGCAGCAGCAGCAATAGCAGCCGATCTAGCCGGCAAAGCTAAGCACGTCGCCAGCTGCAAGCTAGCGGCAACAACCCTGACAGCCACAACAGGGTACTTGCAC 375
Tb427VSG-3076 CAGGACAGGTGTCCGCGGCAGCAGTAGAGGCAGCACTAGCGGCAGATCTAGCGAACAAGGCCCAACAAGTCAGCAGCTGCAAGCTAGCAGCAACAACCCTGGCAGCCAACACAGGTTACTTGCAC 276
VSG121 GGCAGAATAGCTGAATTCCTAGAGGTCATGACAGCTCACAGGGGAACAACAAACAAATACGGCTGCCTTTCAAAAAGCAGATCAGACAACTCAAATTCGATAACGGACAGCGTCGCAAATCTAAA 500
Tb427VSG-1952 GGCAGAATAGCTGAATTCCTAGAGGTCATGACAGCTCACAGGGGAACAACAAACAAATACGGCTGCCTTTCAAAAAGCAGATCAGACAACTCAAATTCGATAACGGACAGCGTCGCAAATCTAAA 500
Tb427VSG-3076 GGCAGAATCGCTGAATTCTTAGAAGTCATGACAGCACACAGGGGCGAGACAACCCAATATGTATGCCTTTCAAAGGCTACGACGCAGGACGCAAATTCAGCAACAGACACGTTGACGAACCTTCG 401
VSG121 AGACAAGTGCAAGCTGACAGTGCAACAAATAAGTCCGGACAACAAAGCAACCGAGCAGATAACCAAAGCAGGCTTCACGAAACTAACGGCGACCGGCGGGCTAACAAGCAGCGACCTGGGAGGCA 625
Tb427VSG-1952 AGACAAGTGCAAGCTGACAGTGCAACAAATAAGTCCGGACAACAAAGCAACCGAGCAGATAACCAAAGCAGGCTTCACGAAACTAACGGCGACCGGCGGGCTAACAAGCAGCGACCTGGGAGGCA 625
Tb427VSG-3076 AGCCAAGTGCAAACTAACAATGCCAAAGATAGCGCCAGGAATAGAGGAAGCCAAACAGATAACGAAAGCAGGCTTCCCAAATCTAAAGGACAGTGACGGGCTTTCGGCGGGCGACATAGGAGGCA 526
VSG121 GCGGCCAAGCAGTCTGCATGATTCTGAGCACCACAGCCAGCGAAGTGGTAAACAACGGAAACCTAGATGAACCGGTCCCTTACGCCGGCGGCTACCTAAGACGTAAACACGACCTGACATCGGAC 750
Tb427VSG-1952 GCGGCCAAGCAGTCTGCATGATTCTGAGCACCACAGCCAGCGAAGTGGTAAACAACGGAAACCTAGATGAACCGGTCCCTTACGCCGGCGGCTACCTAAGACGTAAACACGACCTGACATCGGAC 750
Tb427VSG-3076 GCGGCCAAGCTAAATGCAAGTTACTAAGCACCACCAGCAGCGACGTAGTAGACGACGGCGACCTAGCAGAAGCAGTACCTTACGCCACCGGATACCTACGGCGGAAGATGGACATGGAGCAGAC 651
VSG121 GGTAACGACAACTTAGCGACCATCACAGACTCAGCTTCAGCGCCAGCAACCCGGGCCAAGACAGACCCCTATCTCCAGATCTGGCGGGCCTTCAAAAACCTAGAAGACTGCGAATCCACGTTCAC 875
Tb427VSG-1952 GGTAACGACAACTTAGCGACCATCACAGACTCAGCTTCAGCGCCAGCAACCCGGGCCAAGACAGACCCCTATCTCCAGATCTGGCGGGCCTTCAAAAACCTAGAAGACTGCGAATCCACGTTCAC 875
Tb427VSG-3076 GGCCAAACCACGCTGCACAGCCTCGGGACTACAGCAACAGCAGCAGATACCGCAAGCAAAACCGCGGCATATCACCAACTGTGGACAGCCTACGTGGAGCTGGAAAACTGCGAAACGGCGTTTAA 776
VSG121 AAGCGGCTACAGCCGACCATCGCCAGAAACATTAAAAGCAGCTGACGAAACAAAAACAGCAATTAAAAACTATGTAGTACAGAAGGAAGGCAAATACGACCAGGCTACAGACAAAGAAGACGACT 1000
Tb427VSG-1952 AAGCGGCTACAGCCGACCATCGCCAGAAACATTAAAAGCAGCTGACGAAACAAAAACAGCAATTAAAAACTATGTANTACAGAAGGAAGGCAAATACGACCAGGCTACAGACAAAGAAGACGACT 1000
Tb427VSG-3076 AGCGGGATATACCAGGCCAACAGCCCATACGTTAACAGCAGCTGAAGCACGAAGACGGCGATCAAGAACTATGTATTACACAAGCAAGGCAAATACAATCAGACTAAAGACAAAGAAAGCAATT 901
VSG121 ACAAGAACCTCGACAAGATATTCAAGGATGGCAAGGACTTTTACCCGCAAAAACTCTGGGATGCCATGGATAAGAAAGATCTGCTAAAAGATGCGACGCAAACCAACGAGATCAAAAAACTTGCC 1125
Tb427VSG-1952 ACAAGAACCTCGACAAGATATTCAAGGATGGCAAGGACTTTTACCCGCAAAAACTCTGGGATGCCATGGATAAGAAAGATCTGCTAAAAGATGCGACGCAAACCAACGAGATCAAAAAACTTGCC 1125
Tb427VSG-3076 ACAAGGAACTTGACAAGATATTCAAGAATGGCAAGGACTTTTACCTGCAAAAGCTCTGGGATGCCATGGATAAGAAAGATGTCGCAAAAGAAGCGACGGAACAGCCGAAAGCAAAAAACTTGAT 1026
VSG121 GACATAACCGACAGAAGCGAACTAAACAAGGTTCTGTTATATTACACAAGACAGAAGGAACAAACCTTAACTAAAGAACTAAAAGAAGCACAAGAAAAAGCAACACAGGCGAACCAGAATGACGC 1250
Tb427VSG-1952 GACATAACCGACAGAAGCGAACTAAACAAGGTTCTGTTATATTACACAAGACAGAAGGAACAAACCTTAACTAAAGAACTAAAAGAAGCACAAGAAAAAGCAACACAGGCGAACCAGAATGACGC 1250
Tb427VSG-3076 GAAATAACCGACATAAACGAACTAACGAGGTCTTATTATATTTTACAAGGCAGAAGGAACAAAGCTTAAAAAAGGAAATAAAAGAAGCTCAAGAAAAAGCCA-ACAGGCGAACCAGAACGACGC 1150

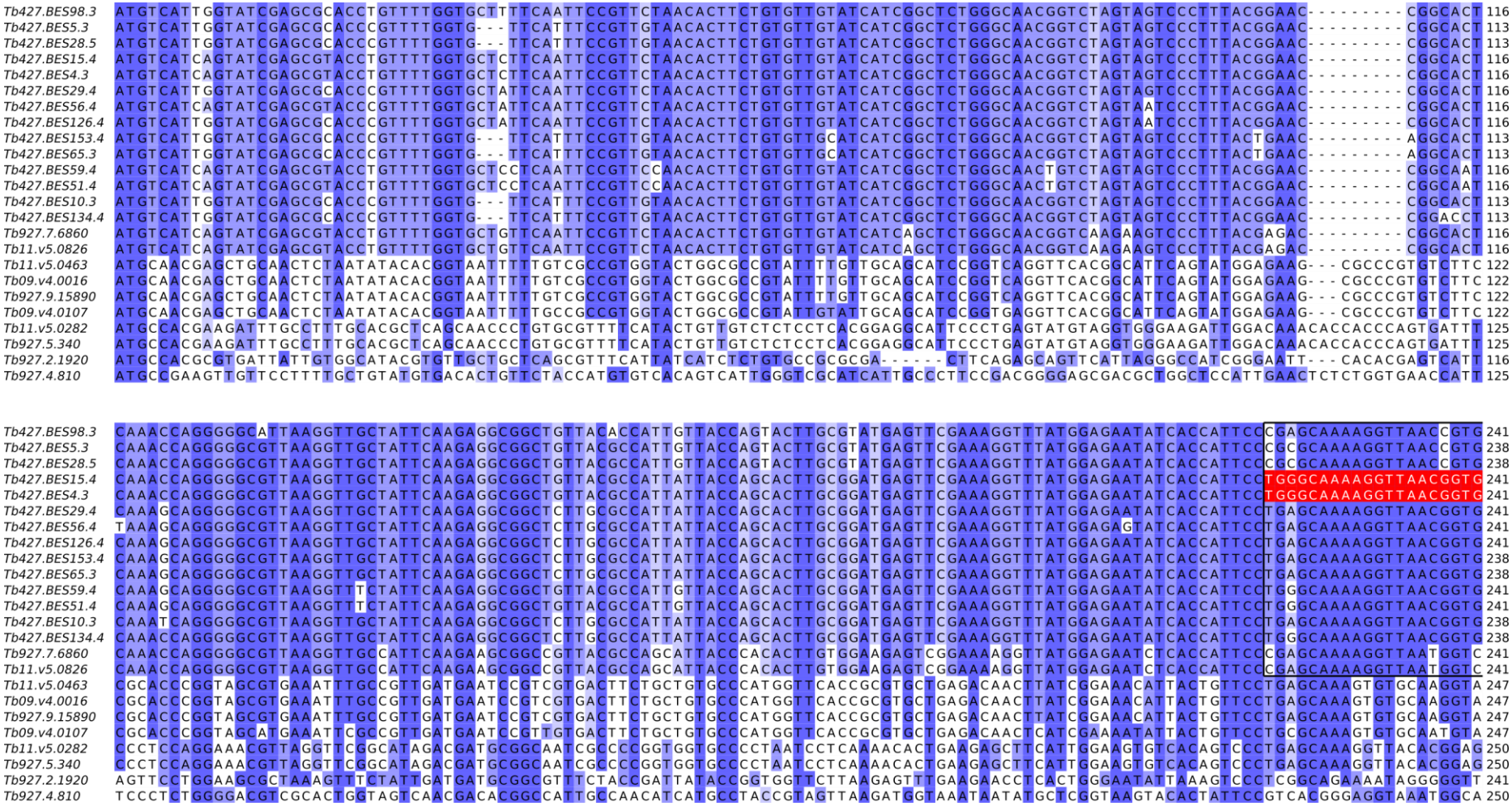


Appendix 6. The *T. brucei* genome encodes divergent copies of *VSG121*. Multiple DNA sequence alignment of *VSG121* (Tb427VSG-6) and related genes. Nucleotides coloured based on consensus, *VSG121* probe sequence highlighted in red. Alignment performed using MAFFT, annotated with Jalview.

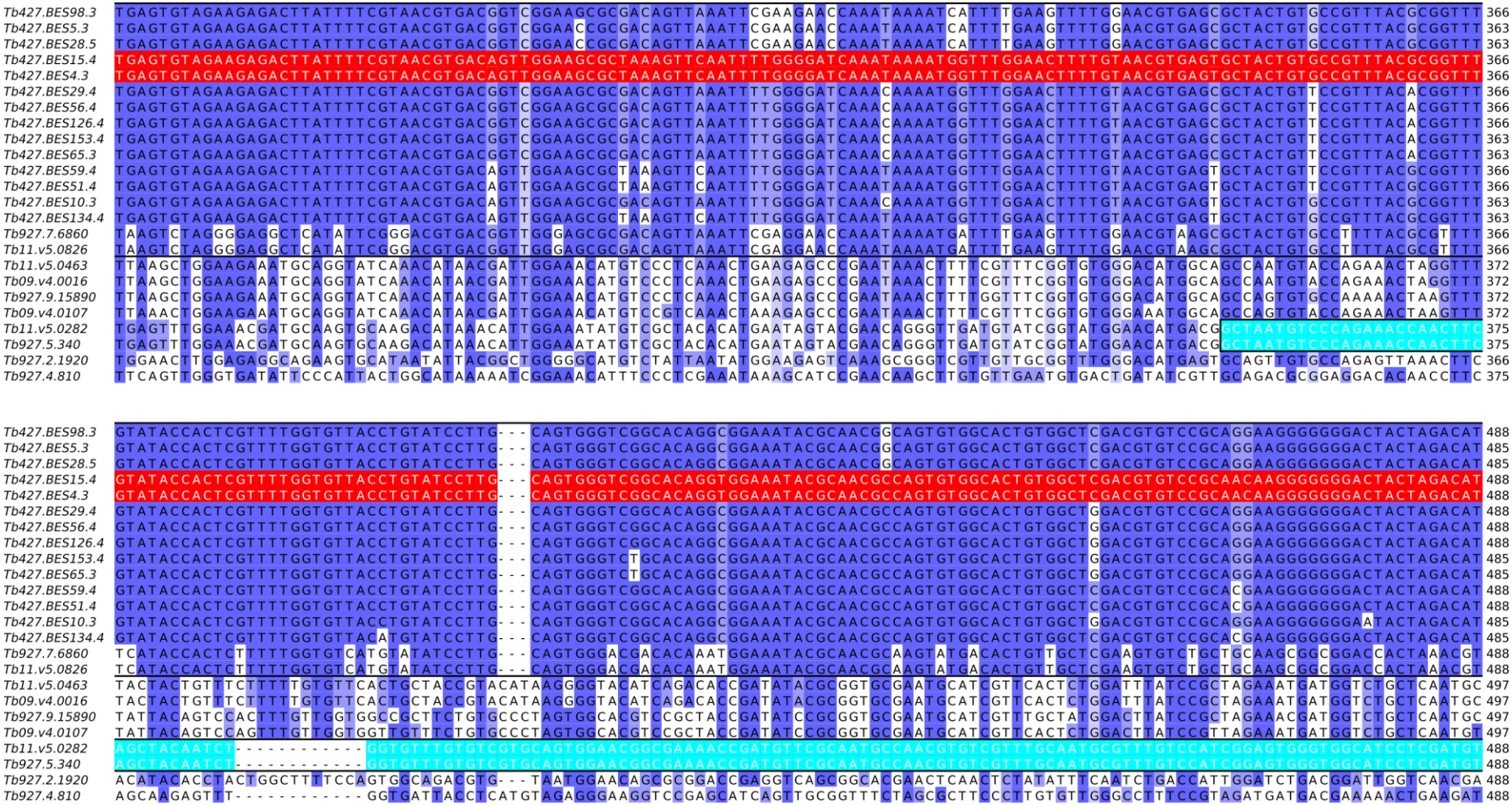
Appendix 7. *GRESAGs* targeted by pan-family RNAi fragments

(GR)ESAG family	Pan-family RNAi fragment				
	1	2	3	4	5
1	Tb927.1.4870 Tb927.1.4910 Tb927.1.5120 Tb927.3.5830 Tb927.9.660				
2	Tb927.1.2040 Tb927.1.4890 Tb927.1.5100 Tb927.3.2510 Tb927.3.570 Tb927.6.470 Tb927.6.500 Tb927.6.540 Tb927.9.7330 Tb927.9.7410 Tb11.01.6240				
3	Tb11.v5.0752				
4	Tb927.10.16190 Tb927.11.17040				
5	Tb927.7.6860 Tb11.v5.0826	Tb927.9.15890 Tb09.v4.0016 Tb09.v4.0107 Tb11.v5.0463	Tb927.5.340 Tb11.v5.0282	Tb927.2.1920	Tb927.4.810
6/7	Tb927.7.3250 Tb927.7.3260				
8	N/A				
10	Tb927.8.3620 Tb927.8.3630 Tb927.8.3650	Tb927.1.2820 Tb927.1.2850 Tb927.1.2880	Tb927.10.9080		
12	Tb927.7.7510				

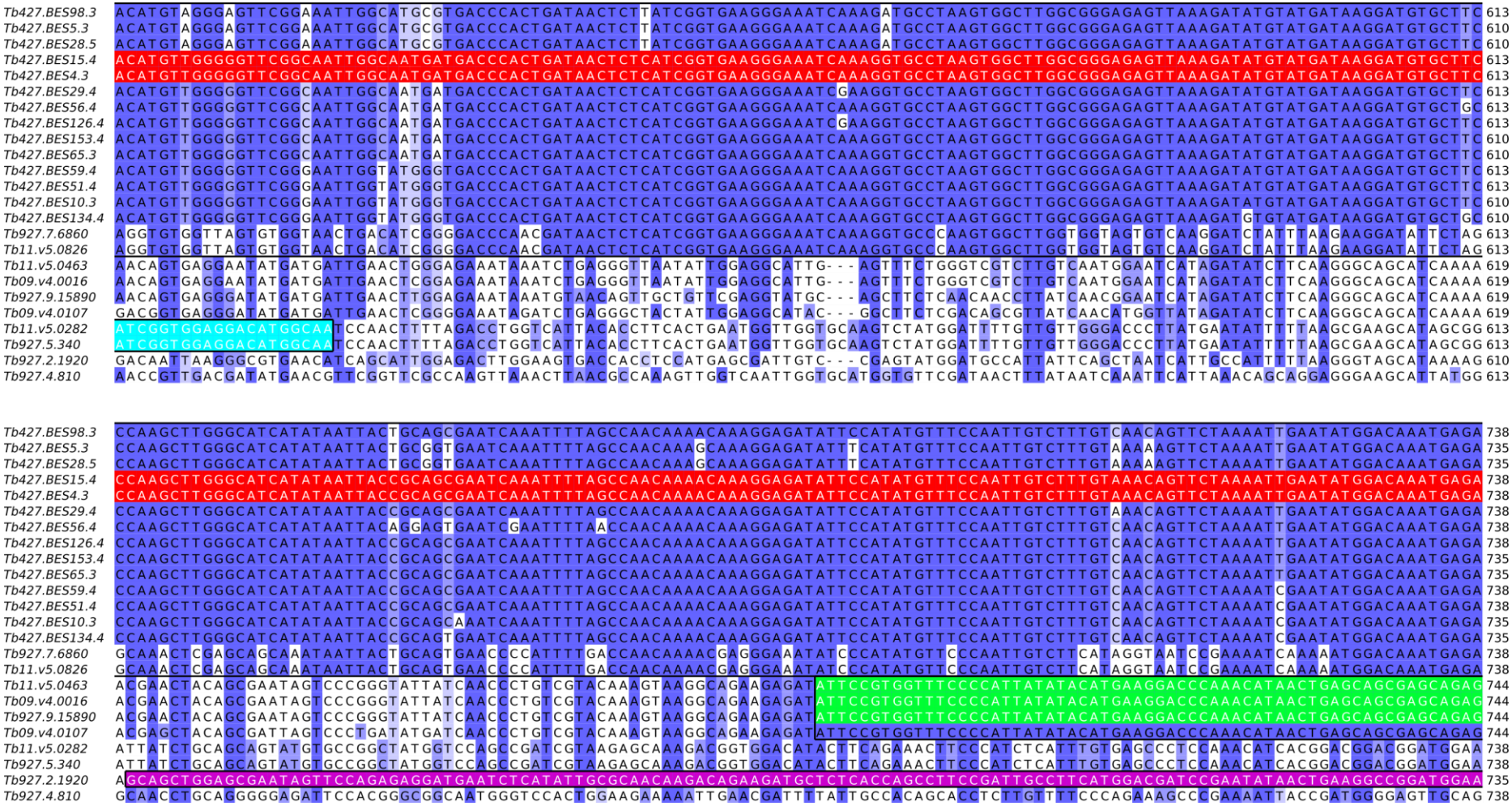
Pan-family RNAi fragment 1 targets all *ESAGs* of the indicated family as well as *GRESAG* genes in all except *(GR)ESAG8*. Greyed out cells indicate the lack of use of additional RNAi fragments.


Tb427.BES98.3 ATGTCATTGGTATCGAGCGCACCTGTTTTGGTGCTTTTCAATTCCGTTCTAACACTTCTGTGTTGTATCATCGGCTCTGGGCAACGGTCTAGTAGTCCCTTTACGGAAC---------CGGCACT 116
Tb427.BES5.3 ATGTCATTGGTATCGAGCGCACCCGTTTTGGTG---TTCATTTCCGTTGTAACACTTCTGTGTTGTATCATCGGCTCTGGGCAACGGTCTAGTAGTCCCTTTACGGAAC---------CGGCACT 113
Tb427.BES28.5 ATGTCATTGGTATCGAGCGCACCCGTTTTGGTG---TTCATTTCCGTTGTAACACTTCTGTGTTGTATCATCGGCTCTGGGCAACGGTCTAGTAGTCCCTTTACGGAAC---------CGGCACT 113
Tb427.BES15.4 ATGTCATCAGTATCGAGCGTACCTGTTTTGGTGCTCTTCAATTCCGTTCTAACACTTCTGTGTTGTATCATCGGCTCTGGGCAACGGTCTAGTAGTCCCTTTACGGAAC---------CGGCACT 116
Tb427.BES4.3 ATGTCATCAGTATCGAGCGTACCTGTTTTGGTGCTCTTCAATTCCGTTCTAACACTTCTGTGTTGTATCATCGGCTCTGGGCAACGGTCTAGTAGTCCCTTTACGGAAC---------CGGCACT 116
Tb427.BES29.4 ATGTCATTGGTATCGAGCGCACCCGTTTTGGTGCTATTCAATTCCGTTCTAACACTTCTGTGTTGTATCATCGGCTCTGGGCAACGGTCTAGTAGTCCCTTTACGGAAC---------CGGCACT 116
Tb427.BES56.4 ATGTCATCAGTATCGAGCGTACCGTTTTGGTGCTATTCAATTCCGTTCTAACACTTCTGTGTTGTATCATCGGCTCTGGGCAACGGTCTAGTAATCCCTTTACGGAAC---------CGGCACT 116
Tb427.BES126.4 ATGTCATTGGTATCGAGCGCACCCGTTTTGGTGCTATTCAATTCCGTTCTAACACTTCTGTGTTGTATCATCGGCTCTGGGCAACGGTCTAGTAATCCCTTTACGGAAC---------CGGCACT 116
Tb427.BES153.4 ATGTCATTGGTATCGAGCGCACCCGTTTTGGTG---TTCATTTCCGTTGTAACACTTCTGTGTTGCATCATCGGCTCTGGGCAACGGTCTAGTAGTCCCTTTACTGAAC---------AGGCACT 113
Tb427.BES65.3 ATGTCATTGGTATCGAGCGCACCCGTTTTGGTG---TTCATTTCCGTTGTAACACTTCTGTGTTGCATCATCGGCTCTGGGCAACGGTCTAGTAGTCCCTTTACTGAAC---------AGGCACT 113
Tb427.BES59.4 ATGTCATCAGTATCGAGCGTACCTGTTTTGGTGCTCCTCAATTCCGTTCCAACACTTCTGTGTTGTATCATCGGCTCTGGGCAACTGTCTAGTAGTCCCTTTACGGAAC---------CGGCAAT 116
Tb427.BES51.4 ATGTCATCAGTATCGAGCGTACCTGTTTTGGTGCTCCTCAATTCCGTTCCAACACTTCTGTGTTGTATCATCGGCTCTGGGCAACTGTCTAGTAGTCCCTTTACGGAAC---------CGGCAAT 116
Tb427.BES10.3 ATGTCATTGGTATCGAGCGCACCCGTTTTGGTG---TTCATTTCCGTTGTAACACTTCTGTGTTGTATCATCGGCTCTGGGCAACGGTCTAGTAGTCCCTTTACGGAAC---------CGGCACT 113
Tb427.BES134.4 ATGTCATTGGTATCGAGCGCACCCGTTTTGGTG---TTCATTTCCGTTGTAACACTTCTGTGTTGTATCATCGGCTCTGGGCAACGGTCTAGTAGTCCCTTTACGGAAC---------CGGACCT 113
Tb927.7.6860 ATGTCATCAGTATCGAGCGTACCTGTTTTGGTGCTGTTCAATTCCGTTCTAACACTTCTGTGTTGTATCATCAGCTCTGGGCAACGGTCAAGAAGTCCCTTTACGAGAC---------CGGCACT 116
Tb11.v5.0826 ATGTCATCAGTATCGAGCGTACCTGTTTTGGTGCTGTTCAATTCCGTTCTAACACTTCTGTGTTGTATCATCAGCTCTGGGCAACGGTCAAGAAGTCCCTTTACGAGAC---------CGGCACT 116
Tb11.v5.0463 ATGCAACGAGCTGCAACTCTAATATATCACGGTAATTTTTGTCGCCGTGGTACTGGCGCCGTATTTTGTTGCAGCATCCGGTCAGGTTCACGGCATTCAGTATGGAGAAG---CGCCCGTGTCTTC 122
Tb09.v4.0016 ATGCAACGAGCTGCAACTCTAATATATCACGGTAATTTTTGTCGCCGTGGTACTGGCGCCGTATTTTGTTGCAGCATCCGGTCAGGTTCACGGCATTCAGTATGGAGAAG---CGCCCGTGTCTTC 122
Tb927.9.15890 ATGCAACGAGCTGCAACTCTAATATATCACGGTAATTTTTGTCGCCGTGGTACTGGCGCCGTATTTTGTTGCAGCATCCGGTCAGGTTCACGGCATTCAGTATGGAGAAG---CGCCCGTGTCTTC 122
Tb09.v4.0107 ATGCAACGAGCTGCAACTCTAATATACACGGTAATTTTTGCCGCCGTGGTACTGGCGCCGTATTGTATTGCAGCATCCGGTGAGGTTCACGGCATTCAGTATGGAGAAG---CGCCCGTGTCTTC 122
Tb11.v5.0282 ATGCCACGAAGATTTGCCTTTGCACGCTCAGCAACCCTGTGCGTTTTCATACTGTTGTCTCTCCTCACGGAGGCATTCCCTGAGTATGTAGGTGGGAAGATTGGACAAACACCACCCAGTGATTT 125
Tb927.5.340 ATGCCACGAAGATTTGCCTTTGCACGCTCAGCAACCCTGTGCGTTTTCATACTGTTGTCTCTCCTCACGGAGGCATTCCCTGAGTATGTAGGTGGGAAGATTGGACAAACACCACCCAGTGATTT 125
Tb927.2.1920 ATGCCACGCGTGATTATTGTGGCATACGTGTTGCTGCTCAGCGTTTCATTATCATCTCTGTGCCGCGCGA------CTTCAGAGCAGTTCATTAGGGCCATCGGGAATT---CACACGAGTCATT 116
Tb927.4.810 ATGCCGAAGTTGTTCCTTTTGCTGTATGTGACACTGTTCTACCATGTGTCACAGTCATTGGGTCGCATCATTGCCCTTCCGACGGGGAGCGACGCTGGCTCCATTGAACTCTCTGGTGAACCATT 125

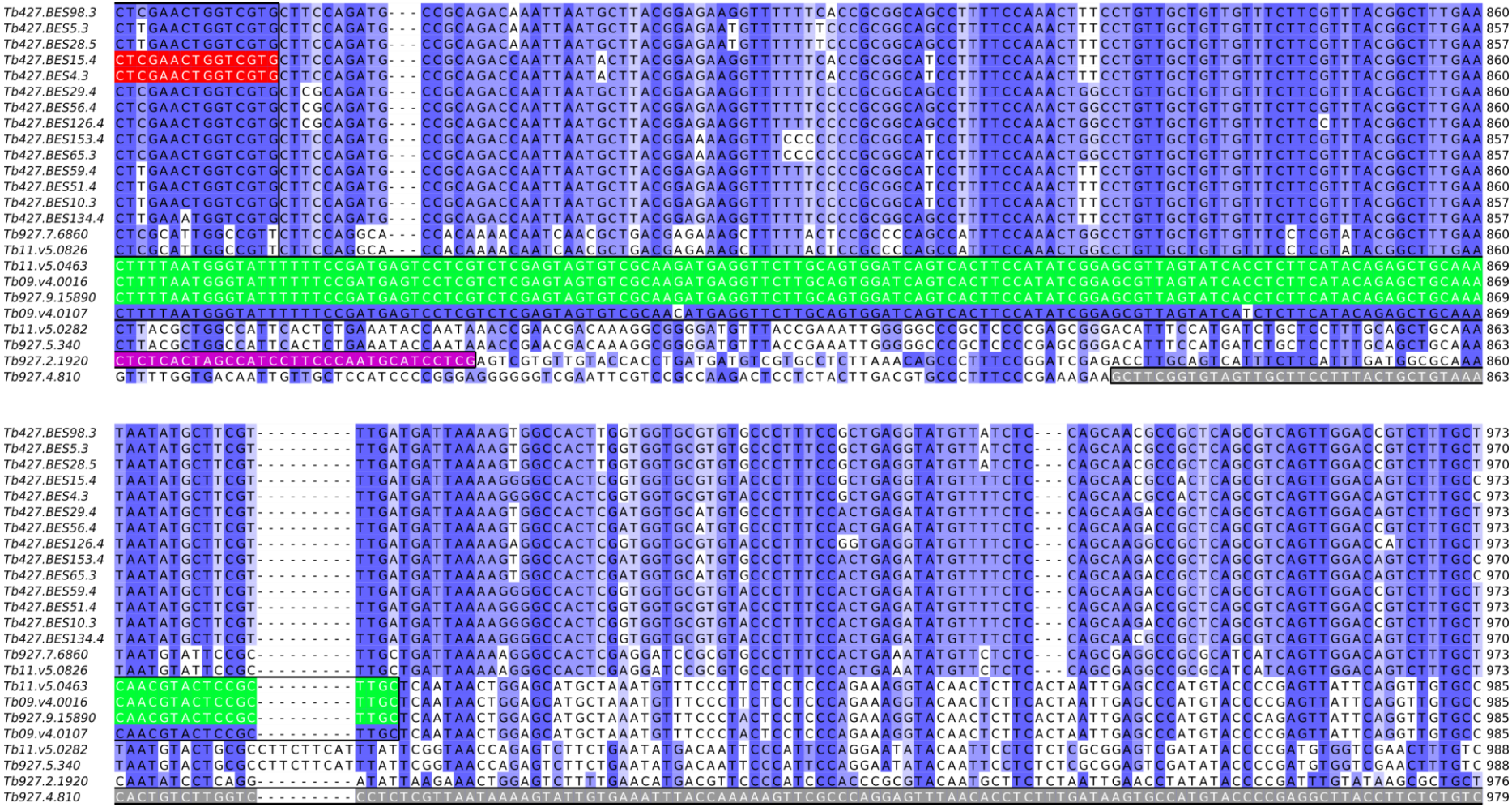
Tb427.BES98.3 CAAACCAGGGGGCATTAAGGTTGCTATTCAAGAGGCGGCTGTTACACCATTGTTACCAGTACTTGCGTATGAGTTCGAAAGGTTTATGGAGAATATCACCATTCCCGAGCAAAAGGTTAACCGTG 241
Tb427.BES5.3 CAAACCAGGGGGCGTTAAGGTTGCTATTCAAGAGGCGGCTGTTACGCCATTGTTACCAGTACTTGCGTATGAGTTCGAAAGGTTTATGGAGAATATCACCATTCCCGCGCAAAAGGTTAACCGTG 238
Tb427.BES28.5 CAAACCAGGGGGCGTTAAGGTTGCTATTCAAGAGGCGGCTGTTACGCCATTGTTACCAGTACTTGCGTATGAGTTCGAAAGGTTTATGGAGAATATCACCATTCCCGCGCAAAAGGTTAACCGTG 238
Tb427.BES15.4 CAAACCAGGGGGCGTTAAGGTTGCTATTCAAGAGGCGGCTGTTACGCCATTATTACCAGCACTTGCGGATGAGTTCGAAAGGTTTATGGAGAATATCACCATTCCTGGGCAAAAGGTTAACGGTG 241
Tb427.BES4.3 CAAACCAGGGGGCGTTAAGGTTGCTATTCAAGAGGCGGCTGTTACGCCATTATTACCAGCACTTGCGGATGAGTTCGAAAGGTTTATGGAGAATATCACCATTCCTGGGCAAAAGGTTAACGGTG 241
Tb427.BES29.4 CAAAGCAGGGGGCGTTAAGGTTGCTATTCAAGAGGCGGCTCTTGCGCCATTATTACCAGCACTTGCGGATGAGTTCGAAAGGTTTATGGAGAATATCACCATTCCTTGAGCAAAAGGTTAACGGTG 241
Tb427.BES56.4 TAAAGCAGGGGGCGTTAAGGTTGCTATTCAAGAGGCGGCTCTTGCGCCATTATTACCAGCACTTGCGGATGAGTTCGAAAGGTTTATGGAGAGTATCACCATTCCTGAGCAAAAGGTTAACGGTG 241
Tb427.BES126.4 CAAAGCAGGGGGCGTTAAGGTTGCTATTCAAGAGGCGGCTCTTGCGCCATTATTACCAGCACTTGCGGATGAGTTCGAAAGGTTTATGGAGAATATCACCATTCCTGAGCAAAAGGTTAACGGTG 241
Tb427.BES153.4 CAAAGCAGGGGGCGTTAAGGTTGCTATTCAAGAGGCGGCTCTTGCGCCATTATTACCAGCACTTGCGGATGAGTTCGAAAGGTTTATGGAGAATATCACCATTCCTGAGCAAAAGGTTAACGGTG 238
Tb427.BES65.3 CAAAGCAGGGGGCGTTAAGGTTGCTATTCAAGAGGCGGCTCTTGCGCCATTATTACCAGCACTTGCGGATGAGTTCGAAAGGTTTATGGAGAATATCACCATTCCTGAGCAAAAGGTTAACGGTG 238
Tb427.BES59.4 CAAAGCAGGGGGCGTTAAGGTTTCTATTCAAGAGGCGGCTGTTACGCCATTGTTACCAGCACTTGCGGATGAGTTCGAAAGGTTTATGGAGAATATCACCATTCCTGGGCAAAAGGTTAACGGTG 241
Tb427.BES51.4 CAAAGCAGGGGGCGTTAAGGTTTCTATTCAAGAGGCGGCTGTTACGCCATTGTTACCAGCACTTGCGGATGAGTTCGAAAGGTTTATGGAGAATATCACCATTCCTGGGCAAAAGGTTAACGGTG 241
Tb427.BES10.3 CAAATCAGGGGGCGTTAAGGTTGCTATTCAAGAGGCGGCTCTTGCGCCATTATTACCAGCACTTGCGGATGAGTTCGAAAGGTTTATGGAGAATATCACCATTCCTGAGCAAAAGGTTAACGGTG 238
Tb427.BES134.4 CAAACCAGGGGGCGTTAAGGTTGCTATTCAAGAGGCGGCTCTTGCGCCATTATTACCAGCACTTGCGGATGAGTTCGAAAGGTTTATGGAGAATATCACCATTCCTGGGCAAAAGGTTAACGGTG 238
Tb927.7.6860 CAAACCAGGGGGCGTTAAGGTTGCCATTCAAGAAGCGGCCGTTACGCCAGCATTACCCACACTTGTGGAAGAGTCGGAAAAGGTTATGGAGAATCTCACCATTCCCGAGCAAAAGGTTAATGGTC 241
Tb11.v5.0826 CAAACCAGGGGGCGTTAAGGTTGCCATTCAAGAAGCGGCCGTTACGCCAGCATTACCCACACTTGTGGAAGAGTCGGAAAAGGTTATGGAGAATCTCACCATTCCCGAGCAAAAGGTTAATGGTC 241
Tb11.v5.0463 CGCACCCGGTAGCGTGAAATTTGCCGTTGATGAATCCGTCGTGACTTCTGCTGTGCCCATGGTTCACCGCGTGCTGAGACAACTTATCGGAAACATTACTGTTCCTGAGCAAAGTGTGCAAGGTA 247
Tb09.v4.0016 CGCACCCGGTAGCGTGAAATTTGCCGTTGATGAATCCGTCGTGACTTCTGCTGTGCCCATGGTTCACCGCGTGCTGAGACAACTTATCGGAAACATTACTGTTCCTGAGCAAAGTGTGCAAGGTA 247
Tb927.9.15890 CGCACCCGGTAGCGTGAAATTTGCCGTTGATGAATCCGTCGTGACTTCTGCTGTGCCCATGGTTCACCGCGTGCTGAGACAACTTATCGGAAACATTACTGTTCCTGAGCAAAGTGTGCAAGGTA 247
Tb09.v4.0107 CGCACCCGGTAGCATGAAATTCGCCGTTGATGAATCCGTTGTGACTTCTGCTGTGCCCATGGTTCACCGCGTGCTGAGACAACTCATCGAAAATATTACTGTTCCTGCGCAAAGTGTGCAATGTA 247
Tb11.v5.0282 CCCTCCAGGAAACGTTAGGTTCGGCATAGACGATGCGGCAATCGCCCCGGTGGTGCCCCTAATCCTCAAAACACTGAAGAGCTTCATTGGAAGTGTCACAGTCCCTGAGCAAAGGTTACACGGAG 250
Tb927.5.340 CCCTCCAGGAAACGTTAGGTTCGGCATAGACGATGCGGCAATCGCCCCGGTGGTGCCCCTAATCCTCAAAACACTGAAGAGCTTCATTGGAAGTGTCACAGTCCCTGAGCAAAGGTTACACGGAG 250
Tb927.2.1920 AGTTCCTGGAAGCGCTAAAGTTTCTATTGATGATGCGGCGTTTCTACCGATTATACCGGTGGTTCTTAAGAGTTTGAAGAACCTCACTGGGAATATTAAAGTCCCTCGGCAGAAAATAGGGGGTT 241
Tb927.4.810 TCCCTCTGGGGACGTCGCACTGGTAGTCAACGACACGGCCATTGCCAACATCATGCCTACCGTAGTTAAGATGGTAAATAATATGCTCGGTAAGTACACTATTCCGTCACGGGAGGTAAATGGCA 250

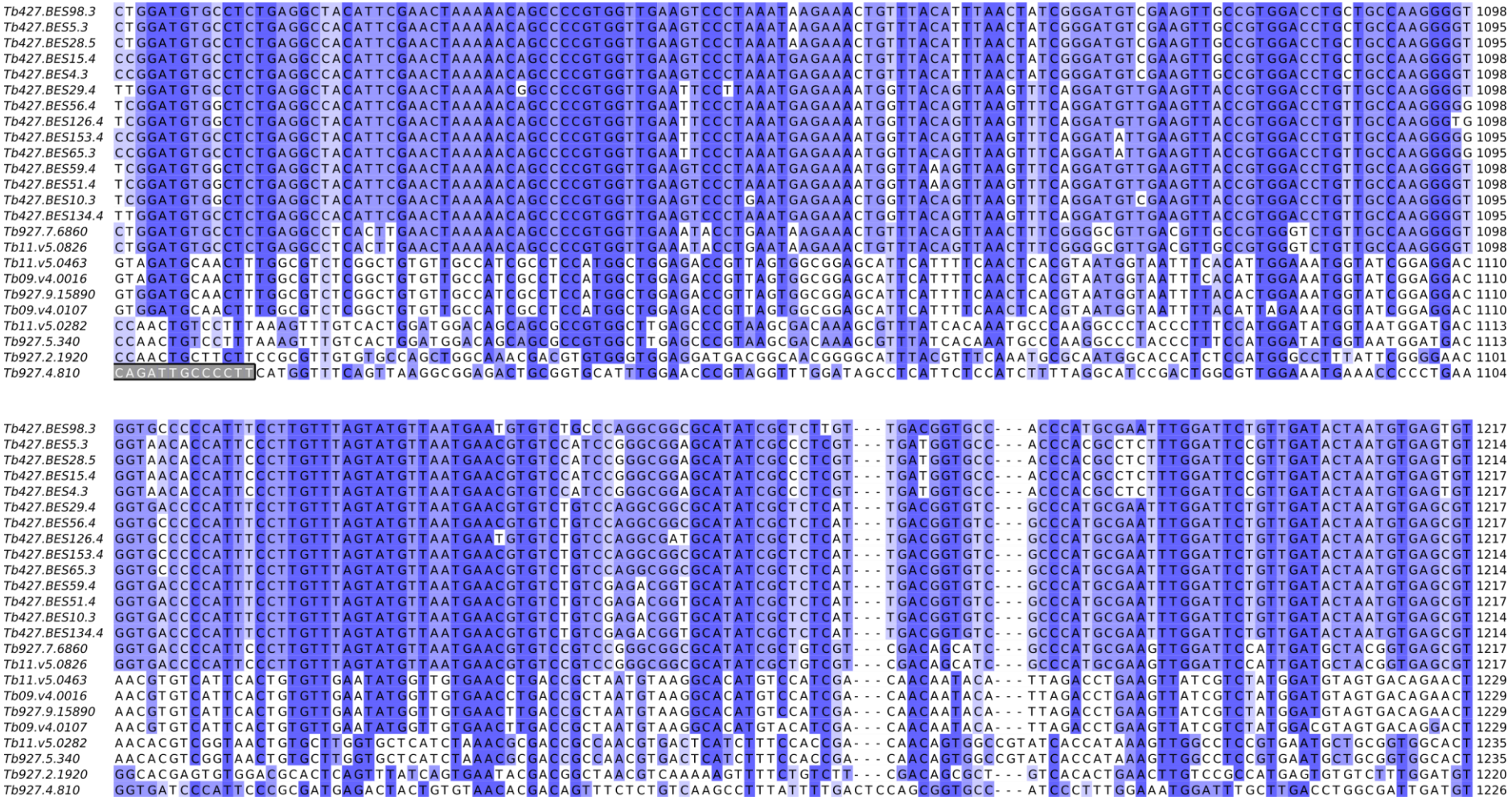

Tb427.BES98.3 TGAGTGTAGAAGAGACTTATTTTCGTAACGTGACGGTCGGAAGCGCGACAGTTAAATTCGAAGAACCAAATAAAATCATTTTGAAGTTTTGGAACGTGAGCGCTACTGTGCCGTTTACGCGGTTT 366
Tb427.BES5.3 TGAGTGTAGAAGAGACTTATTTTCGTAACGTGACGGTCGGAACCGCGACAGTTAAATTCGAAGAACCAAATAAAATCATTTTGAAGTTTTGGAACGTGAGCGCTACTGTGCCGTTTACGCGGTTT 363
Tb427.BES28.5 TGAGTGTAGAAGAGACTTATTTTCGTAACGTGACGGTCGGAACCGCGACAGTTAAATTCGAAGAACCAAATAAAATCATTTTGAAGTTTTGGAACGTGAGCGCTACTGTGCCGTTTACGCGGTTT 363
Tb427.BES15.4 TGAGTGTAGAAGAGACTTATTTTCGTAACGTGACAGTTGGAAGCGCTAAAGTTCAATTTTGGGGATCAAATAAAATGGTTTGGAACTTTTGTAACGTGAGTGCTACTGTGCCGTTTACGCGGTTT 366
Tb427.BES4.3 TGAGTGTAGAAGAGACTTATTTTCGTAACGTGACAGTTGGAAGCGCTAAAGTTCAATTTTGGGGATCAAATAAAATGGTTTGGAACTTTTGTAACGTGAGTGCTACTGTGCCGTTTACGCGGTTT 366
Tb427.BES29.4 TGAGTGTAGAAGAGACTTATTTTCGTAACGTGACGGTCGGAAGCGCGACAGTTAAATTTTGGGGATCAAACAAAATGGTTTGGAACTTTTGTAACGTGAGCGCTACTGTTCCGTTTACACGGTTT 366
Tb427.BES56.4 TGAGTGTAGAAGAGACTTATTTTCGTAACGTGACGGTCGGAAGCGCGACAGTTAAATTTTGGGGATCAAACAAAATGGTTTGGAACTTTTGTAACGTGAGCGCTACTGTTCCGTTTACACGGTTT 366
Tb427.BES126.4 TGAGTGTAGAAGAGACTTATTTTCGTAACGTGACGGTCGGAAGCGCGACAGTTAAATTTTGGGGATCAAACAAAATGGTTTGGAACTTTTGTAACGTGAGCGCTACTGTTCCGTTTACACGGTTT 366
Tb427.BES153.4 TGAGTGTAGAAGAGACTTATTTTCGTAACGTGACGGTCGGAAGCGCGACAGTTAAATTTTGGGGATCAAACAAAATGGTTTGGAACTTTTGTAACGTGAGCGCTACTGTTCCGTTTACACGGTTT 363
Tb427.BES65.3 TGAGTGTAGAAGAGACTTATTTTCGTAACGTGACGGTCGGAAGCGCGACAGTTAAATTTTGGGGATCAAACAAAATGGTTTGGAACTTTTGTAACGTGAGCGCTACTGTTCCGTTTACACGGTTT 363
Tb427.BES59.4 TGAGTGTAGAAGAGACTTATTTTCGTAACGTGACAGTTGGAAGCGCTAAAGTTCAATTTTGGGGATCAAATAAAATGGTTTGGAACTTTTGTAACGTGAGTGCTACTGTTCCGTTTACGCGGTTT 366
Tb427.BES51.4 TGAGTGTAGAAGAGACTTATTTTCGTAACGTGACAGTTGGAAGCGCTAAAGTTCAATTTTGGGGATCAAATAAAATGGTTTGGAACTTTTGTAACGTGAGTGCTACTGTTCCGTTTACGCGGTTT 366
Tb427.BES10.3 TGAGTGTAGAAGAGACTTATTTTCGTAACGTGACAGTTGGAAGCGCGACAGTTAAATTTTGGGGATCAAACAAAATGGTTTGGAACTTTTGTAACGTGAGTGCTACTGTGCCGTTTACGCGGTTT 363
Tb427.BES134.4 TGAGTGTAGAAGAGACTTATTTTCGTAACGTGACAGTTGGAAGCGCTAAAGTTCAATTTTGGGGATCAAATAAAATGGTTTGGAACTTTTGTAACGTGAGTGCTACTGTGCCGTTTACGCGGTTT 363
Tb927.7.6860 TAAGTCTAGGGGAGGCTCATATTCGGACGTGACGGTTGGGAGCGCGACAGTTAAATTCGAGGAACCAAATAAAATGATTTTGAAGTTTTGGAACGTAAGCGCTACTGTGCCTTTTACGCGTTTT 366
Tb11.v5.0826 TAAGTCTAGGGGAGGCTCATATTCGGACGTGACGGTTGGGAGCGCGACAGTTAAATTCGAGGAACCAAATAAAATGATTTTGAAGTTTTGGAACGTAAGCGCTACTGTGCCTTTTACGCGTTTT 366
Tb11.v5.0463 TTAAGCTGGAAGAAATGCAGGTATCAAACATAACGATTGGAAACATGTCCCTCAAACTGAAGAGCCCGAATAAACTTTTCGTTTCGGTGTGGGACATGGCAGCCAATGTACCAGAAACTAGGTTT 372
Tb09.v4.0016 TTAAGCTGGAAGAAATGCAGGTATCAAACATAACGATTGGAAACATGTCCCTCAAACTGAAGAGCCCGAATAAACTTTTCGTTTCGGTGTGGGACATGGCAGCCAATGTACCAGAAACTAGGTTT 372
Tb927.9.15890 TTAAGCTGGAAGAAATGCAGGTATCAAACATAACGATTGGAAACATGTCCCTCAAACTGAAGAGCCCGAATAAACTTTTGGTTTCGGTGTGGGACATGGCAGCCAGTGTGCCAAAAACTAAGTTT 372
Tb09.v4.0107 TTAAACTGGAAGAAATGCAGGTATCAAACATAACGATTGGAAACATGTCCGTCAAACTAAAGAGCCCGAATAAACTTTTCGTTTCGGTGTGGGAAATGGCAGCCAGTGTACCAGAAACTAAGTTT 372
Tb11.v5.0282 TGAGTTTGGAAACGATGCAAGTGCAAGACATAAACATTGGAAATATGTCGCTACACATGAATAGTACGAACAGGGTTGATGTATCGGTATGGAACATGACGGCTAATGTCCCAGAAACCAACTTC 375
Tb927.5.340 TGAGTTTGGAAACGATGCAAGTGCAAGACATAAACATTGGAAATATGTCGCTACACATGAATAGTACGAACAGGGTTGATGTATCGGTATGGAACATGACGGCTAATGTCCCAGAAACCAACTTC 375
Tb927.2.1920 TGGAACTTGGAGAGGCAGAAGTGCATAATATTACGGCTGGGGGCATGTCTATTAATATGGAAGAGTCAAAGCGGGTCGTTGTTGCGGTTTGGGACATGAGTGCAGTTGTGCCAGAGTTAAACTTC 366
Tb927.4.810 TTCAGTTGGGTGATATTCCCATTACTGGCATAAAAATCGGAAACATTTCCCTCGAAATAAAGCATCCGAACAAGCTTGTGTTGAATGTGACTGATATCGTTGCAGACGCGGAGGACACAACCTTC 375

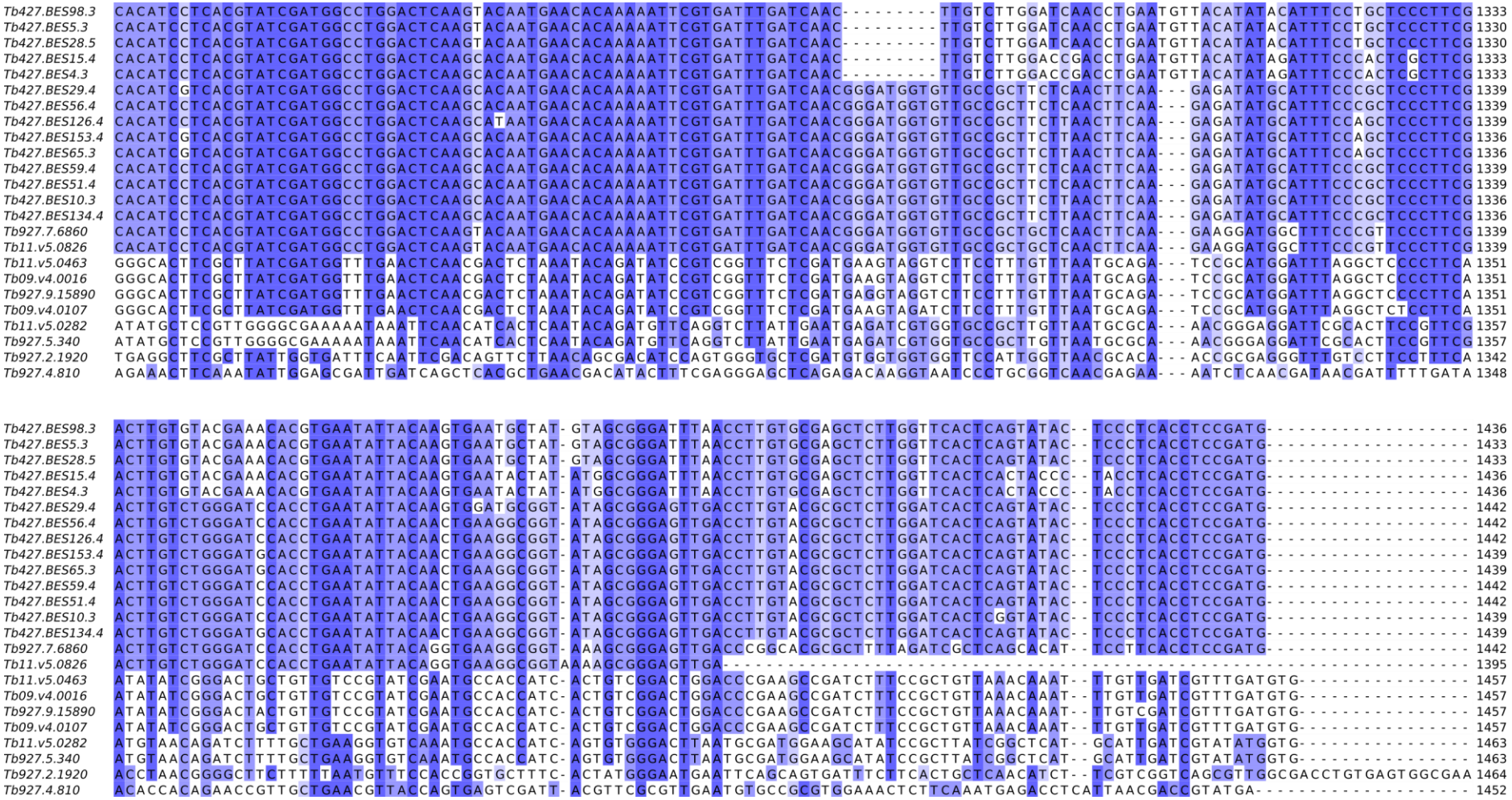
Tb427.BES98.3 GTATACCACTCGTTTTGGTGTTACCTGTATCCTTG---CAGTGGGTCGGCACAGGCGGAAATACGCAACGGCAGTGTGGCACTGTGGCTCGACGTGTCCGCAGGAAGGGGGGGGACTACTAGACAT 488
Tb427.BES5.3 GTATACCACTCGTTTTGGTGTTACCTGTATCCTTG---CAGTGGGTCGGCACAGGCGGAAATACGCAACGGCAGTGTGGCACTGTGGCTCGACGTGTCCGCAGGAAGGGGGGGGACTACTAGACAT 485
Tb427.BES28.5 GTATACCACTCGTTTTGGTGTTACCTGTATCCTTG---CAGTGGGTCGGCACAGGCGGAAATACGCAACGGCAGTGTGGCACTGTGGCTCGACGTGTCCGCAGGAAGGGGGGGGACTACTAGACAT 485
Tb427.BES15.4 GTATACCACTCGTTTTGGTGTTACCTGTATCCTTG---CAGTGGGTCGGCACAGGTGGAAATACGCAACGCCAGTGTGGCACTGTGGCTCGACGTGTCCGCAACAAGGGGGGGGACTACTAGACAT 488
Tb427.BES4.3 GTATACCACTCGTTTTGGTGTTACCTGTATCCTTG---CAGTGGGTCGGCACAGGTGGAAATACGCAACGCCAGTGTGGCACTGTGGCTCGACGTGTCCGCAACAAGGGGGGGGACTACTAGACAT 488
Tb427.BES29.4 GTATACCACTCGTTTTGGTGTTACCTGTATCCTTG---CAGTGGGTCGGCACAGGCGGAAATACGCAACGCCAGTGTGGCACTGTGGCTGGACGTGTCCGCAGGAAGGGGGGGGACTACTAGACAT 488
Tb427.BES56.4 GTATACCACTCGTTTTGGTGTTACCTGTATCCTTG---CAGTGGGTCGGCACAGGCGGAAATACGCAACGCCAGTGTGGCACTGTGGCTGGACGTGTCCGCAGGAAGGGGGGGACTACTAGACAT 488
Tb427.BES126.4 GTATACCACTCGTTTTGGTGTTACCTGTATCCTTG---CAGTGGGTCGGCACAGGCGGAAATACGCAACGCCAGTGTGGCACTGTGGCTGGACGTGTCCGCAGGAAGGGGGGGACTACTAGACAT 488
Tb427.BES153.4 GTATACCACTCGTTTTGGTGTTACCTGTATCCTTG---CAGTGGGTCTGCACAGGCGGAAATACGCAACGCCAGTGTGGCACTGTGGCTGGACGTGTCCGCAGGAAGGGGGGGACTACTAGACAT 485
Tb427.BES65.3 GTATACCACTCGTTTTGGTGTTACCTGTATCCTTG---CAGTGGGTCTGCACAGGCGGAAATACGCAACGCCAGTGTGGCACTGTGGCTGGACGTGTCCGCAGGAAGGGGGGGACTACTAGACAT 485
Tb427.BES59.4 GTATACCACTCGTTTTGGTGTTACCTGTATCCTTG---CAGTGGGTCGGCACAGGCGGAAATACGCAACGCCAGTGTGGCACTGTGGCTCGACGTGTCCGCACGAAGGGGGGGACTACTAGACAT 488
Tb427.BES51.4 GTATACCACTCGTTTTGGTGTTACCTGTATCCTTG---CAGTGGGTCGGCACAGGCGGAAATACGCAACGCCAGTGTGGCACTGTGGCTCGACGTGTCCGCACGAAGGGGGGGACTACTAGACAT 488
Tb427.BES10.3 GTATACCACTCGTTTTGGTGTTACCTGTATCCTTG---CAGTGGGTCGGCACAGGCGGAAATACGCAACGCCAGTGTGGCACTGTGGCTGGACGTGTCCGCAGGAAGGGGGGGGAATACTAGACAT 485
Tb427.BES134.4 GTATACCACTCGTTTTGGTGTTACATGTATCCTTG---CAGTGGGTCGGCACAGGCGGAAATACGCAACGCCAGTGTGGCACTGTGGCTCGACGTGTCCGCACGAAGGGGGGGACTACTAGACAT 485
Tb927.7.6860 TCATACCACTCTTTTTGGTGTCATGTATATCCTTG---CAGTGGGACGACACAAATGGAAATACGCAACGCAAGTATGACACTGTTGCTCGAAGTGTCTGCTGCAAGCGGCGGACCACTAAACGT 488
Tb11.v5.0826 TCATACCACTCTTTTTGGTGTCATGTATATCCTTG---CAGTGGGACGACACAAATGGAAATACGCAACGCAAGTATGACACTGTTGCTCGAAGTGTCTGCTGCAAGCGGCGGACCACTAAACGT 488
Tb11.v5.0463 TACTACTGTTTCTTTTTGTGTTCACTGCTACCGTACATAAGGGGTACATCAGACACCGATATACGCGGTGCGAATGCATCGTTCACTCTGGATTTATCCGCTAGAAATGATGGTCTGCTCAATGC 497
Tb09.v4.0016 TACTACTGTTTCTTTTTGTGTTCACTGCTACCGTACATAAGGGGTACATCAGACACCGATATACGCGGTGCGAATGCATCGTTCACTCTGGATTTATCCGCTAGAAATGATGGTCTGCTCAATGC 497
Tb927.9.15890 TATTACAGTCCACTTTGTTGGTGGCCGCTTCTGTGCCCTAGTGGCACGTCCGCTACCGATATCCGCGGTGCGAATGCATCGTTGCTATGGACTTATCCGCTAGAAACGATGGTCTGCTCAATGC 497
Tb09.v4.0107 TATTACAGTCCAGTTTGTTGGTGGTTGTTTCTGTGCCCTAGTGGCACGTCCGCTACCGATATCCGCGGTGCGAATGCATCGTTCACTCTGGACTTATCCGTTAGAAATGATGGTCTGCTCAATGT 497
Tb11.v5.0282 AGCTACAATCT----------GGTGTTTGTGTCGTGCAGTGGAACGGCGAAAACCGATGTTCGCAATGCCAACGTGTCGTTTGCAATGCGTTTGTCCATCGGAGTGGGTGGCATCCTCGATGT 488
Tb927.5.340 AGCTACAATCT----------GGTGTTTGTGTCGTGCAGTGGAACGGCGAAAACCGATGTTCGCAATGCCAACGTGTCGTTTGCAATGCGTTTGTCCATCGGAGTGGGTGGCATCCTCGATGT 488
Tb927.2.1920 ACATACACCTACTGGCTTTTCCAGTGGCAGACGTG---TAATGGAACAGCGCGGACCGAGGTCAGCGGCACGAACTCAACTCTATATTTCAATCTGACCATTGGATCTGACGGATTGGTCAACGA 488
Tb927.4.810 AGCAAGAGTTT----------GGTGATTACCTCATGTAGAGGGAAGGTCCGAGCATCAGTTGCGGTTTCTAGCGCTTCCCTTGTGTTGGGCCTTTCCGTAGATGATGACGAAAAACTGAAGAT 488

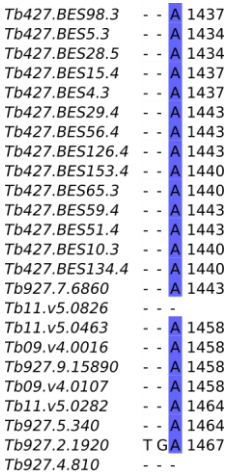

Tb427.BES98.3 ACATGTAGGGAGTTCGGAAATTGGCATGCGTGACCCACTGATAACTCTTATCGGTGAAGGGAAATCAAAGATGCCTAAGTGGCTTGGCGGGAGAGTTAAAGATATGTATGATAAGGATGTGCTTC 613
Tb427.BES5.3 ACATGTAGGGAGTTCGGAAATTGGCATGCGTGACCCACTGATAACTCTTATCGGTGAAGGGAAATCAAAGATGCCTAAGTGGCTTGGCGGGAGAGTTAAAGATATGTATGATAAGGATGTGCTTC 610
Tb427.BES28.5 ACATGTAGGGAGTTCGGAAATTGGCATGCGTGACCCACTGATAACTCTTATCGGTGAAGGGAAATCAAAGATGCCTAAGTGGCTTGGCGGGAGAGTTAAAGATATGTATGATAAGGATGTGCTTC 610
Tb427.BES15.4 ACATGTTGGGGGTTCGGCAATTGGCAATGATGACCCACTGATAACTCTCATCGGTGAAGGGAAATCAAAGGTGCCTAAGTGGCTTGGCGGGAGAGTTAAAGATATGTATGATAAGGATGTGCTTC 613
Tb427.BES4.3 ACATGTTGGGGGTTCGGCAATTGGCAATGATGACCCACTGATAACTCTCATCGGTGAAGGGAAATCAAAGGTGCCTAAGTGGCTTGGCGGGAGAGTTAAAGATATGTATGATAAGGATGTGCTTC 613
Tb427.BES29.4 ACATGTTGGGGGTTCGGCAATTGGCAATGATGACCCACTGATAACTCTCATCGGTGAAGGGAAATCGAAGGTGCCTAAGTGGCTTGGCGGGAGAGTTAAAGATATGTATGATAAGGATGTGCTTC 613
Tb427.BES56.4 ACATGTTGGGGGTTCGGCAATTGGCAATGATGACCCACTGATAACTCTCATCGGTGAAGGGAAATCAAAGGTGCCTAAGTGGCTTGGCGGGAGAGTTAAAGATATGTATGATAAGGATGTGCTGC 613
Tb427.BES126.4 ACATGTTGGGGGTTCGGCAATTGGCAATGATGACCCACTGATAACTCTCATCGGTGAAGGGAAATCGAAGGTGCCTAAGTGGCTTGGCGGGAGAGTTAAAGATATGTATGATAAGGATGTGCTTC 613
Tb427.BES153.4 ACATGTTGGGGGTTCGGCAATTGGCAATGATGACCCACTGATAACTCTCATCGGTGAAGGGAAATCAAAGGTGCCTAAGTGGCTTGGCGGGAGAGTTAAAGATATGTATGATAAGGATGTGCTTC 610
Tb427.BES65.3 ACATGTTGGGGGTTCGGCAATTGGCAATGATGACCCACTGATAACTCTCATCGGTGAAGGGAAATCAAAGGTGCCTAAGTGGCTTGGCGGGAGAGTTAAAGATATGTATGATAAGGATGTGCTTC 610
Tb427.BES59.4 ACATGTTGGGGGTTCGGGAATTGGTATGGGTGACCCACTGATAACTCTCATCGGTGAAGGGAAATCAAAGGTGCCTAAGTGGCTTGGCGGGAGAGTTAAAGATATGTATGATAAGGATGTGCTTC 613
Tb427.BES51.4 ACATGTTGGGGGTTCGGGAATTGGTATGGGTGACCCACTGATAACTCTCATCGGTGAAGGGAAATCAAAGGTGCCTAAGTGGCTTGGCGGGAGAGTTAAAGATATGTATGATAAGGATGTGCTTC 613
Tb427.BES10.3 ACATGTTGGGGGTTCGGGAATTGGTATGGGTGACCCACTGATAACTCTCATCGGTGAAGGGAAATCAAAGGTGCCTAAGTGGCTTGGCGGGAGAGTTAAAGATATGTATGATAAGGATGTGCTTC 610
Tb427.BES134.4 ACATGTTGGGGGTTCGGGAATTGGTATGGGTGACCCACTGATAACTCTCATCGGTGAAGGGAAATCAAAGGTGCCTAAGTGGCTTGGCGGGAGAGTTAAAGATGTGTATGATAAGGATGTGCTGC 610
Tb927.7.6860 AGGTGTGGTTAGTGTGGTAACTGACATCGGGGACCCAACGATAACTCTCATCGGTGAAGGGAAATCAAAGGTGCCCAAGTGGCTTGGTGGTAGTGTCAAGGATCTATTTAAGAAGGATATTCTAG 613
Tb11.v5.0826 AGGTGTGGTTAGTGTGGTAACTGACATCGGGGACCCAACGATAACTCTCATCGGTGAAGGGAAATCAAAGGTGCCCAAGTGGCTTGGTGGTAGTGTCAAGGATCTATTTAAGAAGGATATTCTAG 613
Tb11.v5.0463 AACAGTGAGGAATATGATGATTGAACTGGGAGAAATAAATCTGAGGGTTAATATTGGAGGCATTG---AGTTTCTGGGTCGTCTTGTCAATGGAATCATAGATATCTTCAAGGGCAGCATCAAAA 619
Tb09.v4.0016 AACAGTGAGGAATATGATGATTGAACTCGGAGAAATAAATCTGAGGGTTAATATTGGAGGCATTG---AGTTTCTGGGTCGTCTTGTCAATGGAATCATAGATATCTTCAAGGGCAGCATCAAAA 619
Tb927.9.15890 AACAGTGAGGGATATGATGATTGAACTTGGAGAAATAAATGTAACAGTTGCTGTTCGAGGTATGC---AGCTTCTCAACAACCTTATCAACGGAATCATAGATATCTTCAAGGGCAGCATCAAAA 619
Tb09.v4.0107 GACGGTGAGGGATATGATGATTGAACTCGGGGAAATAGATCTGAGGGCTACTATTGGAGGCATAC---GGCTTCTCGACAGCGTTATCAACATGGTTATAGATATCTTCAAGGGCAGCATCAAAA 619
Tb11.v5.0282 ATCGGTGGAGGACATGGCAATCCAACTTTTAGACCTGGTCATTACACCTTCACTGAATGGTTGGTGCAAGTCTATGGATTTTGTTGTTGGGACCCTTATGAATATTTTTAAGCGAAGCATAGCGG 613
Tb927.5.340 ATCGGTGGAGGACATGGCAATCCAACTTTTAGACCTGGTCATTACACCTTCACTGAATGGTTGGTGCAAGTCTATGGATTTTGTTGTTGGGACCCTTATGAATATTTTTAAGCGAAGCATAGCGG 613
Tb927.2.1920 GACAATTAAGGGCGTGAACATCAGCATTGGAGACTTGGAAGTGACCACCTCCATGAGCGATTGTC---CGAGTATGGATGCCATTATTCAGCTAATCATTGCCATTTTTAAGGGTAGCATAAAAG 610
Tb927.4.810 AACCGTTGACGATATGAACGTTCGGTTCGCCAAGTTAAACTTAACGCCAAAGTTGGTCAATTGGTGCATGGTGTTCGATAACTTTATAATCAAATTCATTAAACAGCAGGAGGGAAGCATTATGG 613

Tb427.BES98.3 CCAAGCTTGGGCATCATATAATTACTGCAGCGAATCAAATTTTAGCCAACAAAACAAAGGAGATATTCCATATGTTTCCAATTGTCTTTGTCAACAGTTCTAAAATTGAATATGGACAAATGAGA 738
Tb427.BES5.3 CCAAGCTTGGGCATCATATAATTACTGCGGTGAATCAAATTTTAGCCAACAAAGCAAAGGAGATATTTCATATGTTTCCAATTGTCTTTGTAAAAAGTTCTAAAATTGAATATGGACAAATGAGA 735
Tb427.BES28.5 CCAAGCTTGGGCATCATATAATTACTGCGGTGAATCAAATTTTAGCCAACAAAGCAAAGGAGATATTTCATATGTTTCCAATTGTCTTTGTAAAAAGTTCTAAAATTGAATATGGACAAATGAGA 735
Tb427.BES15.4 CCAAGCTTGGGCATCATATAATTACCGCAGCGAATCAAATTTTAGCCAACAAAACAAAGGAGATATTCCATATGTTTCCAATTGTCTTTGTAAACAGTTCTAAAATTGAATATGGACAAATGAGA 738
Tb427.BES4.3 CCAAGCTTGGGCATCATATAATTACCGCAGCGAATCAAATTTTAGCCAACAAAACAAAGGAGATATTCCATATGTTTCCAATTGTCTTTGTAAACAGTTCTAAAATTGAATATGGACAAATGAGA 738
Tb427.BES29.4 CCAAGCTTGGGCATCATATAATTACCGCAGCGAATCAAATTTTAGCCAACAAAACAAAGGAGATATTCCATATGTTTCCAATTGTCTTTGTAAACAGTTCTAAAATTGAATATGGACAAATGAGA 738
Tb427.BES56.4 CCAAGCTTGGGCATCATATAATTACAGGAGTGAATCGAATTTTAACCAACAAAACAAAGGAGATATTCCATATGTTTCCAATTGTCTTTGTCAACAGTTCTAAAATTGAATATGGACAAATGAGA 738
Tb427.BES126.4 CCAAGCTTGGGCATCATATAATTACCGCAGCGAATCAAATTTTAGCCAACAAAACAAAGGAGATATTCCATATGTTTCCAATTGTCTTTGTCAACAGTTCTAAAATTGAATATGGACAAATGAGA 738
Tb427.BES153.4 CCAAGCTTGGGCATCATATAATTACCGCAGCGAATCAAATTTTAGCCAACAAAACAAAGGAGATATTCCATATGTTTCCAATTGTCTTTGTCAACAGTTCTAAAATTGAATATGGACAAATGAGA 735
Tb427.BES65.3 CCAAGCTTGGGCATCATATAATTACCGCAGCGAATCAAATTTTAGCCAACAAAACAAAGGAGATATTCCATATGTTTCCAATTGTCTTTGTCAACAGTTCTAAAATTGAATATGGACAAATGAGA 735
Tb427.BES59.4 CCAAGCTTGGGCATCATATAATTACCGCAGCGAATCAAATTTTAGCCAACAAAACAAAGGAGATATTCCATATGTTTCCAATTGTCTTTGTCAACAGTTCTAAAATCGAATATGGACAAATGAGA 738
Tb427.BES51.4 CCAAGCTTGGGCATCATATAATTACCGCAGCGAATCAAATTTTAGCCAACAAAACAAAGGAGATATTCCATATGTTTCCAATTGTCTTTGTCAACAGTTCTAAAATCGAATATGGACAAATGAGA 738
Tb427.BES10.3 CCAAGCTTGGGCATCATATAATTACCGCAGCAAATCAAATTTTAGCCAACAAAACAAAGGAGATATTCCATATGTTTCCAATTGTCTTTGTCAACAGTTCTAAAATCGAATATGGACAAATGAGA 735
Tb427.BES134.4 CCAAGCTTGGGCATCATATAATTACCGCAGTGAATCAAATTTTAGCCAACAAAACAAAGGAGATATTCCATATGTTTCCAATTGTCTTTGTCAACAGTTCTAAAATCGAATATGGACAAATGAGA 735
Tb927.7.6860 GCAAACTCGAGCAGCAAATAATTACTGCAGTGAACCCCATTTTGACCAACAAAACGAGGGAAATATCCCATATGTTCCCAATTGTCTTCATAGGTAATCCGAAAATCAAAAATGGACAAATGAGA 738
Tb11.v5.0826 GCAAACTCGAGCAGCAAATAATTACTGCAGTGAACCCCATTTTGACCAACAAAACGAGGGAAATATCCCATATGTTCCCAATTGTCTTCATAGGTAATCCGAAAATCAAAAATGGACAAATGAGA 738
Tb11.v5.0463 ACGAACTACAGCGAATAGTCCCGGGTATTATCAACCCTGTCGTACAAAGTAAGGCAGAAGAGATATTCCGTGGTTTCCCCATTATATACATGAAGGACCCAAACATAACTGAGCAGCGAGCAGAG 744
Tb09.v4.0016 ACGAACTACAGCGAATAGTCCCGGGTATTATCAACCCTGTCGTACAAAGTAAGGCAGAAGAGATATTCCGTGGTTTCCCCATTATATACATGAAGGACCCAAACATAACTGAGCAGCGAGCAGAG 744
Tb927.9.15890 ACGAACTACAGCGAATAGTCCCGGGTATTATCAACCCTGTCGTACAAAGTAAGGCAGAAGAGATATTCCGTGGTTTCCCCATTATATACATGAAGGACCCAAACATAACTGAGCAGCGAGCAGAG 744
Tb09.v4.0107 ACGAGCTACAGCGATTAGTCCCTGATATGATCAACCCTGTCGTACAAAGTAAGGCAGAAGAGATATTCCGTGGTTTCCCCATTATATACATGAAGGACCCAAACATAACTGAGCAGCGAGCAGAG 744
Tb11.v5.0282 ATTATCTGCAGCAGTATGTGCCGGCTATGGTCCAGCCGATCGTAAGAGCAAAGACGGTGGACATACTTCAGAAACTTCCCATCTCATTTGTGAGCCCTCCAAACATCACGGACGGACGGATGGAA 738
Tb927.5.340 ATTATCTGCAGCAGTATGTGCCGGCTATGGTCCAGCCGATCGTAAGAGCAAAGACGGTGGACATACTTCAGAAACTTCCCATCTCATTTGTGAGCCCTCCAAACATCACGGACGGACGGATGGAA 738
Tb927.2.1920 AGCAGCTGGAGCGAATAGTTCCAGAGAGGATGAATCTCATATTGCGCAACAAGACAGAAGATGCTCTCACCAGCCTTCCGATTGCCTTCATGGACGATCCGAATATAACTGAAGGCCGGATGGAA 735
Tb927.4.810 GCAACCTGCAGGGGGAGATTCCACGGGCGGCAATGGGTCCACTGGAAGAAAAATTGAACGATTTTATTGCCACAGCACCTCTTGTTTTCCCAGAAAGCCCGAAAATTACCGATGGGGAGTTGCAG 738

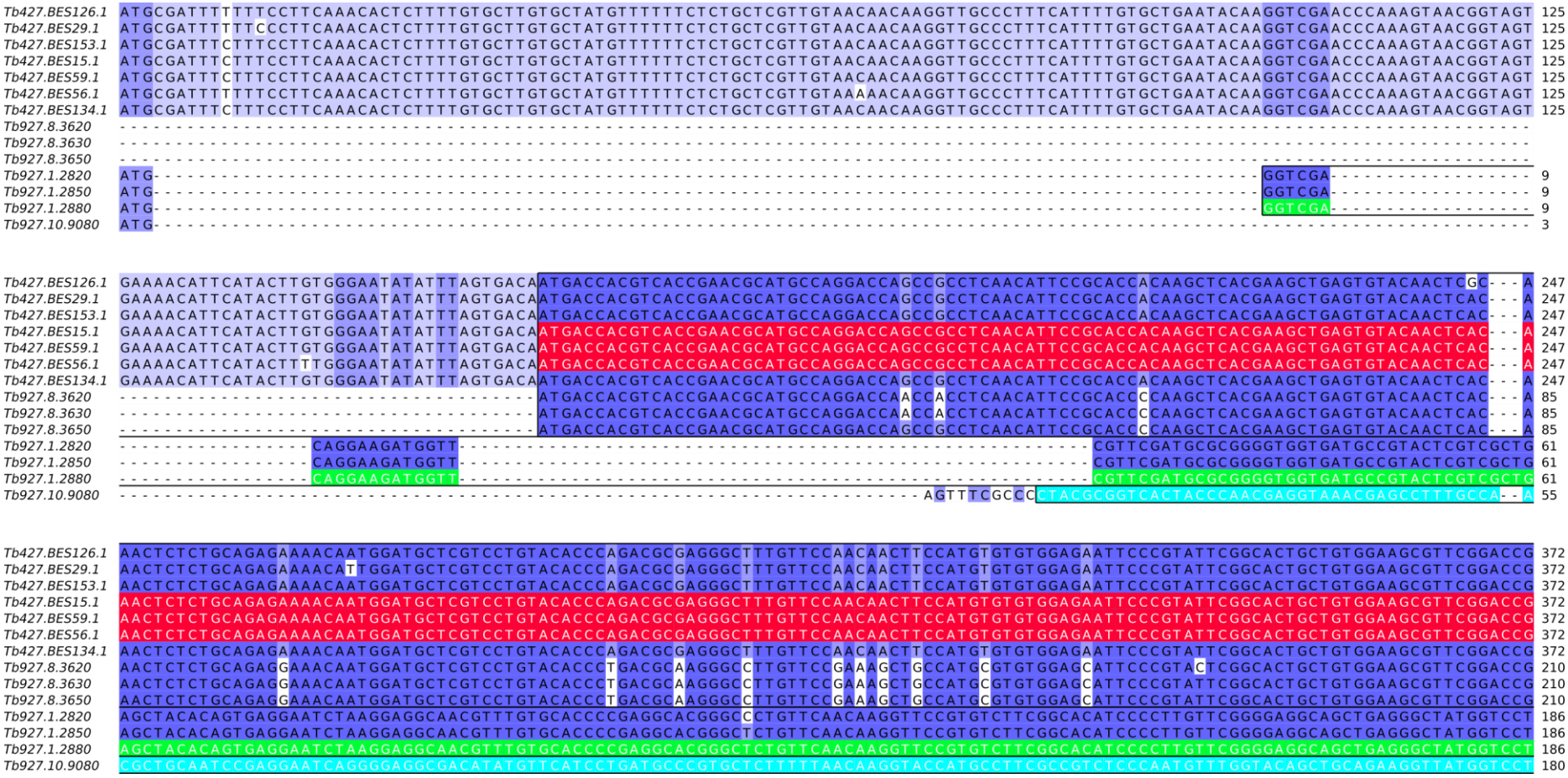

Tb427.BES98.3
Tb427.BES5.3
Tb427.BES28.5
Tb427.BES15.4
Tb427.BES4.3
Tb427.BES29.4
Tb427.BES56.4
Tb427.BES126.4
Tb427.BES153.4
Tb427.BES65.3
Tb427.BES59.4
Tb427.BES51.4
Tb427.BES10.3
Tb427.BES134.4
Tb927.7.6860
Tb11.v5.0826
Tb11.v5.0463
Tb09.v4.0016
Tb927.9.15890
Tb09.v4.0107
Tb11.v5.0282
Tb927.5.340
Tb927.2.1920
Tb927.4.810

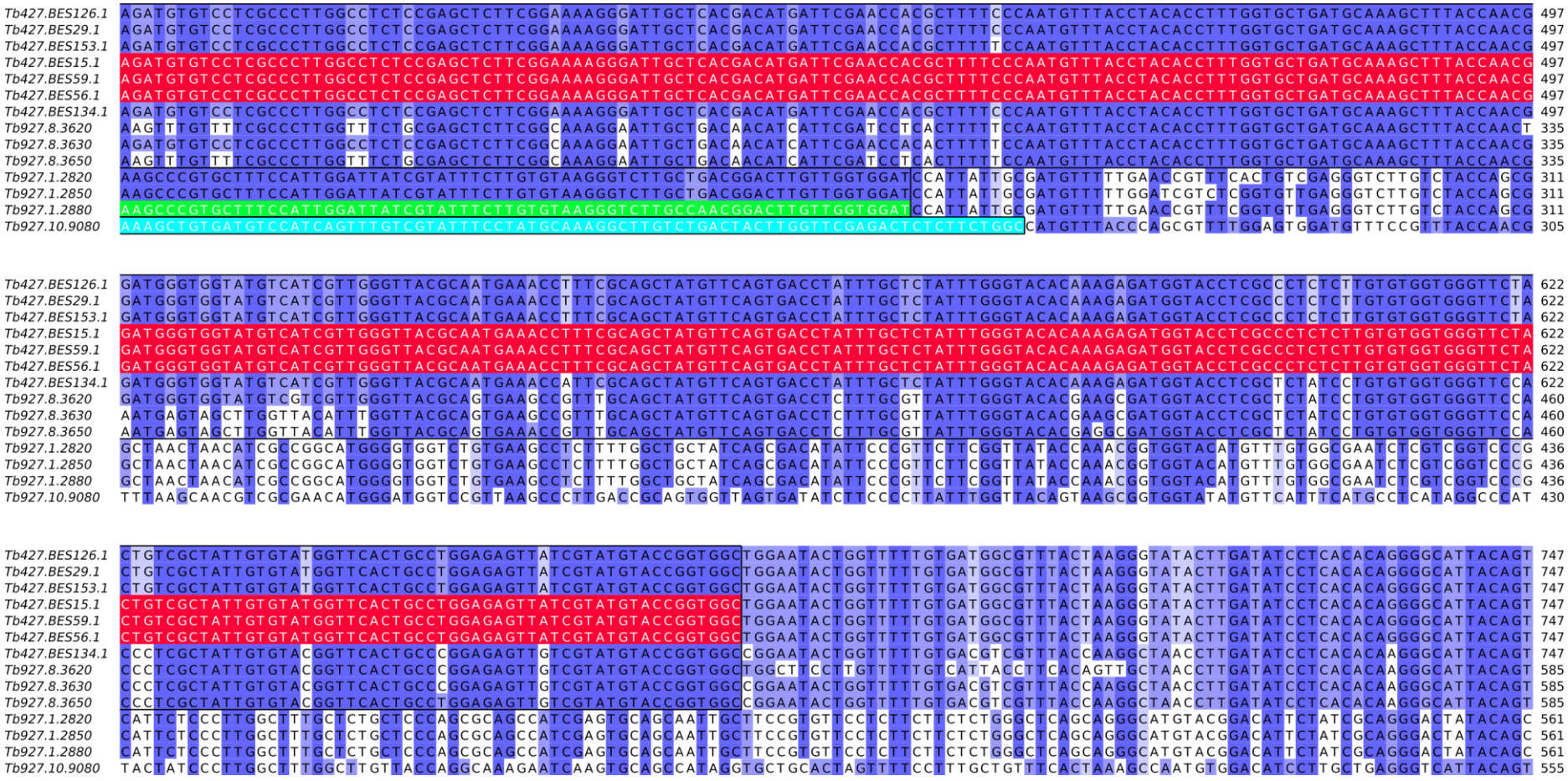

Tb427.BES98.3 CTGGATGTGCCTCTGAGGCTACATTCGAACTAAAAACAGCCCCGTGGTTGAAGTCCCTAAATAAGAAACTGTTTACATTTAACTATCGGGATGTCGAAGTTGCCGTGGACCTGCTGCCAAGGGGT 1098
Tb427.BES5.3 CTGGATGTGCCTCTGAGGCCACATTCGAACTAAAAACAGCCCCGTGGTTGAAGTCCCTAAATAAGAAACTGTTTACATTTAACTATCGGGATGTCGAAGTTGCCGTGGACCTGCTGCCAAGGGGT 1095
Tb427.BES28.5 CTGGATGTGCCTCTGAGGCCACATTCGAACTAAAAACAGCCCCGTGGTTGAAGTCCCTAAATAAGAAACTGTTTACATTTAACTATCGGGATGTCGAAGTTGCCGTGGACCTGCTGCCAAGGGGT 1095
Tb427.BES15.4 CCGGATGTGCCTCTGAGGCCACATTCGAACTAAAAACAGCCCCGTGGTTGAAGTCCCTAAATGAGAAACTGTTTACATTTAACTATCGGGATGTCGAAGTTGCCGTGGACCTGCTGCCAAGGGGT 1098
Tb427.BES4.3 CCGGATGTGCCTCTGAGGCCACATTCGAACTAAAAACAGCCCCGTGGTTGAAGTCCCTAAATGAGAAACTGTTTACATTTAACTATCGGGATGTCGAAGTTGCCGTGGACCTGCTGCCAAGGGGT 1098
Tb427.BES29.4 TTGGATGTGCCTCTGAGGCTACATTCGAACTAAAAACGGCCCCGTGGTTGAATTCCTTAAATGAGAAAATGGTTACAGTTAAGTTTCAGGATGTTGAAGTTACCGTGGACCTGTTGCCAAGGGGT 1098
Tb427.BES56.4 TCGGATGTGGCTCTGAGGCTACATTCGAACTAAAAACAGCCCCGTGGTTGAATTCCCTAAATGAGAAAATGGTTACAGTTAAGTTTCAGGATGTTGAAGTTACCGTGGACCTGTTGCCAAGGGGG 1098
Tb427.BES126.4 TCGGATGTGGCTCTGAGGCTACATTCGAACTAAAAACAGCCCCGTGGTTGAATTCCCTAAATGAGAAAATGGTTACAGTTAAGTTTCAGGATGTTGAAGTTACCGTGGACCTGTTGCCAAGGGTG 1098
Tb427.BES153.4 CCGGATGTGCCTCTGAGGCTACATTCGAACTAAAAACAGCCCCGTGGTTGAATTCCCTAAATGAGAAAATGGTTACAGTTAAGTTTCAGGATATTGAAGTTACCGTGGACCTGTTGCCAAGGGGG 1095
Tb427.BES65.3 CCGGATGTGCCTCTGAGGCTACATTCGAACTAAAAACAGCCCCGTGGTTGAATTCCCTAAATGAGAAAATGGTTACAGTTAAGTTTCAGGATATTGAAGTTACCGTGGACCTGTTGCCAAGGGGG 1095
Tb427.BES59.4 TCGGATGTGGCTCTGAGGCTACATTCGAACTAAAAACAGCCCCGTGGTTGAAGTCCCTAAATGAGAAAATGGTTAAAGTTAAGTTTCAGGATGTTGAAGTTACCGTGGACCTGTTGCCAAGGGGT 1098
Tb427.BES51.4 TCGGATGTGGCTCTGAGGCTACATTCGAACTAAAAACAGCCCCGTGGTTGAAGTCCCTAAATGAGAAAATGGTTAAAGTTAAGTTTCAGGATGTTGAAGTTACCGTGGACCTGTTGCCAAGGGGT 1098
Tb427.BES10.3 TCGGATGTGGCTCTGAGGCTACATTCGAACTAAAAACAGCCCCGTGGTTGAAGTCCCTGAATGAGAAACTGGTTACAGTTAAGTTTCAGGATGTCGAAGTTACCGTGGACCTGTTGCCAAGGGGT 1095
Tb427.BES134.4 TTGGATGTGCCTCTGAGGCCACATTCGAACTAAAAACAGCCCCGTGGTTGAAGTCCCTAAATGAGAAACTGGTTACAGTTTCAGGATGTTGAAGTTACCGTGGACCTGTTGCCAAGGGGT 1095
Tb927.7.6860 CTGGATGTGCCTCTGAGGCCTCACTTGAACTAAAAACAGCCCCGTGGTTGAAATACCTGAATAAGAAACTGTTTACAGTTAACTTTCGGGGCGTTGACGTTGCCGTGGGTCTGTTGCCAAGGGGT 1098
Tb11.v5.0826 CTGGATGTGCCTCTGAGGCCTCACTTGAACTAAAAACAGCCCCGTGGTTGAAATACCTGAATAAGAAACTGTTTACAGTTAACTTTCGGGGCGTTGACGTTGCCGTGGGTCTGTTGCCAAGGGGT 1098
Tb11.v5.0463 GTAGATGCAACTTTGGCGTCTCGGCTGTGTTGCCATCGCCTCCATGGCTGGAGACCGTTAGTGGCGGAGCATTCATTTTCAACTCACGTAATGGTAATTTCACATTGGAAATGGTATCGGAGGAC 1110
Tb09.v4.0016 GTAGATGCAACTTTGGCGTCTCGGCTGTGTTGCCATCGCCTCCATGGCTGGAGACCGTTAGTGGCGGAGCATTCATTTTCAACTCACGTAATGGTAATTTCACATTGGAAATGGTATCGGAGGAC 1110
Tb927.9.15890 GTGGATGCAACTTTGGCGTCTCGGCTGTGTTGCCATCGCCTCCATGGCTGGAGACCGTTAGTGGCGGAGCATTCATTTTCAACTCACGTAATGGTAATTTTACACTGGAAATGGTATCGGAGGAC 1110
Tb09.v4.0107 GTGGATGCAACTTTGGCGTCTCGGCTGTGTTGCCATCGCCTCCATGGCTGGAGACCGTTAGTGGCGGAGCATTCATTTTCAACTCACGTAATGGTAATTTTACATTAGAAATGGTATCGGAGGAC 1110
Tb11.v5.0282 CCAACTGTCCTTTAAAGTTTGTCACTGGATGGACAGCAGCGCCGTGGCTTGAGCCCGTAAGCGACAAAGCGTTTATCACAAATGCCCAAGGCCCTACCCTTTCCATGGATATGGTAATGGATGAC 1113
Tb927.5.340 CCAACTGTCCTTTAAAGTTTGTCACTGGATGGACAGCAGCGCCGTGGCTTGAGCCCGTAAGCGACAAAGCGTTTATCACAAATGCCCAAGGCCCTACCCTTTCCATGGATATGGTAATGGATGAC 1113
Tb927.2.1920 CCAACTGCTTCTTCCGCGTTGTGTGCCAGCTGGCAAACGACGTGTGGGTGGAGGATGACGGCAACGGGGCATTTACGTTTCAAATGCGCAATGGCACCATCTCCATGGGCCTTTATTCGGGGAAC 1101
Tb927.4.810 CAGATTGCCCCTTCATGGTTTCAGTTAAGGCGGAGACTGCGGTGCATTTGGAACCGTAGGTTTGGATAGCCTCATTCTCCATCTTTTAGGCATCCGACTGGCGTTGGAAATGAAACCCCCTGAA 1104
Tb427.BES98.3 GGTGCCCCCATTTCCTTGTTTAGTATGTTAATGAATGTGTCTGCCCAGGCGGCGCATATCGCTCTTGT---TGACGGTGCC---ACCCATGCGAATTTGGATTCTGTTGATACTAATGTGAGTGT 1217
Tb427.BES5.3 GGTAACACCATTCCCTTGTTTAGTATGTTAATGAACGTGTCCATCCGGGCGGAGCATATCGCCCTCGT---TGATGGTGCC---ACCCACGCCTCTTTGGATTCCGTTGATACTAATGTGAGTGT 1214
Tb427.BES28.5 GGTAACACCATTCCCTTGTTTAGTATGTTAATGAACGTGTCCATCCGGGCGGAGCATATCGCCCTCGT---TGATGGTGCC---ACCCACGCCTCTTTGGATTCCGTTGATACTAATGTGAGTGT 1214
Tb427.BES15.4 GGTAACACCATTCCCTTGTTTAGTATGTTAATGAACGTGTCCATCCGGGCGGAGCATATCGCCCTCGT---TGATGGTGCC---ACCCACGCCTCTTTGGATTCCGTTGATACTAATGTGAGTGT 1217
Tb427.BES4.3 GGTAACACCATTCCCTTGTTTAGTATGTTAATGAACGTGTCCATCCGGGCGGAGCATATCGCCCTCGT---TGATGGTGCC---ACCCACGCCTCTTTGGATTCCGTTGATACTAATGTGAGTGT 1217
Tb427.BES29.4 GGTGACCCCATTTCCTTGTTTAGTATGTTAATGAACGTGTCTGTCCAGGCGGCGCATATCGCTCTCAT---TGACGGTGTC---GCCCATGCGAATTTGGATTCTGTTGATACTAATGTGAGCGT 1217
Tb427.BES56.4 GGTGCCCCCATTTCCTTGTTTAGTATGTTAATGAACGTGTCTGTCCAGGCGGCGCATATCGCTCTCAT---TGACGGTGTC---GCCCATGCGAATTTGGATTCTGTTGATACTAATGTGAGCGT 1217
Tb427.BES126.4 GGTGCCCCCATTTCCTTGTTTAGTATGTTAATGAATGTGTCTGTCCAGGCGATGCATATCGCTCTCAT---TGACGGTGTC---GCCCATGCGAATTTGGATTCTGTTGATACTAATGTGAGCGT 1217
Tb427.BES153.4 GGTGCCCCCATTTCCTTGTTTAGTATGTTAATGAACGTGTCTGTCCAGGCGGCGCATATCGCTCTCAT---TGACGGTGTC---GCCCATGCGAATTTGGATTCTGTTGATACTAATGTGAGCGT 1214
Tb427.BES65.3 GGTGCCCCCATTTCCTTGTTTAGTATGTTAATGAACGTGTCTGTCCAGGCGGCGCATATCGCTCTCAT---TGACGGTGTC---GCCCATGCGAATTTGGATTCTGTTGATACTAATGTGAGCGT 1214
Tb427.BES59.4 GGTGACCCCATTTCCTTGTTTAGTATGTTAATGAACGTGTCTGTCGAGACGGTGCATATCGCTCTCAT---TGACGGTGTC---GCCCATGCGAATTTGGATTCTGTTGATACTAATGTGAGCGT 1217
Tb427.BES51.4 GGTGACCCCATTTCCTTGTTTAGTATGTTAATGAACGTGTCTGTCGAGACGGTGCATATCGCTCTCAT---TGACGGTGTC---GCCCATGCGAATTTGGATTCTGTTGATACTAATGTGAGCGT 1217
Tb427.BES10.3 GGTGACCCCATTTCCTTGTTTAGTATGTTAATGAACGTGTCTGTCGAGACGGTGCATATCGCTCTCAT---TGACGGTGTC---GCCCATGCGAATTTGGATTCTGTTGATACTAATGTGAGCGT 1214
Tb427.BES134.4 GGTGACCCCATTTCCTTGTTTAGTATGTTAATGAACGTGTCTGTCGAGACGGTGCATATCGCTCTCAT---TGACGGTGTC---GCCCATGCGAATTTGGATTCTGTTGATACTAATGTGAGCGT 1214
Tb927.7.6860 GGTGACCCCATTCCCTTGTTTAGTATGTTAATGAACGTGTCCGTCCGGGCGGCGCATATCGCTGTCGT---CGACAGCATC---GCCCATGCGAAGTTGGATTCCATTGATGCTACGGTGAGCGT 1217
Tb11.v5.0826 GGTGACCCCATTCCCTTGTTTAGTATGTTAATGAACGTGTCCGTCCGGGCGGCGCATATCGCTGTCGT---CGACAGCATC---GCCCATGCGAAGTTGGATTCCATTGATGCTACGGTGAGCGT 1217
Tb11.v5.0463 AACGTGTCATTCACTGTGTTGAATATGGTTGTGAACCTGACCGCTAATGTAAGGCACATGTCCATCGA---CAACAATACA---TTAGACCTGAAGTTATCGTCTATGGATGTAGTGACAGAACT 1229
Tb09.v4.0016 AACGTGTCATTCACTGTGTTGAATATGGTTGTGAACCTGACCGCTAATGTAAGGCACATGTCCATCGA---CAACAATACA---TTAGACCTGAAGTTATCGTCTATGGATGTAGTGACAGAACT 1229
Tb927.9.15890 AACGTGTCATTCACTGTGTTGAATATGGTTGTGAACTTGACCGCTAATGTAAGGCACATGTCCATCGA---CAACAATACA---TTAGACCTGAAGTTATCGTCTATGGATGTAGTGACAGAACT 1229
Tb09.v4.0107 AACGTGTCATTCACTGTGTTGAATATGGTTGTGAACTTGACCGCTAATGTAAGGCACATGTACATCGA---CAACAATACA---TTAGACCTGAAGTTATCGTCTATGGACGTAGTGACAGGACT 1229
Tb11.v5.0282 AACACGTCGGTAACTGTGCTTGGTGCTCATCTAAACGCGACCGCCAACGTGACTCATCTTTCCACCGA---CAACAGTGGCCGTATCACCATAAAGTTGGCCTCCGTGAATGCTGCGGTGGCACT 1235
Tb927.5.340 AACACGTCGGTAACTGTGCTTGGTGCTCATCTAAACGCGACCGCCAACGTGACTCATCTTTCCACCGA---CAACAGTGGCCGTATCACCATAAAGTTGGCCTCCGTGAATGCTGCGGTGGCACT 1235
Tb927.2.1920 GGCACGAGTGTGGACGCACTCAGTTTATCAGTGAATACGACGGCTAACGTCAAAAAGTTTTCTGTCTT---CGACAGCGCT---GTCACACTGAACTTGTCCGCCATGAGTGTGTCTTTGGATGT 1220
Tb927.4.810 GGTGATCCCATTCCCGCGATGAGACTACTGTGTAACACGACAGTTTCTCTGTCAAGCCTTTATTTTGACTCCAGCGGTGCC---ATCCCTTTGGAAATGGATTTGCTTGACCTGGCGATTGATGT 1226

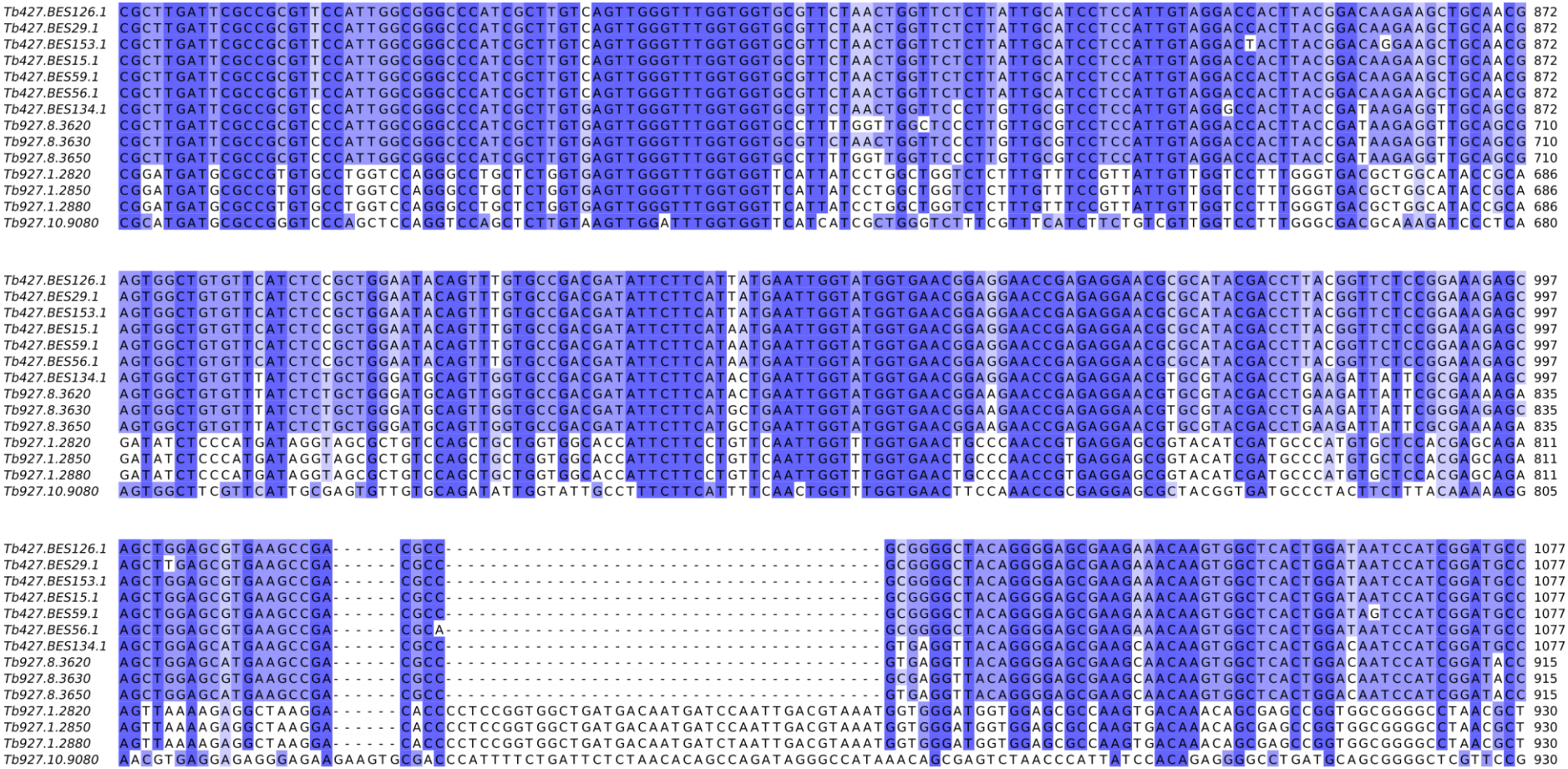

Tb427.BES98.3 CACATCCTCACGTATCGATGGCCTGGACTCAAGTACAATGAACACAAAAATTCGTGATTTGATCAAC---------TTGTCTTGGATCAACCTGAATGTTACATATACATTTCCTGCTCCCTTCG 1333
Tb427.BES5.3 CACATCCTCACGTATCGATGGCCTGGACTCAAGTACAATGAACACAAAAATTCGTGATTTGATCAAC---------TTGTCTTGGATCAACCTGAATGTTACATATACATTTCCTGCTCCCTTCG 1330
Tb427.BES28.5 CACATCCTCACGTATCGATGGCCTGGACTCAAGTACAATGAACACAAAAATTCGTGATTTGATCAAC---------TTGTCTTGGATCAACCTGAATGTTACATATACATTTCCTGCTCCCTTCG 1330
Tb427.BES15.4 CACATCCTCACGTATCGATGGCCTGGACTCAAGCACAATGAACACAAAAATTCGTGATTTGATCAAC---------TTGTCTTGGACCGACCTGAATGTTACATATAGATTTCCCACTCGCTTCG 1333
Tb427.BES4.3 CACATCCTCACGTATCGATGGCCTGGACTCAAGCACAATGAACACAAAAATTCGTGATTTGATCAAC---------TTGTCTTGGACCGACCTGAATGTTACATATAGATTTCCCACTCGCTTCG 1333
Tb427.BES29.4 CACATCGTCACGTATCGATGGCCTGGACTCAAGCACAATGAACACAAAAATTCGTGATTTGATCAACGGGATGGTGTTGCCGCTTCTCAACTTCAA---GAGATATGCATTTCCCGCTCCCTTCG 1339
Tb427.BES56.4 CACATCCTCACGTATCGATGGCCTGGACTCAAGCACAATGAACACAAAAATTCGTGATTTGATCAACGGGATGGTGTTGCCGCTTCTCAACTTCAA---GAGATATGCATTTCCCGCTCCCTTCG 1339
Tb427.BES126.4 CACATCCTCACGTATCGATGGCCTGGACTCAAGCATAATGAACACAAAAATTCGTGATTTGATCAACGGGATGGTGTTGCCGCTTCTTAACTTCAA---GAGATATGCATTTCCAGCTCCCTTCG 1339
Tb427.BES153.4 CACATCGTCACGTATCGATGGCCTGGACTCAAGCACAATGAACACAAAAATTCGTGATTTGATCAACGGGATGGTGTTGCCGCTTCTTAACTTCAA---GAGATATGCATTTCCAGCTCCCTTCG 1336
Tb427.BES65.3 CACATCGTCACGTATCGATGGCCTGGACTCAAGCACAATGAACACAAAAATTCGTGATTTGATCAACGGGATGGTGTTGCCGCTTCTTAACTTCAA---GAGATATGCATTTCCAGCTCCCTTCG 1336
Tb427.BES59.4 CACATCCTCACGTATCGATGGCCTGGACTCAAGCACAATGAACACAAAAATTCGTGATTTGATCAACGGGATGGTGTTGCCGCTTCTCAACTTCAA---GAGATATGCATTTCCCGCTCCCTTCG 1339
Tb427.BES51.4 CACATCCTCACGTATCGATGGCCTGGACTCAAGCACAATGAACACAAAAATTCGTGATTTGATCAACGGGATGGTGTTGCCGCTTCTCAACTTCAA---GAGATATGCATTTCCCGCTCCCTTCG 1339
Tb427.BES10.3 CACATCCTCACGTATCGATGGCCTGGACTCAAGCACAATGAACACAAAAATTCGTGATTTGATCAACGGGATGGTGTTGCCGCTTCTTAACTTCAA---GAGATATGCATTTCCCGCTCCCTTCG 1336
Tb427.BES134.4 CACATCCTCACGTATCGATGGCCTGGACTCAAGCACAATGAACACAAAAATTCGTGATTTGATCAACGGGATGGTGTTGCCGCTTCTTAACTTCAA---GAGATATGCATTTCCCGCTCCCTTCG 1336
Tb927.7.6860 CACATCCTCACGTATCGATGGCCTGGACTCAAGTACAATGAACACAAAAATTCGTGATTTGATCAACGGGATGGTGTTGCCGCTGCTCAACTTCAA---GAAGGATGGCTTTCCCGTTCCCTTCG 1339
Tb11.v5.0826 CACATCCTCACGTATCGATGGCCTGGACTCAAGTACAATGAACACAAAAATTCGTGATTTGATCAACGGGATGGTGTTGCCGCTGCTCAACTTCAA---GAAGGATGGCTTTCCCGTTCCCTTCG 1339
Tb11.v5.0463 GGGCACTTCGCTTATCGATGGTTTGAACTCAACGACTCTAAATACAGATATCCGTCGGTTTCTCGATGAAGTAGGTCTTCCTTTGTTTAATGCAGA---TCCGCATGGATTTAGGCTCCCCTTCA 1351
Tb09.v4.0016 GGGCACTTCGCTTATCGATGGTTTGAACTCAACGACTCTAAATACAGATATCCGTCGGTTTCTCGATGAAGTAGGTCTTCCTTTGTTTAATGCAGA---TCCGCATGGATTTAGGCTCCCTTCA 1351
Tb927.9.15890 GGGCACTTCGCTTATCGATGGTTTGAACTCAACGACTCTAAATACAGATATCCGTCGGTTTCTCGATGAGGTAGGTCTTCCTTTGTTTAATGCAGA---TCCGCATGGATTTAGGCTCCCTTCA 1351
Tb09.v4.0107 GGGCACTTCGCTTATCGATGGTTTGAACTCAACGACTCTAAATACAGATATCCGTCGGTTTCTCGATGAAGTAGATCTTCCTTTGTTTAATGCAGA---TCCGCATGGATTTAGGCTCTCCTTCA 1351
Tb11.v5.0282 ATATGCTCCGTTGGGGCGAAAAATAAATTCAACATCACTCAATACAGATGTTCAGGTCTTATTGAATGAGATCGTGGTGCCGCTTGTTAATGCGCA---AACGGGAGGATTCGCACTTCCGTTCG 1357
Tb927.5.340 ATATGCTCCGTTGGGGCGAAAAATAAATTCAACATCACTCAATACAGATGTTCAGGTCTTATTGAATGAGATCGTGGTGCCGCTTGTTAATGCGCA---AACGGGAGGATTCGCACTTCCGTTCG 1357
Tb927.2.1920 TGAGGCTTCGCTTATTGGTGATTTCAATTCGACAGTTCTTAACAGCGACATCCAGTGGGTGCTCGATGTGGTGGTGGTTCCATTGGTTAACGCACA---ACCGCGAGGGTTTGTCCTTCCTTCA 1342
Tb927.4.810 AGAAACTTCAAATATTGGAGCGATTGATCAGCTCACGCTGAACGACATACTTTCGAGGGAGCTCAGAGACAAGGTAATCCCTGCGGTCAACGAGAA---AATCTCAACGATAACGATTTTTGATA 1348
Tb427.BES98.3 ACTTGTGTACGAAACACGTGAATATTACAAGTGAATGCTAT-GTAGCGGGATTTAACCTTGTGCGAGCTCTTGGTTCACTCAGTATAC--TCCCTCACCTCCGATG---------- 1436
Tb427.BES5.3 ACTTGTGTACGAAACACGTGAATATTACAAGTGAATGCTAT-GTAGCGGGATTTAACCTTGTGCGAGCTCTTGGTTCACTCAGTATAC--TCCCTCACCTCCGATG---------- 1433
Tb427.BES28.5 ACTTGTGTACGAAACACGTGAATATTACAAGTGAATGCTAT-GTAGCGGGATTTAACCTTGTGCGAGCTCTTGGTTCACTCAGTATAC--TCCCTCACCTCCGATG---------- 1433
Tb427.BES15.4 ACTTGTGTACGAAACACGTGAATATTACAAGTGAATACTAT-ATGGCGGGATTTAACCTTGTGCGAGCTCTTGGTTCACTCACTACCC--TACCTCACCTCCGATG---------- 1436
Tb427.BES4.3 ACTTGTGTACGAAACACGTGAATATTACAAGTGAATACTAT-ATGGCGGGATTTAACCTTGTGCGAGCTCTTGGTTCACTCACTACCC--TACCTCACCTCCGATG---------- 1436
Tb427.BES29.4 ACTTGTCTGGGATCCACCTGAATATTACAAGTGGATGCGGT-ATAGCGGGAGTTGACCTTGTACGCGCTCTTGGATCACTCAGTATAC--TCCCTCACCTCCGATG---------- 1442
Tb427.BES56.4 ACTTGTCTGGGATCCACCTGAATATTACAACTGAAGGCGGT-ATAGCGGGAGTTGACCTTGTACGCGCTCTTGGATCACTCAGTATAC--TCCCTCACCTCCGATG---------- 1442
Tb427.BES126.4 ACTTGTCTGGGATCCACCTGAATATTACAACTGAAGGCGGT-ATAGCGGGAGTTGACCTTGTACGCGCTCTTGGATCACTCAGTATAC--TCCCTCACCTCCGATG---------- 1442
Tb427.BES153.4 ACTTGTCTGGGATGCACCTGAATATTACAACTGAAGGCGGT-ATAGCGGGAGTTGACCTTGTACGCGCTCTTGGATCACTCAGTATAC--TCCCTCACCTCCGATG---------- 1439
Tb427.BES65.3 ACTTGTCTGGGATGCACCTGAATATTACAACTGAAGGCGGT-ATAGCGGGAGTTGACCTTGTACGCGCTCTTGGATCACTCAGTATAC--TCCCTCACCTCCGATG---------- 1439
Tb427.BES59.4 ACTTGTCTGGGATCCACCTGAATATTACAACTGAAGGCGGT-ATAGCGGGAGTTGACCTTGTACGCGCTCTTGGATCACTCAGTATAC--TCCCTCACCTCCGATG---------- 1442
Tb427.BES51.4 ACTTGTCTGGGATCCACCTGAATATTACAACTGAAGGCGGT-ATAGCGGGAGTTGACCTTGTACGCGCTCTTGGATCACTCAGTATAC--TCCCTCACCTCCGATG---------- 1442
Tb427.BES10.3 ACTTGTCTGGGATCCACCTGAATATTACAACTGAAGGCGGT-ATAGCGGGAGTTGACCTTGTACGCGCTCTTGGATCACTCGGTATAC--TCCCTCACCTCCGATG---------- 1439
Tb427.BES134.4 ACTTGTCTGGGATGCACCTGAATATTACAACTGAAGGCGGT-ATAGCGGGAGTTGACCTTGTACGCGCTCTTGGATCACTCAGTATAC--TCCCTCACCTCCGATG---------- 1439
Tb927.7.6860 ACTTGTCTGGGATCCACCTGAATATTACAGGTGAAGGCGGT-AAAGCGGGAGTTGACCCGGCACGCGCTTTTAGATCGCTCAGCACAT--TCCTTCACCTCCGATG---------- 1442
Tb11.v5.0826 ACTTGTCTGGGATCCACCTGAATATTACAGGTGAAGGCGGTAAAAGCGGGAGTTGA---------- 1395
Tb11.v5.0463 ATATATCGGGACTGCTGTTGTCCGTATCGAATGCCACCATC-ACTGTCGGACTGGACCCGAAGCCGATCTTTCCGCTGTTAAACAAAT--TTGTTGATCGTTTGATGTG---------- 1457
Tb09.v4.0016 ATATATCGGGACTGCTGTTGTCCGTATCGAATGCCACCATC-ACTGTCGGACTGGACCCGAAGCCGATCTTTCCGCTGTTAAACAAAT--TTGTTGATCGTTTGATGTG---------- 1457
Tb927.9.15890 ATATATCGGGACTACTGTTGTCCGTATCGAATGCCACCATC-ACTGTCGGACTGGACCCGAAGCCGATCTTTCCGCTGTTAAACAAAT--TTGTCGATCGTTTGATGTG---------- 1457
Tb09.v4.0107 ATATATCGGGACTGCTGTTGTCCGTATCGAATGCCACCATC-ACTGTCGGACTGGACCCGAAGCCGATCTTTCCGCTGTTAAACAAAT--TTGTTGATCGTTTGATGTG---------- 1457
Tb11.v5.0282 ATGTAACAGATCTTTTGCTGAAGGTGTCAAATGCCACCATC-AGTGTGGGACTTAATGCGATGGAAGCATATCCGCTTATCGGCTCAT--GCATTGATCGTATATGGTG---------- 1463
Tb927.5.340 ATGTAACAGATCTTTGCTGAAGGTGTCAAATGCCACCATC-AGTGTGGGACTTAATGCGATGGAAGCATATCCGCTTATCGGCTCAT--GCATTGATCGTATATGGTG---------- 1463
Tb927.2.1920 ACCTAACGGGGCTTCTTTTTAATGTTTCCACCGGTGCTTTC-ACTATGGGAATGAATTCAGCAGTGATTTCTTCACTGCTCAACATCT--TCGTCGGTCAGCGTTGGCGACCTGTGAGTGGCGAA 1464
Tb927.4.810 ACCACACAGAACCGTTGCTGAACGTTACCAGTGAGTCGATT-ACGTTCGCGTTGAATGTGCCGCGTGGAAACTCTTCAAATGAGACCTCATTAACGACCGTATGA---------- 1452

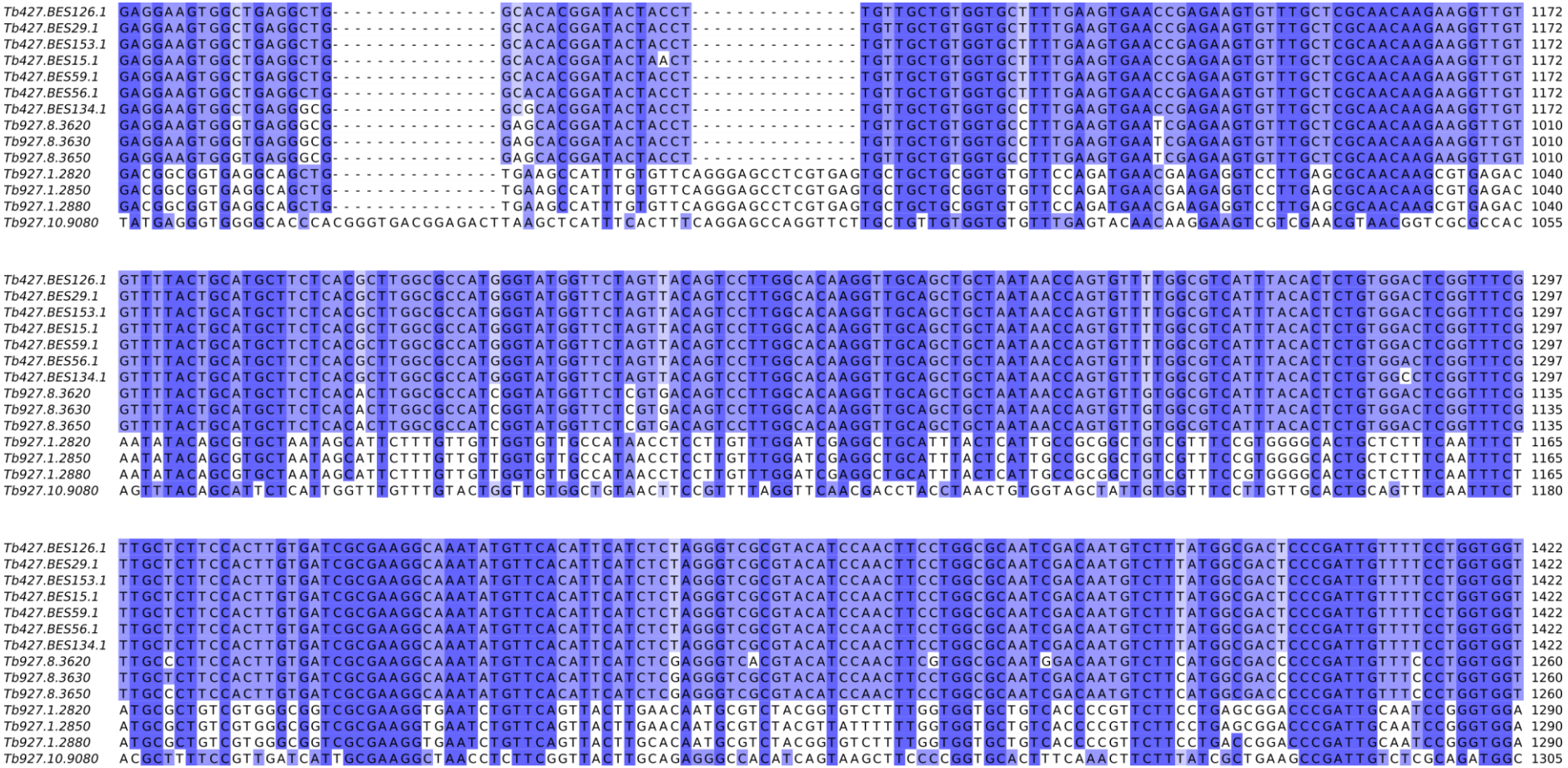


Appendix 8. The *(GR)ESAG5* family is divergent at the DNA level. Multiple DNA sequence alignment of *(GR)ESAG5* family genes. Nucleotides coloured based on consensus over all sequences, except in boxes, where they are coloured based on consensus of sequences within the box. RNAi fragments are highlighted: fragment 1, red; 2, green; 3, light blue; 4, purple; and 5, grey. Black boxes surround the genes targeted by these respective RNAi fragments.


Tb427.BES126.1 125
Tb427.BES29.1 125
Tb427.BES153.1 125
Tb427.BES15.1 125
Tb427.BES59.1 125
Tb427.BES56.1 125
Tb427.BES134.1 125
Tb927.8.3620
Tb927.8.3630
Tb927.8.3650
Tb927.1.2820 9
Tb927.1.2850 9
Tb927.1.2880 9
Tb927.10.9080 3
Tb427.BES126.1 247
Tb427.BES29.1 247
Tb427.BES153.1 247
Tb427.BES15.1 247
Tb427.BES59.1 247
Tb427.BES56.1 247
Tb427.BES134.1 247
Tb927.8.3620 85
Tb927.8.3630 85
Tb927.8.3650 85
Tb927.1.2820 61
Tb927.1.2850 61
Tb927.1.2880 61
Tb927.10.9080 55
Tb427.BES126.1 372
Tb427.BES29.1 372
Tb427.BES153.1 372
Tb427.BES15.1 372
Tb427.BES59.1 372
Tb427.BES56.1 372
Tb427.BES134.1 372
Tb927.8.3620 210
Tb927.8.3630 210
Tb927.8.3650 210
Tb927.1.2820 186
Tb927.1.2850 186
Tb927.1.2880 186
Tb927.10.9080 180

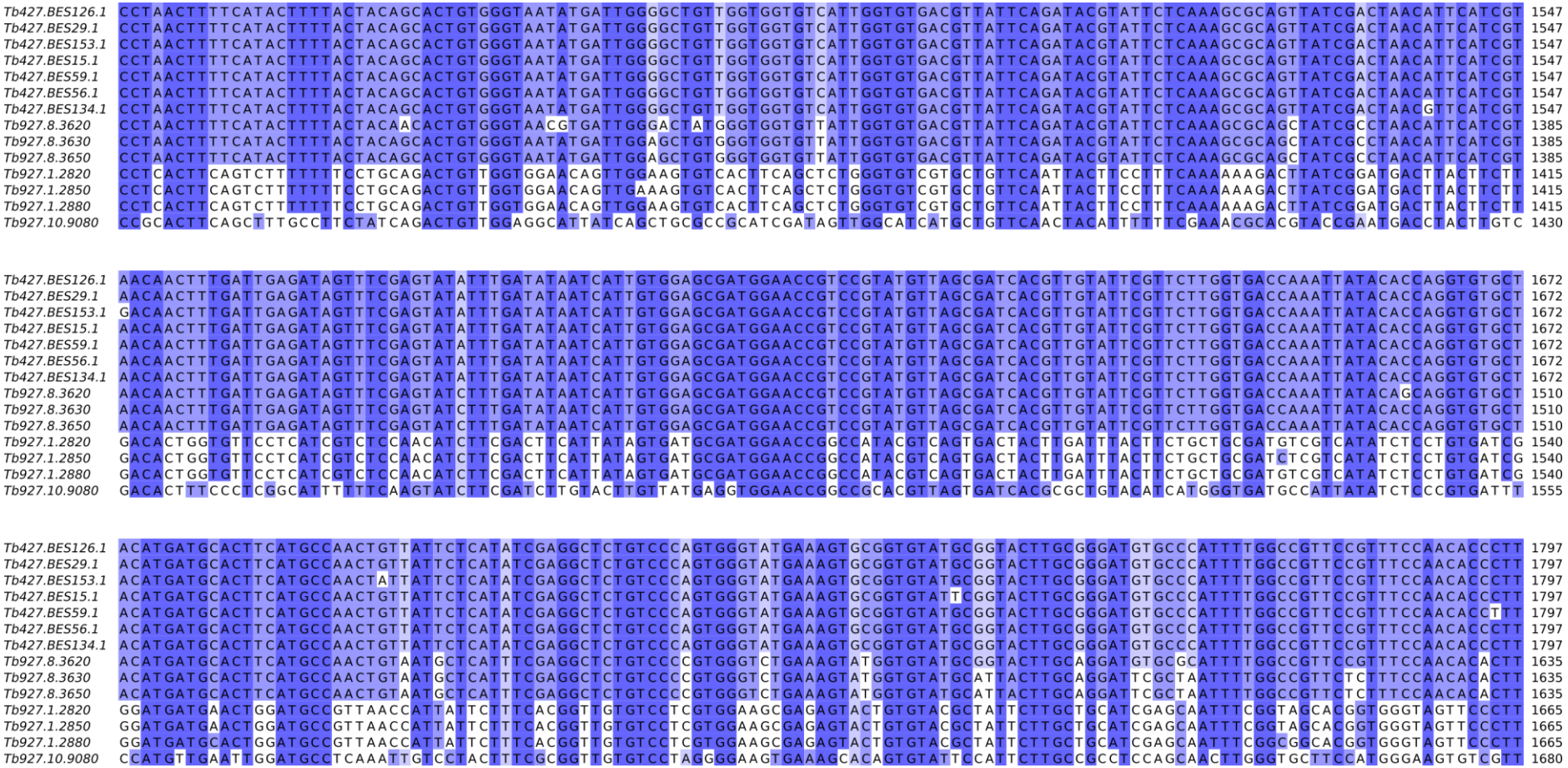

Tb427.BES126.1 497
Tb427.BES29.1 497
Tb427.BES153.1 497
Tb427.BES15.1 497
Tb427.BES59.1 497
Tb427.BES56.1 497
Tb427.BES134.1 497
Tb927.8.3620 335
Tb927.8.3630 335
Tb927.8.3650 335
Tb927.1.2820 311
Tb927.1.2850 311
Tb927.1.2880 311
Tb927.10.9080 305
Tb427.BES126.1 622
Tb427.BES29.1 622
Tb427.BES153.1 622
Tb427.BES15.1 622
Tb427.BES59.1 622
Tb427.BES56.1 622
Tb427.BES134.1 622
Tb927.8.3620 460
Tb927.8.3630 460
Tb927.8.3650 460
Tb927.1.2820 436
Tb927.1.2850 436
Tb927.1.2880 436
Tb927.10.9080 430
Tb427.BES126.1 747
Tb427.BES29.1 747
Tb427.BES153.1 747
Tb427.BES15.1 747
Tb427.BES59.1 747
Tb427.BES56.1 747
Tb427.BES134.1 747
Tb927.8.3620 585
Tb927.8.3630 585
Tb927.8.3650 585
Tb927.1.2820 561
Tb927.1.2850 561
Tb927.1.2880 561
Tb927.10.9080 555

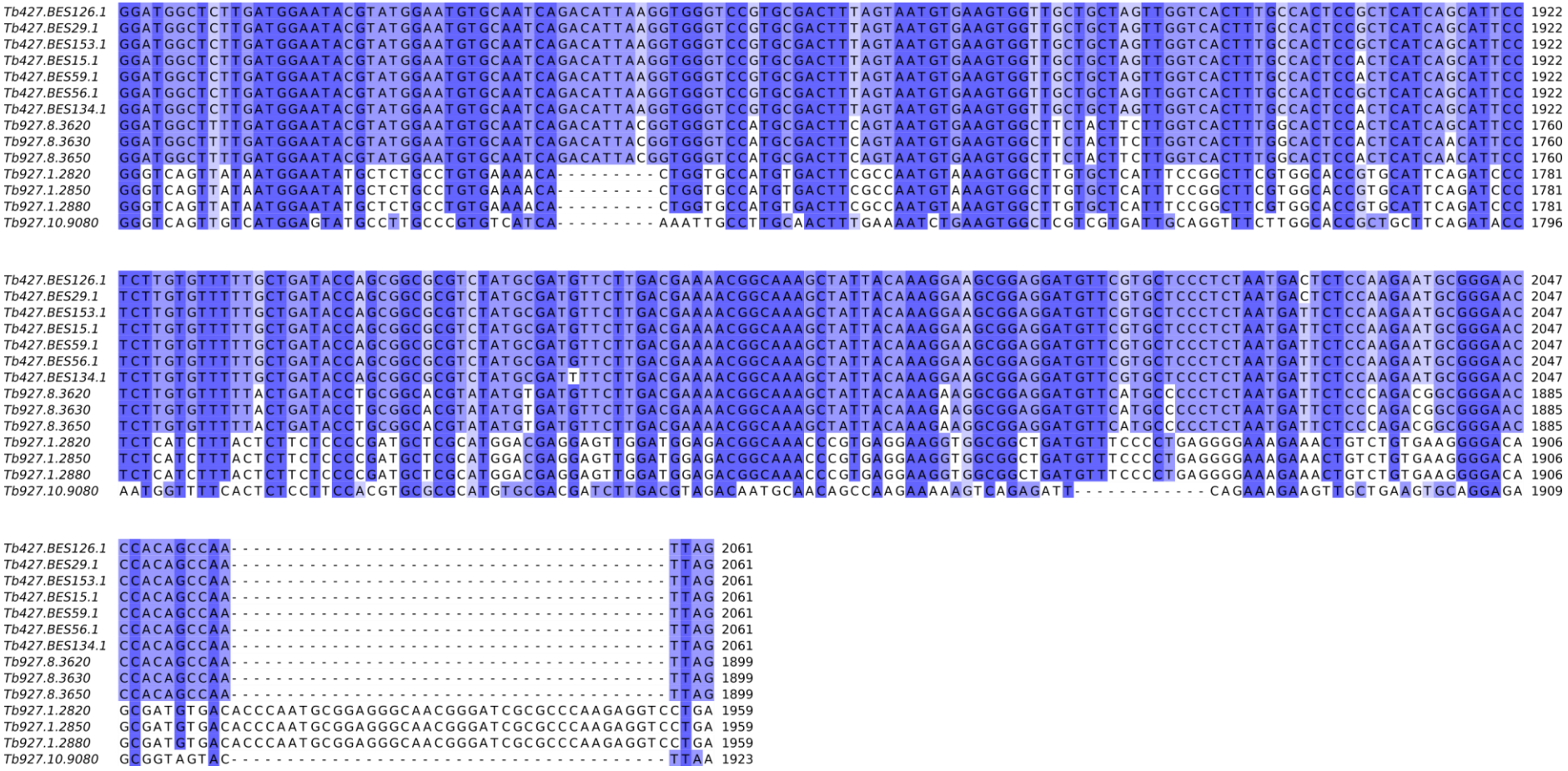

Tb427.BES126.1 872
Tb427.BES29.1 872
Tb427.BES153.1 872
Tb427.BES15.1 872
Tb427.BES59.1 872
Tb427.BES56.1 872
Tb427.BES134.1 872
Tb927.8.3620 710
Tb927.8.3630 710
Tb927.8.3650 710
Tb927.1.2820 686
Tb927.1.2850 686
Tb927.1.2880 686
Tb927.10.9080 680
Tb427.BES126.1 997
Tb427.BES29.1 997
Tb427.BES153.1 997
Tb427.BES15.1 997
Tb427.BES59.1 997
Tb427.BES56.1 997
Tb427.BES134.1 997
Tb927.8.3620 835
Tb927.8.3630 835
Tb927.8.3650 835
Tb927.1.2820 811
Tb927.1.2850 811
Tb927.1.2880 811
Tb927.10.9080 805
Tb427.BES126.1 1077
Tb427.BES29.1 1077
Tb427.BES153.1 1077
Tb427.BES15.1 1077
Tb427.BES59.1 1077
Tb427.BES56.1 1077
Tb427.BES134.1 1077
Tb927.8.3620 915
Tb927.8.3630 915
Tb927.8.3650 915
Tb927.1.2820 930
Tb927.1.2850 930
Tb927.1.2880 930
Tb927.10.9080 930


Tb427.BES126.1 GAGGAAGTGGCTGAGGCTG----------GCACACGGATACTACCT----------TGTTGCTGTGGTGCTTTTGAAGTGAACCGAGAAGTGTTTGCTCGCAACAAGAAGGTTGT 1172
Tb427.BES29.1 GAGGAAGTGGCTGAGGCTG----------GCACACGGATACTACCT----------TGTTGCTGTGGTGCTTTTGAAGTGAACCGAGAAGTGTTTGCTCGCAACAAGAAGGTTGT 1172
Tb427.BES153.1 GAGGAAGTGGCTGAGGCTG----------GCACACGGATACTACCT----------TGTTGCTGTGGTGCTTTTGAAGTGAACCGAGAAGTGTTTGCTCGCAACAAGAAGGTTGT 1172
Tb427.BES15.1 GAGGAAGTGGCTGAGGCTG----------GCACACGGATACTAACT----------TGTTGCTGTGGTGCTTTTGAAGTGAACCGAGAAGTGTTTGCTCGCAACAAGAAGGTTGT 1172
Tb427.BES59.1 GAGGAAGTGGCTGAGGCTG----------GCACACGGATACTACCT----------TGTTGCTGTGGTGCTTTTGAAGTGAACCGAGAAGTGTTTGCTCGCAACAAGAAGGTTGT 1172
Tb427.BES56.1 GAGGAAGTGGCTGAGGCTG----------GCACACGGATACTACCT----------TGTTGCTGTGGTGCTTTTGAAGTGAACCGAGAAGTGTTTGCTCGCAACAAGAAGGTTGT 1172
Tb427.BES134.1 GAGGAAGTGGCTGAGGGCG----------GCGCACGGATACTACCT----------TGTTGCTGTGGTGCCTTTGAAGTGAACCGAGAAGTGTTTGCTCGCAACAAGAAGGTTGT 1172
Tb927.8.3620 GAGGAAGTGGGTGAGGGCG----------GAGCACGGATACTACCT----------TGTTGCTGTGGTGCCTTTGAAGTGAATCGAGAAGTGTTTGCTCGCAACAAGAAGGTTGT 1010
Tb927.8.3630 GAGGAAGTGGGTGAGGGCG----------GAGCACGGATACTACCT----------TGTTGCTGTGGTGCCTTTGAAGTGAATCGAGAAGTGTTTGCTCGCAACAAGAAGGTTGT 1010
Tb927.8.3650 GAGGAAGTGGGTGAGGGCG----------GAGCACGGATACTACCT----------TGTTGCTGTGGTGCCTTTGAAGTGAATCGAGAAGTGTTTGCTCGCAACAAGAAGGTTGT 1010
Tb927.1.2820 GACGGCGGTGAGGCAGCTG----------TGAAGCCATTTGTGTTCAGGGAGCCTCGTGAGTGCTGCTGCGGTGTGTTCCAGATGAACGAAGAGGTCCTTGAGCGCAACAAGCGTGAGAC 1040
Tb927.1.2850 GACGGCGGTGAGGCAGCTG----------TGAAGCCATTTGTGTTCAGGGAGCCTCGTGAGTGCTGCTGCGGTGTGTTCCAGATGAACGAAGAGGTCCTTGAGCGCAACAAGCGTGAGAC 1040
Tb927.1.2880 GACGGCGGTGAGGCAGCTG----------TGAAGCCATTTGTGTTCAGGGAGCCTCGTGAGTGCTGCTGCGGTGTGTTCCAGATGAACGAAGAGGTCCTTGAGCGCAACAAGCGTGAGAC 1040
Tb927.10.9080 TATGAGGGTGGGGCACCCACGGGTGACGGAGACTTAAGCTCATTTCACTTTCAGGAGCCAGGTTCTTGCTGTTGTGGTGTGTTTGAGTACAACAAGGAAGTCGTCCGAACGTAACGGTCGCGCCAC 1055

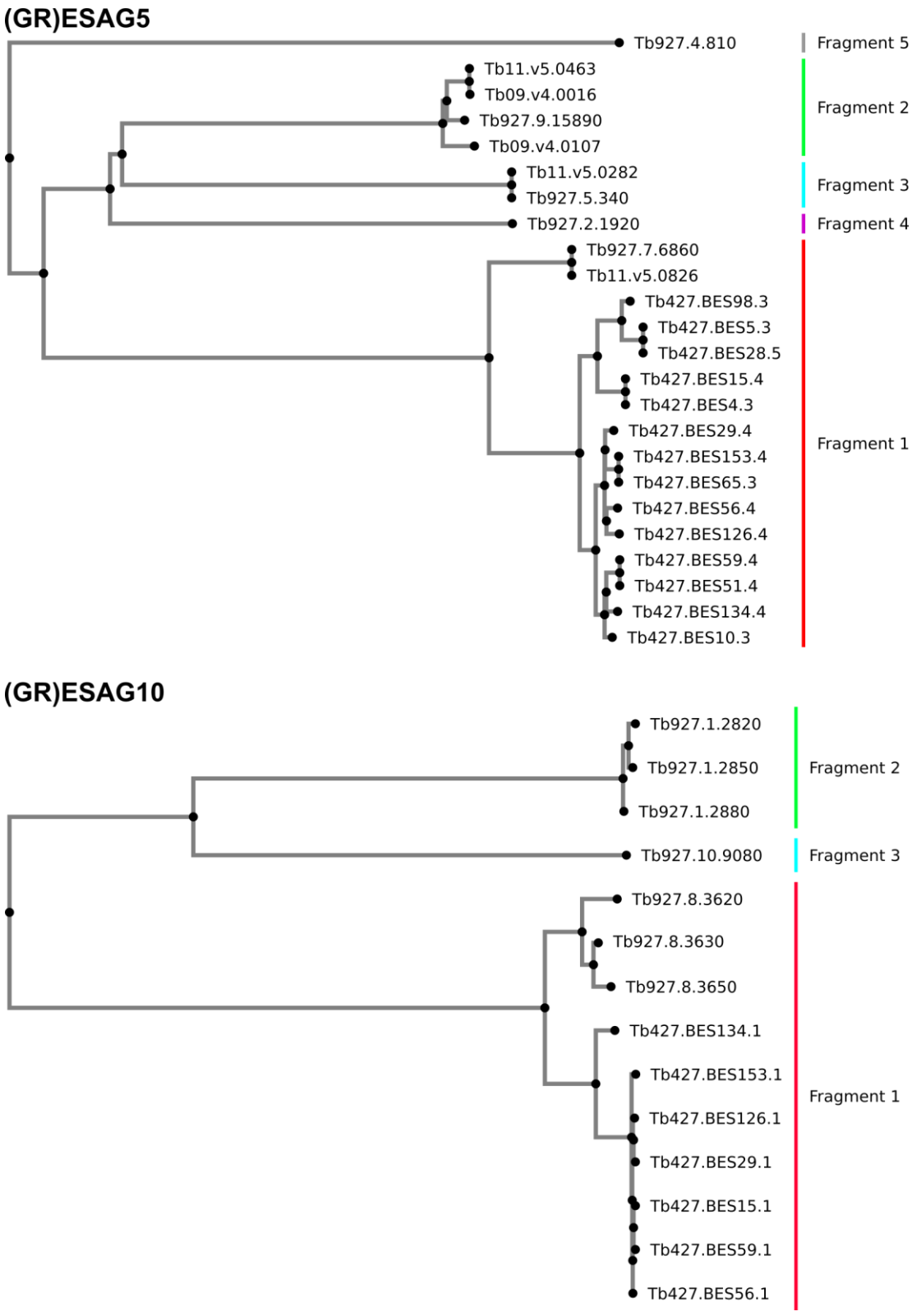
Tb427.BES126.1 GTTTTACTGCATGCTTCTCACGCTTGGCGCCATGGGTATGGTTCTAGTTACAGTCCTTGGCACAAGGTTGCAGCTGCTAATAACCAGTGTTTTGGCGTCATTTACACTCTGTGGACTCGGTTTCG 1297
Tb427.BES29.1 GTTTTACTGCATGCTTCTCACGCTTGGCGCCATGGGTATGGTTCTAGTTACAGTCCTTGGCACAAGGTTGCAGCTGCTAATAACCAGTGTTTTGGCGTCATTTACACTCTGTGGACTCGGTTTCG 1297
Tb427.BES153.1 GTTTTACTGCATGCTTCTCACGCTTGGCGCCATGGGTATGGTTCTAGTTACAGTCCTTGGCACAAGGTTGCAGCTGCTAATAACCAGTGTTTTGGCGTCATTTACACTCTGTGGACTCGGTTTCG 1297
Tb427.BES15.1 GTTTTACTGCATGCTTCTCACGCTTGGCGCCATGGGTATGGTTCTAGTTACAGTCCTTGGCACAAGGTTGCAGCTGCTAATAACCAGTGTTTTGGCGTCATTTACACTCTGTGGACTCGGTTTCG 1297
Tb427.BES59.1 GTTTTACTGCATGCTTCTCACGCTTGGCGCCATGGGTATGGTTCTAGTTACAGTCCTTGGCACAAGGTTGCAGCTGCTAATAACCAGTGTTTTGGCGTCATTTACACTCTGTGGACTCGGTTTCG 1297
Tb427.BES56.1 GTTTTACTGCATGCTTCTCACGCTTGGCGCCATGGGTATGGTTCTAGTTACAGTCCTTGGCACAAGGTTGCAGCTGCTAATAACCAGTGTTTTGGCGTCATTTACACTCTGTGGACTCGGTTTCG 1297
Tb427.BES134.1 GTTTTACTGCATGCTTCTCACGCTTGGCGCCATGGGTATGGTTCTAGTTACAGTCCTTGGCACAAGGTTGCAGCTGCTAATAACCAGTGTTTTGGCGTCATTTACACTCTGTGGCCTCGGTTTCG 1297
Tb927.8.3620 GTTTTACTGCATGCTTCTCACACTTGGCGCCATCGGTATGGTTCTCGTGACAGTCCTTGGCACAAGGTTGCAGCTGCTAATAACCAGTGTTGTGGCGTCATTTACACTCTGTGGACTCGGTTTCG 1135
Tb927.8.3630 GTTTTACTGCATGCTTCTCACACTTGGCGCCATCGGTATGGTTCTCGTGACAGTCCTTGGCACAAGGTTGCAGCTGCTAATAACCAGTGTTGTGGCGTCATTTACACTCTGTGGACTCGGTTTCG 1135
Tb927.8.3650 GTTTTACTGCATGCTTCTCACACTTGGCGCCATCGGTATGGTTCTCGTGACAGTCCTTGGCACAAGGTTGCAGCTGCTAATAACCAGTGTTGTGGCGTCATTTACACTCTGTGGACTCGGTTTCG 1135
Tb927.1.2820 AATATACAGCGTGCTAATAGCATTCTTTGTTGTTGGTGTTGCCATAACCTCCTTGTTTGGATCGAGGCTGCATTTACTCATTGCCGCGGCTGTCGTTTCCGTGGGGCACTGCTCTTTCAATTTCT 1165
Tb927.1.2850 AATATACAGCGTGCTAATAGCATTCTTTGTTGTTGGTGTTGCCATAACCTCCTTGTTTGGATCGAGGCTGCATTTACTCATTGCCGCGGCTGTCGTTTCCGTGGGGCACTGCTCTTTCAATTTCT 1165
Tb927.1.2880 AATATACAGCGTGCTAATAGCATTCTTTGTTGTTGGTGTTGCCATAACCTCCTTGTTTGGATCGAGGCTGCATTTACTCATTGCCGCGGCTGTCGTTTCCGTGGGGCACTGCTCTTTCAATTTCT 1165
Tb927.10.9080 AGTTTACAGCATTCTCATTGGTTTGTTTGTACTGGTTGTGGCTGTAACTTCCGTTTTAGGTTCAACGACCTACCTAACTGTGGTAGCTATTGTGGTTTCCTTGTTGCACTGCAGTTTCAATTTCT 1180

Tb427.BES126.1 TTGCTCTTCCACTTGTGATCGCGAAGGCAAATATGTTCACATTCATCTCTAGGGTCGCGTACATCCAACTTCCTGGCGCAATCGACAATGTCTTTATGGCGACTCCCGATTGTTTTCCTGGTGGT 1422
Tb427.BES29.1 TTGCTCTTCCACTTGTGATCGCGAAGGCAAATATGTTCACATTCATCTCTAGGGTCGCGTACATCCAACTTCCTGGCGCAATCGACAATGTCTTTATGGCGACTCCCGATTGTTTTCCTGGTGGT 1422
Tb427.BES153.1 TTGCTCTTCCACTTGTGATCGCGAAGGCAAATATGTTCACATTCATCTCTAGGGTCGCGTACATCCAACTTCCTGGCGCAATCGACAATGTCTTTATGGCGACTCCCGATTGTTTTCCTGGTGGT 1422
Tb427.BES15.1 TTGCTCTTCCACTTGTGATCGCGAAGGCAAATATGTTCACATTCATCTCTAGGGTCGCGTACATCCAACTTCCTGGCGCAATCGACAATGTCTTTATGGCGACTCCCGATTGTTTTCCTGGTGGT 1422
Tb427.BES59.1 TTGCTCTTCCACTTGTGATCGCGAAGGCAAATATGTTCACATTCATCTCTAGGGTCGCGTACATCCAACTTCCTGGCGCAATCGACAATGTCTTTATGGCGACTCCCGATTGTTTTCCTGGTGGT 1422
Tb427.BES56.1 TTGCTCTTCCACTTGTGATCGCGAAGGCAAATATGTTCACATTCATCTCTAGGGTCGCGTACATCCAACTTCCTGGCGCAATCGACAATGTCTTTATGGCGACTCCCGATTGTTTTCCTGGTGGT 1422
Tb427.BES134.1 TTGCTCTTCCACTTGTGATCGCGAAGGCAAATATGTTCACATTCATCTCTAGGGTCGCGTACATCCAACTTCCTGGCGCAATCGACAATGTCTTTATGGCGACTCCCGATTGTTTTCCTGGTGGT 1422
Tb927.8.3620 TTGCCCTTCCACTTGTGATCGCGAAGGCAAATATGTTCACATTCATCTCGAGGGTCACGTACATCCAACTTCGTGGCGCAATGGACAATGTCTTCATGGCGACCCCCGATTGTTTCCCTGGTGGT 1260
Tb927.8.3630 TTGCTCTTCCACTTGTGATCGCGAAGGCAAATATGTTCACATTCATCTCGAGGGTCGCGTACATCCAACTTCCTGGCGCAATCGACAATGTCTTCATGGCGACCCCCGATTGTTTCCCTGGTGGT 1260
Tb927.8.3650 TTGCCCTTCCACTTGTGATCGCGAAGGCAAATATGTTCACATTCATCTCGAGGGTCGCGTACATCCAACTTCCTGGCGCAATCGACAATGTCTTCATGGCGACCCCCGATTGTTTCCCTGGTGGT 1260
Tb927.1.2820 ATGCGCTGTCGTGGGCGGTCGCGAAGGTGAATCTGTTCAGTTACTTGAACAATGCGTCTACGGTGTCTTTTGGTGGTGCTGTCACCCCGTTCTTCCTGAGCGGACCCGATTGCAATCCGGGTGGA 1290
Tb927.1.2850 ATGCGCTGTCGTGGGCGGTCGCGAAGGTGAATCTGTTCAGTTACTTGAACAATGCGTCTACGTTATTTTTTGGTGGTGCTGTCACCCCGTTCTTCCTGAGCGGACCCGATTGCAATCCGGGTGGA 1290
Tb927.1.2880 ATGCGCTGTCGTGGGCGGTCGCGAAGGTGAATCTGTTCAGTTACTTGCACAATGCGTCTACGGTGTCTTTTGGTGGTGCTGTCACCCCGTTCTTCCTGACCGGACCCGATTGCAATCCGGGTGGA 1290
Tb927.10.9080 ACGCTTTTCCGTTGATCATTGCGAAGGCTAACCTCTTCGGTTACTTGCAGAGGGCCACATCAGTAAGCTTCCCCGGTGCACTTTCAAACTTCTTTATCGCTGAAGCCGATTGTCTCGCAGATGGC 1305

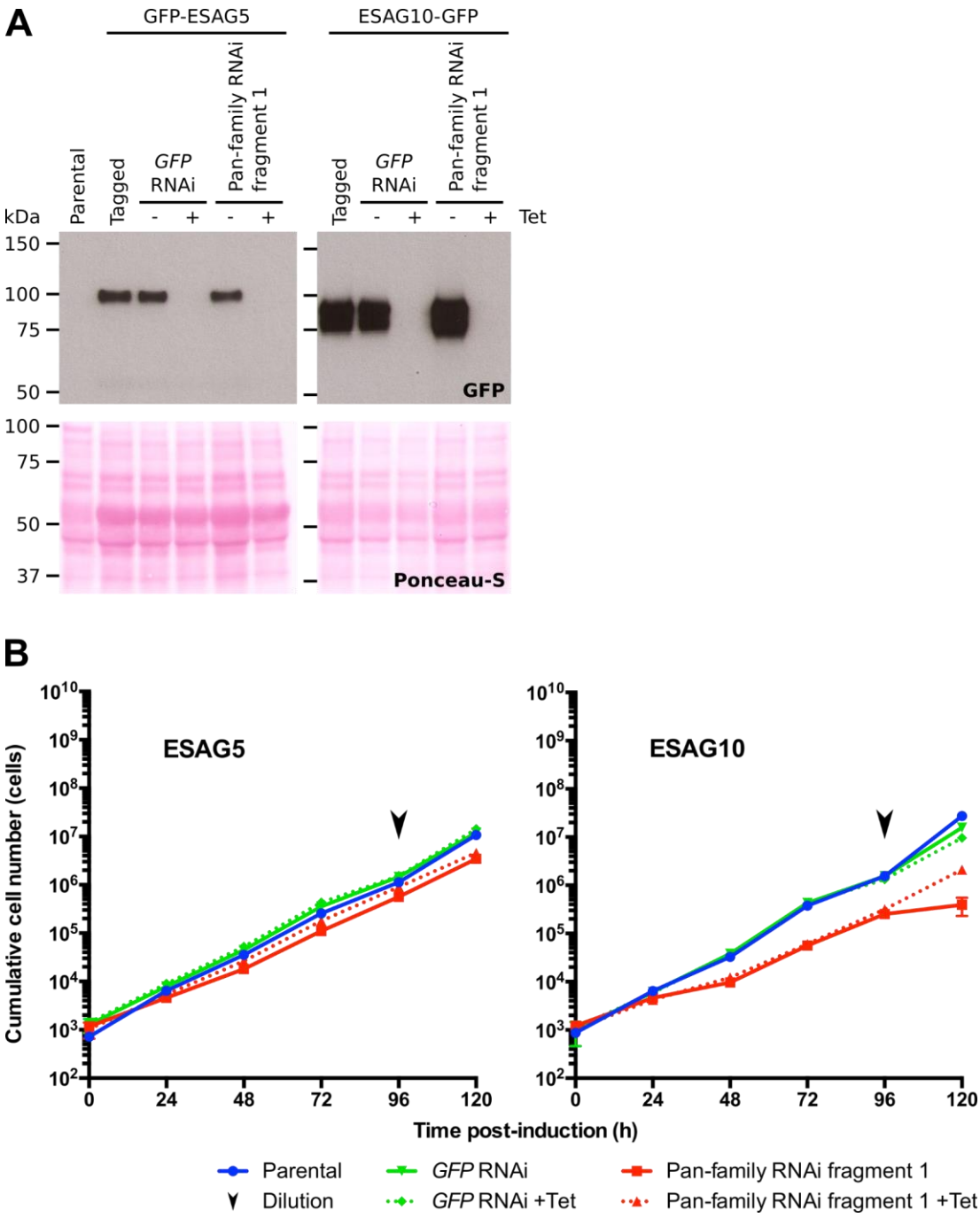

Tb427.BES126.1 CCTAACTTTTCATACTTTTACTACAGCACTGTGGGTAATATGATTGGGGCTGTTGGTGGTGTCATTGGTGTGACGTTATTCAGATACGTATTCTCAAAGCGCAGTTATCGACTAACATTCATCGT 1547
Tb427.BES29.1 CCTAACTTTTCATACTTTTACTACAGCACTGTGGGTAATATGATTGGGGCTGTTGGTGGTGTCATTGGTGTGACGTTATTCAGATACGTATTCTCAAAGCGCAGTTATCGACTAACATTCATCGT 1547
Tb427.BES153.1 CCTAACTTTTCATACTTTTACTACAGCACTGTGGGTAATATGATTGGGGCTGTTGGTGGTGTCATTGGTGTGACGTTATTCAGATACGTATTCTCAAAGCGCAGTTATCGACTAACATTCATCGT 1547
Tb427.BES15.1 CCTAACTTTTCATACTTTTACTACAGCACTGTGGGTAATATGATTGGGGCTGTTGGTGGTGTCATTGGTGTGACGTTATTCAGATACGTATTCTCAAAGCGCAGTTATCGACTAACATTCATCGT 1547
Tb427.BES59.1 CCTAACTTTTCATACTTTTACTACAGCACTGTGGGTAATATGATTGGGGCTGTTGGTGGTGTCATTGGTGTGACGTTATTCAGATACGTATTCTCAAAGCGCAGTTATCGACTAACATTCATCGT 1547
Tb427.BES56.1 CCTAACTTTTCATACTTTTACTACAGCACTGTGGGTAATATGATTGGGGCTGTTGGTGGTGTCATTGGTGTGACGTTATTCAGATACGTATTCTCAAAGCGCAGTTATCGACTAACATTCATCGT 1547
Tb427.BES134.1 CCTAACTTTTCATACTTTTACTACAGCACTGTGGGTAATATGATTGGGGCTGTTGGTGGTGTCATTGGTGTGACGTTATTCAGATACGTATTCTCAAAGCGCAGTTATCGACTAACGTTCATCGT 1547
Tb927.8.3620 CCTAACTTTTCATACTTTTACTACAACACTGTGGGTAACGTGATTGGGACTATGGGTGGTGTTATTGGTGTGACGTTATTCAGATACGTATTCTCAAAGCGCAGCTATCGCCTAACATTCATCGT 1385
Tb927.8.3630 CCTAACTTTTCATACTTTTACTACAGCACTGTGGGTAATATGATTGGAGCTGTGGGTGGTGTTATTGGTGTGACGTTATTCAGATACGTATTCTCAAAGCGCAGCTATCGCCTAACATTCATCGT 1385
Tb927.8.3650 CCTAACTTTTCATACTTTTACTACAGCACTGTGGGTAATATGATTGGAGCTGTGGGTGGTGTTATTGGTGTGACGTTATTCAGATACGTATTCTCAAAGCGCAGCTATCGCCTAACATTCATCGT 1385
Tb927.1.2820 CCTCACTTCAGTCTTTTTTTCCTGCAGACTGTTGGTGGAACAGTTGGAAGTGTCACTTCAGCTCTGGGTGTCGTGCTGTTCAATTACTTCCTTTCAAAAAAGACTTATCGGATGACTTACTTCTT 1415
Tb927.1.2850 CCTCACTTCAGTCTTTTTTTCCTGCAGACTGTTGGTGGAACAGTTGAAAGTGTCACTTCAGCTCTGGGTGTCGTGCTGTTCAATTACTTCCTTTCAAAAAAGACTTATCGGATGACTTACTTCTT 1415
Tb927.1.2880 CCTCACTTCAGTCTTTTTTTCCTGCAGACTGTTGGTGGAACAGTTGGAAGTGTCACTTCAGCTCTGGGTGTCGTGCTGTTCAATTACTTCCTTTCAAAAAAGACTTATCGGATGACTTACTTCTT 1415
Tb927.10.9080 CCGCACTTCAGCTTTGCCTTCTATCAGACTGTTGGAGGCATTATCAGCTGCGCCGCATCGATAGTTGGCATCATGCTGTTCAACTACATTTTTTCGAAACGCACGTACCGAATGACCTACTTGTC 1430
Tb427.BES126.1 AACAACTTTGATTGAGATAGTTTCGAGTATATTTGATATAATCATTGTGGAGCGATGGAACCGTCCGTATGTTAGCGATCACGTTGTATTCGTTCTTGGTGACCAAATTATACACCAGGTGTGCT 1672
Tb427.BES29.1 AACAACTTTGATTGAGATAGTTTCGAGTATATTTGATATAATCATTGTGGAGCGATGGAACCGTCCGTATGTTAGCGATCACGTTGTATTCGTTCTTGGTGACCAAATTATACACCAGGTGTGCT 1672
Tb427.BES153.1 GACAACTTTGATTGAGATAGTTTCGAGTATATTTGATATAATCATTGTGGAGCGATGGAACCGTCCGTATGTTAGCGATCACGTTGTATTCGTTCTTGGTGACCAAATTATACACCAGGTGTGCT 1672
Tb427.BES15.1 AACAACTTTGATTGAGATAGTTTCGAGTATATTTGATATAATCATTGTGGAGCGATGGAACCGTCCGTATGTTAGCGATCACGTTGTATTCGTTCTTGGTGACCAAATTATACACCAGGTGTGCT 1672
Tb427.BES59.1 AACAACTTTGATTGAGATAGTTTCGAGTATATTTGATATAATCATTGTGGAGCGATGGAACCGTCCGTATGTTAGCGATCACGTTGTATTCGTTCTTGGTGACCAAATTATACACCAGGTGTGCT 1672
Tb427.BES56.1 AACAACTTTGATTGAGATAGTTTCGAGTATATTTGATATAATCATTGTGGAGCGATGGAACCGTCCGTATGTTAGCGATCACGTTGTATTCGTTCTTGGTGACCAAATTATACACCAGGTGTGCT 1672
Tb427.BES134.1 AACAACTTTGATTGAGATAGTTTCGAGTATATTTGATATAATCATTGTGGAGCGATGGAACCGTCCGTATGTTAGCGATCACGTTGTATTCGTTCTTGGTGACCAAATTATACACCAGGTGTGCT 1672
Tb927.8.3620 AACAACTTTGATTGAGATAGTTTCGAGTATCTTTGATATAATCATTGTGGAGCGATGGAACCGTCCGTATGTTAGCGATCACGTTGTATTCGTTCTTGGTGACCAAATTATACAGCAGGTGTGCT 1510
Tb927.8.3630 AACAACTTTGATTGAGATAGTTTCGAGTATCTTTGATATAATCATTGTGGAGCGATGGAACCGTCCGTATGTTAGCGATCACGTTGTATTCGTTCTTGGTGACCAAATTATACACCAGGTGTGCT 1510
Tb927.8.3650 AACAACTTTGATTGAGATAGTTTCGAGTATCTTTGATATAATCATTGTGGAGCGATGGAACCGTCCGTATGTTAGCGATCACGTTGTATTCGTTCTTGGTGACCAAATTATACACCAGGTGTGCT 1510
Tb927.1.2820 GACACTGGTGTTCCTCATCGTCTCCAACATCTTCGACTTCATTATAGTGATGCGATGGAACCGGCCATACGTCAGTGACTACTTGATTTACTTCTGCTGCGATGTCGTCATATCTCCTGTGATCG 1540
Tb927.1.2850 GACACTGGTGTTCCTCATCGTCTCCAACATCTTCGACTTCATTATAGTGATGCGATGGAACCGGCCATACGTCAGTGACTACTTGATTTACTTCTGCTGCGATCTCGTCATATCTCCTGTGATCG 1540
Tb927.1.2880 GACACTGGTGTTCCTCATCGTCTCCAACATCTTCGACTTCATTATAGTGATGCGATGGAACCGGCCATACGTCAGTGACTACTTGATTTACTTCTGCTGCGATGTCGTCATATCTCCTGTGATCG 1540
Tb927.10.9080 GACACTTTCCCTCGGCATTTTTTCAAGTATCTTCGATCTTGTACTTGTTATGAGGTGGAACCGGCCGCACGTTAGTGATCACGCGCTGTACATCATGGGTGATGCCATTATATCTCCCGTGATTT 1555
Tb427.BES126.1 ACATGATGCACTTCATGCCAACTGTTATTCTCATATCGAGGCTCTGTCCCAGTGGGTATGAAAGTGCGGTGTATGCGGTACTTGCGGGATGTGCCCATTTTGGCCGTTCCGTTTCCAACACCCTT 1797
Tb427.BES29.1 ACATGATGCACTTCATGCCAACTGTTATTCTCATATCGAGGCTCTGTCCCAGTGGGTATGAAAGTGCGGTGTATGCGGTACTTGCGGGATGTGCCCATTTTGGCCGTTCCGTTTCCAACACCCTT 1797
Tb427.BES153.1 ACATGATGCACTTCATGCCAACTATTATTCTCATATCGAGGCTCTGTCCCAGTGGGTATGAAAGTGCGGTGTATGCGGTACTTGCGGGATGTGCCCATTTTGGCCGTTCCGTTTCCAACACCCTT 1797
Tb427.BES15.1 ACATGATGCACTTCATGCCAACTGTTATTCTCATATCGAGGCTCTGTCCCAGTGGGTATGAAAGTGCGGTGTATTCGGTACTTGCGGGATGTGCCCATTTTGGCCGTTCCGTTTCCAACACCCTT 1797
Tb427.BES59.1 ACATGATGCACTTCATGCCAACTGTTATTCTCATATCGAGGCTCTGTCCCAGTGGGTATGAAAGTGCGGTGTATGCGGTACTTGCGGGATGTGCCCATTTTGGCCGTTCCGTTTCCAACACCTTT 1797
Tb427.BES56.1 ACATGATGCACTTCATGCCAACTGTTATTCTCATATCGAGGCTCTGTCCCAGTGGGTATGAAAGTGCGGTGTATGCGGTACTTGCGGGATGTGCCCATTTTGGCCGTTCCGTTTCCAACACCCTT 1797
Tb427.BES134.1 ACATGATGCACTTCATGCCAACTGTTATTCTCATATCGAGGCTCTGTCCCAGTGGGTATGAAAGTGCGGTGTATGCGGTACTTGCGGGATGTGCCCATTTTGGCCGTTCCGTTTCCAACACCCTT 1797
Tb927.8.3620 ACATGATGCACTTCATGCCAACTGTAATGCTCATTTCGAGGCTCTGTCCCCGTGGGTCTGAAAGTATGGTGTATGCGGTACTTGCAGGATGTGCGCATTTTGGCCGTTCCGTTTCCAACACACTT 1635
Tb927.8.3630 ACATGATGCACTTCATGCCAACTGTAATGCTCATTTCGAGGCTCTGTCCCCGTGGGTCTGAAAGTATGGTGTATGCATTACTTGCAGGATTCGCTAATTTTGGCCGTTCTCTTTCCAACACACTT 1635
Tb927.8.3650 ACATGATGCACTTCATGCCAACTGTAATGCTCATTTCGAGGCTCTGTCCCCGTGGGTCTGAAAGTATGGTGTATGCATTACTTGCAGGATTCGCTAATTTTGGCCGTTCTCTTTCCAACACACTT 1635
Tb927.1.2820 GGATGATGAACTGGATGCCGTTAACCATTATTCTTTCACGGTTGTGTCCTCGTGGAAGCGAGAGTACTGTGTACGCTATTCTTGCTGCATCGAGCAATTTCGGTAGCACGGTGGGTAGTTCCCTT 1665
Tb927.1.2850 GGATGATGAACTGGATGCCGTTAACCATTATTCTTTCACGGTTGTGTCCTCGTGGAAGCGAGAGTACTGTGTACGCTATTCTTGCTGCATCGAGCAATTTCGGTAGCACGGTGGGTAGTTCCCTT 1665
Tb927.1.2880 GGATGATGCACTGGATGCCGTTAACCATTATTCTTTCACGGTTGTGTCCTCGTGGAAGCGAGAGTACTGTGTACGCTATTCTTGCTGCATCGAGCAATTTCGGCGGCACGGTGGGTAGTTCCCTT 1665
Tb927.10.9080 CCATGTTGAATTGGATGCCTCAAATTGTCCTACTTTCGCGGTTGTGTCCTAGGGGAAGTGAAAGCACAGTGTATTTCCATTCTTGCCGCCTCCAGCAACTTGGGTGCTTCCATGGGAAGTGTCGTT 1680



Appendix 9. The *(GR)ESAG10* family is divergent at the DNA level. Multiple DNA sequence alignment of *(GR)ESAG10* family genes. Nucleotides coloured based on consensus over all sequences, except in boxes, where they are coloured based on consensus of sequences within the box. RNAi fragments are highlighted: fragment 1, red; 2, green and 3, light blue. Black boxes surround the genes targeted by these respective RNAi fragments.



Appendix 10. Pan (GR)ESAG5 and (GR)ESAG10 family knockdown requires multiple RNAi fragments for full coverage. Neighbour-joining hierarchical clustering of *(GR)ESAG5* and *(GR)ESAG10* family DNA sequences. Generated using MAFFT and Phylo.io 1.0.0 with the Jukes-Cantor substitution model. Based on 1316 and 1318 conserved sites in *(GR)ESAG5* and *(GR)ESAG10* sequences, respectively. Genes targeted by different RNAi fragments required for pan-family knockdown are shown.



Appendix 11. Pan-family RNA fragment 1 mediates ablation of GFP-ESAG5 or ESAG10-GFP, but does not induce growth defects. A, Immunoblot of whole cell lysate from GFP-tagged ESAG5 and ESAG10, and derivative *GFP* and pan-family RNAi fragment 1 RNAi lines, shows ablation of chimeric protein in both cases. RNAi lines were grown in the absence (-) or presence (+) of tetracycline for 24 hours to induce RNAi. Parental VSG121-positive SmOx sample is included as a negative control. Chimeric protein detected with monoclonal anti-GFP antibodies. Expected approximate chimeric protein sizes (post signal peptide sequence cleavage): GFP-ESAG5, 82 and ESAG10-GFP, 105 kDa. Nitrocellulose membrane stained with Ponceau-S provided as a loading control. **B**, Growth of GFP-tagged ESAG5 and ESAG10 lines, and derivative RNAi lines was monitored for five days. Neither *GFP* nor pan-family RNAi fragment 1 RNAi induce growth defects. Parental, *GFP* RNAi and pan-family RNAi fragment 1 cultures are shown in blue, green and red, respectively. Cultures containing tetracycline are shown with a dashed line, absence of tetracycline is indicated by a solid line. Arrowheads indicate which day cultures were diluted into fresh media to maintain log-phase growth.

Appendix 12.

Professional Internships for PhD Students Reflection Form

Name of Organization

Warwick Ventures

Details of Placement
Please describe your main activities during the placement.

I was tasked with producing a report outlining how the Research Excellence Framework 2014 (REF2014) attributed impact in the History Unit of Assessment (UOA), as well as assisting Warwick Ventures Business Development Managers in their roles supporting and developing spin-out companies from the University of Warwick. I worked with a number of the University's spin-out companies in various roles, these included producing marketing materials, undertaking market research, planning a new lab facilities and liaising with consultants and academics to communicate my findings. Warwick Ventures wished to understand how the REF2014 History panel attributed 'impact' to case studies from this UOA, by understanding this they aim to be able to provide better support to the University, to improve in this category for the next REF assessment. The findings of my report detailed how case studies reporting contributions to media productions and publications that generated wealth, and reached a wide audience, received very high impact scores. I suggested that identifying specific academics or projects whose research could be developed to provide such outcomes should be a focus for Warwick Ventures.

Placement Achievements
Please detail all outcomes from the placement, including any publications, presentations given and reports written etc.

The report I generated into how the REF2014 attributed impact to History case studies was the most significant output of my time at Warwick Ventures. To produce this report I needed to extract information from a representative set of case studies from the History UOA, with a variety of impact scores. This included designing the study, deciding what information was needed, how to go about collecting this, and then critically analysing the data. I produced marketing materials for a spinout company, VirionHealth. These included a datasheet that was presented to potential investors. A plan was put together detailing specifications, equipment and construction works required to relocate Stoli Catalysis to new premises. This required a great deal of research and input from consultants. Finally, I undertook preliminary marketplace research into therapeutic cyclic peptides, for a promising future University spin-out company. GlobalData and

Espacenet databases were mined for current and trial drugs of this class, including patents, which were compiled to provide to an external consultant to conduct further research.

Skill development
Has this Placement helped you developed any new skills or enhanced your previous skill set?

The placement allowed me to apply a number of skills I have learned throughout my academic career, but mainly the PhD, to different and unfamiliar situations. Conducting research and handling datasets are not new skills for me, but the placement highlighted to me that these can be used outside of my day-to-day work in the lab, and have probably been enhanced the experience. Designing how to assess the REF2014 case studies was practice in experimental design, again something I am used to, but not in the context I had to apply this skill during the placement.

New skills include searching GlobalData and patent databases, which I had no experience using. I also had to interact with a number of academics and consultants that work outside of my field, which has increased my confidence.

Future Work
Has this Placement influenced your future career aspirations? If so, in what way?

The placement has not influenced my career aspirations.