Framework for genetic improvement of oil yield and harvestability in oil palm

TEH CHEE KENG

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

September, 2020

Acknowledgement

First of all, I would like to express my deepest gratitude to my supervisors, Prof. Festo Massawe, Dr. Sean Mayes and Dr. David Ross Appleton for all the guidance and advice throughout the PhD programme. I also want to thank them for encouraging me to publish the research findings in books and peer-reviewed journals and we have made it. Besides, I am grateful to Dr. Chew Fook Tim from National University of Singapore who has mentored and always inspired me in scientific research. I am also thankful to my colleagues in Sime Darby Plantation R&D Centre for all the good supports. Without them, the programme would not have been possible. I wish to take this opportunity to thank my beloved wife, Akiko Saito for her strong support, understanding and patience during my busy schedules. Certainly, I will not miss a chance to thank my parents and brothers for their motivations as well. At last, I would like to acknowledge Sime Darby Plantation (including Yayasan Sime Darby) and the University of Nottingham Malaysia for my PhD scholarship.

ABSTRACT

Oil palm (*Elaeis guineensis* Jacq.) contributes 35% of global vegetable oil and has become the most important oil crop in the world. It occupies only 7% of total global oilseed-harvested land with 10 times more oil yield per hectare of cropland than any other oil crop. To further improve the superior oil yield, enormous efforts have been put into selective breeding programmes in the last eight decades. Nevertheless, the breeding progress of oil palm, like other perennial crops, is extremely slow (typically 12 years per breeding cycle) and costly. In this study, major QTL (quantitative trait loci) for trunk height (THT) and correlated yield traits were discovered in low trunk increment breeding populations using single nucleotide polymorphism (SNP) and simple sequence repeat (SSR) markers. Further understanding of the correlation between THT and oil yield traits, and the effect of cumulative positive alleles of QTL can significantly increase the success rate of multiple trait integration through MAS to produce high-yielding oil palms with low height increment. This compares to an estimated 1% probability of success using conventional breeding methods. Furthermore, illegitimacy in breeding crosses and commercial seed production can be a common problem, if mismanagement happens in controlled pollination (given that oil palm is a naturally outcrossing species) and seed processing. For perennial crops such as oil palm, the illegitimacy problem can easily badly effect a 12-year breeding cycle and illegitimacy can lead to a 25-year yield reduction in commercial planting. Therefore, a practical genomic SNP-based legitimacy testing method with accuracy >97% was developed as a standard quality control procedure for seed purity in breeding and commercial seed production. The method has also successfully addressed the inaccuracy and inefficiency of the conventional fruit census test for contamination. The overall objectives of this study and findings are reported in the thesis.

TABLE OF CONTENTS

Contents

Acknowledgement I

ABSTRACT II

TABLE OF CONTENT III

LIST OF TABLES AND FIGURES V

LIST OF ABBREVIATIONS VI

THESIS STRUCTURE VIII

LIST OF PUBLICATIONS IX

Chapter 1: INTRODUCTION 1

1.1 Oil palm as a major oil crop 1

1.2 Oil palm breeding and its constraints 2

1.3 Trait preference shift in oil palm breeding 5

1.4 Multiple trait integration into oil palm breeding and towards marker-assisted selection 6

AIMS AND OBJECTIVES 8

Chapter 2: APPLICATIONS OF GENETIC TOOLS IN OIL PALM BREEDING 11

2.1 Introduction 11

2.2 Methodology 12

2.3 Results and discussion 13

2.4 Conclusion and key recommendation for future works 16

2.5 Published book chapter contributing to the objectives in Chapter 2 17

Chapter 3: A PRACTICAL GENOMIC SNP-BASED LEGITIMACY TEST METHOD FOR OIL PALM BREEDING AND SEED PRODUCTION 28

3.1 Introduction 28

3.2 Methodology 29
3.3 Results & Discussion 30
3.4 Conclusion and key recommendation for future works 31
3.5 Published journal article contributing to the objectives in Chapter 3 32
Chapter 4: MAJOR QTL FOR TRUNK HEIGHT AND CORRELATED AGRONOMIC TRAITS PROVIDE INSIGHTS INTO MULTIPLE TRAIT INTERGRATION IN OIL PALM BREEDING 40
4.1 Introduction 40
4.2 Methodology 41
4.3 Results & Discussion 42
4.4 Conclusion and key recommendation for future work 44
4.5 Published journal article contributing to the objectives in Chapter 4 45
Chapter 5: GENERAL CONCLUSIONS 61
References 64

LIST OF TABLES AND FIGURES

Table 1: A chronological list of publications in a book and peer-reviewed journals included in the main body of the thesis

Figure 1: The single gene inheritance of shell thickness of oil palm

Figure 2: Genetic stratification and genome-wide association study for mesocarp oil yield of 2,045 tenera palms representing Deli dura x AVROS pisifera population and Nigerian dura x AVROS pisifera population. (a) Neighbour-joining tree constructed based on genetic distance. Group I and group II are Deli dura x AVROS pisifera population and Nigerian dura x AVROS pisifera population. (b) A Manhattan plot of compressed mixed linear model (MLM) for Deli dura x AVROS pisifera population. (c) A Manhattan plot of compressed MLM for Nigerian dura x AVROS pisifera population, based on Teh et al. (2016)

Figure 3: Legitimacy reference based on the parentage of Family A at illegitimacy-indicative SNP threshold=3% (the dotted red line), based on Teh et al. (2019 (Accepted))

Figure 4: Distribution of quantitative trait loci (QTL) for trunk height in 2008 (THT_2008), frond length (FL) and total oil yield per palm (O/P) on a linkage map of GM x GM population. KW significant – Kruskal-Wallis at the p = 0.005 threshold; GW significant = genome-wide logarithm of odds, to the base 10 (LOD) = 4.0 threshold; putative – 2.4 ≤ LOD ≤ 4.0 threshold; * non-cofactor, based on Teh et al. (2020)

Figure 5: An overview of key findings and milestones achieved within and across thematic areas in this research programme

Remark: Excluding the tables and the figures published in book chapter and journals.

LIST OF ABBREVIATIONS

bp	Base pair
BPRO	Breeding population of restricted origins
cM	centiMorgan
CP	Cross-pollinated population
DAV	Dumpy AVROS
DxP	Deli *dura* x AVROS *pisifera*
FFB	Fresh fruit bunch
FIPS	Family-and-individual palm selection
FL	Frond length
GCA	General combining ability
GEBV	Genomic estimated breeding value
GM	*Gunung Melayu*
GS	Genomic selection
GWAS	Genome-wide association study
GxO	Interspecific hybridisation between *E. guineensis* and *E. oleifera*
HGP	Human Genome Project
IRGSP	International Rice Genome Sequencing Project
IV	Iodine value

JL	*Johore Labis*
LD	Linkage disequilibrium
LG	Linkage group
MAS	Marker-assisted selection
MLM	Mixed linear model
O/P	Total oil yield per palm
ppm	parts per million
QTL	Quantitative trait loci
RRS	Reciprocal recurrent selection
S/F	Shell-to-fruit ratio
SCA	Specific combining ability
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat
THT	Trunk height
UR	*Ulu Remis*

THESIS STRUCTURE

The thesis is comprised of Five Chapters and also reports on **three published papers** derived from the present research programme.

1. **Chapter 1**: The section introduces the importance of oil palm as an oil crop, selective breeding and the constraints encountered by the industry. The scope includes introduction, problem statements, aims and objectives of this study.

2. **Chapter 2**: The latest development of genomic technology in oil palm is discussed. The chapter also further explains how the technology can be translated to applications for selecting high-yielding oil palms as commercial planting materials.

3. **Chapter 3**: Identification of genetic contamination in breeding populations is discussed and followed by the development of a practical SNP-based legitimacy testing method and compared to the conventional fruit census and the SHELL marker testing method.

4. **Chapter 4**: After ensuring the genetic legitimacy of assayed breeding populations, investigation of the correlation between THT and other agronomical traits, and identification of major QTL for trunk height was performed. The findings provide important insights into multiple trait integration in oil palm breeding.

5. **Chapter 5**: The section provides a general conclusion by linking the development of genomics technology to selective breeding and commercial seed production in the oil palm industry.

LIST OF PUBLICATIONS

Table 1: A list of publications in a book and peer-reviewed journals included in the main body of the thesis

No.	Title	Status
1	Teh C K, Kwong Q B, Ong A L, Mohaimi M, Sukganah A, Chew F T, Mayes S, Appleton D R, Harikrishna K (2017): Application of genomic tools in oil palm breeding. In: Soh A C, Mayes S, Roberts J A (eds) Oil Palm Breeding: Genetics and Genomics. CRC Press, Boca Raton, pp 246 – 256	Published in a book chapter
2	Teh C K, Lee H L, Hafiza A, Ong A L, Mayes S, Chew F T, Appleton D R (2019): A practical genome-enabled legitimacy assay for oil palm breeding and seed production. BMC Plant Biology 19, 470. doi.org/10.1186/s12870-019-2062-x	Published in a peer-reviewed journal
3	Teh C K, Ong A L, Mayes S, Massawe F, Appleton D R (2020): Major QTLs for trunk height and correlated agronomic traits provide insights into multiple trait integration in oil palm breeding. Genes 11, 826.doi: 10.3390/genes11070826	Published in a peer-reviewed journal

Chapter 1: INTRODUCTION

1.1 Oil palm as a major oil crop

Oil palm is a perennial flowering plant in the genus *Elaeis* and is one of the most important edible oil-producing members of Aracaceae (formally known as Palmea) family. The genus is primarily represented by two oil palm species; *E. guineensis* Jacq. and *E. oleifera* (HBK) Cortes that originated from West Africa and Latin America, respectively. Fruit kernel and mesocarp of both species can produce edible oil. To date, only the African oil palm is commercially planted outside of South America due to its superior oil yield with an average of 3.5 tonne hectare^{-1} year^{-1} in Malaysia compared to only 0.4 tonne hectare^{-1} year^{-1} produced by the Latin American species (Hardon, 1969).

Nevertheless, the South American species performs better in term of oil quality including a higher content of carotene with average 4,592 ppm in comparison to *E. guineensis* which has a range of 400 – 1,000 ppm only (Yap et al., 1997). The *E. oleifera* oil is also highly unsaturated due to higher oleic and linoleic contents, reflected in high iodine value (IV) ranging from 70 – 93 compared to a range of 32 – 65 (Corley, 1979; Ong et al., 1981; Rajanaidu et al., 1989; Sharman, 1999). With oil quality >3,000 ppm carotene and >80 IV, the *E. oleifera* oil has a great potential to be directly exploited in pharmaceuticals without any further processing (Mohd et al., 2002). This species, however, is not preferred by the planters because of poor oil yield. Various interspecific hybridisation between *E. guineensis* and *E. oleifera* (GxO) and backcrossing programmes were initiated to transfer the desirable traits from *E. oleifera* to the commercial *E. guineensis* (Amblard et al., 1995) and to retain high oil yield.

To date, the programme has yet to be proven to be successful, mainly due to inconsistent interests of breeders, particularly in Africa and Southeast Asia. The breeders tend

to spend the same resources on further improvement of *E. guineensis*, rather than yield restoration through GxO hybrid and backcross programmes which require 12 years per selection cycle. Commercial *E. guineensis* is yielding 8 to 10 times more than other oil crops, making the oil palm species as the most efficient oil crop in the world year^{-1} hectare^{-1}. This also explains why only 7% of total global oilseed-harvested land is occupied by oil palm yet 35% of the global vegetable oils and fats is from palm oil (OilWorld, 2018). To retain this competitive advantage, constraints encountered in conventional breeding programmes need to be overcome, and genomics technologies can be one of the solutions.

1.2 Oil palm breeding and its constraints

The oil palm breeding programmes in Southeast Asia were solely initiated with the Deli *dura* which originated from the four palms planted in Bogor Botanical Garden in 1848 (Pamin, 1998). The same population were subsequently shared by different breeding programmes and different selection criteria applied over a few generations, leading to genetic differentiation between subpopulations, with each subpopulation termed a Breeding Population of Restricted Origin (BPROs) (Rosenquist, 1986). For instance, The *Ulu Remis* (UR) origin was selected for high bunch number and sex ratio (the ratio of female inflorescences to total inflorescences on a palm) in Marihat Baris, Sumatra (Hartley, 1967). By contrast, the *Johore Labis* (JL) origin was selected in Socfindo, Sumatra for bigger bunches (high bunch weight) and thinner fruit shell, instead of bunch number. Selection for low height increment and high oil-to-mesocarp ratio was also conducted by a Swiss company to form the *Gunung Malayu* (GM) origin in 1919 (Corley and Tinker, 2003).

Three fruit forms, namely *dura*, *tenera* and *pisifera* are naturally found in the African origin, whereas only one *dura*-like fruit form is present in the American origin. In Africa, owing to the poor performance of the available thick-shelled *dura* palms, thin-shelled *tenera* with greater mesocarp oil yield was adopted. This has paved the way for a series of test crosses

among *dura*, *pisifera* and *tenera*, eventually leading to confirmation of the single gene inheritance of the shell thickness trait in 1941 (Beirnaert and Vanderweyen, 1941). In Figure 1, thick-shelled *dura* have thinner mesocarp, thus produce lower overall mesocarp oil, whereas shell-less *pisifera* are usually female sterile. The cross between *dura* (Sh^+ Sh^+) and *pisifera* (Sh^- Sh^-) results in 100% thin-shelled heterozygous *tenera* (Sh^+ Sh^-) with increased mesocarp oil yield and restored fertility. Therefore, the major companies in Southeast Asia subsequently introduced the AVROS *pisifera* from Africa to be crossed with the existing Deli *dura* to produce commercial *tenera* planting materials in this region. The effort has paid off with a 30% increment in commercial oil yield per hectare realised after switching to *tenera* planting in the 60's (Corley and Tinker, 2003; Hardon et al., 1987). Nevertheless, this complicated the oil palm breeding programmes by requiring the separate development of the Deli *dura* and the AVROS *pisifera* as maternal and paternal parent, respectively.

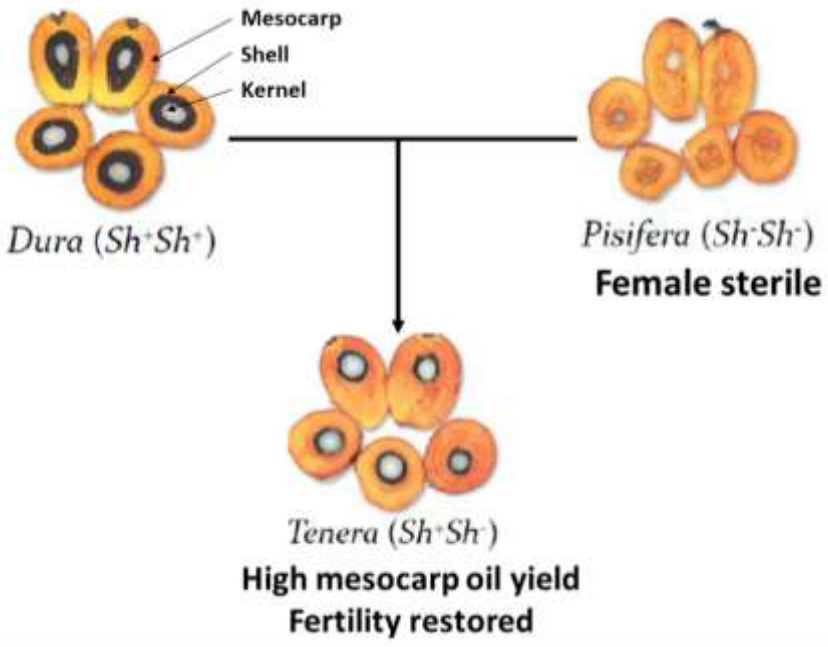


Figure 1: The single gene inheritance of shell thickness of oil palm

Two crossing schemes i.e. reciprocal recurrent selection (RRS) and family-and-individual palm selection (FIPS) are generally adopted in oil palm breeding. The RRS for elite *dura* and *pisifera* palms mainly depends on breeding values (BVs) determined by progeny testing, while the FIPS is directly based on the best performance of individual palms (Corley and Tinker, 2003). As described earlier, the breeding programmes in Southeast Asia was dominated by the Deli *dura* origin before 1960s, thus FIPS was the best method to improve oil and bunch yields at that time. Each selection cycle typically requires 12 years (50% time spent for phenotyping) to complete. The application of RRS was then used to determine general combining ability (GCA) and specific combining ability (SCA) between the Deli *dura* and the AVROS *pisifera* (DxP) families. Breeders are unable to produce all DxP combinations with a large population size for progeny testing, simply due to resource limitation where only 136 – 148 palms are usually planted in a hectare of land. This may result in the best combinations not being detected, and eventually hinder the breeding progress (Hardon, 1970). Hence, most of Malaysian breeding programmes have shifted back to FIPS to focus more on GCA to reduce dependency on progeny test, but selection response for those traits, such as FFB (fresh fruit bunch) with poor heritability is usually compromised. To improve this, DNA marker development has taken place to identify QTL for traits in oil palm since the 90's; however, most of the findings are still in the discovery phase.

Furthermore, incorrect parentage due to genetic (usually pollen) contamination can be a problem for oil palm breeding programmes. Controlled pollination is essential to ensure the DxP progeny testing reflects the true families or origins. However, genetic contamination cannot be totally avoided in controlled pollination and seed production (Corley, 2005), so a robust quality control method should be always in place. Currently, the conventional fruit census (fruit cross-section) refers to the presence of a fibre ring around the kernel and shell thickness to determine genetic purity of a DxP family. Any unexpected *pisifera* and *dura* palms

from a DxP family are identified as non-*tenera* contamination, but the fruit census is laborious and only feasible when the palms start fruiting at 2 – 3 years after field planting. For commercial planting, in field detection happens too late to replace those contaminants, with an ensuing long-term yield loss throughout the 20-25 years of production lifespan. To solve the problem, novel mutations in the *SHELL* gene responsible for fruit form (*dura, pisifera* or *tenera*) could be used as DNA markers for early detection of non-*tenera* contamination at the nursery stage (Singh et al., 2013a). Still, the conventional fruit census and the SHELL testing both have the same limitation, which is the inability to detect illegitimate *tenera* derived from incorrect DxP parentage.

1.3 Trait preference shift in oil palm breeding

As described earlier, the breeding objectives in oil palm still mainly focus on oil yield in the mesocarp and kernel to maximise plantation profit. In the past 10 years, acute labour shortage for collecting FFB and loose fruit in the fields has become a major problem, thus breeders are shifting their focus for harvestability traits. For example, foreign labours represents over 70% of the total 491,339 workforce of oil palm plantations in Malaysia, reflecting alarmingly dependency on migrants, mainly supplied from Indonesia, Bangladesh, the Philippines, Thailand and Myanmar (Hanim, 2012). This foreign workforce is keener to serve the oil palm plantations in their home countries for similar or higher wages. During the COVID-19 pandemic period, the problem has become worse because of government restrictions on employment of foreign workers. Developing problems with labour and adjustments to plantation management practice to account for these partly explains the stagnant oil yield in Malaysia, even though superior planting materials are available in the country. Sarawak might incur up to a RM1 billion loss per year in uncollected FFB because of labour shortage, ranging from 15 to 20% (Diana, 2013) across plantations. Some

mechanisation has been introduced to harvest FFB, but the commercial applicability is still very limited.

Therefore, low height increment or the 'dumpiness' trait has become the second most agriculturally important trait in oil palm, after oil yield. The oil palms with low THT can ease FFB harvest in the field. Interestingly, the trait is also reported to reduce the risk of fruit damage when the FFB hit the ground, leading to higher quality of the oil extracted (Hadi et al., 2009). This motivates breeders to revisit their low height increment breeding populations. One of them is Deli *dumpy* E206, which had unusual large girth and slow stem height increment. The self-pollinated progeny of Deli *dumpy* E206 was found to be uniformly short, indicating that the trait was highly inheritable (Jagoe, 1952). The materials were then included in many breeding programmes, especially in Southeast Asia. For example, low height increment of dumpiness was retained in the Deli *dura* with a three-quarter *dumpy* ancestry x AVROS *pisifera* (Soh, 1981). Another similar breeding programme is being conducted in Sime Darby Plantation R&D Centre by using the AVROS *pisifera* with 25% *dumpy* origin (known as *dumpy* AVROS; DAV) as a pollen source to pollinate another short-stem GM *dura* population in order to further accelerate the THT reduction of *tenera*. However, the inheritance of dumpiness trait is yet to be reported from this breeding programme.

1.4 Multiple trait integration into oil palm breeding and towards marker-assisted selection

The oil palm industry, especially in Malaysia is experiencing labour shortages which have become more severe than other producing countries. Hence, the breeding objective of superior oil yield needs to be expanded to integrate other traits, such as low THT and short frond length (FL) (small canopy) for vegetative compactness to improve harvest and field management efficiency. Multiple trait integration in oil palm has gained breeder's attention in recent years, but it is challenging. In many cases, breeding fitness of a cultivar may drop

when introgression from non-elite of germplasm takes place. This is mainly due to linkage drag of unwanted effects from the wild donor of non-targeted genes or QTL linked to the targeted genes. The only solution is to perform rounds of selection and backcrossing cycles with sufficient progeny to identify the recombinants without unwanted linkages to restore an elite background. Nonetheless, the phenotypic-based selection in perennial crops, such as oil palm is extremely costly and time consuming. The environmental effects on polygenic/complex traits including oil yield and THT are also prominent, presenting a great challenge in capturing the genetic effects for crop improvements. By locating the QTL underlying the targeted traits, the breeders can adopt these QTL as DNA markers to quickly detect and select the best combinations of alleles from each cross at seed stage. The conventional phenotyping cycles eventually can be minimised dramatically using the MAS programmes in oil palm. The reported studies are promising, but mainly stop at discovery phase.

AIMS AND OBJECTIVES

The research project therefore aims to develop a baseline for genetic discoveries for agronomic traits and translation to applications in oil palm breeding and seed production. Aims and objectives below is structured into five chapters.

1. To understand the latest development of genomics study and its applications in oil palm industry. An overview of oil palm breeding and DNA marker development for QTL mapping is reported and published.

2. To identify 'hidden' *tenera* and *dura* contamination in oil palm using SNP markers. A SNP-based legitimacy array which consists of 200 genomic SNP markers was developed and has proven to be efficient in detecting both *tenera* and *dura* contamination. The testing method for genetic purity control in oil palm is reported and published and has since been fully implemented as a quality control measure in Sime Darby Plantation Seed production.

3. To investigate genetics of THT and correlated agronomical traits in two low height increment breeding populations. Correlations between THT and oil yield traits, and their QTL were used to develop a method for multiple trait integration. The findings are reported and published.

Chapter 1: INTRODUCTION

Chapter 2: APPLICATIONS OF GENETIC TOOLS IN OIL PALM BREEDING

This chapter reviews the challenges encountered in oil palm breeding programmes and the potential solutions reported in the research community.

1. To understand the fundamentals of oil palm breeding programmes and their challenges

2. To assess the latest development of genomic technology in oil palm and other organisms

3. To propose MAS programmes and deployment in oil palm breeding programmes

Chapter 3: A PRACTICAL GENOMIC SNP-BASED LEGITIMACY TEST METHOD FOR OIL PALM BREEDING AND SEED PRODUCTION

Without a good control of genetic purity, any breeding programme and commercial seed production programme may end with failure. This chapter includes the development of a practical genome-enabled legitimacy test using SNP markers, with specific objectives given below.

1. To evaluate an existing legitimacy testing method using SHELL markers to the gene that determines the fruit forms

2. To develop and evaluate a new genomic SNP-based legitimacy testing method

Chapter 4: MAJOR QTL FOR TRUNK HEIGHT AND CORRELATED AGRONOMIC TRAITS PROVIDE INSIGHTS INTO MULTIPLE TRAIT INTERGRATION IN OIL PALM BREEDING

The genetic legitimacy of two breeding populations (GM and GM x DAV) with low height increment was confirmed and subjected to QTL mapping for THT and other correlated agronomic traits. The specific objectives under this chapter are given below.

1. To evaluate the trait variation in reduced THT between a GM x GM population and its GM x DAV *tenera* progeny

2. To conduct QTL mapping for THT and THT-correlated agronomic traits in GM x GM population using a high-density OP200K genotyping SNP array (Kwong et al., 2016) and SSR markers

3. To consolidate the identified QTL for selected traits and MAS to accelerate reduced height breeding programmes with better trait predictability

Chapter 5: GENERAL CONCLUSIONS

Chapter 2: APPLICATIONS OF GENETIC TOOLS IN OIL PALM BREEDING

2.1 Introduction

Oil palm breeding was first started in Africa with the objective to obtain high-yielding oil palms for commercial planting. A codominant inheritance of shell thickness was discovered from a series of test crosses among *dura*, *tenera* and *pisifera* (Beirnaert and Vanderweyen, 1941) when breeders in Africa looked for alternative breeding material that can perform similarly to the superior Deli *dura* palms in Southeast Asia. A cross between thick-shelled *dura* and shell-less female-sterile *pisifera* produces 100% thin-shelled *tenera*, due to higher mesocarp percentage and oil yield, with full fertility. Consequently, an immediate 30% mesocarp oil increment was achieved in 1960s when the commercial Deli *dura* planting was shifted to DxP planting in Southeast Asia (Corley and Tinker, 2003; Hardon et al., 1987).

Two selection schemes including FIPS and RRS are commonly practised in oil palm breeding. Interestingly, FIPS was sufficient to select the best Deli *dura* palms based on their parent's phenotypic performance. The Deli *dura* population and the AVROS *tenera*/*pisifera* population were maintained separately after the introduction of DxP breeding programmes. Therefore, RRS has taken place since then to evaluate SCA and GCA between the Deli and the AVROS parents through progeny testing (Corley and Tinker, 2003). The RRS, however, is costly (only 136 – 148 palms planted in a hectare of land) and requires two generations, with more than 20 years to complete a selection cycle. The first decade is just to propagate both the maternal Deli and the paternal AVROS parents, respectively, and followed by subsequent 10 years for progeny testing between the parents. During the process, some elite parents may not be available anymore because of disease or aging, so they are usually replaced by their siblings by practice, unfortunately dilutes the detection power of SCA. Most breeding

programmes in Malaysia decided to put more focus on GCA evaluation through FIPS, so that the progeny testing can be minimised. In return, selection response for those traits with poor heritability, like FFB and stress tolerance is usually ineffective, and this only can be compensated by more selection cycles. Understanding of the QTL underlying agronomic traits for MAS programmes thus has become essential, allowing early selection of elite parents with potentially reduced progeny testing in oil palm.

Rapid advancement in genomics and DNA sequencing technologies have revealed the light at the end of the tunnel. Whole-genome sequencing is now mature enough to present the complete DNA repertoire of an organism's genome, distributed in the nucleus, mitochondria and chloroplasts. By aligning individual genomes, the genetic polymorphism in populations becomes accessible to study the inherited differences in genic or other portions of the genome which are directly responsible for phenotypic variants, influencing an individual's risk of disease and/or fitness in the environment. For instance, the three billion DNA base pairs (bp) of the Human genome were successfully sequenced and completed in 2003 under the Human Genome Project (HGP). With the published genomes, the HGP has facilitated the discovery of more than 1,800 disease-related genes in humans. Similar achievements are also reported in plants, such as rice (*Oryza sativa* L.) under the International Rice Genome Sequencing Project (IRGSP). To follow the same path, the oil palm genome was sequenced and become publicly available in 2013 (Singh et al., 2013b). The published genome sequence has become an important reference for genetic discovery programmes that are critical to expedite breeding progress in oil palm.

2.2 Methodology

The methodology used in this chapter is mainly based on literature search, discussions with colleagues and stakeholders in the oil palm industry. Experience from previous studies in Sime Darby Plantation R&D Centre is also shared. The breeding history of oil palm was compiled to

further understand how two breeding programmes i.e. RRS and FIPS were started and are still being implemented until now. By knowing the pros and cons of each approach, MAS can be incorporated properly to solve the main problems in oil palm breeding i.e. costly phenotyping and low selection response. Also, from previous research experience, genome-wide association study (GWAS) and genomic selection (GS) for oil yield traits has proven to be feasible in oil palm breeding. This serves as a useful reference to further extend the single-trait selection to multiple trait integration programmes.

2.3 Results and discussion

With a reference genome, whole-genome resequencing is the most comprehensive and cost effective method to measure polymorphism in the genome (Feuillet et al., 2011). The common polymorphisms, including SNP and SSR are readily discovered and widely used as molecular markers due to their high abundance, polymorphism, reproducibility and codominance. In oil palm, Sime Darby Plantation R&D Centre resequenced a pool of 132 individual palms representing 59 important breeding populations (*E. guineensis* and *E. oleifera*). The effort has yielded more than seven million raw SNP markers and a total of 200,000 quality-tested SNPs were selected to develop a high-density Illumina OP200K SNP array (Kwong et al., 2016). According to the minimal LD decay of 19.5 Kbp observed among the assayed oil palm populations, the OP200K SNP array with an average density of one SNP for every 11 Kbp provides an excellent mapping resolution for linkage mapping and GWAS.

The OP200K SNP array was utilised to construct an ultrahigh-density linkage map of a commercial Deli *dura* x AVROS *pisifera* family (Ong et al., 2019). The map comprised of 27,890 linked SNPs, spanned a total map length of 1,151.7 centiMorgan (cM) with an average interval of 0.04 cM and was the densest linkage map for oil palm ever reported at that time. The same linkage map was also used to extend the published physical map (Singh et al., 2013b) from 685 Mbp to 969 Mbp by anchoring unplaced scaffolds to the pseudo-molecules. To date, most

of the QTL mapping works in oil palm were based on the linkage analysis, but linkage-based mapping resolution for complex traits is usually limited by small mapping populations in breeder's fields (Billotte et al., 2010; Montoya et al., 2013). A big step forward for oil palm breeding came when the first GWAS for high mesocarp oil yield using the OP200K SNP array and the reference oil palm genome was reported in 2016 (Teh et al., 2016). Unlike linkage analysis, GWAS accesses total meiotic recombination accumulated in a large population because of evolution to increase mapping resolution. The GWAS for mesocarp oil yield was based on a discovery population with 2,045 *tenera* palms. The short-range LD distance allowed good localisation of candidate genes for the tested traits. A major QTL for mesocarp oil yield was found on Chromosome 5 in both Deli *dura* x AVROS *pisifera* population and Nigerian *dura* x AVROS *pisifera* population (Figure 2).

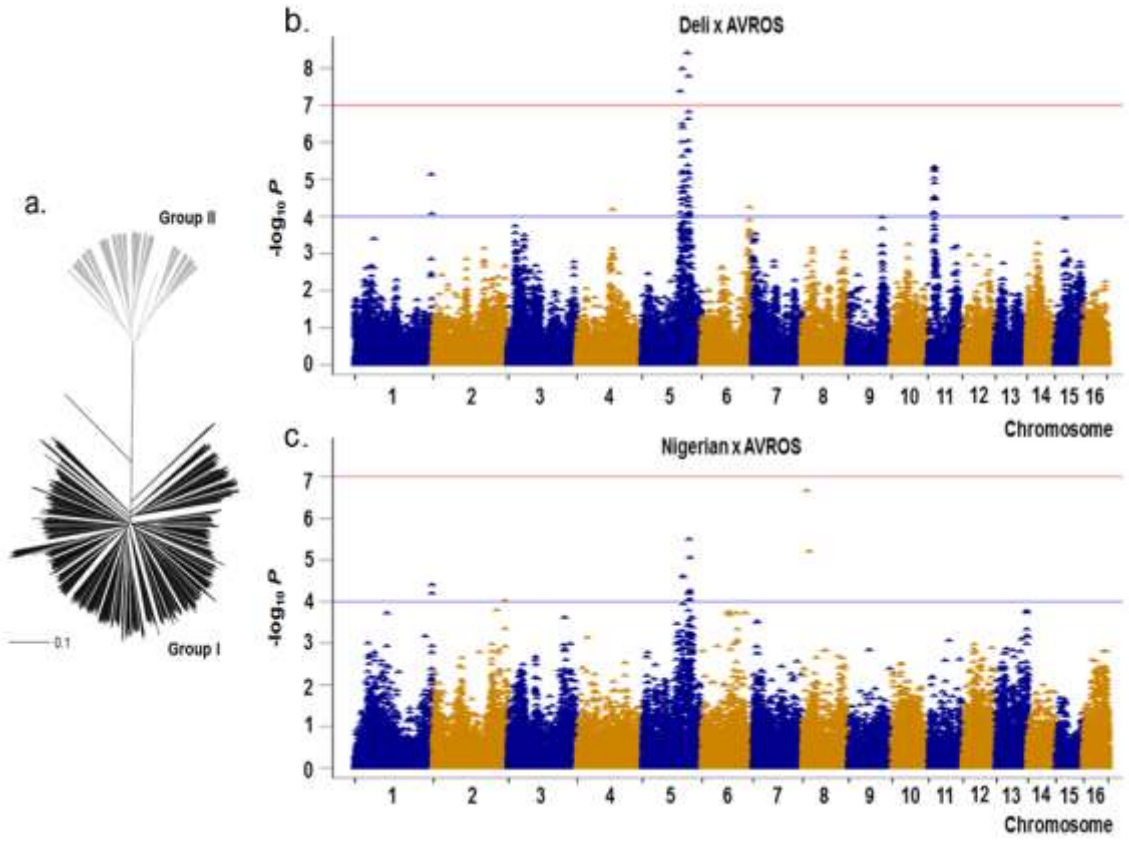


Figure 2: Genetic stratification and genome-wide association study for mesocarp oil yield of 2,045 *tenera* palms representing Deli *dura* x AVROS *pisifera* population and Nigerian *dura* x AVROS *pisifera* population. (a) Neighbour-joining tree constructed based on genetic distance. Group I and group II are Deli *dura* x AVROS *pisifera* population and Nigerian *dura* x AVROS *pisifera* population. (b) A Manhattan plot of compressed mixed linear model (MLM) for Deli dura x AVROS *pisifera* population. (c) A Manhattan plot of compressed MLM for Nigerian *dura* x AVROS *pisifera* population, based on Teh et al. (2016)

For MAS deployment, predictability of breeding value probably remains poor if the identified major QTL only explains small part of the phenotypic variation. Therefore, GS was

then evaluated in order to consolidate the total genetic effect throughout the genome relative to a targeted trait, whereas all QTL are expected to be in LD with at least one marker (Goddard and Hayes, 2007). The first GS study in oil palm using low-density SSR markers was reported (Cros et al., 2015b), and continued with reciprocal genomic selection based on simulated SNP genotypic data (Cros et al., 2015a). A further investigation of GS for a quantitative shell thickness trait (as shell-to-fruit ratio, S/F) in 312 *tenera* palms was carried out using empirical genotypic data generated from the OP200K SNP array (Kwong et al., 2016). Surprisingly, the predictive ability was high (based on correlation between genomic estimated breeding value, GEBV and the observed S/F = 0.74). Nevertheless, the values generated from linkage analysis, GWAS and GS will only be realised if they can be effectively deployed through MAS in oil palm breeding programmes.

Only a few major QTL are used in traditional MAS owing to the expense of genotyping cost, but the genetic gain can be small (Dekkers, 2004). Indeed, GS is an expanded form of MAS which covers the genome-wide breeding values of complex traits to maximise the predictive accuracy (Goddard and Hayes, 2007). Both identified QTL and GS models can be harmonised for optimising the marker density used for routine genotyping. The outcome is to break the bottlenecks in the selection cycle and land use. In Sime Darby Plantation R&D Centre, the company has successfully translated the genetic discoveries to deployment in oil palm breeding programmes and commercial seed production since 2016. The target is to achieve more than a 15% increment in oil yield of commercial DxP planting materials.

2.4 Conclusion and key recommendation for future works

The developments in molecular technology and genomics during the past decade have enabled significant improvement in throughput coupled with a reduction in costs for genome sequencing as well as genetic marker assays. Concurrent improvement in bioinformatics tools that allow detailed quantitative genetic analyses for large datasets means that the oil palm

industry now has access to the tools to develop and deploy molecular markers for selection of the best planting materials. The critical success factors for this technology to provide the much needed yield improvements in coming years will be the adaptation of molecular breeding with the planted trials and commercial seed production. This will be necessary to enable both marker discovery/development process and optimisation of future breeding programmes in order to realise the potential increase in breeding efficiency and desirable trait combinations possible through MAS.

2.5 Published book chapter contributing to the objectives in Chapter 2

Teh C K, Kwong Q B, Ong A L, Mohaimi M, Sukganah A, Chew F T, Mayes S, Appleton D R, Harikrishna K (2017): Application of genomic tools in oil palm breeding. In Soh A C, Mayes S, Robert J A (eds) Oil Palm: Genetics and Genomics. CRC Press, Boca Raton, pp 246 – 256





OIL PALM BREEDING

Genetics and Genomics





Edited by
Aik Chin Soh
Sean Mayes
Jeremy A. Roberts



CRC Press
Taylor & Francis Group

9.4 APPLICATION OF GENOMIC TOOLS IN OIL PALM BREEDING

Chee Keng Teh, Qi Bin Kwong, Ai Ling Ong, Mohaimi Mohamed, Sukganah Apparow, Fook Tim Chew, Sean Mayes, David
Ross Appleton, and Harikrishna Kulaveerasingam

9.4.1 Conventional Breeding Programs

9.4.1.1 Breeding Selection

Oil palm breeding programs were first initiated in Africa with the objective to obtain high-yielding materials for commercial planting. In the early years, apart from the Department of Agriculture in Malaysia, almost all the breeding programs were carried out by private plantation companies. One of the major challenges for the breeding of perennial crops is to achieve sufficient selection intensity and reasonable breeding gains while maintaining genetic variability for selection gains in the development of future generations. Owing to the poor quality of the available *dura* origins in Africa, *tenera* thin-shelled fruiting palms were adopted, making way to a series of carefully conducted crossing experiments that led to the discovery of the single gene inheritance of the shell thickness trait in 1941, from various test crosses made between *dura*, *tenera*, and *pisifera* (Beirnaert and Vanderweyen 1941). The study also found that the cross between *dura* and *pisifera* produced 100% *tenera*, which was the best combination for high oil yield due to a higher bunch number and thicker mesocarp. To date, the same approach is still being practiced for selecting the elite *dura* and *pisifera* parents, based on the performance of their *dura* × *pisifera* (*tenera*) progeny.

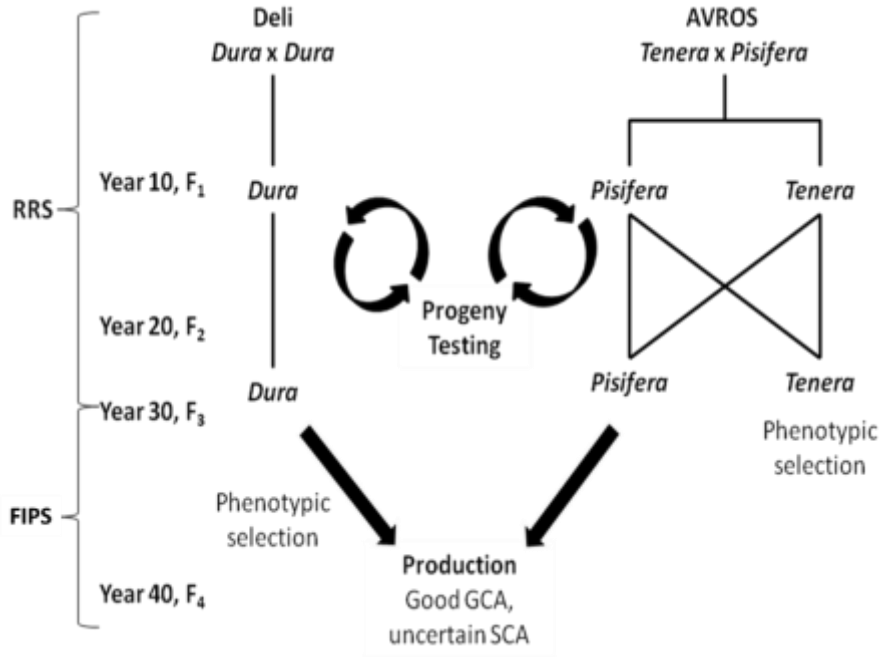


FIGURE 9.7 Conventional selection schemes in oil palm breeding. RRS—reciprocal recurrent selection; FIPS—family-and-individual palm selection; GCA—general combining ability; SCA—specific combining ability.

9.4.1.2 Constrains

Two selection schemes, that is, reciprocal recurrent selection (RRS) and family- and-individual palm selection (FIPS) are generally deployed in oil palm breeding. Figure 9.7 illustrates current breeding practice in oil palm. FIPS predominantly relies on the selection of individual palms with the best performance, while RRS is based on the breeding values determined by progeny testing (Corley and Tinker 2003). Each scheme indeed has its pros and cons and many plantations in effect practice a combination of the pure approaches, partly due to the female sterility present in many *pisifera* origins. Unlike the African-based populations, the Deli palms derived from four dura palms planted at Bogor Botanical Garden in 1848 (Pamin 1998) exhibit better trait uniformity, thicker mesocarp, high bunch number, and high mesocarp oil yield. The oil palm breeding programs in Malaysia thus emphasized the genetic improvement of the Deli *dura* population through pedigree selection to produce high-yielding commercial *dura* materials until the 1960s. At that time, FIPS was the best method to improve yield and bunch component traits by selecting dura palms based on their phenotypic performance with no requirement in dura parents to evaluate their combining abilities. The *dura* × *pisifera* planting material was only accepted in Southeast Asia after its success in Africa. The commercial *dura* × *pisifera* breeding programs kept the Deli *dura* populations and AVROS *tenera*/*pisifera* populations separate. Owing to the direct evaluation of breeding values through progeny testing, RRS permits the exploration of specific combining ability (SCA) as well as general combining ability (GCA). However, the scheme requires a large progeny-testing population (Rosenquist 1990) derived from a small number of parents. Large-scale combinatory crossing for the evaluation of individual *dura* and *pisifera* combinations is not feasible due to resource limitation. This may then result in the best combination for GCA and SCA not being detected, which can hinder genetic improvement (Hardon 1970). RRS requires two generations of more than 20 years to complete a selection cycle. One generation is to propagate the Deli parents and the AVROS parents, respectively, followed by another generation of progeny testing between the parents. The breeders may lose some of the elite Deli and AVROS parents due to aging, disease infection, etc. after the completion of a selection cycle. This has become a more significant issue recently with the growing problem of Ganoderma infection and basal stem rot losses within palm stands. As a solution, selected palm sibs are often used to replace lost palms for the next cycle based on GCA, which unfortunately dilutes the power of RRS. Most breeding programs in Malaysia have shifted back to FIPS, focusing on GCA to reduce the dependency on progeny testing. However, the selection response for this method is usually lower particularly for those phenotypes with poor heritability, such as fresh fruit bunch (FFB). FIPS requires more selection cycles, even though the interval per cycle is shorter in comparison to RRS. In recent decades, DNA markers have been employed to identify QTLs for traits of interests in oil palm. By accessing the level of genetic polymorphism in the genome, palms that possess particular combinations of desired QTLs could be selected at the nursery stage. By using the trait-linked markers, breeders can reduce conventional phenotyping cycles and also enrich for the best combinations of alleles from each cross. So far, genetic marker development in the oil palm industry is still in its discovery phase. This chapter discusses both genetic discoveries and more importantly their applications to oil palm breeding programs and seed production.

9.4.2 Boom in DNA Sequencing

9.4.2.1 Whole-Genome Sequence

Whole-genome sequencing is a laboratory process that can reveal the complete DNA repertoire of an organism's genome, distributed in the nucleus, mitochondria, and chloroplasts. Sequencing of individual genomes provides a better understanding of genetic variation both within and between species. The genetic polymorphism in populations of an organism becomes extremely important when the inherited differences in genic or other portions of the genome are found to be directly responsible for phenotypic variants, influencing

an individual's risk of disease and or affecting their fitness in the environment. Hence, whole-genome sequencing is the essential step to localize DNA variants that may contribute to both major and minor genetic effects of the targeted phenotypes, followed by a systematic and complete functional characterization of the genome. The three billion DNA base pairs (bp) of the human genome were sequenced and completed in 2003 under the Human Genome Project (HGP), an international public project led by the United States. With the published genomes, the HGP has facilitated the discovery of more than 1800 disease genes. The scientific communities are seeking to duplicate the same achievements in agricultural crops. As for commercial crop genomes, rice (Oryza sativa L.) was first sequenced and published by The International Rice Genome Sequencing Project (IRGSP) in 2000. A series of genome improvements were subsequently achieved, by correcting the reassembly, with a reduction in redundancy and an increase in genome completeness. The current release of the rice genome 7.0 was published in the Rice Annotation Project Database (http://rice.plantbiology.msu.edu/annotation_pseudo_current.shtml) in 2012. A total of 373.2 Mbp of non-overlapping rice genome sequence and 55,986 genes (loci) from the 12 rice chromosomes were reported. This represents only a slight increment in genome size and number of identified genes compared to release 4.0 (372.1 Mbp and 42,653 genes) (Ouyang et al. 2007), suggesting that saturating coverage of the genome has been achieved. The oil palm, as the most traded and dominant source of vegetable oil production in the world, was only sequenced and became publicly available in 2013. As for the 1.8-Gbp oil palm genome, 1.535 Gbp of assembled sequence and transcriptome data from 30 tissue types were used to predict at least 34,802 genes (Singh et al. 2013b). About 53% of the genomic scaffolds are represented in a physical map with 16 linkage groups, which are equivalent to the chromosome pairs (diploid number, 2n = 32) in oil palm. The genome appears to be of medium size among monocotyledons. It is larger than banana (0.60 Gbp) (Sagi et al. 2005) and soybean (1.15 Gbp) (Swaminathan et al. 2007), but smaller than barley (5.00 Gbp) (The International Barley Sequencing Consortium, 2012) and maize (2.30 Gbp) (Messing et al. 2004). The published genome sequence has become an important reference for genetic discovery programs that are critical to expedite breeding progress in oil palm.

9.4.2.2 Resequencing of Various Palm Populations

With the existence of the reference genome, whole-genome resequencing is the most comprehensive and cost-effective method to discover or capture polymorphism in the genome (Feuillet et al. 2011). The polymorphism can be categorized as either large or small variations caused by different mutation mechanisms. The popular SNPs and SSRs are widely utilized as molecular markers due to their high abundance, polymorphism, reproducibility, and codominance. Resequencing of pooled individuals provides a quick genome-wide polymorphism survey at very moderate cost and is also useful for the estimation of allelic frequencies, which is essential for population genetic studies. However, this approach does not permit LD analysis, compared to individual resequencing. At the Sime Darby Plantation R&D Centre, 132 individual oil palms (*E. guineensis* and *E. oleifera*) belonging to 59 diverse populations were pooled for resequencing, followed by further individual sequencing of 55 representative palms (one palm per population). The pooled sequencing data yielded more than seven million raw SNPs. About one million of the *E. oleifera*-specific variants were then separated out, based on the individual sequence data. After considering good sequencing depth from the pooled sequencing data, about three million high-quality SNPs were identified for further analysis and validation. The genome and SNP sequence data were then available to facilitate an understanding of the genetic basis of underlying QTLs of important agronomic traits in oil palm

9.4.3 Development of DNA Markers

DNA markers, such as RFLP, AFLP, and RAPD, were initially developed to evaluate the genetic diversity of

oil palm germplasm (Shah et al. 1994; Barcelos et al. 2002) and to construct linkage maps (Mayes et al. 1997). Attention then shifted rapidly to multiallelic SSR markers, which were derived from genomic libraries and EST databases (Billotte et al. 2001; Ting et al. 2010), but this method only provides medium mapping resolution and throughput. The publication of the oil palm genome, together with the reduced cost of high-throughput resequencing provided a better opportunity for high-density SNP and SSR detection in oil palm. In the last decade, high-density genome-wide SNP arrays have been developed for many crops and livestock species, including rice, corn, cattle, horse, and salmon. The first SNP genotyping array for oil palm was a 4.5K custom Illumina SNP array, which has been proven to be useful in the construction and comparison of linkage maps (Ting et al. 2014). Nevertheless, the density of genome coverage of the array for association analysis may be inadequate, especially in highly heterogeneous populations. Hence, a higher-density Illumina OP200K SNP array consisting of 200,000 SNPs was designed from the three million loci detected in the resequencing data and successfully validated in various populations using different genetic analyses (Kwong et al. 2016; Teh et al. 2016). According to the minimal LD decay of 19.5 Kbp observed among the oil palm populations examined, the OP200K SNP array with one SNP every 11 Kbp, on average provided excellent mapping resolution for linkage analysis and genome-wide association study (GWAS).

9.4.4 Development of DNA Markers

9.4.4.1 Linkage Mapping

Methods of genetic mapping were developed to localize genes responsible for traits on the basis of a correlation between phenotype and DNA variation in an unbiased fashion. The first genetic map of fruit flies was successfully constructed through linkage analysis in 1913 (Sturtevant 1913). The method involves characterizing the meiotic recombination events in a family (mapping population) through analysis of the combination of markers (phenotypic or genetic) transmitted from the parents. DNA markers that show strong correlation to the trait variation (showing "linkage") are expected to reside close to the genes controlling variation in the trait. The QTL localization for heritable traits is still mainly dependent on family-based mapping methods. A linkage map with 576 SSR markers and an additional 102 diversity arrays technology (DArT) markers was constructed at Sime Darby for a commercial cross of Deli *dura* × AVROS *pisifera* (branded as Calix 600) using the regression model implemented in JoinMap4 (van Ooijen 2006) at a LOD = 5.0 threshold, using Kosambi's distances. The map consisted of 23 linkage groups, spanning 1308 cM with a 2.1 cM average mapping interval (unpublished), which was one of the densest maps for oil palm at the time. However, the mapping resolution, particularly for complex traits (e.g., oil yield and disease tolerance), was still limited by insufficient recombination within these small populations (Billotte et al. 2010; Montoya et al. 2013; Jeennor and Volkaert 2014). More than 1000 seeds are usually obtained from a single cross of oil palm, but only 16–96 palms are planted and trait-recorded in field trials. The same problem with limited family sizes was also a limitation of linkage analysis in humans, especially for non-Mendelian diseases in the twentieth century (Altshuler et al. 2008).

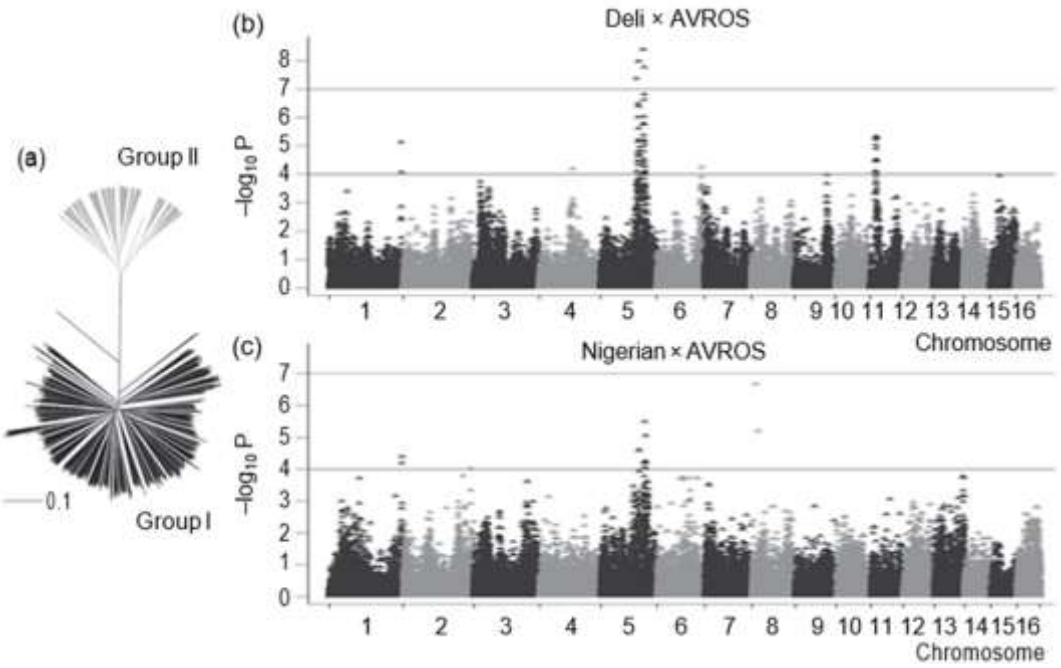


FIGURE 9.8 Genetic stratification and genome-wide association study for mesocarp oil content (%) of 2045 individuals representing Deli × AVROS and Nigerian × AVROS. (a) Neighbor-joining tree (NJ) constructed based on genetic distance. Group I and group II are Deli × AVROS and Nigerian × AVROS. (b) Manhattan plot of compressed MLM model for Deli × AVROS. (c) Manhattan plot of compressed MLM model for Nigerian × AVROS. (Based on Teh et al. 2016.)

9.4.4.2 GWAS and GS

The reference oil palm genomes (Singh et al. 2013b) and high-density OP200K genotyping array (Kwong et al. 2016) has enabled the first GWAS of mesocarp oil content (%) in oil palm (Teh et al. 2016). Unlike linkage analysis, GWAS provides access to the total meiotic recombination accumulated in a large population during evolution to increase the mapping resolution. The GWAS was carried out on 2045 *tenera* palms using the OP200K array. In order to maximize the mapping resolution, the discovery population was selected based on short-range LD distance, which is inherited with long breeding cycles and heterogeneous breeding populations. Figure 9.8 shows a genetic stratification and association study for mesocarp oil content. Deli × AVROS and Nigerian × AVROS palms were grouped through genetic clustering analysis (Figure 9.8a) and subjected to phenotype–genotype association analysis independently using a simple linear model, but genomic inflation was still unsatisfactorily high. The false positives were then significantly reduced in Deli × AVROS (GIF = 1.1) and Nigerian × AVROS (GIF = 1.9) populations using a compressed mixed linear model with population parameters previously determined (P3D) (Zhang et al. 2010). A total of 62 and 18 significant association signals for mesocarp oil content (%) were detected in Deli × AVROS (Figure 9.8b) and Nigerian × AVROS (Figure 9.8c), respectively, according to whole-genome significance thresholds ($\log_{10}$ p-value ≥4.0; Bonferroni's $\log_{10}$ p-value ≥7.0) (Teh et al. 2016).

The major association peaks observed in Chromosome 5 had not been reported in any previous QTL linkage studies. The GWAS results also revealed potential genes responsible for variation in mesocarp oil content (%) of the oil palm populations. GWAS effectively identifies only the major QTLs that exceed the genome-wide significance threshold. The predictability of breeding value remains poor if these QTLs only contribute small effects to the complex traits. Hence, GS was then investigated to evaluate the total genetic

effects of the entire genome relative to a trait, whereas all QTLs are expected to be in LD with at least one marker (Goddard and Hayes 2007). In order to achieve this, the GS accuracy in an oil palm population using a low density of SSR markers was first reported (Cros et al. 2015b) and the work was continued with reciprocal recurrent genomic selection (RRGS) based on simulated SNP markers (Cros et al. 2015a).

With the generated genotypic data, a further investigation of GS accuracy was subsequently carried out for the quantitative residual shell thickness variation (percent shell-to-fruit within *tenera* shell-types, S/F%) trait in 312 *tenera* palms using the high-density OP200K SNP array (Kwong et al. 2016). The prediction accuracy was surprisingly high (correlation between genomic estimated breeding value [GEBV] and observed S/F% trait = 0.74), even with a small training set. Figure 9.9 shows a strong correlation between predicted GEBV and observed S/F%. This is the value that tells a breeder how a palm will be expected to perform in the field according to the genomic profile. GWAS and GS clearly provide the foundation for marker-assisted selection (MAS); however, the value of this can only be realized if they can be effectively deployed in oil palm breeding programs.

9.4.5 Marker Application to Oil Palm Breeding

9.4.5.1 Marker-assisted Selection

MAS has been extensively implemented in animal and plant breeding, but the increases in genetic gain have been small when using only a few major QTLs (Dekkers 2004). GS indeed is an expanded form of MAS by covering the whole genome so that breeding values of complicated traits can be predicted more accurately (Goddard and Hayes 2007). As a cost-optimized solution, both methods could be combined to optimize marker density for routine genotyping. The identified QTLs and GS model can break the cycle time and land use bottlenecks of RRS and FIPS by early selection at the nursery stage leading to a reduction in the number of progeny testing in the field. In Sime Darby Plantation R&D Centre, a large F_2 progeny test population (~1000 palms/trial) derived from multiple Deli *dura* × AVROS *pisifera* crosses was used as a training set for GWAS and GS modeling of oil yield traits. Figure 9.10 shows the scheme for selection of the best palms from current commercial material. All germinated Deli *dura* and AVROS *pisifera* seedlings in the F_3 were screened and selected based on the genetic profiles of the training set; thus allele dropout should be minimal. Only those *dura* and *pisifera* seedlings that potentially contribute the best enrichment of beneficial alleles to their future progeny were field planted. More importantly, the breeders could now assess SCA between dura and *pisifera* parents in the F_3 to produce high-yielding tenera planting materials, without the need of another round of progeny testing. The new MAS model effectively maximizes the genetic gains of oil palm in half the selection cycle compared to conventional selection methods.

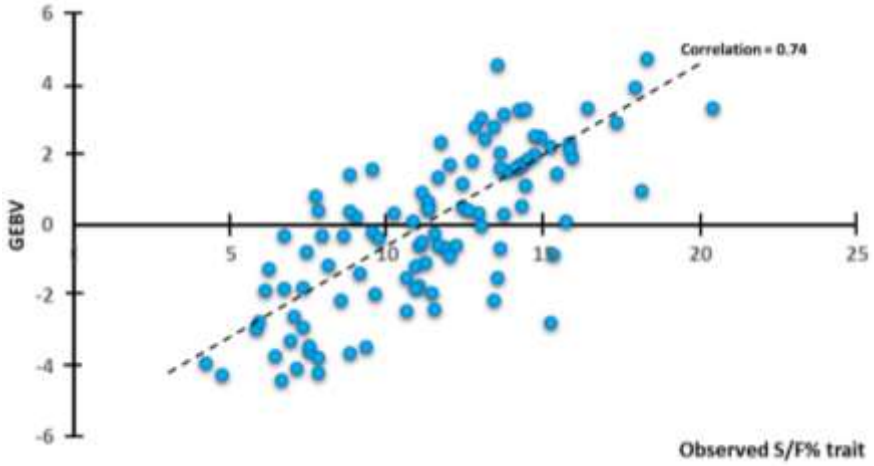


FIGURE 9.9 Representative regression plot of genomic estimated breeding value (GEBV) and observed shell thickness trait variation within *tenera* for the validation set. The correlation coefficient between GEBV and the observed trait was 0.74 with 55% of the variation explained for this representative case. (Based on Kwong et al. 2016.)

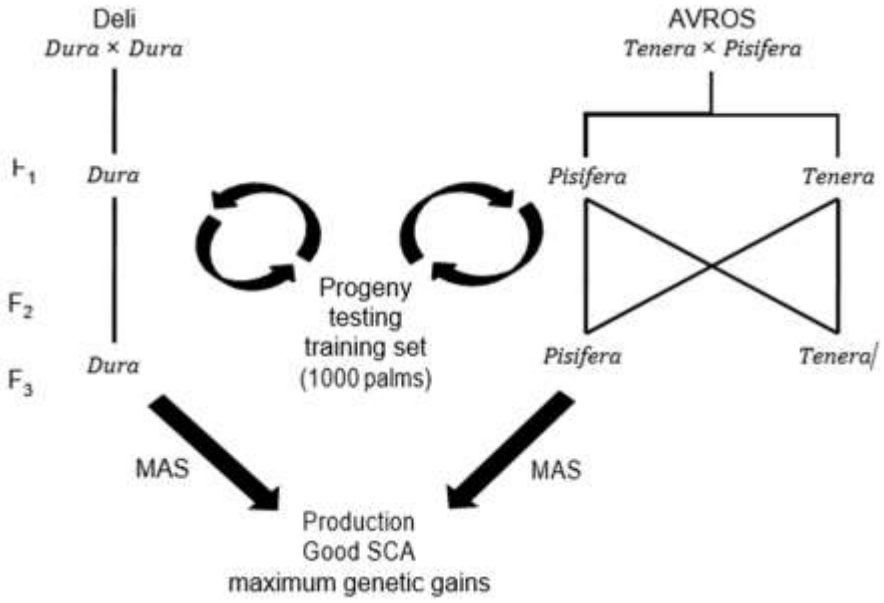


FIGURE 9.10 Marker-assisted selection (MAS) in oil palm selection schemes. SCA—specific combining ability.

9.4.5.2 "Genome Select"

The retrospective validation of the MAS program in oil palm will still require two phenotyping cycles. However, by using the same genetic findings, short-term selection response is measurable within a generation

of *tenera*. A modified GS model for oil yield traits was deployed to "Genome Select" the highest yielders out of Sime Darby's commercial seeds. If the selection response is high, the selected palms should have a yield potential significantly better than the unselected base population, provided the base population possesses sufficient phenotypic and genetic variation. The idea indeed originated from "The Breeder's Equation," which is

$$R = h^2S$$

S is known as selection differential, indicating the difference between population mean of all potential parents before selection and the population mean of the actual selected parents that will be used to generate the next generation. The h^2 is heritability for a trait and R is the selection response, which is the change in means between the population before selection and the population in the next generation. In this relationship, the h^2 of a trait is the link between the within-generation change S and the between-generation change R. However, the commercial *tenera* palms are basically the end products of breeding and hybridization, so h^2 may be less relevant to the Genome Select program. A new selection response R_a only measures the mean differences within the same generation. In this scenario, the main component by which R_a differs from S is the slope of GS regression line between the GEBV and observed traits, instead of h^2, reflecting the accuracy of selection using GEBVs. According to the 0.65 prediction accuracy observed for total oil yield per palm using a carefully selected and optimized SNP panel, yield potential increment greater than 16% is expected by planting the selected best predicted performance palms from the current best commercial *tenera* population. Figure 9.9 shows the expected effect of genome selection on current commercial materials. Composite Figure 9.11 gives details of the implementation of the Genome Select approach. Testing and selection has been conducted at sufficient scale for commercial fields in multiple locations (Figure 9.12a). Practical tools, such as barcode labeling, database, and barcode reading system, were developed to allow sampling, genotyping, and selection on this scale (Figure 9.12b and c). The prospective validation will be feasible once the phenotyping cycle is complete, but early indications should be available within the first three years of yield recording.

9.4.6 Conclusions

The developments in molecular technology during the past 10 years have enabled significant improvements in throughput combined with a reduction in costs for genome sequencing as well as genetic marker assays. Concurrent developments in bioinformatics tools that allow detailed quantitative genetic analysis of large datasets means that the oil palm industry now has access to the tools to develop and deploy molecular markers for the selection of the best planting materials. The critical success factors for this technology to provide the much needed yield improvements in the coming years will be the harmonization of molecular breeding with the planted breeding trials and commercial seed production. This will be necessary to enable both the marker discovery/development process and optimization of future breeding programs in order to realize the potential increase in breeding efficiency and desirable trait combinations possible through MAS.

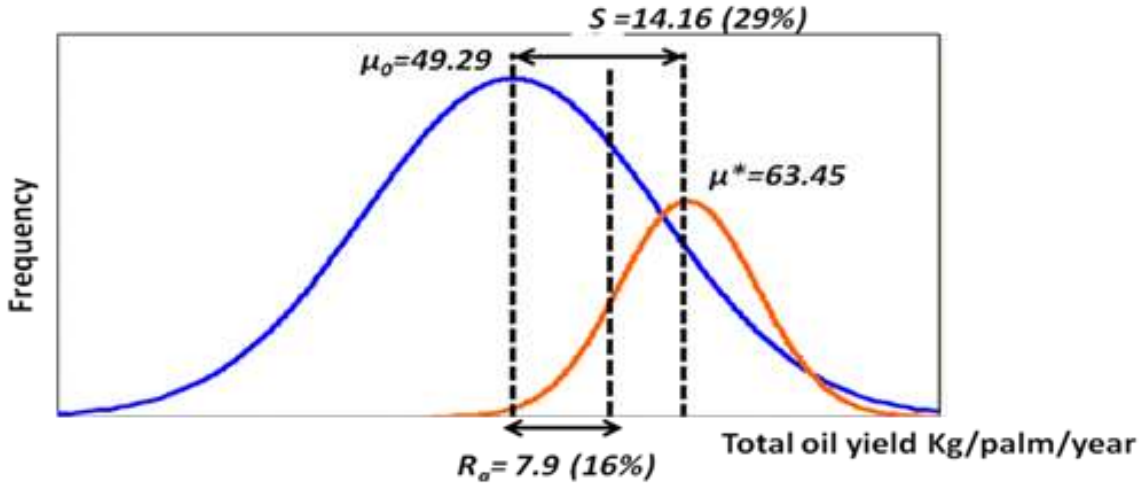


FIGURE 9.11 Example of "selection response" of the genome selected materials compared to the unselected base population. μ_0 is the yield mean of the unselected base population (blue curve). μ^* is the yield mean of the top-30% genome selected population (red curve). S is selection differential. R_a is selection response within generation using GS.

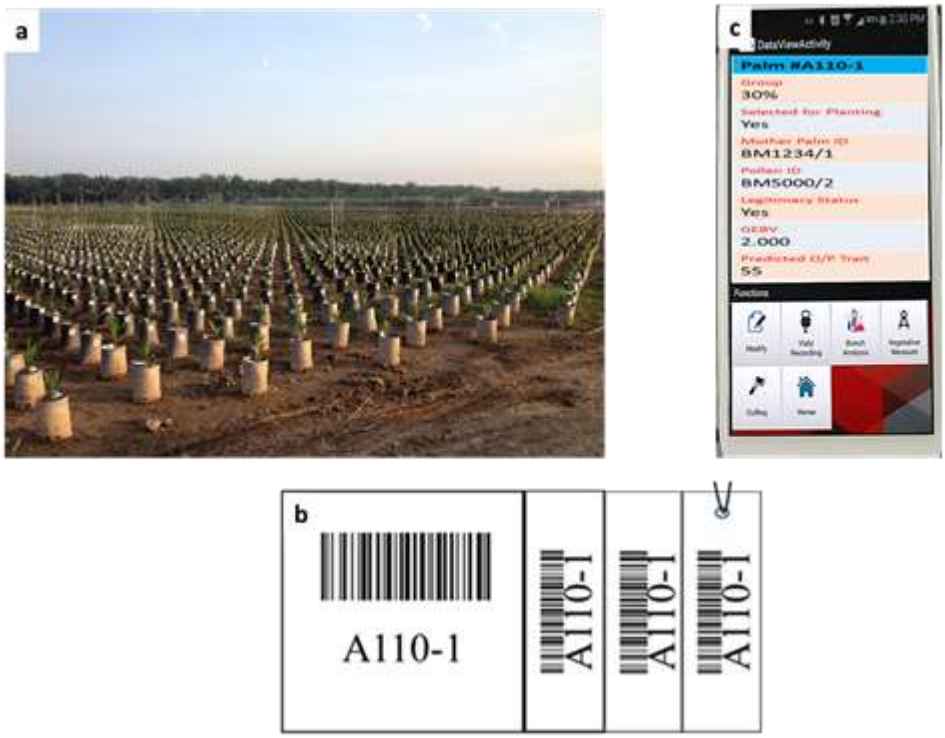


FIGURE 9.12 Deployment of Genome Select on the current best commercial *tenera* population. (a) Tested and barcoded seedlings were planted at the nursery; the genome selected individuals were transplanted to the commercial field. (b) Barcode labelling system was introduced. (c) Barcode reading system was developed as a smartphone app to retrieve palm biodata at nursery for selection purposes, including the predicted GS value of the individual palm.

Chapter 3: A PRACTICAL GENOMIC SNP-BASED LEGITIMACY TEST METHOD FOR OIL PALM BREEDING AND SEED PRODUCTION

3.1 Introduction

Oil palm is an out-pollinating, but monoecious crop. As discussed in Chapter 2, the oil palm breeding programmes maintain the maternal Deli *dura* and the paternal AVROS *pisifera*, separately. The elite parents are selected through progeny testing to produce high-yielding commercial *tenera* seeds. Hence, controlled pollination is essential to ensure the crosses reflect the true families or backgrounds. Nevertheless, complete avoidance of illegitimacy in oil palm is almost impossible, because pollination errors including late bagging, inflorescence bag damage and contaminated pollen source are still evident in oil palm breeding and seed production, with the introduction of the pollinating weevil (*Elaeidobius kamerunicus*) increasing the transfer into improperly bagged inflorescences (Corley, 2005). Occasionally, hermaphrodite flowers also cause the same problem through self-pollination (Beirnaert, 1935).

Conventionally, genetic purity of seed lots in oil palm breeding and seed production is monitored based on fruit cross-section to determine the presence of a fibre ring around the kernel. A DxP cross theoretically produces 100% *tenera* due to the codominant monogenic inheritance of fruit forms (Beirnaert and Vanderweyen, 1941). In this case, any unexpected *dura* and *pisifera* palms present in the DxP crosses are considered as non-*tenera* contamination. Nevertheless, the fruit census is only feasible when the palms start fruiting 2 – 3 years after field planting; the process thus is laborious, costly and does not easily allow errors to be corrected. The yield loss due to planting of non-*tenera* commination can be long term, up to 20 – 25 years when the replacement of contamination is not economically viable. As an estimation, every 1% *dura* contamination may cost RM178.9 million year^{-1} (USD 45.6

million year^{-1}) in Malaysia, given a 25% oil loss produced from *dura*, compared with *tenera* fruits. The problem has become alarming, when average of 10.7% non-*tenera* contamination estimated throughout Malaysia's smallholder plantings (Ooi et al., 2016).

As a solution, novel mutations in the *SHELL* gene responsible for the fruit forms in oil palm have been found (Singh et al., 2013a) and developed as DNA markers for early elimination of the non-*tenera* contamination before field planting. However, both conventional fruit census and the SHELL markers are unable to detect illegitimate *tenera* palms derived from the incorrect DxP crosses, and also do not provide information to trace the main cause of illegitimacy. Aside from the commercial yield loss, *tenera* illegitimacy (whether accidental or deliberate) is believed to have a further negative effect on future generations of selective breeding, such as producing confused interpretations in breeding trials. In this research programme, a decision was made to develop a robust genomic SNP-based legitimacy assay after observing illegitimate palms in the assayed mapping populations. The primary objective of this study is to develop a practical assay for legitimacy using an optimised SNP density, followed by implementation as a commercial quality tool.

3.2 Methodology

A total of 1,000 palms with known fruit forms were used to determine the detection accuracy of the legitimacy testing method developed in this study. In the population, two commercial Deli *dura* x AVROS *pisifera* half-sib *tenera* families (Family A and Family B) and a Deli *dura* family (Family C) were used to simulate possible contamination types in the field/production pipeline. The legitimacy assay development was started with 1,000 genomic SNP markers which were selected from the OP200K Infinium array (Illumina, San Diago, CA, USA) (Kwong et al., 2016) to form a smaller genotyping panel, namely GS1000^{TM}. In addition, pedigree records of the assayed populations were also validated through the SHELL marker test. For simulation purposes, Family B and Family C were considered as *tenera* and *dura*

contamination, respectively for Family A. The legitimacy test was then performed using an in-house developed programmes, termed *LegiPerl 1.0*. The test result generated using the GS1000^{TM} panel was benchmarked to further reduce the marker density. Detection accuracy of each test was quantified according to the illegitimate palms which could be distinguished from the reference.

3.3 Results & Discussion

The result indicates that the GS1000^{TM} panel can easily discriminate the *dura* contamination (Family C) (Figure 3). The genomic SNP-based legitimacy test is solely based on identity by descent of markers, which can still be present even with correct prediction of fruit forms due to origin-specific alleles on the *SHELL* gene. Furthermore, the legitimacy test also successfully distinguished the Family B as *tenera* contamination from Family A, while the conventional fruit census and SHELL test would wrongly declare that the mix of Family A and Family B were legitimate. Marker density was also successfully reduced to 200 SNP markers, which provides a similar detection power to the GS1000^{TM} panel. Apart from marker density, sampling schemes, such as ISO-2859-3 series are also available to define statistically valid sample size for a seed lot, instead of testing every single cross. With this, the legitimacy test developed in this study as now been practically implemented commercially and deployed in breeding programmes and seed production for Sime Darby Plantations.

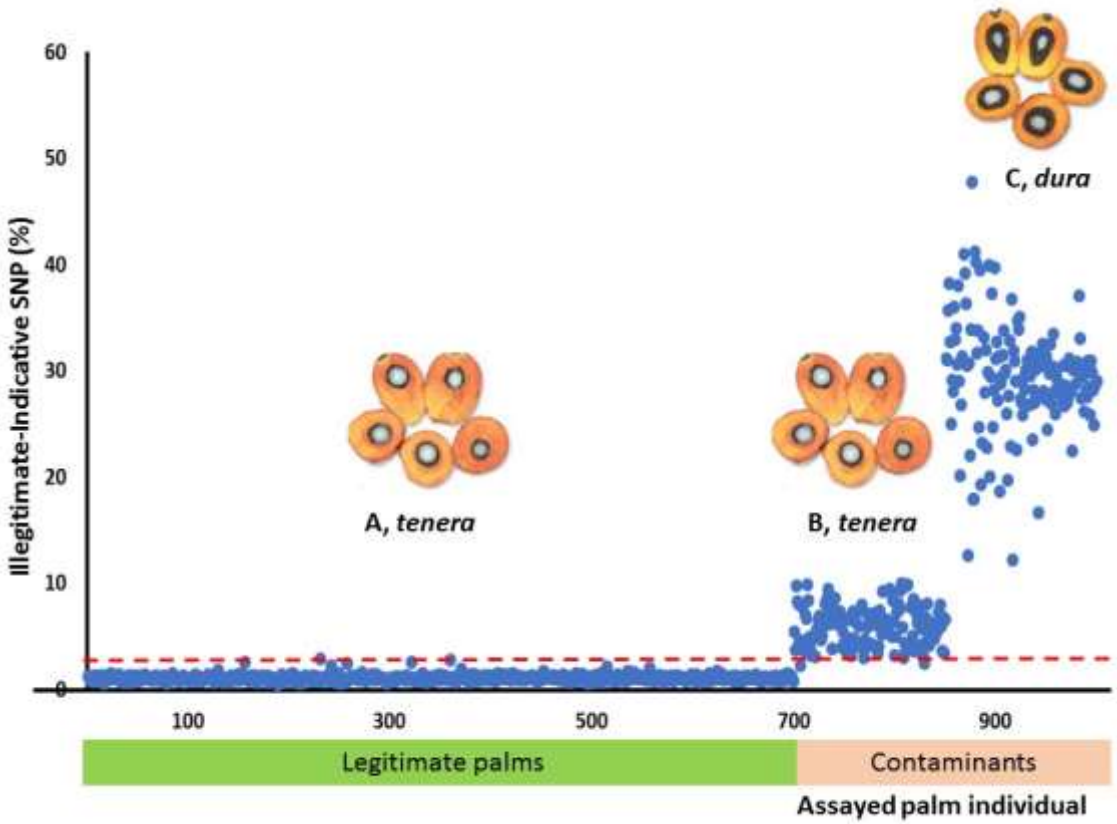


Figure 3: Legitimacy reference based on the parentage of Family A at illegitimacy-indicative SNP threshold=3% (the dotted red line), based on Teh et al. (2019 (Accepted))

3.4 Conclusion and key recommendation for future works

The genomic SNP-based legitimacy test can distinguish both *dura* and *tenera* contamination at or before the nursery stage. With the optimised 200 SNP panel coupled with a suitable sampling scheme, implementation of the legitimacy test can achieve more than 97% detection accuracy and serve as a standard quality control procedure for seed purity in breeding programmes and in commercial seed production. Without the access to pedigree records and parent palms, the SHELL test is still a powerful tool for the smallholders to eliminate the *dura* contamination prior to field planting. For perennial crops such as oil palm,

illegitimacy problem can readily result in major impacts to a 12-year breeding effort and for commercial palms, a 25-year reduction in oil yield potential due to contaminants. Therefore, genetic purity and good agricultural practises should be treated as being equally important to ensure the highest oil productivity of the oil palm industry into the future.

3.5 Published journal article contributing to the objectives in Chapter 3

Teh C K, Lee H L, Hafiza A, Ong A L, Mayes S, Chew F T, Appleton D R (2019): A practical genome-enabled legitimacy assay for oil palm breeding and seed production. BMC Plant Biology 19, 470.doi.org/10.1186/s12870-019-2062-x

Teh et al. BMC Plant Biology (2019) 19:470
https://doi.org/10.1186/s12870-019-2062-x

BMC Plant Biology

RESEARCH ARTICLE Open Access

A practical genome-enabled legitimacy assay for oil palm breeding and seed production

Chee-Keng Teh[1,2*], Heng-Leng Lee[1], Hafiza Abidin[1], Ai-Ling Ong[1,2], Sean Mayes[3], Fook-Tim Chew[4] and David Appleton[1]

Abstract

Background: Legitimacy in breeding and commercial crop production depends on optimised protocols to ensure purity of crosses and correct field planting of material. In oil palm, the presence of three fruit forms permits these assumptions to be tested, although only after field planting. The presence of incorrect fruit forms in a cross is a clear sign of illegitimacy. Given that *tenera* forms produce 30% more oil for the same weight of fruit as *dura*, the presence of low levels of *dura* contamination can have major effect during the economic lifespan of an oil palm, which is around 25 years. We evaluated two methods for legitimacy test 1) The use of SHELL markers to the gene that determines the shell-thickness trait 2) The use of SNP markers, to determine the legitimacy of the cross.

Results: Our results indicate that the SHELL markers can theoretically reduce the major losses due to *dura* contamination of *tenera* planting material. However, these markers cannot distinguish illegitimate *tenera*, which reduces the value of having bred elite *tenera* for commercial planting and in the breeding programme, where fruit form is of limited utility, and incorrect identity could lead to significant problems. We propose an optimised approach using SNPs for routine quality control.

Conclusions: Both *dura* and *tenera* contamination can be identified and removed at or before the nursery stage. An optimised legitimacy assay using SNP markers coupled with a suitable sampling scheme is now ready to be deployed as a standard control for seed production and breeding in oil palm. The same approach will also be an effective solution for other perennial crops, such as coconut and date palm.

Keywords: Contamination, *SHELL*, DNA fingerprinting, Genetic purity, Seed quality control

Background

Oil palm (*Elaeis guineensis* Jacq.) is an out-pollinating, but monoecious crop; however, maternal and paternal lineages for the purpose of seed production are developed independently based on presence/absence of kernel shell in the fruit. Thick-shelled *dura* palms produce lower oil yield than *tenera* heterozygotes, whereas shell-less *pisifera* palms are usually female sterile. The crosses between *dura* and *pisifera* result in thin-shelled *tenera* progenies that exhibit 30% higher oil yield than *dura* parents with full fertility restored. Hence, *tenera* palms are preferred for commercial planting.

Like other plant and animal breeding programmes, understanding the parentage of families and individuals is crucial. In oil palm, the goal is to measure combining ability between *dura* and *pisifera* based on performance of their *tenera* progenies, also known as progeny testing. Controlled pollination is required to ensure the test reflects the true families or backgrounds. However, pollination errors, such as a contaminated pollen source, inflorescence bag damage and late bagging are still evident in oil palm breeding and seed production, leading to illegitimacy [1]. Aside from instances of uncontrolled pollination, hermaphrodite flowers occasionally form whereby self-pollination can occur [2], although the bunches with high number of male

* Correspondence: teh.chee.keng@simedarbyplantation.com
[1]Biotechnology & Breeding Department, Sime Darby Plantation R&D Centre, Banting, Selangor, Malaysia
[2]School of Biosciences, University of Nottingham Malaysia, Semenyih, Selangor, Malaysia
Full list of author information is available at the end of the article

© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

inflorescences can usually be identified and removed. Hence, seed illegitimacy in oil palm can happen at any stage from pollen collection, pollination and seed processing through to field planting. Quality assurance and problem identification are usually difficult, especially when combinatorial errors occur.

To date, the oil palm industry still relies on the shell fruit form and the presence of a fiber ring around the kernel to determine genetic purity of seed lots in breeding and commercial seed production. Seeds derived from *dura* (*Sh* + *Sh*+) x *pisifera* (*sh*- *sh*-) are theoretically 100% *tenera* (*Sh* + *sh*-) due to co-dominant monogenic inheritance [3]. Unexpected *dura* and *pisifera* progeny palms are defined as non-*tenera* contamination. However, the fruit census can only be carried out 2–3 years after field planting. Census methods based on a cross-section of mature fruit is laborious and inefficient. Most importantly, replacement of contaminants at this mature stage is not economically viable, leading to long-term yield loss from these contaminant palms throughout the 20–25 years of production lifespan. To solve this issue, novel mutations in the *SHELL* gene that are responsible for the fruit form have been identified [4] and can be used for detecting non-*tenera* contamination as early as the seed stage. Early elimination of the contaminants before field planting would significantly improve cost effectiveness and reduce yield losses due to poor seed production quality, hence improving oil palm productivity per unit area of land [5, 6]. However, conventional fruit census and SHELL testing both share two key limitations, whereby the methods are unable to identify illegitimate *tenera* that are derived from the wrong crossing or parentage and do not provide informative data as to the production causes of the illegitimacy. The wrong parentage illegitimacy is believed to have further impact on oil yield due to the different genetic combining ability of every resulting progeny seed produced from a pollinated flower. This is especially impactful in breeding programmes where illegitimate parental stock can effect multiple future generations and confuse interpretations from breeding trials.

Established seed producers and plantation companies have dedicated breeding programmes to sustain their planting requirements. Accurate pedigree records enable these producers to utilize legitimacy tests on every cross or seed lot before releasing the seeds to stakeholders, which often include smallholder plantations and farmers. The objective of this study was to develop an improved practical assay for legitimacy test using an optimised number of genomic single nucleotide polymorphism (SNP) markers, followed by implementation as a commercial quality control tool. A commercial population of *dura*, half- and full-sib mother palms was blind tested to determine the accuracies of legitimacy test using 1000 SNPs and marker subsets optimised.

Results

Genetic clustering and fruit form validation

The assayed SNPs had < 5% missing data in the GS1000™ panel for 988 markers. The markers successfully assigned the 1000 palms to three clusters i.e. Family A (700 palms; black), Family B (150 palms; red) and Family C (150 palms; green) in a principle component plot (Fig. 1). Family A and Family B were more genetically related due to their half-sib lineage and were therefore overlapping. A broader distribution of Family C was due to mother palms derived from different lineages. In addition, the fruit forms of all three families were also validated by the SHELL test. Family A and Family B with sh^{Deli}/sh^{AVR06} genotype were *tenera*, whereas Family C with sh^{Deli}/sh^{Deli} were *dura*. Overall, the result indicated a good agreement with known pedigree records and observed fruit form.

Legitimacy reference of Family A

Family A exhibited 0.5–2.9% illegitimacy-indicative SNPs and therefore can be defined as a legitimate population. The two illegitimacy sources were successfully resolved using the GS1000™ panel (Fig. 2). All *dura* contaminants from Family C were clearly distinguished with 12.3–47.7% illegitimacy-indicative SNPs. The second contaminant source, Family B was 98.7% half-sib *tenera* with 3.0–10.0% illegitimacy-indicative SNPs. This information was used as a legitimacy reference standard for subsequent sensitivity analysis.

Sensitivity of legitimacy test across marker densities

In all family combinations, the accuracy improved significantly when more markers were used. Family (A + C) and Family (A+ B + C) had the highest accuracy mean, recorded 0.997 (± 0.001) and 0.993 (± 0.002) at 900 SNPs, respectively (Fig. 3). However, the accuracy plateaued at different marker subset sizes depending on the family combinations. For Family (A + C), the plateau point was at 80 SNPs with 0.959 accuracy mean (± 0.096). The plateau point of Family (A + B + C) was at 200 SNPs with 0.977 accuracy mean (± 0.027).

Discussion

For perennial species such as oil palm, selective breeding of elite *dura* and *pisifera* individuals is based on phenotypic performance of their progenies (reflected in combining ability). Legitimacy of a commercial progeny is declared if non-*tenera* contamination through fruit census is lower than 5%. Nevertheless, excessive *dura* and *tenera* can be commonly found in breeding crossing programmes used to expand and improve parental lines, such as *dura* x *tenera* and *tenera* x *tenera*, causing significant deviation

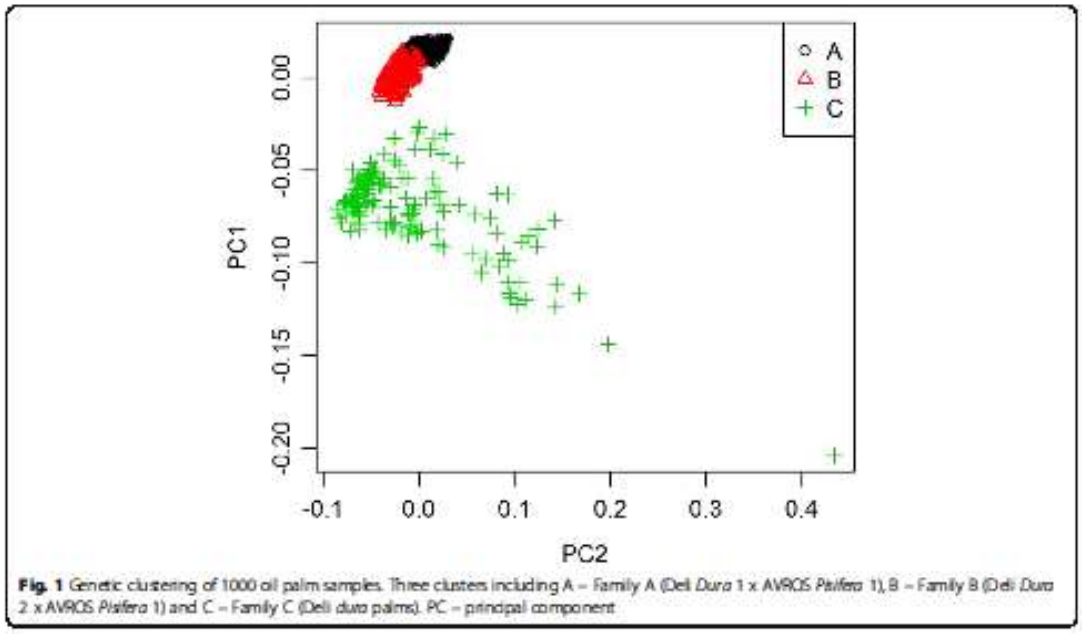


Fig. 1 Genetic clustering of 1000 oil palm samples. Three clusters including A – Family A (Deli *Dura* 1 x AVROS *Pisifera* 1), B – Family B (Deli *Dura* 2 x AVROS *Pisifera* 1) and C – Family C (Deli *dura* palms). PC – principal component

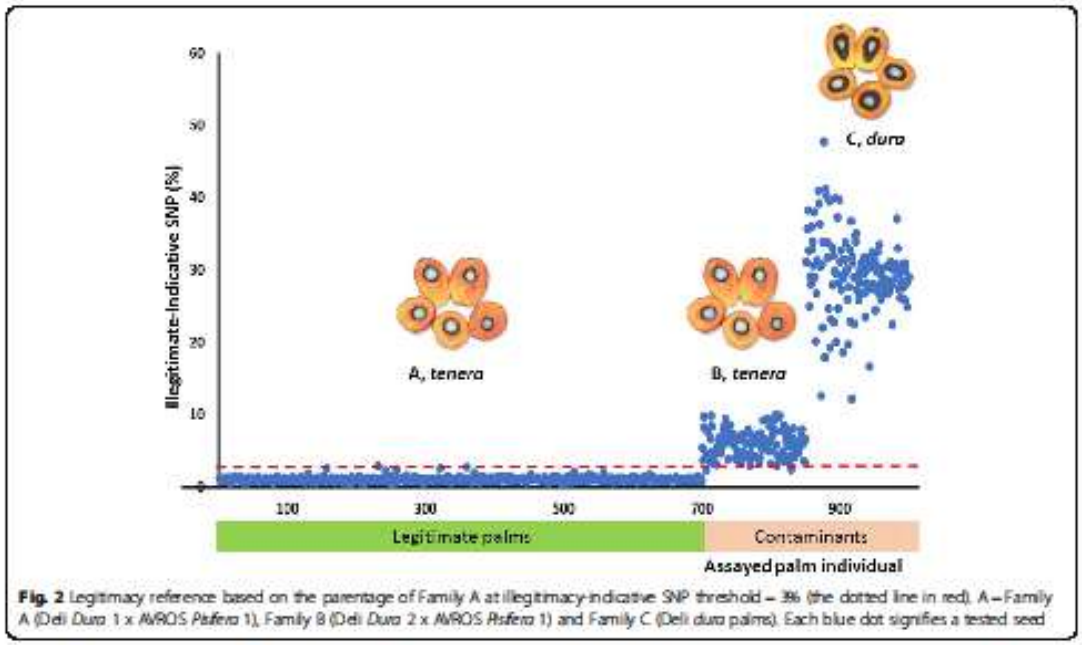


Fig. 2 Legitimacy reference based on the parentage of Family A at illegitimacy-indicative SNP threshold – 3% (the dotted line in red). A – Family A (Deli *Dura* 1 x AVROS *Pisifera* 1), Family B (Deli *Dura* 2 x AVROS *Pisifera* 1) and Family C (Deli *dura* palms). Each blue dot signifies a tested seed

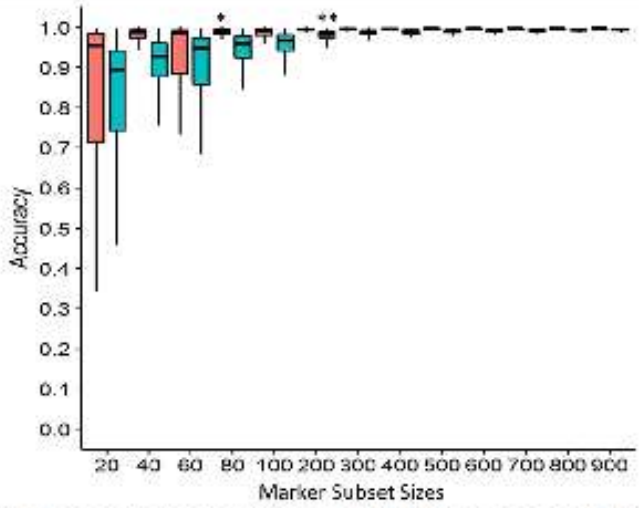


Fig. 3 Sensitivity of legitimacy test across marker subset sizes. Red box – Family (A + C); Blue box – Family (A + B + C). * The plateau point is at 80 SNPs with 0.959 accuracy mean (± 0.096) for Family (A + C). ** The plateau point is at 200 SNPs with 0.977 accuracy mean (± 0.027) for Family (A + B + C). Family (A + C) indicates occurrence of *dura* contamination only and Family (A + B + C) indicates occurrence of both *dura* and *tenera* contamination

from theoretical Mendelian segregation ratios (Chi-square test) [7]. This suggests that 'hidden' illegitimacy can exist at unknown levels, where shell-type is expected to segregate or even where shell-type is consistent with legitimacy, as in commercial *tenera* material.

Estimating 1% *dura* contamination may incur a cost of RM 178.8 million/year (USD 45.6 million/year) based on 28.6 million MT annual crude palm oil (CPO) production in Malaysia [8], given a 25% oil loss yield from *dura* fruits compared with *tenera* fruits and a CPO

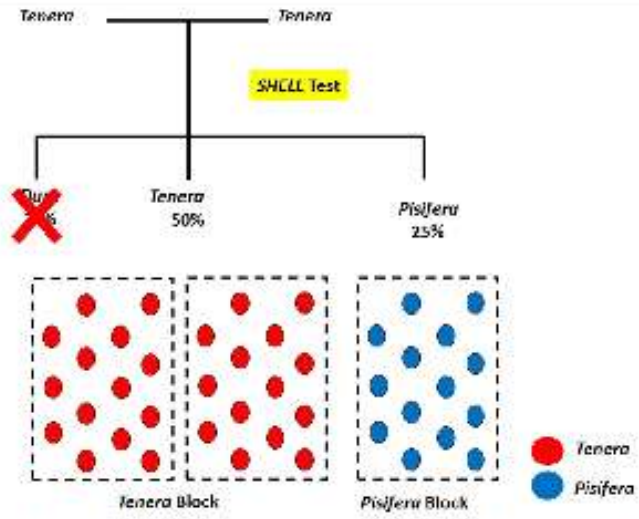


Fig. 4 Application of SHELL markers in a *tenera* x *tenera* cross for *pisifera* selection programmes. The SHELL test enables differentiation of three fruit forms i.e. *dura*, *tenera* and *pisifera* at nursery stage. *Tenera* and *pisifera* with the desired number of palms can be planted in different blocks and 25% of trial land area can be reduced by removing the unwanted *dura*

price of RM 2500/MT (USD 638/MT). Hence, the importance of detection of fruit form before planting using SHELL markers becomes more evident when considering the reported average of 10.7% non-*tenera* contamination in Malaysia's smallholder plantings [6]. The SHELL test also immediately benefits the *pisifera* selection programme. Due to common female sterility, new *pisifera* parents are derived using sib-mating i.e. *tenera* x *tenera* or *tenera* x *pisifera*. In a *tenera* x *tenera* crosses, the fruit form segregation of the progeny usually follows a Mendelian ratio of 1 *dura*: 2 *tenera*: 1 *pisifera* (Fig. 4). The oil palm breeders can remove unwanted *dura* seedlings through the SHELL test to reduce 25% of total land used for a trial, leaving the desired number of *tenera* and *pisifera* to be planted in discrete blocks for easier trial management and data recording in the future. As *pisifera* are usually female sterile, they tend to allocate more resources to vegetative growth and often compete strongly with *dura* and *tenera* palms in the same trial, leading to a potential reduction in performance of the fruit forms with shells. The breeders can also conduct yield evaluation and pollen collection in the separated blocks, instead of spending time tracking the palms in the conventional random planting.

In this study, a total of 1000 palms were successfully assayed using the GS1000™ SNP panel and assigned to three clusters according to their pedigree i.e. Family A, Family B and Family C (Fig. 1). The results further validated the utility of the OP200K genotyping array used in previous studies [9, 10]. Family A and Family B which overlapped, were indeed half-sibs with a common paternal parent, AVROS *Pisifera* 1. This explains the higher genetic relatedness between these two families. Family C originated from multiple lineages of *dura* mother palms and could be clearly distinguished in the principle component plot, indicating the distinct genetic base. A legitimacy reference based on the parentage of Family A was then established (Fig. 2). At a 3% illegitimacy-indicative SNP threshold, for *dura* contamination (Family C) was distinguished through legitimacy test. Although the sh^{AVROS} mutation alone can explain 100% of the fruit form variation in the AVROS-based *tenera* palms in this study, this may not be the case in other origins due to additional reported novel mutations in the *SHELL* gene [4, 6]. Further characterisation on the haploinsufficiency of the *SHELL* gene also unveiled possible novel mutations and the potential for cis-compound mutations yet to be characterised. This may confound shell-type prediction, particularly in introgression hybrids and more exotic germplasm [11]. Legitimacy tests based on identity by descent can effectively address this issue. The legitimacy test developed in this study was also able to distinguish the *tenera* fruit of Family A and Family B, while SHELL markers did not distinguish them and identified them correctly as both 100% *tenera*. With 0% *dura* occurrence, the current Standards and Industrial Research Institute of Malaysia (SIRIM) certification for commercial seed production would declare the mix of Family A and B as fully legitimate, but part of the sample set was actually from incorrect parentage (Family B). The yield loss due to *tenera* contamination can now be prevented or at least accounted for with the help of this practical legitimacy assay protocol. The same scenario occurs in other crops whereby illegitimate individuals cannot always be distinguished by phenotype. For example, in coconut (*Cocos nucifera* L.) breeding, the petiole colour is often used as phenotypic marker to determine genetic purity of Yellow Dwarf x Tall hybrids, but no colour variation is observed within Tall cultivars [12].

The DNA fingerprinting technique was developed in the 1980's after discovering restriction fragment length polymorphism (RFLP) markers in human DNA [13]. More marker systems were then discovered and applied in mammals [14], birds [15] and eventually plants. To increase genotyping throughput and polymorphism, oil palm researchers shifted their preference to microsatellite markers and the results were promising [16–18]. The technique has been strongly recommended as standard practice in all crosses and for tissue culture clones for the past decade. However, full adoption of legitimacy fingerprinting for commercial seed production has not been previously realized, mainly due to high assay cost and unclear sampling method. To address this, a practical method using SNP markers was developed for commercial-scale legitimacy test and the marker set size was optimised down from 1000 to 200 loci, sufficient to identify both *dura* and *tenera* contamination with consistent accuracy based on the test Family (A + B + C) (Fig. 3). Interestingly, only 80 loci were required to distinguish *dura* contaminants with consistent accuracy based on Family (A + C). An oil palm seed producer usually produces more than a million seeds annually. Testing every seed per cross is impractical and economically unjustifiable. However, many established sampling schemes for quality control, such as ISO2859-3 series are widely adopted in the manufacturing sector. These schemes provide a comprehensive reference on effective sample size to achieve the quality acceptance level based on the available seed lot and can also be adopted effectively to provide a quality underpinning to the legitimacy assay developed here for seed production quality control. Re-evaluation of the assay, however, is still necessary when dealing with alternative populations, or if the genetic base of existing

parent populations becomes narrow due to selective breeding.

Conclusion

Legitimacy tests using SNP markers can identify both *dura* and *tenera* contamination at or before the nursery stage. An optimised marker set of 200 SNPs coupled with a suitable sampling scheme will enable implementation of legitimacy test with accuracy more than 97% as a standard quality control procedure for seed purity in breeding and commercial production. Without access to pedigree records and parent palms, the SHELL test is still a powerful tool for the smallholders to discard *dura* contaminations before field planting. For perennial crops such as oil palm, illegitimacy can easily result in failure of a 12-year breeding selection cycle, and for commercial palms, a 25-year reduction in yield potential. Hence, genetic purity and good agricultural practices are equally essential to ensure the highest oil productivity of the oil palm industry into the future.

Methods

Plant materials and DNA preparation

A total of 1000 palms comprising of Family A (700 palms from Deli *Dura* 1 x AVROS *Pisifera* 1), Family B (150 palms from Deli *Dura* 2 x AVROS *Pisifera* 1) and Family C (150 Deli *dura* palms) was selected. Family A and Family B were produced as commercial *tenera* seeds and Family C was used as mother palms for breeding and commercial seed production. Also, the three parent palms, known as Deli *Dura* 1, Deli *Dura* 2 and AVROS *Pisifera 1* were included. All palms were derived from in-house breeding materials and maintained at Sime Darby Plantation R&D Centre, Malaysia. Total genomic DNA was isolated from 0.1 g of young leaf tissue (at frond 0) using the DNAeasy Plant Mini Kit (Qiagen, Germany).

SNP genotyping

The 1003 palms were genotyped for the legitimacy test and fruit form validation. A total of 1000 SNPs were selected from the OP200K Infinium array (Illumina) [9] to form a smaller genotyping panel, namely GS1000™. The probe design for the SNPs was done based on the requirement of the Kompetitive Allele Specific PCR™ (KASP™) genotyping platform. About 0.3 ng of the genomic DNA was used as template. Two fluorophores FAM and HEX were used to distinguish the KASP™ genotyping data. In the clustering plot, the samples marked in red and blue were homozygous for alleles with HEX and FAM, respectively. The heterozygous samples appeared in green. To confirm the fruit forms of assayed palms, the same genotyping method was done using KASP™ probes for sh^{mpob} and sh^{AVROS} mutations [11] at their reported genomic positions, i.e. 3,078,161 bp and 3,078,154 bp on Chromosome 2 [4]. Only the SNPs with lower than 5% missing data were selected for subsequent analysis.

Data analysis

Genetic clustering among three families (A, B and C) was analysed in the R Package *SNPRelate* with default parameters [19]. The SHELL test and genetic relatedness among the assayed palms were used to validate their pedigree records. The legitimacy test was then performed using an in-house developed programme, namely *LegiPerl 1.0*. The programme classified samples as legitimate only if the progeny and parental genotypes followed Mendelian inheritance (Table 1). In this test, the 'legitimate' individuals accepted ≤3% of illegitimacy-indicative SNPs. The KASP™'s average genotyping error in positive DNA samples was reported to be 0.7–1.6% [20].

For the blind test, Family B and Family C were assumed to be contaminants for Family A. The legitimacy tests were conducted in different combinations of families, reflecting possible contamination types in production/field: (i) Family (A + C) (850 palms) for *dura* contamination, (ii) Family (A + B) (850 palms) for half-sib contamination, and (iii) Family (A + B + C) (1000 palms) for multiple contamination sources. A legitimacy reference was constructed from Family (A + B + C) using the GS1000™ panel. This was followed by sensitivity analysis of the legitimacy test using different SNP subsets ranging from 20 to 900 markers with stepwise increments of 20 up to 100 SNPs and then 100 markers up to 900 for each combination of (i) and (iii). Each marker subset was analysed with 1000 iterations of random marker selection from the GS1000™ panel. The accuracy of each test was then determined based on the observed number of illegitimate palms that could not be distinguish from the reference. The descriptive statistics and graph plotting for accuracy were generated in R.

Table 1 Progeny and parental genotypes that follow Mendelian inheritance

Dura Parent		*Pisifera* Parent	*Tenera* Progeny
A/A	X	*a/a*	*A/a*
A/A	X	*A/a*	*A/A, A/a*
A/a	X	*A/A*	*A/A, A/a*
A/a	X	*A/a*	*A/A, A/a, a/a*

A/A and *a/a* = homozygous genotypes; *A/a* = heterozygous genotype

Abbreviations
CPO: Crude palm oil; KASP™: Kompetitive Allele Specific PCR™; RFLP: Restriction fragment length polymorphism; SIRIM: Standards and Industrial Research Institute of Malaysia; SNP: Single nucleotide polymorphism

Acknowledgements
We would like to acknowledge the sample contribution of our breeders from Oil Palm Breeding Section. We also want to thank Molecular Breeding Laboratory for genotyping the samples and analytical supports.

Authors' contributions
CT and HL conceived and designed the experiments; HL and CT analyzed the data; CT, HA, AO, FC, DA and SM wrote the paper. All the authors have read and approved the manuscript.

Funding
This study was fully supported by Sime Darby Plantation Sdn. Bhd. for funding, Sime Darby Plantation R&D Centre for collecting materials, conducting experiment, analyzing data, paper writing and revising the manuscript.

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Biotechnology & Breeding Department, Sime Darby Plantation R&D Centre, Banting, Selangor, Malaysia. [2]School of Biosciences, University of Nottingham Malaysia, Semenyih, Selangor, Malaysia. [3]School of Biosciences, University of Nottingham, Nottingham, UK. [4]Department of Biological Sciences, National University of Singapore, Lower Kent Ridge Rd, Singapore, Singapore.

Received: 23 May 2019 Accepted: 9 October 2019
Published online: 05 November 2019

References

1. Corley RHV. Illegitimacy in oil palm breeding - a review. J Oil Palm Res. 2005;17:64–9.
2. Beirnaert A. Introduction à la biologie florale du palmier à huile (Elaeis guineensis Jacquin), vol. 5. Congo Belge, Ser. Sci. Inst. Nat. Etude agron; 1935. p. 3–42.
3. Beirnaert A, Vanderweyen R. Contribution à l'étude génétique et biométrique des variétés d'Elaeis guineensis Jacq. In: Publ Inst Nat Etude Agron Congo Belge Ser Sci; 1941. p. 1–101.
4. Singh R, Low E-TL, Ooi LC-L, Ong-Abdullah M, Ting N-C, Nagappan J, Nookiah R, Amiruddin MD, Rosli R, Manaf MAA, et al. The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. Nature. 2013;500(7462):340–4.
5. Lim CC, Rao V. DNA marker technology and private sector oil palm breeding. Planter. 2004;80:611–28.
6. Ooi LC-L, Low E-TL, Abdullah MO, Nookiah R, Ting NC, Nagappan J, Manaf MAA, Chan K-L, Halim MA, Azizi N, et al. Non-tenera Contamination and the Economic Impact of SHELL Genetic Testing in the Malaysian Independent Oil Palm Industry. Front Plant Sci. 2016;7(771):1–13.
7. Dumortier F, Van Amstel H, Corley RHV. Oil palm breeding at Binga, Zaire, 1970-1990. London: Unilever Plantations; 1992.
8. Production of crude palm oil in Malaysia for 2016 and 2017. http://bepi.mpob.gov.my/index.php/en/statistics/production/177-production-2017/792-production-of-crude-oil-palm-2017.html. Accessed 15 Sept 2019.
9. Kwong QB, Teh CK, Ong AL, Heng HY, Lee HL, Mohaimi M, Low JZ, Sukganah A, Chew FT, Mayes S, et al. Development and validation of a high-density SNP genotyping Array for African oil palm. Mol Plant. 2016;9(8):1132–41.
10. Teh C-K, Ong A-L, Kwong Q-B, Apparow S, Chew F-T, Mayes S, Mohamed M, Appleton D, Kulaveerasingam H. Genome-wide association study identifies three key loci for high mesocarp oil content in perennial crop oil palm. Sci Rep. 2016;6:19075.
11. Teh C-K, Muaz SD, Tangaya P, Fong P-Y, Ong A-L, Mayes S, Chew F-T, Kulaveerasingam H, Appleton D. Characterizing haplotype efficiency of SHELL gene to improve fruit form prediction in introgressive hybrids of oil palm. Sci Rep. 2017;7(1):3118.
12. Rajesh MK, Thomas RJ, Rijith J, Shareefe M, Jacob PM. Genetic purity assessment of D x T hybrids in coconut with SSR markers. Indian J Genet. 2012;72(4):472–4.
13. Jeffreys AJ, Wilson V, Thein SL. Hypervariable 'minisatellite' regions in human DNA. Nature. 1985;314:67.
14. Jeffreys AJ, Morton DB. DNA fingerprints of dogs and cats. Anim Genet. 1987;18(1):1–15.
15. Burke T, Bruford MW. DNA fingerprinting in birds. Nature. 1987;327:149.
16. Bakoume C, Azlah MY, Praveena T, Teh CK, Suzaini Y, Hamidah M, Mohamad SJ, Mohd NB, Khairudin H, Harikrishna K. DNA Sequence-Based Markers for Verification of Ramet-to-Orbet Relationship in Oil Palm (*Elaeis guineensis* Jacq.). Am J Plant Sci. 2011;2:539–48.
17. Hama-Ali EO, Alwee SSRS, Tan SG, Panandam JM, Ling HC, Namasivayam P, Peng HB. Illegitimacy and sibship assignments in oil palm (Elaeis guineensis Jacq.) half-sib families using single locus DNA microsatellite markers. Mol Biol Rep. 2015;42(5):917–25.
18. Singh R, Jayanthi N, Tan SG, Jothi Malar P, Cheah SC. Development of simple sequence repeat (SSR) markers for oil palm and their application in genetic mapping and fingerprinting of tissue culture clones. Asia Pac J Mol Biol Biotechnol. 2007;15(3):121–31.
19. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28(24):3326–8.
20. Semagn K, Babu R, Hearne S, Olsen M. Single nucleotide polymorphism genotyping using Kompetitive allele specific PCR (KASP): overview of the technology and its application in crop improvement. Mol Breed. 2014;33(1):1–14.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC

Chapter 4: MAJOR QTL FOR TRUNK HEIGHT AND CORRELATED AGRONOMIC TRAITS PROVIDE INSIGHTS INTO MULTIPLE TRAIT INTERGRATION IN OIL PALM BREEDING

4.1 Introduction

Oil palm is the most efficient oil crop in the world and produces 10 times more oil than soybean per hectare of crop land. Since the start of oil palm breeding programmes, oil yield improvement in the kernel and mesocarp has been the major interest of the industry, to maximise plantation profit. However, a shift of trait preference is occurring as labour shortages for collecting FFB and loose fruits have become evident, especially in Malaysia. This partly explains why the national oil yield in Malaysia has been stagnated for many years, even though high-yielding planting materials are produced in the country. To ease harvest operations, some mechanisation has been introduced, but the commercial value is still very restricted due to poor height uniformity of oil palms and unsuitable planting terrain.

Hence, the oil palm breeders are keen to integrate reduced trunk height (THT) and frond length (FL) through selection to produce compact oil palms. A direct benefit is to improve the harvest efficiency in the field, while shorter FL allows higher planting density from 138 up to 180 palms per hectare (Barcelos et al., 2015) to obtain higher FFB yield per area. Currently, two major short THT breeding materials, including *Gunung Melayu* (GM) and Dumpy AVROS (DAV), are maintained in Sime Darby Plantation R&D Centre and are being evaluated in trials. Lee and Pang (2000) reported that no significant difference in THT was observed between GM x AVROS *pisifera* and its commercial DxP control. Hence, the company decided to pursue hybridisation between *GM* and *DAV*, aiming to increase the reduction in THT. However, evaluation of this material is yet to be reported.

In other crops, agronomic fitness of a cultivar usually drops when introgression takes place due to linkage drag by the inferior effects of non-target genes or QTL linked to the genes, which are often inherited from the wild donor (Young and Tanksley, 1989). Furthermore, when selecting for multiple complex traits, breaking the linkage drag in perennial crops such as oil palm via conventional breeding programmes is always restricted by small progeny size with 16 – 96 palms planted per cross. With a planting density of 136 – 146 palms per hectare of crop land, trial maintenance is too costly and labour intensive to normally allow large planting of progenies. This is where the MAS can help. The objectives of this study include (1) to evaluate the trait variation in low height increment between GM x GM population and its GM x DAV *tenera* progeny, (2) to localise QTL for THT and other correlated agronomic traits in the GM x GM population and (3) to consolidate the identified QTL for selected traits for MAS to accelerate short THT breeding programmes with better trait predictability.

4.2 Methodology

Two low trunk height increment breeding populations, namely GM x GM (n = 113 palms) and its GM x DAV progeny (n = 70 palms) maintained at Sime Darby Plantation R&D Centre, Selangor, Malaysia were proven to be legitimate using the developed legitimacy test (in Chapter 3) and selected for this study. Yield recording and bunch analysis were performed according to the industry standard (Blaak et al., 1963) with modifications (Rao et al., 1983) when both populations reached three years old. For THT and other vegetative traits, the recording were done twice at 7 and 14 years after field planting using the default method (Corley et al., 1971). Presence of low THT and short FL were then determined by comparing the GM x GM population and the GM x DAV population to their respective DxP control planted in the same trials. A high-density linkage map of the GM x GM population was successfully constructed using the OP200K SNP array (Kwong et al., 2016) and in-house developed SSR markers. The map served as a reference to localise the QTL for the recorded traits. From the

analysis, the major QTL for THT, FL and total oil yield per palm (O/P) traits were identified, and followed by definition of positive alleles and threshold of cumulative effects of the positive alleles on each trait. Eventually, the threshold numbers of positive alleles and mean trait values were used to perform palm selection for multiple traits visualised on a scatter plot between traits.

4.3 Results & Discussion

To address the first objective, low THT and short FL traits were found to be heritable from GM and DAV origins. The result supports the idea that both THT and FL traits of a *tenera* population can be reduced through the hybridisation between the two origins, instead of using a single genetic donor, such as the GM origin reported earlier (Lee and Pang, 2000). Besides, the correlation analysis among the traits revealed several interesting insights:

1. Significant correlations between the derivative traits and associated component traits validated the robustness of the trait data.

2. The correlations between independent traits support meaningful biological interpretations, such as a decline in number of fruitlets formed in a bunch perhaps caused by diluted pollination efficiency of weevils when too many female inflorescences developed in an area.

3. Trunk growth is not linear across palm ages, which agrees with a previous study in oil palm (Henson, 2006). The mature palms tend to grow taller to compete for sunlight, so the genotype x environment effect on THT might complicate the oil yield at the

post-mature stage. This may explain the lack of a clear correlation between the 14th year THT and the early oil yield traits taken before 7 years after field planting.

The second objective was achieved. A total of 22,190 linked SNP markers were mapped to a high-density linkage map of the GM x GM population. The map with an average mapping interval of 0.05 cM spanned a cumulative length of 1,618.51 cM. However, QTL localisation on such high number of markers in a cross-pollinating population (CP) using MapQTL®6 programmes is technically not feasible (Van Ooijen, 2009). This study has provided a solution, which is reducing the number of linked markers at the threshold of mapping interval <10cM. The result is promising and major QTL for THT, FL and O/P traits were identified (Figure 4). Interestingly, one of the major QTLs locating on linkage group LG9 is likely the hotspot of pleotropic effects, including THT, FL and O/P, supported by the significant correlations among the traits. More importantly, the cumulative effect of quantitative alleles for each QTL now is ready to be used in MAS either for single-trait or multiple-trait selection. For a multiple trait integration programme in oil palm, this probability of success makes it close to 'impossible' to break the linkage drag, especially if the current breeding approach of using limited family size (usually not more than 96 palms assessed per cross) remains unchanged. With the help of MAS, breeders now can fully utilise the best 1% selection (10 – 15 palms selected from a full screening on 1,000 – 1,500 seeds per cross, which is often achieved in a mature fruit bunch).

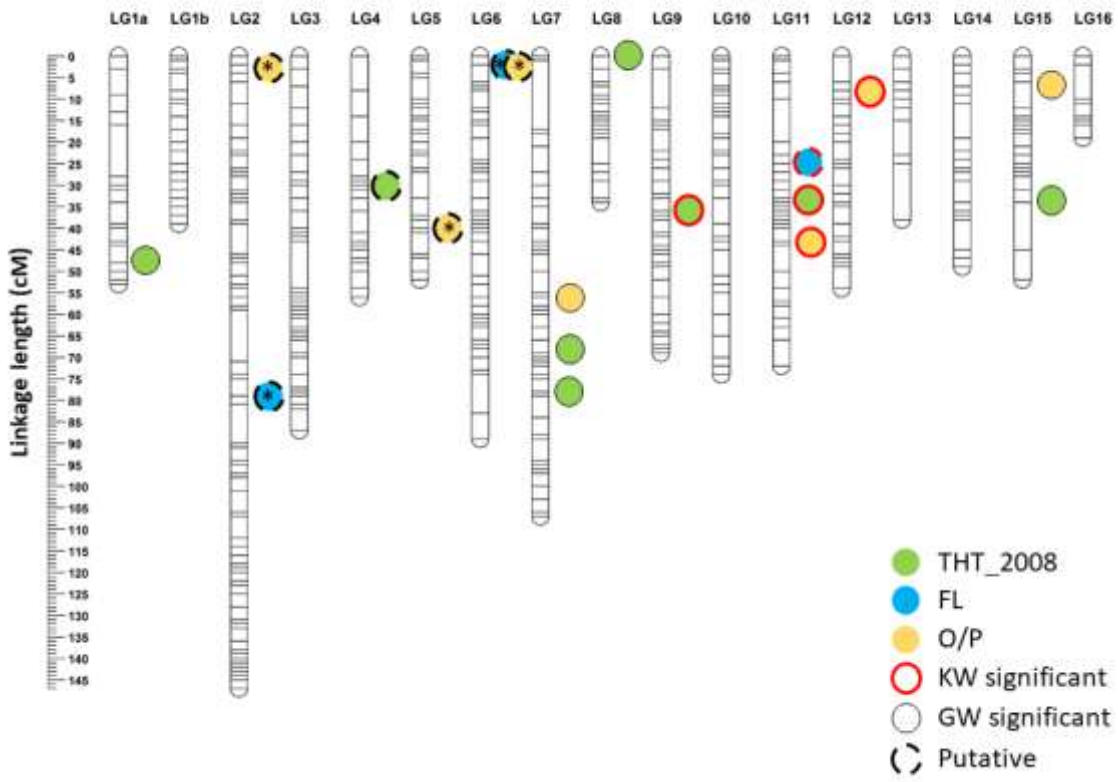


Figure 4: Distribution of quantitative trait loci (QTL) for trunk height in 2008 (THT_2008), frond length (FL) and total oil yield per palm (O/P) on a linkage map of GM x GM population. KW significant – Kruskal-Wallis at the p = 0.005 threshold; GW significant = genome-wide logarithm of odds, to the base 10 (LOD) = 4.0 threshold; putative – 2.4 ≤ LOD ≤ 4.0 threshold; * non-cofactor, based on Teh et al. (2020)

4.4 Conclusion and key recommendation for future work

The reduced THT and FL traits carried by GM and DAV origins has been investigated in this study, showing good potential for selective breeding. Furthermore, a good understanding of

the correlation among the agronomic traits and the effects of the cumulative positive alleles of QTL pave the way for oil palm breeders to conduct multiple trait integration through the MAS method, particularly to produce high-yielding short or compact oil palms. Now, with a 1% probability of success in multiple trait integration, future trials need to be saturated with elite palms selected from a large starting population. Meanwhile, non-linear THT growth across developmental stages should also be taken cautiously, depending on the breeding objectives.

4.5 Published journal article contributing to the objectives in Chapter 4

Teh C K, Ong A L, Mayes S, Massawe F, Appleton D R (2020): Major QTLs for trunk height and correlated agronomic traits provide insights into multiple trait integration in oil palm breeding. Genes 11, 826.doi: 10.3390/genes11070826

genes

Article

Major QTLs for Trunk Height and Correlated Agronomic Traits Provide Insights into Multiple Trait Integration in Oil Palm Breeding

Chee-Keng Teh [1,2,*], Ai-Ling Ong [1,2], Sean Mayes [3], Festo Massawe [2] and David Ross Appleton [1]

1 Biotechnology & Breeding Department, Sime Darby Plantation R&D Centre, Serdang 43400, Selangor State, Malaysia; ong.ailing.sdtc@simedarbyplantation.com (A.-L.O.); david.ross.appleton@simedarbyplantation.com (D.R.A.)
2 School of Biosciences, University of Nottingham Malaysia, Semenyih 43500, Selangor State, Malaysia; festo.massawe@nottingham.edu.my
3 School of Biosciences, University of Nottingham, Sutton Bonington Campus, Leicestershire LE12 5RD, UK; Sean.Mayes@nottingham.ac.uk
* Correspondence: teh.chee.keng@simedarbyplantation.com

Received: 7 May 2020; Accepted: 16 June 2020; Published: 21 July 2020



Abstract: Superior oil yield is always the top priority of the oil palm industry. Short trunk height (THT) and compactness traits have become increasingly important to improve harvesting efficiency since the industry started to suffer yield losses due to labor shortages. Breeding populations with low THT and short frond length (FL) are actually available, such as *Dumpy* AVROS *pisifera* (DAV) and *Gunung Melayu dura* (GM). However, multiple trait stacking still remains a challenge for oil palm breeding, which usually requires 12–20 years to complete a breeding cycle. In this study, yield and height increment in the GM × GM (GM-3341) and the GM × DAV (GM-DAV-3461) crossing programs were evaluated and palms with good yield and smaller height increment were identified. In the GM-3341 family, non-linear THT growth between THT_2008 (seven years old) and THT_2014 (13 years old) was revealed by a moderate correlation, suggesting that inter-palm competition becomes increasingly important. In total, 19 quantitative trait loci (QTLs) for THT_2008 (8), oil per palm (O/P) (7) and FL (4) were localized on the GM-3341 linkage map, with an average mapping interval of 2.01 cM. Three major QTLs for THT_2008, O/P and FL are co-located on chromosome 11 and reflect the correlation of THT_2008 with O/P and FL. Multiple trait selection for high O/P and low THT (based on the cumulative effects of positive alleles per trait) identified one palm from 100 palms, but with a large starting population of 1000–1500 seedling per cross, this low frequency could be easily compensated for during breeding selection.

Keywords: compact palms; marker-assisted selection; *Elaeis guineensis*; linkage maps; dwarfism

1. Introduction

Oil palm (*Elaeis guineensis* Jacq.) is a major source of vegetable oils and fats, contributing 35% of global vegetable oil production [1]. Per hectare of cropland, oil palm is able to yield 10 times more oil per year than any other temperate oil crop, which leads to oil palm cultivation being one of the most efficient agricultural land uses in humid tropical environments [2]. This also explains why only 7% of total global oilseed-harvested land is occupied by oil palms [3] yet ~35% of the global trade is in palm oil. To sustain long-term profit, enormous efforts have been put into improving oil yield per unit area, which involves two main components, i.e., fresh fruit bunch (FFB) production and oil-to-bunch ratio (O/B).

In recent years, an acute labor shortage in Malaysia has become a major problem for collecting FFB and loose fruit in the field. Foreign labor represents over 70% of the total 491,339 workforce of oil palm plantations in Malaysia, which reflects their high dependency on migrants who mainly come from Indonesia, Bangladesh, the Philippines, Thailand and Myanmar [4]. In addition, foreign workers are increasingly choosing to work in oil palm plantations in their home countries for similar or higher wages. This has contributed to stagnant oil yield in Malaysia, which leads to suboptimal estate management, even though high-yielding commercial seeds are available in the country. For example, Sarawak alone incurred losses of up to US$ 240 million per year in uncollected FFB because of labor shortages ranging from 15% to 20% across plantations [5]. Several mechanization approaches to harvest FFB and to collect loose fruit have been tested, but their commercial applicability is still restricted mainly due to the non-uniform palm architecture and topography of estates. Hence, the oil palm industry is keen to integrate oil yield traits and secondary vegetative traits, such as reduced trunk height and increased bunch stalk length [6] through selective breeding for better harvesting efficiency. Compact palms are a typical example. By retaining equivalent levels of inter-palm competition, the short rachis trait allows up to 180 palms per hectare, which is higher than the standard densities (138–143 palms per hectare) [7]. Improvements in yield per area, however, can only be fully realized with an increase in labor supply and/or increased levels of mechanization. Thus, the integration of slow height increment into compact palms is crucial to reduce the additional demand for labor and to facilitate mechanization.

One of the important genetic resources for reduced height increment is a self-pollinated family, Deli *dumpy* E206, which has been found to be uniformly short, indicating that the trait is highly heritable [8]. The family was crossed into many breeding programs, especially in Southeast Asia. In oil palm, the mesocarp oil content is partly determined by the presence/thickness of non-productive kernel shell in the fruit. To achieve 30% more oil, the commercial thin-shelled *tenera* are produced from crosses between thick-shelled *dura* and shell-less *pisifera* (usually female sterile). Hence, Deli *dura* with 75% *dumpy* ancestry × commercial AVROS *pisifera* produced a comparable oil yield to the *dura* × *pisifera* (DxP) control, while retaining lower trunk height growth due to the dumpy character [9]. Breeders from the Golden Hope and Highlands Research Unit (now under the Sime Darby Plantation) began exploiting this material differently. They introgressed the current pollen source, which is derived from AVROS *pisifera* and used very widely in Malaysia, into a new origin including 25% *dumpy* ancestry, termed *dumpy* AVROS, and hybridized these with normal Deli *dura* to generate commercial *tenera* palms with shorter trunk height increment [10]. Also, the company adopted the *Gunung Melayu dura* (GM), which originated from the Deli and was selected for its low height increment and high mesocarp oil content. No significant difference in trunk height between *GM* × AVROS *pisifera* and other Deli *dura* × AVROS *pisifera* was observed [11]. However, to date, evaluation of the hybridization between GM and *dumpy* AVROS *pisifera* (DAV) to reduce trunk height has yet to be reported.

The breeding fitness of a cultivar may drop when introgression from non-elite sources of germplasm occurs. The phenomenon is mostly due to linkage drag of the undesirable effects (normally from the wild donor) of non-targeted genes or quantitative trait loci (QTL) linked to the targeted genes or inherited from other chromosomes during crossing. Multiple selection and backcrossing cycles with sufficient progeny are crucial to identify the individuals that break the linkage drag by recombination between the QTL and undesirable genes to recover an elite background. Breaking of unwanted linkages in perennial crops, such as oil palm, based on phenotypic selection is costly, time consuming and inaccurate for all but the simplest phenotypes. Many fruit agronomic traits of oil palm can only be measured after three years of field planting, followed by another four years of recording to complete a selection cycle. In addition, the environmental effects on polygenic/complex traits, including oil yield and trunk height (THT) are strong, which presents a great challenge to capturing the genetic effects in improved breeding materials. To address this, locating the QTL underlying these traits has become increasingly important. The objectives of this study include: (1) to evaluate the trait variation in reduced trunk height increment between a short-trunk maternal GM × GM population and its *tenera* progeny (GM × DAV), (2) QTL mapping for THT and THT-correlated agronomic traits in a GM × GM

population using a high-density OP200K genotyping array [12] and simple sequence repeats (SSRs), and (3) consolidating the QTL for selected traits and marker-assisted selection (MAS) to accelerate reduced height breeding programs with better trait predictability.

2. Materials and Methods

2.1. Plant Materials and DNA Preparation

A full-sib *dura* family, namely, GM-3341 (n = 113 palms) was obtained from a cross between GM-Parent1 and GM-Parent2. To determine variation in the THT trait, a progeny test family, i.e., GM-DAV-3461 *tenera* (n = 70 palms) derived from the common GM-Parent2 *dura* and DAV-Parent1 *pisifera* was selected and compared. Both GM-3341 and GM-DAV-3461 families were planted in 2001 in a randomized complete block design (RCBD) at a density of 148 palms per hectare in the East Estate (PT132 trial, 10 replicates per family) and the North Estate (PT138 trial, 5 replicates per family) of the Sime Darby Plantation on Carey Island, Malaysia. A total of 76 commercial DxP palms derived from an admixed parentage of Deli *dura* × AVROS *pisifera* were planted as a control in each trial. The control planted in the PT132 trial and PT138 trial were named DxP control (GM) and DxP control (GM-DAV), respectively. The estates generally share similar annual temperatures and rainfall. For linkage analyses, a total of 113 palms of the GM-3341 family were sampled at unopen frond 0 for DNA extraction. Total genomic DNA was isolated from 0.1 g of the leaf tissue using the DNAeasy Plant Mini Kit (Qiagen, Hilden, Germany). The DNA quality was visualized on a 0.8% agarose gel and quantified using the Nanodrop 2000 (NanoDrop Technologies, Wilmington, DE, USA). The DNA samples were stored at −20 °C in a freezer for subsequent genotyping.

2.2. Trait Recording and Statistical Analyses

In oil palm breeding (and research) programs, the industry standard is followed, which usually involve the planting of controlled crosses in RCBD in two locations. Crosses are extensively recorded for certain traits. For example, bunch yield for every bunch (usually one bunch per month) for several years, while others are more time intensive (such as bunch analysis, where a minimum of 3 per palm is standard) or have been deemed of less importance (height increment is one of these). Overall, this tends to reflect the ease of data collection as opposed to the importance of the trait for the breeding program [13]. In this study, yield recording was started when both trials reached three years old. Both the number of bunches (BNO) and average bunch weight (BWT) were recorded to generate 4-year averages for FFB (from 2004 to 2008). At least three bunches per palm were analyzed for O/B and other bunch components according to the industry standard [14] with modifications [15]. The total oil per palm (O/P) was then calculated from the multiplication of the 4-year average FFB and average O/B. For the THT trait, the distance from the ground to the 41st frond was measured in 2008 (as THT_2008) and 2014 (as THT_2014) using an extendable ruler. At the same time, other vegetative traits were also measured based on industry standard methods [16]. A total of 16 traits (Table 1) were compiled for standard descriptive statistics; trait values between 1.5 and 3 times the interquartile range were identified as outliers and removed. Presence of low THT and short frond length (FL) were determined separately in the GM-3341 family at the PT132 trial and the GM-DAV-3461 family at the PT138 trial by comparing them to their respective DxP controls using a two-sample T test at p = 0.05 threshold. A Pearson correlation analysis of the 16 traits in the GM-3341 family was also performed at the p = 0.05 threshold. All statistical tests were done in the Minitab® 17 program [17] and box plots were generated in R.

Table 1. Pearson correlation of trunk height traits, frond length and oil yield-related traits of the GM-3341 family.

Trait	Trait Abbreviation	THT_2008	THT_2014
Fresh fruit bunch	FFB/kg	0.423 *	0.242 *
Bunch number	BNO	0.364 *	0.119
Bunch weight	BWT	0.248 *	0.156
Fruit-to-bunch ratio	F/B	−0.055	−0.015
Mean fruit weight	MFW	−0.068	0.107
Mesocarp-to-fruit ratio	M/F	−0.150	0.103
Shell-to-fruit ratio	S/F	0.156	−0.153
Kernel-to-fruit ratio	K/F	0.033	0.043
Oil-to-dry mesocarp ratio	O/DM	0.237	0.216
Oil-to-wet mesocarp ratio	O/WM	−0.038	−0.042
Oil-to-bunch ratio	O/B	0.126	0.069
Kernel-to-bunch ratio	K/B	0.081	−0.005
Total oil per palm	O/P	0.294 *	0.165
Frond length	FL	0.270 *	0.301 *
Trunk height in 2008	THT_2008	NA	0.595 *
Trunk height in 2014	THT_2014	NA	NA

* Significant correlation at the $p = 0.05$ threshold; NA—not applicable.

2.3. *DNA Marker Genotyping*

A *tenera* palm, namely, Progeny palm 5002 derived from D4-93/76 (Deli *dura*) × 2/2311 (EKONA *pisifera*) was whole-genome sequenced with 30X coverage using the Roche 454 Titanium FLX platform. From the reference genome, a total of 11,653 SSRs were discovered and primers designed. The SSR amplification conditions with fluorescent-labeled M13-tailed forward primers [18] were optimized for each of the primer pairs and followed by genotype scoring using an Applied Biosystems 3730xl DNA Analyser (Applied Biosystems, Carlsbad, CA, USA). The SSR primers formed a multiallelic marker framework. Furthermore, the OP200K Infinium array (Illumina, San Diego, CA, USA) based on 59 breeding origins of oil palms [12] was applied to the population. Only the markers with heterozygous genotypes in at least one of the parents of the GM-3341 family were defined as informative and were used to genotype the mapping population.

2.4. *Construction of Linkage Maps*

The GM-3341 family, which is defined as a cross-pollinator (CP) population, was deployed as a mapping population. A full-density GM × GM linkage map was constructed using Lep-MAP3 [19,20] for large number of SNPs, which is applicable to outbred families, such as oil palm, based on the maximum likelihood method. Marker quality was checked under the *Filtering* module. Informative SNPs with >5% of missing data and/or identical genotype segregation were removed. Subsequently, the *Separate Chromosomes* module was used for binning the assayed SNPs in each parent based on the optimized logarithm of odds, to the base 10 (LOD) thresholds ranging from 4 to 30. To order the binned SNPs, the Kosambi mapping function was adopted for the conversion of recombination frequencies into map distances (in unit cM). The maps of both parents were joined as one using the option "sexAveraged = 1". A subset of SNPs was selected from every 5 cM of the full-density linkage map and loaded into the Joinmap® 5 program [21] for linkage mapping to enable QTL mapping analysis. In addition, a set of 67 informative SSRs were also included. The markers with segregation distortion at chi-square $p < 0.001$ (***) were removed. The selected markers were ordered based on optimized LOD thresholds (6.0–9.0) with the Kosambi map function and regression model, and the linkage map was visualized using the Mapchart 2.32 program [22]. Each LG number was identified with a chromosome of the published physical map [23]. In addition, the linked SNPs were also mapped onto the same physical map to estimate correlation between linkage (in cM) and physical (in Mb) positions in the genome (Tables S1 and S2), which allowed recombination estimates (in cM/Mb)

across the pseudo-chromosomes. Finally, the average mapping interval was calculated at LG and genome-wide levels.

2.5. Quantitative Trait Locus Mapping Analyses

All QTL mapping analyses for THT_2008, THT_2014, O/P and FL traits were carried out using the MapQTL® 6 program [24]. Initially, single marker tests using the rank sum test of Kruskal-Wallis (KW) were done at the $p = 0.005$ threshold, followed by interval mapping (IM) to identify potential QTL peaks. In order to determine the genome-wide LOD threshold, permutation tests were performed with 1000 iterations at $\alpha = 0.05$, but 'putative' QTLs at 1.5 below the genome-wide LOD = 4.0 threshold were still reported. The significant QTLs detected in IM were fixed as cofactors for multiple QTL mapping (MQM) mapping. Any new QTLs detected were considered as cofactors for the next round of MQM mapping and the process were iterated until no change occurred. Then, trait variation explained by the identified QTL was determined based on MQM mapping analysis.

2.6. Cumulative Effects of Quantitative Trait Allele Analysis

The putative and genome-wide significant QTLs were selected to measure the marker effects for THT_2008, FL and O/P traits based on median comparison among alleles in Minitab® 17 program [17]. In each QTL, the positive alleles in each trait that benefit the breeding program were defined in the GM-3341 family (107 palms after outlier removal). The alleles that decreased FL and THT_2008 values were known as positive, whereas the positive alleles for O/P increased the trait value. Next, the total number of positive alleles across the QTLs in individual palms were plotted against each trait to determine the cumulative effects of quantitative trait alleles. The threshold number of positive alleles for each trait selection was defined based on the intersection between the mean line of a trait and the fitted line on a scatter plot. The threshold numbers of positive alleles and mean trait values were then used to perform palm selection for multiple traits on a scatter plot between traits.

3. Results

3.1. Descriptive Statistics and Comparison of Traits between Families

Sixteen traits related to oil yield, THT and FL were recorded in the GM-3341 *dura* family ($n = 107$ palms) and the GM-DAV-3461 *tenera* family ($n = 70$ palms) located in PT132 in the East Estate and PT138 in the North Estate, respectively (Table S3). A commercial DxP control was also included in each trial and served as a control for comparison of traits using a two-sample T test at the $p = 0.05$ threshold. After removal of the outliers, the 16 traits followed normal distributions at the $p = 0.05$ threshold and the dataset was subjected to subsequent analyses. The GM-3341 family produced 19.50 ± 5.00 kg/palm/year of O/P, which was lower than the 26.25 ± 7.44 kg/palm/year produced by the GM-DAV-3461 family. Poorer average performance in both FFB (103.37 ± 21.21 kg/palm/year) and O/B (18.72 ± 2.70%) in the GM-3341 family were the main contributors, compared to 125.90 ± 24.49 kg/palm/year of average FFB and 20.29 ± 3.90% of average O/B in the GM-DAV-3461 family. Shell thickness variation within *tenera* was measured by the shell-to-fruit ratio (S/F). As expected, the GM-3341 family had a thicker shell (average S/F = 30.97 ± 3.24%) than that of the GM-DAV-3461 family (average S/F = 15.59 ± 2.82%), with the GM-3341 being purely *dura* shell-types and the GM-DAV-3461 being purely the thinner *tenera* hybrid. In terms of average fruit-to-bunch ratio (F/B), the GM-3341 family and the GM-DAV-3461 recorded 63.60 ± 4.71% and 60.45 ± 5.63%, respectively, which were within the normal range, i.e., 60–65% [13]. Trunk height and FL were measured twice, once in 2008 (as THT_2008) and also in 2014 (as THT_2014) at PT132 trial, whereas only a single measurement was done at the PT138 trial in 2008 (Table S3). The GM-3341 family at the PT132 trial with a mean THT_2008 = 145.07 cm and mean FL = 534.53 cm were significantly shorter than the DxP control, which had means for THT_2008 = 213.02 cm and FL = 554.83 cm at the $p = 0.05$ threshold for both (Figure 1A,B). The same result was observed in the PT138 trial, where the

GM-DAV-3461 family had means for THT_2008 = 206.37 cm and FL = 565.34 cm and were significantly shorter than the DxP control with means for THT_2008 = 266.90 cm and FL = 584.30 cm at the same p threshold (Figure 1C,D).

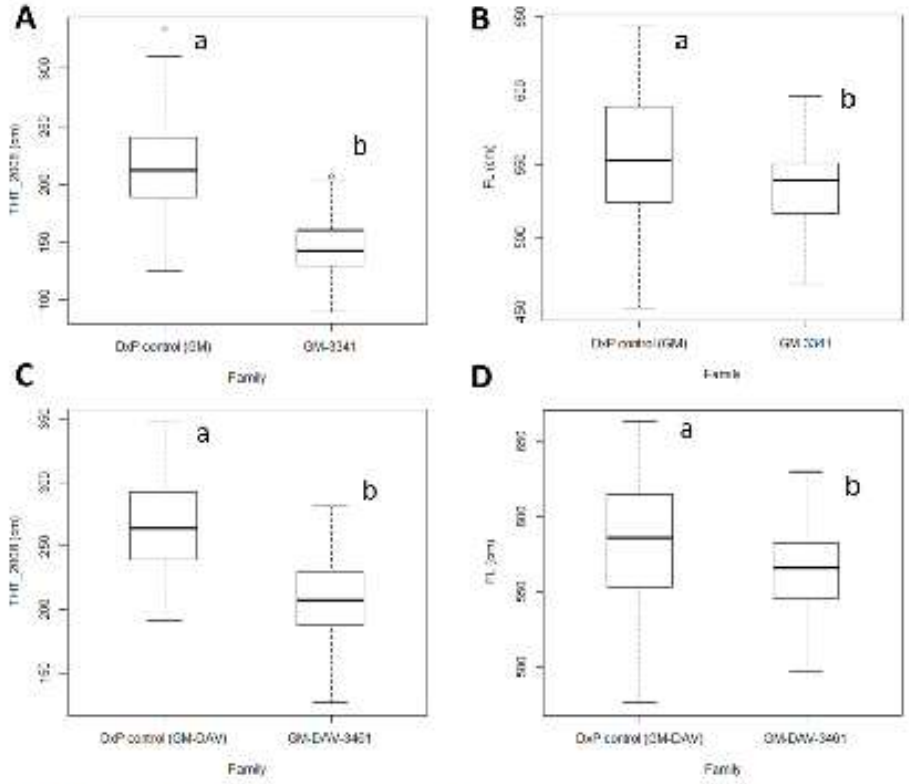


Figure 1. The boxplots represent median values, percentile 25–75 for trunk height measured in 2008 (THT_2008) and frond length (FL) between families. (**A**) THT_2008 comparison between the DxP control and the GM-3341 family in the PT132 trial in the East Estate; (**B**) FL comparison between the DxP control and the GM-3341 family in the PT132 trial in the East Estate; (**C**) THT_2008 comparison between the DxP control and the GM-DAV-3461 family in PT138 in the North Estate; (**D**) FL comparison between the DxP control and the GM-DAV-3461 family in PT138 in the North Estate. The trait comparisons were analyzed based on two-sample t test at the $p = 0.05$ threshold.

3.2. *Pearson Correlation among Traits*

The correlations among all the 16 traits of the GM-3341 family are illustrated in Table S4. As mentioned, O/P is a derivative trait from multiplying FFB and O/B per palm, thus significant correlations with FFB (correlation coefficient, $r = 0.619$; $p = 0.000$) and O/B ($r = 0.332$; $p = 0.007$) were observed. The same effects applied to the component traits of FFB and O/B. Fresh fruit bunch weight significantly correlated with BNO ($r = 0.824$; $p = 0.000$) and average BWT ($r = 0.253$; $p = 0.012$), while O/B significantly correlated with F/B ($r = 0.473$; $p = 0.000$), mesocarp-to-fruit ratio (M/F) ($r = 0.490$; $p = 0.000$) and oil-to-wet mesocarp (O/WM) ($r = 0.444$; $p = 0.000$). In addition, oil palm fruitlets are mostly made up of an outer layer of mesocarp and a central nut consisting of a shell-coated kernel. Hence, each of the fruitlet components is actually a proportion of the mean fruit weight (MFW) and this was reflected in the significant correlations among MFW, S/F, M/F and the kernel-to-fruit ratio (K/F). A significant correlation was also detected between independent oil yield-related traits such as

BNO with F/B ($r = -0.262$; $p = 0.036$). For vegetative traits, there was a correlation between FL and THT (THT_2008, $r = 0.270$; $p = 0.008$ and THT_2014, $r = 0.301$; $p = 0.009$) (Tables 1 and S4). Interestingly, THT recorded in 2008 and 2014 was not highly correlated ($r = 0.595$; $p = 0.000$), suggesting the non-linear THT increment of oil palm over the growing cycle. THT_2008 correlated better with yield-related traits including O/P, FFB, BNO and BWT, compared to THT_2014 and FL. This is likely to be before (THT_2008) and after (THT_2014) palm-to-palm competition becomes important.

3.3. *Construction of Linkage Maps*

The GM-3341 family, as a mapping population was genotyped using 200,000 SNPs. Of these, 33,190 SNPs were considered informative for linkage mapping and were successfully binned into linkage groups (LGs) that were consistent with 16 chromosome pairs of oil palm genome (2n = 32). The constructed linkage map with 33,190 SNPs spanned a cumulative length of 1618.51 centiMorgan (cM) with an average mapping interval of 0.05 cM and was termed the "full-density" map (data not shown). For QTL localization, a subset of 588 SNPs with an average mapping interval of 5 cM were coupled with 67 informative SSRs (Tables S1 and S2) to build a linkage map with reduced density. From this, 565 markers with ≤5% missing data were linked as a linkage map that had a cumulative length of 1093.51 cM and a lower average mapping interval of 2.01 cM (Figure 2, Table 2). The number of LGs, however, increased to 17, because LG1 fragmented into two, namely, LG1a and LG1b. The marker positions between full and reduced-density linkage maps were compared and were generally in good agreement. The latter map was then adopted for estimating an average recombination rate throughout the genome, which was 1.74 cM/Mb, and which ranged from 0.63 cM/Mb (LG1b) to 3.24 cM/Mb (LG10).

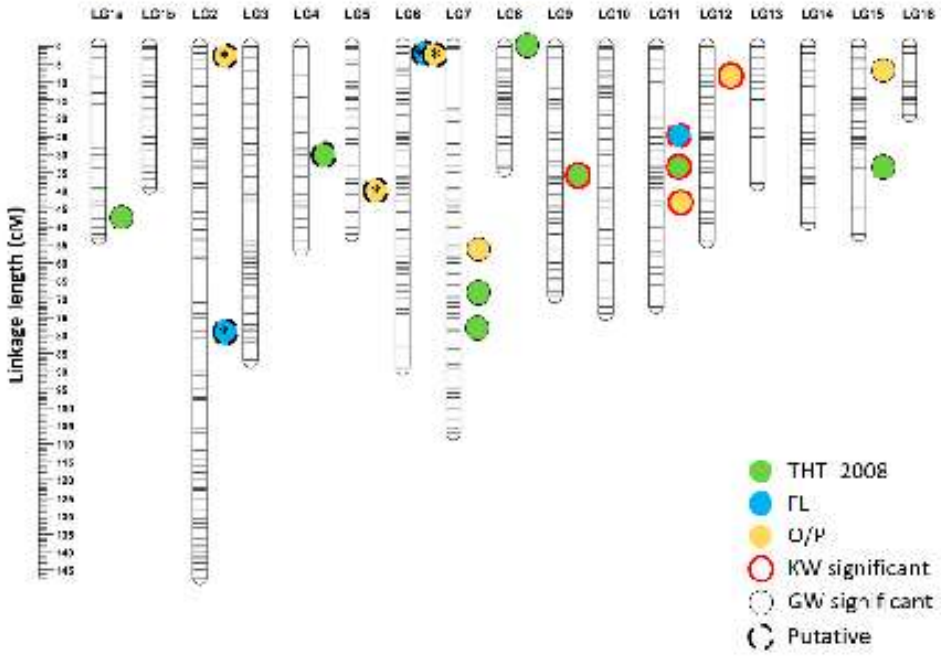


Figure 2. Distribution of quantitative trait loci (QTLs) for trunk height in 2008 (THT_2008), frond length (FL) and total oil per palm (O/P) on a linkage map of the GM-3341 family. KW significant—Kruskal-Wallis at the $p = 0.005$ threshold; GW significant—genome-wide logarithm of odds, to the base 10 (LOD) = 4.0 threshold; putative—2.4 ≤ LOD < 4.0 threshold; * non-cofactor.

Table 2. Linkage length and mapping interval across linkage groups in the GM-3341 family.

No.	Group	Linkage Length (cM)	No. of Linked Markers	Mapping Interval (cM)	No. of Markers in the Physical Map	Linkage Length of Common Markers (cM)	Physical Length (Mb)	Average Recombination Rate (cM/Mb)
1	LG1a	53.44	20	2.67	10	53.44	63.71	0.84
2	LG1b	38.99	20	1.95	18	38.99	61.63	0.63
3	LG2	147.43	73	2.02	36	147.43	64.78	2.28
4	LG3	86.94	48	1.81	27	86.94	58.66	1.48
5	LG4	55.97	32	1.75	9	55.97	34.73	1.61
6	LG5	52.11	34	1.53	20	52.11	37.29	1.40
7	LG6	88.89	55	1.62	22	83.33	41.90	1.99
8	LG7	107.34	49	2.19	27	105.61	42.27	2.50
9	LG8	33.67	28	1.20	22	33.67	39.98	0.84
10	LG9	69.28	38	1.82	23	68.16	34.66	1.97
11	LG10	74.31	35	2.12	19	74.31	22.94	3.24
12	LG11	72.36	37	1.96	13	72.36	26.05	2.78
13	LG12	54.00	28	1.93	26	54.00	26.63	2.03
14	LG13	37.90	11	3.45	7	23.12	25.51	0.91
15	LG14	49.50	23	2.15	18	49.50	23.78	2.08
15	LG15	51.95	22	2.36	16	45.24	22.25	2.03
17	LG16	19.43	12	1.62	9	16.43	17.05	0.96
	GW	1093.51 *	565 *	2.01 *	322 *	1060.61 *	643.82 *	1.74 **

* Sum total; ** average value; GW—genome-wide; LG—linkage map; cM—centiMorgan; Mb—Megabase; each LG identified as a chromosome of the published genome [23].

3.4. Mapping of QTLs for THT, FL and O/P Traits

The QTL mapping for THT and correlated traits started with single marker testing on 655 linked markers using the ranking sum test of KW. A total of 49 significant markers were detected to be effecting THT_2008 (30 markers), THT_2014 (1 marker), FL (13 markers) and O/P (5 markers) at the $p = 0.005$ threshold (Table S5). Subsequently, the lower-density linkage map of the GM-3341 family was used to conduct IM. The detected IM-based QTLs were fixed as cofactors for MQM analyses. LOD = 4.0 was set as the genome-wide significance threshold based on permutation tests with 1000 iterations at $\alpha = 0.05$, whereas the QTLs within $2.5 \leq$ LOD < 4.0 were termed "putative". First of all, no significant QTL for THT_2014 was found in IM and MQM. As shown in Figure 2, however, genome-wide significant QTLs for THT_2008 were detected on LG1a, LG7, LG8, LG9, LG11 and LG15, explaining 7.4–17.0% of the trait variation each (Figure 2, Table 3). Another putative QTL for the same trait was located on LG4 with 4.3% of the variation explained. As for FL, four putative QTLs on LG2 and LG6 and LG11 that explained 10.2–16.5% of the variation were found. Three of the QTLs on LG2 (rs795951455) and LG6 (rs795993306 and ss1810567844) emerged as non-cofactor peaks when rs795990373 on LG11 was fixed as a cofactor. Similarly, by fixing the four genome-wide significant QTLs for O/P residing at 56.03 cM on LG7, 43.79 cM on LG11, 34.73 cM on LG12 and 16.25 cM on LG15 as cofactors, three putative QTLs emerged on LG2, LG6 and LG5. The O/P variation explained by these QTLs ranged from 6.0–28.5%.

3.5. Cumulative Effects of Quantitative Trait Allele Analysis

In this analysis, we adopted 16, 9 and 6 positive alleles for THT_2008, O/P and FL, respectively. Only THT_2008 and O/P traits fit well into quadratic relationships between the number of the positive QTL alleles and trait variation, but in antagonistic directions (Figure 3A–C). The palms carrying more than 13 positive alleles for THT_2008 trait were mainly shorter than the population mean = 145.07 cm, whereas those with more than 7 positive alleles for O/P trait mainly yielded higher than the population mean = 19.50 kg/palm/year. Hence, the number of positive alleles for respective THT_2008 and O/P were used for multiple trait selection (Figure 3D). We found that 77% and 86% of the selected palms had shorter THT and higher O/P than the population mean, suggesting that the single trait selection method was effective. For multiple trait selection, by contrast, only one palm in the GM3341 family with low THT and high O/P (the orange dot shown in Figure 3D) was identified using the QTL markers.

Table 3. Multiple QTL mapping and Kruskal-Wallis test for trunk height recorded in 2008 (THT_2008), frond length (FL) and total oil per palm (O/P) traits in the GM-3341 family.

Trait	Linkage Group	Cofactor	Peak LOD	Peak Position (cM)	Trait Variation Explained (%)	KW
THT_2008	1a	rs795987799	4.55	48.04	7.6	
	4	rs795956088	2.67	30.22	4.3	
	7	rs796035893	6.65	68.67	11.7	
	7	ss1810504935	6.79	78.32	12.0	
	8	rs795967221	5.02	0.00	8.5	
	9	rs796032489	4.42	35.56	7.4	*****
	11	ss1810516343	9.11	33.55	17.0	****
	15	rs795995560	5.99	33.79	10.4	
O/P	2	DA290 ^{NC}	3.74	81.14	7.6	
	5	rs796007949 ^{NC}	2.88	39.55	6.0	
	6	DA358 ^{NC}	3.02	88.89	6.3	
	7	ss1810302303	4.09	56.03	10.9	
	11	rs796015770	4.56	43.79	12.4	****
	12	rs796017329	8.13	34.73	25.2	****
	15	rs795992318	8.92	16.25	28.5	
FL	2	rs795951455 ^{NC}	2.69	79.40	10.2	
	6	rs795993306 ^{NC}	3.11	2.26	11.7	
	6	ss1810567844 ^{NC}	3.11	2.38	11.7	
	11	rs795990373	3.73	24.55	16.5	******

THT_2008—Trunk height in 2008, THT_2014—Trunk height in 2014, O/P—Total oil per palm, ^{NC} non-cofactor, cM—centiMorgan, KW—Kruskal-Wallis statistics, **** 0.005 significance level, ***** 0.001 significance level, ****** 0.0005 significance level.

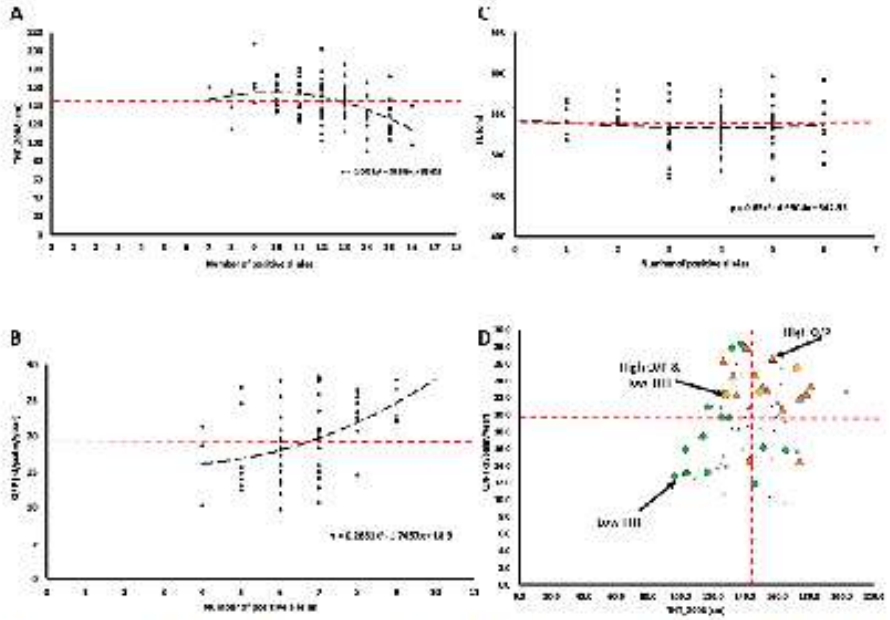


Figure 3. Cumulative effects of quantitative trait alleles for (**A**) THT_2008, (**B**) O/P, (**C**) FL traits, and (**D**) the application for selection of THT_2008 and O/P traits in the GM-3341 Family. Red dotted line—mean THT_2008 (145.07 cm), mean O/P (19.50 kg/palm/ year) and mean FL (534.53 cm); orange dot—High O/P and low THT_2008, red triangle—High O/P, green diamond—Low THT_2008; multiple trait selection in Figure 3D based on threshold numbers of positive alleles for THT_2008 = 13 and for O/P = 7.

4. Discussion

Oil palm breeding in Southeast Asia predominantly started with the Deli *dura*, which is believed to be derived from four palms planted in the Bogor Botanical Garden, Indonesia in 1848 [25,26]. The Deli *dura* were subsequently distributed to various breeding programs, each of which stressed different selection approaches over generations and caused separation between subpopulations, which are known as "breeding populations of restricted origin" (BPRO) [27]. However, these breeding programs still share the same overall objective, that is, superior oil yield, which has been achieved by different selection for traits. In addition, understanding the co-dominant inheritance of the shell thickness gene [28] also paved the way for exploitation of the superior oil content of the thin-shelled *tenera*, which was commercially produced from hybridization between thick-shelled Deli *dura* and shell-less AVROS *pisifera*. At the beginning of the 1960s, planters in the Southeast Asia region successfully realized a 30% increment in oil yield per hectare as *tenera* planting materials replaced the original Deli *dura* materials [13,29]. In this study, we also observed a 34.6% O/P increment from the GM-3314 *dura* family (19.50 ± 5.00 kg/palm/year) to the DAV-GM-3461 *tenera* family (26.25 ± 7.44 kg/palm/year) (Table S3).

Nevertheless, the oil palm industry in Malaysia started to experience labor shortages, particularly for FFB and loose fruit collection, around a decade ago and the problem is escalating. Hence, the breeding objective of superior oil yield is being expanded to include traits, such as low THT and compactness to improve harvest and management efficiency. In this study, low THT and short frond length within the GM-3341 family and its half-sib GM-DAV-3461 family was found to be heritable from GM *dura* and DAV *pisifera* (Figure 1). This finding supports the idea that both traits can actually be improved through hybridization between GM and DAV origins to produce short or even compact *tenera* palms, compared to using a single genetic donor of GM origin as reported earlier [11]. The robustness of trait data was verified based on significant correlations between the derivative traits and associated component traits of the GM-3341 family at the $p = 0.05$ threshold (Table S4). Meaningful biological interpretation was also performed based on the correlations between independent traits. For example, an inverse correlation between BNO and F/B ($r = -0.262$; $p = 0.036$) is likely to be related to the pollination ability of weevils (*Elaeidobius kamerunicus*), which is reflected in the number of insects per receptive female inflorescence per hectare [30]. Decline in "pollination force"' can happen if many receptive female inflorescences are developed in an area with a fixed population of weevils. In such scenarios, the palms still can produce more BNO, but a smaller number of flowers on these female inflorescences were pollinated to form fruitlets, leading to lower FFB. Bunches with more than 60% F/B are considered normal [30,31], so, both the GM-3341 family and the GM-DAV-3461 family did not encounter a fruit set problem.

In the GM-3341 family, trunk growth was found to be non-linear across different ages based on a moderate correlation between THT_2008 (seven years after planting) and THT_2014 (13 years after planting) ($r = 0.595$; $p = 0.000$) (Table 1). This finding agrees with previous studies of trunk growth in oil palm, which are generally affected by three main factors, i.e., palm age, genotype and environmental conditions (such as agronomic practices and soil types) [32]. In the first 3 years, trunk growth is usually negligible [33] and then becomes significant at six years onwards [34]. Furthermore, many studies also reported that THT does increase with planting density, but the effect only becomes observable at 9 years after planting, which is when the canopy starts to overlap [35,36]. The mature palms tend to grow taller to compete for sunlight, so the genotype effect on THT is no longer as clear at that stage. Taller palms will also capture more sunlight, so small differences in height may lead to large differences in FFB. This may have complicated the QTL detection for THT_2014 in IM and MQM. Palm-to-palm competition might affect the yield mechanism in these palms as well, based on the missing correlations between THT_2014 and yield traits (O/P, FFB, BNO and BWT). It is worth noting that the yield recording in this study was conducted around 7 years after planting, so is unlikely to include the effects of competition between palms.

Oil palm breeders have attempted to stack superior O/P, low THT and short FL to produce compact palms. In most breeding programs, multiple trait integration becomes a real challenge when

desirable traits are antagonistic to each other. Such a phenomenon is common in oil palm breeding for high-yielding, compact and low THT palms. We noticed that THT_2008 was significantly correlated to O/P ($r = 0.294$) and FL ($r = 0.270$), which is undesirable from a breeding perspective (Tables 1 and S4). These correlations seem to be biologically sensible in that vigorous vegetative growth of fronds and trunk height help to capture more sunlight for photosynthesis. However, it is not simple as that. In Arabidopsis, the relationship between leaf area growth and plant biomass was reported to be non-linear; it depends on how carbon assimilates produced from photosynthesis are partitioned into new leaf area, roots, reproduction and respiration [37]. In fact, the same findings for oil palm were also reported three decades ago [38,39]. Photosynthesis rate increases in response to sink demand in the crop. Extra assimilates will only be allocated to FFB production (effectively oil yield) after the demand for vegetative growth and maintenance is satisfied [34]. Thus, the moderate correlation between THT_2008 and O/P raises the possibility of selecting palms that buck this trend and have a higher proportion of dry matter in fruit bunches. However, when dealing with multiple complex traits, breaking linkage drag in perennial oil palm through conventional breeding methods will always be hindered by small progeny size with 16–96 palms per cross. With a planting density of 136–146 palms per hectare, the field evaluation of a large number of progenies is too costly and labor intensive—this is where MAS can help.

By using the OP200K genotyping array [12], the GM family with 33,190 linked markers was more genetically diverse than the reported commercial Deli × AVROS family with 27,890 linked markers [40], which presumably reflects the effect of breeding selection pressure in the latter population. High-density linkage maps are known to be a powerful way to improve physical genome assemblies. For instance, a further 311 Mb of scaffolds were successfully anchored to the pseudo-chromosomes of oil palm based on the Deli × AVROS linkage map [40]. The genome of melon (*Cucumis melo* L.) double haploid line DL92 was also extended with 354.8 Mb of sequence using a linkage map with an average mapping interval of 1.99 cM [41]. Nonetheless, QTL localization on a high-density linkage map using the MapQTL® 6 program is always constrained by insufficient degrees of freedom to accommodate many markers (typically > 600) due to uninformative segregation types in cross-pollinating (CP) populations [24], which is common in oil palm. Hence, we only selected markers at around 5 cM apart to construct a lower-density linkage map with 565 linked markers that spanned a cumulative length of 1093.51 cM with an average mapping interval of 2.01 cM, but with good genome coverage (Figure 2, Table 2). The marker order in both maps were highly correlated but the mapping interval was smaller than 5 cM, as expected when a degree of conflicting data is removed. Inflated map length and orientation errors are likely to be reduced after removing some markers with incorrect SNP calling. The genome-wide recombination rate of 1.74 cM/mb calculated in the GM genome was close to the Deli × AVROS genome (1.8 cM/Mb) [40] and the date palm genome (*Phoenix dactylifera*) (1.9 cM/Mb) [42], suggesting this is a reasonable range for palm species.

For QTL mapping, double recombination should rarely occur in the lower-density linkage map, which is still much denser than the 10 cM threshold [43]. Linkage group 1 fragmented into LG1a and LG1b in the lower-density map. In fact, this did not restrict the analysis, because all LGs are traceable to the 16 homologous chromosomes of the reference genome [23]. We preferred MQM analysis because it is a single dimensional search over the genome to test a single segregating QTL as in IM, while simultaneously fitting the selected cofactor, under both the null-hypothesis (H_0) and the alternative hypothesis (H_1). With this approach, the cofactors can reduce the residual variance, therefore, if a QTL explains a large proportion of the total variance, the linked markers used as cofactors in subsequent MQM analyses can significantly improve the power for identifying new segregating QTLs [24]. This method thus enabled detection of new QTLs for THT_2008, O/P and FL traits after fixing the major QTLs on LG9, LG11 and LG12 as cofactors (Figure 2). Interestingly, we found that ss1810516343 as the major QTL for THT_2018 was closely flanked by putative QTL locus rs795990373 for FL and genome-wide significant rs796015770 for O/P on LG9. This marker arrangement strongly reflects the correlation of THT_2008 with O/P and FL, suggesting a possible pleotropic effect. In fact,

the effects of reduced height on increased yield are seen in other crops. A good example is wheat, where the famous Reduced Height (*Rht*) gene has a pleotropic effect, decreasing plant height and increasing grain yield [44]; the subsequent discovery of *Rht-B1b* and *Rht-D1b* successfully improved seed yield by 5–10% with a 20% stem height reduction [45].

This study provides initial QTL locations to enable identification of causal genes, particularly for THT and O/P traits. For this purpose, the QTLs that were genome-wide and KW significant should be given the highest priority (Table 2). The GM-based QTLs for THT identified in this study are in different locations from previous studies based on Deli *dura* × DAV origin [46] and *Dura* × Ghana origin [47], which reveals the possible effects of different genes or gene paralogues across the genomes derived from different origins. Consequently, the application of the current set of QTL genes for height may remain origin specific, unless the genetic causality is fully understood. To prepare for MAS in the GM origin, the cumulative effects of quantitative trait alleles for THT_2008, O/P and FL were used as a model to determine the effects of different numbers of positive alleles for the best selection response. The results show that single trait selections for THT_2008 and FL based on 13 and seven positive alleles were promising for the former but not for FL (Figure 3). The FL variation did not significantly deviate from the population mean (534.53 cm) regardless of how many positive alleles were carried; thus, the trait was omitted for multiple locus selection. Using the same thresholds, we identified only 1% of the GM3341 family with low THT and high O/P to be selected for progeny testing with DAV *pisifera* palms. In a multiple trait integration program in oil palm, this probability of success makes it is close to "impossible" to break the linkage drag, especially if the current breeding approach of using a limited family size (usually not more than 96 palms assessed per cross) remains unchanged. Before field planting, breeders can plant all 1000–1500 seedlings per cross in the nursery and select for the best 100 palms using the markers derived from this research, for field planting for the next breeding cycle. A 1% selection would provide 10–15 palms containing the desired alleles and multiple bunches of the same cross would provide large enough families to field select for additional variation beyond the selected loci.

5. Conclusions

The reduced THT and FL traits from GM and DAV origins have been determined, and show great potential for selective breeding. Further understanding of the correlations between agronomic traits and the effects of the cumulative positive alleles of QTLs pave the way for breeders to conduct multiple trait integration through MAS programs, particularly to produce high-yielding short or compact oil palms. With a 1% probability of breaking linkage drag, future trials need to be saturated with elite palms selected from a large starting population. However, non-linear THT growth across developmental stages should also be taken cautiously, depending on the breeding objectives.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/7/826/s1, Table S1: The 655 selected markers (588 SNPs and 67 SSRs) for linkage map construction. Table S2: Sequences of 67 SSR primer pairs. Table S3: Descriptive statistics of oil yield, frond length and trunk height traits. Table S4: Person correlation among the 16 traits of the GM-3341 family. Table S5: Kruskal-Wallis analyses of trait-affecting markers. All SNPs of the OP200K genotyping array have been deposited in dbSNP under the handle of SDTC_BB with NCBI submitted SNP (ss) accession numbers of 1810069240-1810592638.

Author Contributions: Conceptualization, C.-K.T., S.M. and F.M.; investigation, C.-K.T., A.-L.O. and S.M.; manuscript writing, C.-K.T. and A.-L.O.; writing-review and editing, S.M., F.M. and D.R.A.; supervision, F.M., S.M. and D.A.; funding acquisition, D.R.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was fully supported by internal grants from Sime Darby Plantation, Malaysia.

Acknowledgments: We would like to acknowledge the contribution of the Oil Palm Breeding Section, Sime Darby Plantation R&D Centre for conducting the breeding trials and providing the oil palm materials. We also would like to thank the Sime Darby Plantation, Malaysia for funding and the University of Nottingham Malaysia for the PhD fee scholarship to C.-K.T. and A.-L.O.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. OECD/FAO. *OECD-FAO Agricultural Outlook 2019–2028*; OECD Publishing: Paris, France; Food and Agriculture Organization of the United Nations: Rome, Italy, 2019.
2. Sayer, J.; Ghazoul, J.; Nelson, P.; Klintuni Boedhihartono, A. Oil palm expansion transforms tropical landscapes and livelihoods. *Glob. Food Secur.* **2012**, *1*, 114–119. [CrossRef]
3. ISTA Mielke. *Oil World*; ISTA Mielke: Hamburg, Germany, 2018.
4. Hanim, A. Help for palm oil industry. *Star* **2012**. Available online: https://www.thestar.com.my/business/business-news/2012/05/15/help-for-palm-oil-industry (accessed on 5 May 2020).
5. Diana, R. Manpower still a problem in oil palm industry. *Star* **2013**. Available online: https://www.thestar.com.my/news/community/2013/07/24/manpower-still-a-problem-more-locals-needed-to-mitigate-labour-shortage-in-oil-palm-industry/ (accessed on 5 May 2020).
6. Le Guan, V.; Ouattara, S.; Jacquemard, J.C. Oil palm selection with a view to easier harvesting. Initial results. *Oléagineux* **1990**, *45*, 523–531.
7. Barcelos, E.; Rios, S.d.A.; Cunha, R.N.V.; Lopes, R.; Motoike, S.Y.; Babiychuk, E.; Skirycz, A.; Kushnir, S. Oil palm natural diversity and the potential for yield improvement. *Front. Plant Sci.* **2015**, *6*, 190. [CrossRef] [PubMed]
8. Jagoe, R.B. The *dumpy* oil palm *Malay. Agric. J.* **1952**, *35*, 12.
9. Soh, A.C.; Vanialingam, T.; Taniputra, B.; Pamin, K. Derivatives of the *dumpy* palm—Some experimental results. *Plant* **1981**, *57*, 227–239.
10. Nair, K.P.P. Oil Palm (Elaeis guineensis Jacquin). In *The Agronomy and Economy of Important Tree Crops of the Developing World*; Nair, K.P.P., Ed.; Elsevier: London, UK, 2010; pp. 209–236. [CrossRef]
11. Lee, C.H.; Pang, T.Y. Breeding for short height increment in oil palm. In Proceedings of the International Symposium of Oil Palm Genetic Resources and Utilization, Kuala Lumpur, Malaysia, 8–10 June 2000; p. 12.
12. Kwong, Q.B.; Teh, C.K.; Ong, A.L.; Heng, H.Y.; Lee, H.L.; Mohaimi, M.; Low, J.Z.; Sukganah, A.; Chew, F.T.; Mayes, S.; et al. Development and validation of a high-density SNP genotyping array for african oil palm. *Mol. Plant* **2016**, *9*, 1132–1141. [CrossRef]
13. Corley, R.H.V.; Tinker, P.B. Selection and breeding. In *The Oil Palm*, 3rd ed.; Blackwell: Oxford, UK, 2003; pp. 133–187.
14. Blaak, G.; Sparnaaij, L.D.; Menendez, T. Methods of bunch analysis. In *Breeding and Inheritance in the Oil Palm (Elaeis Guineensis Jacq.) Part II*; Springer: Berlin, Germany, 1963; Volume 4, pp. 146–155.
15. Rao, V.; Soh, A.C.; Corley, R.H.V.; Lee, C.H.; Rajanaidu, N.; Tan, Y.P.; Chin, C.W.; Lim, K.C.; Tan, S.T.; Lee, T.P.; et al. A critical reexamination of the method of bunch analysis in oil palm breeding. *Palm Oil Res. Inst. Malays. Occ. Pap.* **1983**, *9*, 1–28.
16. Corley, R.H.V.; Hardon, J.J.; Tan, G.Y. Analysis of growth of the oil palm (Elaeis guineensis Jacq.) I. Estimation of growth parameters and application in breeding. *Euphytica* **1971**, *20*, 307–315. [CrossRef]
17. Minitab, Inc. *MinitabT7*; Computer Software; Minitab, Inc.: State College, PA, USA, 2016.
18. Blair, M.W.; Hedetale, V.; McCouch, S.R. Fluorescent-labeled microsatellite panels useful for detecting allelic diversity in cultivated rice (Oryza sativa L.). *Theor. Appl. Genet.* **2002**, *105*, 449–457. [CrossRef]
19. Rastas, P.; Paulin, L.; Hanski, I.; Lehtonen, R.; Auvinen, P. Lep-MAP. Fast and accurate linkage map construction for large SNP datasets. *Bioinformatics* **2013**, *29*, 3128–3134. [CrossRef]
20. Rastas, P. Lep-MAP3: Robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* **2017**, *33*, 3726–3732. [CrossRef]
21. Van Ooijen, J.W. *Software for the Calculation of Genetic Linkage Maps in Experimental Populations of Diploid Species*; Kyazma, B.V.: Wageningen, The Netherlands, 2018.
22. Voorrips, R.E. MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **2002**, *93*, 77–78. [CrossRef]
23. Singh, R.; Ong-Abdullah, M.; Low, E.T.L.; Manaf, M.A.A.; Rosli, R.; Nookiah, R.; Ooi, L.C.L.; Ooi, S.E.; Chan, K.L.; Halim, M.A.; et al. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* **2013**, *500*, 335–339. [CrossRef]

24. Van Ooijen, J.W. *MapQTL 6, Software for the Mapping of Quantitative Trait Loci in Experimental Populations of Diploid Species*; Kyazma, B.V.: Wageningen, The Netherlands, 2009.
25. Pamin, K.A. A hundred and fifty years of oil palm in Indonesia: From the Bogor Botanical Garden to the industry. In Proceedings of the International Oil Palm Conference 'Commodity of the Past, Today and the Future', Bali, Indonesia, 23–25 September 1998; pp. 3–23.
26. Kushairi, A.; Rajanaidu, N. Breeding populations, seed production and nursery management. In *Advances in Oil Palm Research*; Basiron, Y., Jalani, B.S., Chan, K.W., Eds.; Malaysian Palm Oil Board: Kuala Lumpur, Malaysia, 2000; Volume 1, pp. 39–98.
27. Rosenquist, E.A. The genetic base of oil palm breeding populations. In Proceedings of the International Workshop on Oil Palm Germplasm and Utilization, Bangi, Malaysia, 26–27 March 1985; pp. 27–56.
28. Beirnaert, A.; Vanderweyen, R. Contribution à l'étude génétique et biométrique des variétiés d'Elaeis guineensis Jacq. *Publ. Inst. Nat. Etude Agron. Congo Belge Ser. Sci.* **1941**, *3*, 95–110.
29. Hardon, J.J.; Corley, R.H.V.; Lee, C.H. *Breeding and Selecting the Oil Palm*; Academic Press: London, UK, 1987; pp. 63–81.
30. Mohd Basri, W.; Norman, K. Role and effectiveness of Elaeidobius kamerunicus, Thrips hawaiiensis and Pyroderces sp. in pollination of mature oil palm in Peninsular Malaysia. *J. Oil Palm Res.* **1997**, *9*, 1–16.
31. Cik Mohd Rizuan, Z.A.; Noor Hisham, H.; Samsudin, A. Role of pollinating weevil (Elaeidobius kamerunicus), seasonal effect and its relation to fruit set in oil palm area of FELDA. In Proceedings of the PIPOC 2013 Conference, Kuala Lumpur, Malaysia, 19–21 November 2013.
32. Henson, I.E. Measuring and modelling oil palm trunk growth. *Oil Palm Bull.* **2006**, *52*, 1–24.
33. Jacquemard, J.C. Contribution to the study of the height growth of the stems of Elaeis guineensis Jacq. Study of the L2T × D10D cross. *Oléagineux* **1979**, *34*, 492–497.
34. Corley, R.H.V.; Tinker, P.B. Growth, flowring and yield. In *The Oil Palm*, 4th ed.; Blackwell Science Ltd.: Oxford, UK, 2016; pp. 89–134.
35. Kwan, B.K.W. The effect of planting density on the first fifteen years of growth and yield of oil palm in Sabah. *Tech. Bull.* **1994**, *11*, 26.
36. Henson, I.E.; Mohd Tayeb, D. Physiological analysis of an oil palm density trial on a peat soil. *J. Oil Palm Res.* **2003**, *15*, 1–27.
37. Weraduwage, S.M.; Chen, J.; Anozie, F.C.; Morales, A.; Weise, S.E.; Sharkey, T.D. The relationship between leaf area growth and biomass accumulation in Arabidopsis thaliana. *Front. Plant Sci.* **2015**, *6*. [CrossRef]
38. Henson, I.E. Photosynthesis and source-sink relationship in oil palm. *Trans. Malays. Soc. Plant Physiol.* **1990**, *1*, 165–171.
39. Henson, I.E. Age-related changes in stomatal and photosynthetic characteristics of leaves of oil palm (Elaeis guineensis Jacq.). *Elaeis* **1991**, *4*, 51–59.
40. Ong, A.-L.; Teh, C.-K.; Kwong, Q.-B.; Tangaya, P.; Appleton, D.R.; Massawe, F.; Mayes, S. Linkage-based genome assembly improvement of oil palm (Elaeis guineensis). *Sci. Rep.* **2019**, *9*, 6619. [CrossRef]
41. Argyris, J.M.; Ruiz-Herrera, A.; Madriz-Masis, P.; Sanseverino, W.; Morata, J.; Pujol, M.; Ramos-Onsins, S.E.; Garcia-Mas, J. Use of targeted SNP selection for an improved anchoring of the melon (*Cucumis melo* L.) scaffold genome assembly. *BMC Genom.* **2015**, *16*, 4. [CrossRef]
42. Mathew, L.S.; Spannagl, M.; Al-Malki, A.; George, B.; Torres, M.F.; Al-Dous, E.K.; Al-Azwani, E.K.; Hussein, E.; Mathew, S.; Mayer, K.F.X.; et al. A first genetic map of date palm (*Phoenix dactylifera*) reveals long-range genome structure conservation in the palms. *BMC Genom.* **2014**, *15*, 285. [CrossRef]
43. Xu, Z.; Zou, F.; Vision, T.J. Improving quantitative trait loci mapping resolution in experimental crosses by the use of genotypically selected samples. *Genetics* **2005**, *170*, 401–408. [CrossRef]
44. Balyan, H.S.; Singh, O. Pleiotropic effects of GA-insensitive Rht Genes on grain yield and its component characters in wheat. *Cereal Res. Commun.* **1994**, *22*, 195–200.
45. Jobson, E.M.; Johnston, R.E.; Oiestad, A.J.; Martin, J.M.; Giroux, M.J. The Impact of the wheat Rht-B1b semi-dwarfing allele on photosynthesis and seed development under field conditions. *Front. Plant Sci.* **2019**, *10*. [CrossRef]

46. Pootakham, W.; Jomchai, N.; Ruang-areerate, P.; Shearman, J.R.; Sonthirod, C.; Sangsrakru, D.; Tragoonrung, S.; Tangphatsornruang, S. Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics* **2015**, *105*, 288–295. [CrossRef]
47. Lee, M.; Xia, J.H.; Zou, Z.; Ye, J.; Rahmadsyah; Alfiko, Y.; Jin, J.; Lieando, J.V.; Purnamasari, M.I.; Lim, C.H.; et al. A consensus linkage map of oil palm and a major QTL for stem height. *Sci. Rep.* **2015**, *5*, 8232. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Chapter 5: GENERAL CONCLUSIONS

An overview of key findings and milestones achieved within and across chapters is visualised in Figure 5. Some key messages are emphasized below:

1. **Breaking the bottleneck**: Conventional oil palm breeding is progressing at 'a snail's pace' compared to annual crops, mainly due to a lengthy breeding cycle and expensive progeny testing. To overcome this bottleneck, the industry begun exploring MAS in oil palm breeding in the 1990s. Since then, DNA marker development has grown enormously, but most of the markers developed rarely make it past the discovery phase into MAS deployed in a breeding programme. Historically, low genotyping throughput has only permitted limited QTL discovery and deployment, which compromises the predictive ability for traits of interest, making the MAS commercially unattractive to the industry. However, this limitation does not hold true anymore given the recent rapid advancement in genomics and DNA sequencing technologies. High throughput and accurate DNA sequencing have built a strong foundation to maximise the predictability for a trait based on the whole-genome assessment. The oil palm industry indeed has deployed single-trait MAS (or GS) to their breeding programmes and commercial seed production in oil palm. The result is promising, but the main interest still remains on oil yield-related traits.

2. **Do-it-right**: Illegitimacy can cause significant losses in oil yield potentially and breeding efforts (also effecting QTL mapping), but the industry usually chooses to ignore the problem. This happens because the conventional fruit census is not accurate in detecting the genetic contamination and consistently underestimates the seriousness. In this study, a genomic SNP-based legitimacy test was successfully developed and has

been deployed as a standard quality control tool in oil palm breeding and commercial DxP seed production for Sime Darby Plantations. Without the access to the parental palms, farmers can still use the SHELL test to eliminate the *dura* contamination. For commercial seed producers, the legitimacy testing method is highly recommended to detect the *tenera* contamination which is due to errors or mixed pollen sources which cannot be detected using the SHELL test and conventional fruit census.

3. **Model for expansion**: The rapid change in environment and market will require oil palm breeding to be more agile and resilient. Hence, oil yield improvement alone will not be sufficient to serve the purpose. In this study, a multiple trait integration using MAS has been developed in a low height increment breeding population (GM x GM). The result clearly reveals the presence of linkage drag between antagonistic traits (O/P against THT and FL) and it constrains multiple-trait selection based on current progeny population size. In this study, the MAS approaches have extended the possibility for breaking the unwanted linkage drags, even where the desired genotype is only expected to occur at frequency of 1% in a cross, through genotyping at the seedling stage of all progeny (generally 1,000 – 1,500 per cross).

By linking all three chapters, this research provides an important framework/model to translate fundamental plant science from genetic discoveries in the lab to field application in MAS deployment and to further expand the MAS for multiple trait integration. With these efforts, the oil palm breeding programmes will be more agile and resilient to the rapid change of environment (e.g. climate change) and market (e.g. labour shortage). In this case, high-yielding and low THT planting material apparently will be one of the game changers for the oil palm industry in the near future.

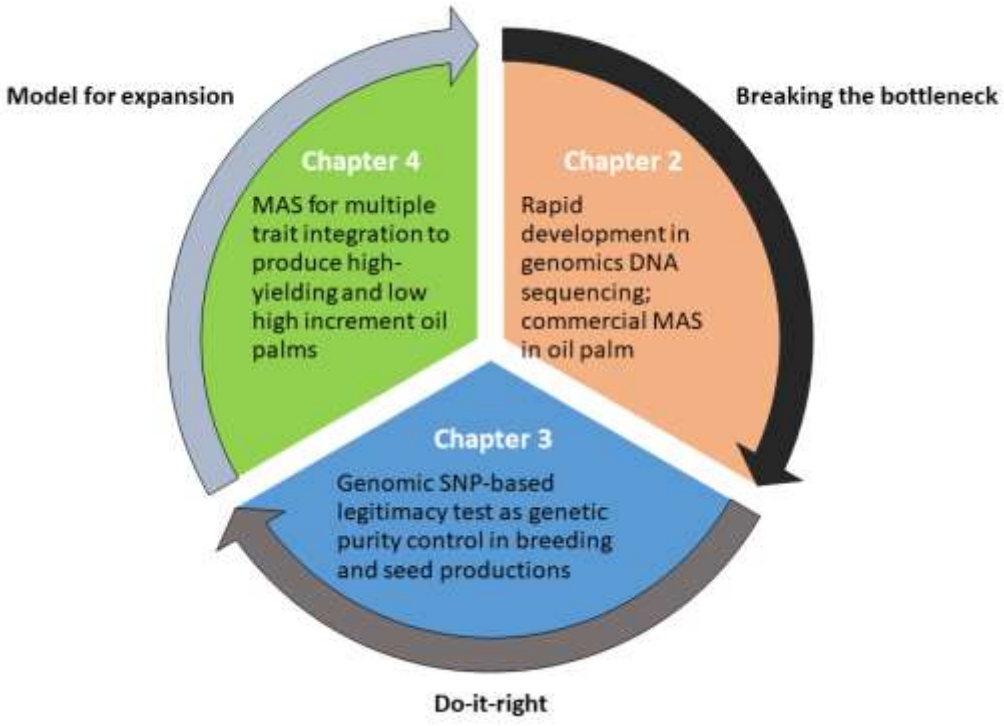


Figure 5: An overview of key findings and milestones achieved within and across thematic areas in this research programme

References

Amblard, P., Noiret, J.M., Kouamé, B., Potier, F., and Adon, B. (1995). Performances comparées des hybrides interspécifiques et du matériel commercial E. guineensis. Oléaginueux Crops gras Lipides 2:335-340.

Barcelos, E., Rios, S.d.A., Cunha, R.N.V., Lopes, R., Motoike, S.Y., Babiychuk, E., Skirycz, A., and Kushnir, S. (2015). Oil palm natural diversity and the potential for yield improvement. Frontiers in Plant Science 6:190-190.

Beirnaert, A. (1935). Introduction à la biologie florale du palmier à huile (Elasie guineensis Jacquin) Congo Belge, Ser. Sci.: Inst. Nat. Etude agron. 3-42.

Beirnaert, A., and Vanderweyen, R. (1941). Contribution à l'étude génétique et biométrique des variétiés d'Elaeis guineensis Jacq. In: Publ. Inst. Nat. Etude Agron. Congo Belge. Ser. Sci. . 1-101.

Billotte, N., Jourjon, M.F., Marseillac, N., Berger, A., Flori, A., Asmady, H., Adon, B., Singh, R., Nouy, B., Potier, F., et al. (2010). QTL detection by multi-parent linkage mapping in oil palm (Elaeis guineensis Jacq.). Theoretical and Applied Genetics 120:1673-1687.

Blaak, G., Sparnaaij, L.D., and Menendez, T. (1963). Methods of bunch analysis. In: Breeding and inheritance in the oil palm (*Elaeis guineensis* Jacq.) Part II: J.W. Afr. Ins. Oil Palm Res. 146 - 155.

Corley, R.H.V. (1979). Palm oil composition and oil palm breeding. Planter 55:467-478.

Corley, R.H.V. (2005). Illegitimacy in oil palm breeding - a review. Journal of Oil Palm Research 17:64-69.

Corley, R.H.V., Hardon, J.J., and Tan, G.Y. (1971). Analysis of growth of the oil palm (Elaeis guineensis Jacq.) I. Estimation of growth parameters and application in breeding. Euphytica 20:307-315.

Corley, R.H.V., and Tinker, P. (2003). Selection and breeding. In: The oil palm Oxford: Blackwell Science Ltd Blackwell Publishing.

Cros, D., Denis, M., Bouvet, J.-M., and Sánchez, L. (2015a). Long-term genomic selection for heterosis without dominance in multiplicative traits: case study of bunch production in oil palm. BMC Genomics 16:1-17.

Cros, D., Denis, M., Sánchez, L., Cochard, B., Flori, A., Durand-Gasselin, T., Nouy, B., Omoré, A., Pomiès, V., Riou, V., et al. (2015b). Genomic selection prediction accuracy in a perennial crop: case study of oil palm (Elaeis guineensis Jacq.). Theoretical and Applied Genetics 128:397-410.

Dekkers, J.C.M. (2004). Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J. Anim. Sci. 82:E-Suppl:E313-E328.

Diana, R. (2013). Manpower still a problem in oil palm industry. In: The Star: The Star Online.

Feuillet, C., Leach, J.E., Rogers, J., Schnable, P.S., and Eversole, K. (2011). Crop genome sequencing: lessons and rationales. Trends in Plant Science 16:77-88.

Goddard, M.E., and Hayes, B.J. (2007). Genomic selection. Journal of Animal Breeding and Genetics 124:323-330.

Hadi, S., Ahmad, D., and Akande, F.B. (2009). Determination of the bruise indexes of oil palm fruits. Journal of Food Engineering 95:322-326.

Hanim, A. (2012). Help for palm oil industry. In: The Star: The Star Online.

Hardon, J.J. (1969). Interspecific hybrids in the genus Elaeis II. vegetative growth and yield of F1 hybrids E. guineensis x E. oleifera. Euphytica 18:380-388.

Hardon, J.J. (1970). Inbreeding in populations of the oil palm (Elaeis guineensis Jacq.) and its effects on selection. Oleagineux 25:449-456.

Hardon, J.J., Corley, R.H.V., and Lee, C.H. (1987). Breeding and selecting the oil palm. London: Academic Press.

Hartley, C.W.S. (1967). Factors affecting growth, flowering and yield. In: The Oil Palm London: Longman. 137 - 193.

Henson, I.E. (2006). Measuring and modelling oil palm trunk growth. Oil Palm Bulletin 52:1-24.

Jagoe, R.B. (1952). The *dumpy* oil palm Malay. Agric. J. 35:12.

Kwong, Q.B., Teh, C.K., Ong, A.L., Heng, H.Y., Lee, H.L., Mohaimi, M., Low, J.Z., Sukganah, A., Chew, F.T., Mayes, S., et al. (2016). Development and Validation of a High-Density SNP Genotyping Array for African Oil Palm. Molecular Plant 9:1132-1141.

Lee, C.H., and Pang, T.Y. (2000). Breeding for short height increment in oil palm. In: International Symposium of Oil Palm Genetic Resources and Utilization Kuala Lumpur. 12.

Mohd, D., Rajanaidu, N., Kushairi, A., Mohd Rafii, Y., Mohd Isa, Z.A., and Noh, A. (2002). PS4-high carotene *E. oleifera* planting materials. MPOB Information Series 154.

Montoya, C., Lopes, R., Flori, A., Cros, D., Cuellar, T., Summo, M., Espeout, S., Rivallan, R., Risterucci, A.M., Bittencourt, D., et al. (2013). Quantitative trait loci (QTLs) analysis of palm oil fatty acid composition in an interspecific pseudo-backcross from Elaeis oleifera (H.B.K.) Cortés and oil palm (Elaeis guineensis Jacq.). Tree Genetics & Genomes 9:1207-1225.

OilWorld. (2018). Oil World. Hamburg: ISTA Mielke.

Ong, A.-L., Teh, C.-K., Kwong, Q.-B., Tangaya, P., Appleton, D.R., Massawe, F., and Mayes, S. (2019). Linkage-based genome assembly improvement of oil palm (Elaeis guineensis). Scientific Reports 9:6619.

Ong, S.H., Chuah, C.C., and Sow, H.P. (1981). The co-dominance theory: genetic interpretations of analyses of mesocarp oils from Elaeis guineensis, Elaeis oleifera and their hybrids. Journal of the American Oil Chemists' Society 58:1032-1038.

Ooi, L.C.-L., Low, E.-T.L., Abdullah, M.O., Nookiah, R., Ting, N.C., Nagappan, J., Manaf, M.A.A., Chan, K.-L., Halim, M.A., Azizi, N., et al. (2016). Non-tenera Contamination and the Economic Impact of SHELL Genetic Testing in the Malaysian Independent Oil Palm Industry. Frontiers in Plant Science 7.

Pamin, K.A. (1998). A hundred and fifty years of oil palm in Indonesia: from the Bogor Botanical Garden to the industry. . In: International Oil Palm Conference 'Commodity of the Past, Today and the Future'. 3-23.

Rajanaidu, N., Rao, V., H., H., and Ong, A.S.H. (1989). Genetic resources – new developments in oil palm breeding. Elaeis 1:1-10.

Rao, V., Soh, A.C., Corley, R.H.V., Lee, C.H., Rajanaidu, N., Tan, Y.P., Chin, C.W., Lim, K.C., Tan, S.T., Lee, T.P., et al. (1983). A critical reexamination of the method of bunch analysis in oil palm breeding. Palm Oil Research Institute Malaysia Occ Paper 9:1 - 28.

Rosenquist, E.A. (1986). The genetic base of oil palm breeding populations. In: The International Workshop on Oil Palm Germplasm and Ultilization. 27-56.

Sharman, M. (1999). Utilization of Nigerian PS1 and PS2 selection in oil palm breeding programmes at UP Bhd. In: Seminar PS1 and PS2 Oil Palm Planting Materials--Rajanaidu, N., and Jalani, B.S., eds. Kuala Lumpur: Palm Oil Research Institute Malaysia.

Singh, R., Low, E.-T.L., Ooi, L.C.-L., Ong-Abdullah, M., Ting, N.-C., Nagappan, J., Nookiah, R., Amiruddin, M.D., Rosli, R., Manaf, M.A.A., et al. (2013a). The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. Nature 500:340-344.

Singh, R., Ong-Abdullah, M., Low, E.T.L., Manaf, M.A.A., Rosli, R., Nookiah, R., Ooi, L.C.L., Ooi, S.E., Chan, K.L., Halim, M.A., et al. (2013b). Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. Nature 500:335-339.

Soh, A.C., Vanialingam, T., Taniputra, B., Pamin, K. (1981). Derivatives of the *dumpy* palm - some experimental results. The Planter 57:227 - 239.

Teh, C.-K., Ong, A.-L., Kwong, Q.-B., Apparow, S., Chew, F.-T., Mayes, S., Mohamed, M., Appleton, D., and Kulaveerasingam, H. (2016). Genome-wide association study identifies three key loci for high mesocarp oil content in perennial crop oil palm. Scientific Reports 6:19075.

Teh, C.-K., Ong, A.-L., Mayes, S., Massawe, F., and Appleton, D.R. (2020). Major QTLs for Trunk Height and Correlated Agronomic Traits Provide Insights into Multiple Trait Integration in Oil Palm Breeding. 11:826.

Teh, C.K., Lee, H.L., Hafiza, A., ONG, A.L., Mayes, S., CHEW, F.T., and Appleton, D. (2019 (Accepted)). A practical genome-enabled legitimacy assay for oil palm breeding and seed production. BMC Plant Biology.

Van Ooijen, J.W. (2009). MapQTL 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species Wageningen, Netherlands: Kyazma B. V.

Yap, S.C., Choo, Y.M., Hew, N.F., and Goh, S.H. (1997). Distribution of dietary palm carotenes and their metabolites in the rabbit. Nutrition Research 17:1721-1731.

Young, N.D., and Tanksley, S.D. (1989). RFLP analysis of the size of chromosomal segments retained around the Tm-2 locus of tomato during backcross breeding. Theor Appl Genet 77:353-359.