UNEDITED

PROCEEDINGS OF THE

International Seminar on 100 YEARS OF TECHNOLOGICAL ADVANCEMENT in Oil Palm Breeding & Seed Production

Monday, 13 November 2017
Kuala Lumpur Convention Centre (KLCC),
Kuala Lumpur, Malaysia

Jointly organised by:

ISOPB

MPOB

www.isopb.org www.mpob.gov.my

UNEDITED

PROCEEDINGS OF THE

International Seminar on
100 YEARS OF TECHNOLOGICAL ADVANCEMENT
in Oil Palm Breeding & Seed Production

Jointly organised by:

ISOPB

MPOB

www.isopb.org www.mpob.gov.my

Abbre. Title: Proc. Int. Sem. 100 Years of Tech. Advan. in OP Breeding & Seed Production

Malaysian Palm Oil Board (MPOB), 2017

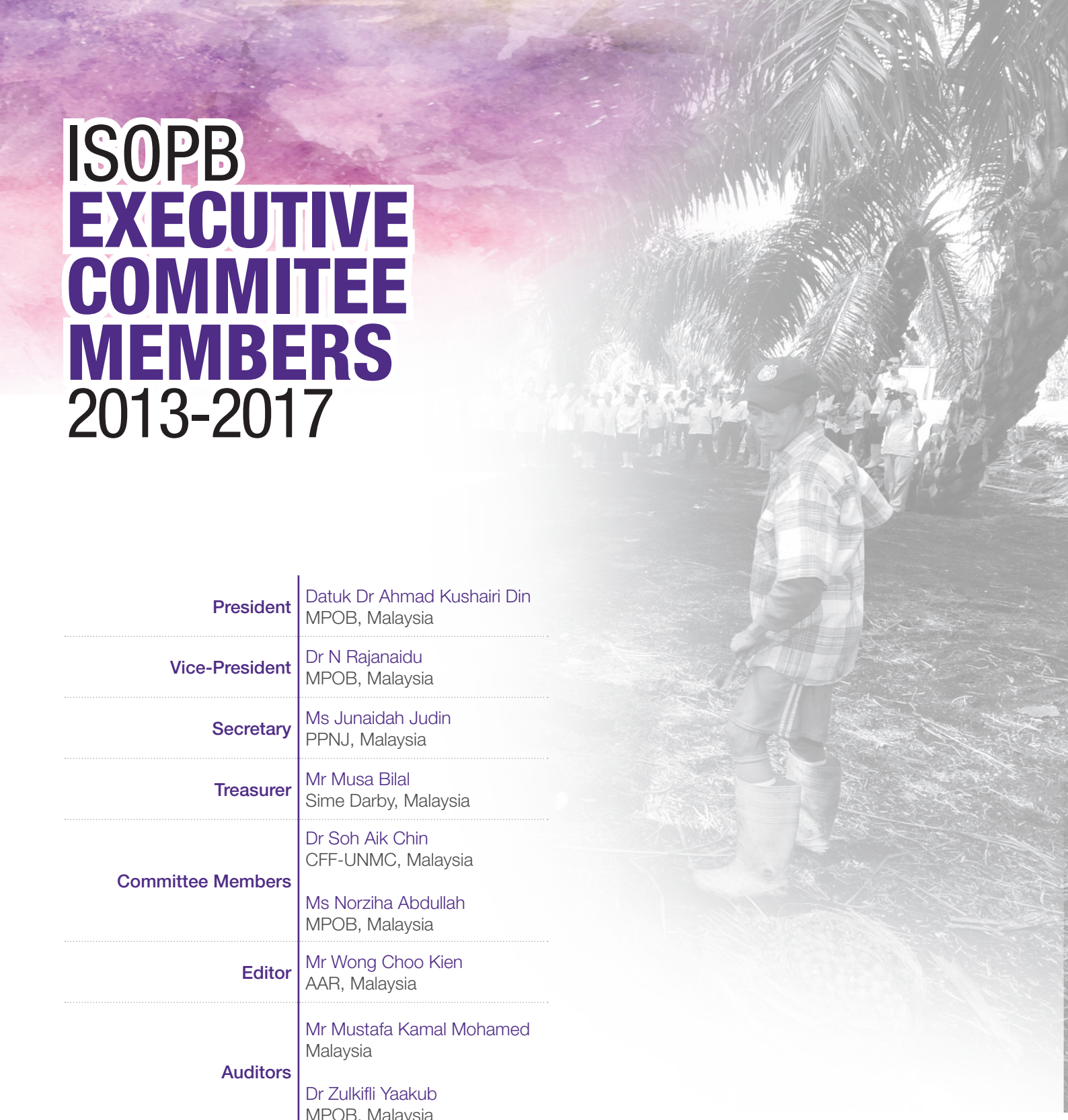
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of the publisher.

Published in 2017 by the Malaysian Palm Oil Board.

SEMINAR ORGANISING COMMITTEE

Advisor	Datuk Dr Ahmad Kushairi Din MPOB, Malaysia
Chairperson	Dr N Rajanaidu MPOB, Malaysia
Co-Chairperson	Dr Ahmad Parveez Ghulam Kadir MPOB, Malaysia
Secretary I	Dr Zulkifli Yaakub MPOB, Malaysia
Secretary II	Mr Mustafa Kamal Mohamed Malaysia
Treasurer	Mr Musa Bilal Sime Darby, Malaysia
Members	Mr Wong Choo Kien AAR, Malaysia Dr Tan Joon Sheong FGV, Malaysia Ms Norziha Abdullah MPOB, Malaysia
Secretariat	Ms Ruba'ah Masri MPOB, Malaysia Ms Rozana Abu Bakar MPOB, Malaysia Ms Norhayati Md Fahimy MPOB, Malaysia Ms Azleha Dollah MPOB, Malaysia Mr Norhafizal Abdul Halim MPOB, Malaysia

ISOPB EXECUTIVE COMMITEE MEMBERS 2013-2017



President	Datuk Dr Ahmad Kushairi Din MPOB, Malaysia
Vice-President	Dr N Rajanaidu MPOB, Malaysia
Secretary	Ms Junaidah Judin PPNJ, Malaysia
Treasurer	Mr Musa Bilal Sime Darby, Malaysia
Committee Members	Dr Soh Aik Chin CFF-UNMC, Malaysia Ms Norziha Abdullah MPOB, Malaysia
Editor	Mr Wong Choo Kien AAR, Malaysia
Auditors	Mr Mustafa Kamal Mohamed Malaysia Dr Zulkifli Yaakub MPOB, Malaysia
Regional Rep. Asia-Pacific	Dr Ir Abdul Razak Purba IOPRI, Indonesia
Africa	Dr Ir Claude Bakoume Cameroon
Europe	Dr Tristan Durand-Gasselin PalmElit SAS, France
Americas	Dr Hernan M Romero Cenipalma, Colombia

ISOPB Official Web-Site

http://www.isopb.org

CONTENTS

OPENING REMARKS

Datuk Dr Ahmad Kushairi Din

Director-General
Malaysian Palm Oil Board (MPOB) and

President
International Society for Oil Palm Breeders (ISOPB)

Distinguished guests, speakers and participants;

Ladies and gentlemen,

Good Morning.

On behalf of the International Society for Oil Palm Breeders (ISOPB) and the Malaysian Palm Oil Board (MPOB), I wish a warm welcome to our distinguished guests, speakers, members of the society and participants to our annual ISOPB event, the International Seminar on 100 Years of Technological Advancement in Oil Palm Breeding and Seed Production.

While the Malaysian oil palm industry is celebrating a century of success in oil palm cultivation, various remembrance gestures were observed, be it national level or society level. At the national level, these include launching of stamps and philately items as well as commemorative coins; while at the society level, seminars are commonly observed which include our ISOPB seminar today which carries the word 100 years in its title.

Ladies and gentlemen,

The first Malaysian commercial oil palm plantation, established in 1917, is located at the Tennamaram Estate in Bestari Jaya, Selangor using seeds from Rantau Panjang, Kuala Selangor which marked the foundation of the oil palm industry in Malaysia.

Seeds or planting materials, in general, are critical to such an event. Without seeds, there would be no plantation. Hence, the oil palm breeders' role, to improve genetic yield potential of oil palm, is critical to the success of the industry.

Oil palm breeders, working with a perennial crop, which had a long breeding cycle, indeed, are facing great challenges. They need to be technically strong and somewhat 'psychic' to predict what trait is important in future. At the same time, they need to be equipped with soft skills like patience to observe the fruits of their labour.

Ladies and gentlemen,

Today's seminar, we have gathered oil palm breeders from various countries and organisations, to share their knowledge with us, typically towards the latest technologies they have deploy to assist in their oil palm breeding research and seed production.

Here, you will experience a range of tools and equipments deployed, which include tools based on physics principles, statistics, information technology, molecular genetics and genomics, and the latest, genome-editing.

For genome-editing, we are fortunate to have Prof Kai-Jun Zhao from Chinese Academy of Agriculture Sciences, Beijing, China, to share with us this technology, which had created excitement among breeders working on annual crop. Do interact with Prof Kai-Jun for those of you interested to learn more on Genome Editing as he will be joining us for PIPOC 2017.

In keeping abreast in scientific knowledge and advancement, it is important that breeders should examine, what is available out there, in order to sharpen our tools for oil palm improvement so that the industry could continue to advance significantly, striving for many more 100 years of oil palm excellence in Malaysia and in the world.

A monograph about oil palm germplasm, describing about almost half a century long forward-looking efforts of Malaysian oil palm breeders, with credit goes to Dr Rajanaidu, will be launched tomorrow for all to treasure and the most important to benefit from it.

Ladies and gentlemen,

Last but not least, I would like to thank a few groups of people: the speakers for sparing their time to share their knowledge with us; the organising committee and the secretariat, for ensuring all aspects of the seminar is in orderly manner, and most importantly, to all participants who had come near or far, making this seminar possible and a successful one.

I now officially declare this ISOPB International Seminar on 100 Years of Technological Advancement in Oil Palm Breeding and Seed Production officially open.

Thank you.

KEYNOTE ADDRESS

Genome Editing Technologies and Their Potential Applications in Crop Improvement

Rukmini Mishra[1] and Kai-Jun Zhao[1#]

ABSTRACT

Crop improvement is very essential to meet the increasing global food demands and enhance food nutrition. Conventional crop breeding methods like random mutation, cross breeding methods and use of genetically modified (GM) crops have contributed a lot towards world food security. But they have certain limitations such as taking lot of time and resources, and causing biosafety concerns. These limitations could be overcome by the recently emerged genome editing tools that can precisely modify DNA sequences at the genomic level by using sequence specific nucleases (SSNs). The artificially engineered SSNs such as Zinc finger nucleases (ZFNs), transcriptional activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease Cas9 (CRISPR/Cas9) have proven to be highly effective for genome editing in a wide variety of organisms including plants. The CRISPR/Cas9 system is the most recently developed targeted genome modification system and seems to be more efficient, inexpensive, easy, user friendly and rapidly adopted genome editing tool. Large-scale genome editing has not only improved the yield and quality but also has enhanced the disease resistance ability in several model and other major crops. Increasing case-studies suggest that genome editing is an efficient, precise and powerful technology that can accelerate basic and applied research towards crop improvement. In this review, we briefly overviewed the structure and mechanism of genome editing tools and their potential applications in crop plants. We have also discussed the broad applications of the targeted genome editing tools and their future prospects towards the improvement of agronomic traits in crops.

[1]National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI), Institute of Crop Science, Chinese Academy of Agriculture Sciences (CAAS), Beijing 100081, China.
[#] Corresponding author: Kaijun Zhao (zhaokaijun@caas.cn)

SESSION 1

Agrigenomics in the Breeder's Toolbox: Latest Advances towards an Optimal Implementation of Genomic Selection in Oil Palm

Jacob, F[1]; Cros, D[2]; Cochard, B[1] and Durand-Gasselin, T[1]

ABSTRACT

PalmElit implements the genetic improvement and marketing programs for CIRAD® oil palm seeds. The commercial seeds embody 80 years of genetic improvement work undertaken by IRHO, CIRAD and PalmElit in conjunction with several partners of excellence located on each of the continents where oil palm is grown. An increase of more than 60% in oil yields was achieved since 1960. This result illustrates the efficiency of the recurrent reciprocal selection (RRS) underlying the conducted breeding program. So far, assessment of parental breeding values has largely relied on progeny testing, which is an efficient but time- and money-consuming step within the RRS scheme. With the recent development of oil palm genomic resources, genomic selection (GS) appears as an attractive strategy to increase the efficiency of oil palm breeding programs. On a theoretical point of view, GS has the potential to increase the rate of genetic gain by shortening the breeding cycle and/or increasing the selection intensity.

PalmElit, together with its research partner CIRAD, has been leading research for nearly 10 years in order to develop and assess the implementation of GS in oil palm breeding. Some of the key achievements have been shared with the scientific community since 2015 (Cros et al., 2015a, 2015b, 2017a; Marchal et al., 2016) which corroborate the potential of GS in terms of increased genetic gain. Further research is still ongoing to answer the simple -but critical- question: what is the optimal use of GS in terms of genetic gain vs time- and cost-efficiency? In this paper, following a brief review on the GS history and key concepts, we present our latest results which address critical aspects such as prediction accuracy and optimal use of GS within breeding schemes. We extend and discuss our conclusions in light of the literature available in oil palm and other crop species. Finally, we summarize the perspectives and challenges for successful implementation of GS in oil palm.

---

[1]PalmElit S.A.S, 2214 Boulevard de la Lironde, 34980 Montferrier-sur-Lez, France.
[2]CIRAD, UMR AGAP (Genetic Improvement and Adaptation of Mediterranean and Tropical Plants Research Unit), 34398 Montpellier, France.

INTRODUCTION

Pros and cons of the classical recurrent reciprocal selection

Oil palm varieties typically consists in *tenera* hybrid crosses between heterotic group A (mostly Deli origin, *dura* palms) and group B (mostly African origins, *pisifera* palms). Selection and breeding among the parental populations usually relies on progeny testing since hybrid performances might not be accurately predicted based on parental performances (Corley and Tinker, 2015a). In order to achieve an efficient and sustained improvement of its commercial hybrids, PalmElit employs a recurrent reciprocal selection (RRS) strategy for both group A and B parental populations (Baudouin et al., 1997). This strategy aims at improving the general combining abilities (GCA) of the parental population along the successive breeding cycles. Pros and cons of the RRS in oil palm have been already debated (Corley and Tinker, 2015a). According to Gallais (Gallais and Poly, 1990), the main advantages of recurrent selection are:

- increasing the frequency of genes and associations favoring the type of variety to be developed
- enabling effective recombination, hence highly effective multi-trait breeding
- preventing an over-rapid loss of variation, provided it is carried out correctly
- partially fixing heterosis
- ensuring continuous, long-term progress
- providing outputs directly applicable for varietal creation

When RRS is applied in oil palm, one breeding cycle extends over a long period of time (~20 years) in contrast with some annual crops (e.g. 3 months in rice). Despite this long cycling time, a high genetic gain rate has been achieved since 1960 (~+1%/year for yield, Durand-Gasselin et al., 2010), highlighting the potential of oil palm in terms of genetic improvement. The main limitation in terms of cost- and time-efficiency relates to estimation of the parental GCA and hybrid values since it traditionally requires progeny testing for each parent. Thus, techniques allowing faster and/or cheaper GCA or hybrid value estimation could greatly improve oil palm breeding, including RRS.

New tools of the agrigenomics era: marker assisted selection and genomic selection

As more genetic and genomic resources become available for oil palm, new breeding tools become available such as marker assisted selection (MAS, reviewed for crops in Collard and Mackill, 2008). In MAS, molecular marker data can be used to predict phenotype(s), based on known association between the chosen marker(s) and phenotype(s). Marker-phenotype associations can be identified using approaches such as quantitative trait loci (QTL) mapping. In that case, markers linked with the strongest QTLs can be selected and used for predicting the associated phenotype. This selection method can be efficient provided that:

- MAS is faster and/or cheaper than the conventional phenotypic screening
- QTLs are accurately identified (appropriate experimental design to guarantee a high detection power and to limit the risk of false positives and of QTL effect overestimation)
- linkage between markers and QTLs is strong

- association between marker(s) and the phenotype(s) is conserved in the population and the environment where the selection will be carried out
- a limited number of QTLs accounts for a sufficient part of the phenotypic variation (e.g. the trait is essentially mono- or oligogenic)

The latter point defines one major drawback of the classical MAS strategy since many agronomic traits are quantitative and thus likely influenced by a large number of loci. Genomic selection (GS) was developed as a specific case of MAS designed for quantitative traits (Meuwissen et al., 2001). In genomic selection, individuals are genotyped over a dense set of genome-wide markers that can ideally account for all QTLs in the genome. Based on marker data, a genomic estimated breeding value (GEBV, with BV and GCA generally linked by BV = 2*GCA) can be assigned to each genotyped individual provided that the model was calibrated using an appropriate training set (TS) which combines genotypic and phenotypic data for the trait(s) of interest.

GS was first developed and implemented for cattle breeding and has later found its way to plant breeding. Publication trends clearly illustrate the research expansion for GS in plants as of 2009 (Figure 1). Despite the amount of research conducted, and the growing evidence for its potential in hybrid breeding (Marulanda et al., 2016; Zhao et al., 2015), practical implementation of GS has remained limited to a few species including wheat, maize, rye, pines, cassava, and recently oil palm (Cros et al., 2017a; Kwong et al., 2017).

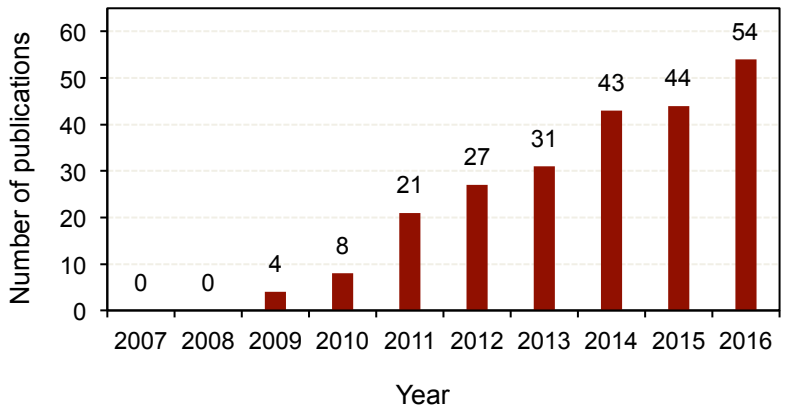


Figure 1: Publication trends for GS in plants. Publications trends were estimated by counting the number of publications in the plant field that contains "genomic selection" in the title and are referenced on Google Scholar.

Potential of GS in oil palm breeding

As evoked earlier, GS could improve many aspects of the oil palm breeding programs:

- the estimation of the value of hybrid crosses which have not been phenotyped. In that respect, GS can directly support the identification and selection of commercial hybrid with higher agronomic value

- the estimation of the GCA of individuals among the germplasm. In that case, GS can assist the process of recombination within the germplasm to increase the genetic value of the parental population
- the duration of the breeding cycle (reduction) by replacing part of, or the entire phenotyping process
- the selection intensity (increase) for both hybrid crosses and parental populations by including individuals for which only genotypic data is available

In the following article, we review the latest results which address critical aspects such as prediction accuracy and optimal use of GS within breeding schemes. We extend and discuss our conclusions in light of the literature available in oil palm and other crop species. Finally, we summarize the perspectives and challenges for successful implementation of GS in oil palm.

CURRENT STATUS FOR GS IN OIL PALM

From classical breeding to genomic selection at PalmElit: past, present, and future

In the past, for CIRAD® germplasm, GCA estimation was based on appropriate statistical analysis of genetic trials with complex experimental designs (e.g. incomplete and unbalanced factorial designs). Lately, we implemented a pedigree-based best linear unbiased prediction (T-BLUP) approach to improve parental GCA estimation in the context of these complex genetic trial designs. By borrowing information from the pedigree (under the form of a kinship matrix), pedigree-based BLUP could also estimate GCA of individuals which are not tested but are related to progeny-tested individuals (P-BLUP, Figure 2A-B). This example illustrates how appropriate statistics can estimate GCA of untested individuals, provided that a suitable training set (TS) is available for calibrating the model.

Study on seven yield components indicated that GCA prediction accuracy using P-BLUP is intermediate to high depending on the trait and the heterotic group considered (ranging from 0.22 to 0.82, Table 1, Cros et al., 2017a). However, this approach requires accurate knowledge of the germplasm pedigree (Corley and Tinker, 2015a) and cannot account for Mendelian sampling. This is illustrated by the fact that pedigree information cannot discriminate individuals within full-sib families although these have distinct genotypes as a result of Mendelian segregation (Figure 2A-B). Thus, P-BLUP-based GCA estimation is not suitable for intra-family selection. To overcome this limitation, we decided to test whether GS could perform better than P-BLUP (Figure 2). For genome-based predictions, we used a similar BLUP model, that we designated as G-BLUP. The G-BLUP method was successfully applied for hybrid prediction in various species, including maize, soybean, rice, triticale and sunflower (Zhao et al., 2015). Moreover, a previous study of Cros et al. indicated that this model performs similarly to several other tested models when applied on empirical oil palm data (Cros et al., 2015a).

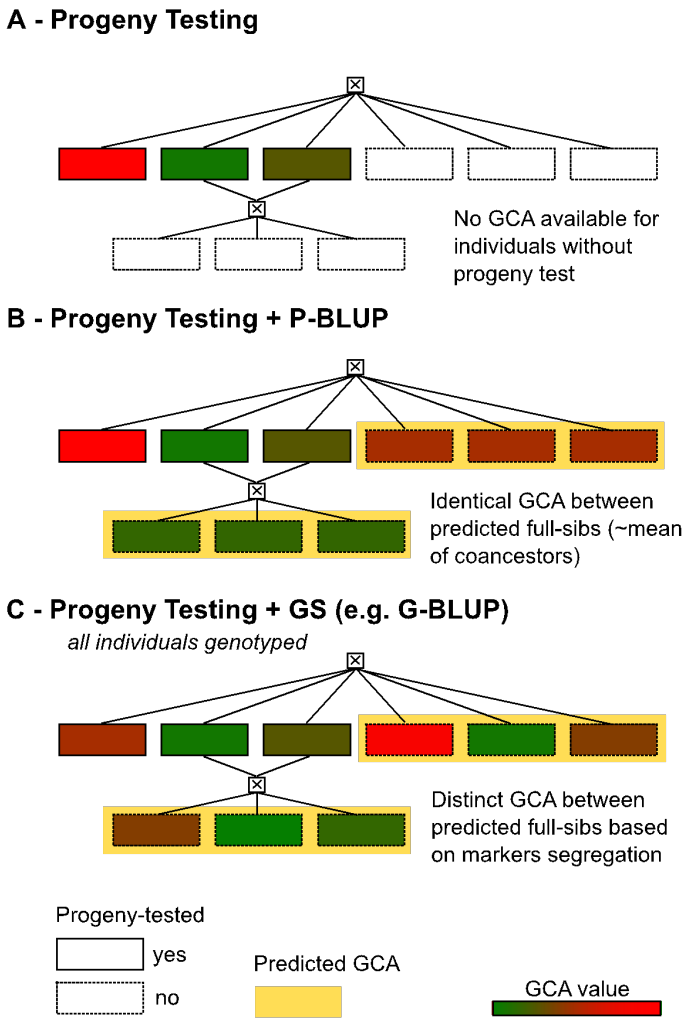


Figure 2: Comparison of GCA calculation methods based on a simple case.

The general model used for GCA prediction can be written as follows:

$$Y = X\beta + Zb + Z_A g_A + Z_B g_B + Z_D s_{AB} + e$$

where Y is the vector of the phenotypes of the hybrid individuals, β and b are the vectors of fixed and random effects due to the experimental design, respectively, X and Z their associated incidence
matrices, g_A and g_B are the vectors of GCA (additive effects) of A and B parents, respectively, s_{AB} is the vector of SCA (dominance effects) of crosses, Z_A, Z_B and Z_D their incidence matrices and *e* is the vector of residual effects. Covariance definition for GCAs defines the main difference between P-BLUP and G-BLUP. For P-BLUP, the covariance is derived from genealogical relationships (pedigree information) whereas for G-BLUP, it is derived from genomic relationships (marker data).

A training set (TS) corresponding to ~500 crosses from 150 A parents and 156 B parents grown in one site in Indonesia was used to predict values for a validation set (VS) of ~200 crosses from 67 A parents and 42 B parents grown in another location in Indonesia (for

details, see Cros et al., 2017a). The parents of both TS and VS were genotyped using genotyping-by-sequencing (GBS) which produced >5000 high quality SNPs suitable for GS. The hybrid crosses were phenotyped but not genotyped. Comparison of the prediction accuracies between P-BLUP and G-BLUP indicated that G-BLUP can perform better than P-BLUP depending on the group and the trait (Table 1 and Figure 3). This observation holds true for parental GCA and hybrid value prediction. The best improvement was obtained for FFB.

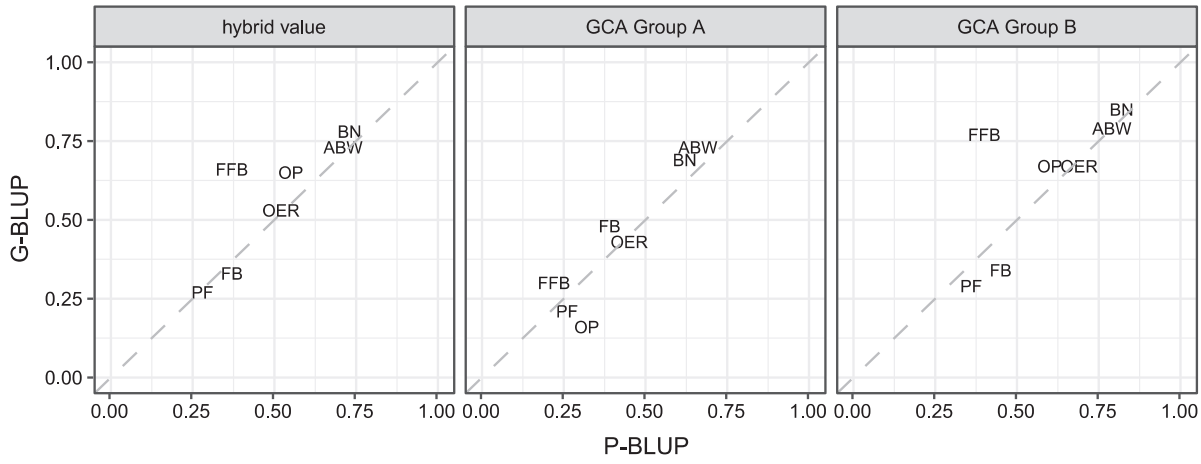


Figure 3: Correlation between P-BLUP and G-BLUP prediction accuracies. The plots are based on the same data as shown in Table 1.

TABLE 1: ACCURACY OF P-BLUP AND G-BLUP FOR YIELD COMPONENTS IN THE STUDY OF CROS et al. 2017A

	Prediction accuracy across populations (sites)					
	Hybrid value		GCA Group A		GCA Group B	
Yield component	P-BLUP	G-BLUP	P-BLUP	G-BLUP	P-BLUP	G-BLUP
FFB	0.37	**0.66**	0.22	**0.30**	0.40	**0.77**
BN	0.73	**0.78**	0.62	**0.69**	0.82	**0.85**
ABW	0.71	**0.73**	0.66	**0.73**	0.79	0.79
FB	0.37	0.33	0.39	**0.48**	0.45	0.34
PF	0.28	0.27	0.26	0.21	0.36	0.29
OP	0.55	**0.65**	0.32	0.16	0.60	**0.67**
OER	0.52	**0.53**	0.45	0.43	0.69	**0.67**

FFB: annual cumulative fresh fruit bunch, in kg
BN: annual cumulative bunch number
ABW: annual average bunch weight, in kg
FB: fruit-to-bunch ratio, in kg
PF: pulp-to-fruit ratio, in %
OP: oil-to-pulp ratio, in %
OER: oil extraction rate, in %
For G-BLUP, the accuracy corresponds to the accuracy obtained with the maximum number of SNPs
Bold: G-BLUP prediction accuracy higher than P-BLUP

A simulation study was performed to assess the potential gain when employing GS as a preselection step on FFB within the classical RRS scheme (Figure 4). For example, FFB in the top 100 hybrid crosses could be increased by ~11% when applying the preselection on a breeding population of 5000 A and 5000 B palms This example demonstrates that a simple preselection step using GS can already greatly improve the genetic gain in commercial hybrids.

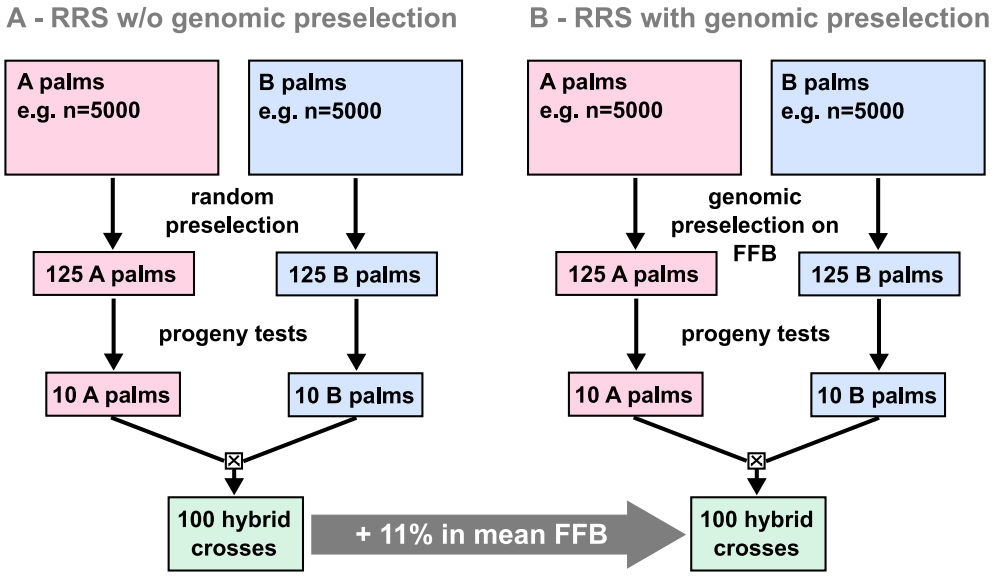


Figure 4: Representation of the simulation design used to estimate the genetic gain of genomic preselection on FBB (B) compared to a classical RRS scheme (A). The analysis is described in Cros et al. 2017a

Overview of the research published by other entities

Until now, very few studies have been published on GS applied to oil palm. We briefly summarize here what is published besides research conducted within PalmElit's network (Cros et al., 2015a, 2015b, 2017a; Marchal et al., 2016).

The first publication on GS in oil palm was presented by Wong and Bernardo (Wong and Bernardo, 2008). This work was conducted in association with Applied Agricultural Resources Sdn. Bhd (AAR). Based on simulated data for a small oil palm parental population derived from a single cross, Wong and Bernardo demonstrated the potential of genomic selection compared to phenotypic selection and QTL-based marker-assisted selection. The study also provided the first estimates of gain depending on parameters such as the size of the breeding population, the number of replications in phenotypic assays, and the heritability of the trait. The cost per unit gain and the time per unit gain were calculated to assess the efficiency of each breeding strategies conducted over 37-38 years (corresponding to 2 cycles of classical phenotypic selection or 4 cycles of marker-assisted or genomic selection). The

improvement obtained with GS (up to +25% in the response to selection with a population size N=70, and cost per unit gain reduced by at least 26% compared to phenotypic selection) was mainly attributed to the shorter generation time when selection was based solely on genotypic data (6 years vs 19 years for a traditional selection cycle). This analysis also suggests that increasing the number of parental palms tested could be more efficient than increasing the number of replication in field tests. However, since this study was conducted with simulated data under specific assumptions, the results need to be validated with empirical data.

A recent study by Sime Darby reported interesting results related to the implementation of GS for early selection among commercial hybrid populations (Kwong et al., 2017). 1,218 commercial hybrids were genotyped and phenotyped for 6 production traits with varying heritability. The GS strategy applied was to use part of the hybrid population as TS to predict the value of the other part (=VS). This study mainly focused on optimizing the marker set and statistical method to maximize the prediction accuracy while reducing the number of markers (potentially leading to reduced genotyping costs). The results of Kwong et al. are further discussed below.

IMPLEMENTATION AND OPTIMIZATION OF GS IN OIL PALM

Considerations for optimizing the use of GS

Several studies provide evidence for the efficiency of genomic prediction for increasing the gain in agronomic traits (e.g. in oil palm: Cros et al., 2017a; Kwong et al., 2017). Still, there is a large gap between the experimental studies and GS routine implementation in breeding. Several facts can explain such discrepancy:

- Additional costs due to genotyping might render the method less profitable. Thus, GS implementation implies an accurate estimation of both gain per time and gain per costs.
- Optimal use of GS could require profound changes in the breeding scheme, with an impact on the traditional breeding practices.
- GS might not be the most optimal selection method for all agronomic traits.
- Commercialization of planting material selected solely on the basis of its genomic estimated value might be problematic due to the absence of phenotypic records to demonstrate its real agronomical value and check secondary traits which have not been selected for.

Many studies have focused on determining the optimal parameters for maximizing the gain while fewer also included cost considerations (e.g. Rajsic et al., 2016; Wong and Bernardo, 2008). Figure 5 summarizes some of the main parameters which affect GS gain and cost, some of which will be further discussed below.

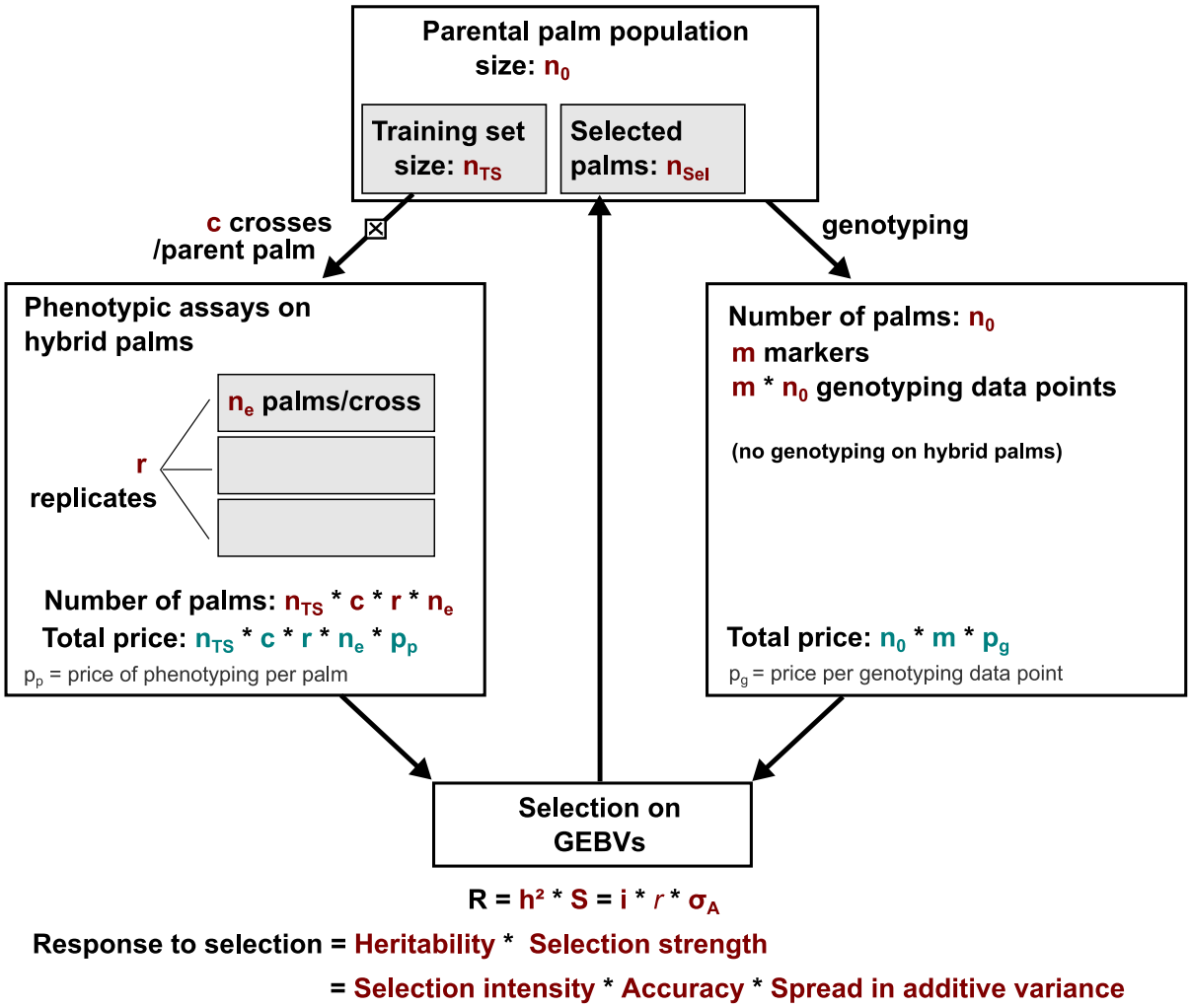


Figure 5: Parameters affecting the gain and cost of GS. Parameters affecting the gain and cost are indicated in brown and cyan respectively. The link between cost and the indicated parameters is direct and given by the formula indicating the total cost for both phenotyping (left) and genotyping (right). The link between the design parameters and the gain (response to selection) is indirect with n_{Sel} and n_0 affecting i, and nearly all design parameters affecting the prediction accuracy r. Omitted here is the generation time, which impacts the gain rate and can be reduced by skipping part or all of the longest phenotyping assays.

Key parameters for gain maximization

As shown in the formula used to estimate the response to selection (Figure 5), i (selection intensity) and *r* (selection accuracy) are key parameters which can be adjusted to optimize GS while σ_A is mainly an intrinsic genetic feature of the trait under selection within the considered breeding population. Overall, increasing the size and diversity within the breeding population, and increasing the prediction accuracy positively contribute to the GS efficiency. The gain rate could also be increased by shortening the generation time, which would imply to decrease the time spent on phenotypic tests, e.g. by implementing progeny tests at lower frequency throughout the cycles.

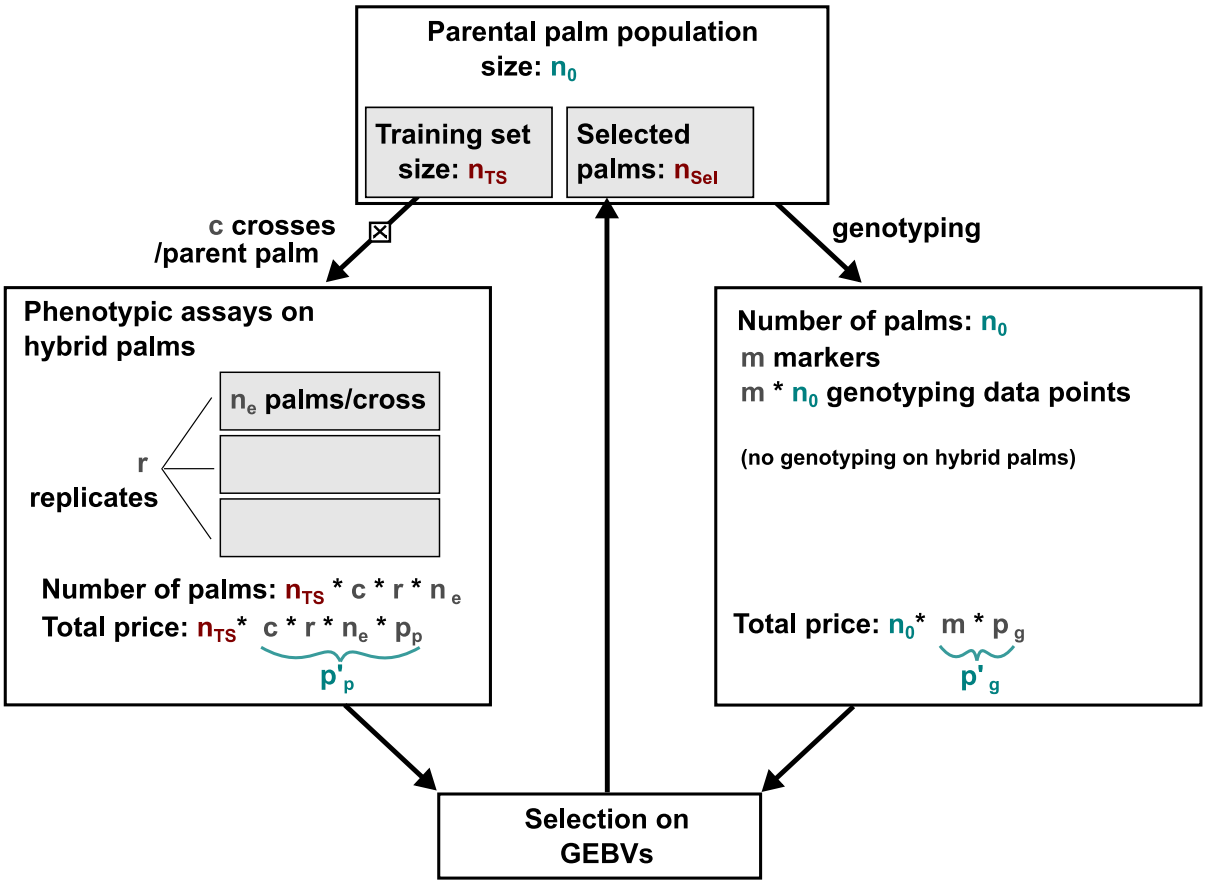
Cost minimization

It has been a general concern among breeders that GS would increase the breeding costs. Such concern should however be allayed by several facts:

- Genotyping costs remains much smaller compared to phenotyping costs, and both technological progress and increasing labor costs will contribute to widen the gap between them in the future. To illustrate such difference, progeny testing in one of our on-going standard genetic trial in Nigeria costs 6,400-22,000€ per parental palm while genotyping cost will likely be below 50€/palm for a large-scale implementation and a reasonable number of SNPs (examples of current estimates are ~50-60€/palm by GBS and ~200-300€/palm with the OP300K SNP). Thus, even with the highest genotyping cost (300€) and the lowest phenotyping costs (6,400€), the ratio of genotyping/phenotyping costs lays below 1/20.
- The cost increase related to GS can be compensated by a decrease in the cost of the phenotypic assays. For example, a similar selection accuracy could be achieved using a smaller but better designed training set (Rincent et al., 2017; Wolfe et al., 2017).
- Considering the current genotyping costs per individual applying GS for direct selection in commercial hybrids seems not economically viable, unless the selected hybrids can provide a sufficient return on investment (e.g. by cloning them). It is possible to focus the genotyping effort on the parental palms, allowing both parent and hybrid prediction, while reducing the genotyping costs (Cros et al., 2015a, 2015b, 2017a; Marchal et al., 2016) compared to strategies where hybrid individuals are also genotyped (Kwong et al., 2017).

In silico breeding approach for GS design optimization

Based on the same scheme described in Figure 5, which illustrates a simple use of GS for breeding within a parental population, we tested the impact of several parameters on costs and genetic gain. The analysis design is described in Figure 6 and the results in Figure 7.



Scenario 1: parental palm selection without GS
Scenario 2: parental palm selection with GS

Response to Scenario 2 (R_2) / Response to Scenario 1 (R_1)

$$R_2/R_1 = i_2/i_1 * r_2/r_1$$

Fixed parameters
n_{TS} = 100
n_{Sel} = 25

Varying parameters
n_0 = 100 - 5000 = number of palms in the parental population
$p'_p = c * r * n_e * p_p$ = 6400 - 22000 € = phenotyping costs per parental palm
$p'_g = m * p_g$ = 50 - 300 € = genotyping costs per palm
i_2 is directly derived from n_0
r_2/r_1 = 0.8 - 1.8 (estimation derived from Cros et al. 2017)

Figure 6: Design used to test the influence of several parameters on GS gain and cost. Parameters with a unique fixed value are indicated in brown. Parameters with variable values are indicated in cyan, the range allowed is delimited by extreme values estimated based on empirical data. Parameters which are not included are in grey. Parental palm selection without GS (Scenario 1) consists in progeny-testing n_{TS}=100 parent palms and selecting the top n_{Sel}=25 palms. Parental palm selection with GS (Scenario 2) consists in progeny-testing n_{TS}=100 parent palms, estimating breeding values for n_0 breeding palms including the training set after having all of them genotyped, and selecting the top n_{Sel}=25 palms. C: number of crosses per tested palm. r: number of replicates for each tested cross. n_e: number of palms tested per cross. p_p: phenotyping cost per hybrid palm in the field. m: number of markers. p_g: cost per marker data point. p'_p and p'_g: phenotyping and genotyping costs per breeding palm respectively. i: selection intensity. r: selection accuracy.

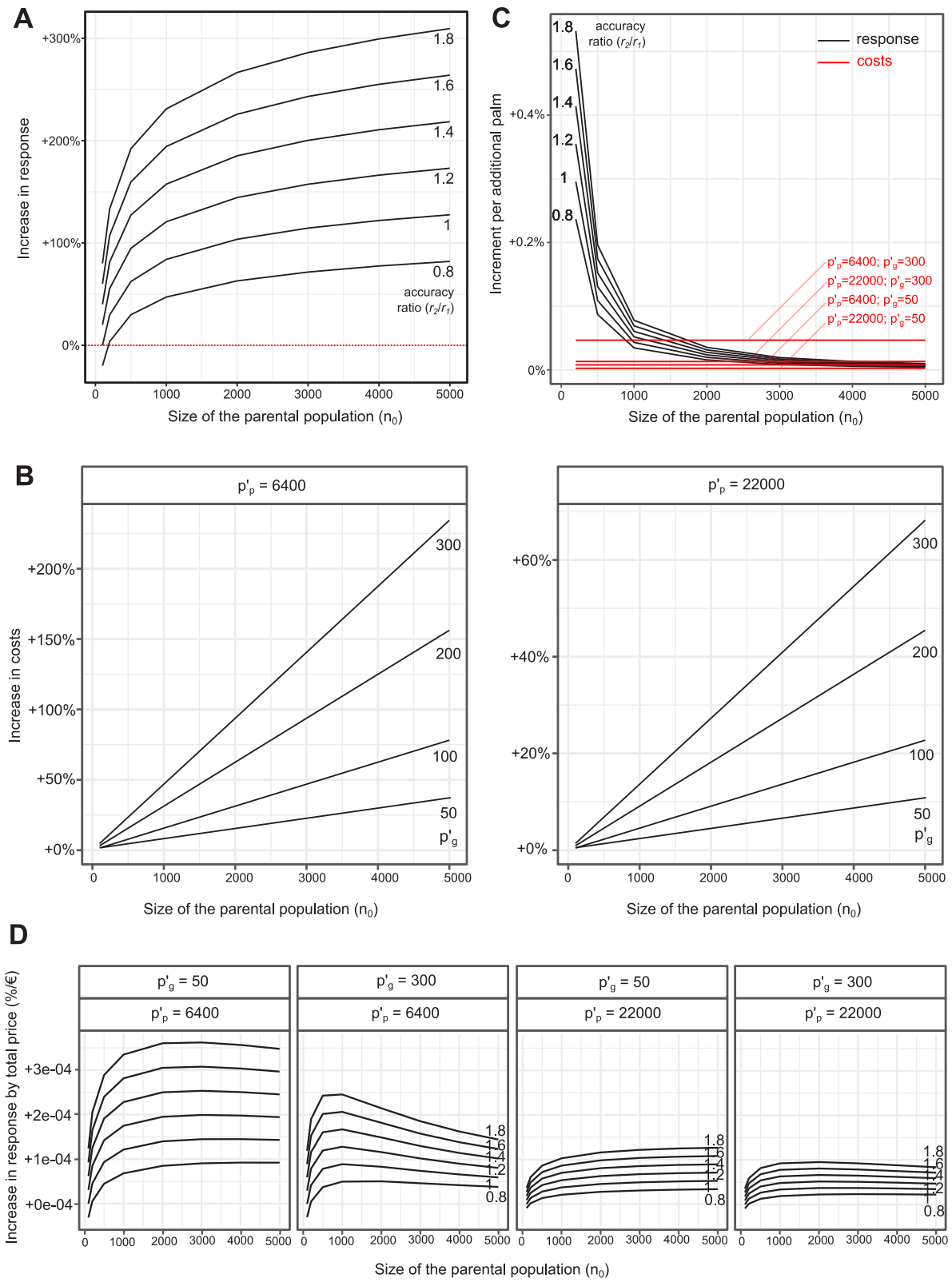


*Figure 7: Influence of several parameters on GS gain and cost. The parameters tested are described in Figure 6: increase in response to GS as compared to phenotypic selection ($(R_2-R_1)/R_1$), genomic vs phenotypic selection accuracy ratio (r_2/r_1), phenotyping and genotyping costs per breeding palm (p'_p and p'_g respectively), and size of the breeding population (n_0) **A.** Additional genetic gain when using GS as compared to selection based on traditional progeny testing for different accuracy ratios between GS and the traditional method, and*

depending on the population size. ***B.*** *Increase in costs depending on the population size, phenotyping costs, and genotyping costs.* ***C.*** *Cost and gain increment when increasing the population size, depending on the accuracy ratio, phenotyping costs, and genotyping costs.* ***D.*** *Increase in genetic gain per total cost, depending on the population size, phenotyping and genotyping costs, and the accuracy ratio.*

This analysis suggests that:

- Even with a lower accuracy, the response obtained with GS will in most cases (provided that n_0 is large enough) outperform that of traditional selection (Figure 7A).
- The cost increase (due to genotyping) remains low compared to the traditional cost. This is due to the low genotyping/phenotyping cost ratio. Moreover, the genotyping costs could be compensated by further decreasing the phenotyping costs (e.g. less frequent phenotypic tests, or decreasing the number of replicates) (Figure 7B).
- When the size of the genotyped population increases, the gain increases non-linearly while the cost increases linearly. Thus, the effect on the response of adding more palms to the test population is low above 1000 individuals while the genotyping cost increment remains constant (Figure 7C).
- Consistent with the point above, the increase in response per cost peaks for a fixed population size, which depends on the genotyping and phenotyping costs (Figure 7D).

Reduction of the genotyping costs

Since GS requires dense markers at low costs and genomic resources are now available, SNP markers are now favored over other marker types such as SSRs. So far, exploratory studies in oil palm employed SNP genotyping techniques which provide very large numbers of SNPs at relatively high costs (several thousand SNPs, for ~50-300€/palm, which are realistic estimates for the genotyping costs with GBS in Cros et al., 2017a and with the OP300K SNP array in Kwong et al., 2017, respectively). It has been proposed that reducing the number of SNPs used in GS could contribute to reduce the costs. However, this strategy has downsides:

- Reducing the marker number results in the prediction being influenced more by realized relatedness rather than by QTL effects, thereby decreasing the advantage compared to pedigree-based prediction (Jannink et al., 2010).
- Although approaches have been proposed to define the optimal marker sets and these sometimes even improve the prediction accuracy compared to prediction using all available markers (Cros et al., 2017a; Kwong et al., 2017), the defined marker sets are trait-specific, and thus, not necessarily optimal for multi-trait breeding.

In addition to marker number, other technical aspects of genotyping can be optimized. For example, adapting the genotyping technique depending on the number of markers and samples, and minimizing the labor-intensive steps in sample collection and handling.

Optimization of the GS accuracy

As illustrated in Figure 7, prediction accuracy is a key factor for GS efficiency. The prediction accuracy reflects how well the model deduced from the training set can predict the

genotypic and/or phenotypic value of the tested population. Many factors can affect the accuracy. A brief overview is given below.

Selected trait

Many studies have already highlighted the influence of the trait genetic architecture on GS accuracy. Critical parameters are for example: the QTL number, the heritability, the respective proportion of genetic additivity, dominance and epistasis. Theoretically, the accuracy positively correlates with heritability and this has already been confirmed in oil palm (Kwong et al., 2017) as well as in other species (Covarrubias-Pazaran, 2016; Duangjit et al., 2016; Wolfe et al., 2017). Additive effects are easier to estimate compared to dominant and epistatic effects. Because the part of additivity is generally larger in hybrids, models based solely on additivity can perform well (reviewed in Zhao et al., 2015). In oil palm, this general principle seems to hold true for several yield traits (Cros et al., 2017a; Kwong et al., 2017; Marchal et al., 2016).

Training set

Besides the quality of both genotypic and phenotypic data, the training set design represents a critical factor. The training set combines phenotypic and genotypic data in order to calibrate the model used for prediction. Ideally, the training set should be large and cover all the genetic diversity present in the test population in an unbiased manner (topic reviewed in Zhao et al., 2015). This implies that the relatedness between the training set and the test population must be as high as possible (Cros et al., 2015a; Zhao et al., 2015) while population structure must remain low (exemplified in Duangjit et al., 2016). As a consequence, the training set needs updating along the breeding cycles. Since compliance with these rules can prove difficult when dealing with typical breeding populations, some methods have been proposed to optimize the training set design (Rincent et al., 2017; Wolfe et al., 2017). For breeding companies, a good knowledge of the history and genetics of the breeding population can significantly support the training set design.

Statistical models

A range of statistical methods are available for GS and their efficiency has been already compared in several studies (Covarrubias-Pazaran, 2016; Cros et al., 2015a; Heslot et al., 2012; Jannink et al., 2010; Kwong et al., 2017; Zhao et al., 2015). In many cases, the models display similar performances. On a theoretical point of view, some models might be better suited than others depending on the genetic architecture of the trait considered. Some Bayesian models, for example, can potentially better account for traits which are affected by QTLs with varying effect variances (e.g. a few QTLs with large variance and many with smaller variance), contrary to BLUP which assumes equal effect variance for all QTLs. So far, we have privileged the use of G-BLUP model (which is analogous to rr-BLUP) implemented in ASReml®, since this model has proven its robustness and efficiency for a diversity of trait and species (Covarrubias-Pazaran, 2016; Heslot et al., 2012; Jannink et al., 2010; Zhao et al., 2015), including yield traits in oil palm (Cros et al., 2015a, 2017a).

At the instar of FFB, BN, and ABW in oil palm (Figure 8A-B, Tisné et al., 2015), agronomic traits are often correlated. This implies that independent selection on each correlated trait

might not yield the best results. Multivariate GS, together with index selection, has the potential to overcome this issue. Multivariate GS can for example increase the accuracy as shown for BN and ABW in oil palm (Marchal et al., 2016).

Genotypic data

When many markers are available (e.g. several thousand), using all of them can decrease the accuracy and heritability. A similar result is obtained with too few markers (Cros et al., 2017a; Kwong et al., 2017). The optimal marker number and density is determined by factors such as relatedness, effective population size, and genetic diversity (Jannink et al., 2010; Zhao et al., 2015).

Several studies assessed the impact of marker selection based on criteria such as the marker distribution, LD, and association with the trait. For example, Kwong et al. showed that marker selection based on association and LD can lead to improved accuracy (Kwong et al., 2017). One drawback of this strategy, however, is that the marker set defined is trait- and population-specific.

The quality of the genotypic data can affect the prediction accuracy. For example, missing data is undesirable though imputing can compensate for it, especially when using pedigree data (Cros et al., 2017a). In that respect, genotyping techniques which yield high-quality data at low missing rates should be privileged.

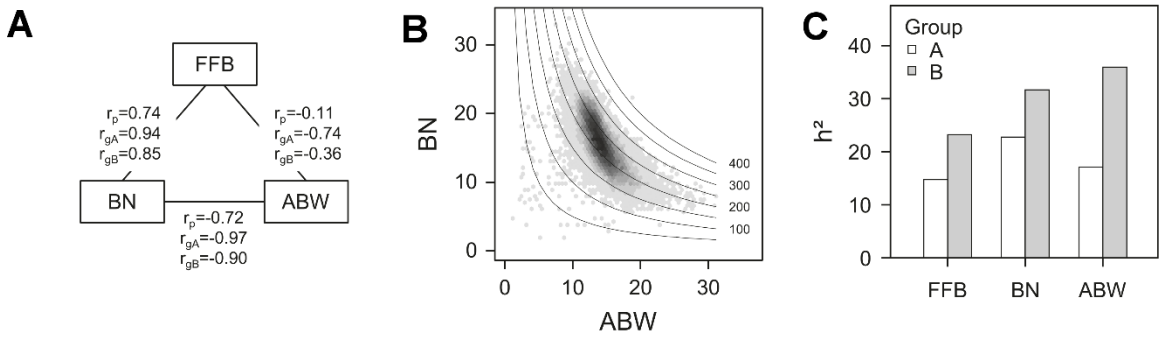


Figure 8 : Heritability and correlations for BN, ABW and FFB depending on the population considered. Figure extracted from Tisné et al., 2015. ***A.*** *Phenotypic correlations (r_p) and genotypic correlations in heterotic groups A (r_{gA}) and B (r_{gB}) between FFB, BN and ABW.* ***B.*** *Relationship between average bunch weight (ABW) and bunch number (BN) in a A × B population. The grey scale indicates the density of points with similar BN and ABW values. Isoproduction curves are drawn with corresponding FFB values given on the right of the curves.* ***C.*** *Narrow sense heritability (h^2) for the three production traits, i.e. FFB, BN, and ABW estimated from A × B individuals.*

Integrating GS within effective breeding schemes

How to optimally integrate GS within the selection scheme is a long-standing question. In the context of a breeding program, where resources are limited, implementing GS without any cost increase implies a resource reallocation. Few published studies report investigations on strategies to apply GS for hybrid breeding (Endelman et al., 2014; Longin et al., 2015; Lorenz, 2013; Marulanda et al., 2016; Riedelsheimer and Melchinger, 2013). At least three types of scenarios can be envisaged: within population GS, across population GS, and across generation GS. A comparison of the three scenarios was performed in cassava, which highlights the tradeoff between selection accuracy and distance between training set and test population (Wolfe et al., 2017).

In oil palm, Cros et al. showed the interest of adding GS as a within population pre-selection step to the conventional RRS (Cros et al., 2017a). Though more efficient in terms of genetic gain than classical RRS, the generation time in this scenario does not decrease compared to RRS. In a more recent simulation study, Cros et al. assessed the gain for FFB in breeding strategies where the training set is updated only every second or third generation, and includes individuals from one or two generations (Figure 9, Cros et al., 2017b). As expected, the selection accuracy and gain decreases with the number of generations between GS selection candidates and training set (Figure 10). In this simulation, updating the training set every second cycle by progeny testing and aggregation of data from two cycles performed best (Figure 10B). Thus, the generation time can be dramatically shortened every second cycle which compensates for the slight decrease in prediction accuracy.

This scheme can certainly be further improved, since many other parameters can be modulated to improve the overall efficiency. The most promising strategies can then be implemented in the field to determine their actual performance.

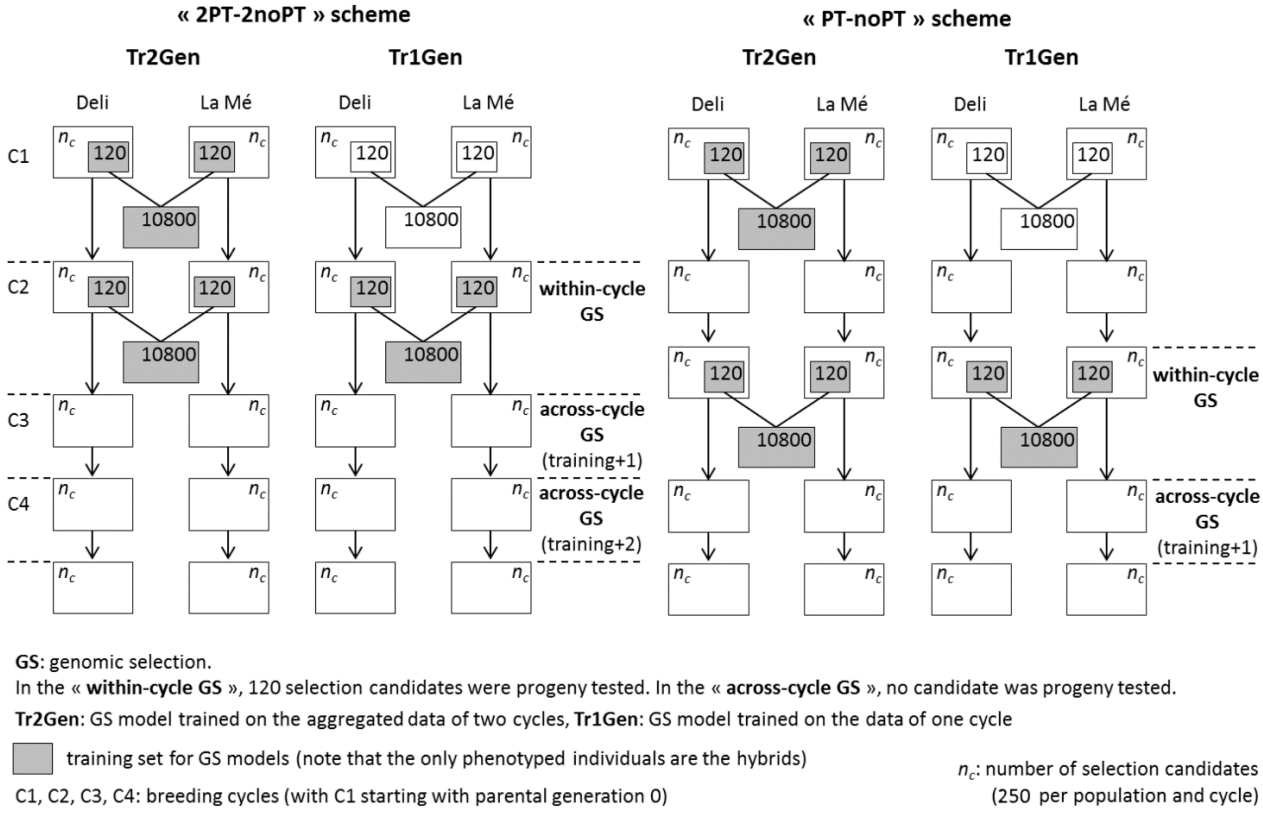


Figure 9: Comparison of breeding strategies involving a training set based on generation (Tr1Gen) or two successive generations (Tr2Gen) and updated every second (PT-noPT) or third generation (2PT-2noPT). 18 individuals were selected within each population at each cycle. Figure from Cros et al., 2017b.

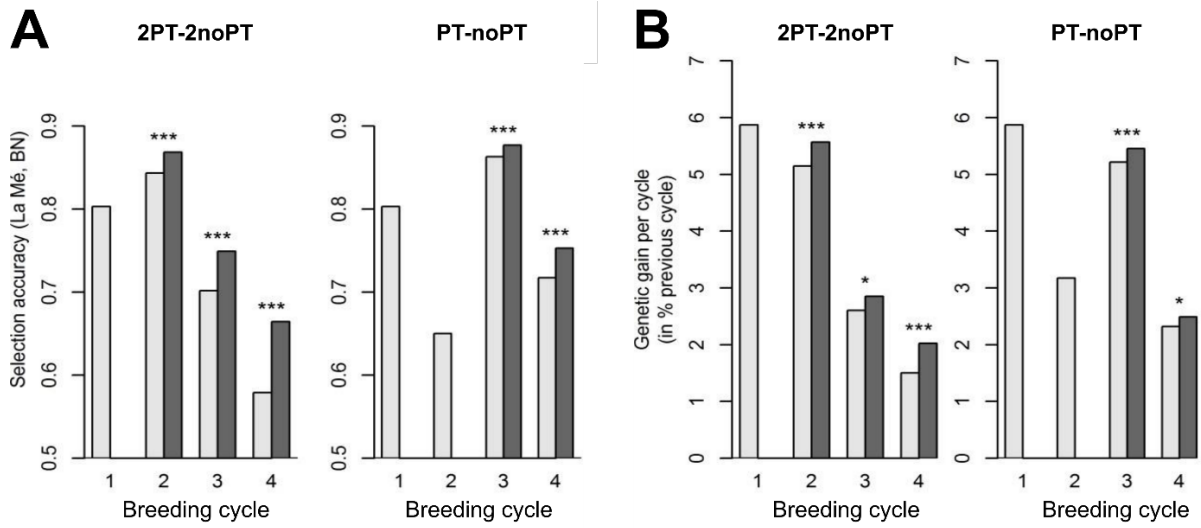


*Figure 10: Selection accuracy for BN in group B **(A)** and genetic gain for FFB **(B)** for the breeding schemes described in Figure 9. The breeding population size was fixed to 250 individuals per population and cycle. 18 individuals were selected within each population at each cycle. Data for Tr1Gen and Tr2Gen are presented in light grey and dark grey respectively. Figure from Cros et al., 2017b.*

CONCLUSION AND DISCUSSION

Concluding remarks

Similar to what was already shown in other hybrid species (Zhao et al., 2015), GS has the potential to increase the breeding efficiency in oil palm. Simple strategies such as the ones described above could significantly increase the genetic progress in oil palm. Using GS as a pre-selection step can already increase the FFB by 11% over one RRS cycle compared to traditional phenotypic selection (Figure 4, Cros et al., 2017a). Similar to what was observed in black spruce and maritime pine (Bartholomé et al., 2016; Lenz et al., 2017), the selection accuracy is not significantly increased with GS as compared to pedigree-based selection for some agronomic traits (Cros et al., 2017a). For these, a higher gain can only be obtained if GS is associated with a reduction in the generation time and/or an increased selection intensity. So far, oil palm studies have focused on GS for yield traits. However, as suggested by a study in wheat, GS could also be efficient with other traits such as disease resistance (Juliana et al., 2017). This needs to be tested in the future.

Since genotyping generates additional costs, resource reallocation (i.e. by minimizing progeny testing) might be necessary to compensate for those, as proposed in Figure 9. For this, we also concentrate the genotyping effort on the breeding population, thus limiting the number of individuals to genotype, while the value of commercial hybrids (not genotyped) can be accurately predicted based the parents' genetic value. From our own data, we note that the cost of progeny-testing one individual is far above its genotyping cost, and the gap is expected to widen in the future as the genotyping costs are gradually decreasing. The optimal strategy is difficult to determine since many parameters need to be taken into account. Some of them are illustrated in Figure 6. Thus, GS will likely be implemented differently depending on the economic and technical constraints applying to the oil palm breeding companies.

Challenges and perspectives for GS in oil palm

Estimating the gain of GS in a simple breeding scenario and for a unique trait is rather simple (e.g. the analysis presented in Figures 6 and 7). Designing and assessing a breeding strategy that integrates GS while allowing efficient selection for all agronomic traits can be more complex, especially when some of these agronomic traits are correlated (example of FFB, BN and ABW in Figure 8A-B). Since the selection strategy needs to take into account the genetics of the breeding population and of the traits under selection, the reality might prove even more complex in oil palm where the A and B breeding populations have fairly distinct history and characteristics, and the agronomic traits have various genetic architectures (Figure 8C and Corley and Tinker, 2015b). Consistent with this, the study by Cros et al. highlights the differential advantage of GS in oil palm depending on the population and the trait (Cros et al., 2017a). Fine tuning will be necessary to develop a breeding strategy which can efficiently select for traits with various genetic architecture and heritability. Since selection on genomic values is relevant for a subset of the agronomic traits only (e.g. quantitative traits), it is then more appropriate to use the term "genomics-assisted selection" to describe the full breeding strategy, which encompasses several other selection steps such as MAS for mono- and oligogenic traits. There are certainly things to learn from other hybrid crop species but breeding in highly valuable perennial crops differs on several aspects from breeding in low-value annual crops. Thus, solutions implemented in other hybrid crops might not be directly applicable to oil palm. In that respect, research in non-hybrid perennial species

such as forest trees can be useful (Bartholomé et al., 2016; Isik et al., 2016; Lenz et al., 2017; Resende et al., 2017), even though there is no report on practical GS implementation so far.

On the other hand, there is still some uncertainty about the long-term progress with GS. Reports on GS over many cycles based on empirical data are so far limited. In a multi-parental population of tropical maize, a high genetic gain was achieved using rapid cycling GS over four cycles (Zhang et al., 2017). Predictions can be easily derived from the breeder's formula but this does not take into account genetic drift, random changes in allele frequencies and dominance effects. A loss of genetic variation after each selection cycle is unavoidable. Since the genetic progress correlates with the genetic diversity, the response to selection is expected to decrease over the selection cycles. Thus, managing diversity is a key point for enhancing the long-term progress. By increasing the selection intensity and shortening the breeding cycle, GS could exacerbate this phenomenon, thereby increasing the short-term progress at the expense of the long-term one. This is even more relevant for oil palm where the base population is very narrow.

ABBREVIATIONS

ABW: annual average bunch weight
BLUP: based linear unbiased prediction
BN: annual cumulative bunch number
BV: breeding value
CPO: crude palm oil
FB: fruit-to-bunch ratio
FFB: annual cumulative fresh fruit bunch
G-BLUP: genomic best linear unbiased prediction
GBS: genotyping by sequencing
GCA: general combining ability
GEBV: genomic estimated breeding value
GS: genomic selection
KPO: kernel palm oil
MAS: marker-assisted selection
OER: oil extraction rate
OP: oil-to-pulp ratio
P-BLUP: pedigree-based best linear unbiased prediction
PF: pulp-to-fruit ratio
QTL: quantitative trait locus
rr-BLUP: random regression best linear unbiased prediction
RRS: recurrent reciprocal selection
SCA: specific combining ability
SNP: single nucleotide polymorphism
T-BLUP: traditional best linear unbiased prediction
TS: training set
VS: validation set

REFERENCES

Bartholomé, J., Van Heerwaarden, J., Isik, F., Boury, C., Vidal, M., Plomion, C., and Bouffier, L. (2016). Performance of genomic prediction within and across generations in maritime pine. BMC Genomics *17*, 604.

Baudouin, L., Baril, C., Clément-Demange, A., Leroy, T., and Paulin, D. (1997). Recurrent selection of tropical tree crops. Euphytica *96*, 101–114.

Collard, B.C.., and Mackill, D.J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos. Trans. R. Soc. B Biol. Sci. *363*, 557–572.

Corley, R. h. v., and Tinker, P. b. (2015a). Selection and Breeding. In The Oil Palm, (John Wiley & Sons, Ltd), pp. 138–207.

Corley, R. h. v., and Tinker, P. b. (2015b). The Origin and Development of the Oil Palm Industry. In The Oil Palm, (John Wiley & Sons, Ltd), pp. 1–29.

Covarrubias-Pazaran, G. (2016). Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. PLoS ONE *11*.

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

Cros, D., Denis, M., Bouvet, J.-M., and Sánchez, L. (2015b). Long-term genomic selection for heterosis without dominance in multiplicative traits: case study of bunch production in oil palm. BMC Genomics *16*, 651.

Cros, D., Bocs, S., Riou, V., Ortega-Abboud, E., Tisné, S., Argout, X., Pomiès, V., Nodichao, L., Lubis, Z., Cochard, B., et al. (2017a). Genomic preselection with genotyping-by-sequencing increases performance of commercial oil palm hybrid crosses. Theor. Appl. Genet. *in press*.

Cros, D., Tchounke, B., and Nkague-Nkamba, L. (2017b). Training genomic selection model across multiple breeding cycles increases genetic gain in oil palm. In Séminaire SelGen 2017, "La Sélection Génomique - Bilan et Perspectives," (INRA, Paris), p.

Duangjit, J., Causse, M., and Sauvage, C. (2016). Efficiency of genomic selection for tomato fruit quality. Mol. Breed. *36*, 29.

Durand-Gasselin, T., Blangy, L., Picasso, C., Franqueville, H. de, Breton, F., Amblard, P., Cochard, B., Louise, C., and Nouy, B. (2010). Sélection du palmier à huile pour une huile de palme durable et responsabilité sociale. Ol. Corps Gras Lipides *17*, 385–392.

Endelman, J.B., Atlin, G.N., Beyene, Y., Semagn, K., Zhang, X., Sorrells, M.E., and Jannink, J.-L. (2014). Optimal Design of Preliminary Yield Trials with Genome-Wide Markers. Crop Sci. *54*, 48–59.

Gallais, A., and Poly, J. (1990). Théorie de la sélection en amélioration des plantes (Masson).

Heslot, N., Yang, H.-P., Sorrells, M.E., and Jannink, J.-L. (2012). Genomic Selection in Plant Breeding: A Comparison of Models. Crop Sci. *52*, 146–160.

Isik, F., Bartholomé, J., Farjat, A., Chancerel, E., Raffin, A., Sanchez, L., Plomion, C., and Bouffier, L. (2016). Genomic selection in maritime pine. Plant Sci. *242*, 108–119.

Jannink, J.-L., Lorenz, A.J., and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics *9*, 166–177.

Juliana, P., Singh, R.P., Singh, P.K., Crossa, J., Huerta-Espino, J., Lan, C., Bhavani, S., Rutkoski, J.E., Poland, J.A., Bergstrom, G.C., et al. (2017). Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. Theor. Appl. Genet. *130*, 1415–1430.

Kwong, Q.B., Ong, A.L., Teh, C.K., Chew, F.T., Tammi, M., Mayes, S., Kulaveerasingam, H., Yeoh, S.H., Harikrishna, J.A., and Appleton, D.R. (2017). Genomic Selection in Commercial Perennial Crops: Applicability and Improvement in Oil Palm (Elaeis guineensis Jacq.). Sci. Rep. *7*, 2872.

Lenz, P.R.N., Beaulieu, J., Mansfield, S.D., Clément, S., Desponts, M., and Bousquet, J. (2017). Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (Picea mariana). BMC Genomics *18*, 335.

Longin, C.F.H., Mi, X., and Würschum, T. (2015). Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. TAG Theor. Appl. Genet. Theor. Angew. Genet. *128*, 1297–1306.

Lorenz, A.J. (2013). Resource Allocation for Maximizing Prediction Accuracy and Genetic Gain of Genomic Selection in Plant Breeding: A Simulation Experiment. G3 GenesGenomesGenetics *3*, 481–491.

Marchal, A., Legarra, A., Tisné, S., Carasco-Lacombe, C., Manez, A., Suryana, E., Omoré, A., Nouy, B., Durand-Gasselin, T., Sánchez, L., et al. (2016). Multivariate genomic model improves analysis of oil palm (Elaeis guineensis Jacq.) progeny tests. Mol. Breed. *36*, 2.

Marulanda, J.J., Mi, X., Melchinger, A.E., Xu, J.-L., Würschum, T., and Longin, C.F.H. (2016). Optimum breeding strategies using genomic selection for hybrid breeding in wheat, maize, rye, barley, rice and triticale. Theor. Appl. Genet. *129*, 1901–1913.

Meuwissen, T.H., Hayes, B.J., and Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics *157*, 1819–1829.

Rajsic, P., Weersink, A., Navabi, A., and Pauls, K.P. (2016). Economics of genomic selection: the role of prediction accuracy and relative genotyping costs. Euphytica *210*, 259–276.

Resende, R.T., Resende, M.D.V., Silva, F.F., Azevedo, C.F., Takahashi, E.K., Silva-Junior, O.B., and Grattapaglia, D. (2017). Assessing the expected response to genomic selection of individuals and families in Eucalyptus breeding with an additive-dominant model. Heredity *119*, 245–255.

Riedelsheimer, C., and Melchinger, A.E. (2013). Optimizing the allocation of resources for genomic selection in one breeding cycle. TAG Theor. Appl. Genet. Theor. Angew. Genet. *126*, 2835–2848.

Rincent, R., Charcosset, A., and Moreau, L. (2017). Predicting genomic selection efficiency to optimize calibration set and to assess prediction accuracy in highly structured populations. Theor. Appl. Genet. 1–17.

Tisné, S., Denis, M., Cros, D., Pomiès, V., Riou, V., Syahputra, I., Omoré, A., Durand-Gasselin, T., Bouvet, J.-M., and Cochard, B. (2015). Mixed model approach for IBD-based QTL mapping in a complex oil palm pedigree. BMC Genomics *16*.

Wolfe, M.D., Carpio, D.P.D., Alabi, O., Egesi, C., Ezenwaka, L.C., Ikeogu, U.N., Kawuki, R.S., Kayondo, I.S., Kulakow, P., Lozano, R., et al. (2017). Prospects for genomic selection in cassava breeding. BioRxiv 108662.

Wong, C.K., and Bernardo, R. (2008). Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. TAG Theor. Appl. Genet. Theor. Angew. Genet. *116*, 815–824.

Zhang, X., Pérez-Rodríguez, P., Burgueño, J., Olsen, M., Buckler, E., Atlin, G., Prasanna, B.M., Vargas, M., Vicente, F.S., and Crossa, J. (2017). Rapid Cycling Genomic Selection in a Multi-parental Tropical Maize Population. G3 Genes Genomes Genet. g3.117.043141.

Zhao, Y., Mette, M.F., and Reif, J.C. (2015). Genomic selection in hybrid breeding. Plant Breed. *134*, 1–10.

Genomic Selection: A Method for Accelerated Improvement of Oil Palms

Harikrishna Kulaveerasingam; Teh Chee Keng; Kwong Qi Bin; Ong Ai Ling and David Appleton

ABSTRACT

Oil palm is the most productive oil crop producing 37 % of total vegetable oil from just 18 million ha, or 6% of total agricultural land cultivated with oil crops (Oil World, 2016). Despite having yields greater than 12 MT/oil/ha/yr in breeding trials (Corley and Tinker, 2003), commercial yields have stagnated over the past 25 years at around 3-4 MT oil/ha/yr, on average. In contrast, dramatic yield increases have been realized commercially for soybean and corn through the use of biotechnology techniques. Sime Darby Plantation has invested in an extensive genomics research programme and has developed tools to facilitate the study of the genetic components of oil yield. The trait is complex and therefore suitably large populations are required for this study. For oil palm, this often poses a problem as only small populations are usually available, thereby restricting the trait associated information that can be gained. However, using large multi-parental populations, high-throughput SNP genotyping has enabled successful genetic association studies for both annual crops (Huang et al., 2010; Jia et al., 2013; Li et al., 2013) and oil palm (Teh et al., 2016). Genomic selection (GS), a genome-wide marker approach has been described as a superior method for genetic gain, especially with complex traits and has proved very successful in animal breeding. We have deployed genomic selection for oil palm to facilitate shortening of the breeding cycle through removing the need for progeny testing and dura phenotyping in the case of mother palm selection. Our paper describes the development of a SNP genotyping array, genome-wide association analysis (GWAS) for oil yield components, GS methodology followed by deployment for selective oil palm breeding.

Sime Darby Research Sdn. Bhd., R&D Centre- Carey Island, Lot 2664, Jalan Pulau Carey, 42960, Carey Island, Selangor, Malaysia.

INTRODUCTION

The increasing world population places incredible pressure on the global food production system. The remarkable achievements of the green revolution have passed with commercial yields of many crops stagnating in recent years. Although oil palm (*Elaeis guineensis*) is the most productive oil-producing crop with oil being up to 90% of its mesocarp dry mass (Murphy, 2014), the same challenge persists. For the past 80 years, oil palm breeders selected and crossed oil palms for desired traits to improve oil yield. However, these conventional breeding methods on their own are unable to deliver the dramatic genetic gains required to feed the rapidly growing population.

Modern sequencing technologies have enabled novel genomics approaches to be considered and deployed in the case of annual crops and livestock. This has conferred significant improvements in genetic estimated breeding value (GEBV) and translated to higher commercial yields. However, the success of marker-based technologies in perennial crops, such as oil palm remain less impactful. Classical linkage mapping has been widely deployed in oil palm to identify genetic regions of interests, especially for Mendelian traits. For instance, the genes responsible for fruit form (*Shell*) and fruit colour (*Virescens*) were located (Singh et al., 2013; Singh et al., 2014). Subsequently, the genetic architecture underlying *Shell* was found to be more complicated than expected (Ooi et al., 2016; Teh et al., 2017). Most commercially important traits, such as oil yield, bunch characteristics and abiotic stress tolerance are complex in nature. The study of quantitative trait loci (QTLs) for these traits that include genetic-environment interactions, requires huge mapping populations to provide sufficient statistical resolving power to associate markers with relevant traits. In the case of oil palm, biparental population size is always small resulting in limited recombination, so our strategy was forced to shift from the classical approach to GWAS. We have assayed thousands of progeny palms representing Sime Darby's breeding stocks using a custom 200,000 SNP genotyping array, namely OP200K (Kwong et al., 2016). The important QTLs for mesocarp oil yield and shell thickness traits were first identified and successfully validated in other populations (Kwong et al., 2016; Teh et al., 2016). Nevertheless, the identified QTLs do not explain the major trait variation, so we decided to improve this further through GS.

Genomic selection is a form of marker assisted selection using markers distributed across the entire genome. The first GS without any *a priori* knowledge of association and genome position of individual markers was carried out in cattle (Hayes et al., 2009). The approach has been evaluated and adapted in plants, inclusive of maize and rice (Crossa et al., 2014; Spindel et al., 2015). For oil palm, GS was initially evaluated based on simulated data. Superiority of GS methods over traditional marker-assisted selection (MAS) and phenotypic selection was concluded (Cros et al., 2015a; Cros et al., 2015b; Wong and Bernardo, 2008). We have continued the investigation using our empirical data, providing a better understanding of the applicability for GS on oil palm (Kwong et al., 2017). Genomic selection models for various yield traits were validated and available for deployment. With deep optimization on analytical throughput and production cost, the first 100 ha planting of GenomeSelect™ materials which potentially yield 15% more oil, was successfully launched in 2016. Efficiency improvements of GenomeSelect™ is still ongoing in order to cater for the full replanting operation in Sime Darby Plantation.

RESULTS & DISCUSSION

Most genetic discoveries in oil palm are still mainly based on classical linkage mapping using different marker systems with modest density. The approach is proven to be reasonably powerful for detecting major QTLs for heritable traits. However, good mapping resolution for complex traits, such as oil yield requires significantly larger populations, which is always a constraint for oil palm. Another common problem when using linkage analysis approaches is population-specific QTLs. This occurrs because genetic polymorphisms within a biparental family do not represent the entire origin of oil palms. This can also result in false positives being identified. The outcome being that QTL's often cannot be replicated or used successfully across multiple parental lines, even within restricted families. To address this, more populations, again are needed making replication-validation very iterative and inefficient. Therefore, we decided to embark on GWAS using the genetic variations accumulated across the breeding stock of Sime Darby Plantation. A total of 200,000 SNPs derived from 59 origins was placed on a custom array, namely OP200K. The array was proven to be applicable for a wide range of genetic analyses (Kwong et al., 2016). For instance, the genetic relationship among 312 *tenera* palms was effectively assigned to five major clusters: I. *Ulu Remis* (UR) x AVROS; II. *Johore Labis* (JL) x AVROS; III. Nigerian x AVROS; IV. *Gunung Melayu* (GM) x *Dumpy* AVROS (DA); and V. JL x DA (Fig. 1a). The clustering was in a good agreement with pedigree records. Moreover, the LD decay for each cluster also revealed differences in breeding histories (Fig. 1b). The LD decayed slowest at 120 kb and 146 kb in both the commercial UR x AVROS and JL x AVROS clusters, respectively, compared with 19.5 kb in the semi-wild Nigerian x AVROS cluster. The result indicated lower genetic diversity and potential effects of higher selection pressure in the commercial materials. More importantly, the mapping resolution for whole-genome linkage and GWAS using OP200K array was also proven to be the highest in oil palm research to date, and sufficient for accurate selection. Genotyping was then extended to more than 7,000 progeny palms from heterogeneous breeding populations.

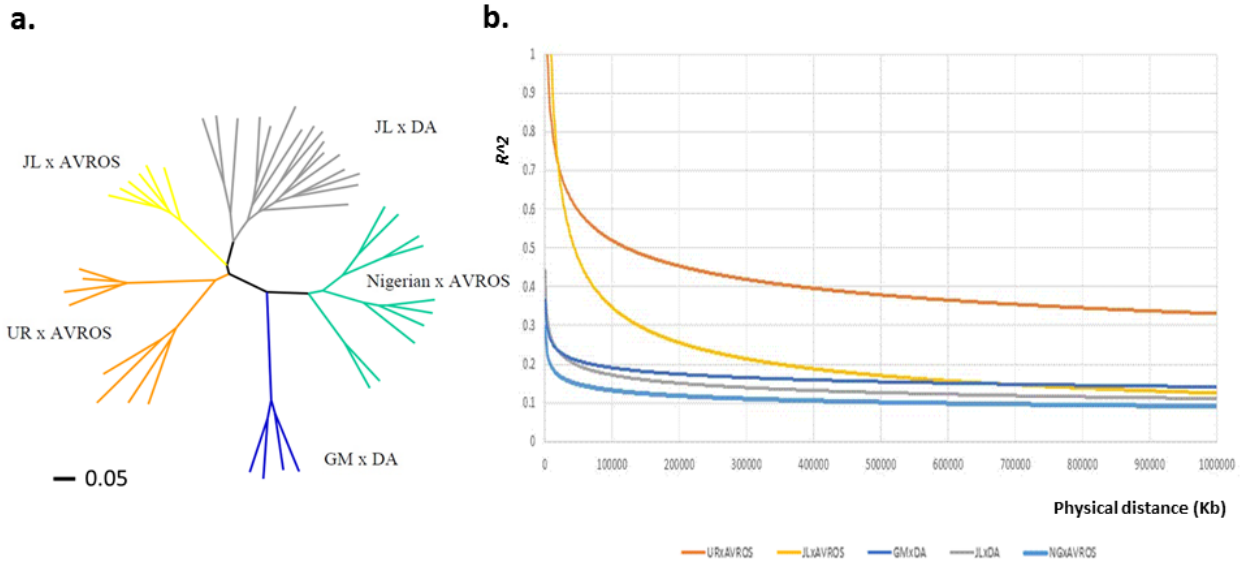


Figure 1: Genetic stratification of 312 tenera palms. (a) A neighbour-joining dengrogram based on pairwise genetic distance. (b) Genome-wide decay across the five tenera clusters estimated from correlation (r^2) between SNPs in a 1-Kb window size within chromosomes.

The first GWAS for high mesocarp oil yield was successfully performed based on 2,056 palms from two important breeding populations. We effectively controlled the genomic inflation due to population structure and cryptic relatedness. In total, 62 and 18 significantly associated SNPs for mesocarp oil yield were mainly clustered on Chromosome 5 and Chromosome 11 in Deli x AVROS and Nigerian x AVROS guided by a genome-wide threshold (p) of 10^{-4} and a Bonferroni threshold of 10^{-7} (Teh et al., 2016). More importantly, we started learning about the adaptability of MAS to reciprocal recurrent selection in oil palm. The breeders potentially can select the elite maternal *dura* and paternal *pisifera* without progeny testing to produce high-yielding *tenera*, so that the breeding selection cycle can be at least halved and made to be more accurate. On the other hand, the QTL regions flagged by the three replicated SNPs in both populations are also being fine mapped to identify the causal genes responsible for mesocarp oil content.

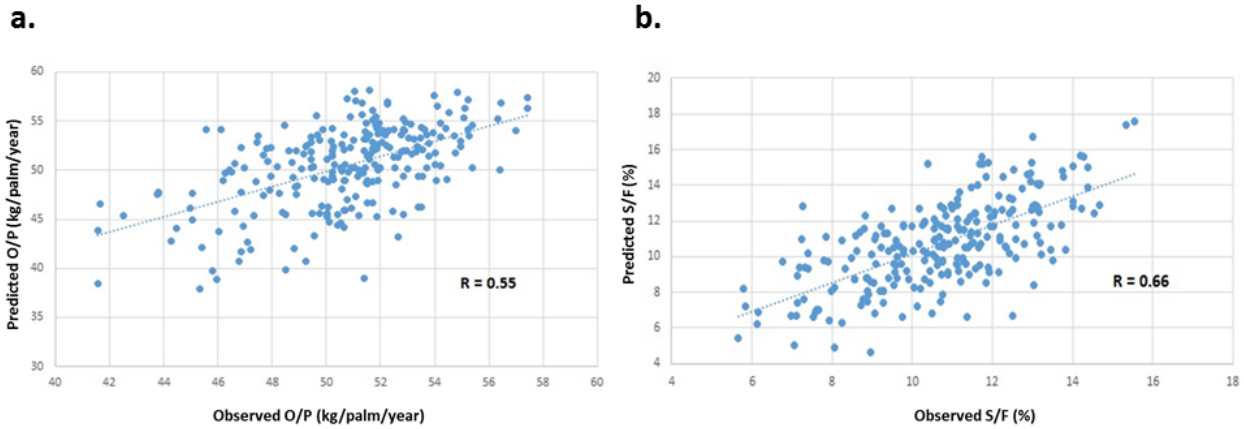


Figure 2: Representative regression plots of predicted and observed yield traits (a) Prediction accuracy for oil-to-palm (O/P, kg/palm/year). (b) Prediction accuracy for shell-to-fruit ratio (S/F, %).

Traditional MAS only considers the major QTLs, but some of them especially for complex traits may explain a small portion of phenotypic variation, leading to poor selection response (Dekkers, 2004). This has encouraged us to attempt GS by unfolding the whole-genome effects of 200,000 SNPs again. Different models for mesocarp-to-fruit ratio (M/F), shell-to-fruit ratio (S/F), kernel-to-fruit ratio (K/F), fruit per bunch (F/B), oil-to-mesocarp ratio (O/M), oil per bunch (O/B) and oil per palm (O/P) were evaluated in a total of 1,218 UR x AVROS palms (Kwong et al., 2017). The result indicated similar prediction accuracy among RR-BLUP and Bayesian models. As expected, the accuracy per trait was found to be correlating with the genomic heritability per trait (r=0.98). Figure 2a and Figure 2b showed prediction accuracy for the highly heritable S/F trait was much higher than the complex O/P trait. Ultilising all 200K SNPs would be impractical and extremely costly, so we further improved the GS model through RR-BLUP B where GWAS profiles and pairwise LD were taken into consideration. Surprisingly, the prediction accuracy for O/P trait peaked at 1,000

ranked SNPs and started to decline with greater number of SNPs (Figure 3). This was contrary to previous reports that showed improvement in prediction accuracy with increasing marker density. The markers with negative effect were probably masked by random marker selection used for other GS models.

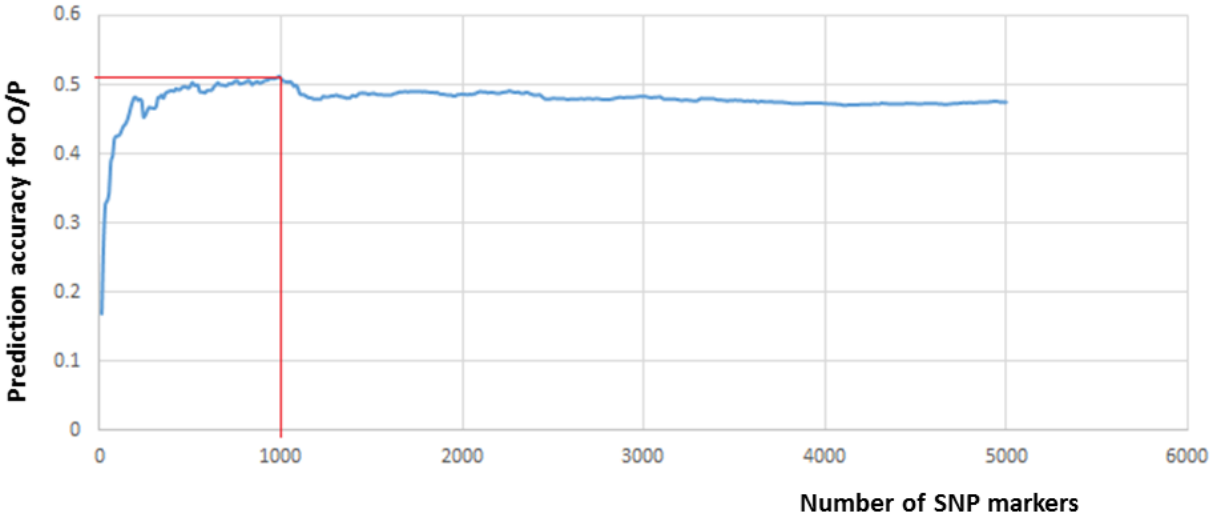


Figure 3: Average prediction accuracy for oil per palm (O/P) trait across different SNP densities using RR-BLUP B model. The prediction accuracy was determined based on correlation (r^2) between predicted and observed O/P traits.

The first GS deployment was to predict and select the highest yielders from the commercial seeds of Sime Darby Plantation. The planting of these materials, termed GenomeSelect™ will potentially yield 15% more oil using the optimised GS model with 0.65 prediction accuracy. A sizable seed producer usually produces more than a million seeds annually. This led us to the next question of scalability and the right choice of genotyping platforms. The current platforms are generally based on polymerase chain reaction (PCR), array and genotype-by-sequencing (GBS) approaches. Cost per data point of GBS and array are known to be lower because of high multiplex ability. The feature is attractive to genetic discovery programmes, which always require high marker density. Nevertheless, PCR methods are still cost effective for 1,000 SNPs or less, which fulfills our required marker density for commercial deployment. The platform also provides lower rates of variant calling error and missing data, compared to GBS and array base. After establishing the PCR-based platform in Sime Darby Plantation, a total of 18,000 GenomeSelect™ palms (Fig. 4a) was field planted in two 50 ha blocks, located at inland and costal estates in year 2016. The palms with pedigree and predicted performance are barcoded and stored in our database. The 1-year old palms at the coastal site are already producing bunches, of which initial fruits have been removed by ablation (Fig. 4b). Yield recording will commence in the 24th month after field planting and prospective validations will follow.

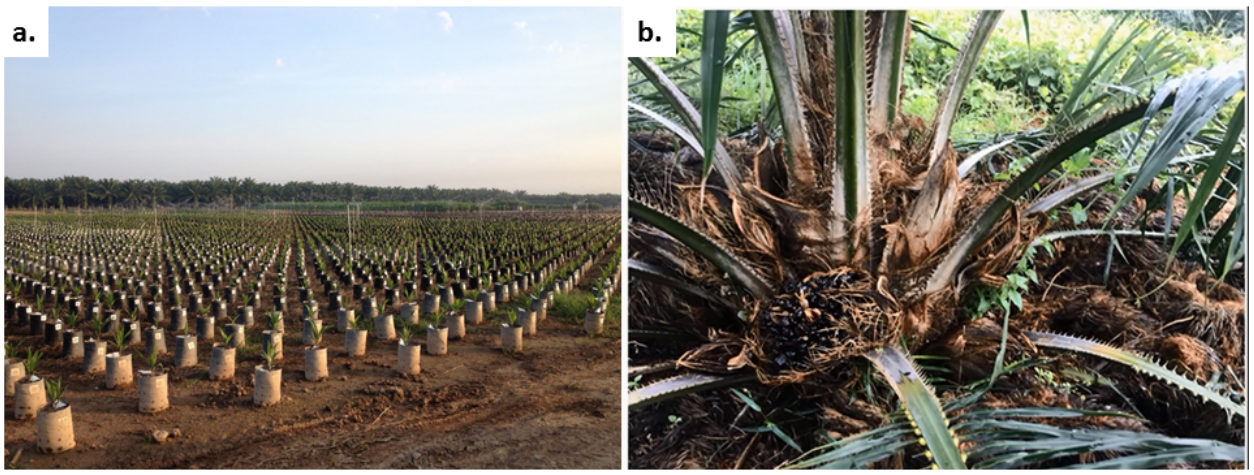


Figure 4: The preparation of GenomeSelect™ materials and current status. (a) A total of 18,000 barcoded GenomeSelect™ was placed in the nursery. (b) Early bunch production of GenomeSelect™ materials after one year of field planting.

CONCLUSION

The intricacies of oil palm breeding and population structure meant new methods of genetic selection needed to be developed. A hybrid of the latest crop and animal genomic selection methods along with human GWAS approaches was used to provide comparatively good trait predictability, even for complex traits such as total yield. These new tools provided optimum accuracy and significantly more feasible methods of testing for commercial application, and were used to select and plant 100 ha, or 18,000 plants for large scale observation. Subsequently, we have used the key learnings from the development of this platform to enable selection of new elite parental lines and seed production capacity to be built to cater for full replanting activities in Sime Darby Plantation and beyond. New traits related to climate tolerance have also been identified and included in the selection protocols, and in the near future these techniques can be used to introduce new secondary traits into commercial populations that can address important challenges for oil palm commercial yield improvement.

REFERENCES

1. Corley, R.H.V., Tinker, P.B., 2003. Selection and breeding, The Oil Palm 3ed. Blackwell, pp. 133 - 187.
2. Cros, D., Denis, M., Bouvet, J.-M., 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.
3. Cros, D., Denis, M., Sánchez, L., Cochard, B., Flori, A., Durand-Gasselin, T., Nouy, B., Omoré, A., Pomiès, V., Riou, V., Suryana, E., Bouvet, J.-M., 2015b. Genomic selection prediction accuracy in a perennial crop: case study of oil palm (Elaeis guineensis Jacq.). Theoretical and Applied Genetics 128, 397-410.
4. Crossa, J., Perez, P., Hickey, J., Burgueno, J., Ornella, L., Ceron-Rojas, J., Zhang, X., Dreisigacker, S., Babu, R., Li, Y., Bonnett, D., Mathews, K., 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity 112, 48-60.

5. 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.
6. Hayes, B.J., Bowman, P.J., Chamberlain, A.J., Goddard, M.E., 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. Journal of Dairy Science 92, 433-443.
7. Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, K., Huang, T., Zhou, T., Jing, Y., Li, W., Lin, Z., Buckler, E.S., Qian, Q., Zhang, Q.-F., Li, J., Han, B., 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42, 961-967.
8. Jia, G., Huang, X., Zhi, H., Zhao, Y., Zhao, Q., Li, W., Chai, Y., Yang, L., Liu, K., Lu, H., Zhu, C., Lu, Y., Zhou, C., Fan, D., Weng, Q., Guo, Y., Huang, T., Zhang, L., Lu, T., Feng, Q., Hao, H., Liu, H., Lu, P., Zhang, N., Li, Y., Guo, E., Wang, S., Wang, S., Liu, J., Zhang, W., Chen, G., Zhang, B., Li, W., Wang, Y., Li, H., Zhao, B., Li, J., Diao, X., Han, B., 2013. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (Setaria italica). Nat Genet 45, 957-961.
9. Kwong, Q.B., Ong, A.L., Teh, C.K., Chew, F.T., Tammi, M., Mayes, S., Kulaveerasingam, H., Yeoh, S.H., Harikrishna, J.A., Appleton, D.R., 2017. Genomic Selection in Commercial Perennial Crops: Applicability and Improvement in Oil Palm (Elaeis guineensis Jacq.). Scientific Reports 7, 2872.
10. Kwong, Qi B., Teh, Chee K., Ong, Ai L., Heng, Huey Y., Lee, Heng L., Mohamed, M., Low, Joel Z.-B., Apparow, S., Chew, Fook T., Mayes, S., Kulaveerasingam, H., Tammi, M., Appleton, David R., 2016. Development and Validation of a High-Density SNP Genotyping Array for African Oil Palm. Molecular Plant 9, 1132-1141.
11. Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., Han, Y., Chai, Y., Guo, T., Yang, N., Liu, J., Warburton, M.L., Cheng, Y., Hao, X., Zhang, P., Zhao, J., Liu, Y., Wang, G., Li, J., Yan, J., 2013. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet 45, 43-50.
12. Murphy, D.J., 2014. The future of oil palm as a major global crop: Opportunity and challenges. . Journal of Oil Palm Research 26, 1-24.
13. Oil World, 2016. Oil World, Oil World Annual. ISTA Mielke, Hamburg.
14. 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., Omar, W., Murad, A.J., Lakey, N., Ordway, J.M., Favello, A., Budiman, M.A., Van Brunt, A., Beil, M., Leininger, M.T., Jiang, N., Smith, S.W., Brown, C.R., Kuek, A.C.S., Bahrain, S., Hoynes-O'Connor, A., Nguyen, A.Y., Chaudhari, H.G., Shah, S.A., Choo, Y.-M., Sambanthamurthi, R., Singh, R., 2016. Non-tenera Contamination and the Economic Impact of SHELL Genetic Testing in the Malaysian Independent Oil Palm Industry. Frontiers in Plant Science 7.
15. 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., Chan, K.-L., Halim, M.A., Azizi, N., Lakey, N., Smith, S.W., Budiman, M.A., Hogan, M., Bacher, B., Van Brunt, A., Wang, C., Ordway, J.M., Sambanthamurthi, R., Martienssen, R.A., 2013. The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. Nature 500, 340-344.
16. Singh, R., Low, E.T.L., Ooi, L.C.L., Ong-Abdullah, M., Nookiah, R., Ting, N.C., Marjuni, M., Chan, P.L., Ithnin, M., Manaf, M.A.A., Nagappan, J., Chan, K.L., Rosli, R., Halim, M.A., Azizi, N., Budiman, M.A., Lakey, N., Bacher, B., Van Brunt, A., Wang, C., Hogan, M., He, D., MacDonald, J.D., Smith, S.W., Ordway, J.M.,

Martienssen, R.A., Sambanthamurthi, R., 2014. The oil palm VIRESCENS gene controls fruit colour and encodes a R2R3-MYB. Nat Commun 5.

17. Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., Atlin, G., Jannink, J.-L., McCouch, S.R., 2015. Correction: Genomic Selection and Association Mapping in Rice (Oryza sativa): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. PLOS Genetics 11, e1005350.
18. Teh, C.-K., Muaz, S.D., Tangaya, P., Fong, P.-Y., Ong, A.-L., Mayes, S., Chew, F.-T., Kulaveerasingam, H., Appleton, D., 2017. Characterizing haploinsufficiency of SHELL gene to improve fruit form prediction in introgressive hybrids of oil palm. Scientific Reports 7, 3118.
19. Teh, C.-K., Ong, A.-L., Kwong, Q.-B., Apparow, S., Chew, F.-T., Mayes, S., Mohamed, M., Appleton, D., 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.
20. Wong, C.K., Bernardo, R., 2008. Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor Appl Genet 116.

SESSION 2

Using Near Infrared Spectroscopy (NIRS) for Rapid Determination of Oil Content in Dried Oil Palm Mesocarp

Sudarno[1]*; Divo D. Silalahi[1]; F. Davrieux[2]; Yong Yit Yuan[1] and Jean P. Caliman[1]

ABSTRACT

The bunch analysis technique is widely used for studying bunch and fruit traits of oil palm trees by oil palm breeders. The bunch analysis method as first proposed by Blaak et. al. (1963) has been reviewed many times over the years but remained basically the same with minor modifications, especially in the sub-sampling procedures. The whole procedure is laborious and time-consuming. One of the important stages of bunch analysis is the oil content determination which is traditionally conducted using Soxhlet extraction method. Due to the toxicity and harmfulness of N-Hexane as employed in the Soxhlet oil extraction procedure, an alternative method more environment-friendly and rapid will be most welcomed by oil palm breeders. In this paper, we share our experience in the development and employment of near infrared reflectance spectroscopy (NIRS) for rapid, high accuracy, precision and minimal chemical risk method for the determination of oil content of dried oil palm mesocarp. A total of 2,891samples were collected during four years (2013 to2016). The oil content of each sample was determined by the conventional Soxhlet extraction method and, correspondingly, its corresponding NIRS reflectance obtained by scanning using NIRS instrument (XDS –FOSS). Samples of varying oil contents ranging from 1.3 and 88.0% on dry matter basis were used. These data were further analysed and processed chemo-metrically to develop and test various calibration models. By using modified partial least square (MPLS) regression, data sets of each year (2013, 2014 and 2015) and its combination were used to generate four calibration models i.e. K-13, K-14, K-15 and K-13-14-15. For model assessment, external validation was conducted using samples of 2016. Oil contents of validation samples determined by Soxhlet extraction and those generated by NIRS calibration models were compared and studied statistically. The results showed that the model generated from the combination data sets (K-13-14-15) provided the best fit with r^2 = 0.977, SEP= 1.42, and RPDp = 6.27. The correlation between reference and predicted values of validation samples were in agreement with high linearity, accuracy and robustness. The results demonstrated that NIR spectroscopy can be used as an alternative. rapid and reliable technique to estimate the oil content in the mesocarp. NIRS is being routinely used in our bunch analysis laboratory as an alternative to the Soxhlet oil extraction method which is currently being employed in most bunch analysis laboratories.

Keywords: *Oil content, mesocarp, oil palm fruit, NIRS*

[1]SMART Research Institute (SMARTRI), PT SMART Tbk, Riau, Jl. Teuku Umar 19, Pekanbaru 28112 Riau, Indonesia.
[2]CIRAD, Performance of Tropical Production and Processing Systems Department, UMR QUALISUD, Montpellier Cedex 5, France.

INTRODUCTION

Oil palm is an important crop cultivated in the tropical and sub-tropical regions. Two different species, namely *Elaeis guineensis Jacq.* and *Elaeis oleifera,* originally from West Africa and Latin Amerika, respectively, are commercially cultivated[1]. Compared to other commodities such as soybean, rapeseed and canola[2, 3], oil palm (*E. guineensis)* is known to be the highest oil producer with an average plantation yield 3.68 tonnes ha^{-1} yr^{-1}. Currently, Indonesia and Malaysia are the top producers of palm oil, collectively supplying about 57 million tonnes of oil which represent more than 85% of the world supply of palm oil or almost 35% of the entire production of all vegetable oils[4]. However, it was projected that the human population will reach about 9-10 billion in 2050 and the world demand of palm oil about 240 Million tonnes[5, 6]. To deal with the escalating demand, efforts must be made to improve oil production by increasing the potential yield above 10 tonnes ha^{-1} yr^{-1}. Research and development are required in oil palm breeding to develop superior varieties of oil palm with superior bunch and fruit component that collectively contribute to high oil yield.

Bunch analysis, first developed by Blaak *et al. in* 1963[7], is currently being used for studying the bunch and fruit traits of oil palm trees. The method has been reviewed and modifications recommended by Rao *et al*. in 1983[8] has been generally adopted by plant breeders[9]. A principal step of the bunch analysis is oil content determination in dried oil palm mesocarp that was conventionally conducted through Soxhlet extraction method using hexane as a solvent. The conventional method is considerably less efficient since it is time consuming, harmful and highly toxic due to n-hexane usage. An alternative and more sustainable method with advance technology is most welcomed to meet the demand of breeding trials. Near infrared reflectance spectroscopy (NIRS) is expected to be one of the most reliable techniques replacing the Soxhlet extraction method in oil content determination of oil palm mesocarp.

In the last few years, NIRS has been introduced and applied as a secondary measurement method in wide ranging applications such as determination of quality parameters in food and feed industries e.g.: rice[10], winter grain and maize[11]. NIRS was also successfully used for ripeness assessment of palm oil fruit[12]. NIRS method provides opportunities in terms of simpler sample preparation, faster analysis and no chemical usage. Due to the superiority and benefits of NIRS, this study was initiated and aimed at developing NIRS as a robust, rapid and routine method for determination of oil content in dried oil palm mesocarp (O/DM).

MATERIALS AND METHOD

Oil Palm Mesocarp Sample

A total of 2891 samples of oil palm fruits derived from various planting materials and maturity levels were collected from several plantations of PT SMART Tbk. (Sinar Mas Agribusiness Resources and Technology) during four years (2013, 2014, 2015 and 2016). The preparation of samples followed the bunch analysis procedure with randomised selection

of fruits. Each set of fruit sample was cleaned from any foreign materials (dust, dirt etc.) and subsequently the mesocarp peeled and sliced from the nut using a sharp knife. The mesocarp was dried in oven at 80°C for 24 h and ground using a laboratory blender (Waring)[13]. The samples were dried again in oven in order to remove the remaining moisture absorbed during preparation before analysis.

Laboratory Analysis of Oil to Dry Mesocarp (O/DM)

The oil content of dried mesocarp (O/DM) was determined using the Soxhlet extraction. Approximately 2.5g of sample was weighed and packed into a stapled filter-paper sachet (Whatman grade 91[14]). The packed samples were inserted into an extraction column of Soxhlet apparatus with 5000 ml capacity, followed by oil extraction with refluxing system using n-Hexane for 10 h. The extracted samples were then dried overnight at 80°C to remove n-hexane. The percentage oil content was calculated on dry matter basis derived from the difference in weight of sample before and after extraction. Standard error of O/DM (%) using Soxhlet extraction method was 0.98%. The distribution of measured O/DM (%) of all samples is shown in Fig.1.

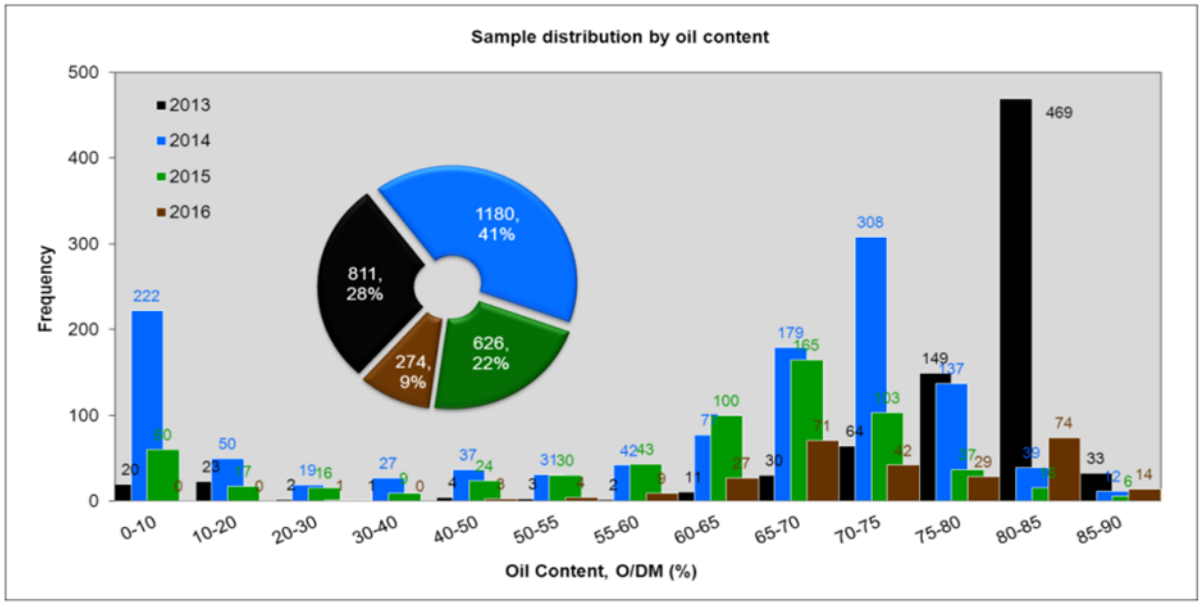


Fig. 1: Sample distribution by oil to dry mesocarp (O/DM).

Spectra Acquisition

Spectral reflectance of samples was acquired using a FOSS NIRS-XDS rapid content analyser[15] and operated with ISIScan IV software v.4.5 (Infrasoft International, port Matilda, USA). Approximately 3 to 5 g of dried-ground sample was inserted into a small ring cup of 36 mm inner diameter. Each sample was scanned three times and between scans the sample was remixed and repacked to get different angle of scan. For each spectrum, 32 scans were acquired, averaged and recorded as reciprocal of reflectance logarithm (log 1/R)[16, 17]. The spectral data with a wavelength range of 400 - 2498 nm at 2 nm intervals were captured. Before used for calibration the spectra of each sample were averaged for associating with the corresponding with O/DM value obtained from the Soxhlet extraction.

NIRS Calibration Process

This study used WinISI software v.4.5 (Infrasoft International, Port Matilda, PA,USA) for developing NIRS calibration. Initially, the spectral data were analysed through principal component analysis (PCA) with a mathematical treatment of second derivatisation (2:5:5:1) and a scatter correction of standard normal variate detrend (SNV-D) for sample discrimination. This study defined the normal data population as selected samples with Mahalanobis distance (GH value) ≤ 3and the outlier samples with GH value >3 as followed in previous studies[10, 17]. The obtained selected samples were then classified by year of sampling and divided randomly into calibration and validation data sets (Table 2). The samples classified as calibration sets were then used for developing NIRS calibration.

Modified partial least square regression (MPLS) method was applied to develop the NIRS calibration models for predicting O/DM of dried mesocarp samples. Initially, some strategies were applied to create calibration equation including wavelength segment selection, derivation and scatter correction using all samples of calibration set. Furthermore, by using the best strategy, three sub calibrations for each year were also developed in order to study how the calibration performance annually. Each developed calibration model consisted of several statistical variables that could be used for evaluation with coefficient of determination for calibration (R^2), root mean standard error of calibration (SEC), standard error of cross-validation (SECV), coefficient of determination in cross-validation (1-VR). Model with R^2 value closest to 1and the smallest standard error can be generally considered as the optimum model. The ratio of standard deviation (SD) to SECV defined as ratio performance to deviation of calibration (RPDc) is also important to be used for the calibration assessment[18]. The generated models were also validated through prediction test using both calibration and validation data sets in order to verify their accuracy and precision between actual and predicted O/DM. Several statistics acquired after the test were used for evaluation including standard error of prediction (SEP), coefficient of determination for prediction (r^2), bias, slope, and ratio performance to deviation of prediction (RPDp)[17]. Similar with RPDc, RPDp is calculated from the SD divided by SEP. The robustness of model can be assessed at a glance based on the RPD value[19]. For natural product, RPD is classified into five levels which are excellent (>4.0),very good (3.5-4.0), good (3.0-3.4), fair (2.5-2.9), poor (2.0-2.4) and very poor (0.0-1.9)[18]. Besides using the calibration and validation data set, the prediction test of each model was also performed using the sample set of 2016.

RESULT AND DISCUSSION

PCA, Sample Classification and Spectra Analysis

PCA was applied for the spectra of raw data set with the aim to get normal data population and to remove outliers. PCA can analyse the complex data variation of the spectra, explain the structure of variables and simplify the data into a model with significant and linearly relevant variables called principal components (PCs). The obtained PCs were then used to identify the population centre and compute the distance from each sample in the population

which is expressed as the Mahalanobis distance (Global H or GH)[12, 17, 20]. The PCA processed 2891 data for sample discrimination and found there were 42 samples which had GH value > 3. Fig.2(a) presents the GH value distribution of all samples. Most samples had GH value between 0-3 (%) expressed as blue colour bar and categorised as normal data population. The remaining sample that had GH value higher than 3 expressed as red bar colour and categorised as outliers.

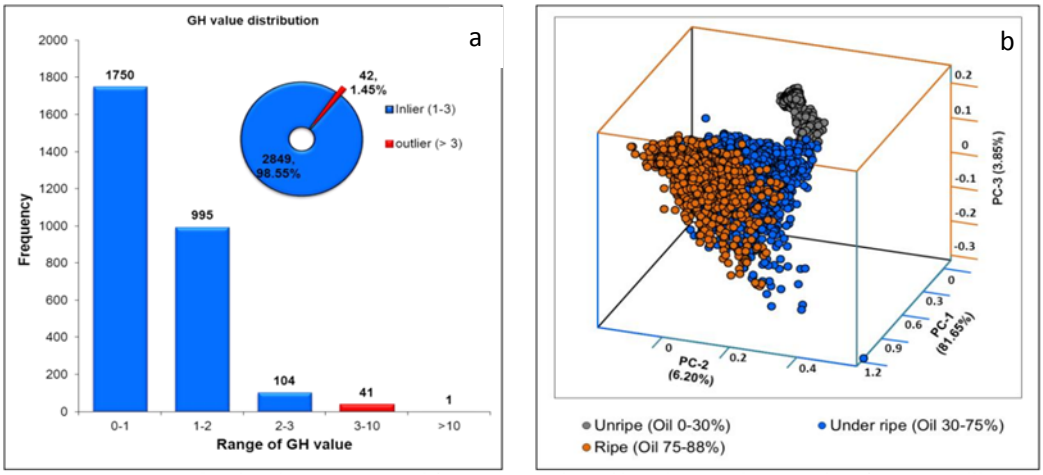


Fig. 2: GH value distribution generated from PCA (a) and 3-D ordination plot of PC 1-2-3 (b).

Each sample also had score value of each PC obtained from the PCA process. Fig.2 (b) shows the 3D ordination plots of score values of all samples using the first three PCs which were classified based on the maturity levels. As can be seen, the first three PCs represent the dominant variability of sample which of PC-1 81.65%, PC-2 6.20 and PC-3 3.85% with the total of 91.70%. For distinguishing the samples, three levels of maturity had been defined: unripe, under ripe and ripe-over ripe with ranges of O/DM content 0-30%, 30-75% and 75-88%, respectively. It is clearly seen that the distribution plots of the samples are clumped in group in accordance with their maturity level. This suggested that there was a linear correlation between the spectral data and O/DM value of mesocarp samples.

TABLE 1: SUMMARY OF OIL TO DRY MESOCARP (O/DM) OF SAMPLE STUDIED.

Sample Classification	Data set	n	Min (%)	Max (%)	Mean (%)	SD[d] (%)	CV[e] (%)	Model name
All samples	Raw	2891	1.32	88.04	62.49	24.31	38.9	-
	Outlier[a]	42	5.16	85.92	35.49	28.96	81.6	-
	Selected	2849	1.32	88.04	62.89	24.01	38.19	-
Data set for NIRS calibration and validation								
Samples of 2013	Cal[b]	567	4.71	86.99	76.38	15	19.64	K-13
	Val[c]	230	7.09	86.05	76.16	16.89	22.17	
Samples of 2014	Cal	794	1.32	86.07	54.21	26.78	49.4	K-14
	Val	361	3.38	87.08	54.2	28.04	51.73	
Samples of 2015	Cal	443	2.07	87.24	57	22.15	38.86	K-15
	Val	182	2.07	84.34	58.43	21.17	36.24	
Combined samples of 2013, 2014 and 2015	Cal	1804	1.32	87.24	61.86	24.58	39.74	K-13-14-15
	Val	773	2.07	87.08	61.73	25.44	41.22	
Samples of 2016	Val	272	28.61	88.04	72.95	72.95	12.52	

[a]Samples with GH value > 3.00, [b]Calibration set, [c] Validation set, [d] Standard deviation, [e]Coefficient variation

Table 1 presents the reference values of oil content of dried mesocarp samples that were analysed in four years; 2013, 2014, 2015 and 2016 with a total sample of 2891 together with their classified data sets. The selected samples represented high variability of O/DM values between 1.32% and 88.04% with standard deviation of 24.31%, suggested that this data was acceptable for further NIRS calibration. The sample of 2013, 2014 and 2015 of the selected data were randomly divided into calibration and validation data sets with a portion about 70% and 30%, respectively. Subsequently the two data sets were separated by years in order to develop annual calibration models. Additionally, samples of 2016 were used for external validation for all developed calibration models. Between calibration and validation set, the range, mean and SD value are relatively similar for each year, with CV value mostly of over 19%, with exception of 2016 samples. Torres *et. al.*[21] has noted that sample size and sample distribution of calibration set is a substantial factor since these influence the range of prediction value. Different SD and CV values of each year data sets indicated different sample distribution as depicted in Fig. 1. As can be seen that the samples dominantly had high O/DM between 60-85% and such sample distribution was a representative of the actual routine sample analysed for bunch analysis. However, those varying data sets were reliable for the NIRS calibration.

Fig. 3(a) shows the averaged spectra of raw near infrared of the mesocarp samples representing the maturity levels. It is clearly seen that each sample class has different absorbance (Log 1/R) which corresponds to the order of maturity level. Samples with lower O/DM values exhibited lower absorbance and vice versa for sample with higher O/DM values showed higher absorbance. This result is in line with the previous study for the spectra of fresh palm fruits[12]. A number of main absorption peaks were observed at wavelengths 1210, 1432, 1725, 1760, 2144, 2306, and2348 nm. Those peaks indicated the oil or fatty acid absorption interaction[22]. In detail, two peak regions between 1110-1298 nm and 1300-1630 nm were designated as the second overtone of C-H stretching vibration and the first overtone

of O-H stretching vibration, respectively. The variation peak around 1725-1760 nm indicated the first overtone of -CH stretching vibration. Moreover, two combination bands were found at 2144nm referring to the stretching bands interaction of C-H and C=C and at 2306 nm ascribing with the interaction of C-H stretching and C-H deformation[23]. Furthermore, some interference including broad peaks, baseline shifts (vertical offsets) and parallel shifts may be due to the raw spectra scattering of samples. Thus, spectra pre-treatment is required to correct and to minimize the spectra data variations which are not correlated with any constituent[24]. This study used math treatment 2,5,5,1 (second derivative with a gap of 5 data point, first and second smoothing over 5 and 1 data points) with scatter correction of standard normal variate and detrend (SNV-D). The spectra after math treatment and scatter correction for mesocarp and kernel samples are shown in Fig.3(b).

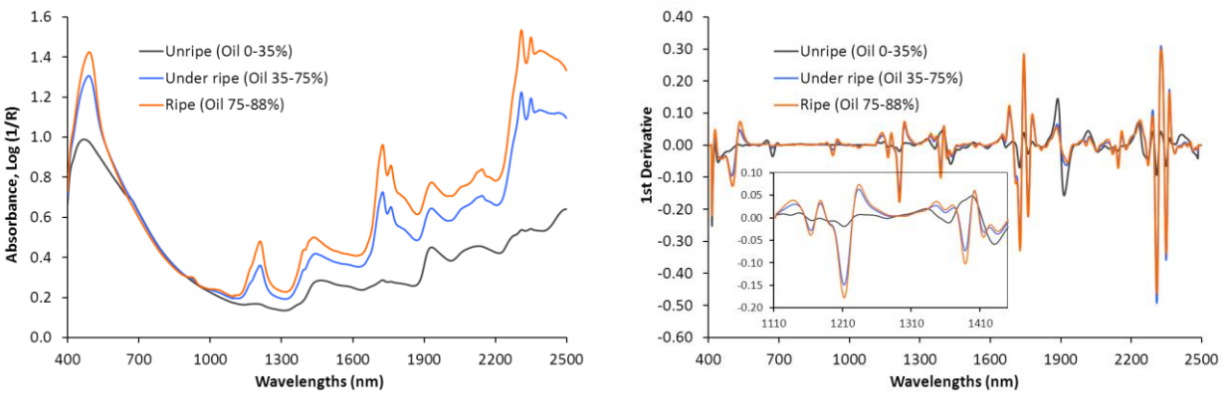


Fig. 3: Raw near infrared spectra (a) and treated spectra (b) of the averaged – maturity level of oil palm mesocarp.

Development of the Calibration Model

The development of the calibration model employed the modified partial least square (MPLS) regression technique. Table 2 presents the initial strategy for getting best models with selecting wavelength region, mathematical treatment and scatter correction. The first strategy used full wavelength segment (400-2500 nm) without any spectra treatment and the output model had bad value of SEC (1.85%), SECV (1.88%) and SEP (2.46%) despite high R^2c value (0.994). All of the strategy generated calibration models with a perfect linearity (R^2 >0.99) and high RPD (> 10). However, the wavelength selection and spectra treatment showed a positive effect on the standard error values. The use of near infrared wavelength only (1110-2488) slightly reduced the SEC and SECV. A previous work reported that spectral information from wavelengths below1100 nm being irrelevant for a robust model[25]. Therefore, the next strategy used the wavelength selection between 1110 – 2488 nm. The use of 1st and 2nd derivation increased the model performance indicated by the decrease of SEC, SECV and increment of R^2, RPD. Significant improvements were observed after applying both spectra pre-treatments of mathematical and scatter correction. However, the best and excellent model was obtained through the use of 2551math treatment and SNVD scatter correction with $R^2$0.997, SEC 1.30, SECV 1.34, SEP1.47 and RPDc 16.65. This corresponded with the previous explanation that math treatment and scatter correction

enhanced the spectra data information by correcting spectra error and improved the model performance[17, 26].

TABLE 2: SUMMARY OF CALIBRATION DEVELOPMENT PROCESS FOR O/DM PREDICTION IN DRIED OIL PALM MESOCARP THROUGH MODIFIED PARTIAL LEAST SQUARE REGRESSION.

Parameter	Regression for NIRS calibration					
	Strategy-1	Strategy-2	Strategy-3	Strategy-4	Strategy-5	Strategy-6
Wavelength segment (nm)	400-2500	1110-2488	1110-2488	1110-2488	1110-2488	1110-2488
Spectra pretreatment	none	none	1551	1551-SNVD	2551	2551-SNVD
No. latent variable (PLS no.)	15	15	15	11	14	11
N[a]	1618	1603	1633	1566	1657	1626
SD (%)	23.09	22.65	23.58	19.26	23.8	22.33
R^{2} [b]	0.994	0.994	0.995	0.996	0.995	0.997
SEC (%)[c]	1.85	1.83	1.68	1.28	1.75	1.30
SECV (%) [d]	1.88	1.86	1.73	1.31	1.82	1.34
RPDc[e]	12.28	12.18	13.63	14.70	13.08	16.65
SEP (%)[f]	2.46	2.47	2.15	1.55	2.32	1.47

[a] Number of sample used for calibration, [b] coefficient of determination for calibration, [c] standard error of calibration,
[d] Standard error of cross-validation, [e] ratio performance to deviation of calibration, [f] Standard error of prediction

The criteria of the best calibration model were then employed to develop sub-models that based on yearly calibration sets. The data used for the yearly models is shown in Table 2. Table 3 displays the statistical results of the generated calibration model by year. Calibration model of 2013 dataset (K-13) has the lowest SD and SEC compared to the other models. This could be due to limited sample distribution processed for the calibration with samples restricted to narrow and high O/DM ranges. The indication was the shorter estimation range between 67-92% with low RPD value 5.311. Consequently, the model was relevant for predicting only sample with high O/DM, despite it was categorised as excellent calibration model. K-14 model showed a significant difference compared to K-13 model. SD of K-14 (25.244) was several times higher than that of K-13 (4.26), as well as for the standard error SEC, SECV. This could be due to more number of samples processed in calibration of K-14 with wider O/DM range which in accordance to the estimation range (0-135%). It should be noted that the estimation range is just a relevance for prediction but the possible O/DM of sample is only between >0% and <90%. Similar result was also observed for K-15 model with slightly lower of SD, SEC, and SECV compared to K-14. Calibration data sets of 2013, 2014 and 2015 were combined and processed to generate into K-13-14-15 calibration model. It could be shown that K-13-14-15 model showed good performance with low standard errors, high RPD and wider range of sample prediction. The result showed that a large number and sample distribution used for calibration influenced the performance of calibration model. As used in statistical analysis, a large number of samples is necessary since this may provide positive effect for developing good calibration[25]. This could be proven by testing those models with sample data set of 2016 which will be discussed. However, all generated calibration models were categorised as excellent calibration with RPD value higher than 4.

TABLE 3: STATISTICAL RESULTS OF CALIBRATION MODEL DEVELOPED FOR OIL TO DRY MESOCARP (O/DM) FROM SEVERAL DATA SETS OF VARIOUS YEARS.

Calibration Model	PLS No.	SD	Estimation range		Calibration		Cross-Validation		RPDc [e]	Model Assessment
			Min	Max	R^2 [a]	SEC [b]	SECV [c]	1-VR [d]		
K-13	9	4.26	67.92	92.92	0.969	0.75	0.80	0.965	5.31	Excellent
K-14	6	25.22	0.00	132.21	0.996	1.53	1.59	0.996	15.84	Excellent
K-15	9	20.45	0.00	119.94	0.996	1.25	1.37	0.996	14.94	Excellent
K-13-14-15	12	22.33	0.00	131.47	0.997	1.30	1.34	0.996	16.65	Excellent

[a]Coefficient of determination for calibration, [b]standard error of calibration,[c]Standard error of cross-validation,
[d]Coefficient of determination in cross validation, [e] Ratio performance to deviation of calibration

Prediction Test Of Calibration Model

The generated calibration models were evaluated in order to know their prediction performance.This evaluation was conducted through prediction test for both calibration and validation datasets. The statistical parameters resulted from the prediction test are summarised in Table 4. The prediction test of sample 2013 using the K-13 model showed a good result with SEP values i.e. 1.84 and 2.02 for the calibration and validation sets, respectively. TheSEP value of the test was two times higher than that of the SECV value of K-13 model. This was due to the high residual value between actual and predicted value of several samples with low O/DM that were not covered in the range of estimation K-13 model.

The SEP values obtained from the prediction tests of K-14and K-15 models were lower than that of K-13. These were likely due to the models have better estimation range that may cover any predicted samples with O/DM and consequently provide less residual value or less SEP value. Such models are reliable for prediction for all sample set that have various O/DM values from the lowest to the highest. RPDp value could also be used as indicator to evaluate the performance of the prediction test. K-13-14-15 model was thus chosen as the best model since it showed good predictive performance, representing the entire sample with the highest RPD value. However, all the prediction results except for K-13, showed good performance with high linearity between actual and predicted values indicated by low bias, slope and r^2 tended to one. Prediction results of each model and each year are presented in Fig. 4. It was clearly seen that there was an empty gap of sample prediction with O/DM between 15-65% observed in prediction of K-13 model (Fig. 4(a)). This is the evidence that K-13 model have limited predictive performance especially for low O/DM sample. Meanwhile, three other prediction models (Fig. 4(b), 4(c) and 4(d)) showed better relationship between reference and predicted values from the lowest to the highest O/DM of samples.

TABLE 4: PREDICTION RESULT OF THE DEVELOPED CALIBRATION MODELS APPLIED FOR VARIOUS DATA SETS

Calibration Model	Sample Year	Data set	n	SEP	Bias	Slope	SD	r^{2} [c]	RPDp [d]
Prediction test for samples of each year data set									
K-13	2013	Cal [a]	567	1.84	-0.4	0.904	13.61	0.994	7.42
		Val [b]	230	2.02	-0.47	0.907	15.35	0.994	7.62
K-14	2014	Cal	794	1.62	0.04	0.997	26.76	0.996	16.50
		Val	361	1.58	0.18	0.997	27.99	0.997	17.69
K-15	2015	Cal	443	1.38	0.07	0.994	22.07	0.996	15.96
		Val	182	1.51	-0.03	1.005	21.34	0.995	14.14
K-13-14-15	2013, 2014, 2015	Cal	1804	1.47	0.05	0.997	24.56	0.996	16.68
		Val	773	1.51	0.05	1.001	25.51	0.996	16.88
Prediction test for samples of 2016									
K-13	2016	Val	272	2.13	-0.97	0.84	7.77	0.976	3.65
K-14	2016	Val	272	1.72	0.07	1.022	9.49	0.968	5.52
K-15	2016	Val	272	1.56	0.55	1.006	9.31	0.975	5.98
K-13-14-15	2016	Val	272	1.42	0.28	0.96	8.87	0.977	6.27

[a]Calibration set, [b]Validation set,[c]Coefficient of determination, [d]Ratio performance to deviation of prediction

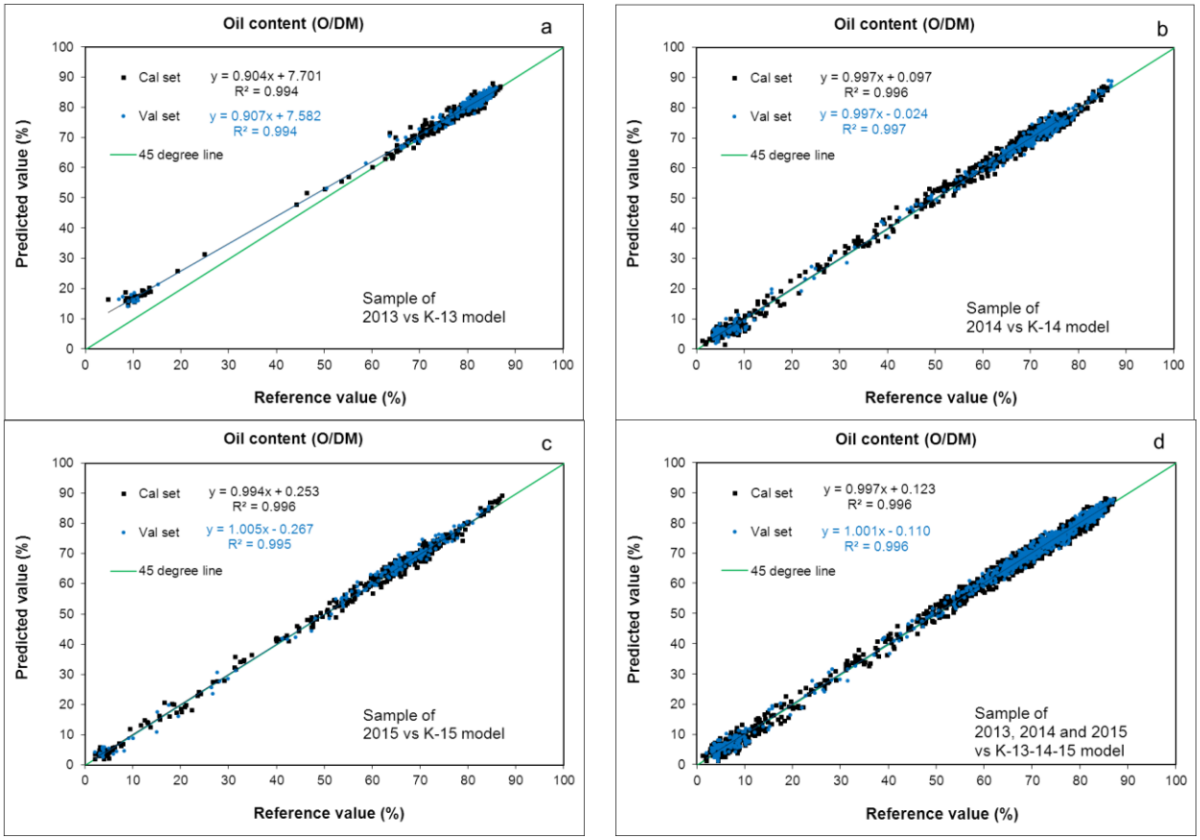


Fig. 4: Regression plot showing the correlation between Soxhlet extraction and NIRS prediction value of calibration and validation data sets; K-13 (a), K-14 (b), K-15 (c) and K-13-14-15 (d) prediction models.

Samples of 2016 used as external validation data set, were also tested to allgenerated models. The statistical results are in Table 4. Gradual enhancements were found from the test using K-13, K-14, K-15 and K-13-13-15 models. The best prediction result of sample 2016 belonged to K-13-14-15 model with SEP, r^2, and RPD were 1.42, 0.977 and 6.27, respectively. In comparison, the obtained SEP (1.42 %) of the external prediction test of K-13-14-15 model was close to its SECV value (1.34 %) and less than 1.5 times to the SEL value (0.98 %). This revealed that K-13-14-15 model is reliable to be used for external prediction of mesocarp samples with all range of O/DM altering the conventional Soxhlet extraction method. Fig. 5(a) shows the regression plot for sample of 2016 prediction using the K-13-14-15 model. Fig. 5(b) presents the comparison diagram of predictive performance of SEP using the four generated models for sample of 2016.

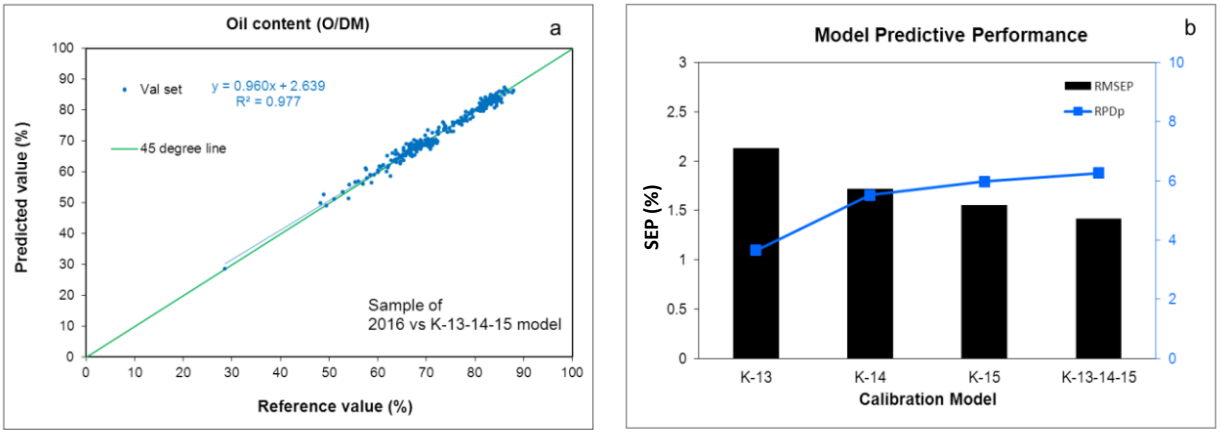


Fig. 5: Regression plot of sample of 2016showing the correlation between Soxhlet extraction and NIRS prediction values using K-13-14-15 model (a) and comparison of predictive performance of SEP using various calibration models

CONCLUSION

This study demonstrated that NIRS calibration models for a robust prediction of oil content (O/DM) in dried oil palm mesocarp could be successfully developed through MPLS regression method. However, the number of samples and O/DM ranges used for calibration influenced the prediction performance, especially the standard error and coefficient of correlation. In this case, calibration model generated from the combination data sets of samples 2013, 2014 and 2015 namely K-13-14-15 was favourably chosen as the best MPLS model. Prediction test of the external validation set of samples of 2016 using K-13-14-15 model showed the best results with lowest SEP and highest RPDp compared to other models. The present study demonstrated that NIRS can be a suitable alternative method used for determining oil content of dried oil palm mesocarp for supporting breeding programs since its beneficial in term of rapid analysis, cost saving and environmental advantages. The NIRS method for such application is currently and routinely being used in our bunch analysis laboratory.

ACKNOWLEDGEMENT

The authors gratefully acknowledge PT. SMART Tbk. management for support and encouragement. While, special thanks are also extended to all laboratory and field team members for making this collaborative study a success.

REFERENCES

1. R.H.V. Corley and P. Tinker, *The oil palm.* 4th ed Blackwell Science, (2007).
2. D.A. do Amaral Teles, M.F. Braga, R. Antoniassi, N.T.V. Junqueira, J.R. Peixoto, and J.V. Malaquias, "Yield Analysis of Oil Palm Cultivated Under Irrigation in the Brazilian Savanna", *Journal of the American Oil Chemists' Society*, **93,** 2 (2016). doi:10.1007/s11746-015-2765-6
3. I.W. Arolu, M.Y. Rafii, M. Marjuni, M.M. Hanafi, Z. Sulaiman, H.A. Rahim, M.I.Z. Abidin, M.D. Amiruddin, A.K. Din, and R. Nookiah, "Breeding of high yielding and dwarf oil palm planting materials using Deli dura × Nigerian pisifera population", *Euphytica*, **213,** 7 (2017). doi:10.1007/s10681-017-1943-z
4. USDA, *Oilseeds: World Markets and Trade* in *Foreign agricultural service, circular series*. 2017, United States Department of Agriculture.
5. D.J. Murphy, "The future of oil palm as a major global crop: opportunities and challenges", *J Oil Palm Res*, **26,** 1 (2014).
6. M. Muhamad Saufi, W. Wan Ishak, and T. LimHo, "Oil palm fruit classifications by using near infrared images", *Research Journal of Applied Sciences, Engineering and Technology*, **7,** 11 (2014).
7. G. Blaak, L.D. Sparnaaij, and T. Menendes, "Breeding and inheritance in oil palm, Part II. Methods of bunch analysis", *J. W. Afr. Inst. Oil Palm Res*, (1963).
8. V. Rao, A.C. Soh, and R.H.V. Corley, *A Critical Reexamination of the Method of Bunch Quality Analysis in Oil Palm Breeding.* P.O.R.I.M., (1983).
9. J. Junaidah, A. Kushairi, B. Jones, L. Kho, Z. Isa, and J. Rusmin, "Innovation for Oil Extraction Method Using NMR in Bunch Analysis", *International Seminar on Breeding for Sustainability in Oil Palm*, (2011).
10. J. Wu and C. Shi, "Calibration model optimization for rice cooking characteristics by near infrared reflectance spectroscopy (NIRS)", *Food Chemistry*, **103,** 3 (2007). doi:https://doi.org/10.1016/j.foodchem.2006.07.063
11. M. Viljoen, T. Brand, D. Brandt, and L. Hoffman, "Prediction of the chemical composition of winter grain and maize with near infrared reflectance spectroscopy", *South African Journal of Plant and Soil*, **22,** 2 (2005). doi:10.1021/jf00005a015
12. D.D. Silalahi, C.E. Reaño, F.P. Lansigan, R.G. Panopio, N.C. Bantayan, F. Davrieux, J. Caliman, Y.Y. Yuan, and Sudarno, "Near infrared spectroscopy: a rapid and non-destructive technique to assess the ripeness of oil palm (Elaeis guineensis Jacq.) fresh fruit", *Journal of Near Infrared Spectroscopy*, **24,** 2 (2016). doi:10.1255/jnirs.1205
13. *Waring Blender*. Available from: http://www.waringcommercialproducts.com/catalog.php?pcID=84_95&products_id=30
14. *Whatman® qualitative filter paper, Grade 91, creped*. Available from: http://www.sigmaaldrich.com/catalog/product/aldrich/z699632?lang=en&region=ID.
15. *FOSS NIR spectroscopy-XDS rapid content analyzer*. Available from: https://www.fossanalytics.com/en.

16. J. Shenk and M. Westerhaus, "Analysis of agriculture and food products by near infrared reflectance spectroscopy", *Infrasoft International, Port Matilda, PA*, (1993).

17. A. Quampah, Z.R. Huang, J.G. Wu, H.Y. Liu, J.R. Li, S.J. Zhu, and C.H. Shi, "Estimation of oil content and fatty acid composition in cottonseed kernel powder using near infrared reflectance spectroscopy", *Journal of the American Oil Chemists' Society*, **89,** 4 (2012). doi:10.1007/s11746-011-1945-2

18. P. Williams, "The RPDstatistic: a tutorial note", *NIR news*, **21,** 1 (2014). doi:10.1255/nirn.1419

19. L.E. Agelet and C.R. Hurburgh Jr, "A tutorial on near infrared spectroscopy and its calibration", *Critical Reviews in Analytical Chemistry*, **40,** 4 (2010). doi:http://dx.doi.org/10.1080/10408347.2010.515468

20. A. Vergnoux, M. Guiliano, Y. Le Dréau, J. Kister, N. Dupuy, and P. Doumenq, "Monitoring of the evolution of an industrial compost and prediction of some compost properties by NIR spectroscopy", *Science of the Total Environment*, **407,** 7 (2009).

21. I. Torres, D. Pérez-Marín, M.-J.D.l. Haba, and M.-T. Sánchez, "Fast and accurate quality assessment of Raf tomatoes using NIRS technology", *Postharvest Biology and Technology*, **107** (2015). doi:http://dx.doi.org/10.1016/j.postharvbio.2015.04.004

22. B. Pérez-Vich, L. Velasco, and J.M. Fernández-Martínez, "Determination of seed oil content and fatty acid composition in sunflower through the analysis of intact seeds, husked seeds, meal and oil by near-infrared reflectance spectroscopy", *Journal of the American Oil Chemists' Society*, **75,** 5 (1998). doi:10.1007/s11746-998-0064-1

23. B.G. Osborne, T. Fearn, and P.H. Hindle, *Practical NIR spectroscopy with applications in food and beverage analysis.* Longman scientific and technical, Harlow, UK (1993).

24. M. Blanco, J. Coello, I. Montoliu, and M.A. Romero, "Orthogonal signal correction in near infrared calibration", *Analytica Chimica Acta*, **434,** 1 (2001). doi:http://dx.doi.org/10.1016/S0003-2670(01)00820-0

25. S.F.S. Matsimbe, S.Y. Motoike, F.A.d.C. Pinto, H.G. Leite, and G.E. Marcatti, "Prediction of oil content in the mesocarp of fruit from the macauba palm using spectrometry", *Revista Ciência Agronômica*, **46,** 1 (2015).

26. L. Salguero-Chaparro, V. Baeten, J.A. Fernández-Pierna, and F. Peña-Rodríguez, "Near infrared spectroscopy (NIRS) for on-line determination of quality parameters in intact olives", *Food Chemistry*, **139,** 1–4 (2013). doi:http://dx.doi.org /10.1016/ j.foodchem.2013.01.002

Applications of Near Infrared Spectroscopy to Determine Oil to Dry Mesocarp, Iodine Value and ß-Carotene for Oil Palm Breeding

Shenyang Chin and Choo Kien Wong

ABSTRACT

The usage of NIRS to determine oil quality parameters for palm oil is not entirely novel as studies have been carried out to determine the suitability of using NIR wavelength for Iodine Value (IV), free fatty acids (FFA), peroxide value, β-carotene, palm oil residue in palm kernel methyl esters and palm oil methyl esters. It has also been used to study oil loss in palm oil mills by analysing residual oil content in press fibre, palm oil sludge, steriliser condensate and empty fruit bunches. The application of NIRS for routine analysis in oil palm breeding is an interesting prospect given (i) the larger range of values obtained in analysis due to genetic variation compared to samples obtained from commercial bulk, (ii) the ability to perform analysis with minimal sample preparation and without the usage of organic solvents, (iii) the advantage of analysing multiple parameters for crude palm oil samples simultaneously. In this study, we developed calibration models for Oil to Dry Mesocarp (ODM), Iodine Value (IV) and β-carotene content using conventional methods as reference data. Furthermore, a validation step was performed using independent set of samples. In brief, the sample number (n), coefficient of determination (r2) for (i) calibration, (ii) validation and (iii) standard error prediction (SEP) were satisfactory i.e. n=573, r2=0.959; n=442, r2=0.889; SEP=1.284 (%) for ODM, n=1238, r2=0.978; n=3051, r2=0.892; SEP=0.978 (unit) for IV and n=728, r2=0.978; n=142, r2=0.952; SEP=54.491 (ppm) for β-carotene.

Advanced Agriecological Research Sdn. Bhd., Jalan Teknologi 3/6, Taman Sains Selangor 1, Kota Damansara 47810 Petaling Jaya, Selangor, Malaysia.

INTRODUCTION

With the arrival of powerful computing ability of consumer grade CPU, availability of calibration/analytical software with interactive graphical user interface and relatively affordable spectroscopy machines, it's now possible to carry out in-house analysis of traits with a single click of a button, without using chemicals that are potential health hazards e.g. n-hexane. The ease of performing analysis using this platform however requires prior effort to develop and to validate calibration models. As with any analysis using secondary method, the accuracy and precision of the analysis is highly dependent on the reference data used in calibration.

MATERIALS AND METHODS

Sample preparation. Mesocarp that were dried and grounded were used to calculate Oil to Dry Mesocarp (ODM) was prepared and calculated according to Blaak et al., (1963) with slight medication using soxhlet extraction. Oil samples were prepared according to Hexane free method Wong et al., (2011), β-carotene content was determined via spectrophotometer following MPOB Test Method p2.6 2006 and iodine value (IV) was sent to an external laboratory for analysis using Wijs method. These methods were used to obtain reference values for both calibration and validation.

NIRS Analysis. Foss NIRS DS2500 was used for analysis. About 28g of mesocarp samples were placed in a small cup holder while ~5ml of oil samples were placed in a slurry cup quartz with a 0.5mm gold reflector placed above the sample. Both samples were scanned from 400-1099.5 and 1100-2499.5 wavelengths at 0.5 intervals.

Calibration and Validation. Calibration was performed using WinISI III Project Manager v1.61 using default values. Modified partial least squares were used develop calibration equation for ODM and IV while β-carotene content was developed using multiple linear regression. Validation was also performed using WinISI III by comparing NIRS predicted value and reference values obtained via conventional methods.

DISCUSSION

The validation results indicate that the calibration (Table 1) was sufficiently robust and suitable for routine use. Guidelines under ISO 12099:2010 (ISO, 2010) states "what is acceptable depends on such criteria as the performance of the reference method, the range covered, and the purpose of the analysis and is up to the parties involved to decide".

In order to compare the performance between the two methods, it is important to determine the standard error of laboratory of the reference method (SEL). As seen below (Table 2) standard error prediction (SEP) for ODM and β-carotene content were slightly higher than SEL while little difference was observed for IV. In the case for β-carotene, the amount of oil used in analysis between the two methods differs by an order of magnitude i.e. 0.1g using spectrophotometer and ~5ml or ~5g using the NIRS; this could be one of the possible reasons for difference between SEL and SEP. However, it's interesting to note that goodness of fit in the validation is highest for β-carotene.

The lower range of values predicted by NIRS for ODM is consistently higher than reference values, this was observed in both the calibration and validation step, and is suspected to be

due to the possible increased hygroscopic nature of mesocarp with lower ODM that has more dry matter content. We also observed better calibration and validation by increasing the number of subscans for ODM as the samples are relatively heterogeneous which also reflected in the standard deviation for subscans ~1 unit (data not shown). Moreover, due to the hyrgroscopic nature of mesocarp (Isa et al., 2011), precautions need to be set in placed to ensure the accuracy of reference data used for calibration. Similarly, NIRS calibration and analysis of β-carotene using spectrophotometer needs to be carried out simultaneously because of the sensitivity of β-carotene degradation during freeze thaw cycles. Contrastingly, calibration for IV using NIRS and analysis using Wijs reagent can be performed at different periods due to the stability of saturated and unsaturated fatty acids.

TABLE 1. LINEAR REGRESSION PARAMETERS FOR CALIBRATION MODEL

Product	Parameters	Sample number	Range (Min)	Range (Max)	r^2
Mesocarp	ODM	573	64.24	89.11	0.959
Crude Palm Oil	IV	1238	46.06	64.08	0.978
	β-carotene	728	62.73	2156.43	0.952

TABLE 2. LINEAR REGRESSION PARAMETERS FOR VALIDATION

Product	Parameters	Sample number	Range (Min)	Range (Max)	r^2	Standard Error Prediction	Standard Error Laboratory
Mesocarp	ODM	442	52.36	84.76	0.889	1.28	1.00
Crude Palm Oil	IV	3051	45.9	68.3	0.892	0.98	1.09
	β-carotene	142	157.02	1315	0.952	54.49	32.94

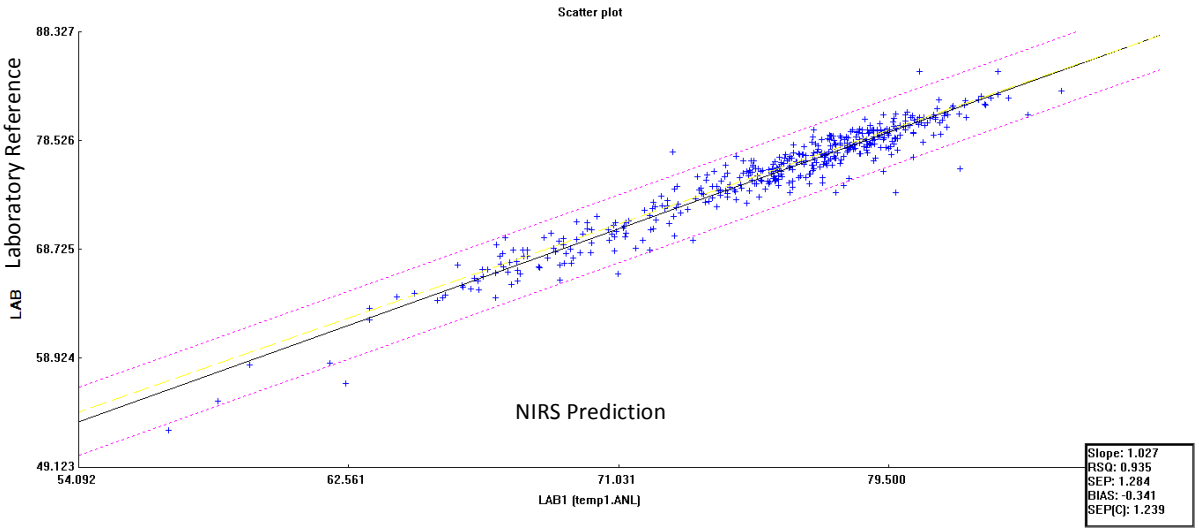


Figure 1. Validation for ODM

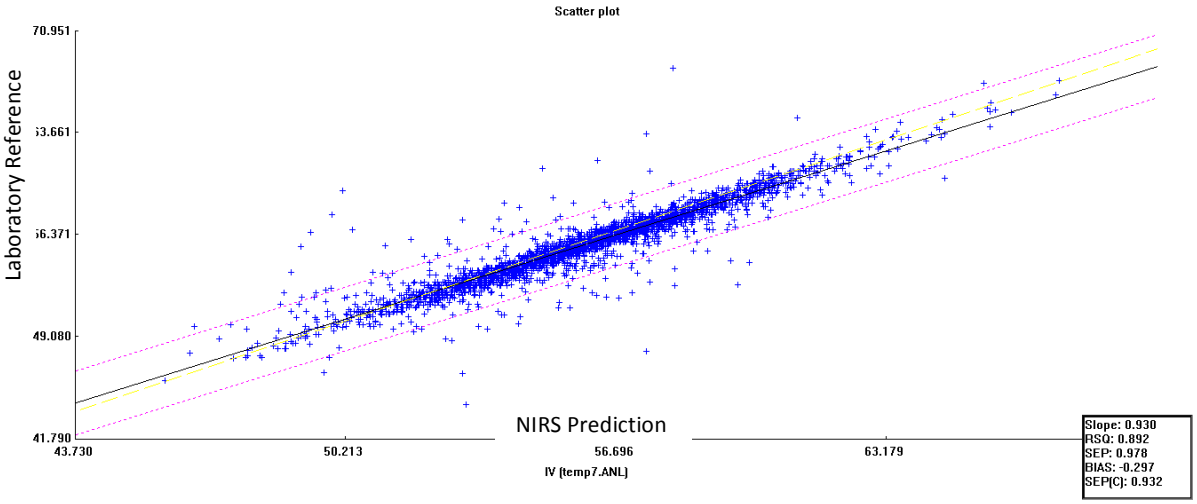


Figure 2. Validation for IV

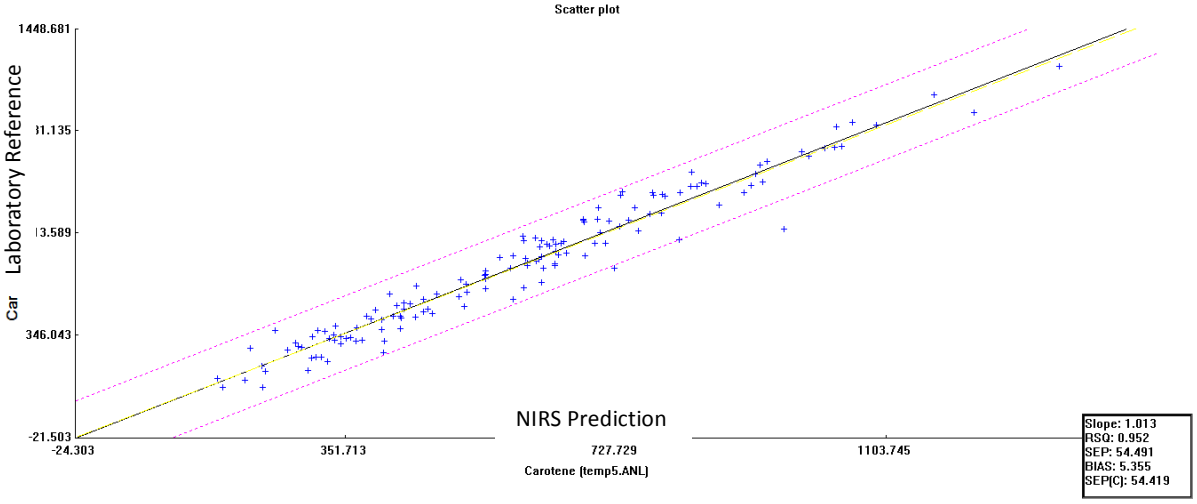


Figure 3. Validation for β-carotene

ACKNOWLEDGEMENT

We would like to thank our Principals Messr. Kuala Lumpur Kepong Bhd., Boustead Plantations Bhd. and AAR Director of Research Mr. Goh Kah Joo for their permission to publish the paper.

REFERENCES

Isa ZA, Kushairi A, Mohd Din A, Suboh O, Junaidah J, Noh A, Choo Kien W and Musa B. (2011). A Critical Re-Examination of the Method of Bunch Analysis in Oil Palm Breeding – An Update. Paper presented at: International Seminar On Breeding For Sustainability In Oil Palm. Kuala Lumpur, 2011. Retrieved from: http://isopb. mpob.gov.my/index.php?r= site/page&view=publication_2011

Wong, C K. and Mariam, M S. (2011). Hexane Free Oil Extraction for Estimation of Iodine Value and Carotene Content. Paper presented at: International Seminar On Breeding For Sustainability In Oil Palm. Kuala Lumpur, 2011. Retrieved from: http://isopb.mpob.gov.my/index.php?r=site/page&view=publication_2011

International Organization for Standardization. (2010). Animal feeding stuffs, cereals and milled cereal products Guidelines for the application of near infrared spectrometry. (ISO Standard No. 12099). Retrieved from: https://www.iso.org/standard/51197.html

Application of Nuclear Magnetic Resonance (NMR) for Oil Extraction in Oil Palm Bunch Analysis - A Critical Testing

Mohd Isa, Z A[1]; Kho, L E[2]; Norziha, A[3] and Nurul Nabilah, Z[1]

ABSTRACT

Analyzing oil content with a pulsed Nuclear Magnetic Resonance (NMR) analyzer is applied in this study to provide breeders with an option for oil analysis with green measurement without any usage of n-hexane. A total of 100 oil palm bunches from various crosses were sampled and analyzed using both Nuclear Magnetic Resonance (NMR) and Blaak's Soxhlet methods. The general objective of this study was to provide other plausible method in the determination of O/DM results in the bunch analysis process. In this study, five replicates per sample were used for both the NMR and soxhlet methods. The utilization of the NMR analyzer in oil content analysis was found very beneficial. Through this study, the results of the NMR method for O/DM was evaluated by comparing the estimated mean values derived from NMR samples against the estimated values obtained by Blaak's Sohxlet method. Besides that, the advantages and disadvantages of soxhlet methods were also critically discussed in this paper. It is hoped that the NMR method could be used in the future and the operators' safety and health will be better taken care of in the long run, cost saving, while the integrity of the data is not compromised.

[1]Kulim (Malaysia) Bhd. P.O. Box 141, 81900 Kota Tinggi, Johor, Malaysia
[2]PLT Scientific Sdn. Bhd., No. 7, Jalan BP 4/8, Bandar Bukit Puchong, 47120, Puchong, Selangor, Malaysia.
[3]Malaysian Palm Oil Board, No. 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia.

Leveraging Agritechnology for Posterity of the Oil Palm Industry

Meilina Ong-Abdullah[1]; Rajinder Singh[1]; Eng-Ti Leslie Low[1]; Mohamad Arif Abd Manaf[1]; Ooi Leslie Cheng-Li[1]; Azizah Mokri[1]; Siew-Eng Ooi[1]; Rajanaidu Nookiah[1]; Norashikin Sarpan[1]; Nuraziyan Azimi[1]; Jared M Ordway[2]; Nan Jiang[2]; Steven W Smith[2]; Nathan Lakey[2]; Ahmad Parveez Ghulam Kadir[1]; Robert A Martienssen[3] and Ravigadevi Sambanthamurthi[1]

ABSTRACT

The year 2017 marks a significant milestone for the Malaysian oil palm industry. The industry has grown from an unassuming ornamental palm to a flourishing commodity in the region and thus, this centennial celebration is indeed special to Malaysia. Aside from the oil palm being blessed with an inherent productive characteristic, the industry's growth has been phenomenal fuelled by global demands for edible oils. Agricultural products are expected to increase by 70% and net exports of oilseeds and vegetable oils are likely to triple by 2050 to meet the projected population growth of ~9 billion. Thus, there is a need to intensify productivity amidst the pressure of a shrinking land resource whilst balancing the environmental and social goals. In this context, agricultural biotechnology provides the necessary tools to address these challenges. The advent of genome technologies and the declining cost of sequencing has tremendously accelerated oil palm genomics research. By leveraging the germplasm resource as well as a rigorous breeding programme, SHELL and VIR genes were discovered and we are now working towards other more complex economically important traits such as height increment, disease resistance and yield. Besides conventional breeding, cloning of elite oil palms is essentially entrenched in breeding programmes as a way to outpace the long breeding cycle of the oil palm which generally results in slow implementation of new improved planting materials into the commercial pool. However, in the course of the tissue culture process, the emergence of a common oil palm fruit somaclonal variant known as mantled, which severely reduces yield, generally halted the progress of the clonal oil palm industry. Through the adoption of a high-resolution epigenome technology, coupled with progress made in oil palm genome research, the team was able to pin down the MANTLED gene that causes mantling. The tissue culture process is also plagued by poor rate of conversion of callus to embryoids. To date, the embryogenesis rate has been hovering at 6% as reported by most tissue culture labs. Hypothesis-driven as well as global approaches have been employed to uncover potential biomarkers that may predict for embryogenesis performance of the selected palm. While some of the findings have since been developed into ready-to-use technologies, some are undergoing refinement and others have yet to be uncovered. Technologies derived from these findings thus far have proven to benefit the industry in living up to its productivity and sustainability goals in the long run. However, its implementation needs to be wide spread for better impact. Developments in agribiotechnology certainly hold considerable promises and for the industry to continue to thrive, it needs to adopt and adapt.

[1]Malaysian Palm Oil Board, No. 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia.
[2]Orion Genomics, 4041 Forest Park Avenue, St Louis, Missouri 63108, USA.
[3]Howard Hughes Medical Institute-Gordon and Betty Moore Foundation, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.

Advancing Genome-based Molecular Precision Agriculture in Oil Palm

Nathan Lakey[1]; Rajinder Singh[2]; Meilina Ong-Abdullah[2]; Eng-Ti Leslie Low[2]; Mohd Arif Manaf[2]; Rajanaidu Nookiah[2]; Steven W Smith[1]; Jared M Ordway[1]; Robert A Martienssen[3] and Ravigadevi Sambanthamurthi[2]

ABSTRACT

In 2013, a team led by MPOB reported the reference genome sequences of E. guineensis *and* E. oleifera*. This milestone rapidly resulted in the identification of the genetic basis of three traits critical to the oil palm industry. The discovery of the VIR gene, responsible for oil palm fruit color, facilitates the use of marker assisted selection (or VIR gene testing) in the development of elite breeding lines with a natural color indicator for fruit ripening. The identification of Karma, a transposon inserted into the MANTLED gene, whose DNA methylation state controls the MANTELD phenotype, paves the way for epigenetic testing of nursery material to identify and cull somaclonal palms destined to yield unproductive abnormal fruits years after field planting. Finally, the discovery of the SHELL gene, responsible for increased yields in tenera hybrids, allows genetic testing and culling of non-*tenera *(low-yielding) palms at the nursery stage, thus enabling the exclusive field planting of high yielding* tenera *palms. SHELL gene testing will also allow for improved efficiency in breeding programmes, especially in the development of male* pisifera *lines. A first study published by MPOB, tested more than 10 000 palms in independent nurseries and smallholdings throughout Malaysia and reported a 10.9% non-*tenera *contamination rate. Expanding upon this earlier work, a current program has tested close to 300 000 additional palms in the commercial sector representing Peninsular Malaysia, Sabah and Sarawak, and found a similar level of contamination. Contrary to common belief, non-tenera contamination was not confined to smallholdings, but was present in similar levels throughout all industry sectors. This work further demonstrates that a novel molecular precision agriculture approach involving SHELL gene testing will result in a significant increase in oil yield from existing planted area. This form of molecular precision agriculture will increase wealth creation among Malaysia's poorest farmers, while also growing plantation profitability, gross national income and national tax revenues. These discoveries made in the Malaysian Oil Palm Genome Project, and more to come, are already helping to achieve sustainability for the most important oil crop worldwide.*

[1]Orion Genomics, 4041 Forest Park Avenue, St Louis, Missouri 63108, USA
[2]Malaysian Palm Oil Board, No. 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia
[3]Howard Hughes Medical Institute-Gordon and Betty Moore Foundation, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

SESSION 3

Utilisation of Barcode System on Seed Production in Socfindo

Indra Syahputra[1]; Dadang Afandi[1]; Chandra Ady Passa[1] and Nicolas Turnbull[2]

ABSTRACT

PT Socfindo has more than 100 years of experience in the oil palm plantation industry, and has contributed to more than 570 million oil palm seeds to plantations both in Indonesia and worldwide. This is equivalent to 2.8 million hectares of oil palm planted. The oil palm breeding work has been ongoingsince 1913 and has led to the current the elite planting material ofSocfindo. The technical quality of seeds produced can lead to a considerable percentage of weak seeds due to uncontrolled proceduresleading topollen contamination. PT Socfindo constantly look forsustainable improvements and innovationsfor its production management system and quality controlin order to guarantee the finest quality of its produced planting material. All the plants used in the progeny test, parental gardens and seed gardens for seed production are checked through biomolecular techniques to confirm the offspring's true from the parents. PT Socfindo implemented Quality Management System (ISO 9001: 2000) in seed production process, ISO 14001: 2004 and OHSAS 18000 since 2002 and 2008 respectively. Furthermore, PT Socfindo has been using barcode information system in every step of its seed production from the field toseed packaging to guarantee the traceability of seeds purchased. Moreover,the barcode systemis also used as a major tool for quality control purposes. The barcode system gives many benefits during data input process, which is faster, has high accuracy, traceability, and with a minimum effective cost.

[1]PT Socfindo Medan PO Box 1254, Medan 20001, North Sumatera, Indonesia,
[2]PalmElit

INTRODUCTION

As one of the prospective sources of non-fossil energy, palm oil is now widely seen by entrepreneurs as a good opportunity for investment. Many companies are expanding their business to palm oil plantations as one of its business support. The prospective palm oil plantation business is characterize by the increasing area of oil palm plantations from year to year-leading toan increasing number of oil palm seeds produced. There are currently 16 oil palm seed producersin Indonesia and expected to be 20 seeds producers by the year 2020.

The development of the oil palm seed industry is facing some challenges:Sustainability of the oil palm plantation industry, the ability to read and satisfy the customer's needs, the capability of creating the best quality planting material, and ability of using efficient resources/technology and human resources capacity (Asmono, 2006).

The Continual improvement of the seed production At PT Socfindo comes through the upgrading of the technical and management systems,the infrastructure, and the efficiency of management and processing using, automation and mechanization to reduce human error.

SOCFINDO BREEDING AND SEED PRODUCTION

a) Development of the breeding programs

PT Socfindo has more than 100 years of experience in the oil palm plantation industry, and has contributed to more than 570 million oil palm seeds to plantations both in Indonesia andworldwide. This is equivalent to 2.8 million hectares of oil palm planted.

The oil palm breeding program at PT Socfindo has started as early as 1913 with the first selections done by Adrien Hallet in Deli Serdang to find the elite planting material used to create the Deli Socfin lines. The Breeding program has been scaled up since the 1970's at Bangun Bandar and Aek Loba estates, now known as Socfindo seeds Production and Laboratories (SSPL) in collaboration with IRHO (Institut de Recherches pour les Huiles et Oleagineux), now PalmElit (CIRAD subsidiaryfocused in breeding and seeds production & sales activities). This partnership has led to the development of many trials and 2 major progeny test projects which enabled theselectionof the best DxP planting material from its plantationsin order to develop its breeding program and elite lines.

b) Current projects

The new millennium has seen the development of many new projects at P.T. Socfindo in collaboration with PalmElit: a new genetic project of 290 ha planted between 2005 and 2012 aimed at testing over 660 crosses; the creation of an extensive collection of palms from all over Africa (Angola, Cameroun, Sierra Leone, Pamol, Binga etc.); the development of a ganoderma screening facility as well as a tissue culture lab and a biomolecular lab. These projects should enable to develop new varieties with unique characteristics in the near future.

c) Varieties of PT Socfindo

Currently, Socfindo provides the best quality DxP seeds with 3 superior varieties: DxP Unggul Socfindo La Mé, DxP Unggul Socfindo Yangambi, and DxP Socfindo MT Gano. This enables to supply specific material with different characteristicsbut allshowing high yields in production.

QUALITY OF SEED PRODUCTION

The good seed producers are those who can copy their best crosses in progeny trials to produce as commercial DxP seeds. To ensure the commercial seeds produced have the same superiority as their progeny trials, the following steps should be taken: Id-Checking, technical production, and quality control.

a) ID-Checking

The first essential pointis to make sure that the superiority of the commercial seeds is at the same level with in the progeny trials; the identity of the parents has to checkedand confirmed. For this, PT Socfindo has implemented an ID-Checking program of verification of legitimacy for all the families used in breeding and seeds production (Turnbull*et al*, 2016). ID checking is use for two purposes (Gasselin *et al*, 2009):

1. For breeding purpose, every parent is controlled.
2. For seed production, every male parent is controlled individually as they are used to produce a large number of seeds. For female parents, legitimacy is controlled at the family level using a significant sample ofpalms (13 is a good number).

Currently, the markers used in PT Socfindo are microsatellite type markers (SSR) developed by N. Billotte (Billotte *et al*, 2002, 2005).

b) Technical Production

Most seeds producer “guarantee” over 99% purity of the seed they produce, but **the reality of that claim** can only truly be judged two to three years after planting when the palm produce identifiable fruit (Gasselin*et al*, 2006). In the seeds production process, the contamination of seeds can occur in the field and the laboratories.

In the field, contaminations will produce illegitimate seeds that will give dura or tenera illegitimate palms. Cross contamination occursafterdamage to the isolation bags, presence of pollen-bearing weevils or insectsduring bagging or development of accompanying male flowers on female inflorescences which may produce viable pollen (Corley, 2005; Ali *et al*, 2014).

On the other hand, contamination in the laboratories will produce illegitimate crosses because of mix-ups between pollens, but this will still give tenera palms. Mistakes can also happen during seedpreparation orgermination process. Finally, it may happen to have confusion between commercial seeds and breeding seeds. To avoid this, commercial and breeding activitymust be separated.

The quality of planting material is related to the technicality and management of seeds production process (Asmono, 2006). To guarantee superiority and quality of the produced

planting material, PT Socfindo follows standard operating procedures in production management system and quality control which is certified. Quality Management System (ISO 9001: 2000) in seed production process, ISO 14001: 2004 and OHSAS 18000 have been implemented since 2002 and 2008 respectively. With those systems, seeds production and management techniques are well executed, consistent, evaluated and continuously improved with innovations.

c) Quality Control

The standard operating procedures setup for the technical production of seeds is monitored through detailed quality control at all levels of work

Socfindo is very strict with the quality control at all the stage of production in the field.In the pollination process in the field; we have 10 quality controls (QC). There are 2 QC in isolation process, 2 QC in the pollination process, 4 QC after pollination process and 2 QC for blank pollination. Controlled pollination is not only done by special QC team (foreman), but by staff and the breeders assisted with our partners.

Quality control at post-harvest is also a major key aspect of the quality of the seeds. It is in place for seed preparation and germination process with the aim to produce the healthiest and high-quality seeds in term of morphology, physiology and genetic purity. This is based on embryo checking, and size control following SNI standards. The purity of the varietiesis ensured by the continuoustraceability of identityusing a specific barcode system.

IMPLEMENTATION OF BARCODE SYSTEM IN PT SOCFINDO

Barcodes are printed horizontal strips of vertical bars used for identifying specific items; a scanning device reads the barcode by moving a beam across the symbol. The first barcode system was developed around the 1940s and 1950s, since then people have become very accustomed to their use, through common applications such as in retail and grocery markets (Cathie, 2004). Wrigley's Juicy Fruit became the first product scanned with a universal product code (UPC), now even traditional Chinese medicine has started using this technology (Cai, 2016).

The benefits that barcode technology brings to the industry are numerous, but can be classified into three main categories:

1. Decrease data-entry errors. Barcodes reduce mistakes to an estimated 1 in 1 million keystrokes – well over 3,000 times the accuracy of a human.
2. Decrease inspection time. When a company has many inspectionsand part programs to choose from, time is wasted searching for the correct one. There is also the potential for human error, leading to an incorrect program or revision level being selected.
3. Increase traceability. When barcodes identify components and materials, traceability is greatly increased. If there are any concerns at a later date, it is simple to traceability.

a) Hardware & Software at PT Socfindo

The system regroups two types of Portable Data Transfer apparatus (PDT) which have a screen and a small keyboard to enter information (picture xxx). Every seed production workers (pollinator), foremen and staff have they own individual PDT which can display different information based on position in the hierarchy.

This PDT is used to feed information to the database using a specific software developed inhouse by connecting to the computer at the end of the day and transferring the data.

b) Quality control

The barcode system was implemented in the field and the germination process for quality control purposes in 2009.

Field system

All the genitors that are used as parents are labeled with a barcode containing the information of identity (selection number), family number and field position of the palm. This barcode label serves as the login (registration) for the quality controls to be performed on each stage of pollination on this individual genitor. This barcode also can be used to control the number of isolated inflorescences and pollinated inflorescences for work follow up and prediction of production.

In the inflorescences, all new bagginghas a barcode label as the bagging label containing the bag id number, the identity of the bagger/pollinator, the identity of genitor and date of bagging. This label will be used for quality control of isolation/bagging process.

In the pollinated bunch, a special barcode label as also used as apollination label that contains thepollination id number, identity of the pollinator, identity of the parents (male and female), category of crossing, and date of pollination. This label will be used for quality control of pollination process and also as id number of the bunchesto follow during the next process.

a. **Barcode label on the Palm (Genitor Barcode Label)**

b. **Quality control recording in the field (barcode & reader)**

c. **Barcode Label on the polinated bunch (Pollination barcode label)**

During QC control, the foreman or staff will first scan the palm label for login using the PDT and then will scan the label of the inflorescence or pollination that is to be controlled. For every differentverification (hole in bag, insects, rotten bunch etc.) a specific code exists that will be entered in the PDT. This information is then grouped in the database at the end of the day by the operator. This enables to have nearly instant reporting of the quality of the work.

Germination system

A specific quality control for water content exists in the germination process. Every bag is labeled with a unique barcode which is used for weekly control of the water content during the heating step. The bags are scanned and then weighed with an electronic scale connected to a computer. A database records the weekly evolution of the water content and links it to the bag identity. This enables to control precisely the quality of heating for optimum germination.

c) Traceability

PT Socfindo started using barcode information system in the field and seed packaging to guarantee the traceability of seeds purchased since 2008.This system includes the seed production and preparation, seeds storage, seeds germination and seeds packaging. For this purpose, the same barcode label used in pollination is also used from the field until seed storage. In the storage, a specific barcode label is added to the sacks containing a certain amount of bunches from a samecrossing category. A last barcode label is used on the storage rack to easily find the desired seed to be process. Once germination batches are made by mixing bunches for the wanted amount of seeds for the germination process, a new barcode label is added in each bag of every batch.

d. Barcode label on the sacks (goni Sacks)

e. Barcode Label on the rack In storage

f. Barcode label in each bag for DXP Lame & DxP MTG

g. Barcode label in each bag for DxP Yangambi

The Barcode label are also used in the packaging and delivery of germinated seeds to customers.

h. Barcode Label on the germinated seeds bag and packaging box

The database enables to follow the information from the batch of germinated seeds sent to the customer all the way back to the bunch used to make this batch. Therefore, there is 100% traceability during all the steps of production.

d) Client supply optimization

All the administration, reporting and document sending are made using data from barcode system. This system gives many benefits such as faster data input, high accuracy, traceability, effective cost and gives high confidence to any customer about the quality of the work in Socfindo. It is part of the image of this company.It enables to prepare seed orders in the shortest time to optimize the speed for sending to customers and insure the genuine quality of Socfindo seeds and minimize counterfeiting.

Companies adopt barcode technology typically custom design a system to fit their needs based on the information they must capture. Surprisingly, only very few companies have carried the technology through to their quality control departments especially for seed production businesses. The application of barcodetechnology to quality control unlocks the potential to improve procedures quality and help error-proof the final steps before products are shippedhand.

CONCLUSION

To guarantee the finest quality of producedplanting material, PT Socfindo always go through sustainable improvement and innovation in technical production, management system and quality control. As previously explained, all the palmsused in progeny tests, parental gardens and seed gardensare checked through biomolecular techniques in order to confirm the offspring's true from the parents. Quality Management System (ISO 9001: 2000) in seed production process, as well as ISO 14001: 2004 and OHSAS 18000 enable to have standard procedures to minimize contamination risks. The barcode information system in the field, germination and seed packaging enables to guarantee the traceability of seeds purchased,as well as having detailed quality control purposes. This system gives many advantages during data input process, which

isfaster, with high accuracy, traceability, and effective cost.Although this methodology works well, Socfindo is already looking to implement a new system using android technology and smartphone as dataloggers. In addition, barcode system might be soon implemented for trials. Nevertheless, for the time being, only strict technical organization and rigorous internal control enable seeds producers to establish a reputation for quality and reliability in this field year after year.

REFERENCES

Ali EOH, Alwee SSRS, Tan SG, Panandam JM,Ling HC, Namasivayam P, Peng HB. 2014.*Illegitimacy and Sibship Assignments in Oil Palm (Elaeisguineensis Jacq.) Half-sib Families Using Single Locus DNA Microsatellite Markers*. Springer Science, Business Media Dordrecht 2014.

Asmono D. 2006. *Development of Oil Palm Seeds Industry by Private Sector in Indonesia*. IOPC 19 – 23 June 2006. Nusa dua Bali, Indonesia. P(1-15).

Cai Y, Li3 X, Wang R,1, Yang Q, Li P, Hu H. 2016. *Quality Traceability System of Traditional Chinese Medicine Based on Two Dimensional Barcode Using Mobile Intelligent Technology*. Plos One. P (1-13).

Cathie ML. 2004. *The Advantages and Disadvantages of Barcodes and Radio Frequency Identification in Supply Chain Management*. University of Wollongong Research Online. 123p.

Corley RHV. 2005. *Illegitimacy in Oil Palm Breeding - a Review*. Journal of Oil Palm Research. Vol.17 (1) pp.64-69.

Gasselin TD, Baskett P, Jacquemard JC, Cochard B. 2006. *Planting Material: What strategy for Planters? DxP Seed is not only DxP Seeds*. IOPC 19 – 23 June 2006. Nusa dua Bali, Indonesia. P(1-12).

Gasselin TD, Billotte N, Pomies V, Mastin G, Potier F, Amblard P, Flori A, Cochard B. 2009. *ID Checking by Microsatellite Type Markers (SSR) During the oil Palm Variety Selection and Production Processes*. ISOPB [International Society for Oil Palm Breeders] Seminar, 04-05 Nov 2009, Kuala Lumpur, Malaysia. P (1-8).

Turnbull N, Cazemar M, Guerin C, Amblard P, Cochard B, Gaselin DT. 2016. *Oil Palm Breeding and Seeds Production in PalmElite*. Proceedings of the ISOPB. 29-30 September 2016. Kisaran, Indonesia. P(8-20).

Data Logger and Its Contributions in Oil Palm Breeding: Sawit Kinabalu's Experience

Jayne, J J and Zuraini, S

ABSTRACT

A data logger or a data recorder is an electronic device that is capable to collect and store data. It is a new device that enhanced field data collection in the oil palm breeding research. The automated data logging software (ADLS) developed was to store the yield recording automatically in the data logger which than customized and integrated with the oil palm breeding software system. Sawit Kinabalu's oil palm breeding research has been collected oil palm breeding data and processed manually since it was started its research in year 1995. The utilization of data logger for the past two years has contributed significant impact to the overall breeding research programme. Not only data collection processes have been very much improved but highly accurate data were able to be generated. Accurate data became an important factor in assessing the success of data information derived in oil palm breeding in order to select parental palms for commercialization. The usage of data logger helps the company increase efficiency and profitability and also reduces costs.

Sawit Kinabalu Seeds Sdn Bhd, Locked Bag 28, 91009 Tawau, Sabah, Malaysia.

INTRODUCTION

Sawit Kinabalu Seeds Sdn. Bhd. (SKSSB) initiated its oil palm breeding research since 1995 with the objective to produce high yielding Sawit Kinabalu D x P planting materials. To ensure its achievable, selections of parental palms based on phenotypic characters consisting of yield, bunch characteristics and vegetative traits are made compulsory. These phenotypic characters are to be made available as it's become an important indicator to the overall performance of our D x P high yielding planting materials which also described by Corley *et al.(2003).*

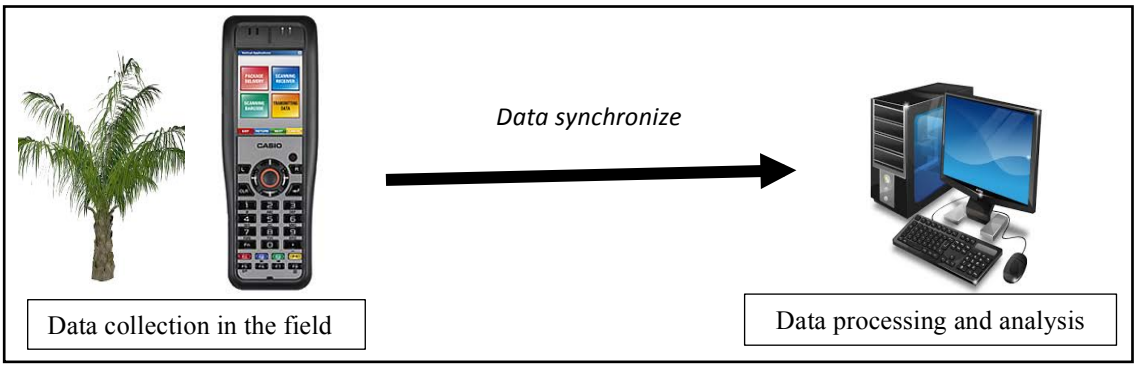
An extensive and comprehensive oil palm breeding selection in SKSSB started with development of germplasm collection from different populations. These germplasm materials are tested in our progeny test trials and continuously evaluated its Good Combining Ability (GCA) to make sure SKSSB stay forefront in its breeding research programme especially during the process identification in the progeny test trials. To-date, five SKSSB's germplasm materials has been tested namely Calabar, Ekona, AVROS, Nigerian, and Deli Duras.

SKSSB's germplasm materials were evaluated based on their individual performances. Among those evaluated are yield characters which are bunch weight (BWT) and bunch number (BNO). Formerly, yield traits were recorded manually. Since breeding programme involved enormous volume database therefore, it may cause data inaccuracy and discrepancies. In addition, manual data collection method requires more labors. Few errors detected in our manual method are wrongly recorded information of palm number, bunch weight and bunch number. Mistakes were found partly due to the attitude of the supervisor and field recorder. Realizing that may cause significant effect to our selection, SKSSB decided to create innovation on our data recording to further improve the current method.

In 2014, SKSSB started its data collection improvement programme by incorporate electronic device into its yield recording activities. The first project electronic device used namely *Casio* data logger (CDL). CDL is device that records data over time or in relation to location either with a built in instrument or sensor or via external instruments. It is small, battery powered, portable, and equipped with a microprocessor, internal memory for data storage, and sensors. Using CDL is an advantage and according to Corley *et al*. (2003) hand-held data logger has obvious advantages in terms of avoiding transcription errors. This paper will share Sawit Kinabalu's experience in using hand-held data logger in our yield recording data collection improvement programme.

METHODOLOGY

The usage of hand-held data logger device in SKSSB was started with yield recording activity. It was built in and customized with hand-held yield data collection system. This system is then being synchronized with the Sawit Kinabalu Yield Recording Data Module in SKSSB Oil Palm Breeding Software. *Picture 1* shows the yield data transmission from the data logger to the software.



Picture 1. Yield data synchronization and transmission

Data Logger

The hand-held data logger used by SKSSB is the one with keypad and screen as shown in *Picture 2*. It is also attached with barcode sensor.

Picture 2. Picture of the model of data logger used by Sawit Kinabalu

The screen provided the user to view instructions by the system and enable recorder to key-in the data information. The keypad consists of menu and numeric buttons. The recorders are required to press the buttons for data entries. Barcode sensor embedded in the device is use to read the palm number by scanning the barcode attached on the palms.

Data Capturing Process

The yield recording (YR) process flow using data logger for data capturing method is shown in *Figure 1* with its details described in **TABLE 1**.

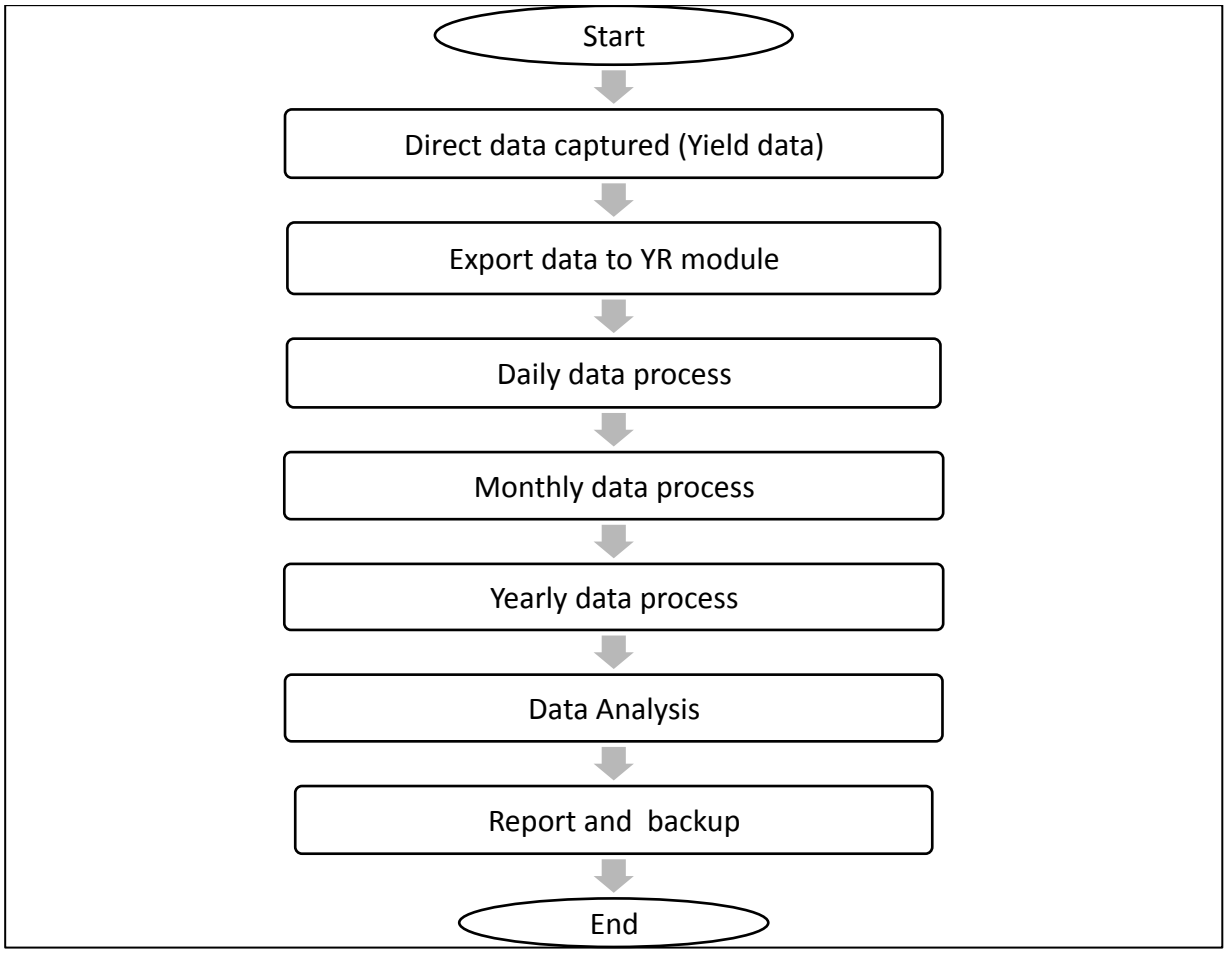
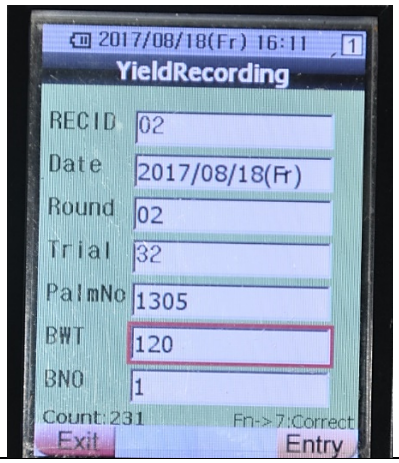
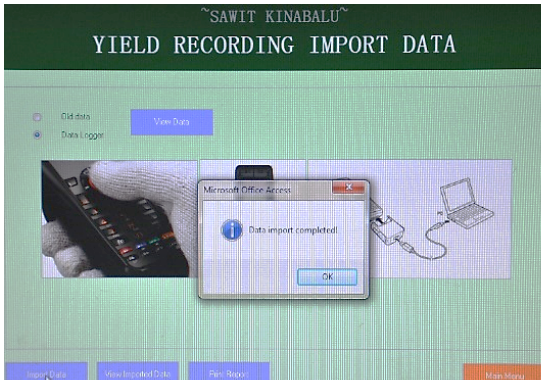
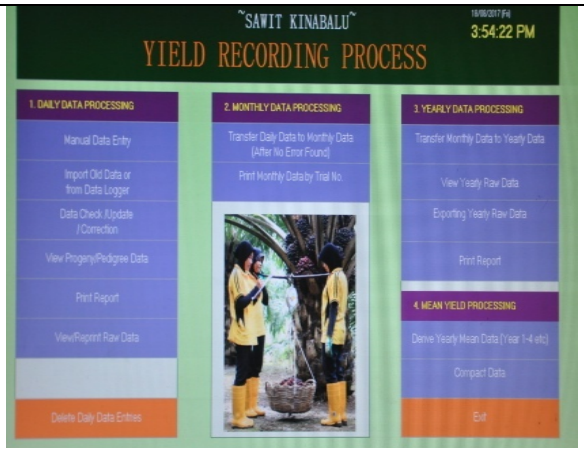


Figure 1. Yield components data flow using hand-held data logger

TABLE 1. TABLE OF YIELD RECORDING PROCESS FLOW DETAILS

No	Picture	Descriptions
1		The software display data information which consists of field recorder identification, date of recording, harvesting round, trial number, palm number, bunch weight and bunch number. The software also automatically recorded number of data that have been recorded daily for each device.
2		The field recorder will scan the barcode that attach to the palm until the palm number appeared in the screen which usually takes about one second.
3		The field recorder will enter the number of bunch and total bunch weight for each individual palm into the data logger.

No	Picture	Details
4		The mobile printer is one of the accessories for the data logger. Field supervisor will bring the mobile printer to the field. The printer equipped with the bluetooth connectivity to the data logger that enables the field supervisor to print the data after yield recording done on that day. The printed data is for the supervisor to do spot check in the field for reliable data.
5		Yield recording is one of the modules in SKSSB Oil Palm Breeding Software. The data in the data logger will be transferred into the YR module in the computer by using the data cable. The computer operator will click on the menu 'Import data from data logger' to transfer the data. The caption "Data Import Completed" will pop-out on the computer screen when the data completely transferred.
6		The oil palm breeding software yield recording module consists of four main menus (Daily data processing, Monthly data processing, yearly data processing, and mean yield processing). The menu enables the breeders to view the data according to their needs.

IMPACTS

The implementation of hand-held data logger has contributed significant impacts to overall SKSSB Breeding data collection as follows:

Reduce Costs

In SKSSB, at least 800 yield recording data were entered to the computer every day. After converted to data logger capturing method, data entries process has been eliminated which finally reduce 85% of overall process involve from initially eight hours to only one hour. The one hour time recorded was use to focus on the checking of outliers data. The total cost of RM 15,000 was saved in managing the recording trials through the reduction of data operator from two to only one. In addition to that, approximately RM 5,000 cost saving was also attained after the cut down of 9,360 papers usage.

Avoid Double Data Entry

The data logger capturing devices has overcome the double data entries by field recorders and data operator since entering process was made once by the field recorders which then immediately transferred to the computer.

Hassle Free

Positive feedback was given by field recorders from the utilization of data logger as follows:

3.3.1 Less materials carried to the field.
3.3.2 No incident of losing data due to missing paper.
3.3.3 Immediate data can be printed, check and corrected in the field.
3.3.4 Less verification requested by data operators.

Increase data accuracy

The usage of barcodes system attach to each palm prevent the occurrence of wrong palm number entered. This is due to the auto-detection and stored function built in that allows capturing of palm number and its basic information to the device by scanning process. Besides accelerating the data recording activity, the barcode system avoid wrongly palm number entered as the field recorders only need to key-in the bunch weight and bunch number information to the device.

Increase Data Efficiency

Efficient handling of the recorded data in the data logger has increased the work productivity from the field records until transferring data to the computer. The data can be viewed immediately by the breeder and data analysis can be done by daily, weekly,

and monthly. Moreover, the recorded data in the hand-held data logger is adaptable to the 'Sawit Kinabalu Oil Palm Breeding Software'.

CONCLUSION

The usage of the hand-held data logger has contributed significant enhancement in SKSSB oil palm breeding research and selection. Not only data collection processes have been improved, highly accurate data were also recorded. The data generated from SKSSB data logger capturing method were adaptable to SKSSB Oil Palm Breeding Software. In addition, simple and user-friendly system built in the data logger permits all levels of education users. The switch from manual to direct data capturing helps the company increase its productivity and efficiency. The positive outcome of this application has now broad up to the second stage of implementations to other activities such as vegetative measurement and seed production.

AKNOWLEDGEMENT

The authors wish to thank the management of Sawit Kinabalu Sdn. Bhd for its permission to present this paper.

REFERENCES

CORLEY,R.H.V. & TINKER, P.B.H. (2003). The Oil Palm (Fourth Edition). p.154-155

ZURAINI,S. JAYNE,J,J, & HOONG,H.W (2013). Exploitation of Sabah Breeding Programme in Enhancing the Performance of Sawit Kinabalu DxP Planting Materials. p.71

FGV Integrated Breeding System (FIBS): Managing Integration of Breeding Data and Operation

Muhamad Farid Abdul Rahim; Lee Yang Ping; Mohd Nasruddin Mohamad; Mohd Latif Kamarudin, Mohd Mahfuz Roslan; Naderman Samin; Nur Adibah Ishak; Suthashinikisan Krishnan; NoorSusilawati Mandangan; Nurul Fatiha Farhana Hanafi; Leao Ling Jiun and Tan Joon Sheong

ABSTRACT

FGV is one of the largest oil palm seed producers in Malaysia generating about 25 million oil palm seeds annually for both local and overseas market. After 5 decades of extensive breeding programme where the first germplasm trial was planted at Felda Taib Andak in early 1960's. Since then, tremendous amount of breeding data has been generated and collected from the field. This continue to grow until today. Various information such as origin, lineage, crossing history, trial information and data were stored manually in multiple locations while the data was processed conventionally. This conventional practise is posing a potential risk to the management who had spent a huge investment in the breeding programme if the storage of data is not archived properly and less effective in data mining. Thus, the advent of information technology is crucial to manage and integrate the breeding data with latest technologies for both operation and breeders' needs. Our breeding software module also known as FGV Integrated Breeding System (FIBS) is an integration of sub-modules from information management, crossing design, trial design, crossing and seed production management, nursery management to field data management to help breeders and operation team to monitor the entire progress of breeding research in efficient and systematic way. The application of information technology in breeding research work is potentially improvise and enable data network integration with an option to integrate with other fields of research such as molecular markers (QTL / gene / protein) information, superior palm selection for ortet and mother palm, agronomy data and crop protection observation and even artificial intelligent system for better understanding of oil palm breeding.

FGV Research and Development, Unit Informatic, FGV Information Centre, Bandar Enstek, Nilai, Negeri Sembilan, Malaysia.

INTRODUCTION

FGV is one of the largest oil palm seed producer in Malaysia generating about 25 million oil palm seed annually for both local and overseas market. After 5 decades of extensive breeding programme where the first germplasm trial was planted at Felda Taib Andak in the early 1960's. Since then, tremendous amount of breeding data had been generated and collected from the field. We foresee this trend will continue to grow in the future too.

Data generated from breeding programmes were involving various genetic backgrounds and characteristics of particular gene pool for breeding research from phenotypic data that was observed from field. From there, the information was further expanded for further improvement and eventually leading towards commercial seed production.

Managing huge data set is never an easy task where every single parameter measured in the field is essential to breeder for decision making. Therefore, managing information generated through breeding research, involving a certain genetic lineage / origins, trial and location of which the palms were planted. From there, advance descendent will be generated through breeding crosses either through selfing, sib mating or intercrossing is very dependent on the purpose of improvement. Improper storage or archiving of information might post some difficulties for tracking a lineage / ancestor / origin.

The growth of data in breeding program such as in Apple breeding program by The Washington State University, have demand on developing an efficient database system for perennial crops to manage increasing number of data generated through yearly field observation instead of using ready-made application in the market such as Agrobase that focussing on annual crops like maize. (*Evans et al*., 2013)

Also, the rapid growth of genotypic studies and advancement of DNA sequencing available to be access through the web, thus, advancement of technologies for phenotypic data acquisition is a must to have better understanding phenotypic-genotypic relation to give better understanding and interpretation to get better crops for the future. (*Rahaman* et al. 2015)

Different research mode between perennial crop and annual crops in term of life span and years of observation makes oil palm a unique path of study fields. Thus, the need for information system technology customized with research nature will optimize the function and usage of the database system.

In Felda Global Ventures Research & Developments Sdn Bhd (FGVRD), we have developed an oil palm breeding database that has the capability to map all information such as origin, parental, trial information as well as the performance observed in the field not limited to yield, bunch analysis, vegetative growth and census but also to digitally map research plots that representing the geo-location of the breeding trials or crosses. All of this can be linked together to be essential breeding information and performance can be accessed and evaluated in a more efficient way as compared to conventional method mostly interpreted from spreadsheet or excel.

From there, breeder will go through selection of parental palm to generate new generation of germplasm, introgression of germplasm for manipulating traits of interest or even for cloning purpose. These parental materials will be used to produce new progenies for future improvement or evaluation. It will go through seed production process which involve flower census and bagging process, male flower process and pollen store, control pollination process, seed process and seed store, seed germination process and nursery planting. Figure 1 shows a complex workflow of seed production start from flower census and bagging until nursey planted. Each process need to record its date for traceability especially pollen and seed store to track the movement of usage.

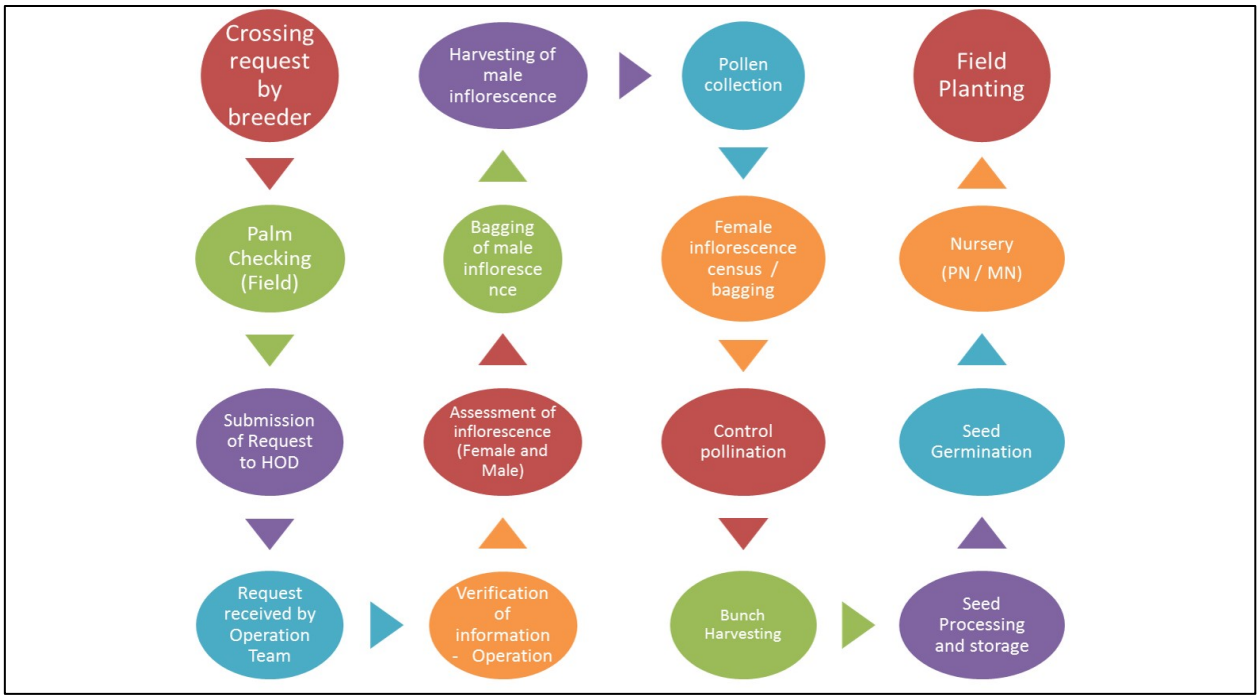


Figure 1.0. Complex workflow involve with seed production process.

Thus, the needs of system to manage and monitoring all the process is essential and at the same time connected the information with the breeding data and information. It will help breeder in monitoring their new crosses created through their selection and monitoring the progress of seed production to be ready for planted new trial at field. This function will improve breeder's time management for preparing new trial planting.

Oil palm breeding programs generate large amount of data and, data require for computation from diverse sources. There are commercial available database packages for other crops, for example, Agrobase®.

However customized workflow is required for the timely and labour intensive activities like oil palm breeding. Hundreds of crosses that producing thousands of seeds are created every year.

These seed go through several rounds of selection. About 96 seedlings per progeny are finally planted in field for experimental evaluation.

The data recording per trial usually takes up to 8 years for several traits collection. Environmental data is also crucial to evaluate the individual palm response to environmental effect. More than 20 types of data are collected. Readily access required data is getting significant to fulfil the increase need for in-house genomic tool development. An integrated breeding software and data management platform is needed for breeding research at FGVRD for experimental design, operation management, database for storage and quality check to data interpretation for breeder's general needs. It will increase the understanding to the complexity of oil palm physiology and its relationship with the environment and increase the probability of developing new varieties.

OBJECTIVES

i) To develop database system for better management and storage of breeding information.
ii) To develop a monitoring system for seed production link with breeding information.

APPROACH

The development of database system for oil palm breeding was largely focused on three main components, namely, (i) information management that tackle the origin of our germplasm collections and lineages; (ii) breeding operation process monitoring for cross creation for breeding seed production and trial creation for trial study ; and (iii) research data management for breeding field data collection and management. Based on these main requirements, a web-based interface was created that communicated with a relational database that linked all the information of any given breeding crosses and its origins or lineages. All essential information was linked and available at breeders' fingertips instantly.

Breeding - Data and selection

Identifying parameters involved such as origin background, trial information, location, and lineage for any given germplasms. Managing the newly selected parental materials from germplasm line for further improvement. All information need to be linked together through generation from ancestor to its descendent either inbred or out cross.

Breeding Seed Production Process

Once a breeder had identified any potential palm(s) to be selected as parental line based on its known trait or yield performance, conventional breeding method such as NCM I, NCM II, MRS and etc. can be deployed. Operational team will go through the process starting from palm and flower census, bagging, control pollination, seed process, seed germination process, nursery process and field planting. There are several parameters to be observed / recorded prior to quality

checking before proceeding to the next stage. It is a continuous process to be monitored and two critical processes are monitored (e.g.; pollen storage and fresh seed storage).

Breeding - Field Data

There are several set of parameters being observed in the field such as yield monitoring particularly sex ratio, bunch weigh and bunch number, bunch analysis consist of collecting raw data from lab and calculated from listed formula, vegetative growth involve several observation such as height, rachis length, number of leaflet and other vegetative anomalies. All of the collected data must be correlated back with an individual palm to better understand its specific and general combining ability as well as the any additive effect observed.

RESULTS

FGV Integrated Breeding System (FIBS)

There are three main components (Figure 2) in FGV Integrated Breeding System structure: (i) Information Management; (ii) Breeding Operation Process Monitoring; (iii) Research Data Management. These components are linked with each other in our relational database to form a complicated network of information for better and faster interpretation.

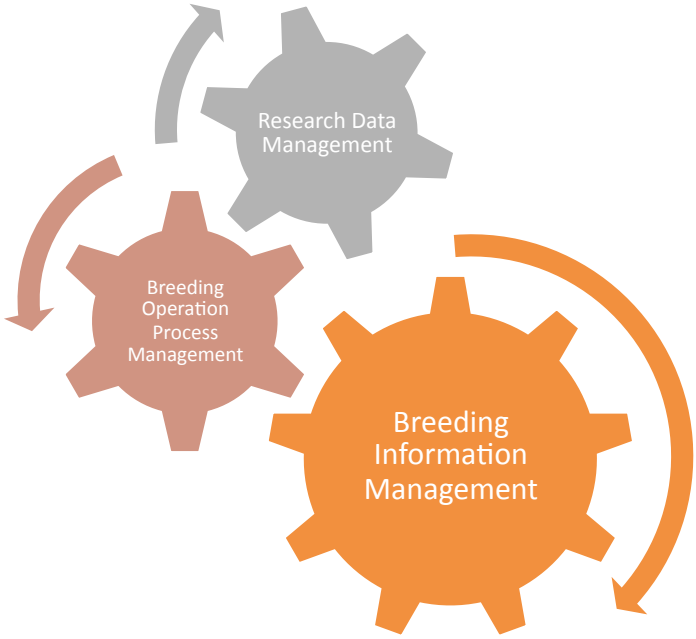


Figure 2.0. Three main components of FIBS that are integrated together to provide a single point of reference for oil palm breeding research.

Information management (IM) component are built based on managing the germplasm origin, parental palm, project and trial information to producing structure and useful information for analysis. While breeding operation process management was processing the information from information management and generating new information later will be used for further breeding program by breeders.

Next, breeding operation (BO) process component involves step where crossing design and trial layout by breeder and execution by operation team involving processes such as control pollination, seed production and field planting. This is to ensure project timeline can be monitored and executed accordingly by both breeders and operation.

While research data management (RDM) function to manage phenotypic data where it will be stored in a database system prior to any analysis. On the other hand, it also allow better planning for field activities with regards to data collection without missing any important census within observation plots.

The information linkage between these three components are being identified to ensure all data collected will be fully utilised for better understanding and interpretation of specific breed behaviour and its interaction with others.

Component I: FIBS Breeding Information Management System

Breeding information have a complex structure consist of several parameter data that linked together to form information involve breeding. For an example, there are various origins such as Deli-Banting, Deli-Serdang, Deli- IRHO, Yangambi, Avros and Lá Mé. These data will correlate with a specific trial information and its respective pedigrees information.

By using associative entity method in database system, FIBS can generate information from different set of data. Figure 3 shows how the relationship are created:

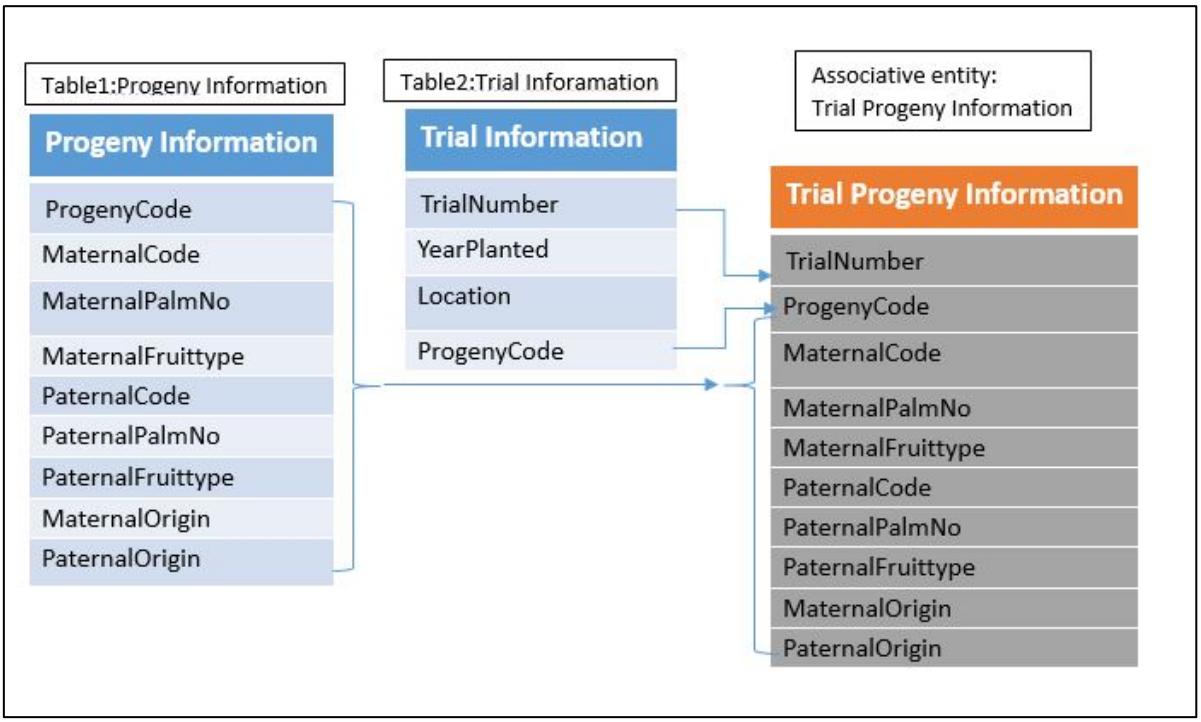


Figure 3.0. Relationship of different attributes and parameters progeny information.

From the Figure 3, Table 1 for Progeny information and Table 2 was Trial information, from there, system will help to construct relationship between the two entities from different data table to organize as one useful information for breeder. Through this technique, breeder can group a set of data and relate it with one another easily by using primary key as connection between tables instead of using manual technique in spreadsheet. This will avoid human error during data information gathering manually while increasing productivity and efficiency in data processing.

These features are useful during studies between background of progenies or germplasms before creating new crosses for further evaluation and analysis. By using this method, it will enable breeder to access all relevant and essential information at ease.

In additional to this, other modules in the system can be developed such as Project and Trial Management, Inventory for Parental and ortets palm, Crossing Matrix, progeny tree construction and linked with two main modules. It will train us to standardize the record we have in one format and it can be linked with all progenies starting from its initial genetic lineage all the way to the current breeding cycles within FGV. For instance, 5 decades of oil palm breeding improvement span across 3 - 4 of breeding cycles as shown in Figure 4 and 5.

+ Add Progeny

Query | Progeny Code | Search

Show 10 entries

Progeny Code	Origin	Maternal Code	Maternal Palm No	Maternal Fruit Type	Paternal Code	Paternal Palm No	Paternal Fruit Type	Trial	Progeny Tree	Action
TK991	OLEIFERA-KLM-KULAI-PISIFERA-Deli-IRHO-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi	INB	30	T	ML	161	P	-	View	Delete
TK992	OLEIFERA-KLM-KULAI-PISIFERA-Deli-IRHO-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi	INB	30	T	GMH	43	P	T236	View	Delete
TK992	OLEIFERA-KLM-KULAI-PISIFERA-Deli-IRHO-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi	INB	30	T	GMH	43	P	-	View	Delete
TK992	OLEIFERA-KLM-KULAI-PISIFERA-Deli-IRHO-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi	INB	30	T	GMH	43	P	T245	View	Delete
TK992	OLEIFERA-KLM-KULAI-PISIFERA-Deli-IRHO-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi	INB	30	T	GMH	43	P	-	View	Delete
TK993	Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi	IHP	18	T	FQ2	57	T	T236	View	Delete
TK993	Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi	IHP	18	T	FQ2	57	T	-	View	Delete
	Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-Irho-Yangambi-									

Figure 4.0. Inventory of Progeny complete with origin and trial information.

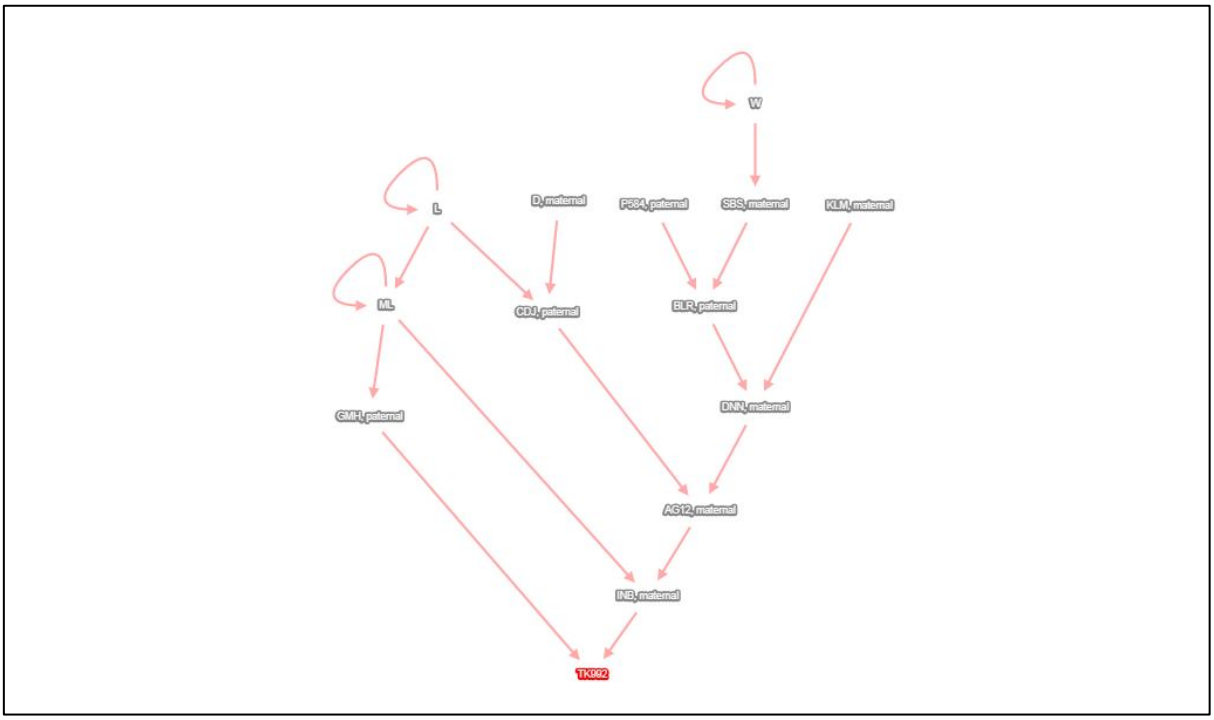


Figure 5.0. Progeny Tree function that has been construct through information in database.

Besides that, another important function for breeders is to monitor the entire crossing design (NCMI, NCMII or others depending on their objectives). By using the FIBS system, it will help breeders to retrace their project or trial either they have fulfil the mating design or need to continue further study from previous breeder. With this, breeders will get better insights in

overall breeding work and continue the legacy of the past breeder. Figure 6 is the example of function of FIBS in constructing **crossing matrix** for breeder's reference.

Parents	CKP	TT3/J25	CDP	ML	CHH	AB1	CEN	TT38/J25	HRU9	TT4	FNK	CEO	GMH
CEB	1												
GKN		1		1									8
DPE			1										
C39		1		2									1
DNN					1								
AG7						1							
CDN	1						3					4	
FPL	1			2				1					
FOP	1			4		1			1				
GJH				2				1			1		
FQE	1			1								1	
FPO	1			2				1		1			
FPN	1			1				1					1
GKE		1		1									
E22	1												

Figure 6.0. Crossing matrix construction by FIBS to monitor the completed mating design for past research consist of several trials.

Through the utilisation of FIBS, breeding project and trial can be organized properly (Figure 7) and information can be retrieved in a secure, faster and accurate manner (Figure 8). This is because the registration of information and data are only granted to the authorized personnel by controlling the in-flow and out-flow of data to ensure data integrity.

This centralized systematic approach will prevent loss of data as all information will be automatically stored in a mirror server. Other interactive features such as integration with Google map function to determine the actual location of the trial as shown in Figure 10.

No	Project	No of Trial	No of Progeny	No of Palm
1	P1: Breeding Trial & Genetic Block	153	1483	15096
2	P2: Clonal Trial & Genetic Block,P Ortet	7	23	73
3	P3: Yangambi Yield Improvement	3	2036	0
4	P5: NPM Yield Improvement	9	930	0
5	P6: Potential Ganoderma Tolerance	1	26	0

Figure 7.0. Project can be organized according to its purpose and objective.

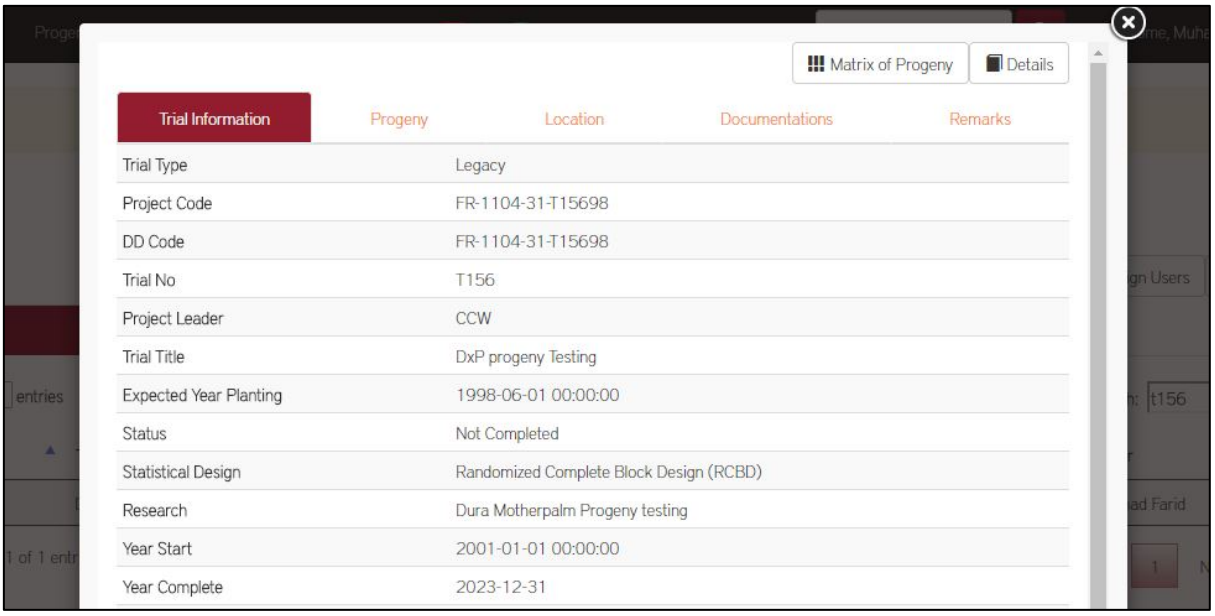
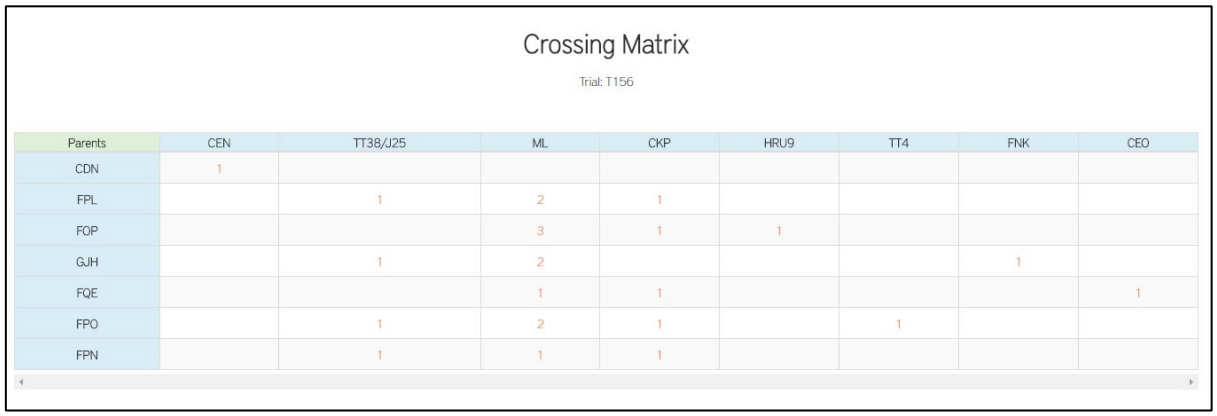


Figure 8.0. Each trial can fill up the relevant information, and also uploading document if necessary. Information about progeny also listed in the trial information.

Furthermore, instead of viewing crossing matrix by combining several trials, the system is capable of generating single trial crossing matrix for breeder to study the objective of trial either by single trial or multiple trial at single time (Figure 9). This will help breeders to understand the past breeders works and continue its legacy for further evaluation and improvement of certain breeds. In addition, we have eliminated the compatibility issues using web-based application software where web browser platform approach is the answer for fast growing technologies information system nowadays and the future.



Parents	CEN	TT38/J25	ML	CKP	HRU9	TT4	FNK	CEO
CDN	1							
FPL		1	2	1				
FOP			3	1	1			
GJH		1	2				1	
FQE			1	1				1
FPO		1	2	1		1		
FPN		1	1	1				

Figure 9.0. Each trial can monitor its crossing matrix to meet the objective of the trial example progeny performance evaluation.

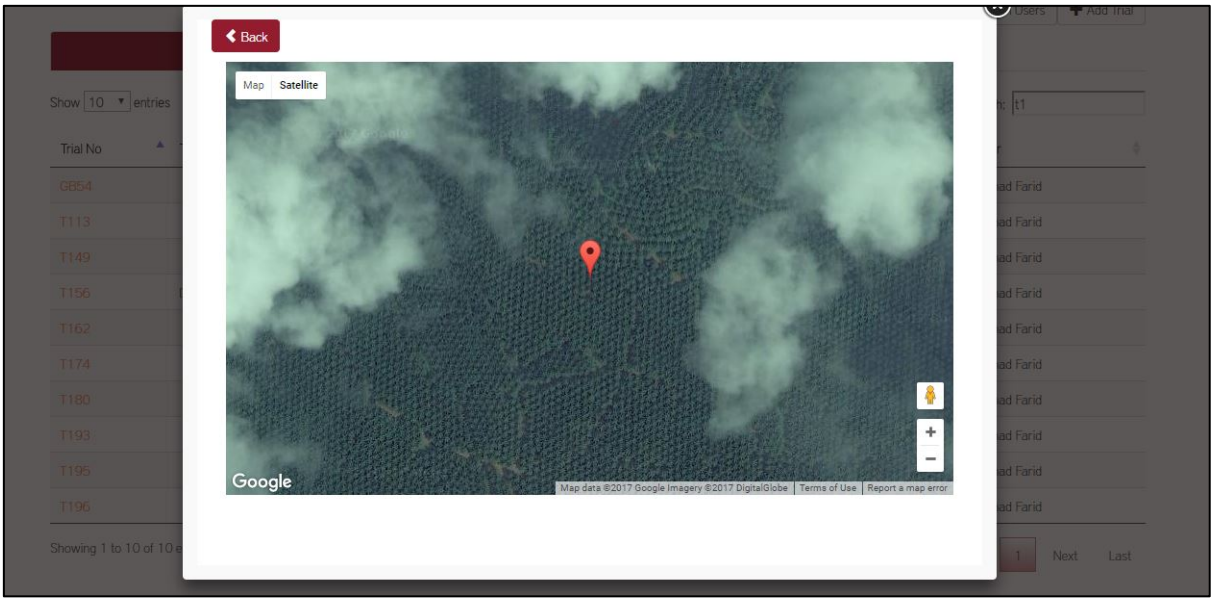


Figure 10.0. The location information also linked with the google maps function to know the location trial planted.

On a separate note, geoinformation of our research trial plots can be tagged with Google maps (Figure 10) to view trial and location where it was planted in FIBS. This will help breeder to understand more about the area and precise location of the trial.

Component II: FIBS Breeding Operation Process Management

Breeding research involve selection of potential parental palm to moving forward to the next generation and field operation responsible to execute the process to produce seed for planting. This two work was a continuous process that play important roles in breeding research to make breeding research alive.

The process involve breeder to go through selection process from data screening and field census palm that fulfil criteria according to its breeding program. Then, field operation team will proceed with palm and flower census to identify correct palm selected to be parental palm by breeders. All the information details was stored centralised at FIBS Information Management component. The process will be continued as showed in Figure 11, where the process was fully run by field operation, start from control pollination process, and seed process until the field planting.

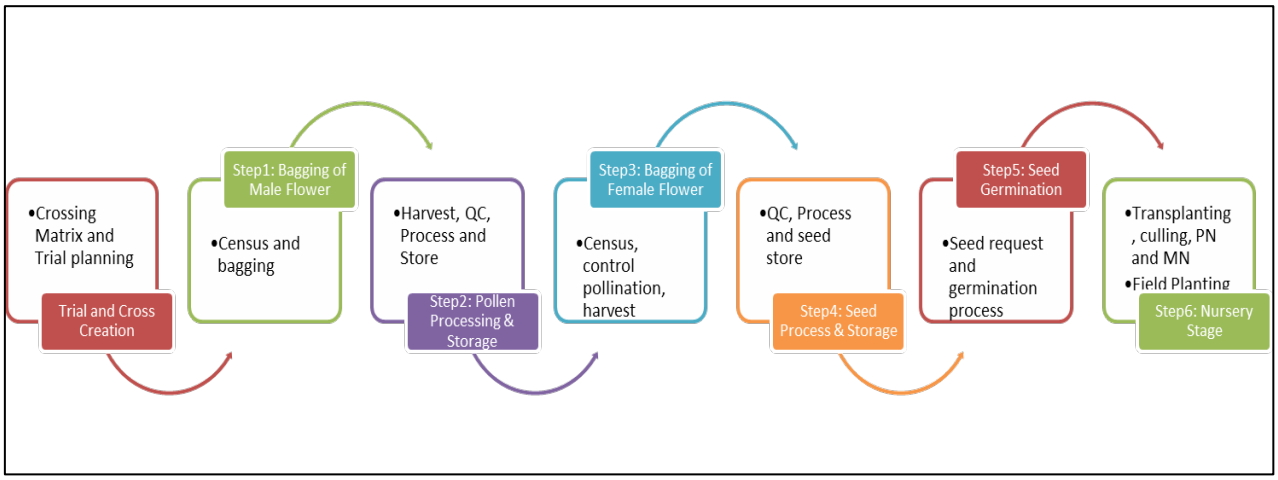


Figure 11.0. Summary workflow of the FIBS Breeding Operation Process management that consist of sub-processes in every steps.

Through FIBS, the process has been grouped and summarized for efficient monitoring by breeding operation and breeders. This component has the most complex system structure due to details sub-processes and steps in the workflow that involve the collection of pollen from male inflorescent and control pollination process for female inflorescent continuous until seed process and nursery process.

Other importance sub module from this operation workflow is pollen and seed inventory in store. This sub module function to record inventory movement of pollen and seed that has been prepared and ready to be used for next step which is control pollination for pollen and seed germination and nursery for seed store. An advantage to this is its potential to be used in commercial seed production process. It can monitor the production of Dura mother palm productivity in producing seed and amount of fresh seeds kept in the storage. While, pisifera pollen inventory will help the pollinator to plan their control pollination work at the same time recording the usage of pollen producing by father palm. The purpose of this module is to increase efficiency in field operation monitoring and higher quality commercial seeds being produced. Integration with breeding information management will help the operation to have an overview background information of different planting material history and its past performance at field. Figure 12 shows the monitoring seed production progress by breeders to make sure project are following the schedule. While figure 13 show the monitoring process by BO and instruct for a next process to be execute by field workers and figure 14 progress by single crosses and its deadline to be planted at nursery and field.

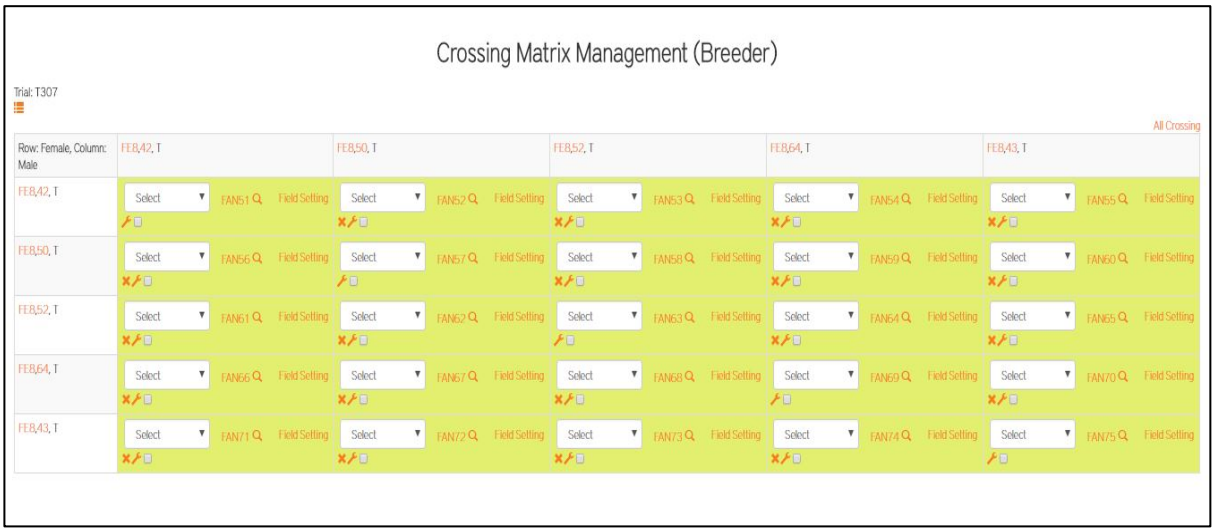


Figure 12.0. Crossing matrix to monitor seed production progress for breeder trial.

Figure 13.0. Breeding operation to monitor and instruct a process to be executed by field workers.

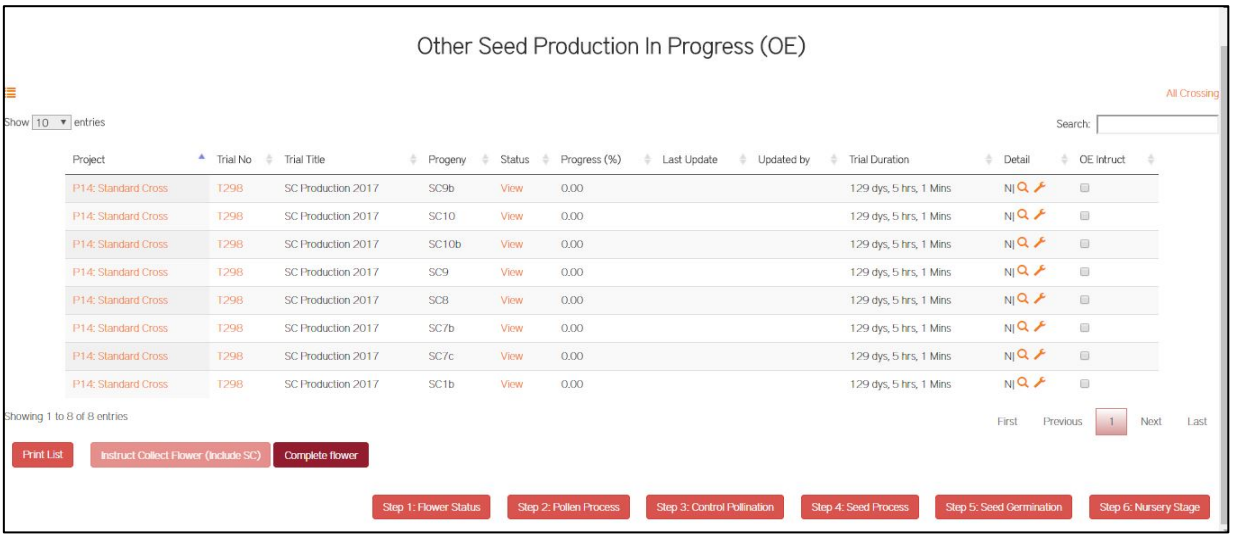


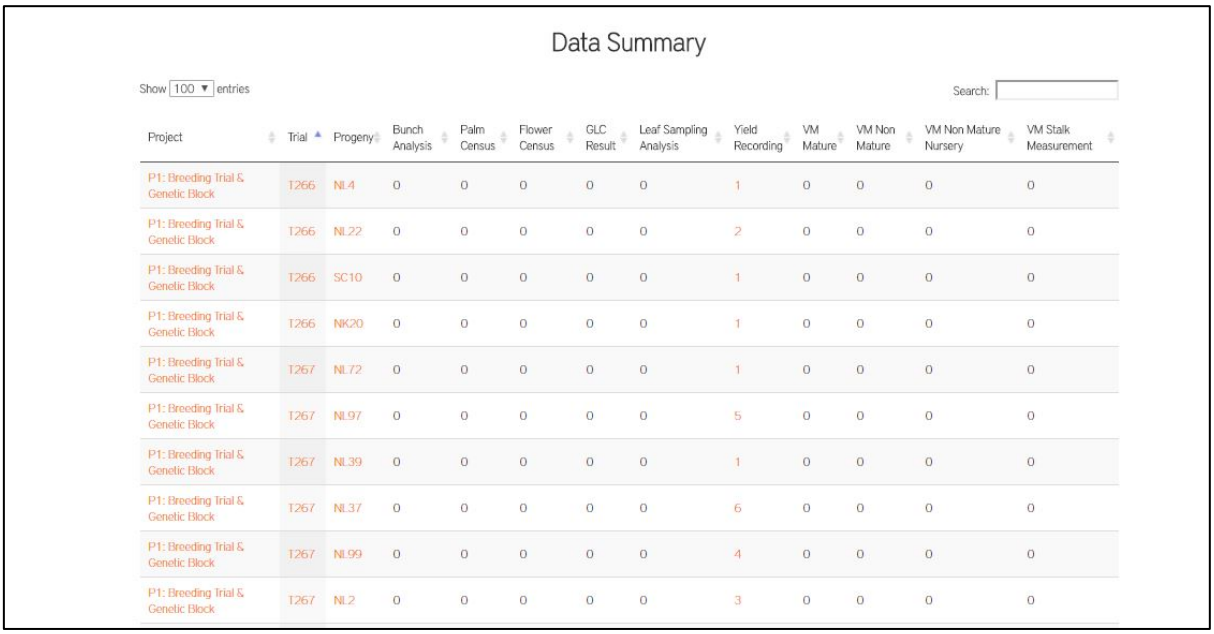
Figure 14.0. Monitor single crosses progress and deadline to be planted.

Component III: FIBS Data Management

The third main component for breeding research is data and parameter observed in the field. This phenotypic data will be used as an indication during progeny performance (testing), parental palm selection for commercial seed production.

The main parameter observed for oil palm is oil yield which consist of various components such as bunch number and bunch weight while oil yield consist of observation on oil percentage per bunch, fruit ratio, fruit size and other. Other parameter usually being observed is vegetative growth, oil nutrient content and agronomical aspect. All of this observation data will be collected for a minimum of 7 to 20 years of planting. Increasing number of data volume collected every year expected to reach about more than two million data point and needs of data to be kept forever for further studies in long term conventional breeding program. Thus, with increasing number of breeding trial and program, the needs of an efficient system to manage all the data and linked with the breeding information system to have better understanding by breeders. It will ease the breeder during background studies for further evaluation and improvement program.

Figure 15 shows, summary data being collected based on progeny type and trials for breeder to further analyse and interpret. This also will help breeding operation to plan their work in collecting field data and work efficiently at the same time supervise field work as breeder needs and performance evaluation.



Project	Trial	Progeny	Bunch Analysis	Palm Census	Flower Census	GLC Result	Leaf Sampling Analysis	Yield Recording	VM Mature	VM Non Mature	VM Non Mature Nursery	VM Stalk Measurement
P1: Breeding Trial & Genetic Block	T266	NL4	0	0	0	0	0	1	0	0	0	0
P1: Breeding Trial & Genetic Block	T266	NL22	0	0	0	0	0	2	0	0	0	0
P1: Breeding Trial & Genetic Block	T266	SC10	0	0	0	0	0	1	0	0	0	0
P1: Breeding Trial & Genetic Block	T266	NK20	0	0	0	0	0	1	0	0	0	0
P1: Breeding Trial & Genetic Block	T267	NL72	0	0	0	0	0	1	0	0	0	0
P1: Breeding Trial & Genetic Block	T267	NL97	0	0	0	0	0	5	0	0	0	0
P1: Breeding Trial & Genetic Block	T267	NL39	0	0	0	0	0	1	0	0	0	0
P1: Breeding Trial & Genetic Block	T267	NL37	0	0	0	0	0	6	0	0	0	0
P1: Breeding Trial & Genetic Block	T267	NL99	0	0	0	0	0	4	0	0	0	0
P1: Breeding Trial & Genetic Block	T267	NL2	0	0	0	0	0	3	0	0	0	0

Figure 15.0. Summarize data being collected by progeny and trial.

Data collected can be monitored on a yearly basis by both breeding operation and breeders. FIBS allows the breeder to analyse the past and present trials utilising the simplified comparative tools. It offers a faster way to compare and study the progenies of interest and to obtain an overview performance. Figure 16, 17 and 18 show analytical tools available in FIBS and the results can be exported into Excel spreadsheet and pdf to be used as presentation image.

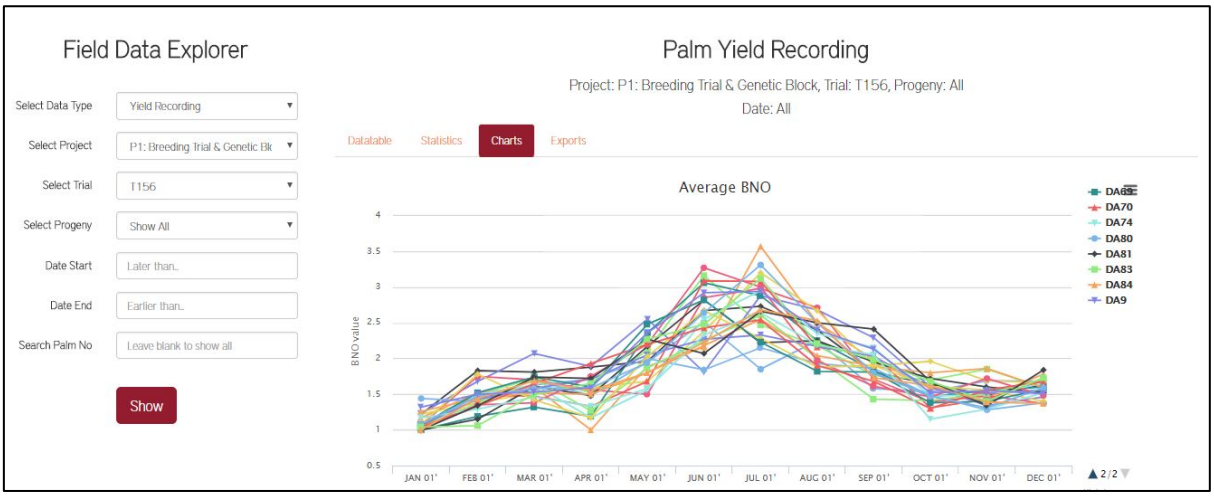


Figure 16.0. Analytical graph projected by FIBS for yield.

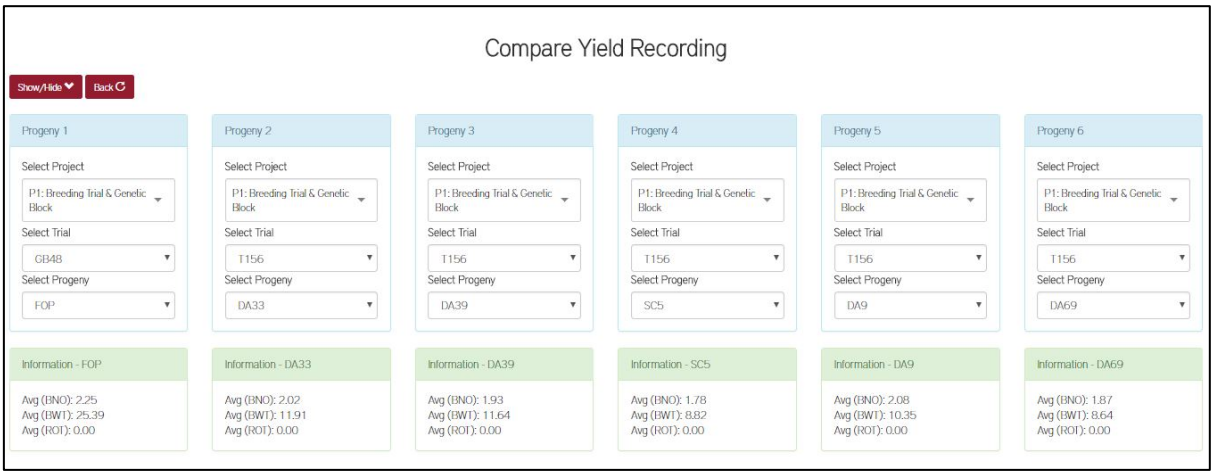


Figure 17.0. Overview of performance comparison between progenies before proceed to a more details analysis by breeder.

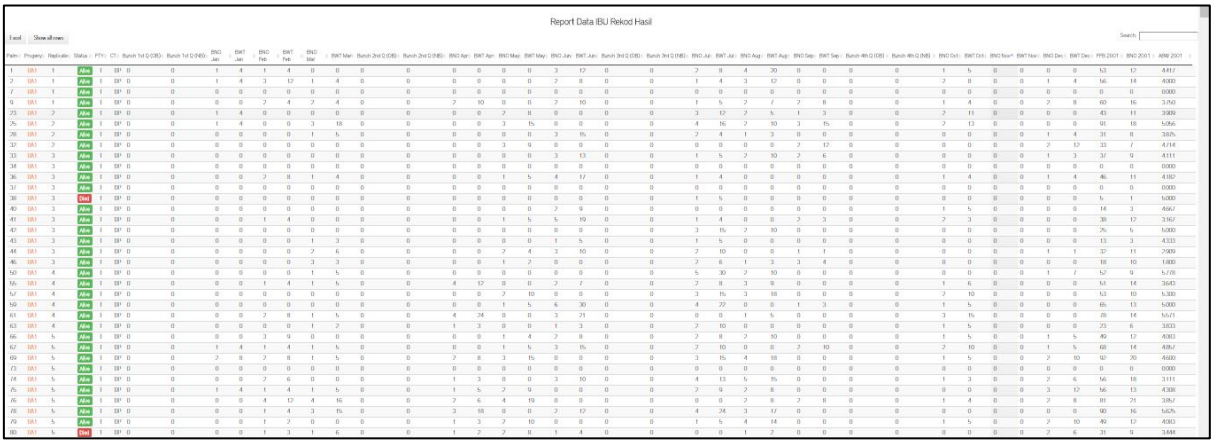


Figure 18.0. Yield recording and monitoring by month and year for each individual palm.

DISCUSSION

1. The development of centralised system can be further improved via the integration with other research field such as molecular breeding and agronomy.
2. This integration will benefit breeder for a better understanding of crops behaviour and pave way towards precision agriculture.
3. FIBS can also be integrated in commercial seed production where the utilisation of both *Dura* and *Pisifera* palms can be fully exploited to its maximum potential.

CONCLUSION

A web-based breeding software and database system (FIBS) had been developed to modernise FGV's breeding research information management. FIBS is essential in our daily breeding operation and research management. The system is now fully operational and had become our standard of operation to ensure the production of quality planting materials for both Malaysia and overseas market demand. The system is so versatile that it can be upgraded or integrated with other modules not limited to marker assisted breeding and precision agriculture.

REFERENCES

Awada L., Smyth S., Phillips P. (2017) Phenotyping and the Future of Plant Breeding. 21st International Consortium on Applied Bioeconomy Research (ICABR), University of California, Berkeley.

Coffman W.R, Lesser W.H., McCouch S.R. (May 2003) Commercialization and the Scientific Research Process: The Example of Plant Breeding

Corley R.H.V., Tinker P.B.H. (2003) The Oil Palm. Forth Edition. Blackwell Science Ltd,Oxford

Evans K., Jung S., Lee T., Brutcher L., Cho I., Peace C., Main D. (2013) Addition of a breeding database in the Genome Database for Rosaceae

Exner V., Hoffmann M., Gruissem W., Hennig L. (2008) PlantDB – a versatile database for managing plant research

Köhl K., Gremmels J. (2015) A software tool for the input and management of phenotypic data using personal digital assistants and other mobile devices. (2015) 11:25 DOI 10.1186/s13007-015-0069-3

Rahaman M., Chen D., Gillani Z., Klukas C., Chen M. (2015) Advanced phenotyping and phenotype data analysis for the study of plant growth and development. DOI: 10.3389/fpls.2015.00619

McLaren C.G., Bruskiewich R., Portugal A., Cosico A. (2005) The International Rice Information System. A Platform for Meta-Analysis of Rice Crop Data. International Rice Research Institute, Manila, The Philippines Plant Physiology, October 2005, Vol. 139, pp. 637–642, www.plantphysiol.org _ 2005 American Society of Plant Biologists.

MPOB Breeding Information System (MPOB BIS™)

Mohd Din, A; Rajanaidu, N; Kushairi, A; Marhalil, M and Zaharah, R;

ABSTRACT

MPOB Breeding Information System (MPOB BIS™) is a computer system that was developed to enable more efficient handling of oil palm breeding data (specifically concerning collection and management of data). Integration and sharing of data from different work stations became easier and more systematic. This system comprised data entry, record checking, data analyses, preparation of reports and storage of data. BIS application was developed using Client/Server Architecture technology where data is stored in the database of the computer server whereas the application is stored in a personal computer. MPOB BIS™ is capable of reducing time of checking and entry of research data from 4.5 - 7 days to 1.4 – 4.5 days. This system has been tested to compare with manual methods of collection, processing, checking and analysis of research data. Data could be retrieved immediately anytime since processing of raw data could be done easily. Time savings in data processing until generation of reports from 6-25 days could be achieved. MPOB BIS™ is capable of storing and channeling information effectively to the client whenever required in real time and within a short period. The search engine of MPOB BIS™ is capable of searching for required information immediately as compared to the manual technique, which is very time consuming in tracing the location and position of files Hence, MPOB BIS™.is capable of reducing working time and subsequently reduce other costs such as labour, energy and material.

Malaysian Palm Oil Board, No. 6 Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia.

INTRODUCTION

Breeding research and development (R&D) activities of a perennial crop like oil palm requires massive data collection. Routine data collection activities normally involved fresh fruit bunch yield, analysis of bunch quality traits and measurement of vegetative traits. Additional collection of other data such as fatty acid composition, carotene, and specialized measurements such as stalk length, etc may also be involved. The MPOB Breeding Information System (MPOB BIS™) was developed by MPOB for efficient handling of R&D data for oil palm breeding (Fig 1). The integration and sharing of data from various sites will be facilitated and made more systematic. MPOBIS comprises data entry, record checking, data analysis and data storage. Reports will be produced in tabular form together with statistical outputs. BIS is based on Client/Server Architecture technology, where data is stored in the database on a computer server and accessible through a personal computer (PC).

METHODOLOGY

BIS comprises 2 modules:

1. Data Processing Module – Data collection from breeding research activities such as recording of bunch yield, bunch analysis, vegetative measurement, fatty acid composition and vitamin E, which will be keyed-in and checked. Raw data will be used to generate derived data. Reports on population mean, family mean and ANOVA will be produced (Fig 2).
2. Breeding Process Module – Germplasm data collection will be captured in the module followed by data from the crossing program, which includes source of pollen, pollination, seed germination, nursery, field planting and passport data. Reports on activities, parents and pedigree will be produced (Fig 3).

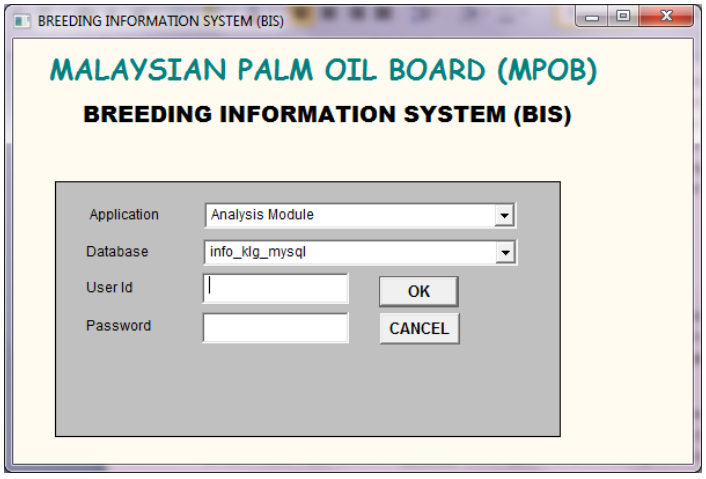


Fig 1. BIS Start Menu

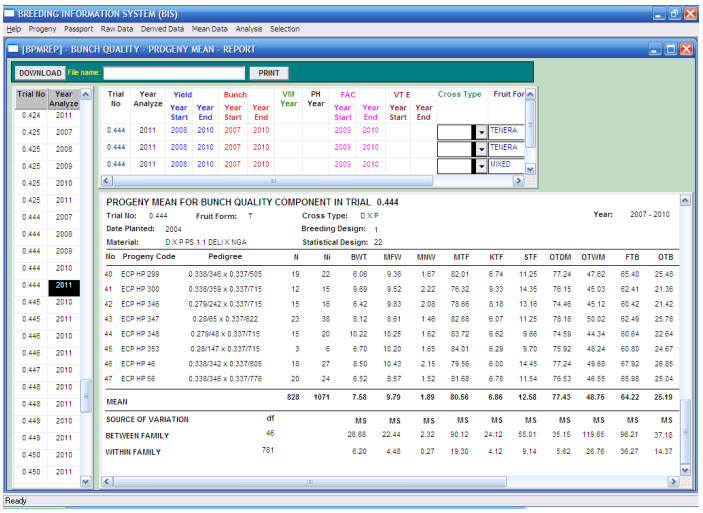


Fig 2. BIS output on Data Processing Module

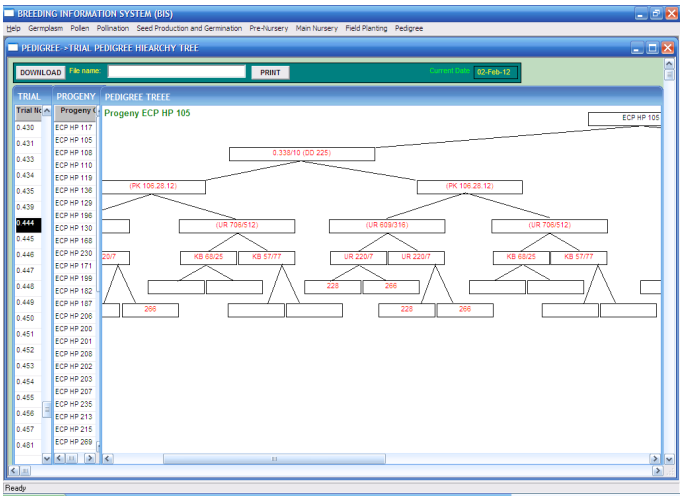


Fig .3 . BIS output on Breeding Process Module

BENEFITS

MPOB BIS™ provides an infrastructure for delivery of information in a user-friendly manner and with easy and quick access of information, whenever and wherever required. It ensures that outputs are generated in a timely, accurate and measureable manner. It also has as an efficient backup system to safeguard against data loss or damage.

REQUIREMENT

Software

- Power Builder
- Database MySQL

Computer Server

- CPU Intel Xeon 2.4 GHz Quad Core
- RAM 8GB RAM DDR3, 1333MHz
- Hard Drives : 2 x 250GB w/RAID 1 for data protection
- Software LINUX Enterprise (CentoS) MySQL, PhpMyAdmin, Apache)

Computer Desktop

- Intel Core 2 Quad Processor Q8400
- 2.66 Ghz, 4MB Cache, 1333FSB
- 320 GB SATA HDD
- 2GB RAM
- Window XP and above
- DVD RW Writer

Proposed Business Model

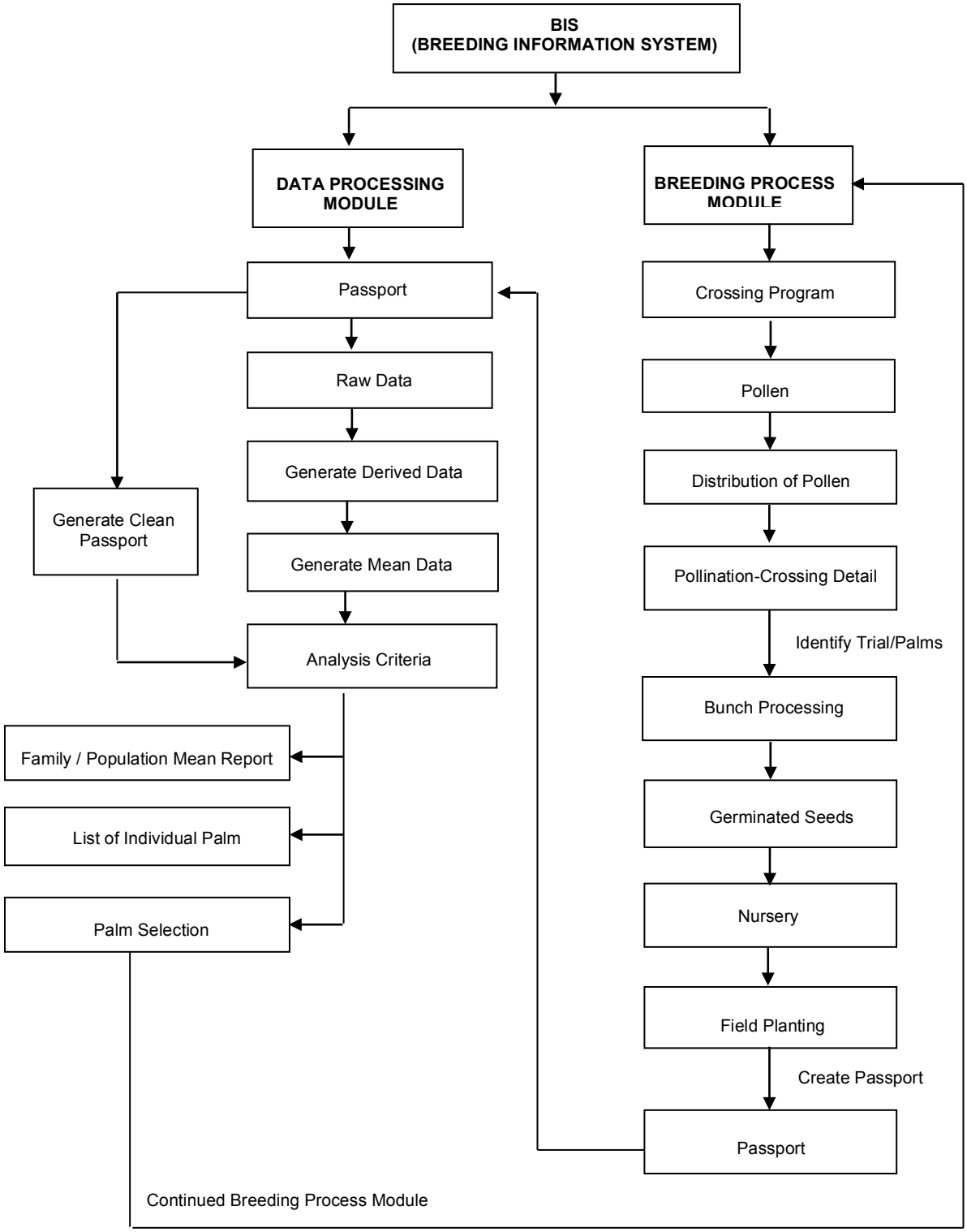
MPOB proposes to license this technology to an IT company who will provide the following services such as installation, training, data migration and technical after sales service

Who will benefit

Members of the industry involved in plant breeding activities will benefit from using MPOB BIS™

Economics

The time required for data checking and data entry reduced from 5-7 to 1-4 days. Data can be accessed in real time and immediately. The time required for data and generation of report reduced from 6-25 to 1-7 days. Labour and material cost could be reduced drastically.



Flow Chart of Breeding Information System

REFERENCES

KUSHAIRI, A. and RAJANAIDU, N. (2000). Chapter 3: Breeding populations, seed production and nursery management. *Advances in Oil Palm Research. Vol. 1.* MPOB, Bangi. p. 39-88.

KUSHAIRI, A., MOHD DIN, A. and RAJANAIDU, N. (2010) Chapter 3: Oil palm breeding and seed production. *Further Advances in Oil Palm Research. Vol. 1.* MPOB, Bangi. p. 47-101.

RAJANAIDU, N; KUSHAIRI, A., RAFII, M., MOHD DIN, A., MAIZURA, I. and JALANI, B.S. (2000). Chapter 6: Oil palm breeding and genetic resources. *Advances in Oil Palm Research Vol. 1.* MPOB, Bangi. p. 171-237.

Protoplast Transformation System as a Potential Platform for Oil Palm Genetic Engineering

Mat Yunus Abdul Masani[1*]; Ghulam Kadir Ahmad Parveez[1]; Gundula A. Noll[2]; Ravigadevi Sambanthamurthi[1] and Dirk Prüfer[2]

ABSTRACT

Protoplasts are a valuable starting material for oil palm genetic engineering because they are totipotent, and chimeras can be avoided by regenerating transgenic plants from a single transformed cell. Here, we report on oil palm protoplast isolation and plant regeneration as well as protoplast transformation using polyethylene glycol (PEG)-mediated transfection and DNA microinjection. We observed efficient protoplast isolation from suspension cells digested with enzyme solution comprising cellulase, pectinase, pectolyase, potassium chloride, calcium chloride and mannitol. We then validated the optimal factors that influenced the plant regeneration from oil palm protoplasts, which depends strongly on the medium composition, plant growth regulator (PGR) and cultivation procedure. For protoplast transformation, we found that concentrations of DNA, PEG and magnesium chloride, and application of heat shock treatment were the important determinants of efficient PEG-mediated transfection. Whereas protoplast embedded in alginate layer cultured for three days and injected with 100 ng/ul DNA solution were the optimal factors for microinjection that enabling us to successfully regenerate transgenic microcalli expressing a green fluorescent protein (GFP) as a visible marker.

[1] Malaysian Palm Oil Board, No. 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia.
[2] Fraunhofer Institut für Molekularbiologie und Angewandte Ökologie, Schlossplatz 8, 48143 Münster, Germany.

INTRODUCTION

The oil palm genetic engineering programme has been initiated at Malaysian Palm Oil Board (MPOB) in early 1990s (Cheah *et al*., 1995). The main objective of this programme is to produce transgenic oil palm with high oleic oil content. Modification of oil quality such as increasing stearic acid, value-added oil such as synthesizing of palmitoleic acid and ricinoleic acid, and novel products such a biodegradable plastics have also been targeted. It was postulated for oil palm; up to 80 % reduction in time could be achieved for those target products through the combination of genetic engineering and tissue culture (Majid and Parveez, 2007). In addition, since the oil palm is a perennial crop, those products will be continuously produced for at least 30 years, making it an ideal candidate for plant genetic engineering.

Particle bombardment and *Agrobacterium*-mediated transformation are commonly used for insertion of valuable genes into oil palm genome. Even though particle bombardment often causes the integration of multiple copies of transgene into the genome of transgenic plants (Stanton, 1998), this direct DNA delivery approach has been used as the first method for oil palm transformation (Parveez, 1998). The main factors influenced particle bombardment for oil palm transformation such as the appropriate selectable marker and promoters, and also optimal physical and biological parameters have been recognised (Parveez and Christou, 1998). Succeeding these factors, the procedure for the production of Basta resistance transgenic oil palm was successfully developed (Parveez *et al*., 2000). Since this success, thousands of embryogenic calli have been bombarded with various genes such as genes involved in the fatty acid biosynthesis pathway to increase oleic acid and stearic acid, polyhydroxybutyrate (PHB) and polyhydroxyvalerate (PHBV) genes for the production of biodegradable plastics (Parveez *et al*., 2015).

Agrobacterium-mediated transformation has been verified to insert either single or low copies of transgenes, which was described from a number of studies such as rice (Raineri *et al*., 1990) and maize (Gould *et al*., 1991). However, *Agrobacterium*-mediated oil palm transformation still restricts with some limitations such as low transformation frequency since oil palm as monocots is not a natural host for *Agrobacterium* unlike dicots plants. Nevertheless, several studies were carried out to improve this method such as the use of immature embryos (IE) as target tissue (Ruslan *et al*., 2005), the determination of best transformation parameters (Masli *et al*., 2009) and selection using 2-deoxyglucose (2-DOG) as positive selection system (Izawati *et al*., 2015) .

The works described above revealed that nearly five years are required to generate transgenic oil palm plants by particle bombardment or *Agrobacterium*-mediated transformation. This long selection and regeneration process encourages the growth of non-transformed cells which subsequently tends to produce escapes and chimeric plants (Nurfahisza *et al*., 2014). It is possible that the optimization of DNA delivery and selection could overcome such challenges but an alternative approach is to use protoplasts as transformation targets because transgenic plants can then be regenerated from a single transformed cell.

In the present study, we describe a systematic procedure for the preparation of oil palm protoplasts and the regeneration of viable oil palm plants. We then have investigated and determined the best parameters to achieve reliable and efficient transformation procedure employing oil palm protoplasts by using PEG-mediated transformation and DNA microinjection.

METHODS

Plant material

Oil palm embryogenic cell suspension cultures were cultivated in 100 ml-flask containing 50 ml Y35N5D2iP liquid medium (Masani *et al*., 2013). The suspension cultures were incubated in the dark at 28ºC on a rotary shaker and agitated at 120 rpm. Half of the Y35N5D2iP liquid medium in the flask culture was discarded and replaced with fresh medium every 14 days.

Protoplast isolation and purification

Protoplasts were isolated from 3-month old oil palm suspension cultures. Cells were collected by filtration through a 300-µm nylon mesh, and 0.5 g fresh weight (fwt) of cells was transferred into a 50-ml centrifuge tube containing 15 ml filter-sterilized enzyme solution [2% (v/v) cellulase (Sigma), 1% (v/v) pectinase (Sigma), 0.5% (w/v) cellulase onuzuka R10 (Duchefa), 0.1% (w/v) pectolyase Y23 (Duchefa), 3% (w/v) KCl, 0.5% (w/v) $CaCl_2.2H_2O$ and 3.6% (w/v) mannitol, pH 5.6]. The cells were resuspended by inverting the tube 6–10 times and then incubated in the dark without shaking at 26ºC for 14 h. The mixture was diluted with 15 ml filter-sterilized washing solution [3% (w/v) KCl, 0.5% (w/v) $CaCl_2.2H_2O$, 3.6% (w/v) mannitol, pH 5.6], resuspended by inverting the tube 3–5 times, filtered through a sterilized double layer of miracloth and collected in a 50-ml centrifuge tube. The filtration step was repeated 2–3 times until all undigested tissues, cell clumps and cell debris were removed. The mixture was centrifuged at 100 × g for 5 min at 22ºC and the supernatant was discarded. The protoplast pellet was resuspended by adding 10 ml washing solution and mixing by inversion, followed by centrifugation as above. The supernatant was removed completely and the protoplast pellet was resuspended in 10 ml filter-sterilized rinse solution [3 % (w/v) KCl, 3.6 % (w/v) mannitol, pH 5.6], and then centrifuged at 100 x g for 5 minutes at 22ºC. After three cycles, the supernatant was removed by leaving 3 ml and stored at room temperature for further experiments.

Media optimization

The purified protoplasts were cultured using different media either in liquid or embedded in agarose solidified media. Three ml rinse solution containing the purified protoplasts was allowed to settle for 20 min at room temperature. For liquid culture, the rinse solution was replaced with liquid media and for agarose solidified cultures, the protoplasts pellet was resuspended with a double concentration of liquid media. Equal volumes of protoplasts suspension and agarose were mixed by adjusting the final concentration to 0.6 % (w/v) of agarose, and then 2 ml each of the mixtures was dispensed into 24 wells culture plate. The culture plate was placed at room temperature for an hour for agarose solidification. The protoplasts embedded in agarose solidified media in each well were covered with 500 µl of the same liquid media was used for preparation of agarose solidified cultures. The culture plates containing liquid or agarose solidified cultures were sealed and incubated at 28ºC in the dark. The culture was monitored microscopically every day to observe the first and second cell division, and seven days intervals for microcolonies and microcalli formations.

Agarose bead culture

The mixture of protoplasts and agarose was prepared by using the same procedure for preparation of agarose solidified. Agarose beads were prepared by dropping 200 µl of the mixture into a 90 x 15 mm petri dish. After agarose solidification, the agarose beads were cultured in Y3A liquid medium in shaking condition at 50 rpm by refreshing the medium at 14-

day intervals. The cultures were continued until the microcalli grew to embryogenic calli. The agarose beads were transferred to a Y3A solid medium when microcalli developed to 5-10 mm in size of whitish and yellowish embryogenic calli. The agarose beads were maintained on Y3A solid medium supplemented with different concentrations of a combination of NAA (0.5-10 µM) and BA (0.1-5 µM) until the formation of embryos was observed. The agarose beads containing the embryogenic calli were incubated at 28°C in the dark and were subcultured every 30 days in fresh medium. The embryos were transferred onto ECI solid medium (Masli et al., 2009) supplemented with the optimum PGR, and then were incubated at 28°C in the light until small plantlets were produced.

Polyethylene glycol (PEG) mediated transformation

One ml of protoplast suspension was incubated at room temperature for 10 min or heat shock treated by incubation at 45ºC for 5 min, immediately placed on ice for 1 min and then incubated at room temperature for 10 min. A 500 µl aliquot of the protoplast suspension was then placed as a single droplet in the middle of a 60 mm x 15 mm petri dish. Then, five drops of 100 µl PEG-MgCl solution [25-50% (w/v) PEG 4000 (Sigma), 10-100 µM $MgCl_2.6H_20$ (Sigma), 3% (w/v) KCl, 3.6% (w/v) mannitol, 0.05% (w/v) 2-N-morpholinoethanesulfonic acid (MES); pH 6.0] were placed surrounding the protoplast drop. Twenty five or 50 µg of plasmid DNA was added slowly to the protoplasts drop, mixed by stirring with 200 µl-tip and incubated at room temperature in the dark. After incubation for 10-30 min, the DNA-protoplasts drop was sequentially mixed with each of PEG-MgCl drops by stirring with 200 µl-tip and incubated for another 30 min, then 4 ml washing solution was added drop by drop and incubated in the dark at 26ºC. PEG-mediated transfection efficiency was calculated as the percentage of the number of GFP positive protoplasts divided by the total number of protoplast in one representative microscope field. The calculation was performed three times for a total of not less than 200 protoplasts.

Alginate thin layer preparation

After leaving the protoplasts suspension for 20 min, the supernatant was removed completely and the protoplasts pellet was resuspended in 3 ml filter-sterilized alginate solution consisted of 1 % (w/v) alginic acid sodium salt (A2158, Sigma) dissolved in Y3A liquid medium. Alginate-embedded protoplasts were distributed as a thin layer onto supporting media comprising 1.5 ml filter-sterilized Y3A [5.5 % (w/v) sucrose and 11.9 % (w/v) glucose supplemented with 0.1 % (w/v) $CaCl_2.2H_20$] solidified with 1 % (w/v) agarose sea plaque, in 35 mm x 10 mm petri dish.

Microinjection workstation

The microinjection workstation consisted of a Leica DM LFS upright microscope (Leica Microsystems Wetzlar GmbH, Germany) with a joystick controlled motorized objective revolver for HCX APOL U-V-I water immersion objectives (10x, 20x, 40x and 63x), mounted on a fixed table and placed in a laminar. The microscope was equipped with a Luigs and Neumann Manipulator set with a control system SM-5 and SM-6 (Luigs and Neumann, Germany).

Microinjection of oil palm protoplasts

A plate containing alginate layer was placed on the microscope stage, and the vitality of embedded protoplasts was confirmed by using the 10x objective. The objective was raised to maximum position to freely allow the needle tip to reach the center of the field view with the X-

and Y-axis controller (Control system SM-5) of the manipulator. The needle was lowered as close as possible to the alginate layer with the Z-axis controller and the cytoplasm or nucleus of target protoplast was identified by adjusting the 20x objective to optimal resolution and contrast, after which the needle tip was moved to right above the protoplast with the X- and Y-axis hand wheel controller. The needle tip was then inserted into the alginate layer just next to the protoplast by using Z-axis hand wheel controller and penetrated into the protoplast by using the X-axis hand wheel controller. The DNA injection solution was slowly injected into the protoplast by using a microinjector CellTram vario, which was confirmed by the fluorescence illumination. The needle tip was carefully withdrawn from the protoplast and moved to the next target protoplast. The injected protoplasts were monitor periodically by using Leica MZ16F fluorescent stereomicroscope with GFP3 filter (Leica Microsystems Wetzlar GmbH, Germany).

Alginate layer culture

Following microinjection, the plates containing the alginate layer were incubated in the dark at 28°C for 5 days. The alginate layers were then separated from supporting media and transferred into 60 mm x 15 mm petri dishes containing 3 ml Y3A liquid medium consisted of 5.5 % (w/v) sucrose and 8.2 % (w/v) glucose supplemented with 10 µM NAA, 2 µM 2,4-D, 2 µM IBA, 2 µM GA3, 2 µM 2iP and 200 mg/l ascorbic acid. The dishes were incubated in the dark by shaking at 50 rpm at 28°C. After 2 weeks, the medium was replaced with similar Y3A liquid medium but the concentrations of sucrose and glucose were decreased to 4 % (w/v) and 7.2 % (w/v), respectively. The alginate layers were cultured in this medium for a month by refreshing the medium at 14-days intervals, then replaced with Y3A liquid medium comprising of 4 % (w/v) sucrose until the microcalli were observed.

RESULTS AND DISCUSSION

Protoplast Culture

Protoplast isolation from oil palm cell suspension cultures

Protoplasts were successfully isolated from callus tissue derived from a 3-month-old cell suspension culture sampled at 4, 7 and 14 days after subculturing. We achieved protoplast yields of 0.9–1.14 x 10^6 per g fresh weight with an average viability of 82% throughout the subculture period, but the size of the protoplasts varied from 5–14 µm at 4 days (Figure 1A), to 15–25 µm at 7 days (Figure 1B) and 25–35 µm at 14 days (Figure 1C). Protoplasts at 7 days with an average size of 15–25 µm contained a dense cytoplasm concentrated around the nucleus (Figure 1E), whereas older protoplasts with an average size of 25–35 µm were fully packed with cytoplasm (Figure 1F), and more than 50% of the younger protoplasts less than 15 µm in size contained cytoplasm but lacked a nucleus (Figure 1D). The protoplasts isolated 7 days after subculture were used for further experiments.

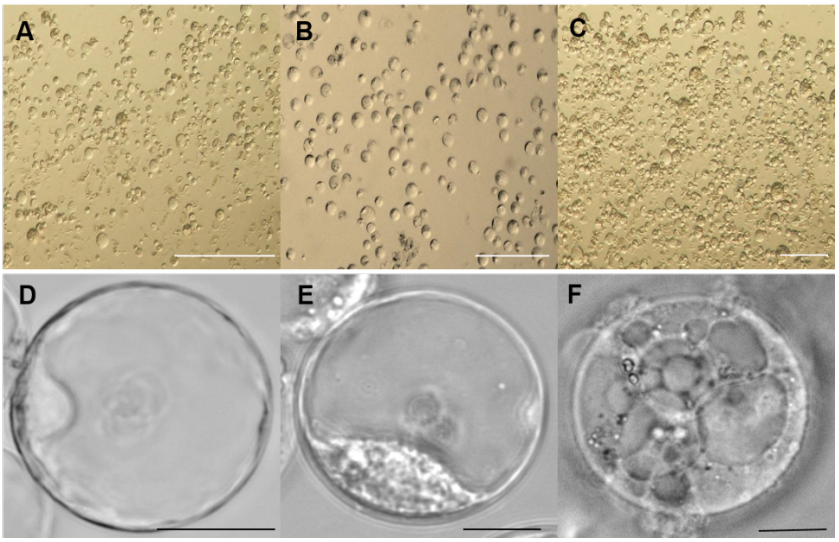


Figure 1. Protoplast derived from oil palm suspension cells

Freshly-isolated protoplasts from cell suspension callus after subculture for 4 days (A), 7 days (B) and 14 days (C). Microscopic images showing the cytoplasm structure of protoplasts isolated after 4 days (D), 7 days (E) and 14 days (F). Scale bar = 100 µm in (A)-(C), (K) and 10 µm in (D)-(F).

Selection of the optimal growth medium

To identify the most suitable medium for the cultivation of oil palm protoplasts, we compared 14 media formulations (Masani *et al*., 2013). The EC media series (½ECI–ECVI) was based on MS basal medium (Murashige and Skoog, 1962), except for ECII, which was based on KM basal medium (Kao and Michayluk, 1975). Meanwhile, the Y3 media series (Y3–Y3F) was based on modified Y3 basal medium (Teixeira *et al*., 1995). Other media components were based on media used for other plant species. Initial experiments indicated that liquid media did not support protoplast division but promoted extensive aggregation, causing most of the protoplasts to die within two weeks. However, protoplasts embedded in solidified agarose remained viable, formed normal cell walls and underwent division in all the media we examined. As shown in Figure 2, the Y3A medium was superior in its capability to induce cell wall formation (5–7 days), initial cell division (9–12 days) (Figure 2D) and second cell division (17–21 days) (Figure 2E).

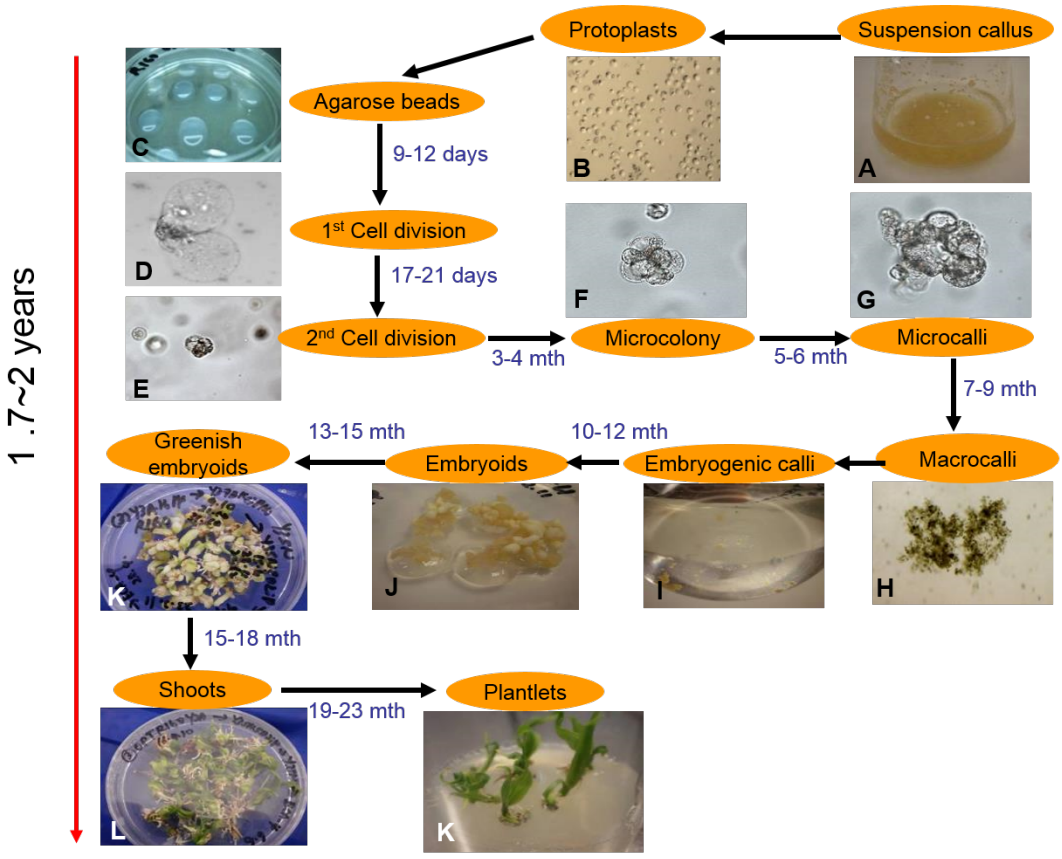


Figure 2. Regeneration of oil palm plantlets from embryogenic calli originating from protoplasts

A 3-month-old oil palm cell suspension culture in Y35N5D2iP liquid medium (A) was used for preparation of totipotency protoplasts (B). Agarose beads containing protoplasts (C) were cultured in Y3A liquid medium supplemented with 200 mg/l ascorbic acid, initiating first cell division (D) and second cell division (E), developing into microcolony (F) subsequently microcalli (G), macrocalli (H) and embryogenic calli (I) within 9 months. The agarose beads were then cultured on solid medium for 2 months to generate embryos (J). After 3 months on solid medium, the whitish embryoids became green (K) and then developed shoots (L) and finally into plantlets (K).

Selection of optimal plant growth regulators (PGR)

Eleven combinations of PGR were tested in the initial experiment (Table 1). Various concentrations of the auxins: naphthalene acetic acid (NAA), 2,4-Dichlorophenoxyacetic acid (2,4-D), indole-3-acetic acid (IAA) and indole-3-butyric acid (IBA) were combined with various concentrations of the cytokinins: zeatin (Zea), gibberellic acid (GA_3), 6-benzylaminopurine (BAP) and 2-γ-dimethylallylaminopurine (2iP), in order to optimize protoplast growth in Y3A medium. Table 1 shows that PGR combinations 1–6 inhibited protoplast growth completely in the medium. If the concentration of 2,4-D, IBA, GA_3 and 2iP was higher than 7 µM, then protoplast division was inhibited and the cells died within one week. Microcolonies formed in Y3A medium after 12–16 weeks when PGR combinations 9–11 were used. The protoplasts cultured in Y3A medium supplemented with PGR combination 10 (10 µM NAA, 2 µM 2,4-D,

10 µM IAA, 2 µM IBA, 10 µM Zea, 2 µM GA_3, 10 µM BA and 2 µM 2iP) achieved the highest frequency of division (18.33%) and microcolony formation (8.86%).

We next studied how each of the individual regulators in PGR combination 10 influenced the development of microcolonies into microcalli (15–30 cells). We tested seven combinations of PGR in Y3A medium by systematically excluding each of the regulators in combination 10 in turn, thus producing PGR combinations 12–18 as shown in Table 1. The number of microcolonies increased significantly when the protoplasts were cultured with PGR combinations 12 (11.3%), 13 (11.9%) and 15 (13%), which excluded 2,4-D, IAA and Zea, respectively. In contrast, the exclusion of any other PGR (combinations 14 and 16–18) reduced the frequency of microcolony formation to a range between 3.56% and 5.03%. After 20–24 weeks in Y3A medium with PGR combinations 12, 13, 15 and 17, the microcolonies developed into microcalli visible with the naked eye. The highest frequency of microcallus formation was 6.13% (PGR combination 15) and the lowest was 1.6% (PGR combination 12). The absence of IBA, GA_3 and 2iP in PGR combinations 14, 16 and 18 had an adverse impact on the formation of microcolonies and no microcalli were produced even after 26 weeks.

Based on the results presented above, we cultivated oil palm protoplasts using agarose bead cultures (Figure 2C) by mixing protoplasts with Y3A medium supplemented with 10 µM NAA, 2 µM 2,4-D, 2 µM IBA, 2 µM GA_3 and 2 µM 2iP, and solidified with 0.6% agarose, which we describe as PGR combination 19 (Table 1).

TABLE 1. EFFECT OF PGR ON PROTOPLASTS CULTURED IN Y3A MEDIUM

PGR	Auxin (µM)				Cytokinin (µM)				Y3A medium		
									Solidified agarose medium		
No	NAA	2,4D	IAA	IBA	Zea	GA_3	BA	2iP	Division frequency	Microcolony frequency	Microcallus frequency
1	1	11	1	11	1	11	1	11	0	0	NA
2	2	10	2	10	2	10	2	10	0	0	NA
3	3	9	3	9	3	9	3	9	0	0	NA
4	4	8	4	8	4	8	4	8	0	0	NA
5	5	7	5	7	5	7	5	7	0	0	NA
6	6	6	6	6	6	6	6	6	0	0	NA
7	7	5	7	5	7	5	7	5	3.51±1.09	1.82±0.46	NA
8	8	4	8	4	8	4	8	4	6.64±2.24	2.98±0.61	NA
9	9	3	9	3	9	3	9	3	9.25±1.01	4.40±0.78	NA
10	10	2	10	2	10	2	10	2	18.33±0.44	8.86±0.85	0
11	11	1	11	1	11	1	11	1	14.37±1.83	6.39±0.93	NA
No	NAA	2,4D	IAA	IBA	Zea	GA_3	BA	2iP	Agarose bead culture		
12	10		10	2	10	2	10	2	NA	11.3±2.42	1.6±0.7
13	10	2		2	10	2	10	2	NA	11.9±1.55	5.5±2.13
14	10	2	10		10	2	10	2	NA	3.56±2.64	0
15	10	2	10	2		2	10	2	NA	13.06±2.95	6.13±1.58
16	10	2	10	2	10		10	2	NA	6.6±0.96	0
17	10	2	10	2	10	2		2	NA	5.03±1.95	2.73±2.12
18	10	2	10	2	10	2	10		NA	4.53±1.65	0
19	10	2		2		2		2	NA	22.2±4.75	9.8±1.87

*The values represent the mean of three independent experiments.
NA: not available

Culturing the agarose beads surrounding with Y3A liquid medium promoted cell division and the development of microcolonies (Figure 2F) at a frequency of 22.2%, which is higher than the 13% achieved with PGR combination 15 in solid agarose medium (Table 1). The Y3A liquid

medium was then supplemented with 4.5% (w/v) sucrose, resulting in the development of microcalli (Figure 2G) at a frequency of 9.8% (compared to 6.13% with PGR combination 15). Simultaneously, the osmotic potential surrounding the agarose beads was gradually reduced, which doubled the number of microcolonies and microcalli.

Plant regeneration from microcalli derived from protoplasts

The microcalli (Figures 2G and 2H) failed to produce embryogenic calli, and the microcalli became discolored probably due to the accumulation of phenolic compounds released from the cells and the reduction of PGR present in the agarose beads. Adding ascorbic acid, silver nitrate ($AgNO_3$) or activated charcoal to the medium surrounding the agarose beads along with the PGRs reduced the discoloration and promoted embryogenesis. The agarose beads were therefore cultured in Y3A liquid medium with the addition of different concentrations of ascorbic acid, $AgNO_3$ and activated charcoal. The microcalli began to turn yellow and then developed into embryogenic calli (Figure 2I), indicating further cell growth and development, particularly in media supplemented with 200 mg/l ascorbic acid. The agarose beads were transferred onto Y3A solid medium (Figure 2J) supplemented with different concentrations of NAA and BA to induce somatic embryogenesis. Among the five different concentrations of NAA and BA we tested, only Y3A solid medium supplemented with 1 µM NAA and 0.1 µM BA (designed as Y3A-4), was able to promote the growth of embryogenic calli. The agarose beads were subcultured at 4-week intervals on Y3A-4 solid medium until all the embryogenic calli had developed into somatic embryos. After 44-48 weeks, off-white embryoids appeared on the surface of the agarose beads (Figure 2J). These were transferred onto ECI solid medium (Masli *et al*., 2009) supplemented with 1 µM NAA and 0.1 µM BA. Greenish embryoids developed after 8 weeks under illumination (Figure 2K) and these regenerated into shoots (Figure 2L) subsequently plantlets (Figure 2K) after a further 12-24 weeks.

Protoplast Transformation

PEG-mediated transfection of oil palm protoplasts

The stable integration of exogenous DNA into the genome of oil palm protoplasts following PEG-mediated transfection, electroporation or microinjection, could facilitate the generation of stable transgenic lines because the plants would be regenerated from a single transformed cell. This requires the optimization of transformation protocols that maintain the viability of oil palm protoplasts, promote the uptake of DNA and demonstrate the efficiency of transgene expression. Because PEG-mediated transfection is a standard method for gene transfer to protoplasts that allows the rapid analysis of transient reporter gene expression, this method was investigated as a first step in the development of an efficient transformation protocol.

The concentration of Mg^{2+} is the most important determinant of efficient PEG-mediated transient gene expression in plant protoplasts. Similarly, we found that the concentration of Mg^{2+} greatly influenced the transfection efficiency of oil palm protoplasts and the intensity of GFP fluorescence. We investigated the impact of Mg^{2+} ions on transfection efficiency by incubating oil palm protoplasts as above for 10 min in the presence of 10 µg of CFDV-hrGFP plasmid DNA mixed with 40% (w/v) PEG dissolved in Rinse solution, but this time we varied the concentration of Mg^{2+} ions by preparing solutions containing 10 mM, 25 mM, 50 mM and 100

mM $MgCl_2$ (data not shown). The presence of 10 mM $MgCl_2$ increased the transfection efficiency by four-fold to 0.39% compared to a PEG solution lacking magnesium (< 0.1%), but higher concentrations were even more beneficial, and the greatest efficiency (2.5%) was achieved in the presence of 50 mM $MgCl_2$. GFP fluorescence was more intense in the protoplasts transfected at higher Mg^{2+} concentrations, indicating the more efficient uptake of exogenous DNA.

Having established the optimal Mg^{2+} concentration for transfection, we next varied the incubation time following the addition of plasmid DNA but prior to the addition of the PEG/$MgCl_2$ solution. Prolonging the incubation period to 15 min or 30 min reduced the transfection efficiency to 1.42% and 0.65%, respectively. Therefore we reverted to the original incubation period of 10 min. Next we investigated the impact of DNA concentration on transfection efficiency by incubating protoplasts in the presence of 25μg or 50μg of CFDV-hrGFP plasmid DNA using the 40% (w/v) PEG/50 mM $MgCl_2$ solution discussed above. High transfection efficiencies were achieved in both cases, but the lower DNA concentration was less efficient (2.05%) than the original 50-μg dose (2.73%). The GFP fluorescence was also more intense in protoplasts transfected with higher concentrations of DNA probably because more was taken up into the cell.

We also investigated the effects of different PEG concentrations, varying the (w/v) concentration of PEG 4000 from 25%, to 40% and also 50%. In each case, the different PEG concentrations were tested with the optimal DNA and $MgCl_2$ concentrations and 10-min DNA incubation time established above. The corresponding transfection efficiencies were 3.74%, 2.02% and 1.66%, showing that 25% (w/v) PEG is optimal for the transformation of oil palm protoplasts. There was no difference in terms of GFP fluorescence regardless of the PEG concentration, suggesting that PEG does not affect hrGFP gene expression but may instead affect the viability of the oil palm protoplasts at concentrations higher than 25%.

Finally, we investigated the effect of heat shock treatment by incubating the protoplasts at 45ºC for 5 min and then cooling on ice for 1 min before adding 50 μg of CFDV-hrGFP plasmid DNA, incubating for 10 min as above and then adding 25% (w/v) PEG in 50 mM $MgCl_2$ (Figure 3). This treatment increased the transfection efficiency even further to 4.76% indicating that a heat shock significantly improves DNA uptake. Fluorescent protoplasts were observed continuously for 9 days indicating that hrGFP fluorescence remains stable following transfection, although the frequency declined over time from 4.42% on day 6 to 4.35% on day 9.

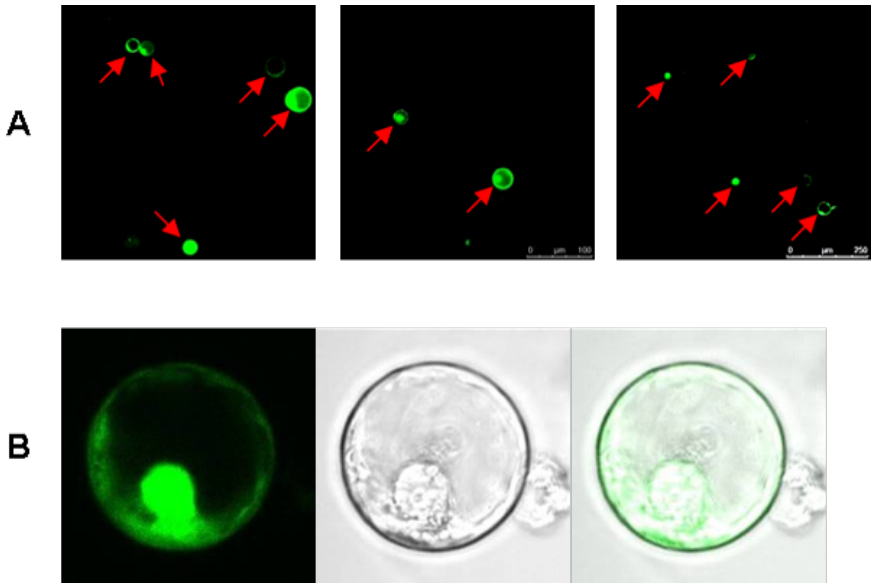


Figure 3. Effect of heat shock treatment on PEG transfection efficiency.

Oil palm protoplasts were incubated at 45ºC for 5 min and then on ice for 1 min before mixing with 50 µg of CFDV-hrGFP plasmid DNA and then PEG-$MgCl_2$ solution (25% PEG, 50 mM $MgCl_2$, 3% KCl and 3.6% mannitol, pH 6.0). The protoplasts were incubated at 26ºC for 3 days (A). Microscopic images are shown representing GFP fluorescence, bright field and three-layer images (Merged) of the protoplasts (B). Red arrows indicate protoplasts showing GFP fluorescence.

Transformation of oil palm protoplasts by DNA microinjection

A DNA microinjection protocol for oil palm protoplasts was developed using protoplasts embedded in an alginate layer (Figure 4B) because microinjection is facilitated if the protoplasts are immobilized in a single plane. Different concentrations of alginate, ranging from 0.5% to 2%, were dissolved in Y3A liquid to prepare the substrate. We found that 1% alginate was ideal for immobilizing the protoplasts, whereas they remained mobile if lower concentrations were used and higher concentrations promoted the formation of clumps.

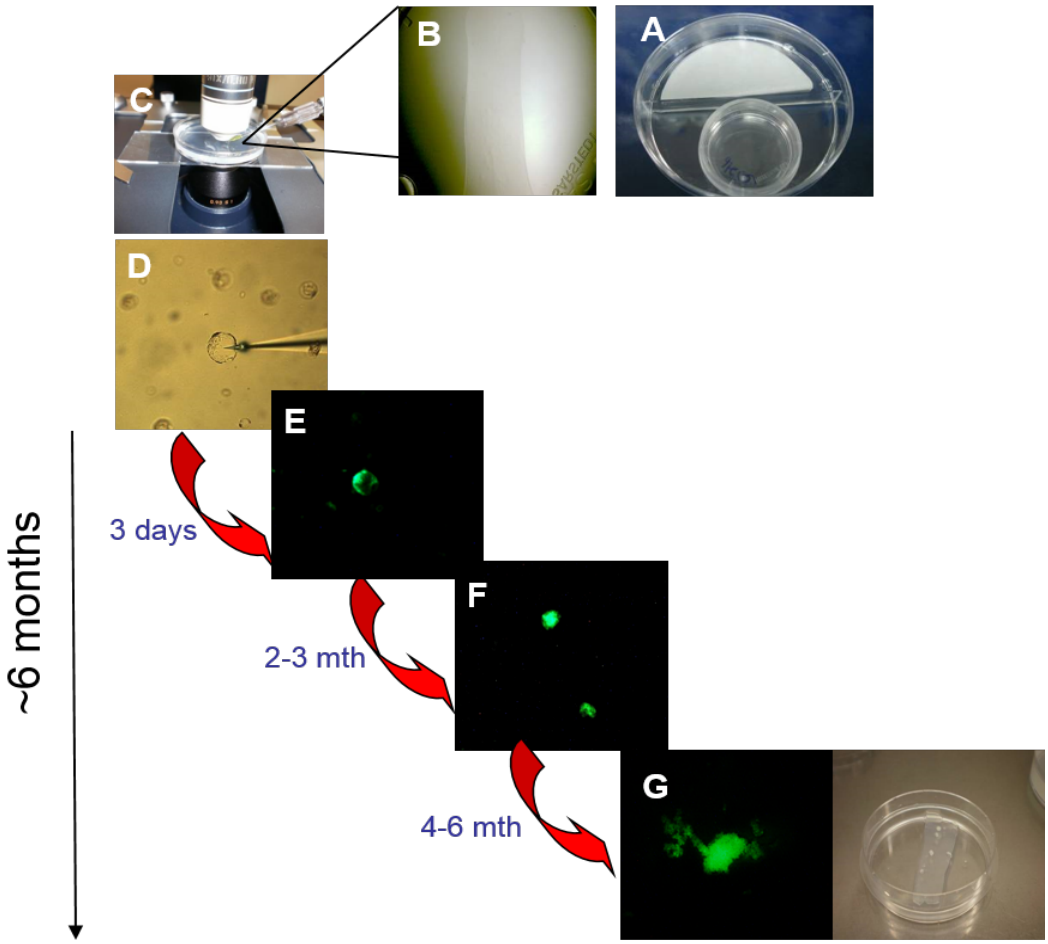


Figure 4. Microinjection of DNA into oil palm protoplasts.

Oil palm protoplasts were isolated from a 3-month-old cell suspension culture after subculture for 7 days, mixed with 1% alginate solution in Y3A medium and distributed as a thin layer onto supporting medium (A). The embedded protoplasts were arranged in a single planar layer (B) and placed on the microscope stage for DNA microinjection (C). The DNA solution was injected into the protoplast (D) and GFP fluorescence was detected in the cytoplasm after 3 days (E). The alginate layer was transferred to Y3A liquid medium and cultured at 28ºC to allow the development of microcolonies (F) and then microcalli (G).

The embedded protoplasts were cultured for 3–4 days in a two-compartment dish (Figure 4A) allowing the partial development of the cell wall, which was the ideal time for DNA microinjection. Freshly-embedded protoplasts were damaged by the procedure, demonstrating that the fragile plasma membrane alone cannot withstand penetration by the needle tip. On the other hand, if the protoplasts were left for 5 or more days, the efficiency of microinjection was limited because the cell wall was by this stage fully developed. A single micromanipulator was used to inject all protoplasts because they were immobilized within the alginate layer (Figure 4C).

We then determined the optimal DNA concentration for injection of protoplasts by comparing the transformation efficiencies achieved when injecting 50 embedded protoplasts with DNA solution at concentrations of 100 ng/µl, 500 ng/µl and 1000 ng/µl (data not shown). After one month, we recorded corresponding transfection efficiencies of 78% (39/50), 40% (20/50) and 10% (5/50), indicating that 100 ng/µl is the optimal concentration of microinjected DNA. Microcolonies developing from the protoplasts injected with 100 ng/µl DNA were observed for 2–3 months, by which time the proportion of colonies expressing GFP had fallen to 34% (17/50) (Figure 4F). GFP expression was maintained for a further 2 months but the proportion of colonies expressing GFP fell to 14% (7/50) after 6 months, when microcalli began to develop (Figure 4G).

Based on the above experiments, the injection of DNA solution at concentration of 100 ng/µl into the cytoplasm of protoplasts embedded in an alginate layer was identified as the optimal platform for the transformation of oil palm protoplasts. This resulted in approximately 14% of the injected protoplasts developing into microcalli that continued to express GFP. Although this is the first report of genetic engineering in oil palm by DNA microinjection, the transformation efficiency of 14% is far higher than that achieved using other approaches such particle bombardment (1%) (Parveez, 1998) and *Agrobacterium*-mediated transformation (0.7%) (Masli *et al*., 2009).

CONCLUSION

We successfully developed an improved protocol for the efficient isolation of high-quality protoplasts from cell suspension cultures, and have regenerated oil palm plants from protoplast cultures. The protoplasts formed cell walls after 5–7 days, sustained primary cell division after 9–12 days, developed into microcolonies in 12–16 weeks, formed microcalli in 20–24 weeks, formed embryogenic calli in 28–36 weeks, developed into greenish embryoids in 52–60 weeks and finally yielded plantlets in 76–92 weeks. We then developed a reliable PEG-mediated transformation protocol for oil palm protoplasts by establishing and validating optimal transformation parameters and the transfection procedure. We also demonstrated the transformation of oil palm protoplasts by DNA microinjection and successfully regenerated transgenic microcalli expressing green fluorescent protein. These novel transformation approaches particularly DNA microinjection, offer new routes to improve the efficiency and applicability of genetic engineering in oil palm.

ACKNOWLEDGEMENT

The authors wish to thank the Director General of MPOB for permission to publish this article. The authors would also like to thank Dr. Ahmad Tarmizi Hashim (Clonal Propagation Group Leader, MPOB) for the supply of oil palm cell suspension calli. Finally, the authors would like to acknowledge all members of the Transgenic Technology Group (MPOB) for their assistance.

REFERENCES

Cheah SC, Sambanthamurthi R, Siti Nor Akmar A, Abrizah O, Manaf MAA, Umi Salamah R, Parveez GKA (1995) Towards genetic engineering oil palm *(Elaeis guineensis* Jacq.). In: Plant Lipid Metabolism. Edited by Kader JC, Mazliak P. Netherlands, Kluwer Academic Publishers. 570-572.

Gould J, Devery M, Hasegawa O, Ulian EC, Peterson G, Smith RH (1991) Transformation of *Zea mays* L. using *Agrobacterium tumefaciens* and shoot apex. *Plant Physiol.* 95: 426-434.

Izawati AMD, Masani MYA, Ismail I, Parveez GKA (2015) Evaluation on the effectiveness of 2-deoxyglucose-6-phosphate phosphatase (DOGR1) gene as a selectable marker for oil palm (*Elaeis guineensis* Jacq.) embryogenic calli transformation mediated by *Agrobacterium tumefaciens*. *Front. Plant Sci*., 6:727. doi:10.3389/fpls.2015.00727.

Kao KN, Michayluk MR (1975) Nutritional requirements for growth of *Vicia hajastana* cells and protoplasts at a very low population density in liquid media. *Planta* 12:105-110.

Majid NA, Parveez GKA (2007) Evaluation of green fluorescence protein (GFP) as a selectable marker for oil palm transformation via transient expression. *Asia Pac. J Mol. Biol. Biotech.* 15: 1-8.

Masani MYA, Noll G, Parveez GKA, Sambanthamurthi R, Prüfer D (2013) Regeneration of viable oil palm plants from protoplasts by optimizing media components, growth regulators and cultivation procedures. *Plant Sci.* 210: 118-127.

Masli DIA, Kadir APG, Yunus AMM (2009) Transformation of oil palm using *Agrobacterium tumefaciens*. *J Oil palm Res. 21*: 643-652.

Murashige T, Skoog F (1962) A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plantarum* 15:473-497.

Nurfahisza AR, Rafiqah MA, Masani MYA, Hanin AN, Rasid OA, Parveez GKA, Ismail I (2014) Molecular analysis of transgenic oil palm to detect the presence of transgenes. *J Oil palm Res.* 26:73-80.

Parveez GKA (1998) Optimization of parameters involved in the transformation of oil palm using the biolistics method. PhD Thesis, Universiti Putra Malaysia.

Parveez GKA, Christou P (1998) Biolistic-mediated DNA delivery and isolation of transgenic oil palm (*Elaeis guineensis* Jacq) embryogenic callus cultures. *J Oil Palm Res.* 10: 29-38.

Parveez GKA, Rasid O, Zainal A, Masri MM, Majid NA, Fadillah HH, Yunus AMM, Cheah SC (2000) Transgenic oil palm: production and projection. *Biochem. Soc. T.* 28: 969-972.

Parveez GKA, Rasid OA, Masani MYA, Sambanthamurthi R (2015) Biotechnology of oil palm: strategies towards manipulation of lipid content and composition. *Plant Cell Rep.* 34(4): 533-543.

Raineri DM, Bottino P, Gordon MP, Nester EW (1990) *Agrobacterium tumefaciens*-mediated transformation of rice (*Oryza sativa* L.). *Bio/Technol.* 8: 33-38.

Ruslan A, Zainal A, Heng WY, Li LC, Beng YC, Soo Ping WY, Li Huey Y, Rashdan MM, Joseph JL, Jusoh SA, Phing LM, Sirajuddin SA (2005) Immature embryo: A useful tool for oil palm (*Elaeis guineensis* Jacq.) genetic transformation studies. *Electron. J Biotech.* 8: 25-34.

Stanton GB (1998) The introduction and expression of transgenes in plants. *Curr. Opin. Biotech.* 9: 227-232.

Teixeira JB, Sondahl MR, Nakamura T, Kirby EG (1995) Establishment of oil palm cell suspensions and plant regeneration. *Plant Cell Tiss. Org.* 40:405-411.

Transcriptomic Analyses of Water Deficit Tolerance in Oil Palm: New Ways for the Old Selection Process

Hernán Mauricio Romero[1,2,*]; Seyed Jazayeri[1]; Cristihian Bayona[1] and Leonardo Araque[1]

ABSTRACT

The development of molecular tools has opened a new era of advancements in different areas of biology, and breeding is not the exception. Genomics, transcriptomics, proteomics, metabolomics are words that we are more and more used to hear in our daily work as breeders. The use of such tools brings a promise of high achievements in terms of rate of success and the speed at which we can get the results: the new cultivars.

The breeding program at Cenipalma started to research on water deficit tolerance of oil palm because in the climate change scenario, Colombia is predicted to be strongly affected. In this way, we are expecting severe changes in the rainfall regimes which will bring with them droughts in some parts of the country, and floods in others. Oil palm is planted on regions in Colombia which will suffer from the two extreme conditions, but mostly from water deficit.

To speed up the breeding for tolerance to water deficit, we studied the physiological and biochemical changes of oil palm to water deficit. We found Elaeis guineensis genotypes with different degrees of tolerance, and found some biochemical and physiological parameters which the statistical analyses by principal components showed as the main factors for tolerance. Using the defined parameters we screened several commercial cultivars and found susceptible and tolerant cultivars. We exposed plants of these genotypes to water deficit, took RNA samples and performed a RNAseq analysis.

The results showed 2215 differentially expressed genes in tolerant vs susceptible cultivars. The general response of oil palm to water deficit involves abscisic acid signaling, stomata closure, and adjustment to osmotic and oxidative stress. The identified candidate genes, together with the parameters determined by the principal component analyses are now in used in the routine screening of progenies for water deficit tolerance in the breeding program at Cenipalma.

Overall, the results showed that the response of oil palm to water deficit is complex and involves several genes, mechanisms and processes through which the plant is able to respond to the stress.

[1]Biology and Breeding Research Program, Colombian Oil Palm Research Center, Cenipalma.
[2]Department of Biology, Universidad Nacional de Colombia.

CLOSING REMARKS

Dr N Rajanaidu

Vice-President
International Society for Oil Palm Breeders (ISOPB)

Dear Colleagues,

The theme of the International Seminar is 100 Years of Technological Advancement in Oil Palm Breeding and Seed Production. The President of ISOPB, Datuk Dr Ahmad Kushairi Din, in his Opening Remarks of the Seminar, highlighted 100 years of commercial oil palm cultivation in Malaysia. The theme of our Seminar too addressed the journey of technological advancement in the field of oil palm breeding and seed production for the past 100 years and charting future direction.

The Seminar started off with the Keynote Address by Prof Kai-Jun Zhao of National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI), Institute of Crop Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing. Prof Kai-Jun covered emerging genome editing tools that can precisely modify DNA sequences at the genomic level. In addition, he overviewed the structure and mechanism of genome editing tools and their potential applications in crop breeding. In the case of oil palm, it is worthwhile to apply the technique to overcome male sterility problem in inter-specific hybrids between Oleifera x Guineensis.

The main Seminar covered topics related to Genomic Selection, New tools and Equipment to estimate oil in dry mesocarp, molecular markers to identify fruit forms and tissue culture abnormal plantlets at the nursery stage and genome-based large- scale screening of seedling to estimate non-*tenera* contamination. Other topics in the programme are Bar Code System for Seed Production, Data Logger to capture data in field breeding trials, FGV Breeding System for managing breeding data, MPOB Breeding Information System and Oil Palm Protoplast Transformation System.

Two papers were presented on Genomic Selection (GS). GS has the potential to make oil palm breeding more efficient. An efficient and effective implementation of GS requires more integrated approach. Relevant training, test sets, high-throughput genotyping on large scale and interaction between breeders, molecular geneticists, bioinformaticians and statisticians is vital. CIRAD paper reviews their research on GS in oil palm for the past 10 years. Apparently, it is

cost effective to preselect the *pisifera/tenera* group for progeny test by applying GS to key yield traits therefore increasing the selection intensity.

Sime Darby GS paper highlights a method for accelerated improvement of oil palm. It describes the development of a SNP genotyping array, genome-wide association analysis (GWAS) for yield components, GS methodology and deployment for selective oil palm breeding. At present, a total of 18 000 barcoded GenomeSelect^{TM} seedlings are in the nursery and one year old palms in the field are bearing bunches. Oil palm breeders look forward to the yield data between GS/GWAS select palms and RRS/MRS based DxP in five years.

Next presentations dealt with oil content in dry mesocarp. The SMART and CIRAD paper on Near Infrared Spectroscopy (NIRS) for Rapid Determination of Oil Content in Dried Oil Palm Mesocarp. Their study showed significant correlation (r^2= 0.977) between reference and predicted value. NIRS is used in their bunch analysis laboratory as an alternative to the Soxhlet methos which uses N-Hexane which is toxic and harmful.

AAR paper on Application of Near Infrared Spectroscopy (NIRS) to determine Oil to Dry Mesocarp and Iodine Value for Oil Palm Breeding. NIRS is being used for analysing multiple parameters for crude palm oil samples simultaneously. Initially, AAR developed calibration model for Oil/Dry Mesocarp, and Iodine Value using conventional method as reference. A validation step was performed using independent set of samples. AAR results showed coefficient of determination (R^2) for O/DM=0.889-0.959; I.V = 0.892-0.959.

Dr Meilina Ong Abdullah delivered a paper on Leveraging Agritechnology for Posterity of Oil Palm Industry. SureSawit Shell Kit^{R} commercialised MPOB used to determine *dura, tenera, pisifera* forms and nigrescens, virecens fruit colours at seedling stage. In addition, Dr Meilina outlined the tissue culture plantlets production process, abnormality fruits (mantling), epigenetics mechanism and KARMA markers to screen tissue culture ramets to weed out abnormal plantlets before field planting.

Mr Nathan Lakey of Orion Biosains shared information on the screening facilities at their laboratory at Puchong, Selangor and the level of non-*tenera* contamination (10%) based on extensive screening of nursery seedling in Sabah, Sarawak and Peninsular Malaysia. Presentations by Mr Indra Syahputra, Socfindo on Bar Code System for Seed Production; Ms Zuraini Shaharudin of Sawit Kinabalu on Data Logger; Mr Muhammad Farid, FGV on Managing Integration of Breeding Data and Operation; and MPOB Breeding Information System by Dr Mohd Din Amiruddin provide overview of tools and software used to manage breeding data efficiently.

The paper on Protoplast Transformation System as a Potential Platform for Oil Palm Genetic Engineering by Dr Abdul Masani Mat Yunus has an implication for oil palm breeding especially modifying the fatty acid composition of palm oil.

ISOPB wish to thank members of the Organising Committee for their time and effort especially the Secretary, Dr Zulkifli Yaakub. The Executive Committee has not decided the theme and venue for the next Seminar. If the members have a proposal, please e-mail to us.

Thank you.

INTERNETIONAL SEMINAR ON **100 YEARS OF TECHNOLOGICAL ADVANCEMENT** IN OIL PALM BREEDING & SEED PRODUCTION

MPOB

MALAYSIAN PALM OIL BOARD
Ministry of Plantation Industries and Commodities, Malaysia
www.mpob.gov.my