Development and applications of the Oil Palm 78K Infinium® HD SNP Array for linkage analysis and chromosome scanning

Ting, N. C., Ordway, J. M., van de Weg, E., Mohamed Serdari, N., Low, E. T. L., Mustaffa, S., Wischmeyer, C., Smulders, M. J. M., Sambanthamurthi, R., & Singh, R.

Ting, N. C., Ordway, J. M., van de Weg, E., Mohamed Serdari, N., Low, E. T. L., Mustaffa, S., Wischmeyer, C., Smulders, M. J. M., Sambanthamurthi, R., & Singh, R. (2023). Development and applications of the Oil Palm 78K Infinium® HD SNP Array for linkage analysis and chromosome scanning. *Scientia Horticulturae*, *318*, Article 112104. DOI: 10.1016/j.scienta.2023.112104

Development and applications of the Oil Palm 78K Infinium® HD SNP Array for linkage analysis and chromosome scanning

Ting Ngoot-Chin[1], Jared M Ordway[2], Eric Van de Weg[3], Norhalida Mohamed Serdari[1], Leslie Low Eng Ti[1], Suzana Mustaffa[1], Corey Wischmeyer[2], Marinus J. M. Smulders[3], Ravigadevi Sambanthamurthi[1], Rajinder Singh[1+*]

[1]Malaysian Palm Oil Board (MPOB), Advanced Biotechnology and Breeding Centre, 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia

[2]Orion Genomics, Saint Louis, Missouri, USA 63108

[3]Plant Breeding, Wageningen University and Research, Wageningen, The Netherlands

*Corresponding author email: rajinder@mpob.gov.my; ORCID ID: 0000000199334884; telephone: +60387694501

Ting Ngoot-Chin: tnching@mpob.gov.my
Jared M Ordway: jordway@oriongenomics.com
Eric Van de Weg: eric.vandeweg@wur.nl
Norhalida Mohamed Serdari: norhalida@mpob.gov.my
Leslie Low Eng Ti: lowengti@mpob.gov.my
Suzana Mustaffa: suzana.mustaffa@mpob.gov.my
Corey Wischmeyer: cwischmeyer@oriongenomics.com
Marinus J. M. Smulders: rene.smulders@wur.nl
Ravigadevi Sambanthamurthi: ravigadevi@gmail.com
Rajinder Singh*: rajinder@mpob.gov.my

ABSTRACT

High-throughput and high-density (HD) genetic marker genotyping systems are critical to optimize the efficiency of oil palm breeding and improvement programmes. This study reports the development of the 78K Infinium® HD customized SNP array, which was used to genotype a thousand palms of a commercial Deli *dura* x AVROS *pisifera* family. A total 64,108 (~82.0 %) polymorphic SNP markers were identified of which, 57, 465 (89.6 %) were mapped onto the genetic map that has the largest number of markers published so far in oil palm and holds 14,781 SNPs on 2,363 orphan scaffolds (whose chromosomal locations were unknown), which will improve the existing oil palm reference genome (EG5.1). The SNPs were highly informative based on the parent-to-progeny allelic inheritance analysis. The data demonstrated that 4.3 % of the progeny resulted from unintentional self-fertilization of the *dura* female parent. These unintended 'selfs' are highly inbred, which will affect their yield. This study also for the first time, describes the homozygosity in the Deli *dura* and AVROS *pisifera*, two important parental lines widely used in commercial seed production. As expected, both the parental palms were highly homozygous, having 138 Mb homozygous regions in common, with 70.3 % identical alleles. Such a detailed genetic analysis of the individual palm has been made possible with this customized HD SNP array, which will be a valuable tool for routine application in oil palm improvement programmes. The strategy used to design and apply the array will also be of interest for wider scientific research.

Keywords: Elaeis guineensis; advanced breeding lines; germplasm; chromosomal aberration; inbreeding; marker-assisted selection

1. Introduction

The oil palm (*Elaeis guineensis* Jacq.) is a monoecious diploid ($2n = 2x = 32$) producing a large fruit bunch of 15 – 25 kg with 500 – 4,000 fruits (Hartley, 1967). Its valuable products include the oils extracted from the fruit mesocarp (palm oil) and kernel (palm kernel oil). The majority of the world's populace relies on only a handful of vegetable crops for their daily oils and fats requirement, indicating the importance of these oil crops in meeting the dietary requirement of humans. Plants contribute to almost 89.0 % of the world’s edible oils and fats (209 of 236 million tonnes), with oil palm being the single largest plant-based contributor (82 million tonnes; Oil World, 2021).

Moving forward, the oil palm yield is key to supplying the increasing demand for oil from the growing global population (expected increase from the current 7.8 to 9.7 billion by 2050; https://www.worldometers.info/world-population/world-population-projections/). Increasing productivity is highly desirable since the agricultural sector is facing severe challenges such as a lack of arable land for expansion and the effects of climate change, which are impacting yield (Zhang and Cai, 2011; Sarkar et al., 2020). Increasing productivity while better managing available natural resources including land is critical for long term viability and mitigating the negative environmental effects of production. As such, there is a push to develop new and better varieties, having higher yields per unit area and increased resilience to pests and diseases as well as tolerance to climate change (Ahmad Parveez et al., 2021). However, improving perennial crops such as oil palm has always been a challenge because of its long breeding cycle of at least 10 years in conventional breeding (Rajanaidu et al., 2000). Although oil palm is by far the most productive oil crop (current average oil yield 4 – 5 tonnes/hectare/year, about 10 times more than other crops; Ahmad Parveez et al. (2020 and 2022)), it is still far below its theoretical potential of 18.2 tonnes/hectare/year (Corley, 1998). Selective breeding of oil palm was only initiated about 100 years ago (Ahmad Malike et al., 2019; Yue et al., 2021). It is thus highly amenable to further yield improvement via genomics-guided breeding (Nyouma et al., 2020; Ithnin et al., 2021; Kalyana et al., 2021). Early prediction of phenotypes using DNA markers will improve the speed and precision of breeding and hasten the development of new planting materials with desired traits.

A major step forward in oil palm genomic research was the release of its genetic blueprint as a draft whole-genome sequence (WGS; Singh et al., 2013a). This paved the way for effectively utilizing the genome information to expedite the development of improved planting materials (Babu et al., 2021; Yue et al., 2021). This strategy has been successfully utilized in other crops, where the availability of genetic blueprints has accelerated the identification of DNA markers linked to important traits and as such, facilitated marker-assisted selection (MAS) or genomics-guided breeding programmes (Yang et al., 2015; Soriano, 2020). Of the markers, single nucleotide polymorphisms (SNPs) are the most popular, not only in plants but also in humans and animals, due to their abundance in the genome. Advances in sequencing have accelerated SNP discovery, while improvements in genotyping technologies have enabled the development of high-throughput assays of SNPs. SNP arrays with increasing density have been reported for potato (8.3 – 20K), apple (8 – 480K), maize (55 – 600K), rice (7 – 700K) and wheat (50 – 820K) (Hamilton et al., 2011; Unterseer et al., 2014; Vos et al., 2015; Chagné et al., 2012; Bianco et al., 2016; McCouch et al., 2016; Xu et al., 2017; Winfield et al., 2016; Morales et al., 2020; Lv et al., 2021).

SNP markers in oil palm were initially developed exploratorily with the first pilot array only containing 96 markers (Singh et al., 2011), which was later expanded to 4,451 (Ting et al., 2014, Low et al., 2020). The release of the draft WGS (Singh et al., 2013a) has facilitated the development of a HD array with 170,860 SNPs (Kwong et al., 2016), which enabled the identification of the markers for oil-to-dry-mesocarp content (Teh et al., 2016) and trunk height (Teh et al., 2020). Apart from arrays, oil palm research has also employed some alternative methods to detect SNP profiles including genotyping-by-sequencing (GBS) (Pootakham et al., 2015; Gan et al., 2018; Bai et al., 2018; Xia et al., 2019; Herrero et al., 2020), again made possible by the availability of the oil palm genome build (Singh et al., 2013a), to anchor the sequences. The essential characteristics of stable inheritance from one generation to another (Ye et al., 2016; Ooi et al., 2019), amenability to high-throughput analysis and more importantly, their frequent occurrence in oil palm (every 99 bp; Jin et al., 2016), have enabled quick adoption of SNP markers in oil palm research. Among the many SNP genotyping strategies available, the array-based platform produces high-quality data that can be easily shared and reproduced among laboratories performing similar research. This quality enables the establishment of consortiums, such as the International RosBREED SNP Consortium (Chagné et al., 2012), to develop, validate and apply functional SNPs as a common resource for all researchers and companies working on a crop, an example that can be a model for the oil palm community.

Array-based SNP data have been used to generate accurate HD genetic maps that allowed the identification of genomic loci for important economic traits for potential application in breeding. In wheat, the QTLs associated with height and grain/kernel were identified (Cui et al., 2017; Lv et al., 2021), and in oil palm those associated with fatty acid composition in interspecific hybrids have been transferable to its backcross generations (Ting et al., 2016). In addition, HD genetic maps together with the availability of WGS are facilitating a better understanding of the genome, including the evolutionary and recombination landscapes of crop plants. The combined information of genome sequences and SNPs in HD genetic maps have been especially effective in comparative genomics, for example, between radish and its close relatives in the *Brassicaceae* family, where the syntenic regions and rearrangements in the radish genome were precisely determined (Luo et al., 2020). SNP genetic maps with high-quality annotation have also led to the identification of candidate gene(s) regulating important traits in the woody trees *Prunus mume*, *Catalpa bungee* and *Camellia sinensis* (Zhang et al., 2015; Lu et al., 2019; Wei et al., 2022), wheat (Cui et al., 2017) as well as oil palm (Ting et al., 2016 and 2021a; Ong et al., 2020). SNP array data can also be used to identify selfed offspring in a progeny population e.g. in tetraploid rose (Vukosavljev et al., 2016), to validate parent-offspring relationships (Muranty et al., 2020) and to reconstruct direct and distant ancestors (Van de Weg et al., 2018; Howard et al., 2021).

In this study, we present the development of a customized Infinium® HD SNP array with 78,196 SNPs and its use for a detailed analysis of a large F_1 oil palm mapping family including a scan of chromosomal aberrations and construction of a high-resolution whole-genome linkage map which, for the first time, reveals details of levels of inbreeding in an advanced breeding population. The array, with SNPs that are polymorphic in a wide collection of germplasm and advanced breeding lines (ABLs), represents an important genomic platform for the oil palm industry, with the strategies applied to its design and use of interest to the larger research community.

2. Materials and methods

2.1. Mapping family

The oil palm F_1 family studied comprised 1,011 palms from a cross between a Deli *dura* (Palm 0.338/391) and an 'Algemene Vereniging van Rubberplanters ter Oostkust van Sumatra' (AVROS) *pisifera* (Palm 0.174/655), hereinafter called the PUP family. Both the parental palms are among the most widely used parents in commercial seed production, derived from a systematically designed modified recurrent selection (MRS) breeding scheme (Rosenquist, 1990) and were planted at MPOB

Kluang, Johor. The PUP progeny are grown at KULIM Plantations Berhad, Kota Tinggi, Johor. Culling was not carried out as usual in the pre- and main-nursery; rather the palm abnormalities were recorded and this facilitated a study of the chromosomal aberrations in the off-type palms as described by Ting et al. (2021b). Leaves collected from the seedlings were first cleaned with paper towel, cut into small pieces (diameter ~0.5 cm) using a paper hole puncher. Three leaf pieces per sample were placed in the 96-well plate for DNA extraction. For storage, the leaves were cut into small pieces (~7.0 x 3.0 cm), packed in the labelled plastic bag (~10.0 x 9.0 cm) and stored at -80 ºC.

2.2. Development of the oil palm 78K Infinium® HD SNP array

The oil palm 78K Infinium® HD SNP array was designed to capture not only informative SNPs within the PUP family described herein, but also variation across other mapping families, ABLs as well as the MPOB's vast germplasm collection. The selected palms were sequenced to an average of 50X raw whole genome coverage (median 50X, minimum 30X and maximum 68X) using a combination of 300 bp paired-end Illumina MiSeq v3 sequencing, 150 bp paired-end Illumina MiSeq v2 sequencing and 100 bp paired-end Illumina HiSeq 2500 sequencing. High quality trimmed reads were aligned to a revised version of the reference *E. guineensis* genome (Singh et al., 2013a; Sanusi et al., 2023) namely, EG5.1 using Bowtie (Langmead et al., 2009) and candidate SNPs were called using SAMtools (Li, 2011).

Candidate SNPs selection was performed with an in-house customized PERL script. To capture a wide range of genetic variation, initial selection criteria included heterozygosity in one or both parents of the mapping families (in particular PUP) and/or polymorphism among the germplasm cohort. This was performed by examining both sequence uniqueness within the reference genome and absence of other SNP among the germplasm cohort, across a 60 bp window spanning each candidate SNP position. Subsequently, the reference genome was divided into 15 Kb sequence windows and one SNP per window was selected with a preference given, when possible, to those that were both informative in PUP and germplasm. The selection also included a set of 1,291 duplicated SNPs (3.5 %) and 3,744 SNPs (SNPMxxxxx) from the previous oil palm Infinium® SNP array − OPSNP3 (Ting et al., 2014; Low et al., 2020), to ensure compatibility with earlier research.

2.3. DNA isolation and SNP genotyping

Genomic DNA was extracted using a modified CTAB method optimized to a semi-robotic or manual grinding system. In the semi-robotic system, freshly collected green leaves were smashed with beads in a 96-well plate and the extraction was carried out by a service provider (the Centre for Marker Discovery and Validation, Malaysia). For a subset of the PUP progeny, DNA was extracted from the frozen leaves (stored at -80 ºC) using the conventional method as described by Rahimah et al. (2006) and Suzana et al. (2015). The extracted DNA was purified and checked for quality via A_{260}/A_{280} ratio (> 1.8) and electrophoresis on a 0.8 % agarose gel. Downstream DNA processing and hybridizations on the Infinium® HD iSelect BeadChips array (Illumina Inc., USA) were performed by the Genome Technology Access Center (GTAC) at Washington University, USA.

2.4. Genotype calling and data evaluation

Genotype calls were made using the Genome Studio® software (http://support.illumina.com/array/array_software/genomestudio/downloads.html). The sample sheet, intensity data and SNP manifest files generated from the iScan™ system were imported as described in the software manual and the auto-clustering function was used to call the SNPs as *aa*, *ab* or *bb*. A GenTrain (GT) score closer to 1.0 indicates better clustering for the SNPs. However, for the SNPs with GT < 0.7, their clustering was re-examined and improved manually. The genotype calls of SNPs that did not amplify or failed to form a well-separated cluster, were excluded from the analysis. In addition, the independent genotype calls for the two replicates of each parental palm were evaluated to ensure reproducibility of the calls for the SNPs.

The genotype data was further evaluated using the factorial analysis and NJ tree implemented via DARwin 6.0.21 (Perrier et al., 2003). A panel of 77 representative SNPs across 16 chromosomes, as assigned in EG5.1 were selected to determine the fidelity of the controlled cross. The panel consisted of three to seven evenly distributed SNPs on each chromosome, selected from an average interval of 9.3 – 12.6 Mb, depending on chromosome size (Supplementary Table A.1). For both the factorial analysis and NJ tree, four foreign (i.e., not related to PUP) palms genotyped with the SNP array were included as a reference to identify contaminants. Also, all PUP palms known to have divergent ploidy levels in at least one of their chromosomes (Ting et al., 2021b) were included as a reference to identify chromosomal aberrant palms. Of the 77 SNP markers, 57 (74.0 %) were polymorphic in the F_1 family with an average polymorphic information content (PIC) of 0.31. The remaining 20 were monomorphic in the PUP family. Computation of the PIC and observed and expected heterozygosity for each marker was carried out using CERVUS 3.0.7 (Kalinowski et al., 2007). Inbreeding coefficient (= $1-\frac{number\ of\ observed\ heterozygosity}{number\ of\ expected\ heterozygosity}$) was calculated assuming that the population was in Hardy-Weinberg equilibrium.

2.5. Evaluation of chromosomal aberration by SNP B-allele frequency (BAF) distribution

Evaluation of chromosomal aberration was carried out as per Chagné et al. (2015) and Ting et al. (2021b). In brief, the BAF value for each SNP called from Genome Studio® was plotted (i.e., scatter-plot) to its corresponding position on the pseudo-chromosome of EG5.1. For good visualization, only the SNP markers with a GT ≥ 0.7 were used. The expected BAF values in a standard diploid palm for the genotypes *aa*, *ab* and *bb* of ideal performing markers are 0, 0.5 and 1.0, respectively, whereas values of ~0, 0.33, 0.66 and 1.0 are likely for the trisomy genotypes *aaa*, *aab*, *abb* and *bbb*, respectively.

2.6. Construction of genetic linkage maps

Genetic linkage maps for the two parental palms and their integrated map were constructed using JoinMap® 5 (Van Ooijen, 2018). Grouping of markers to form each linkage group (LG) was performed at ≥ LOD 7, as at this LOD score, markers in the group had remained intact, up to the maximum LOD of 14. For calculation of LG, the Maximum Likelihood (ML) mapping and Haldane map distancing functions were applied. As a large number of SNP markers were involved, only one marker representing each cluster of co-localizing markers was selected from the initial map constructed and used for subsequent optimization. For optimizing marker order in each LG, ML calculation was repeated 10 – 90 rounds (depending on number of markers grouped) at maximum chain lengths of 200. Only markers with nearest neighbour stress (N.N. Stress) of < 1.0 cM and highly plausible order were retained in the final genetic map. Using the default setting in JoinMap® 5, the statistical counts (ranging from 1 to 500 in each parent) of plausible positions for each marker in the estimated map order could be examined using the *Plausible Positions P1* and *P2* tabsheets, after which, other co-localizing markers were placed accordingly in the final resulting map.

3. Results

3.1. Oil palm 78K Infinium® HD SNP array

A total of 108,543 SNPs were selected for probe design, of which 59,395 and 49,148 were identified from the mapping families/ABLs (including PUP) and the wide germplasm cohort, respectively following the selection criteria described in Materials and methods. Following the assessment of probe quality and bead design, 78,196 SNP markers (named as OPHRMxxxxxx) consisting of 72.1 % transition (28,142 T/C and 28,239 A/G) and 27.9 % transversion (2,403 C/G, 4,076 A/T, 7,683 A/C and 7,653 T/G) SNPs were synthesized. Two probes per SNP target were designed for 63 A/T (T/A), 27 C/G (G/C), 951 A/G (T/C) and 250 A/C (T/G) resulting in a total of 1,291 duplicated SNPs which could be used to evaluate reproducibility of the genotype data generated for the PUP mapping population.

Most of the SNPs (83.3 %) had common alleles between germplasm and mapping families/ABLs (Table 1) and only a small proportion of about 9.8 % were unique either to the germplasm cohort (8.5 %) or the mapping families/ABLs (1.3 %). In the current SNP array, of the 78,196 SNPs, 51,921 (66.4 %) were well distributed across the 16 pseudo-chromosomes, based on the EG5.1 reference genome (Fig. 1) whereas, the remaining 33.6 % SNPs were located in various orphan scaffolds.

3.2. Evaluation of PUP

Approximately 73K SNPs (94.0 %) could be called when the 78K SNP array was used to genotype the 1,011 palms of the PUP family. Looking closer at the 1,291 duplicated SNPs, comparison was made for 1,035 pairs because the others (256 pairs) were either not of comparable quality (either having too many e.g. ≥ 100 missing data points in one of the SNPs or only one of the replicates was scorable). Complete reproducibility of genotype calls among the PUP progeny was observed in 964 (~93.1 %) of the duplicated SNPs that were compared whereas, 44 pairs (4.3 %) showed reproducibility ranging from 97.0 – 99.9 %. It is worth noting that 27 pairs of the duplicated SNPs (2.6 %) failed to reproduce the

SNP cluster plots although the normalized intensity (Norm R) values were good. Among these SNP pairs, one showed a clear heterozygote clustering while the other duplicated SNP formed an unseparated cluster suggesting either as *aa*, *ab* or *bb* genotype call.

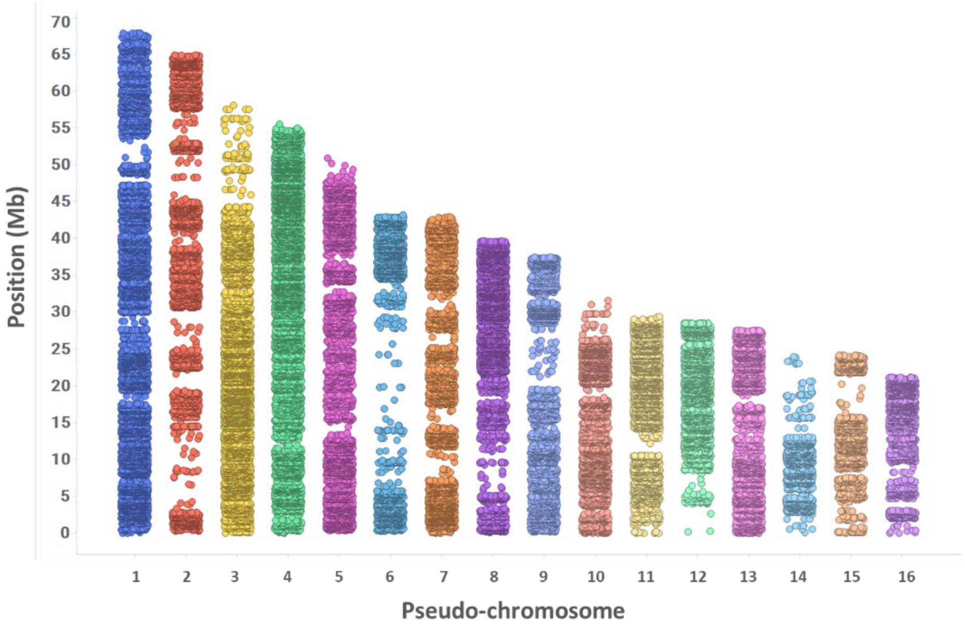

Fig. 1. Distribution of designed SNPs in the 16 pseudo-chromosomes in oil palm EG5.1. Data points are stretched along the x-axis (pseudo-chromosome), but not the y-axis (bp, position) to better display the local SNP density. Note that regions of lower SNP density generally correspond to regions of high repeat content (Singh et al., 2013a). The SNPs on orphan scaffolds are not shown here.

Table 1. Composition of the Oil Palm 78K Infinium® HD SNP Array.

SNP origin	EG5.1	Oil Palm 78K Infinium® HD SNP Array	
		Number of SNP	Percentage (%)
Unique to germplasm	Pseudo-chromosome	2,009	2.57
	Orphan scaffold	4,622	5.91
	Sub-total	6,631	8.48
Unique to mapping family/ABL	Pseudo-chromosome	488	0.62
	Orphan scaffold	515	0.66
	Sub-total	1,003	1.28
Common to germplasm and mapping family/ABL	Pseudo-chromosome	45,971	58.79
	Orphan scaffold	19,145	24.48
	Unmapped	8	0.01
	Sub-total	65,124	83.28
Other	Pseudo-chromosome	578	0.74
	Orphan scaffold	1,116	1.43
	Sub-total	1,694	2.17
OPSNP3	Pseudo-chromosome	2,875	3.68
	Orphan scaffold	869	1.11
	Sub-total	3,744	4.79
Total	Total	78,196	100.00

The genetic relationships and diversity among the PUP progeny and their parental palms were first evaluated using a panel of 77 representative SNP markers (see Materials and methods and Supplementary Table A.1). The factorial analysis revealed an inertia dissimilarity of only 6.57 %, indicating strong relationships among the family members. The percentage inertia increased slightly to 6.61 % when four unrelated palms were included as a reference to identify contaminant(s) in the factorial analysis; the difference was clearly contributed by these four palms. Interestingly, 43 PUP palms formed an isolated cluster with the Deli *dura* maternal palm. A similar clustering profile was also observed in the Neighbour-Joining (NJ) tree (pink, Fig. 2a and b). In addition, nine palms with aberrant chromosomes (brown, Fig. 2b) showed genotype differences and could be distinguished from their other siblings on the NJ tree. Considering the 57 polymorphic SNPs (*aa* x *ab*, *ab* x *aa* and *aa* x *bb*, *ab* x *ab*) from the representative panel of 77 SNPs, 34 (*aa* x *ab*) were homozygous (*aa*) in the 43 palms clustering with the *dura* maternal palm, while the average heterozygosity level observed in the other PUP progeny palms was 0.52 and a value of 0.57 was recorded in the nine suspicious palms. Furthermore, the remaining 23 of the 57 polymorphic SNPs revealed that these 43 *dura*-like palms also had a much higher inbreeding coefficient (-0.023) than those observed in the other PUP palms (-0.265) as well as in the nine suspicious palms (-0.352). Further parentage analysis of the 43 palms of the isolated cluster using CERVUS 3.0.7 revealed negative LOD scores from -8.96 to -38.89, suggesting it was unlikely that the AVROS *pisifera* was the paternal palm. Instead, these 43 palms likely resulted from self-fertilization of the *dura* mother ($p < 0.05$), as under this model their trio LOD scores were positive from 15.71 – 29.04.

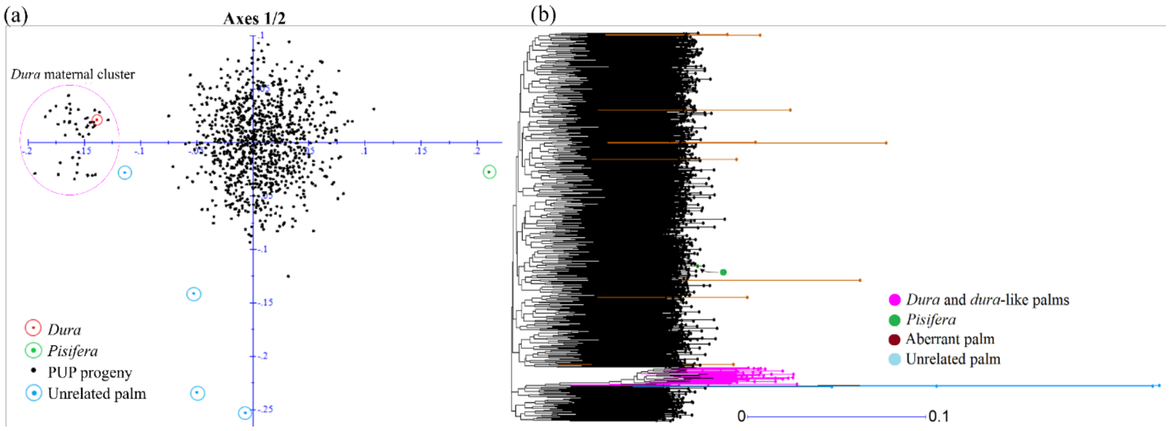

Fig. 2. Genetic relationships and diversity in PUP and between them (black dots) and their two parents (*dura* and *pisifera*) revealed by factorial analysis (a) and NJ tree (b). Four unrelated palms (represented by light-blue encircled dots) were included for comparison.

We expanded the analysis on the 43 maternal-like palms using a larger set of 39,908 good-quality (GT ≥ 0.7) SNPs where, three profiles were observed – three profiles were observed - *aa* x *bb* (170 SNPs), *aa* x *ab* (24,881 SNPs) and *ab* x *aa* (14,857 SNPs) (Supplementary Table A.2). Of these, 170 SNPs were genotyped as *aa* in the *dura* mother and *bb* in the intended *pisifera* father (classified as the *aa* x *bb* segregation profile in the two parents) and thus, should theoretically show *ab* genotype in the progeny. Interestingly, genotypes scored in all the 43 maternal-like progeny were *aa* which, could only result from the unintended selfing (*aa* x *aa*) of the maternal palm. A similar observation was made for these 43 palms with respect to 24,881 SNPs which, clearly revealed the *aa* and *ab* genotypes in the maternal and paternal palm, respectively. However, only one of the SNPs (0.004 %) showed an *ab* genotype, even that only in one (PUP2053/306) of the 43 palms. The observation that all these SNPs revealed an *aa* genotype in the 43 palms (with the single exception stated earlier) further supports the fact that these 43 palms were a result of selfing of the maternal palm. Moreover, analysis of 14,857 SNPs where the observed parental genotypes were *ab* and *aa*, revealed the presence of all three possible genotypes - *ab*, *aa*, and *bb* - in the 43 palms, as expected from selfing of the heterozygous maternal palm. Further examination of the distribution of BAF showed a substantial loss of heterozygosity (LOH), similar to that observed in *dura* maternal palm. Homozygous segments (HS) of various sizes were found in 12 – 16 chromosomes (Fig. 3). A few chromosomes were almost entirely homozygous, like CHR05, where the LOH appeared fixed in these maternal-like palms (discussed in greater detail below). A validation was carried out by testing six randomly chosen palms (Supplementary Table A.2) with the *SHELL* gene assay (Ooi et al., 2016; Singh et al., 2020), in which they were found to be indeed *dura*, and later all the 43 palms were confirmed *dura* by the physical examination of their fruit form.

A closer analysis on the identity of the nine suspicious palms using CERVUS 3.0.7 and the panel of 77 SNP markers showed that eight were legitimate, i.e., the Deli *dura* and AVROS *pisifera* offspring ($p < 0.05$) based on their progeny-maternal-paternal trio LOD scores (4.68 – 9.19) and there was no sign of maternal selfing (LOD -51.84 – -77.75). The exception was PUP2053/726B where the negative trio LOD scores implied that neither the *dura* (-61.25) nor the *pisifera* (-7.20) were its parents. A previous examination of its BAF profile (Ting et al., 2021b) had shown that most of its chromosomes occurred in triplicate which could result in unexpected parent-to-progeny genotypes. For an example, for each of the *ab* calls in PUP2053/726B, the parental genotypes were not *aa* or *bb*. In fact, Ting et al. revealed that all the nine palms had chromosome aberration including aneuploidy and triploidy. Therefore, they were removed from the linkage analysis.

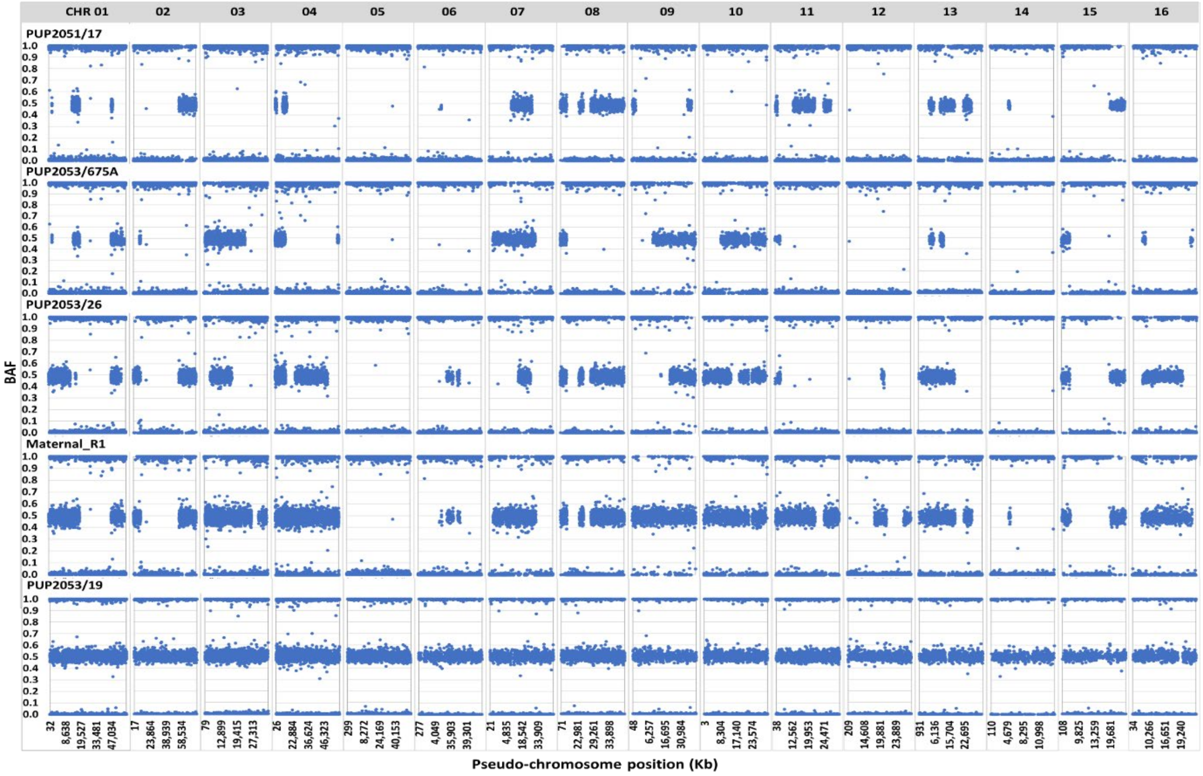

Fig. 3. The BAF distributions across the 16 pseudo-chromosomes of three *dura*-clustered palms in comparison with the *dura* mother (Maternal_R1) and a normal heterozygous palm (PUP2053/19), randomly selected from the other offspring. BAF distribution for all the 43 *dura*-clustered palms is available in Supplementary Fig. A.1.

3.3. PUP genetic linkage map

The genetic linkage map was constructed in two steps. An initial frame work map included 694 of 1,011 palms, where those showing the following characteristics were excluded: i. low heterozygosity (i.e., the 43 *dura* 'selfs'); ii. loose clusters in SNP calling (including the nine aberrant palms); iii. high recombination count in any LG; and iv. with > 5.0 % missing data. This genetic map was used as a reference for reinstating 264 out of the 317 earlier rejected palms from categories ii – iv, resulting in a total of 958 palms for the second round of map construction and leaving out only those with unexpectedly high recombination count in a specific LG.

The array only had a small proportion (6.4 %) of SNPs not called, where a total of 73K were successfully genotyped in the samples. The homozygous and irreproducible markers in the two parental palms were then discarded. This resulted in 64,108 informative markers with the following segregation types: *ab* x *aa* (23,300 SNPs), *aa* x *ab* (35,100 SNPs) and *ab* x *ab* (5,708 SNPs). This SNP collection was used to construct the individual LGs. Of these, 57,465 (89.6 %) were assembled into 16 chromosomes, although some had two to four sub-LGs (Tables 2, 3 and Fig. 4). A highly consistent marker order was observed when each of the PUP genetic LGs was compared to other published DxP genetic maps e.g. P2 and KULIM DxP (Zolkafli et al., 2021) using anchor SNP (SNPMxxxx) markers (Fig. 4). As expected, the present PUP LGs have a notably higher density of SNPs with many co-localizing, resulting in 1,664 clusters, each containing ≥ 10 SNPs. The PUP integrated LGs comprised 49,070 SNP markers, with a total map length of 1,521.0 cM whereas the parental sub-LGs in CHR01, 02, 05, 06, 07, 14 and 15 (not included in the integrated map) were 566.3 cM in total length. They could not be integrated due to the absence of SNP segregating from both parents (*ab* x *ab*). This is demonstrated in Fig. 5 where the BAF of markers on 16 chromosomes of the two parental palms is linked to their corresponding LGs.

Aligning the PUP LGs to their corresponding pseudo-chromosomes in the EG5.1 genome build also revealed 43 possible segmental duplications, with at least one SNP mapping to different chromosomes (Fig. 5). Of the duplications, only four (CHR02 and 08, CHR01 and 06, CHR04 and 11, and CHR12 and 16) were similar/close to those reported by Singh et al. (2013a) while the others were either specific to PUP or not observed in EG5.1. In addition, the alignment also revealed 2,363 orphan scaffolds, represented by 14,781 SNPs placed on the linkage map (Table 3), which now allows these scaffolds to be assigned to their respective pseudo-chromosomes.

Table 2. Summary of number of polymorphic SNP markers scored from the Oil Palm 78K Infinium® HD SNP Array and mapped in the PUP genetic linkage maps.

78K SNP	EG5.1	Polymorphic SNP		Number of SNP mapped in			
		Number of SNP	Percentage (%)	Parental map	Percentage (%)	Integrated map	Percentage (%)
All	Pseudo-chromosome	45,912	58.71	42,405	54.23	36,146	46.22
	Orphan scaffold	18,190	23.26	15,055	19.25	12,828	16.40
	Unmapped	6	0.01	5	0.01	5	0.01
	Total	64,108	81.98	57,465	73.49	48,979	62.64
SNPMxxxxx only	Pseudo-chromosome	602	0.77	602	0.77	499	0.64
	Orphan scaffold	110	0.14	110	0.14	90	0.12
	Total	712	0.91	712	0.91	589	0.75

Table 3. Summary of PUP genetic linkage map. Integrated linkage groups are labelled Cxx, *dura* and *pisifera* parental linkage groups that could not be integrated as DCxx and PCxx, respectively.

Chromosome (CHR)	Linkage group (LG)	Number of SNP marker	Map length (cM)	Number of cluster (containing ≥ 10 co-localized SNPs)	Number of orphan scaffold in EG5.1
1	C01_1	1,850	47.2	50	13
	C01_2	3,017	161.4	101	31
	DC01_1b	91*	3.9	2	5
	DC01_2b	16	34.7	0	2
2	C02	4,611	208.3	129	219
	DC02_1	320	21.4	10	4
	DC02_2	12	18.4	0	9
3	C03	5,174	120.9	168	147
4	C04	7,770	161.9	229	375
5	DC05_1	18	146.8	0	5
	DC05_2	77	0.7	1	4
	PC05	4,404	147.7	137	185
6	C06	1,049	31.4	32	24
	PC06_1	1,084	42.1	35	9
7	C07	2,047	66.2	66	34
	PC07_1	798	45.1	25	23
8	C08	3,949	105.7	114	318
9	C09	2,812	118.0	62	204
10	C10	3,238	115.3	102	103
11	C11	4,532	90.6	122	298
12	C12	2,205	74.8	62	13
13	C13	3,851*	105.3	86	149
14	C14	960	44.3	32	13
	DC14_1	30	3.3	0	0
15	DC15	920	72.2	17	95
	PC15	716	34.3	22	6
16	C16	1,914	65.8	60	75
Sub-total	Integrated LGs	48,979	1,521.0	1,417	2,021
	Maternal LGs (*dura* DCxx)	1,484	297.0	28	119
	Paternal LGs (*pisifera* PCxx)	7,002	269.3	219	223
Total		**57,465**	**2,087.5**	**1,664**	**2,363**

*including 76 common SNPs

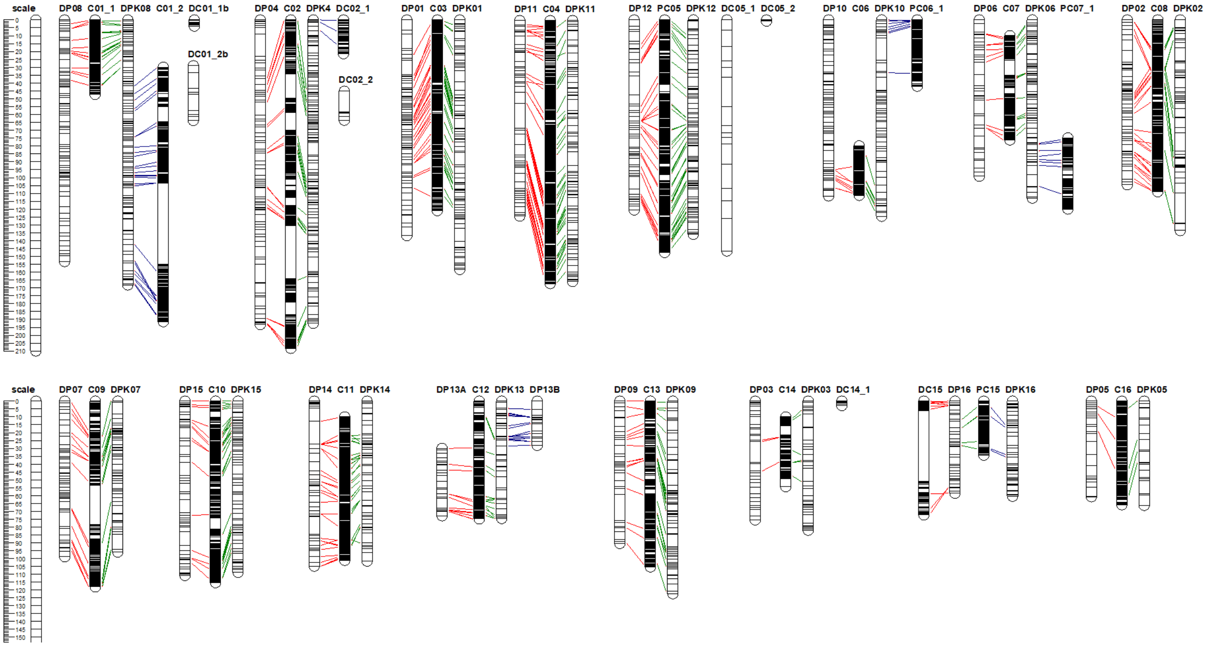

Fig. 4. Aligning the PUP to other DxP genetic linkage maps e.g. P2 (labelled as DP01 – 16) and KULIM DxP (DPK01 – 16) published by Zolkafli et al. (2021) using common SNPMxxxxx markers.

3.4. Homozygous segments (HS) in parental palms and progeny

In this study, the BAF distribution in EG5.1 revealed extensive levels of homozygosity in the two parental palms, which was reflected in the PUP genetic linkage maps. HS contain only monomorphic SNP markers with BAF of 0 or 1, and as such cannot be placed on the genetic map. The HS varied in size (Fig. 5) and segments of at least 400,000 bp long were extracted and shown in Fig. 6. On the genetic map, only gaps ≥ 3.0 cM (assuming 1.0 cM = 1.0 Mb) were considered HS (Fig. 6).

In the *dura* maternal palm, 81 HS of 581,163 – 37,784,289 bp in length were observed on the 16 chromosomes, and 15 of them (HS) could be visualized on 11 LGs. Almost the entire CHR05 (99.6 %), 06 (92.9 %) and 14 (97.0 %) were homozygous, likely because of either insufficient linkage between the small sub-LGs or the extremely large gaps between markers on the LGs. In comparison, the *pisifera* palm had more HS (90) with sizes ranging from 401,250 – 15,213,371 bp, but the total genomic region covered by its HS was only 287,872,604 bp, smaller than in the *dura* (357,976,728 bp). Only CHR09 (91.7 %) in the *pisifera* was very highly homozygous as visualized on both the EG5.1 and genetic map (Fig. 4 and 5). Heterozygous markers were mostly found in two regions, 160,175 – 2,036,315 and

30,246,439 – 36,485,591 bp of EG5.1, and they were mapped at positions 1.7 – 4.4 and 89.7 – 97.0 cM, respectively, in LGC09.

It is important to note that most of the specific HS in one parent are heterozygous in the other parent (Fig. 6). This is consistent with the high number of *aa* x *ab* and *ab* x *aa* SNPs and partly explain why so few (8.9 %) were *ab* x *ab*. Crossing the homo- and heterozygous loci (*aa* x *ab* or *ab* x *aa*) resulted in HS (and thereby LOH) in a number of the PUP palms, as reported previously by Ting et al. (2021b). In the cases where both parental palms had a HS in common, this produced gap in the integrated LGs. In this study, of the 58 common HS (≥ 400,000 bp) covering a total genomic region of 138 Mb, 70.3 % (97 Mb) contained SNP markers with identical homozygous genotypes in both parental palms (identical by descent, IBD) i.e., *aa* x *aa* or *bb* x *bb*, while the remainder had *aa* x *bb*, producing *ab* genotype. Tracing back to the parental palms, the common HS with IBD SNPs in the progeny were 27.1 and 33.7 % of the total HS observed in the *dura* and *pisifera*, respectively. As such, although the *dura* was more homozygous than the *pisifera*, the progeny inherited more HS from the latter.

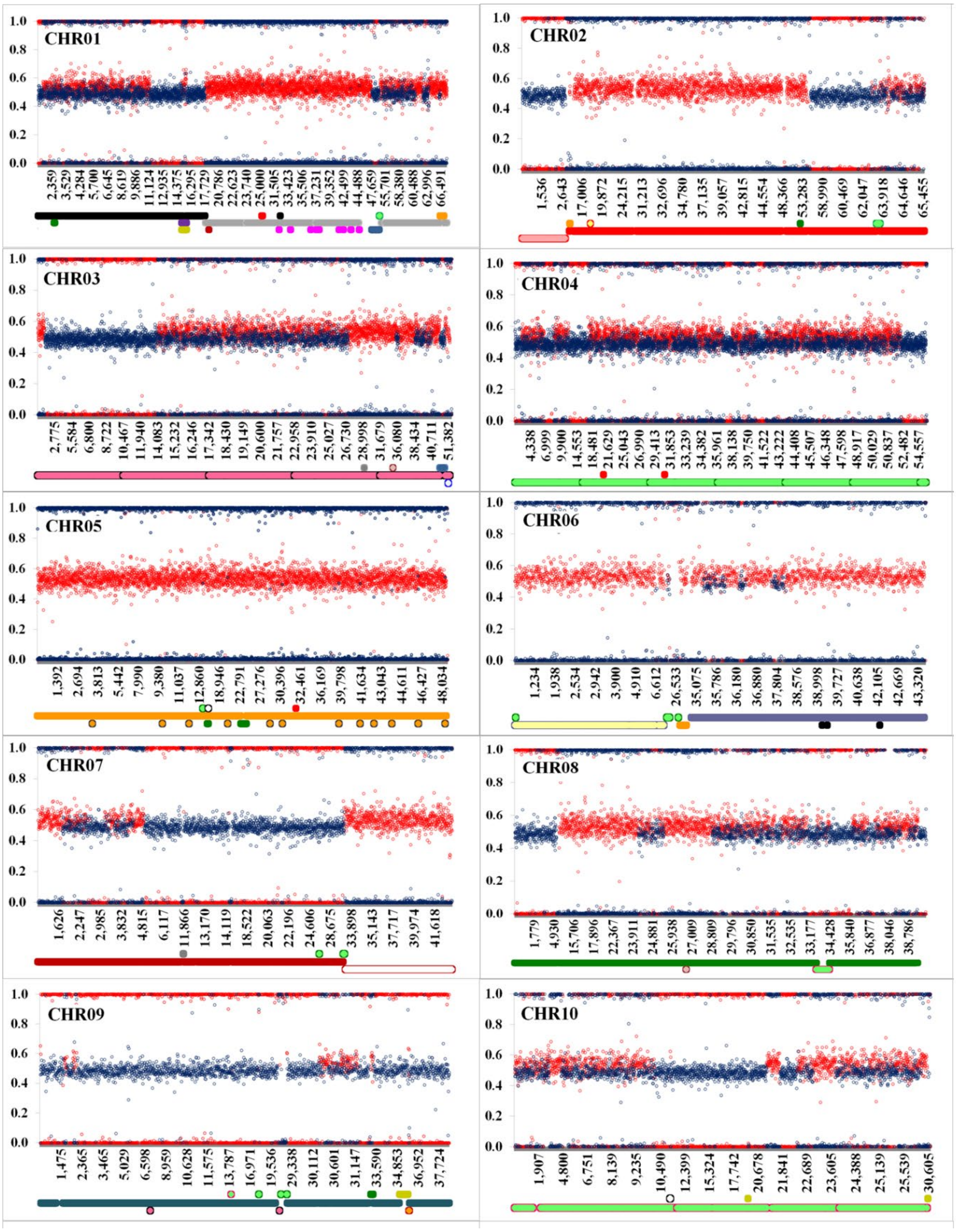

Fig. 5. SNP BAF (range 0.0 – 1.0, y-axis) distribution across the 16 chromosomes (CHR01 – 16) in the *dura* maternal (dark blue dots) and the *pisifera* paternal (red dots) palms. X-axis is the pseudo-chromosome position of SNP markers (Kb in EG5.1). Markers mapped in their corresponding or different linkage groups (LGs) are presented in bars with colour codes, placed under the BAF plot of each chromosome.

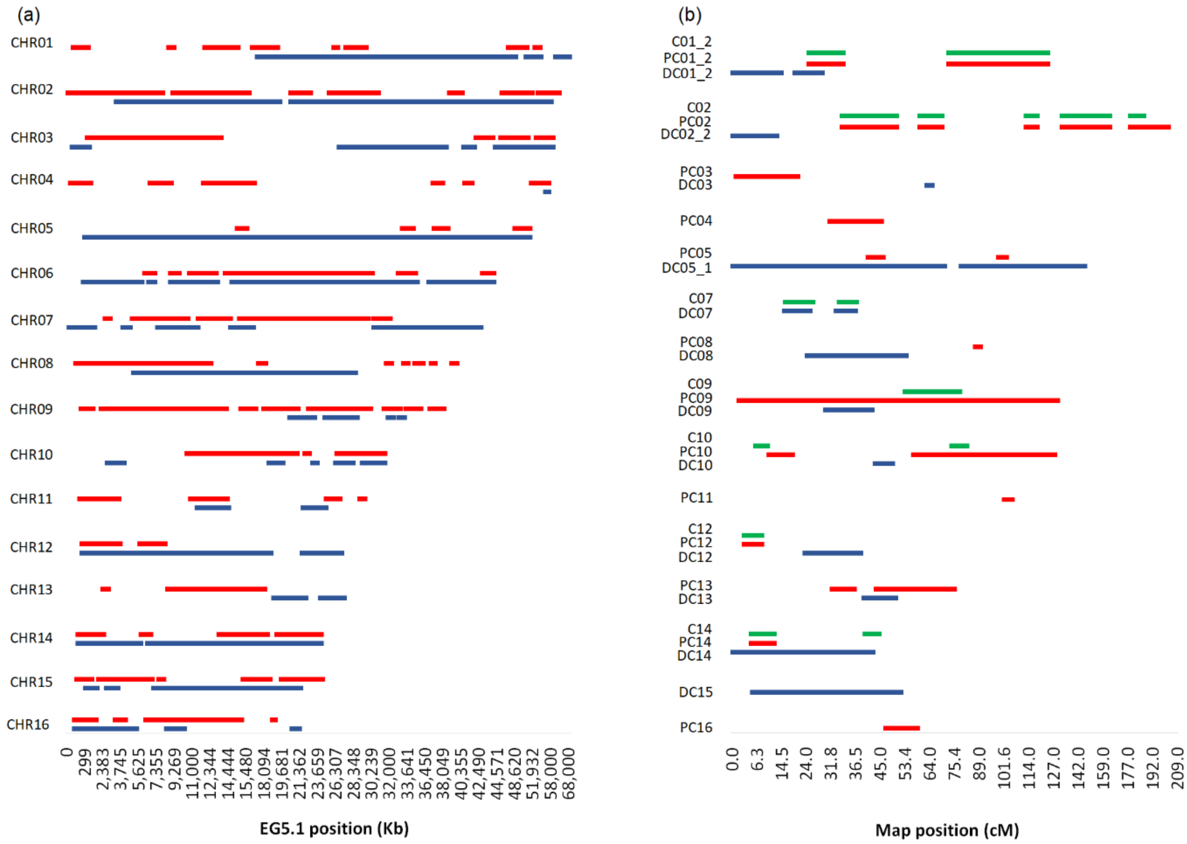

Fig. 6. Distribution of homozygous segments (HS) on EG5.1 pseudo-chromosomes (a) and PUP genetic map (b). Each coloured bar represents a HS of ≥ 400,000 bp in a chromosome and ≥ 3.0 cM in a linkage group (LG) where *dura* parental LG is labelled as DCxx (blue), PCxx for *pisifera* (red) and Cxx for the integrated LG (green).

4. Discussion

The oil palm 78K Infinium® HD SNP array presented in this study was highly effective in generating good quality genotype data and facilitated the generation of a HD genetic linkage map that could be used to further validate and improve the existing EG5.1 genome build (Singh et al., 2013a; Sanusi et al., 2023). As such, the array was designed to contain mostly those 76,918 SNPs (98.4 %) that could be mapped with high confidence to the pseudo-chromosomes and orphan scaffolds in EG5.1. As the SNP array consists of SNPs from ABLs and germplasm cohorts, the application of the SNP array has been extended to genotype more diverse populations developed from the germplasm, such as the population generated from the selfing of the palm T128 (unpublished data) from MPOB's Nigerian collection, which has a higher unsaturated oil content than current commercial material (Singh et al., 2013b), as well as a diverse set of palms originating from different African countries (unpublished data). In comparison, for the previously reported 200K SNP array for oil palm, of the 170,860 SNPs, only 70,804 (41.4 %) could be mapped to the existing reference genome (Kwong et al., 2016). While similar in the number of SNPs mapped to the genome build, the two arrays share only 3,955 SNPs due to different design and selection strategies. The 200K array was designed to focus on major breeding origins available at Sime Darby plantation (Kwong et al., 2016) while the present 78K SNP array focuses more broadly on variation across a wide range of ABLs as well as the oil palm germplasm collection at MPOB.

The Deli *dura* x AVROS *pisifera* hybrid (*tenera*) is among the highest-yielding oil palm varieties (Noh et al., 2012; Ong et al., 2019 and Swaray et al., 2020). Both the Deli *dura* and AVROS *pisifera* breeding lines have gone through several generations of selfing or sib-mating over the past decades. The level of inbreeding is more severe in Deli *dura* as it was derived from an extremely narrow genetic base consisting of only four seedlings (Hartley 1967), whereas the AVROS *pisifera* originated from a small population of palms. In general, the current Deli *dura* x AVROS *pisifera* materials have already achieved high uniformity in the expression of their yield traits (Noh et al., 2010 and 2012). In other words, this indicates low genetic variability which is obvious from the low polymorphism rate as reported previously – 2.2 % (11,394 polymorphic SNPs from the 510,251 SNPs identified) by Bai et al. (2018) and 20.1 % (34,315 polymorphic SNPs identified from the total 170,860 SNPs genotyped) by Ong et al. (2019) for the Deli *dura* x AVROS *pisifera* mapping families. In this study, a new batch of Deli *dura* x AVROS *pisifera* cross consisting of 1,011 F_1 palms was specifically created, which to our knowledge, is the largest oil palm family used for genetic linkage analysis. Although the high number of seedlings in this cross are useful for improving the mapping resolution, there is no denying that collating information on meiosis from multiple families is also important, a process which can be readily facilitated by the SNP array and the information already available in this study. In the past, only field-planted palms were studied, limiting the families to only 80 – 300 palms (Ukoskit et al., 2014; Bai et al., 2018; Ong et al., 2020; Herrero et al., 2020; Zolkafli et al., 2021; Kamaruddin et al., 2021).

In this study, a large number of SNPs (57,465) from the new 78K SNP array were mapped onto the Deli *dura*, AVROS *pisifera* parental and their integrated linkage maps. The marker density is 2.1 – 5.7X higher than in other genetic maps constructed from a similar genetic background (2,938.2 cM/10,023 markers by Bai et al., 2018 and 1,151.7 cM/27,890 markers by Ong et al., 2019), and 1.7 – 1.8X higher than in the *dura* x *dura* genetic maps (1,268.3 cM/ 6,920 markers and 1,646.9 cM/ 32,650 markers by Ong et al., 2020 and 1,618.5 cM/33,190 markers by Teh et al., 2020). On the PUP genetic maps, at least 29.0 % of the markers mapped in clusters to the same genetic position on the LGs, suggesting that the map resolution was limited by the number of observable recombination events rather than by the number of markers. However, a few LGs were nevertheless fragmented due to long stretches of HS in the two parental palms. They were retained as sub-LGs, as forcefully combining them would likely inflate the genetic map as well as inconsistently invert/orientate the sub-LGs. Only the sub-LGs without SNPs segregating in both parents (*ab* x *ab*) were retained in the respective parental maps. The total length of the sub-LGs was similar in both the *dura* (297.0 cM) and *pisifera* (269.3 cM), and if these could be integrated, ~280 cM more segment would have been added to the combined map (1,521 cM), which would likely increase the total coverage of the genetic map to ~1,800 cM.

The integration of the two parental maps could be improved by the generation of additional co-segregating markers e.g. *ab* x *ac* and *ab* x *cd* genotype calls in the parental palms. One strategy could be to aggregate and convert the existing genotype calls (in predetermined haplotype blocks) using Haploblock Aggregator as described by Kruisselbrink et al. *(*2016). An additional approach could be to use the feature in the Infinium® array that allows detection and calling of SNPs with null alleles that were excluded in this study. Montanari et al. (2013) had suggested that manual editing of the SNPs with low or no calls or showing strong deviation from Mendellian inheritance, could allow the SNPs to be called and provide the required information.

A HD genetic map with good genome coverage can assist in genome assembly and improve the existing genome build, such as EG5.1. The pseudo-chromosomes of EG5.1 were first assembled using genetic maps with < 2,000 markers (Singh et al., 2013a) and later, 3,422 markers (Sanusi et al., 2023). In comparison, the current genetic map has 28X higher density, including 14,781 SNPs specifically developed from 2,363 orphan scaffolds, which will allow them (scaffolds) to be placed on the pseudo-chromosomes. The number of orphan scaffolds identified from the genetic map constructed in the present study is notably higher than the 1,323 and 1,968 reported using a single (Ong et al., 2019) and combination of three genetic maps (Ong et al., 2020). However, integration of all these results into an updated assembly would be highly beneficial for the oil palm industry as a whole.

By integrating the genetic map with genome data, the large set of SNP markers facilitated a detailed scan of the individual chromosomes in both the Deli *dura*, AVROS *pisifera* and their progeny. In addition to various chromosomal aberrations reported previously (Ting et al., 2021b), we localized homozygous regions on each chromosome of individual palms with great precision. In particular, the high homozygosity in the two parents is consistent with the low phenotypic variability in their progeny. In general, the *dura* had longer stretches of HS than the *pisifera* e.g. in CHR01, 02, 05, 06, 08, 12, 14 and 15. This could be because the Deli *dura* came from a narrow genetic background and the MRS breeding to improve it involved more and more inbreeding, resulting in such a homozygous genome. Although homozygosity is desirable in the shorter term (e.g. seedling production) giving more uniform DxP progeny, breeders have to be cautious in subsequent improvement programmes as too much homozygosity may result in severe inbreeding depression and expression of deleterious recessive loci (Neaves et al., 2015; Wang, 2019) possibly resulting in increased rate of culling required in the nursery due to higher rate of off-type seedlings.

More importantly, the location and distribution of the HS (Fig. 3 and 5), especially those containing IBD alleles, may indicate genomic regions of interest that have been under selection pressure for alleles associated with desired traits. In such a scenario, meaningful breeding and selection programmes can focus on the HS of interest, either to retain homozygosity or introduce variability (heterozygosity). As these regions are demarcated by known markers, genetic improvement through MAS should be possible, especially if the HS confer a desirable or even an undesirable phenotype. This information will also facilitate editing the parental genomes in the on-going effort to develop designer palms (Zulkifli et al., 2017; Yue et al., 2021).

Interestingly, self-fertilization of the Deli *dura* was observed in 43 PUP progeny although the controlled cross was generated using an established protocol, including blank pollination to detect stray pollination. Nevertheless, the self-fertilization presumably occurred by pollen from rudimentary stamens (staminodes) at the bases of the female flowers. The formation of these stamens can be facilitated by changes in the weather pattern and other factors that result in increased stress to a palm.

They are generally thought not to contribute to pollination as their pollen was found to be infertile, and typically shed before the female flowers are receptive for pollination (Adam et al., 2005). However, there are contrasting reports that some viable pollen is occasionally produced to cause stray pollination (Beirnaert, 1935; Rao and Kushairi, 1999; Corley, 2005). Although blank pollination is a standard procedure employed by breeders, the frequency at which this is done over the productive lifecycle of the maternal palms may differ from one plantation to another. Infrequent validation by blank pollination can increase the probability of self-fertilization.

The selfing rate of 4.3 % as observed in this study is slightly lower than the 5.3 − 30.0 % in *Medicago sativa* L., a tetraploid outcrosser (Dieterich Mabin et al., 2021). However, even with a lower degree of selfing, diploids (such as oil palm) may suffer more severe inbreeding depression than tetraploids because of their lesser genetic buffering from having only 2x *vs*. 4x chromosome sets (Husband and Schemske, 1997). Thus, if these selfed palms are not identified and eliminated early, they can cause loss in fruit yield in the plantation due to inbreeding depression. This dilutes the ultimate purpose of commercial seed production to produce palms with higher yield due to the hybrid vigour resulting from the inter-crossing of two separate and somewhat inbred parental lines (Singh et al., 2020). It would be cost-effective to apply the *SHELL* gene assay (Ooi et al., 2016; Singh et al., 2020 and 2021) that can differentiate the three fruit forms of oil palm, as a quality control tool, to identify and remove *dura* seedlings in the nursery. This will ensure the fidelity of controlled crosses generated using selected *dura* and *pisifera* lines for planting in commercial fields.

5. Conclusions

In conclusion, the SNP array developed in this study consists of highly informative markers across the genome, allowing the construction of a high-resolution genetic map for a Deli *dura* x AVROS *pisifera* cross. The genetic map can improve the existing oil palm genome build by incorporating almost 400 more scaffolds than in other similar initiatives to date. However, a number of orphan scaffolds could not be linked to the genetic map due to the presence of HS in both parents. To incorporate these scaffolds, genetic maps generated from different families will be required. The BAF distribution based on a set of markers ordered in both the physical and genetic maps, identified for the first time, specific chromosomes nearly homozygous in the Deli *dura*. The blocks of homozygosity were obvious across the genomes of both the parental palms and were confirmed by their pattern of inheritance observed in the progeny. This homozygosity admittedly caused some gaps in the LGs, and a small number of sub-LGs could not be merged into the integrated map. Moving forward, the identification of HS and, with it, the IBD alleles, can facilitate MAS, especially if the HS harbours genes of interest. The presence of HS in the *dura* female parent begs the important question of whether sib-crossing or crossing with an unrelated *dura* would be the best way forward to improve the next generation maternal breeding lines. In addition, the SNP marker set proved useful in detecting 'selfs' and aneuploids, which, if undetected, would result in a severe loss of future yield due to inbreeding depression. Although there are standard procedures to detect rudimentary anther pollination, they are not completely robust. Genomic tools like the current SNP array could contribute to the fidelity of controlled crosses and to fully appreciate the benefits of hybrid vigour in the commercial cultivation of oil palm.

Funding

This work was funded by the Eleventh Malaysia Plan (RMKe-11: 201101807) and MPOB [ABBC23-2019].

CRediT authorship contribution statement

Ting Ngoot-Chin: Conceptualization, Methodology, Formal analysis, Investigation, Writing - Original draft preparation. **Jared M Ordway:** Conceptualization, Investigation, Writing - Original draft preparation, review & editing. **Eric Van de Weg:** Formal analysis, Supervision, Writing - Review & editing. **Norhalida Mohamed Serdari:** Investigation. **Leslie Low Eng Ti:** Investigation. **Suzana Mustaffa:** Investigation. **Corey Wischmeyer:** Investigation. **Marinus J. M. Smulders:** Conceptualization, Supervision, Writing - Review & editing. **Ravigadevi Sambanthamurthi:** Conceptualization, Funding acquisition, Writing - Review & editing. **Rajinder Singh:** Conceptualization, Methodology, Funding acquisition**,** Project administration, Supervision, Writing - Review & editing.

Declaration of Competing Interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

All relevant data can be found within the manuscript and its Supplementary data. The oil palm EG5.1 genome build (labelled as AM_EG5.1) and the SNP information are available at the Genomsawit website (access link for AM_EG5.1: http://genomsawit.mpob.gov.my/index.php?track=30&nu=1&info=1 and access link for SNP markers: http://genomsawit.mpob.gov.my/index.php?track=30&nu=5&info=13)

Acknowledgements

The authors thank the Director-General of the Malaysian Palm Oil Board (MPOB) for permission to publish this paper. We would also like to thank Andy Chang Kwong Choong for critical reading of the manuscript, Maizura Azwanie Mohd Zarawi for technical assistance in DNA extraction and Hamizan Omar and the team of Breeding and Genetics Group, MPOB for planting the palms.

Appendix A. Supplementary data

Fig. A.1 BAF distribution for the 43 *dura*-clustered palms.

Table A.1 The representative panel of genome-wide distributed SNPs for construction of NJ tree and factorial analysis.

Table A.2 Examination of genotypes for 39,908 SNPs with GT ≥ 0.7 of the 43 maternal-clustered palms in PUP. The three observed segregation types *aa* x *bb* (170 SNPs), *aa* x *ab* (24,881 SNPs) and *ab* x *aa* (14,857 SNPs).

References

Adam H, Jouannic S, Escoute J, Duval Y, Verdeil JL, Tregear JW (2005). Reproductive developmental complexity in the African oil palm (*Elaeis guineensis*, Arecaceae). Am. J. Bot. 92, 1836-52. https://doi.org/10.3732/ajb.92.11.1836.

Ahmad Malike F, Amiruddin MD, Yaakub Z, Marjuni M, Abdullah N, Abu Bakar NA, Mustaffa S, Mohamad MM, Hassan MY, Ong-Abdullah M, Ahmad Parveez GK, Kushairy A (2019). Oil palm (*Elaeis* spp.) breeding in Malaysia. In: Al-Khayri J, Jain S, Johnson D (Eds) Advances in Plant Breeding Strategies: Industrial and Food Crops. Springer Cham, pp. 489-535. https://doi.org/10.1007/978-3-030-23265-8_13.

Ahmad Parveez GK, Hishamuddin E, Loh SK, Ong-Abdullah M, Mohamed Salleh KM, Zanal Bidin MNI, Sundram S, Hasan ZAA, Idris Z (2020). Oil palm economic performance in Malaysia and R&D progress in 2019. J. Oil Palm Res. 32, 159-190. https://doi.org/10.21894/jopr.2020.0032.

Ahmad Parveez GK, Ahmad Tarmizi AH, Sundram S, Loh SK, Ong-Abdullah M, Palam KDP, Mohamed Salleh KM, Mohd Ishak S, Idris Z (2021). Oil palm economic performance in Malaysia and R&D progress in 2020. J. Oil Palm Res. 33, 181-214. https://doi.org/10.21894/jopr.2021.0026.

Ahmad Parveez GK, Kamil NN, Zawawi NZ, Ong-Abdullah M, Rasuddin R, Loh SK, Selvaduray KR, Hoong SS, Idris Z (2022). Oil palm economic performance in Malaysia and R&D progress in 2021. J. Oil Palm Res. 34, 185-218. https://doi.org/10.21894/jopr.2022.0036.

Babu BK, Mathur RK, Anitha P, Ravichandran G, Bhagya HP (2021). Phenomics, genomics of oil palm (*Elaeis guineensis* Jacq.): way forward for making sustainable and high yielding quality oil palm. Physiol. Mol. Biol. Plants 27, 587-604. https://doi.org/10.1007/s12298-021-00964-w.

Bai B, Wang L, Zhang YJ, Lee M, Rahmadsyah R, Alfiko Y, Ye BQ, Purwantomo S, Suwanto A, Chua NH, Yue GH (2018). Developing genome-wide SNPs and constructing an ultrahigh-density linkage map in oil palm. Sci. Rep. 8, 691. https://doi.org/10.1038/s41598-017-18613-2.

Beirnaert A (1935). Introduction à la biologie florale du palmier à huile (*Elaeis guineensis* Jacquin). Publ. Inst. Nat. Etude agron. Congo Belge, Ser. Sci. 5, 3-42.

Bianco L, Cestaro A, Linsmith G, Muranty H, Denancé C, Théron A, Poncet C, Micheletti D, Kerschbamer E, Di Pierro EA, Larger S, Pindo M, Van de Weg E, Davassi A, Laurens F, Velasco

R, Durel CE, Troggio M (2016). Development and validation of the Axiom® Apple480K SNP genotyping array. Plant J. 86, 62-74. https://doi.org/10.1111/tpj.13145.

Chagné D, Crowhurst RN, Troggio M, Davey MW, Gilmore B, Lawley C, Vanderzande S, Hellens RP, Kumar S, Cestaro A, Velasco R, Main D, Rees J, Iezzoni A, Mockler T, Wilhelm L, Van de Weg E, Gardiner SE, Bassil N, Peace C (2012). Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS One 7, e31745. https://doi.org/10.1371/journal.pone.0031745.

Chagné D, Kirk C, Whitworth C, Erasmuson S, Bicknell R, Sargent DJ, Kumar S, Troggio M (2015). Polyploid and aneuploid detection in apple using a single nucleotide polymorphism array. Tree Genet. Genomes 11, 94. https://doi.org/10.1007/s11295-015-0920-8.

Corley RHV (1998). What is the upper limit to oil extraction ratio? In: Rajanaidu N, Henson IE, Jalani BS (Eds.) Proceedings of the 1996 International oil palm conference “Oil and kernel production in oil palm-a global perspective: 27-28 September 1996; Kuala Lumpur.

Corley RHV (2005). Illegitimacy in oil palm breeding - a review. J. Oil Palm Res. 17, 64-69.

Cui F, Zhang N, Fan X-l, Zhang W, Zhao C-h, Yang L-j, Pan R-q, Chen M, Han J, Zhao X-q, Ji J, Tong Y-p, Zhang H-x, Jia J-z, Zhao G-y, Li J-m (2017). Utilization of a Wheat660K SNP array-derived high-density genetic map for high-resolution mapping of a major QTL for kernel number. Sci. Rep. 7, 3788. https://doi.org/10.1038/s41598-017-04028-6.

Dieterich Mabin ME, Brunet J, Riday H, Lehmann L (2021). Self-fertilization, inbreeding, and yield in Alfalfa seed production. Front. Plant Sci. 12, 700708. https://doi.org/10.3389/fpls.2021.700708.

Gan ST, Wong WC, Wong CK, Soh AC, Kilian A, Low EL, Massawe F, Mayes S (2018). High density SNP and DArT-based genetic linkage maps of two closely related oil palm populations. J. Appl. Genet. 59, 23-34. https://doi.org/10.1007/s13353-017-0420-7.

Hamilton JP, Hansey CN, Whitty BR, Stoffel K, Massa AN, Van Deynze A, De Jong WS, Douches DS, Buell CR (2011). Single nucleotide polymorphism discovery in elite North American potato germplasm. BMC Genom. 12, 302. https://doi.org/10.1186/1471-2164-12-302.

Hartley CWS (1967). The oil palm. Longmans, Green And Co Ltd., London.

Herrero J, Santika B, Herrán A, Erika P, Sarimana U, Wendra F, Sembiring Z, Asmono D, Ritter E (2020). Construction of a high-density linkage map in oil palm using SPET markers. Sci. Rep. 10, 9998. https://doi.org/10.1038/s41598-020-67118-y.

Howard NP, Peace C, Silverstein KAT, Poets A, Luby JL, Vanderzande S, Durel CE, Muranty H, Denancé C, Van de Weg E (2021). The use of shared haplotype length information for pedigree reconstruction in asexually propagated outbreeding crops, demonstrated for apple and sweet cherry. Hortic. Res. 8, 202. https://doi.org/10.1038/s41438-021-00637-5.

Husband BC, Schemske DW (1997). The effect of inbreeding in diploid and tetraploid populations of *Epilobium angustifolium* (onagraceae): implications for the genetic basis of inbreeding depression. Evolution. 51, 737-746. https://doi.org/10.1111/j.1558-5646.1997.tb03657.x.

Ithnin M, Vu WT, Shin M-G, Suryawanshi V, Sherbina K, Zolkafli SH, Serdari NM, Amiruddin MD, Abdullah N, Mustaffa S, Marjuni M, Nookiah R, Kushairi A, Marjoram P, Nuzhdin SV, Chang PL, Singh R (2021). Genomic diversity and genome-wide association analysis related to yield and fatty acid composition of wild American oil palm. Plant Sci. 304, 110731. https://doi.org/10.1016/j.plantsci.2020.110731.

Jin J, Lee M, Bai B, Sun Y, Qu J, Rahmadsyah R, Alfiko Y, Lim CH, Suwanto A, Sugiharti M, Wong L, Ye J, Chua NH, Yue GH (2016). Draft genome sequence of an elite *dura* palm and whole-genome patterns of DNA variation in oil palm. DNA Res. 23, 527-533. https://doi.org/10.1093/dnares/dsw036.

Kalinowski ST, Taper ML, Marshall TC (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16, 1099-1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x.

Kalyana BB, Mathur RK, Venu MVB, Sandip S, Ravichandran G, Anita P, Bhagya HP (2021). Genome-wide association study (GWAS) of major QTLs for bunch and oil yield related traits in *Elaeis guineensis* L. Plant Sci. 305, 110810. https://doi.org/10.1016/j.plantsci.2020.110810.

Kamaruddin K, Ithnin M, Ting N-C, Zulkifli Y, Sanusi NSNM, Singh R (2021). Anchoring a genetic map of an interspecific backcross two family to the genome builds of *Elaeis*. J. Genet. 100, 7. https://doi.org/10.1007/s12041-020-01240-8.

Kruisselbrink JW, Di Pierro EA, Gianfranceschi L, Bink MCAM, Van de Weg WE (2016). Haploblock Aggregator: combining segregation patterns of con-secutive SNP markers into haploblock segregation patterns for generating integrated genetic linkage maps. Wageningen UR, the Netherlands. Available at http://www.wageningenur.nl/en/show/HaploblockAggregator.htm.

Kwong QB, Teh CK, Ong AL, Heng HY, Lee HL, Mohamed M, Low JZ, Apparow S, Chew FT, Mayes S, Kulaveerasingam H, Tammi M, Appleton DR (2016). Development and validation of a high-density SNP genotyping array for African oil palm. Mol. Plant 9, 1132-1141. https://doi.org/10.1016/j.molp.2016.04.010.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

Li H (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987-2993. https://doi.org/10.1093/bioinformatics/btr509.

Low E-TL, Azizi N, Ab Halim MA, Sanusi NSNM, Chan K-L, Amiruddin N, Nagappan J, Ong-Abdullah M, Singh R, Sambanthamurthi R, Manaf MAA, Kushairi A (2020). Oil palm genome: strategies and applications in: Ithnin M, Kushairi A (Eds.) The oil palm genome. Compendium of Plant Genomes. Springer Cham, pp. 83-115.

Lu N, Zhang M, Xiao Y, Han D, Liu Y, Zhang Y, Yi F, Zhu T, Ma W, Fan E, Qu G, Wang J (2019). Construction of a high-density genetic map and QTL mapping of leaf traits and plant growth in an interspecific F_1 population of *Catalpa bungei* × *Catalpa duclouxii* Dode. BMC Plant Biol. 19, 596. https://doi.org/10.1186/s12870-019-2207-y.

Luo X, Xu L, Wang Y, Dong J, Chen Y, Tang M, Fan L, Zhu Y, Liu L (2020). An ultra-high-density genetic map provides insights into genome synteny, recombination landscape and taproot skin colour in radish (*Raphanus sativus* L.). Plant Biotechnol. J. 18, 274-286. https://doi.org/10.1111/pbi.13195.

Lv D, Zhang C, Yv R, Yao J, Wu J, Song X, Jian J, Song P, Zhang Z, Han D, Sun D (2021). Utilization of a Wheat50K SNP microarray-derived high-density genetic map for QTL mapping of plant height and grain traits in wheat. Plants 10, 1167. https://doi.org/10.3390/plants10061167.

McCouch SR, Wright MH, Tung C-W, Maron LG, Mcnally KL, Fitzgerald M, Singh N, DeClerck G, Agosto-Perez F, Korniliev P, Greenberg AJ, Naredo MEB, Mercado SMQ, Harrington SE, Shi Y, Branchini DA, Kuser-Falcão PR, Leung H, Ebana K, Yano M, Eizenga G, McClung A, Mezey J (2016). Open access resources for genome-wide association mapping in rice. Nat. Commun. 7, 10532. https://doi.org/10.1038/ncomms10532.

Montanari S, Saeed M, Knäbel M, Kim Y, Troggio M, Malnoy M, Velasco R, Fontana P, Won K, Durel C-E, Perchepied L, Schaffer R, Wiedow C, Bus V, Brewer L, Gardiner SE, Crowhurst RN, Chagné D (2013). Identification of *Pyrus* single nucleotide polymorphisms (SNPs) and evaluation for genetic mapping in European pear and interspecific *Pyrus* hybrids. PLoS One 8, e77022. https://doi.org/10.1371/journal.pone.0077022.

Morales KY, Singh N, Perez FA, Ignacio JC, Thapa R, Arbelaez JD, Tabien RE, Famoso A, Wang DR, Septiningsih EM, Shi Y, Kretzschmar T, McCouch SR, Thomson MJ (2020). An improved 7K SNP array, the C7AIR, provides a wealth of validated SNP markers for rice breeding and genetics studies. PLoS One 15, e0232479. https://doi.org/10.1371/journal.pone.0232479.

Muranty H, Denancé C, Feugey L, Crépin JL, Barbier Y, Tartarini S, Ordidge M, Troggio M, Lateur M, Nybom H, Paprstein F, Laurens F, Durel CE (2020). Using whole-genome SNP data to reconstruct a large multi-generation pedigree in apple germplasm. BMC Plant Biol. 20, 2. https://doi.org/10.1186/s12870-019-2171-6.

Neaves LE, Eales J, Whitlock R, Hollingsworth PM, Burke T, Pullin AS (2015). The fitness consequences of inbreeding in natural populations and their implications for species conservation – a systematic map. Environ. Evid. 4, 5. https://doi.org/10.1186/s13750-015-0031-x

Noh A, Rafii M, Saleh G, Kushairi A (2010). Genetic performance of 40 Deli *dura* × AVROS *pisifera* full-sib families. J. Oil Palm Res. 22, 781-795.

Noh A, Rafii MY, Saleh G, Kushairi A, Latif MA (2012). Genetic performance and general combining ability of oil palm Deli *dura* x AVROS *pisifera* tested on inland soils. Sci. World J. 2012, 792601. https://doi.org/10.1100/2012/792601.

Nyouma A, Bell JM, Jacob F, Riou V, Manez A, Pomiès V, Nodichao L, Syahputra I, Affandi D, Cochard B, Durand-Gasselin T, Cros D (2020). Genomic predictions improve clonal selection in oil palm (*Elaeis guineensis* Jacq.) hybrids. Plant Sci. 299, 110547. https://doi.org/10.1016/j.plantsci.2020.110547.

Oil World (2021). Oil world annual 2021, vol. 2, 11 June 2021. ISTA Mielke GmbH, Germany.

Ong A-L, Teh C-K, Kwong QB, Tangaya P, Appleton DR, Massawe F, Mayes S (2019). Linkage-based genome assembly improvement of oil palm (*Elaeis guineensis*). Sci. Rep. 9, 6619. https://doi.org/10.1038/s41598-019-42989-y.

Ong A-L, Teh C-K, Mayes S, Massawe F, Appleton DR, Kulaveerasingam H (2020). An improved oil palm genome assembly as a valuable resource for crop improvement and comparative genomics in the *Arecoideae* subfamily. Plants 9, 1476. https://doi.org/10.3390/plants9111476.

Ooi LCL, Low ETL, Abdullah MO, Rajanaidu N, Ting N-C, Nagappan J, Manaf MAA, Chan K-L, Halim MA, Azizi N, Omar W, Murad AJ, Lakey N, Ordway JM, Favello A, Budiman MA, Van Brunt A, Beil M, Leininger MT, Jiang N, Smith SW, Brown CR, Kuek ACS, Bahrain S, Hoynes-O'Connor A, Nguyen AY, Chaudhari HG, Shah SA, 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. Front. Plant Sci. 7, 771. https://doi.org/10.3389/fpls.2016.00771.

Ooi LC-L, Low E-TL, Ordway J, Marjuni M, Yaakub Z, Jiang N, Smith S, Bacher B, Garner PA, Leinenger MT, Sander N, Chan P-L, Ong PW, Abdullah MO, Nookiah R, Manaf MAA, Lakey N, Sambanthamurthi R, Singh R (2019). SureSawitTM true-to-type – a high throughput universal single nucleotide polymorphism panel for dna fingerprinting, purity testing and origin verification in oil palm. J. Oil Palm Res. 31, 561-571. https://doi.org/10.21894/jopr.2019.0048.

Perrier X, Flori A and Bonnot F (2003). Data analysis methods. In: Hamon P, Seguin M, Perrier X, Glaszmann JC. (Eds.) Genetic diversity of cultivated tropical plants. Enfield, Science Publishers. Montpellier, pp.43-76.

Pootakham W, Jomchai N, Ruang-areerate P, Shearman JR, Sonthirod C, Sangsrakru D, Tragoonrung S, Tangphatsornruang S (2015). Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). Genomics 105, 288-295. https://doi.org/10.1016/j.ygeno.2015.02.002.

Rahimah AR, Cheah SC, Singh R (2006). Freeze-drying of oil palm (*Elaeis guineensis*) leaf and its effect on the quality of extractable DNA. J. Oil Palm Res. 18, 296-304.

Rajanaidu N, Kushari A, Rafii M, Mohd Din A, Maizura I and Jalani BS (2000). Oil palm breeding and genetic resources. In: Basiron Y, Jalani BS, Chan KW (Eds.) Advances in oil palm research. Malaysian Palm Oil Board, Bandar Baru Bangi, Malaysia, pp. 171-237.

Rao V, Kushairi A (1999). Quality of oil palm planting material. In: Rajanaidu N, Jalani BS (Eds.) Proceedings of the seminar on sourcing of oil palm planting materials for local and overseas joint-ventures, 22-23 July 1996, Subang Jaya, Malaysia, pp. 188-197.

Rosenquist EA (1990). An overview of breeding technology and selection in *Elaeis guineensis*. In Basiron Y et al. (Eds.) Proceedings of PORIM international palm oil development conference, module agriculture: 5-9 September 1989; Kuala Lumpur.

Sanusi NSNM, Rosli R, Chan K-L, Ab Halim MA, Ting N-C, Singh R, Low E-TL (2023). Integrated consensus genetic map and genomic scaffold re-ordering of oil palm (*Elaeis guineensis*) genome. Comput. Biol. Chem. 102, 107801. https://doi.org/10.1016/j.compbiolchem.2022.107801.

Sarkar MSK, Begum RA, Pereira JJ (2020). Impacts of climate change on oil palm production in Malaysia. Environ. Sci. Pollut. Res. 27, 9760–9770. https://doi.org/10.1007/s11356-020-07601-1.

Singh R, Madon M, Low LET, Ooi LC-L, Chan P-L, Rosli R, Ting N-C, Ithnin M (2011). Oil palm genomics: a foundation for improving agricultural productivity. In: Wahid MB, Choo YM, Chan KW (Eds.) Further advances in oil palm research 2000-2010. Malaysian Palm Oil Board, Bandar Baru Bangi, Malaysia, pp. 202-238.

Singh R, Ong-Abdullah M, Low ETL, Manaf MAA, Rosli R, Rajanaidu N, Ooi LCL, Ooi SE, Chan KL, Halim MA, Azizi N, Nagappan J, Bacher B, Lakey N, Smith SW, He D, Hogan M, Budiman MA, Lee EK, DeSalle R, Kudrna D, Goicoechea JL, Wing RA, Wilson RK, Fulton RS, Ordway JM, Martienssen RA, Sambanthamurthi R (2013a). Oil palm genome sequence reveals divergence of infertile species in old and new worlds. Nature 500, 335-339. https://doi.org/10.1038/nature12309.

Singh R, Low ETL, Ooi LCL, Ong-Abdullah M, Ting N-C, Nagappan J, Rajanaidu N, Amiruddin MD, Rosli R, Manaf MAA, Chan K-L, Halim MA, Azizi N, Lakey N, Smith SW, Budiman MA, Hogan M, Bacher B, Van Brunt A, Wang C, Ordway JM, Sambanthamurthi R, Martienssen RA (2013b)

The oil palm *SHELL* gene controls oil yield and encodes a homologue of *SEEDSTICK*. Nature 500, 340-344. https://doi.org/10.1038/nature12356.

Singh R, Ooi LCL, Ting, N-C, Low ETL, Ong-Abdullah M, Rajanaidu N, Marjuni M, Mustaffa S, Manaf MAA, Garner P, Reed J, Wischmeyer C, Beil M, Bacher B, Lakey ND, Ordway JM, Sambanthamurthi R (2021). A genetic platform for predicting and reducing non-*tenera* contamination in oil palm (*Elaeis guineensis*) seed supply. Tree Genet. Genomes 17, 45. https://doi.org/10.1007/s11295-021-01527-6.

Singh R, Low E-TL, Ooi LC-L, Ong-Abdullah M, Ting N-C, Nookiah R, Ithnin M, Marjuni M, Mustaffa S, Yaakub Z, Amiruddin MD, Manaf MAA, Chan K-L, Halim MA, Sanusi NSNM, Lakey ND, Sachdeva M, Bacher B, Garner P, MacDonald JD, Smith SW, Wischmeyer C, Budiman MA, Beil M, Stroff C, Reed J, Van Brunt A, Berg H, Ordway JM, Sambanthamurthi R (2020). Variation for heterodimerization and nuclear localization among known and novel oil palm SHELL alleles. New Phytol. 226, 426-440. https://doi.org/10.1111/nph.16387.

Soriano JM (2020). Molecular marker technology for crop improvement. Agronomy 10, 1462. https://doi.org/10.3390/agronomy10101462.

Suzana M, Rahimah AR, Maizura I, Singh R (2015). A simple and rapid protocol for isolation of genomic DNA from oil palm leaf tissue. J. Oil Palm Res. 27, 282-287.

Swaray S, Amiruddin MD, Rafii MY, Jamian S, Ismail MF, Jalloh M, Marjuni M, Mustakim MM, Yusuff O (2020). Influence of parental dura and pisifera genetic origins on oil palm fruit set ratio and yield components in their D x P progenies. Agronomy 10, 1793. https://doi.org/10.3390/agronomy10111793.

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. Sci. Rep. 6, 19075. https://doi.org/10.1038/srep19075.

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

Ting N-C, Jansen J, Mayes S, Massawe F, Sambanthamurthi R, Ooi LC-L, Chin CW, Arulandoo X, Seng TY, Alwee SSRS, Ithnin M, Singh R (2014). High density SNP and SSR-based genetic maps of two independent oil palm hybrids. BMC Genom. 15, 309. https://doi.org/10.1186/1471-2164-15-309.

Ting N-C, Yaakub Z, Kamaruddin K, Mayes S, Massawe F, Sambanthamurthi R, Jansen J, Low LET, Ithnin M, Kushairi A, Arulandoo X, Rosli R, Chan K-L, Amiruddin N, Sritharan K, Lim CC, Nookiah R, Mohd Din A, Singh R (2016). Fine-mapping and cross-validation of QTLs linked to fatty acid composition in multiple independent interspecific crosses of oil palm. BMC Genom. 17, 289. https://doi.org/10.1186/s12864-016-2607-4.

Ting N-C, Mayes S, Massawe F, Sambanthamurthi R, Chan K-L, Sritharan K, Singh R (2021a). Candidate genes linked to QTL regions associated with fatty acid composition in oil palm. Biologia 76, 267-279. https://doi.org/10.2478/s11756-020-00563-2.

Ting N-C, Zulkifli MA, Van de Weg E, Zaki NM, Serdari NM, Mustaffa S, Zainol Abidin MI, Sanusi NSNM, Smulders MJM, Low ETL, Ithnin M, Singh R (2021b). Detection of ploidy and chromosomal aberrations in commercial oil palm using high-throughput SNP markers. Planta 253, 63. https://doi.org/10.1007/s00425-021-03567-7.

Ukoskit K, Chanroj V, Bhusudsawang G, Pipatchartlearnwong K, Tangphatsornruang S, Tragoonrung S (2014). Oil palm (*Elaeis guineensis* Jacq.) linkage map, and quantitative trait locus analysis for sex ratio and related traits. Mol. Breed. 33, 415-424. https://doi.org/10.1007/s11032-013-9959-0.

Unterseer S, Bauer E, Haberer G, Seidel M, Knaak C, Ouzunova M, Meitinger T, Strom TM, Fries R, Pausch H, Bertani C, Davassi A, Mayer KF, Schön CC (2014). A powerful tool for genome analysis in maize: development and evaluation of the high density 600K SNP genotyping array. BMC Genom. 15, 823. https://doi.org/10.1186/1471-2164-15-823.

Van de Weg E, Di Guardo M, Jänsch N, Socquet-Juglard D, Costa F, Baumgartner I, Broggini GAL, Kellerhals M, Troggio M, Laurens F, Durel CE, Patocchi A (2018). Epistatic fire blight resistance QTL alleles in the apple cultivar 'Enterprise' and selection X-6398 discovered and characterized through pedigree-informed analysis. Mol. Breed. 38, 5. https://doi.org/10.1007/s11032-017-0755-0.

Van Ooijen (2018). JoinMap® 5, Software for the calculation of genetic linkage maps in experimental populations of diploid species. Kyazma B.V., Wageningen, Netherlands.

Vos PG, Uitdewilligen JG, Voorrips RE, Visser RG, Van Eck HJ (2015). Development and analysis of a 20K SNP array for potato (*Solanum tuberosum*): an insight into the breeding history. Theor. Appl. Genet. 128, 2387-2401. https://doi.org/10.1007/s00122-015-2593-y.

Vukosavljev M, Arens P, Voorrips RE, Van't Westende WPC, Esselink GD, Bourke PM, Cox P, Van de Weg E, Visser RGF, Maliepaard C, Smulders MJM (2016). High-density SNP-based genetic maps for the parents of an outcrossed and a selfed tetraploid garden rose cross, inferred from admixed progeny using the 68K rose SNP array. Hortic. Res. 3, 16052. https://doi.org/10.1038/hortres.2016.52.

Wang SQ (2019). Inbreeding and inbreeding depression of *Paeonia decomposita* (Paeoniaceae), a threatened endemic plant to China. Bot. Stud. 60, 28. https://doi.org/10.1186/s40529-019-0276-8.

Wei K, Wang X, Hao X, Qian Y, Li X, Xu L, Ruan L, Wang Y, Zhang Y, Bai P, Li Q, Aktar S, Hu X, Zheng G, Wang L, Liu B, He W, Cheng H, Wang L (2022). Development of a genome-wide 200K SNP array and its application for high-density genetic mapping and origin analysis of *Camellia sinensis*. Plant Biotechnol. J. 20, 414-416. https://doi.org/10.1111/pbi.13761.

Winfield MO, Allen AM, Burridge AJ, Barker GL, Benbow HR, Wilkinson PA, Coghill J, Waterfall C, Davassi A, Scopes G, Pirani A, Webster T, Brew F, Bloor C, King J, West C, Griffiths S, King I, Bentley AR, Edwards KJ (2016). High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. Plant Biotechnol. J. 14, 1195-206. https://doi.org/10.1111/pbi.12485.

Xia W, Luo T, Zhang W, Mason Annaliese S, Huang D, Huang X, Tang W, Dou Y, Zhang C, Xiao Y (2019). Development of high-density SNP markers and their application in evaluating genetic diversity and population structure in *Elaeis guineensis*. Front. Plant Sci. 10, 130. https://doi.org/10.3389/fpls.2019.00130.

Xu C, Ren Y, Jian Y, Guo Z, Zhang Y, Xie C, Fu J, Wang H, Wang G, Xu Y, Li P, Zou C (2017). Development of a maize 55K SNP array with improved genome coverage for molecular breeding. Mol. Breed. 37, 20. https://doi.org/10.1007/s11032-017-0622-z.

Yang H, Jian J, Li X, Renshaw D, Clements J, Sweetingham MW, Tan C, Li C (2015). Application of whole genome re-sequencing data in the development of diagnostic DNA markers tightly linked to a disease-resistance locus for marker-assisted selection in lupin (*Lupinus angustifolius*). BMC Genom. 16, 660. https://doi.org/10.1186/s12864-015-1878-5.

Ye Y, Cai M, Ju Y, Jiao Y, Feng L, Pan H, Cheng T, Zhang Q (2016). Identification and validation of SNP markers linked to dwarf traits using SLAF-seq technology in lagerstroemia. PLoS One 11, e0158970. https://doi.org/10.1371/journal.pone.0158970.

Yue GH, Ye BQ, Lee M (2021). Molecular approaches for improving oil palm for oil. Mol. Breed. 41, 22. https://doi.org/10.1007/s11032-021-01218-z.

Zhang X, Cai X (2011). Climate change impacts on global agricultural land availability. Environ. Res. Lett. 6, 014014. https://doi.org/10.1088/1748-9326/6/1/014014.

Zhang J, Zhang Q, Cheng T, Yang W, Pan H, Zhong J, Huang L, Liu E (2015). High-density genetic map construction and identification of a locus controlling weeping trait in an ornamental woody plant (*Prunus mume* Sieb. et Zucc). DNA Res. 22, 183-91. https://doi.org/10.1093/dnares/dsv003.

Zolkafli SH, Ting N-C, Sanusi NSNM, Ithnin M, Mayes S, Massawe F, Sambanthamurthi R, Ismail I, Zainol Abidin MI, Roowi SH, Lee YP, Hanafi NFF, Singh R (2021). Comparison of quantitative trait loci (QTLs) associated with yield components in two commercial *dura* × *pisifera* breeding crosses. Euphytica 217, 104. https://doi.org/10.1007/s10681-021-02825-9.

Zulkifli Y, Norziha A, Naqiuddin MH, Fadila, AM, Nor Azwani AB, Suzana M, Samsul KR, Ong-Abdullah M, Singh R, Ahmad Parveez GK, Kushairi A (2017). Designing the oil palm of the future. J. Oil Palm Res. 29, 440-455. https://doi.org/10.21894/jopr.2017.00015.