

Identification of candidate genes responsible for morphological and yield-related traits via genome-wide association analysis (GWAS) in oil palm (*Elaeis oleifera* x *Elaeis guineensis*)

Jaime Osorio

AGROSAVIA

Gina Garzón

AGROSAVIA

Paola Delgadillo

AGROSAVIA

Silvio Bastidas

AGROSAVIA

Leidy Moreno

AGROSAVIA

Felix Enciso

AGROSAVIA

Omar Comejo

Washington State University

Luz Stella Barrero (✉ lbarrero@agrosavia.co)

AGROSAVIA

Research article

Keywords: Association mapping, *Elaeis guineensis*, *Elaeis oleifera*, genotyping-by-sequencing, plant architecture, yield.

Posted Date: August 19th, 2019

DOI: <https://doi.org/10.21203/rs.2.13122/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published on December 3rd, 2019. See the published version at <https://doi.org/10.1186/s12870-019-2153-8>.

Abstract

Background The genus *Elaeis* has two species of economic importance for the oil palm agroindustry: *Elaeis oleifera* (O), native to the Americas, and *Elaeis guineensis* (G), native to Africa. The breeding program in Colombia relies on interspecific OxG crossing populations with tolerance to pests and diseases, high oil quality, and acceptable fruit bunch production. The identification of loci associated to morphological and yield-related traits and the dissection of their genetic architecture will provide essential insights for oil palm breeding strategies. **Results** The genotypes of 471 oil palms, including 62 *E. oleifera* (O), 31 *E. guineensis* (G) and 378 OxG samples were analyzed in this study. A total of 3,776 single nucleotide polymorphisms (SNP) were detected across the 16 oil palm chromosomes using the genotyping-by-sequencing (GBS) technique. The genetic variation and population structure analyses grouped the samples into two clades according to the parental relatedness. A genome wide association analysis (GWAS) was conducting using the OxG hybrid population, resulting in 12 SNPs significantly associated with ten different morphological and yield-related traits. **Conclusions** The work presented herein provides to our knowledge the first association mapping study in an interspecific OxG hybrid population of oil palm. We provide new insights on candidate genes involved in tissue development and plant architecture associated to traits such as: rachis length, trunk diameter, bunch number, and bunch weight. The genes identified in our analysis are putative candidates for future targeted functional analysis. They are valuable resources for the development of marker-assisted selection in oil palm breeding. **Keywords:** Association mapping, *Elaeis guineensis*, *Elaeis oleifera*, genotyping-by-sequencing, plant architecture, yield.

Background

The oil palm is an important crop with a higher quality oil and yield potential compared to other vegetable oil crops [1]. Colombia is the fourth-largest producer worldwide with 3.8 tons of palm oil per hectare, which positions Colombia above the world's average yield [2]. Within the Arecaceae family, the oil palm species (*Elaeis guineensis*), native to West Africa, is the main source of most vegetable oil [3]. However, another oil palm species native to the tropical Central and South America, known as *Elaeis oleifera*, is recognized for its high yield production [3]. The palm is a perennial monocot with a lifespan of approximately 25 years [4], which results in a slow breeding progress. The Colombian Corporation of Agricultural Research (Agrosavia) breeding program has focused on developing OxG interspecific crosses (*E. oleifera* x *E. guineensis*). OxG hybrids are characterized by a slow trunk growth [5], as well as tolerance to bud rot [6–9] and red ring diseases [10] in comparison to their parents. Additionally, OxG hybrids inherit the parthenocarpic fruit development of *E. oleifera* species, which allows the production of seedless fruits [11].

Over the last 20 years, a genetic map of the oil palm have been constructed [12]. Saturated genetic linkage maps are important for the identification of genomic regions associated with major genes and quantitative trait loci (QTLs) controlling agronomic traits. The first marker-based genetic maps for oil palm were generated using restriction fragment length polymorphisms (RFLPs) and amplified fragment length polymorphisms (AFLPs) [13, 14]. Dense genetic maps were subsequently constructed using simple sequence repeats (SSRs) and single nucleotide polymorphism (SNP) markers, which have also been used for QTL identification. Thus, Jeennor and Volkaert [15] identified a QTL associated with bunch weight using a mapping population of 69 samples and a genetic map constructed with 89 SSRs and 101 SNPs. Billotte *et al.* [16], used a multi-parent linkage map generated with 251 SSRs and reported QTLs associated to bunch traits. Similar approaches have enabled the identification of 164 QTLs associated with 21 oil yield component using SSR, AFLP, and RFLP markers [17].

In recent years, advances in next-generation sequencing technology have lowered the cost of DNA sequencing to the point that we can obtain millions of SNPs compared with other technologies [18, 19]. The technique genotyping by sequencing (GBS) is a rapid, low-cost, and robust approach to screen breeding populations using hundreds or thousands of SNPs [20]. Pootakham *et al.* [21] constructed an oil palm map using an F₂ population and 1,085 SNPs derived from GBS and were able to identify QTLs for height and fruit bunch weight. Similarly, a genome-wide association analysis (GWAS), using a larger number of SNPs (4,031) derived from GBS across a diverse panel of *E. guineensis*, allowed to identify novel QTLs associated with the increasing in the trunk height [22].

GWAS have been proposed as a robust approach over QTL linkage mapping [23]. The use of a wide range of genetic background in GWAS analyses increases the possibility to detect QTL regions associated with traits of interest, compared to the limited genetic variation of a bi-parental mapping population [24]. However, GWAS limitations such as the effect of population structure can lead to spurious associations between a candidate marker and a phenotype [25]. To solve this, the mixed linear model incorporates the structure data (Q) and the relative kinship effects (K), resulting in the reduction of false positive associations [26].

Given the food, industrial, and medical purposes, palm oil has experienced rapid growth in economic importance and nowadays is considered the second most traded vegetable oil crop in the world after soybean [27, 28]. The increasing demand of this crop is caused by a shift away from trans-fats to healthier alternatives [29], and because its residues can be processed to produce biofuel [28]. For these reasons, the identification of specific genes involved in morphological traits such as height, foliar area, and its relationship with productivity, is becoming more critical for this crop.

Although previous studies have identified QTLs controlling morphological and yield-related traits in oil palm, these QTLs were detected using F₁ intraspecific populations. Our study is the first report in which molecular markers were mapped through association analysis in an interspecific OxG hybrid populations. Our study aims to: (i) genotype an OxG oil palm mapping population; (ii) analyze genetic diversity, population structure and linkage disequilibrium; (iii) perform GWAS to identify loci or candidate genes involved in yield and other morphological traits for future use in breeding programs.

Methods

Plant material

The plant materials used for current GWAS study came from the National Germplasm Collection of Colombia, maintained at the Colombian Corporation of Agricultural Research (Agrosavia). All accessions were collected following national regulations according to the genetic resources agreement for scientific research without commercial interest No. 74, signed between Agrosavia and the Ministry of Environment and Sustainable Development from Colombia.

A population of 471 oil palm samples consisted of (62) *E. oleifera*, (31) *E. guineensis* and (378) OxG F₁ interspecific hybrid was used in this study. The OxG samples were generated through eight different crossings, however the parents of these crossing are already deceased. For the purpose of this study, other *E. oleifera* genotypes native to the northern of Colombia and *E. guineensis* genotypes that have a Yamgambi and Deli origin were used to estimate the genetic relationships. The OxG and *E. oleifera* samples were collected from the Research Center El Mira and the *E. guineensis* samples were collected from the Research Center La Libertad of Agrosavia [54]. Details of the plant materials can be found in Table S1.

Genotyping and SNP calling

Genomic DNA was extracted from young leaf tissue using the DNeasy Plant Mini Kit (QIAGEN, Germany). All 471 samples were genotyped using the GBS protocol following the procedure cited by Elshire *et al.* [20]. GBS library preparation and sequencing were performed at the Institute of Genomic Diversity (Cornell University, Ithaca, NY, United States). Briefly, samples were digested with the methylation-sensitive restriction enzyme *Pst*I, which has a six base pair recognition site (CTGCAG). Sequencing was performed with 100-bp single-end reads using the Illumina HiSeq 2000 platform (Illumina Inc., United States).

The raw data was demultiplexed using the standard pipeline from the Tassel v4.5.9 software [55]. Then, reads were mapped to the oil palm reference genome of *E. guineensis* [56] using Bowtie2 [57] with the very-sensitive option. SNP calling was performed using the following parameters: minor allele frequency (MAF) < 5%, minimum locus coverage (mnLCov) of 0.9, minimum site coverage (mnScov) of 0.7 and minimum taxon coverage (mnTCov) of 0.5. Finally, SNPs were filtered to retain 5% of missing data and biallelic SNPs with VCFtools v0.1.13 software [58]. SNP data was integrated with phenotypic data to perform the GWAS analysis. The physical positions of SNP markers used in the association analysis were obtained from the Genomsawit Website of the International Malaysian Oil Palm Genome Programme (http://gbrowse.mpob.gov.my/fgb2/gbrowse/Eg5_1/).

Genetic diversity and population structure

High quality SNP markers were used to assess oil palm genetic diversity in 471 samples with breeding coefficient (F) values and germplasm relatedness through a Neighbor-Joining tree according to the Nei's genetic distance matrix. Genetic diversity (π), Tajima's D, and population pairwise F-statistics (F_{ST}) were calculated using VCFtools v0.1.13 [58]. Similarly, the population structure of the 471 samples was estimated using the Admixture v1.3.0 software [59] in both unsupervised and supervised modes.

Phenotypic analysis

The OxG hybrid population (378 samples) has been planted in field and randomly distributed using a randomized complete block design with four blocks. The experimental field is located in the Research Center El Mira. The field was planted in a quincunx or triangular system with 10 meters between the plants; this planting system allowed a density of 115 oil palm trees per hectare. All plants were grown under the same standard agronomic practices.

Phenotypic data of seven morphological measurements and three yield-related traits were used in this study (Table 1). Each trait was measured according to the methodology proposed by Corley *et al.* [60] and Breure [61]. To assess the relationship among the studied traits, a principal component analysis (PCA) and a Pearson's correlation were calculated. A hierarchical cluster analysis using Ward's method was carried out to analyze the genetic relationship among samples by the use of all phenotypic variables. All analyses were performed using the R statistical package [62].

Marker-trait association analysis

Marker-trait association analysis was performed on morphological and yield-related traits using the OxG hybrid population. A unified mixed linear model with a kinship matrix and PCA results were used in the R package GAPIT (Genome Association and Prediction Integrated Tool) [63]. To remove any possible bias caused by population structure, we included the first five principal components calculated using the R package SNPrelate [64], and a relatedness (kinship) matrix in the mixed linear model. Association mapping model evaluations were based on visual observations of the Manhattan plots. Q-Q plots were plotted of the observed $-\log_{10}$ P-values and the expected $-\log_{10}$ P-values to study the appropriateness of the GWAS model. A false discovery rate (FDR) [65] was used to correct for spurious associations.

the quantile-quantile

(QQ) plot supported the appropriateness of the GWAS

model (Fig. 3a).

the quantile-quantile

(QQ) plot supported the appropriateness of the GWAS

model (Fig. 3a).

The heatmap of LD was investigated with a custom script by plotting pairwise R^2 values against the physical distance (base pairs) between markers on the same chromosome.

Candidate genes identification

Genes annotations under the candidate gene regions were determined using published genome information of *E. guineensis* [56]. To assign putative biological functions of significant SNP markers associated with the traits, the flanking sequences of SNPs were queried against databases, such as: HMMER (<https://www.ebi.ac.uk/Tools/hmmer/>), NCBI (<http://www.ncbi.nlm.nih.gov/>), European Molecular Biology Laboratory (<http://www.ebi.ac.uk/>) and European Nucleotide Archive (<http://www.ebi.ac.uk/ena>).

Results

Improving oil quality and increasing yield per hectare in oil palm is a major concern in the oil processing industry. Agrosavia's palm breeding program has focused on interspecific crosses of OxG. According to Bastidas [31] the OxG hybrids of Agrosavia present heterosis in traits such as resistance to diseases, fruit number and weight, leaf length and trunk diameter. To our knowledge, this study represents the first geno-phenotypic and GWAS analysis of an OxG hybrid population. Our 378 OxG hybrid population was screened for a set of morphological and yield-related traits and genotyped with 3,776 high-quality SNPs uncovered by a GBS approach. Correlation analysis results among yield-related traits indicated that BN could be a potential and better selection criterion for production than BW in the hybrid population. It is known that the leaf emission rate determines the bunch number in a palm, being BN negative correlated with number of leaves [32]. In our study, no significant correlations were found between yield and leaf-related traits (FA, LA, LDW, LXL, RL), however, a previous study in *E. oleifera* and OxG found that BN can be greater than number of leaves just when oil palms carry in their genotype a prolific trait [33]. Increases in BN and BW are also expected to correlate with increased mesocarp and kernel oil yields as shown in other oil palm germplasm studies [34]. Future studies specifically related to oil yields for the OxG hybrid population should be conducted considering their importance in oil palm breeding.

GBS and its combination with GWAS has allowed the genetic dissection of variation in complex traits in many plant species [35–37]. Specifically, in oil palm the GBS technique has been used for identifying candidate genes in intraspecific populations related to oil bunch [38], average bunch weight [21, 38] and stem height [22] with the enzyme *ApeKI*, meanwhile the enzymes *PstI-MspI* have been used for oil quality traits studies [39]. We used GBS with the enzyme *PstI* in morphological and yield-related traits in an interspecific OxG hybrid population. The use of this enzyme allowed the discovery of new genetic variants, which according to Chung *et al.* [40] is one of the most important advantages of GBS.

In the present study, association mapping resulted in the identification of 12 SNPs related to 10 morphological and yield-related traits (Table 2). For morphological traits, a significant association was found for LDW on chromosome 3 explaining 10% of the phenotypic variation. This SNP was located in a mechanosensitive (MS) ion channel protein 10-like (*MSL10*) gene. In plants, MS ion channels have been proposed to play a wide array of roles, from the perception of touch and gravity to the osmotic homeostasis of intracellular organelles [41]. Besides, mechanoperception genes are essential for normal cell and tissue growth and development as well as for the proper response to an array of biotic and abiotic stresses [42]. A second gene associated with TD was identified on chromosome 15. This gene is involved in nucleic acid binding and has a C2H2-type Zinc finger domain. The C2H2-ZF gene family has been proposed to be involved in the formation of wood and shoot and cambium development in species such as Poplar, as well as playing a role in stress and phytohormone response [43].

For HT trait, different studies have reported associated QTLs in chromosomes 2, 6, 7 and 9 [22, 34]. In our study, we reported three candidate genes on chromosome 15, which is similar to the results reported by Pootakham *et al.* [21]. However, our candidate genes were positioned in the vicinity of the ones reported by Pootakham *et al.* [21], which highlights the importance of this region (from 19.3 to 23.6 Mbp) in the phenotypic variance of HT (Table 2). The closest gene to the SNPs S15_22553489 and S15_22553493 SNPs, corresponds to a *STYK* gene, which is involved in the control of stomatal movement in response to CO₂ [44]. Recent studies also showed the role of *STYK* gene in stem diameter by increasing the number of xylary fibers in species such as *Bambusa balcooa* [45].

For RL and LXL traits QTLs have been reported on chromosomes 4, 2, 10 and 16 [34]. In our study, four SNPs were associated with four different candidate genes for RL on chromosome 13. The SNP S13_20856724 is the closest to the *AGC3* gene and encodes different G proteins. G proteins had been reported to be involved in a wide range of developmental and physiological processes, having a high potential for yield improvement in crops such as rice [46]. The other significant association was found with the SNP S13_23674227, which is located in an extracellular ribonuclease gene (*RNase* gene). *RNase* genes have been studied for years in plants, playing an important role in plant defense mechanism [47] or plant development due to their ability to modify RNA levels, and thereby influence protein synthesis [48]. Other candidate genes were also found for RL and LXL, but further studies are necessary to determine their role in regulating these traits.

For yield-related traits, previously studies reported associated SNPs in chromosomes 1, 3, 4 and 6 [21, 38, 49]. In our study, other major associations peaks were observed on chromosomes 5 and 10, explaining 11% of the phenotypic variance. A significant SNP related to Yield and BN was located in the gene p5.00_sc00003_p0367, coding for a cation/H(+) antiporter gene. Antiporter proteins function as regulators of monovalent ions, pH homeostasis, and developmental processes in plants [50]. On chromosome 10, the gene p5.00_sc00004_p0097, associated with BW, encodes the zinc finger protein 8. The zinc finger proteins (ZFP) are a large protein family, involved in plant development, regulation of plant height, root development, flower development, seed germination, secondary wall thickening, anther development, and fruit ripening [51]. Studies conducted by Wu *et al.* [52] demonstrated that silencing a gene related to ZFP hampered fruit development in *Nicotiana benthamiana* [52]. This *ZFP* gene might play an important role on yield-related traits in oil palm, as shown in other plants, where overexpression of zinc finger proteins are related with higher yields in crops [53] although, further analysis are needed to determine its role in bunch weight and yield in oil palm.

In oil palm, harvesting of fruit bunches after certain age is a very difficult labor due to their tallness. For this reason, genotypes with less HT and TD are preferred among oil palm farmers. Likewise, larger foliar (RL and LDW) is related to a greater photosynthetic production which could be involve in higher productivity. But most importantly, increasing the number and weight of fruits means higher productivity per palm and therefore higher incomes for farmers. For this reason, leveraging QTLs or genes related to these traits could contribute to develop plant breeding strategies such as marker-assisted selection (MAS), that helps to select promising accessions in in earlier stages (greenhouse conditions), and therefore reduce the breeding cycle. Further work needs to be focus on the biological functions of the set of candidate genes found in our research, though the correlations identified in association studies cannot be dubbed as causations.

Conclusions

The present study is the first one to report significant SNPs associated with morphological and yield-related traits based on GWAS on an interspecific OxG population. Nine of these SNPs are located within chromosomes reported in previous mapping studies, however, the set of genes presented in our analysis could be of value to locate with more precision the intervals of the reported QTLs on chromosomes 13 and 15. Also, candidate genes discovered on chromosomes 3, 5 and 10 have not been reported for the studied traits. The findings from the present study will provide the groundwork for the development of marker assisted breeding in oil palm and will serve as a strong base for future functional studies to dissect high yield production in oil palm.

Abbreviations

FDR, False-Discovery-Rate; GBS, Genotyping By Sequencing; GWAS, Genome-Wide Association Studies; LD, Linkage disequilibrium; MAS Marker-Assisted Selection; PCA, Principal Component Analysis; QTL, Quantitative Trait Loci; SNP, Single Nucleotide Polymorphism; RFLPs, Restriction Fragment Length Polymorphisms; AFLPs, Amplified Fragment Length Polymorphisms; SSRs, Simple Sequence Repeats; TD, Trunk Diameter; HT, Trunk Height; RL, Rachis Length; LDW, Leaf Dry Weight; FA, Foliar Area; LA, Leaf Area; LXL, Leaflet Per Leaf; BW, Bunch Weight; BN, bunch number.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent to publish

Not applicable.

Availability of data and materials

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

Competing interests

The authors declare they have no competing interests.

Funding

Publication of this article has been funded by TV-17 supported by the Ministry of Agriculture and Rural Development of the Republic of Colombia. The funding entities had no role in study design, data collection and analysis, interpretation, decision to publish, or preparation of the manuscript.

Authors' contributions

LSB, FEER and SBP conceptualized and conceived the project and its components. SBP, LPD, GAGM and LPM collected the samples and the trait data. JAOG and GAGM carried out the genotypic and phenotypic analysis with the supervision of OEC. JAOG and GAGM wrote the manuscript and LSB, OEC and FEER corrected and edited it. All authors reviewed and contributed to draft the manuscript as well as read and approved the final manuscript.

Acknowledgements

The authors would like to acknowledge to William Tolosa for his support on sample collection, to Jhon Berdugo for his support in the data analysis and Marco Antonio Lopez to provide the script for the LD heatmap analysis. The authors thank Joanna Kelley for assistance in revising the final version of the manuscript.

References

1. Murphy D. Oil palm: Future prospects for yield and quality improvements. 2009.
2. Pacheco P, Gnych S, Dermawan A, Komarudin H, Okarda B. The palm oil global value chain: Implications for economic growth and social and environmental sustainability. Bogor, Indonesia; 2017.
3. Barcelos E, Rios S de A, Cunha RN V, Lopes R, Motoike SY, Babiychuk E, et al. Oil palm natural diversity and the potential for yield improvement. *Front Plant Sci.* 2015;6:190. doi:10.3389/fpls.2015.00190.
4. Srestasathien P, Rakwatin P. Oil Palm tree detection with high resolution multi-spectral satellite imagery. *Remote Sens.* 2014;6:9749–74. doi:10.3390/rs6109749.
5. Escobar R, Alvarado A. Estrategias para la producción comercial de semillas y clones de palmas de aceite compactas. *Rev Palmas.* 2004;25:293–305. <https://publicaciones.fedepalma.org/index.php/palmas/article/view/1093>.
6. Turner PD. Oil Palm Diseases and Disorders. Oxford University Press; 1981. <https://books.google.com.co/books?id=mAnyXwAACAAJ>.
7. Amblard P, Billotte N, Cochard B, Durand-Gasselin T, Jacquemard JC, Louise C, et al. El mejoramiento de la palma de aceite *Elaeis guineensis* y *Elaeis oleifera* por el Cirad-CP. *Rev Palmas.* 2002:306–10.
8. Zambrano JE. Los híbridos interespecíficos *Elaeis oleifera* HBK. x *Elaeis guineensis* Jacq.: una alternativa de renovación para la Zona Oriental de Colombia. *Rev Palmas.* 2004;25:339–49. <http://publicaciones.fedepalma.org/index.php/palmas/article/view/1098>.
9. Chinchilla C. Tolerancia y resistencia a las pudriciones del cogollo en fuentes de diferente origen de *Elaeis guineensis*. *Rev Palmas.* 2007;28:273–84.
10. Moura J. Manejo integrado das pragas das palmeiras. Ilheus, BA: Centro de Pesquisas do Cacau; 2017.
11. Hartley CWS. The oil palm (*Elaeis guineensis* Jacq.). Second. 1967.
12. Ong A-L, Teh C-K, Kwong Q-B, Tangaya P, Appleton DR, Massawe F, et al. Linkage-based genome assembly improvement of oil palm (*Elaeis guineensis*). *Sci Rep.* 2019;9:6619. doi:10.1038/s41598-019-42989-y.
13. Mayes S, Jack PL, Corley RH V, Marshall DF. Construction of a RFLP genetic linkage map for oil palm (*Elaeis guineensis* Jacq.). *Genome.* 1997;40:116–22.
14. Purba AR, Noyer JL, Baudouin L, Perrier X, Hamon S, Lagoda PJL. A new aspect of genetic diversity of Indonesian oil palm (*Elaeis guineensis* Jacq.) revealed by isoenzyme and AFLP markers and its consequences for breeding. *Theor Appl Genet.* 2000;101:956–61. doi:10.1007/s001220051567.
15. Jeennor S, Volkaert H. Mapping of quantitative trait loci (QTLs) for oil yield using SSRs and gene-based markers in African oil palm (*Elaeis guineensis* Jacq.). *Tree Genet genomes.* 2014;10:1–14.
16. Billotte N, Marseillac N, Risterucci A-M, Adon B, Brottier P, Baurens F-C, et al. Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). *Theor Appl Genet.* 2005;110:754–65.
17. Seng T-YY, Ritter E, Mohamed Saad SH, Leao L-JJ, Harminder Singh RS, Qamaruz Zaman F, et al. QTLs for oil yield components in an elite oil palm (*Elaeis guineensis*) cross. *Euphytica.* 2016;212:399–425. doi:10.1007/s10681-016-1771-6.
18. Yadav P, Vaidya E, Rani R, Yadav N, Singh B, K. Rai P, et al. Recent perspective of next generation sequencing: applications in molecular plant biology and crop improvement. 2016.

19. Nguyen K Le, Grondin A, Courtois B, Gantet P. Next-generation sequencing accelerates crop gene discovery. *Trends Plant Sci.* 2019;24:263–74. doi:<https://doi.org/10.1016/j.tplants.2018.11.008>.
20. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS One.* 2011;6:1–10. doi:10.1371/journal.pone.0019379.
21. Pootakham W, Jomchai N, Ruang-areerate P, Shearman JR, Sonthirod C, Sangsrakru D, et al. Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics.* 2015;105:288–95. doi:<https://doi.org/10.1016/j.ygeno.2015.02.002>.
22. Babu BK, Mathur RK, Ravichandran G, Venu MVB. Genome-wide association study (GWAS) for stem height increment in oil palm (*Elaeis guineensis*) germplasm using SNP markers. *Tree Genet Genomes.* 2019;15:1–8.
23. Huang X, Han B. Natural variations and genome-wide association studies in crop plants. *Annu Rev Plant Biol.* 2014;65:531–51. doi:10.1146/annurev-arplant-050213-035715.
24. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods.* 2013;9:29. doi:10.1186/1746-4811-9-29.
25. Burghardt LT, Young ND, Tiffin P. A guide to genome-wide association mapping in plants. *Curr Protoc Plant Biol.* 2017;2:22–38. doi:10.1002/cppb.20041.
26. Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, et al. Mixed linear model approach adapted for genome-wide association studies. *Nat Genet.* 2010;42:355. <https://doi.org/10.1038/ng.546>.
27. FAO - Trade and market division. Oilcrops. 2014. http://www.fao.org/fileadmin/templates/est/COMM_MARKETS_MONITORING/Oilcrops/Documents/Food_outlook_oilseeds/Food_Outlook_May_2014_OI
28. Kurnia JC, Jangam S V, Akhtar S, Sasmito AP, Mujumdar AS. Advances in biofuel production from oil palm and palm oil processing wastes: A review. *Biofuel Res J.* 2016;3:332–46. doi:10.18331/BRJ2016.3.1.3.
29. World Growth. The economic benefit of palm oil to Indonesia. World Growth Palm Oil Green Dev Campaign. 2011; February:1–27. http://worldgrowth.org/site/wp-content/uploads/2012/06/WG_Indonesian_Palm_Oil_Benefits_Report-2_11.pdf.
30. Sato S, Tabata S, Hirakawa H, Asamizu E, Shirasawa K, Isoe S, et al. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature.* 2012;485:635–41.
31. Bastidas Perez S. Avances en el desarrollo de materiales genéticos resistentes a la PC. *Rev Palmas.* 2013;34:135–41.
32. Breure CJ. Desarrollo de las hojas en la palma de aceite (*Elaeis guineensis*) y determinación de la tasa de apertura de las hojas. *Rev Palmas.* 1996;17. <https://publicaciones.fedepalma.org/index.php/palmas/article/view/563>.
33. Bastidas S, Hurtado PYL. Evaluación de palmas prolíficas en la especie *Elaeis oleifera* e híbridos interespecíficos de *E. oleifera* x *E. guineensis*. 1993:55–60.
34. Ithnin M, Xu Y, Marjuni M, Serdari NM, Amiruddin MD, Low E-TL, et al. Multiple locus genome-wide association studies for important economic traits of oil palm. *Tree Genet Genomes.* 2017;13:103. doi:10.1007/s11295-017-1185-1.
35. Guo D-L, Zhao H-L, Li Q, Zhang G-H, Jiang J-F, Liu C-H, et al. Genome-wide association study of berry-related traits in grape [*Vitis vinifera* L.] based on genotyping-by-sequencing markers. *Hortic Res.* 2019;6:11. doi:10.1038/s41438-018-0089-z.
36. Otto L-G, Mondal P, Brassac J, Preiss S, Degenhardt J, He S, et al. Use of genotyping-by-sequencing to determine the genetic structure in the medicinal plant chamomile, and to identify flowering time and alpha-bisabolol associated SNP-loci by genome-wide association mapping. *BMC Genomics.* 2017;18:599. doi:10.1186/s12864-017-3991-0.
37. Hu X, Zuo J, Wang J, Liu L, Sun G, Li C, et al. Multi-locus genome-wide association studies for 14 main agronomic traits in Barley. *Front Plant Sci.* 2018;9:1683. doi:10.3389/fpls.2018.01683.
38. Babu BK, Mathur RK, Ravichandran G, Anita P, Venu MVB. Genome wide association study (GWAS) and identification of candidate genes for yield and oil yield related traits in oil palm (*Elaeis guineensis*) using SNPs by genotyping-based sequencing. *Genomics.* 2019. doi:<https://doi.org/10.1016/j.ygeno.2019.06.018>.
39. Bai B, Wang L, Lee M, Zhang Y, Rahmadsyah, Alfiko Y, et al. Genome-wide identification of markers for selecting higher oil content in oil palm. *BMC Plant Biol.* 2017;17:93. doi:10.1186/s12870-017-1045-z.
40. Chung YS, Choi SC, Jun TH, Kim C. Genotyping-by-sequencing: a promising tool for plant genetics research and breeding. *Hortic Environ Biotechnol.* 2017;58:425–31.
41. Hamilton ES, Schlegel AM, Haswell ES. United in diversity: mechanosensitive ion channels in plants. *Annu Rev Plant Biol.* 2015;66:113–37. doi:10.1146/annurev-arplant-043014-114700.
42. Haswell ES, Peyronnet R, Barbier-Brygoo H, Meyerowitz EM, Frachisse JM. Two MscS homologs provide mechanosensitive channel activities in the *Arabidopsis* root. *Curr Biol.* 2008;18:730–4.
43. Liu Q, Wang Z, Xu X, Zhang H, Li C. Genome-wide analysis of C2H2 zinc-finger family transcription factors and their responses to abiotic stresses in poplar (*Populus trichocarpa*). *PLoS One.* 2015;10:1–25.
44. Hashimoto M, Negi J, Young J, Israelsson M, Schroeder JI, Iba K. Arabidopsis HT1 kinase controls stomatal movements in response to CO₂. *Nat Cell Biol.* 2006;8:391–7. doi:10.1038/ncb1387.
45. Ghosh JS, Chaudhuri S, Dey N, Pal A. Functional characterization of a serine-threonine protein kinase from *Bambusa balcooa* that implicates in cellulose overproduction and superior quality fiber formation. *BMC Plant Biol.* 2013;13:128. doi:10.1186/1471-2229-13-128.
46. Botella JR. Can heterotrimeric G proteins help to feed the world? *Trends Plant Sci.* 2012;17:563–8. doi:10.1016/j.tplants.2012.06.002.

47. Sangaev SS, Kochetov A V, Ibragimova SS, Levenko BA, Shumny VK. Physiological role of extracellular ribonucleases of higher plants. *Russ J Genet Appl Res.* 2011;1:44–50. doi:10.1134/S2079059711010060.
48. Tvorus EK. Plant ribonucleases. *Sov plant Physiol.* 1976.
49. Billotte N, Jourjon MF, Marseillac N, Berger A, Flori A, Asmady H, et al. QTL detection by multi-parent linkage mapping in oil palm (*Elaeis guineensis* Jacq.). *Theor Appl Genet.* 2010;120:1673–87. doi:10.1007/s00122-010-1284-y.
50. Ma Y, Wang J, Zhong Y, Geng F, Cramer GR, Cheng Z-M (Max). Subfunctionalization of cation/proton antiporter 1 genes in grapevine in response to salt stress in different organs. *Hortic Res.* 2015;2:15031. <https://doi.org/10.1038/hortres.2015.31>.
51. Zang D, Li H, Xu H, Zhang W, Zhang Y, Shi X, et al. An Arabidopsis zinc finger protein increases abiotic stress tolerance by regulating sodium and potassium homeostasis, reactive oxygen species scavenging and osmotic potential. *Front Plant Sci.* 2016;7:1272. doi:10.3389/fpls.2016.01272.
52. Wu W, Cheng Z, Liu M, Yang X, Qiu D. C3HC4-type RING finger protein NbZFP1 is involved in growth and fruit development in *Nicotiana benthamiana*. *PLoS One.* 2014;9:e99352–e99352. doi:10.1371/journal.pone.0099352.
53. Zhou B, Lin JZ, Peng D, Yang YZ, Guo M, Tang DY, et al. Plant architecture and grain yield are regulated by the novel DHHC-type zinc finger protein genes in rice (*Oryza sativa* L.). *Plant Sci.* 2017;254:12–21. doi:<https://doi.org/10.1016/j.plantsci.2016.08.015>.
54. P. SB, R. EAP, C. RR. Genealogía del germoplasma de palma de aceite (*Elaeis guineensis* Jacq.) del proyecto de mejoramiento genético de Corpoica. *Rev Palmas.* 2003;24. <https://publicaciones.fedepalma.org/index.php/palmas/article/view/950>.
55. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics.* 2007;23:2633–5. doi:10.1093/bioinformatics/btm308.
56. Singh R, Ong-Abdullah M, Low E-TL, Manaf MAA, Rosli R, Nookiah R, et al. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature.* 2013;500:335. <https://doi.org/10.1038/nature12309>.
57. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10:R25. doi:10.1186/gb-2009-10-3-r25.
58. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27:2156–8. doi:10.1093/bioinformatics/btr330.
59. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19:1655–64. doi:10.1101/gr.094052.109.
60. Corley RH V, Hardon JJ, Tan GY. Analysis of growth of the oil palm (*Elaeis guineensis* Jacq.) I. Estimation of growth parameters and application in breeding. *Euphytica.* 1971;20:307–15. doi:10.1007/BF00056093.
61. Breure CJ. Factors associated with the allocation of carbohydrates to bunch dry matter production in oil palm (*Elaeis guineensis* Jacq.). *Landbouwniversiteit*; 1987.
62. R development core team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2008. <http://www.r-project.org>.
63. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: genome association and prediction integrated tool. *Bioinformatics.* 2012;28:2397–9. doi:10.1093/bioinformatics/bts444.
64. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A High-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics.* 2012;28:3326–8. doi:10.1093/bioinformatics/bts606.
65. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc.* 1995;57:289–300.

Tables

Table 1. Mean values, standard deviation (SD) and minimum and maximum values from the phenotypic traits used in this study

Category	Trait	Abbreviation	Unit	Mean	SD	Minimum value	Maximum value
Morphological	Trunk Diameter	TD	cm	88.5	6.0	62.4	102
	Trunk Height	HT	cm	250.3	29.5	133.3	327
	Rachis length	RL	cm	421.5	35.3	275.5	530
	Leaf dry weight	LDW	kg	2.2	0.3	1.3	3.7
	Foliar area	FA	m ²	385	78.2	141.3	617.1
	Leaf area	LA	m ²	8.6	1.3	4.7	12.7
Yield	Leaflet per leaf	LXL	unit	234.8	14.8	184	294
	Bunch weight	BW	kg	6.1	1.8	1	19.5
	Bunch number	BN	unit	8.8	5.0	1.0	27
	Yield per palm	Yield	kg	56.5	39.1	1.8	233

Table 2. Significant marker-trait associations for morphological and yield related traits using a mixed linear model approach.

Category	Trait	SNP	Chromosome	Position (kb)	P-value	MAF	R ²	FDR Adjusted P-values	Candidate Gene	SNP position relative to the candidate gene	Candidate gene annotation	
Morphological	LDW	S3_30467222	3	30467222	1.E-04	0.106	0.107	0.050	p5.00_sc00100_p0017	0	Mechanosensitive ion channel protein 10-like (MSL10)	
	TD	S15_21239833	15	21239833	1.98E-05	0.493	0.097	0.010	p5.00_sc00036_p0097	+21.8 kb	Nucleic acid binding	
	HT	S15_22347191	15	22347191	7.94E-05	0.496	0.098	0.084	p5.00_sc00036_p0145	+0.3 kb	Paired amphipathic helix protein (PAH)	
	TD	S15_22553489	15	22553489	7.94E-05	0.496	0.098	0.084	p5.00_sc00036_p0152	-0.6 kb	Serine threonine-protein kinase (STYK)	
	FA											
	LA	S15_22553493	15	22553493	8.90E-05	0.495	0.098	0.084				
		S15_23645020	15	23645020	7.94E-05	0.496	0.098	0.084	p5.00_sc00036_p0217	0	Class E vacuolar protein-sorting machinery protein (VPS)	
	RL	S13_20856724	13	20856724	3.22E-05	0.499	0.074	0.041	p5.00_sc00035_p0180	-12.0 kb	Guanine nucleotide-binding protein subunit gamma (AGC3)	
	LXL		S13_23674227	13	23674227	3.22E-05	0.499	0.074	0.041	p5.00_sc00035_p0078	0	Extracellular ribonuclease (RNAse)
			S13_25522088	13	25522088	3.22E-05	0.499	0.074	0.041	p5.00_sc00128_p0001	0	Imbr1 domain-containing protein (IMBR1)
	S13_24474516		13	24474516	7.14E-05	0.497	0.070	0.067	p5.00_sc00035_p0043	0	Probable ran guanine nucleotide release factor-like (RANGRF)	
Production	Yield BN	S5_41396842	5	41396842	1.E-04	0.132	0.059	0.410	p5.00_sc00003_p0367	+29 kb	Cation h(+) antiporter	
	BW	S10_21597426	10	21597426	3.E-04	0.499	0.054	0.316	p5.00_sc00036_p0097	-8.0 kb	Zinc finger protein 8-like (ZFP)	

* SNP position relative to the candidate gene: SNPs upstream and downstream of candidate genes are specified with "-" and "+", respectively. 0 indicates that SNPs are located within candidate gene.

Figures

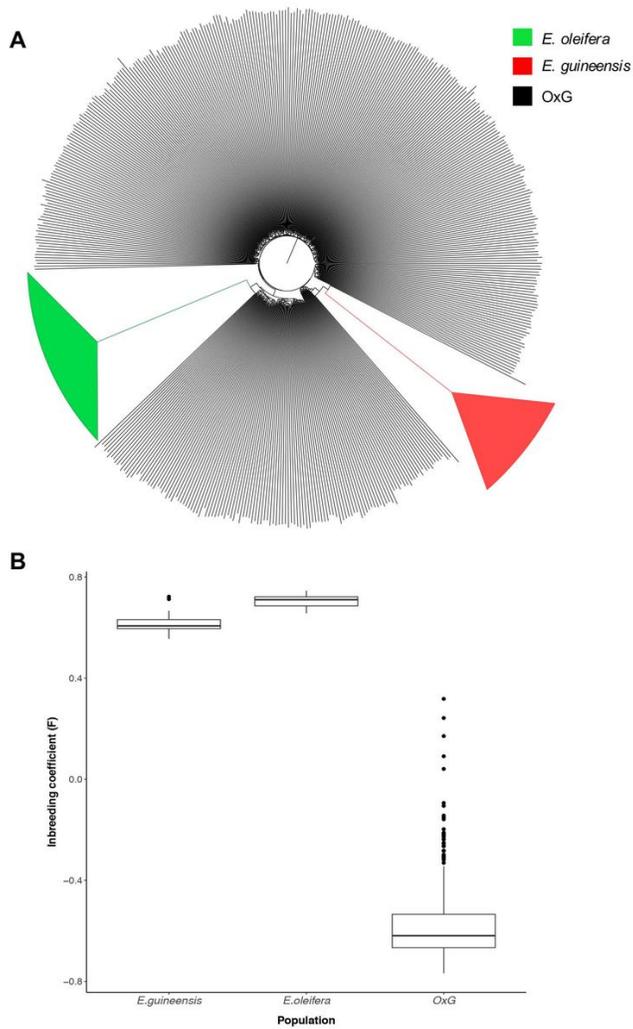


Figure 1

A. Neighbor-Joining tree based on Nei's genetic distance of oil palm clustered into three subgroups, B. Box plots showing the distribution of inbreeding coefficient estimates within each population.

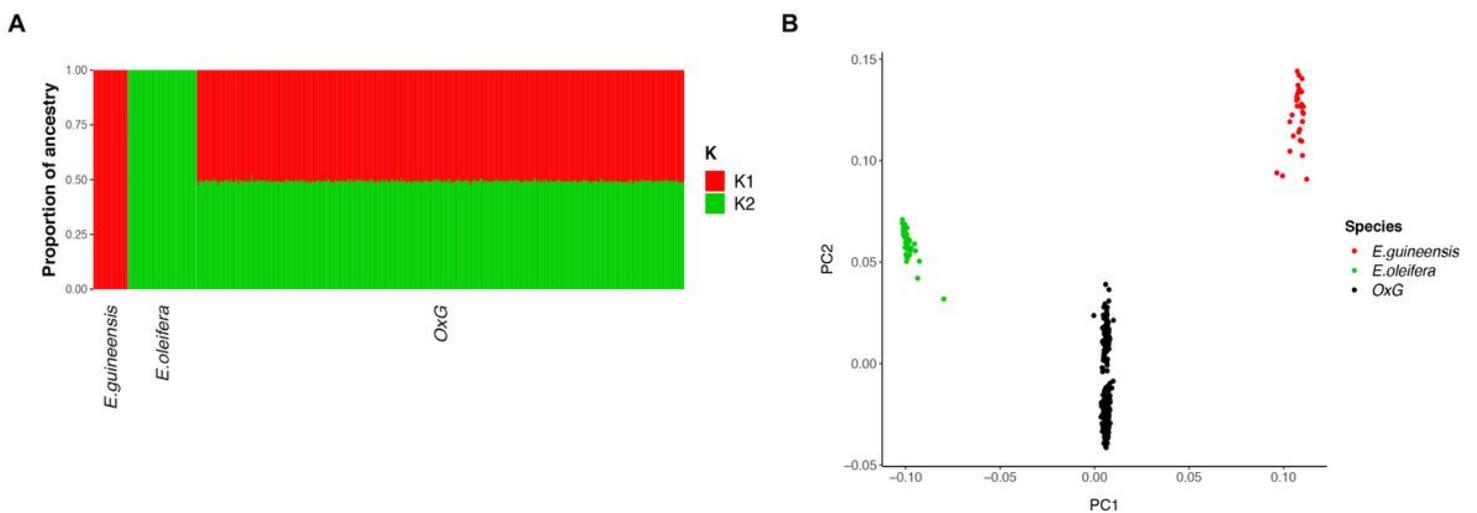


Figure 2

Genetic differentiation of 471 oil palm accessions using 3,776 SNP markers. A. Clustering of samples (k=2) using the supervised mode on Admixture software, B. Principal coordinate analysis (PCA) separating the mapping population into three groups according to their breeding origin.

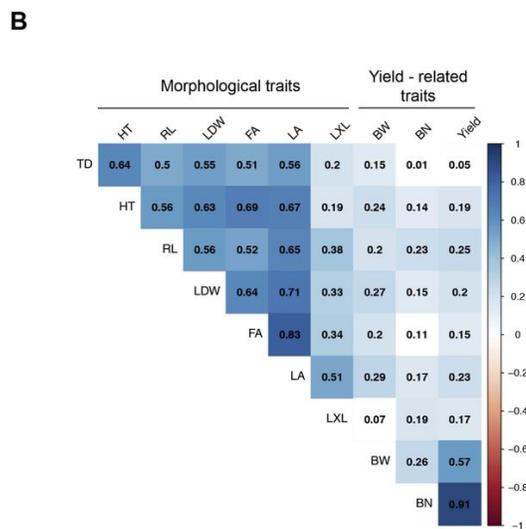
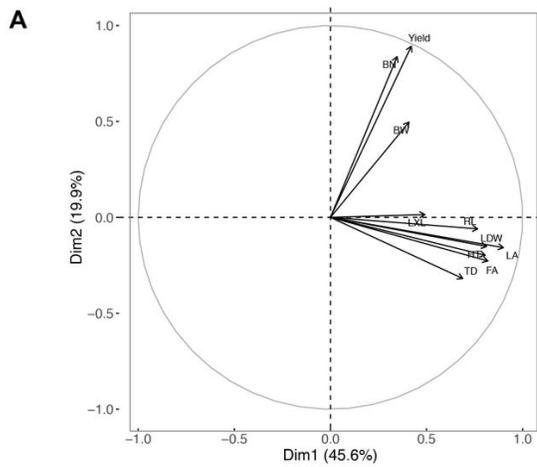


Figure 3

A. Principal component analysis for 10 oil palm morphological and yield-related traits B. Correlation among traits of the whole population. Color boxes indicate significant correlations ($p \leq 0.01$), while white boxes indicate coefficients with a $p \geq 0.01$.

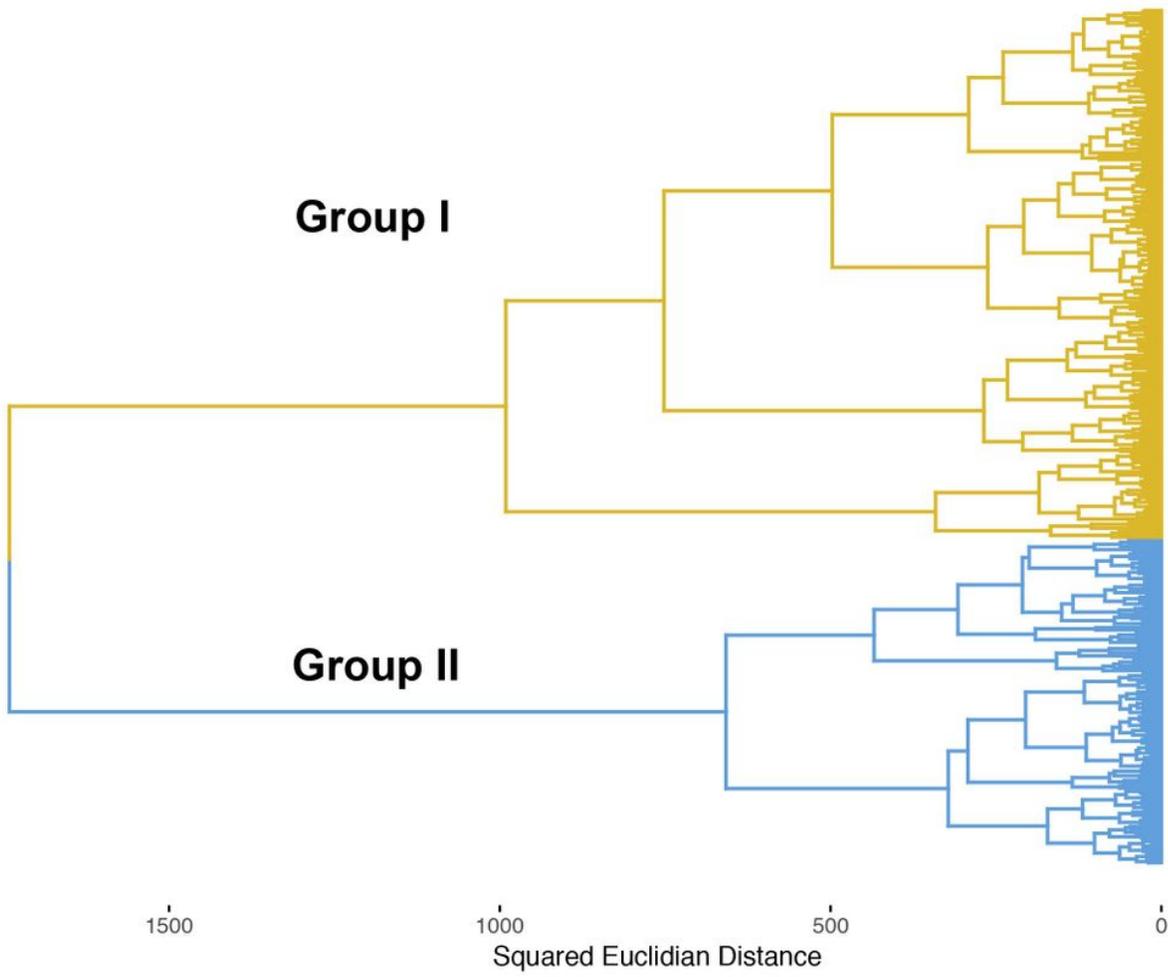


Figure 4

Hierarchical cluster analysis. Samples were clustered using the Ward's method and the squared Euclidean distance.

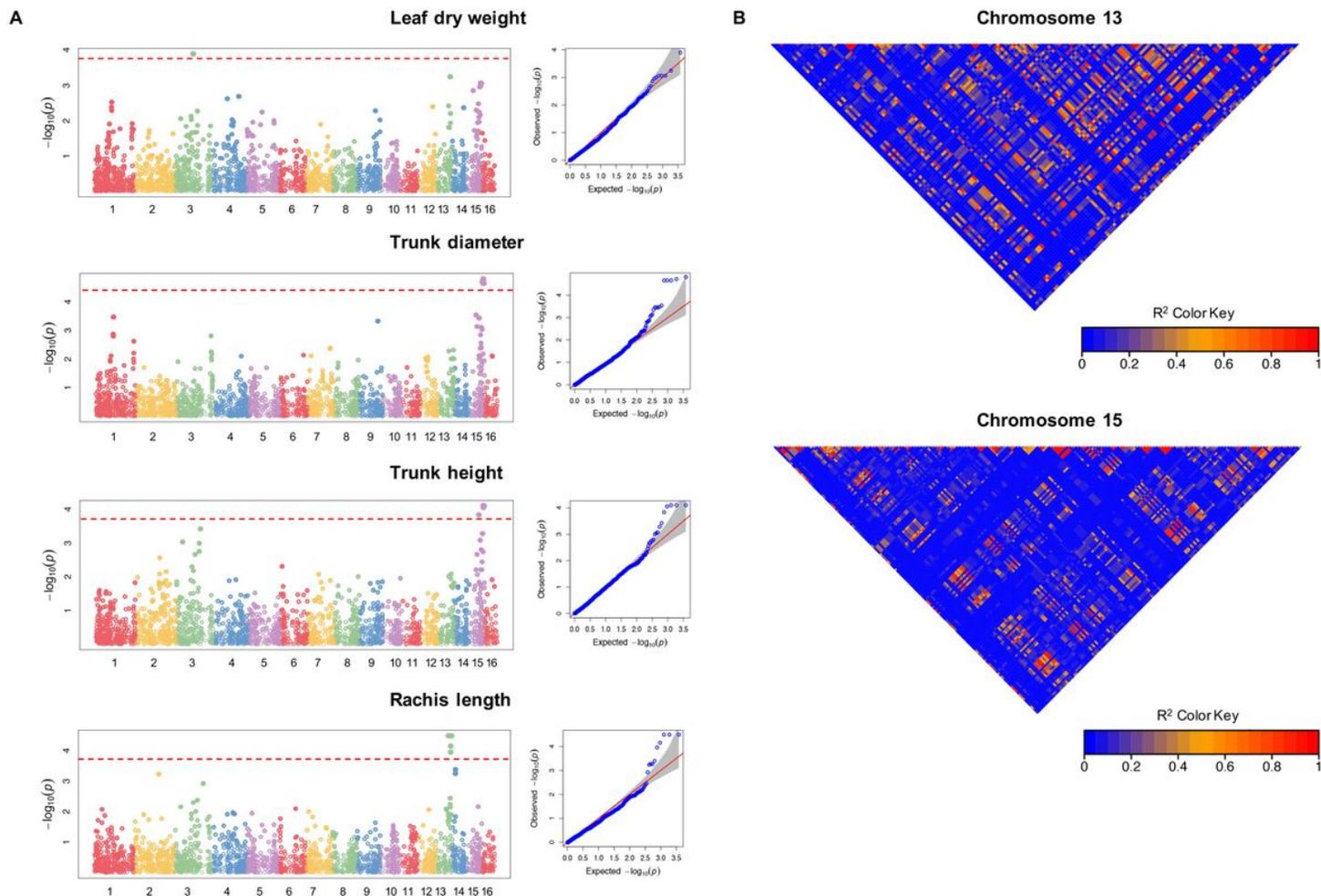


Figure 5

A. Manhattan and Q-Q plots indicating genomic regions associated with leaf dry weight (LDW), trunk diameter (TD), trunk height (HT) and rachis length (RL). The red horizontal line indicates the significant association threshold ($-\log_{10} P \geq 4$), B. Pattern of linkage disequilibrium (LD heatmap) in the chromosomes 13 and 15, which present the most associated SNPs for morphological-related traits.

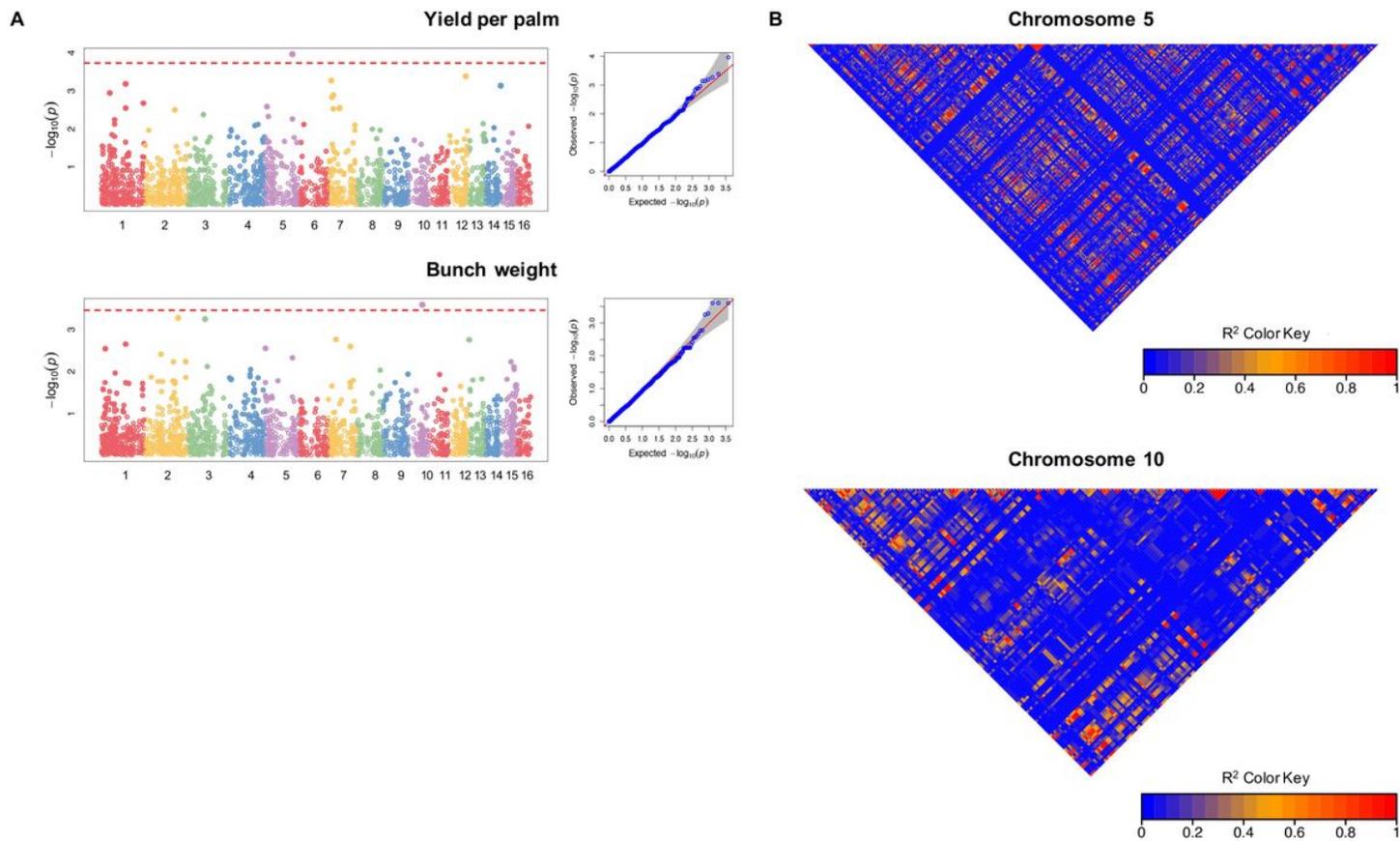


Figure 6

A. Manhattan and Q-Q plots indicating genomic regions associated with yield and bunch weight (BW). The red horizontal line indicates the significant association threshold ($-\log_{10} P \geq 4$). B. Pattern of linkage disequilibrium (LD heatmap) in the chromosomes 5 and 10, which present the most associated SNPs for yield-related traits.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement1.xlsx](#)
- [supplement2.xlsx](#)
- [supplement3.tif](#)
- [supplement3.tif](#)