Int.J.Curr.Res.Aca.Rev.2024; 12(3): 1-13

EXCELLENT PUBLISHERS

International Journal of Current Research and Academic Review

ISSN: 2347-3215 (Online) Volume 12 Number 3 (March-2024)

Journal homepage: http://www.ijcrar.com

doi: https://doi.org/10.20546/ijcrar.2024.1203.001

Genome-Wide Genetic Diversity, Population Structure, and Admixture Analysis in Oil Palm Populations of Cameroon

Mondjeli Constantin[1*], Ndiang Owona Zenabou[2], Sadou Ismaël[1], Sobir Ridwani[3], Muhamad Syukur[3] and Willy Bayuardi Suwarno[3]

[1]*Institute of Agricultural Research for Development, (IRAD), P.O. Box 243 Douala, Cameroon*
[2]*University of Douala, Department of Plant Biology, P.O. Box: 2701 Douala*
[3]*Bogor Agricultural University, Department of Agronomy and Horticulture, Plant Breeding and Biotechnology, P.O. Box 16680 Bogor, Indonesia*

**Corresponding author*

Abstract

Oil palm genetic material bank of Cameroon is interested and well distributed in a lot of oil palm improvement research and producer centers of tropical area of the world. Using the Next Generation Sequencing, we sequenced the genome of 169 wild accessions, 25F1 tested progenies, 09 advanced parents and 01 *oleifera* individual. The assemble genome sequence contained 11,578 scaffolds for a total genome length candidates of 10.034362 Mb. A total of 21,302 nucleotide polymorphism or genetic variations were discovered, from which 20,959 single nucleotide substitutions or SNPs and 343 insertions/deletions or Indels were discovered, indicating the evidence of great diversity of nucleotides variants across the studied genome. The frequency of the single nucleotide substitution observed was 1 SNP in 471 bp. The number of transition point mutation events was more detected compare to the number of transversion, with the ratio of transition/transversion equal to 9.65. The investigation on the genetic population structure determined the best suitable K value at 3, referring the number of the major subpopulations. In addition, the result revealed that most of the individual studied genomes were admixed. Our data provide useful and valuable oil palm resource material for accelerating oil palm genetic improvement breeding program progress.

Article Info

Received: 20 January 2024
Accepted: 28 February 2024
Available Online: 20 March 2024

Keywords

Genome-wide genetic diversity, NSG, polymorphic loci, population structure, admixture.

Introduction

Oil palm (*Elaeis guineensis* Jacq.) is one of the most important oil bearing crops in the world. Palm oil is used in various products ranging from cooking oil and margarine to animal feeds, soaps and cosmetics. A recent escalation in global demand for palm oil is largely driven by its emerging role as a feed stock in biodiesel production (Mekhilef *et al.,* 2011). In terms of yield per unit of planting area, oil palm is the most efficient oil yielding crop, producing roughly four tons of oil per hectare per year, exceeding the production of soybean by nearly tenfold. The rapid ascent of palm oil production in the past few years has eventually led to it over taking soy bean oil as the world's leading vegetable oil. Palm oil currently dominates the global vegetable oil economy, contributing about 36 % of the total world oil production (Gupta, 2011). Oil palm is a diploid monocotyledon with a chromosome complement of 2n =32 and belongs to the genus *Elaeis* and the family Palmae. There are two

commonly known species within the genus *Elaeis*: an economically important *E. guineensis* originating from West Africa and *E. oleifera* native to tropical Central and South America (Soh *et al.,* 2010). Oil palm is an outcrossing species due to asynchronous maturation of male and female inflorescences. Based on a flow cytometric analysis, oil palm has an estimated haploid genome size of about 1800 Mb (Singh *et al.,* 2013). Commercial breeding in oil palm began shortly after the introduction of palm seeds to Southeast Asia. Since then, considerable improvements have been achieved in both yield and quality characters (Corley and Lee, 1992). It has been estimated that 70% of the yield increases in oil palm plantation in the second half of the 19th century was attributed to cultivar improvement (Davidson, 1993). Conventional breeding through recurrent selection has played a crucial role in yield enhancement. Unfortunately, oil palm requires at least 8 to10 years to achieve one breeding cycle. Its long generation time is considered a major obstacle in new cultivar development. The ability to identify individuals possessing desirable traits at a nearly stage will help save time and resources needed to develop superior varieties. One of the quantitative traits in oil palm that contributes directly to increase oil yield is fruit bunch weight (BW). Billotte *et al.,* in 2010 attempted to locate the quantitative trait loci (QT L) affecting BW using a multi-parent linkage map consisting of 251 microsatellite markers. Nine QTL affecting BW were detected using a within-family or an across family model. More recently, Jeennor and Volkaert in 2014 employed a genetic map consisting of 89 microsatellites and 101 single nucleotide polymorphism (SNP) markers constructed from a mapping population of 69 progeny to identify a QTL associated with BW. Another valuable, yet often neglected, characteristic in oil palm is vertical height.

Genetic diversity serves as a way for populations to adapt to changing environments. With more variation, it is more likely that some individuals in a population will possess variations of alleles that are suited for the environment. Those individuals are more likely to survive to produce offspring bearing that allele. The population will continue for more generations because of the success of these individuals (Sahney *et al.,* 2010). It can be measured by the gene diversity representing the proportion of polymorphic loci across the genome; Also by the heterozygosity, which is the fraction of individuals in a population that are heterozygous for a particular locus. In addition, it can be also determined by the nucleotide diversity represented by the extent of nucleotide polymorphisms within a population, and is commonly measured through molecular markers such as micro- and minisatellite sequences, mitochondrial DNA and single-nucleotide polymorphisms (Kawabe *et al.,* 2014). When a population is divided into subpopulations, there is less heterozygosity than there would be if the population was undivided. High heterozygosity means lots of genetic variability. Low heterozygosity means little genetic variability. Often, the observed level of heterozygosity will be compared to what we expect under Hardy-Weinberg equilibrium (HWE). If the observed heterozygosity is lower than expected, we seek to attribute the discrepancy or divergence to forces such as inbreeding. If heterozygosity is higher than expected, we might suspect an isolate-breaking effect means the mixing of two previously isolated populations. Genetic data can be used to infer population structure and assign individuals to groups that often correspond with their self-identified geographical ancestry. Clustering of individuals is correlated with geographic origin or ancestry (Jorde and Wooding, 2004). Population stratification occurs when the study samples comprise multiple groups of individuals who differ systematically in both genetic ancestry and the phenotype under investigation. Two terms are relevant to a review of the estimation of ancestry: local ancestry and global ancestry. Local ancestry estimates the extent to which each person's genome is divided into chromosome segments of definite ancestral origin. Global ancestry estimates the proportion of ancestry from each contributing population, considered as an average over the individual's entire genome (Alexander *et al.,* 2009). Spurious apparent associations would be due to differences in ancestry rather than true association of alleles to phenotypes (Cardon and Palmer, 2003). Admixture occurs when individuals from two or more previously isolated populations interbreed. The previously isolated populations are referred to as ancestral or parental and the newly formed population is referred to as admixed. In non-human species, the same process is often referred to as hybridization. Admixed or hybrid individuals have mixed ancestry. Admixture mapping is a method for capitalizing on recent admixture to correlate ancestry at genetic loci with a phenotype. From the admixture process or events, which should be started after two generations of interbreeding between previously isolated parental populations, chromosomes in admixed individuals are mosaics of ancestry (Shriner, 2013). The advent of DNA-based genetic markers enabled plant breeders to accelerate their breeding programs through the utilization of marker-assisted selections. In the past two decades, several types of molecular markers have been developed in oil pa lm

(restriction fragment length polymorphisms, random amplified polymorphic DNAs, simple sequence repeats) and used in the construction of genetic linkage maps and various phylogenetic studies (Billote *et al.,* 2001; 2005). Recently, attention has been geared toward the use of single nucleotide polymorphisms (SNPs) as genetic markers. The ubiquity of SNPs in eukaryotic genome and their usefulness as genetic markers has been well established over the last decade. SNP markers typically occur at frequencies of one per~100-500 bp in plant genomes, depending on the species, e.g. 1 SNP/ 490 bp in soybean (Choi *et al.,* 2007) and 1 SNP/540 bp in pea (Leonforte *et al.,* 2013). With rapid advancement in sequencing throughput together with an overall decrease in sequencing cost, next generation sequencing technologies have been applied to SNP identification in various plant species (Ganal *et al.,* 2009). However, it remains costly to employ whole-genome sequencing to evaluate multiple individuals in a mapping population, especially for organisms with large genomes such as oil palm (1800 Mb). Reduced representation methods are extremely useful, not only because of their cost reducing aspects, but also because many research questions can be answered with a small set of markers and do not require every base of the genome to be sequenced. There are several techniques employed to reduce genome complexity and to capture only a fraction of the genome to be sequenced. Restriction -site -associated DNA sequencing (RAD-seq) was introduced by Miller *et al.,* in 2007 and adapted to incorporate barcoding for multiplexing by Baird *et al.,* in 2008. More recently, a less complicated method for constructing highly multiplexed reduced representation genotyping-by-sequencing (GBS) libraries was proposed by Elshire *et al.,* in 2011.

GBS is an efficient strategy that can simultaneously detect and score tens of thousands of molecular markers. This technique has successfully been applied to high-density genetic map construction and QTL mapping in several plant species (Poland *et al.,* 2013). Currently available GBS protocols have also been customized to work with multiple sequencing platforms, including the Illumina GAII, Illumina HiSeq, Ion Torrent PGM and Proton (Mascher *et al.,* 2013; Poland *et al.,* 2013). More currently, the Next Generation Sequencing library preparation was performed by Next RAD (Nextera-tagmented reductively-amplified DNA) to provide high significantly unbiased genotype data of loci scattered across the genome, giving high-quality SNP calls on limited available samples of interest (Mondjeli *et al.,* 2024).

Genetic variation is the basis of genetic improvement. The draft genomes of Pisifera form of the species *E. guineensis,* and another species *E. oleifera* oil palm are available. The genome size of oil palm is 1.8 Gb. The genomic study of the oil palm germplasm collection of Cameroon has not been reported yet and the implementation of Next Generation Sequencing library preparation performed by Next RAD has lacking. Considerable contribution efforts oil palm populations have been confirmed in oil palm improvement breeding program of most of the oil palm producer countries. New explorations of wild oil palm population were done recently across the diverse regions of Cameroon. However, information about genome-wide genetic variation is still lacking. The objectives of this study were to determine the genome-wide overview of genetic variation and Stratification through genome-wide discovery of SNP markers. To our knowledge, this is the first attempt to apply the Next RAD technique to identify SNP markers in oil palm.

Materials and Methods

Collection of samples

Two explorations of seven different geographical and ecological regions (forest and savannah area) of Cameroon were recently held in Cameroon by Colombian, Indonesian and Cameroonian researchers to collect and create a useful representative wild oil palm population in their respective oil palm research Centre for germplasm collection issue. A total of 177 wild oil palm accessions were collected and conserved in the germplasm. For this study, 204 oil palm leaflet samples (169 wild accessions, 25 F1 tested progenies, 09 advanced parents and 01 *oleifera* individual) were collected from the Specialized Station for oil palm research La-dibamba of Cameroon and send to Indonesia in the biomolecular laboratory research Centre of PT. Astra Agro Lestari in Pangkalan Bun, Kalimantan.

DNA Extraction

DNA isolation was carried out using the modified Qiagen method in the Tissue Lyser II equipment. Oil palm fresh leaflet blade was cut for about 100 mg into small pieces using a pair of scissors, which was rinsed after each use. The piece of each samples was transferred into 1.2 ml collection tube containing 600 ul of GP1 buffer, introduced also a bead into the same collection tube and seal that tube using collection tube cap. All the

oil palm collection tube samples was arranged into the collection tube holder, which was put into the holder adaptor of the Tissue Lyser II and run the apparatus after previously set the frequency at 30 Hz for 10 minutes.

The running time was appreciated using a timer or chronometer. The collection tube holder was then released from the adaptor and introduced into the water bath at 65oC for 20 minutes. After the first 10 minutes interval, the collection tube holder was removed and inversed by shaking using hands (10 times) and inserted again into the water bath for the last 10 minutes. The incubation of TE buffer bottle was done by introducing it at the time with the collection tube holder into the same water bath for later use. The collection tube holder and the TE buffer bottle was taken out from the water bath and cooled down to room temperature. Each whole mixture was be now transferred into the new sterile collection tube of 2ml microtube. The Chloroform was added at 600 ul into each tube mixture after their introduction in the laminar air flow cabinet. The tubes of this new mixture was then inversed for 10 times. The tubes was then transferred to centrifuge and run at 12000 rpm for 5 minutes. The liquid only was then introduced into new sterilized 2 ml microtube and centrifuged again at 12000 rpm for 5 minutes.

The supernatant was carefully transferred into 2 ml microtube and 600 ul of GP2 was added to mix with the supernatant and inversed the tubes for 10 times. The whole mixture was then transferred into CB3 spin column which was previously inserted into 2 ml collection tube. They were both centrifuged at 12000 rpm for 1 minute. The filtered supernatant was then discarded and the remaining mixture was transferred again into the same GB3 spin column then centrifuged also again at 12000 rpm for 1 minute and discard once more the filtered supernatant.

The GD buffer was added for 500ul then centrifuged the tubes at 12000 rpm for 1 minute. The filtered supernatant was discarded again so that 600 ul of PW buffer was added into the tubes which was centrifuged at 12000 rpm for 1 minute. The filtered supernatant was discarded and this step should be repeated once more by adding 600 ul PW buffer into the same tubes and centrifuged at 12000 rpm for 2 minutes. The CB3 spin column was now transferred into 1.5 ul microtube then 50 ul TE buffer was added into the center of the column and let stand it for 5 minutes on the collection tube holder, then centrifuged the tubes at 12000 rpm for 2 minutes. A total of 204 oil palm fresh leaflet extracted DNA was obtained and the stored into the freezer at - 20^{o}C.

DNA quantification and qualification

DNA concentrations (ng/ul) was determined using the *Nano Drop 2000* spectrophotometer by measuring the DNA absorbance at UV wavelength λ= 260.0 nm and 280.0 nm. The optical density (OD) ratio at 260/280 was calculated to determine the DNA purity, which is related to DNA protein contamination. The electrophoresis process was run during 30 minutes with an electrical current tension of 100 volt. The gel was removed and stained in a solution of ethidium bromide for 5 to 10 minutes, then rinsed with aquades and observed in the UV Transilluminator using a camera.

Genome sequencing, assembly and quality Assessment

High-quality genomic DNA was extracted from 204 leaf samples of wild types and advanced Dura and Tenera palms was prepared following the protocol recommendation of the NGS library preparation of Next RAD for genotype sequencing.

Detection of SNPs and genetic variations in 204 oil palm genome candidates

The detection of SNPs was performed using TASSEL 5.2.33 software (Trait Analysis by Association, Evolution and Linkage), which was able to provide the genome information related to the nucleotide polymorphism, number and proportion of not missing variants, number and proportion of missing variants, number of gametes, number and proportion of not missing gametes, number and proportion of missing gametes, number and proportion of heterozygous loci, and the average minor allele frequency (MAF).

To analyze the genetic structure and relationship between different oil palm genome candidates, Evanno Cross-Entropy method was used to determine the best subpopulation K value of the studied population (Rubinstein, 1997; Rubinstein and Kroese, 2004), while SNiPlaysoftware was used to visualize the subpopulations or clusters with the same similarity across the genome. We performed a Principal Coordinate Analysis (PCoA) to examine the population structure.

Results and Discussion

Genetic diversity in the Oil palm population genome of the study

The genetic diversity is the total number of genetic characteristics in the genetic makeup of a species. It can be measured by the gene diversity representing the proportion of polymorphic loci across the genome, the heterozygosity, and also by the nucleotide diversity represented by the extent of nucleotide polymorphisms or SNPs variants within a population.

SNP discovery and genotyping using Next RAD

It is important to study the patterns and frequencies of mutations in the genome as these mutation events, which include nucleotide substitutions, insertions and deletions, provide the molecular basis of gene and genome evolution. The detection of all the genetic variation presents in a given oil palm genome was included: single nucleotide polymorphisms SNPs, small insertions and deletions (INDELs) and larger forms of structural variation.

By using the next generation sequencing and the oil palm genomic DNA from 204 individuals (169 wild accessions, 25 F1 tested progenies, 09 advanced parents and 01 *oleifera* individual), the result of the study revealed that, the assemble genome sequence of the oil palm population candidate tree contained 11,578 scaffolds for a total genome length candidates of 10.034362 Mb. The raw results were able to display a total of 912,447,003 quality control (QC) passed reads covering length of sequence data, with an average of 4,472,779.43 reads per sample and the mean read depth of 58.96.

From the raw hapmap (Figure 1) read genome sequence, 89.85% of the total sequence bases were able to align unique locations on the reference genome and 10.15% were missing. This current hapmap of the Cameroon oil palm DNA genome can be used as genetic fingerprint markers to identify oil palm population from Cameroon germplasm. In addition, the proportion of heterozygous alleles was 13.51% (Table 2) which is higher than the one reported by Jin *et al.,* (2016) in oil palm elite Dura (1.01%), that in soybean (0.57%) by Lam *et al.,* (2010) and in date palm (0.47%) by Al-Dous *et al.,* (2011). Jin *et al.,* (2016) reported that the genomic candidates of Dura and Pisifera palms used in producing the hybrid Tenera in Southeast Asia showed lowest number of SNPs and genetic diversity. The genomic overall average proportion of the Minor Allele Frequency (MAF), which is the frequency of the less (or least) frequent allele in a given locus and a given population, was estimated for 16,88% of the alleles detected (Table 1).

Single nucleotide polymorphisms (SNP) and insertions/deletions (INDELs) are the most abundant type of DNA sequence polymorphisms and can be theoretically found within every genomic sequence (Rafalski, 2002a, b). The results revealed a total of 21,302 genetic variants, from which 20,959 single nucleotide substitutions and 343 insertions/deletions or INDELS were, discovered (Table 2).

The fraction of INDELS recorded from this study was around 1.61% and this level is lower than the previous estimate (10.4%) in Riju *et al.,* (2007), which utilized oil palm expressed sequence tags (ESTs) available in the NCBI database and Pootakham *et al.,* (2015) who had used genotyping-by-sequencing (GBS) technology to identify putative SNP markers and their association with some agronomic oil palm traits. We focused only on single nucleotide variations and excluded the INDELS and variants involving more than one nucleotideas they were more likely derived from sequencing or alignment errors.

The frequency of the single nucleotide substitution observed was 1 SNP in 471 bp, which is noticeably lower than the previously reported frequency of 1 SNP in 74 bp (Riju *et al.,* 2007) but higher than the one reported, 1 SNP in 665 bp by Pootakham *et al.,* (2015). The frequency observed here was likely to be quite overestimate since these SNPs were identified mainly from the wild oil palm population of the new germplasm collection and additional F1 population derived from progeny test crosses, while SNPs discovered in Riju *et al.,* (2007) were obtained from mining publicly available ESTs derived from several unrelated individuals and those reported by Pootakham *et al.,* (2015) were detected from F1 and F2 progeny.

Moreover, with merely 576 ESTs analyzed, the rate of SNP discovery reported in Riju *et al.,* (2007) may not represent the actual frequency of SNP occurrences across the genome. In this studied oil palm population, the most proportion of the nucleotide polymorphisms (89.15%) detected was related to the change of a purine base to another purine base or pyrimidine base to another pyrimidine base, referring to transitions (G/A, A/G or T/C, C/T).

The two types of genome substitution mutations related to the interchanges of bases were diagnosed. The transversion point mutation events, which are the substitution of a single purine base to pyrimidine base (A/C, C/A, T/A, A/T, G/C, C/G, T/G and G/T) accounted for 9.24% (Figure 2). The most prevalent variation was detected from the point mutation or substitution between two purine bases namely adenine and guanine A/G (44.93%) and the least common type of change was observed on the substitution of pyrimidine base to purine base: cytosine to guanineor their inverse (C/G), representing just 1.73% of total polymorphisms. In the set of SNP calls, the number of transition to the number of transversion for a pair of sequences: transition/transversion ratio was 9.65, which is very high compared to the earlier estimates reported for oil palm, 1.77, 1.55, and 1.67 by Pootakham *et al.,* (2013); Ting *et al.,* (2014) and Pootakham *et al.,* (2015) respectively. A high mean number of alleles and allelic richness imply great allelic or gene diversity which could have been influenced by cross breeding or admixture. Bar charts can be created for individual breeds to show variability in allelic distributions at loci (El-Mousadik and Petit, 1996; Cochard *et al.,* 2009).

Heterozygosity in the oil palm population of Cameroon

The fraction of wild oil palm individuals in the studied population those are heterozygous for a particular locus in the genome is represented in Figure 3. The number of heterozygous genotype loci, distributed across the genome of individual was varied among the studied accessions. The degree of heterogeneity was least for some individuals and greatest to another. The current oil palm population was heterogeneous. Thus, confirming the evidence of nucleotide diversity across the genomes. The oil palm genome of A97_14.18 accession was identified to have more specific heterozygous loci across its genome (Figure3). The result indicated that, although the studied population was wild type oil palm natural population, there is a presence of genetic destruction situation in the current natural population. That is, the gene flow or gene migration which is the transfer of alleles or genes from one population or an individual to another was occurred in their natural environment. The migration into or out of a population may be responsible for a marked great change in allele frequencies (the proportion of members carrying a particular variant of a gene) observed in this study by the heterozygoty proportion of 13.51% (Table 1) and confirmed by the null rejected hypothesis of Hardy-Weinberg equilibrium test also observed in this research. Suggesting that precautions should be taken into account during selection of any oil palm accession candidate with the traits of interest for breeding program purposes.

Missing data in the studied population

In the whole genome sequence polymorphism analysis, the number of missing data should be determined in terms to avoid biases in the genetic parameter estimation (Hellmann *et al.,* 2008; Lynch, 2008; Jiang *et al.,* 2009). In this study, missing number of nucleotide variants across the chromosome candidates was performed. The distribution of the missing genetic variants across the genome of the studied individual oil palm population, revealed that the overall missing nucleotide variants across the genome of individual was ranging from 0 to 0.2 and most of the missing nucleotide variants number was recorded with a frequency of 2%, while the minority of genomic missing data was equal to 0 (Figure 4). The missing variants were relatively found in most of the studied individual genome candidate (Figure 5), suggesting that advices were taken by filtering those missing data from the analyzed genome to control spurious associations.

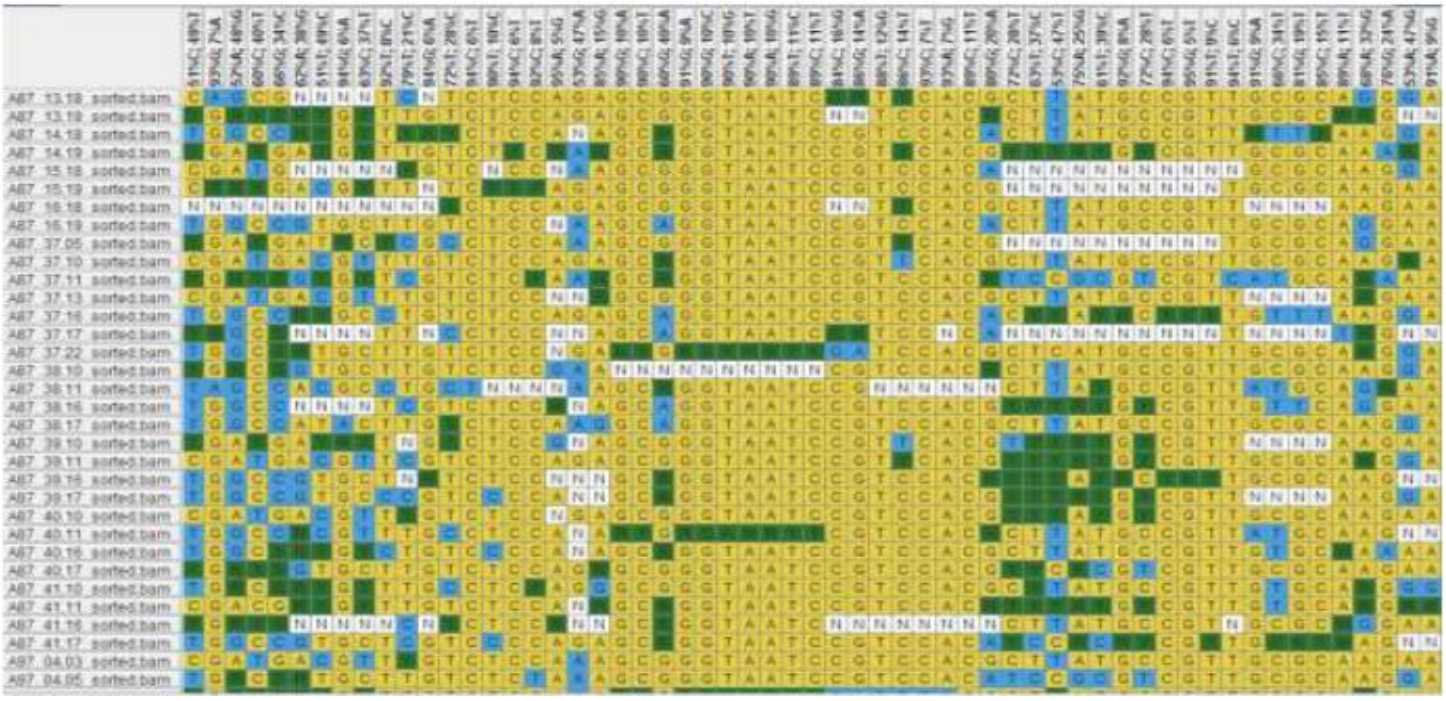
Population Structure of the studied population

The population structure of the studied 204 accessions was classified after excluding the SNPs missing data from the genome candidate. Then, the remaining SNPs were used to construct the population structure using SNipaly program and Evanno cross-entropy plot to determine the best K number of the subpopulation groups (Figure 6). The result from ancestry and candidate genome relationship revealed that the best optimal value of was equal to K= 3 (Figure 7). The investigation on the genetic structure with suitable K value of the studied population revealed that three major genetic clusters or groups were found in this study. The genomic similarity relationship between the reference ancestry and individuals showed that most of studied individuals were formed with at least two extra genomic sets of ancestries (Figure 7). The admixed individuals have mixed ancestries. These results indicate that there is substantial genetic differentiation among individual tested genome and across the genome loci of each candidate. We found that several genomes showed evidence of admixture. These chromosomes are passed on to subsequent generations by inheriting part of alleles or set of chromosome from source population through genetic recombination or point mutations.

Table.1 Overall summary of the studied raw parameter of oil palm genome result from Tassel analysis

State Type	Value
Number of Taxa	204
Number of Sites	21302
Sites x Taxa	4.35E+06
Number Not Missing	3.90E+06
Proportion Not Missing	0.89851
Number Missing	441037
Proportion Missing	0.10149
Number Gametes	8.69E+06
Gametes Not Missing	7.81E+06
Proportion Gametes Not Missing	0.89851
Gametes Missing	882074
Proportion Gametes Missing	0.10149
Number Heterozygous	587136
Proportion Heterozygous	0.13511
Average Minor Allele Frequency	0.16877

Figure.1 Hapmap of SNP genetic fingerprint markers of some major and minor nucleotide variants detected in oil palm population of Cameroon



Note: Nucleotide Codes, (Derived from IUPAC). A A:A; C C:C ; G G:G; T T:T; R A:G; Y C:T; S C:G; W A:T; K G:T; M A:C; + +:+ (insertion); 0 +:- , - -:- (deletion); N Unknown

Table.2 Identified polymorphisms in Cameroon oil palm germplasm

Items	Number of alleles	Percentage (%)
Total number of polymorphisms	21302	
Indels	343	1.61
SNPs	20959	98.39
Transition	18990	89.15
A/G, G/A	9572	44.93
T/C, C/T	9418	44.21
Transversion	1969	9.24
A/C, C/A	449	2.11
A/T, T/A	621	2.92
G/T, G/T	531	2.49
C/G, G/C	368	1.73

Figure.2 Proportion of single nucleotide polymorphism in Cameroon oil palm population

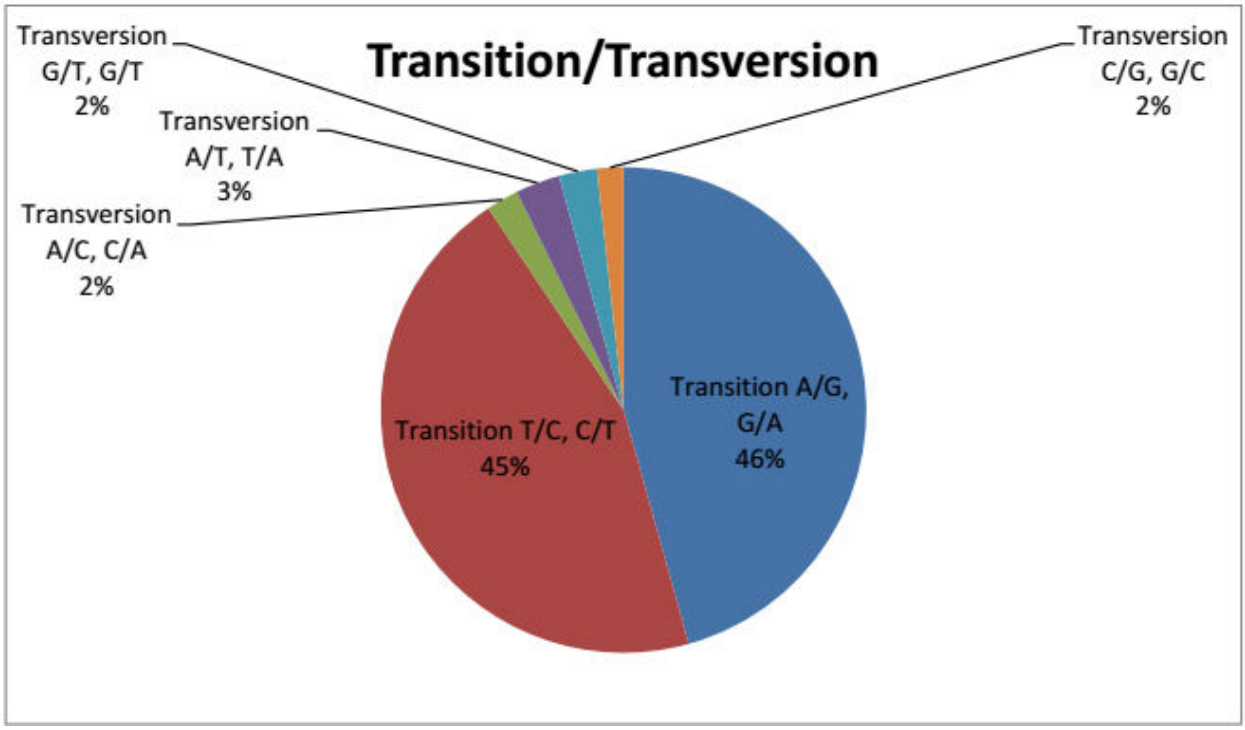


Figure.3 Heterozygosity distribution in the oil palm population of Cameroon

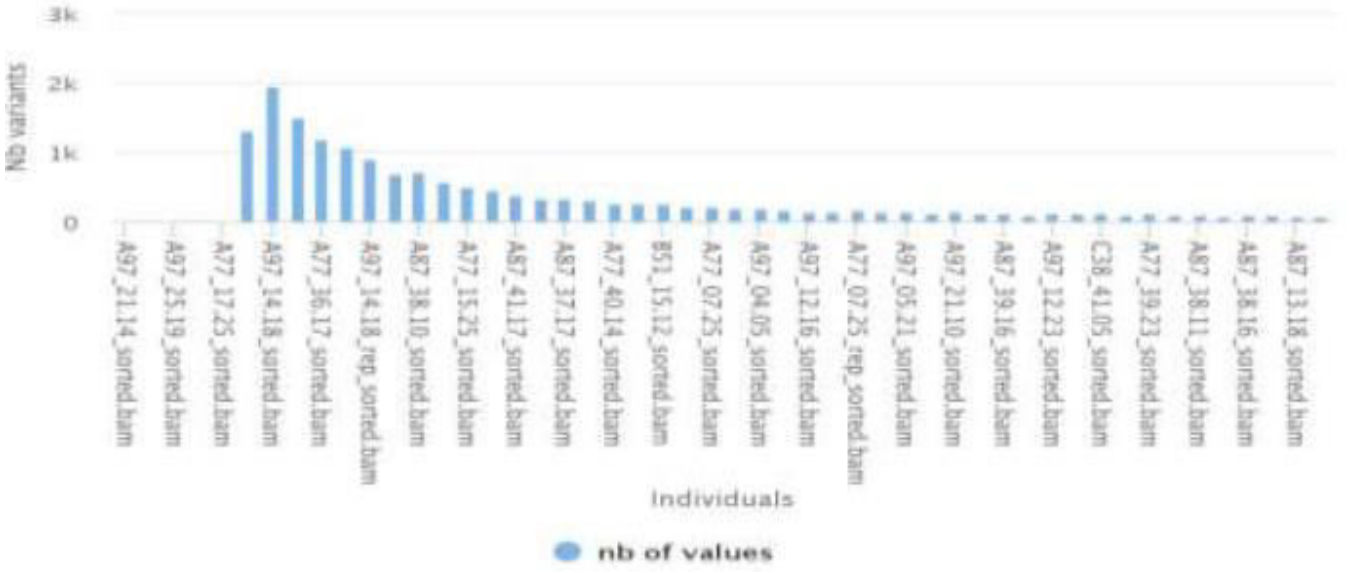


Figure.4 Distribution of missing genetic variants across the overall studied genome of oil palm population of Cameroon

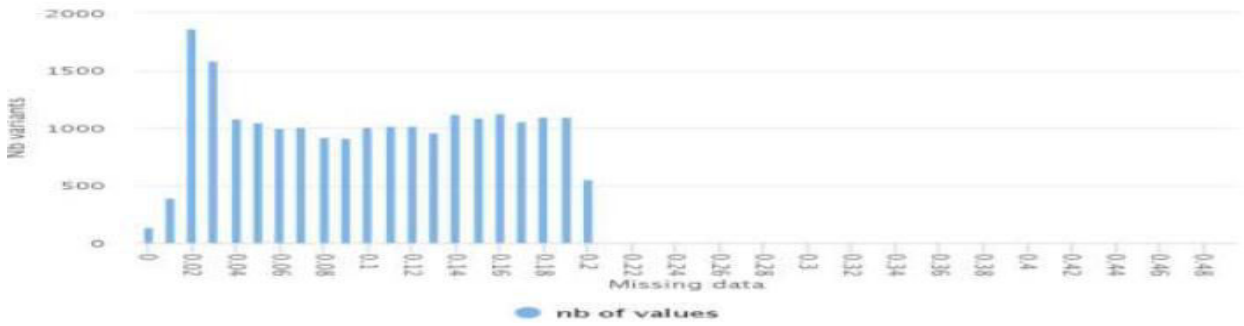


Figure.5 Distribution of number of missing genetic variants per oil palm accession of Cameroon population

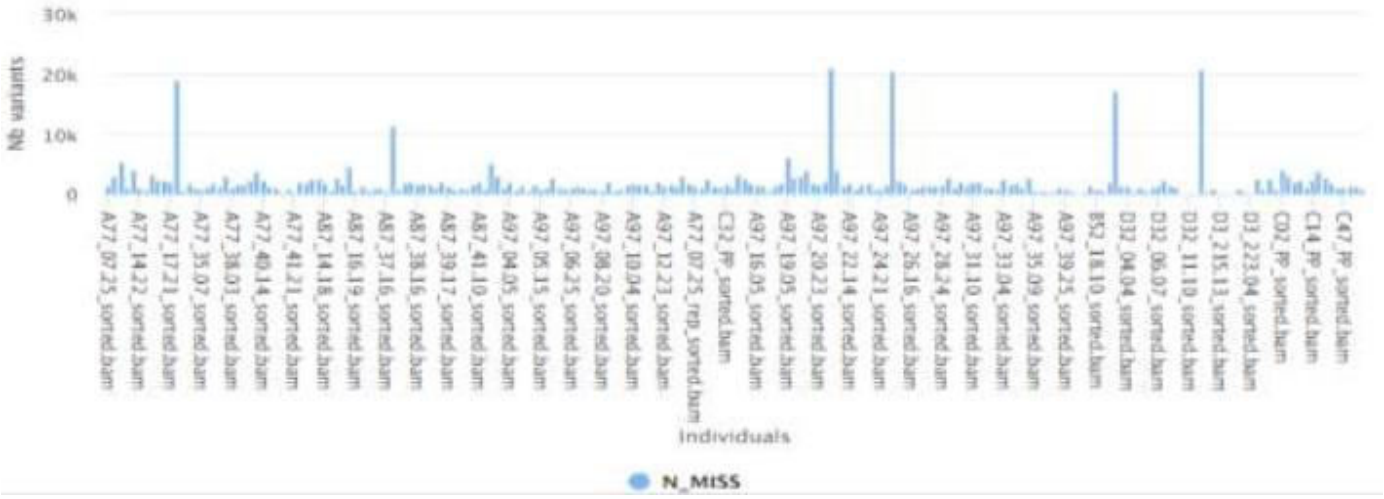


Figure.6 Evanno cross-entropy plot representing the best K number of the subpopulation groups

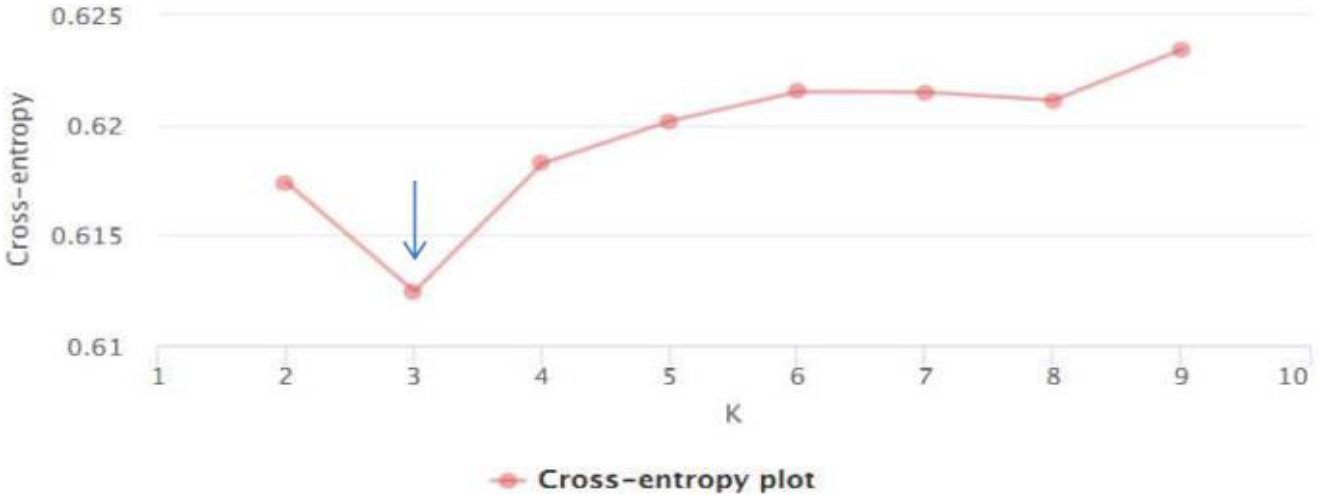


Figure.7 Local Ancestry proportion genome segments in some individuals of the new oil palm germplasm of Cameroon

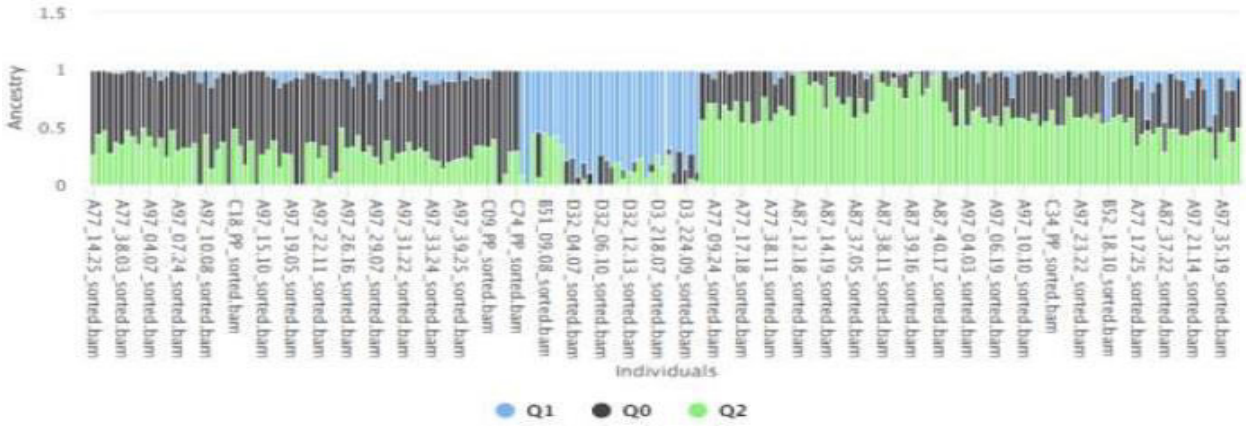
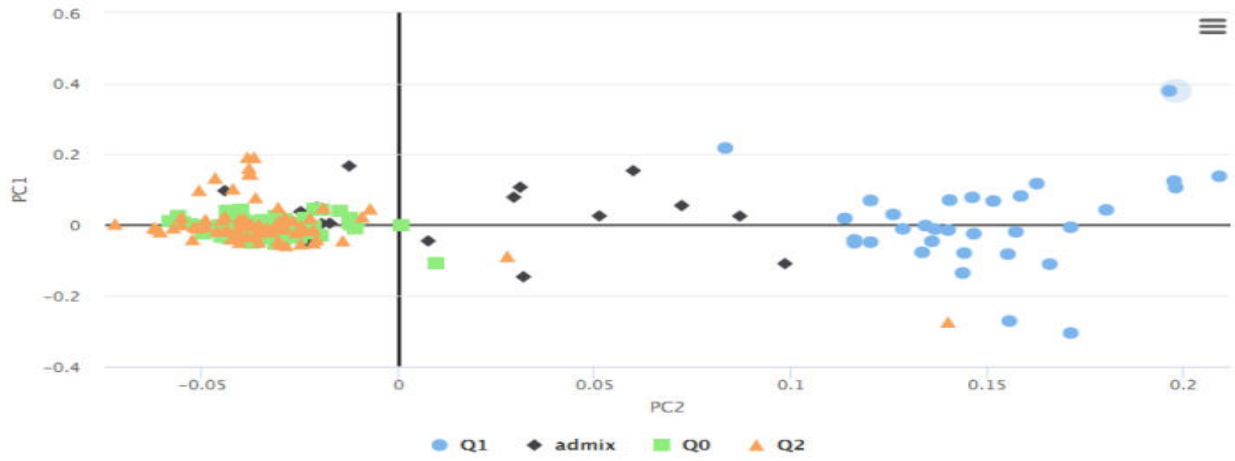


Figure.8 The sub-populations in a population of 204 accessions, suggested by Principal Coordinate Analysis (PCoA) genetic diversity of the new germplasm of Cameroon



The result suggested also the evidence for the wild oil palm individual genome candidate being able to adapt to changing circumstances in their environment. Hellenthal *et al.,* (2014) reported that DNA from the modern-day admixed population resembles a mosaic of small chunks or sets from the source population. The subpopulations Q0 and Q2 were mainly composed of wild accessions from diverse prospected geographical region origins of Cameroon. The subpopulation Q1 was clustered with most of advanced elite Dura and Tenera individuals from progeny test of the new 3rd selection cycle, associated with few accessions.

The principal coordinate analysis (PCoA) was performed to examine this population structure so that the accessions with an admixture region in the genome or SNPs loci can also be visualized. The subpopulation Q1 was quite genomically different to the subpopulations Q0 and Q2. However, Q0 and Q2 were more clustered to each other related to their genome. Most of admixed individuals were relatively similar in terms of genome resemblance SNP loci contents to subpopulation Q2 and Q0. The scattering of individuals of the same subpopulation Q1 and admixture indicate also their relatively genome content differences (Figure 8).

The results revealed that 37.75% of the overall population were composed of cluster Q2 in which 77 accessions were identified, 37.25% was accounted by cluster Q0 with 76 accessions, while 15,69% was recorded by cluster Q1 with 32 accessions and 9.31% by the admixture group where 19 accessions were found. Shriner in 2013 reported that admixture occurs when individuals from two or more previously isolated populations interbreed. The previously isolated populations are referred to as ancestral or parental and the newly formed population is referred to as admixed. In oil palm, the admixture of investigated trees was reported by Jin in 2016.

Conclusion

The Next Generation Sequencing using Next RAD method successfully discovered a great Genome-wide diversity of single nucleotide polymorphism across the studied genomes. The results revealed that the studied oil palm population was divided into three major groups, in which most of individuals were detected as admixture of chromosome segments, indicating the chronological mutation or recombination events that occur generation to generation into the genome of these wild oil palm populations in their natural environment. Our data provide useful and valuable oil palm resource material to accelerate oil palm genetic improvement breeding program progress.

Acknowledgements

This work was supported financially by PT ASTRA AGRO Lestari Tbk Indonesia and Institute of Agricultural Research for Development (IRAD) / Specialized Station for Oil Palm Research of La Dibamba Cameroon. Bogor Agricultural University (IPB) of Indonesia provided supervision of this study.

Disclosure of Conflict of Interest

No conflict of interest to be disclosed.

References

Al-Dous, E. K., George, B., Al-Mahmoud, M. E., *et al.,* 2011, De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). Nat.Biotechnol. 29, 521-7. https://doi.org/10.1038/nbt.1860

Alexander D. H., November J., Lange K. 2009. Fast Model-based estimation of ancestry in unrelated individuals. Genome Research 19(9):1655-64. https://doi.org/10.1101/gr.094052.109.

Baird N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, E. A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers, PLoS ONE 3 (10) e3376. https://doi.org/10.1371/journal.pone.0003376

Billotte N., A. M. Risterucci, E. Barcelos, J. L. Noyer, P. Amblard, F. C. Baurens. 2001. Development, characterisation, and across-taxa utility of oil palm (*Elaeis guineensis* Jacq.) microsatellite markers, Genome 44: 413–425. https://doi.org/10.1139/g01-017

Billotte N., N. Marseillac, A. M. Risterucci, B. Adon, P. Brottier, F. C. Baurens, R. Singh, A. Herran, H. Asmady, C. Billot, *et al.,* 2005. Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.), Theor. Appl. Genet. 110 (4) 754–765. https://doi.org/10.1007/s00122-004-1901-8

Billotte N., M. F. Jourjon, N. Marseillac, A. Berger, A. Flori, H. Asmady, B. Adon, R. Singh, B. Nouy, F. Potier. 2010. QTL detection by multi-parent linkage mapping in oil palm (*Elaeis guineensis* Jacq.), Theor. Appl. Genet. 120 (8) 1673–1687. https://doi.org/10.1007/s00122-010-1284-y.

Cardon L R, Palmer L J. 2003. Population stratification and spurious allelic association. Lancet. 361: 598–604. https://doi.org/10.1016/S0140-6736(03)12520-2

Choi I. Y., D. L. Hyten, L. K. Matukumalli, Q. Song, J. M. Chaky, C. V. Quigley, K. Chase, K. G. Lark, R. S. Reiter, M. S. Yoon. 2007. A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis, Genetics 176 (1) 685–696. https://doi.org/10.1534/genetics.107.070821

Cochard B., Benjamin Adon. Samah Rekima, Norbert Billotte. Rochesmier de Chenon, Anatole Koutou, Bruno Nouy, Alphonse Omoré, Abdul Razak Purba, Jean-Christophe Glazsmann, Jean-Louis Noyer. 2009. Geographic and genetic structure of African oil palm diversity suggests new approaches to breeding. Tree Genetics & Genomes. Volume 5, Issue 3, pp 493–504 https://doi.org/10.1007/s11295-009-0203-3

Corley R. H. V., C. H. Lee. 1992. The physiological basis for genetic improvement of oil palm in Malaysia, Euphytica 60 179–184. https://doi.org/10.1007/BF00039396

Davidson L. 1993. Management for efficient cost-effective and productive oil palm plantations, PORIM International Palm Oil Conference, Palm Oil Research Institute of Malaysia, Kuala Lumpur, pp. 153–167.

El Mousadik A., R. J. Petit. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. 1996. Theoretical and Applied Genetics.Volume 92, Issue 7, pp 832–839 https://doi.org/10.1007/BF00221895

Elshire R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler, S. E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (G B S) approach for high diversity species, PLoS ONE 6 (5) e19379. https://doi.org/10.1371/journal.pone.0019379

Ganal M., T. Altmann, M. Roder. 2009. SNP identification in crop plants. Curr. Opin. Plant Biol. 12: 211–217. https://doi.org/10.1016/j.pbi.2008.12.009

Gupta S. K. 2011. Technological Innovations In Major World Oil Crops, Volume 1: Breeding, Springer https://doi.org/10.1007/978-1-4614-0356-2

Hellenthal G., George B. J. Busby, Gavin Band, James F. Wilson, Cristian Capelli, Daniel Falush, and Simon Myers. 2014. A genetic atlas of human admixture history**.** Science. 14; 343(6172): 747–751. https://doi.org/10.1126/science.1243518

Hellmann I., Mang Y., Gu Z., Li P., Francisco M., *et al.,* 2008. Population genetic analysis of shotgun assemblies of genomic sequences from multiple individuals. Genome Res. 18: 1020–1029. https://doi.org/10.1101/gr.074187.107

Jeennor and H. Volkaert. 2014. 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 10 (1) 1–14. https://doi.org/10.1007/s11295-013-0655-3

Jiang R., Tavaré S., Marjoram P. 2009. Population genetic inference from resequencing data. Genetics181: 187–197. https://doi.org/10.1534/genetics.107.080630.

Jin J., May Lee, Bin Bai, Yanwei Sun, Jing Qu, Rahmadsyah, Yuzer Alfiko, Chin Huat Lim, Antonius Suwanto, Maria Sugiharti, Limsoon Wong, Jian Ye, Nam-Hai Chua and Gen Hua Yue. 2016. Draft genome sequence of an elite Dura palm and whole-genome patterns of DNA variation in oil palm. Dna Research. 23(6), 527–533. https://doi.org/10.1093/dnares/dsw036

Jorde and Wooding. 2004. Genetic variation, classification and 'race. Nature Genetics36 (11s): S28–33. https://doi.org/10.1038/ng1435

Kawabe, K.; Worawut, R.; Taura, S.; Shimogiri, T.; Nishida, T.; Okamoto, S. 2014. Genetic Diversity of mtDNA D-loop Polymorphisms in Laotian Native Fowl Populations. Asian-Australasian Journal of Animal Sciences. 27 (1): 19–23. https://doi.org/10.5713/ajas.2013.13443

Lam, H. M., Xu, X., Liu, X., *et al.,* 2010, Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat. Genet., 42, 1053–9. https://doi.org/10.1038/ng.715

Leonforte A., S. Sudheesh, N. Cogan, P. Salisbury, M. Nicolas, M. Materne, J. Forster, S. Kaur. 2013. SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.), BMC Plant Biol. 13 (1) 1–14. https://doi.org/10.1186/1471-2229-13-161

Lynch M., 2008. Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. Mol. Biol. Evol. 25: 2409–2419. https://doi.org/10.1093/molbev/msn185

Mascher M., S. Wu, P. S. Amand, N. Stein, J. Polan. 2013. Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley, PLoS ONE 8 (10) (2013) e76925. https://doi.org/10.1371/journal.pone.0076925

Mekhilef S., S. Siga, R. Saidur. 2011. A review on palm oil biodiesel as a source of renewable fuel, Renew. Sust. Energ. Rev. 15 (4) 1937–1949. https://doi.org/10.1016/j.rser.2010.12.012

Miller M. R., J. P. Dunham, A. Amores, W. A. Cresko, E. A. Johnson. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers, Genome Res. 17 (2) 240–248. https://doi.org/10.1101/gr.5681207

Mondjeli C., Ardha A., Sadou I., Sobir R., Ajambang W. N., Sobda G., Muhamad S., Ebouelle N. A. B., Willy B. S. 2024. Genome wide association study reveals key genic regions controlling oil palm vegetative characters in wild population of Cameroon. Global Journal of Engineering and Technology Advances, 18(02), 042–048. https://doi.org/10.30574/gjeta.2024.18.2.0025

Poland J. A., P. J. Brown, M. E. Sorrells, J. L. Jannink. 2013. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach, PLoS One 7 (2) e32253. https://doi.org/10.1371/journal.pone.0032253

Pootakham W., P. Uthaipaisanwong, D. Sangsrakru, T. Yoocha, S. Tragoonrung, S. Tangphatsornruang. 2013. Development and characterization of single nucleotide polymorphism markers from 454 transcriptome sequences in oil palm (*Elaeis guineensis*), Plant Breed. 132 711–717.

Pootakham W., Ruang-Areerate P., Jomchai N., Sonthirod C., Sangsrakru D., Yoocha T., Theerawattanasuk K., Nirapathpongporn K., Romruensukharom P., Tragoonrung S., Tangphatsornruang S. 2015. Construction of a high-density integrated genetic linkage map of rubber tree (*Hevea brasiliensis*) using genotyping-by-sequencing (GBS). Frontiers in Plant Science.; 6: 367. https://doi.org/10.3389/fpls.2015.00367

Rafalski A. 2002a. Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biology 5(2):94-100. https://doi.org/10.1016/s1369-5266(02)00240-6

Rafalski J A. 2002b. Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Science. 162(3):329-333. https://doi.org/10.1016/S0168-9452(01)00587-8

Riju A., A. Chandrasekar, V. Arunachalam. 2007. Mining for single nucleotide polymorphisms and insertions/deletions in expressed sequence tag libraries of oil palm, Bioinformation 2 (4) 128–131. https://doi.org/10.6026/97320630002128

Rubinstein, R. Y. 1997. Optimization of Computer simulation Models with Rare Events, *European Journal of Operations Research*, 99, 89–112. https://doi.org/10.1016/S0377-2217(96)00385-2

Rubinstein, R. Y. and Kroese, D. P. 2004. The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning, Springer-Verlag, New York. https://doi.org/10.1007/978-1-4757-4321-0

Sahney, S.; Benton, M. J. & Ferry, P. A. 2010. Links between global taxonomic diversity, ecological diversity and the expansion of vertebrates on land. Biology Letters. 6 (4): 544–547. https://doi.org/10.1098/rsbl.2009.1024

Shriner D. 2013. Overview of Admixture Mapping. CurrProtoc Hum Genet. Chapter: Unit1.23. https://doi.org/10.1002/0471142905.hg0123s76

Singh R., M. Ong-Abdullah, E.T. Low, M. A. Manaf, R. Rosli, R. Nookiah, L. C. Ooi, S. E. Ooi, K. L. Chan, M. A. Halim. 2013. Oil palm genome sequence reveals divergence of interfertile species in old and new worlds, Nature 500 (7462) 335–339. https://doi.org/10.1038/nature12309

Soh A. C., G. Wong, T. Y. Hor, C. C. Tan, P. S. Chew. 2010. Oil palm genetic improvement, Plant Breeding Reviews, John Wiley & Sons, Inc., pp. 165–219. https://doi.org/10.1002/9780470650202.ch4

Ting N. C., J. Jansen, S. Mayes, F. Massawe, R. Sambanthamurthi, O. L. Cheng-Li, C. W. Chin, X. Arulando o, T. Y. Seng, S. S. Alwee. 2014. High density SNP and SSR- based genetic maps of two independent oil palm hybrids, Genomics 15 (1). https://doi.org/10.1186/1471-2164-15-309

How to cite this article:

Mondjeli Constantin, Ndiang Owona Zenabou, Sadou Ismaël, Sobir Ridwani, Muhamad Syukur and Willy Bayuardi Suwarno. 2024. Genome-Wide Genetic Diversity, Population Structure, and Admixture Analysis in Oil Palm Populations of Cameroon. *Int.J.Curr.Res.Aca.Rev.* 12(3), 1-13.
doi: https://doi.org/10.20546/ijcrar.2024.1203.001