

Genome wide association study reveals key genic regions controlling oil palm vegetative characters in wild population of Cameroon

Mondjeli Constantin ^{1,*}, Ardha Apriyanto ², Sadou Ismaël ¹, Sobir Ridwani ³, Ajambang Walter Nchu ¹, Sobda Gonne ¹, Muhamad Syukur ³, Ebouelle Ngothy Arsene Belmont ¹ and Willy Bayuardi Suwarno ³

¹ Department of Crop Production, Institute of Agricultural Research for Development, (IRAD), P.O. Box 243 Douala, Cameroon.

² Department of Molecular Biology, PT Astra Agro Lestari Tbk, P.O. Box 13930 Jakarta Timur, Indonesia.

³ Department of Agronomy and Horticulture, Bogor Agricultural University, Plant Breeding and Biotechnology, P.O. Box 16680 Bogor, Indonesia.

Global Journal of Engineering and Technology Advances, 2024, 18(02), 042-048

Publication history: Received on 30 December 2024; revised on 07 February 2024; accepted on 10 February 2024

Article DOI: <https://doi.org/10.30574/gjeta.2024.18.2.0025>

Abstract

Cameroon's oil palm germplasm is contributing a lot towards the improvement of genetic variation in the base oil palm population used in breeding programs around the world. There is a new oil palm germplasm collection at Dibamba constituted of 169 wild accessions but whose genomic study was not implemented before now. Genome-wide-association study is considered as a powerful tool to identify genomic specific allele variants controlling the expression of important agronomic traits in crops. This approach could potentially accelerate varietal improvement in plant breeding programs. The objective of this study was to identify genic regions across the investigated oil palm genomes which are conferred to control natural variation for oil palm vegetative traits such as leaflet width and leaf area. The allele and genotype frequencies were significantly detected to be in Hardy-Weinberg equilibrium ($P < 0.01$) for these vegetative traits. Across the genome, important numbers of single nucleotide polymorphisms were associated with the oil palm leaflet width and leaf area characters. This suggests that the considered regions may contain genes controlling the phenotype variation expression of the trait of interest and should be useful under positive selection in subsequent breeding of the oil palm.

Keywords: Genome wide association; Genic regions; Hardy-Weinberg equilibrium; SNP

1. Introduction

Oil palm is a perennial tropical palm tree which has become the world's most efficient and important oil yielding crop by ranking first among vegetable oilseeds followed by soybean. It belongs to the Arecaceae family, and is a member of the species *Elaeis guineensis* which originated from tropical western and central Africa. It is allogamous and propagated via seeds. Oil palm has a large diploid genome of 1.8 Gb distributed in 32 chromosomes (Singh *et al.*, 2013). Two vegetable oils, crude palm oil (CPO) and palm kernel oil (PKO), are extracted from the mesocarp pulp and kernel of the fruit respectively. The commercial hybrid tree can produce thousands of potential fruits ranging from 3000 to 3500 in a compact bunch. In terms of oil yield per unit of planting area, by producing approximately four tons of oil per hectare per year, oil palm exceeds the production of soybean by nearly ten-fold. Palm oil is used in various products ranging from cooking oil and margarine to animal feeds, soaps and cosmetics. A recent high global demand for palm oil is largely driven by its emerging role as a feed stock in biodiesel production (Mekhilef *et al.*, 2011).

* Corresponding author: Mondjeli Constantin

Oil palm germplasm from Cameroon is contributing a lot to the improvement of genetic variation of the base oil palm population in various breeding programs around the world. High genetic diversity and production potential of Cameroon's oil palm genotypes were also reported (Alvarado *et al.*, 2010; Mondjeli *et al.*, 2017). There now exists a new oil palm germplasm collection in Cameroon though not yet characterized through genomic studies. This paper presents part of work done in this wise using Genome-Wide-Association studies (GWAS). In fact, GWAS are studies that search for a statistical association between a phenotype and a particular allele by screening loci most commonly by genotyping Single Nucleotide Polymorphisms (SNPs) across the entire genome. It is one of the tools of Marker-assisted selection that can play a pivotal role by facilitating or accelerating efficient breeding for agronomically important characters in oil palm breeding programs. The achievement of reference genome sequences for many important crops and model plants remain a great challenge for breeders because they provide plant genomic contribution potentials in a tremendously accelerated strategy for crop improvement and plant breeding programs (Paterson *et al.*, 2005).

Maximizing the use of genomic data for crop improvement is of fundamental importance if we are to continue increasing crop production in the face of growing human populations and changing climates while minimizing the environmental impact of agricultural activity. High-density genetic markers are being used in GWASs and can also be exploited for genomic selection (Morrell *et al.*, 2011). Most studies on association mapping or linkage disequilibrium (LD) in oil palm reported to date are based on microsatellite markers or small numbers of single nucleotide polymorphisms (SNPs) covering one or only a few chromosomes (Billote *et al.* 2005, 2010; Jeennor and Volkaert, 2014; Ajambang *et al.*, 2016). Genome-wide association studies (GWAS) was mainly reported in maize (Flint-Garcia *et al.*, 2005; Yan *et al.* 2010; Prasanna *et al.*, 2010; Ganai *et al.*, 2011; Zila *et al.*, 2013) and rice (Goff *et al.*, 2002; Yang *et al.* 2014).

Despite the considerable GWAS efforts that have been made in plants (Singh *et al.*, 2013; Mariette *et al.*, 2016; Mangin *et al.*, 2012), information about Hardy-Weinberg equilibrium and genome-wide association mapping are still lacking in oil palm. The objective of this study was to investigate on particular allelic or genic regions of the whole genome sequence associated to oil palm leaflet width and leaf area characters (agronomic traits) from Cameroon's wild oil palm population by screening genome loci genotyped by NextRAD technique. Our data will provide a valuable resource to accelerate oil palm genetic improvement and to investigate mechanisms underlying phenotypic variations of important traits in oil palms.

2. Materials and Methods

2.1. High throughput Genotyping using NextRAD

A total of 169 wild oil palm leaflet accession samples were brought from Cameroon to Indonesia for DNA extraction, quantification and qualification processes. Genomic DNA samples were packing and sent to SNPsaurus, LLC (Oregon, USA) for NextRAD genotyping service (www.snpsaurus.com). The nextRAD method uses selective PCR primers to amplify genomic loci consistently between samples. Genomic DNA (10 ng or less depending upon extraction yield) was first fragmented with Nextera reagent (Illumina, Inc), which also ligates short adapter sequences to the ends of the fragments. Those fragments were amplified using selective primers. Then, those specific amplified fragments were sequenced by Illumina (Illumina, Inc), (150 bp) Paired end reads with 20X coverage. The mapping of this sequencing result was conducted using available Oil palm genome draft (Singh *et al.*, 2013). The identified variants were then filtered by removing loci that had more than the expected maximum of two alleles and those that were present in less than 10% of all samples. All of missing variant calling data from each sample was imputed using LD KNNi method (Money *et al.*, 2015) with TASSEL Software (Bradbury *et al.*, 2007).

2.2. Phenotypic data compilation and genome-wide-association study (GWAS)

Interesting oil palm vegetative characters such as leaflet width and leaf area were observed and collected from individual palm trees by the plant breeding department at CEREPAH IRAD, Cameroon. All individual palms were phenotyped to produce a reliable data for analysis as per standard industry practice with modifications.

The single SNP analysis using the SNPs exact test for Hardy-Weinberg equilibrium was implemented on approximately 6000 SNP markers of the studied individuals in order to evaluate the allele frequencies and genotype frequencies from the wild nature population of the germplasm collection.

Association analysis of genotypic and phenotypic data was conducted based on the Enriched Compressed Mixed Linear Model (ECMLM) with P3D analysis in the R package GAPIT (Lipka *et al.*, 2012). We accounted for the genetic sub-structure resulting from cryptic relatedness by including a kinship matrix as a random effect in the compressed MLM

method. The whole-genome significance cut-offs were fixed at $p \leq 10^{-3}$, and $p \leq 10^{-5}$, based on a Bonferroni correction method. The Quantile-Quantile plots and Manhattan plots were then constructed using the R package GAPIT.

3. Results

3.1. Hardy-Weinberg equilibrium of the studied oil palm population

The single SNP analysis using the SNPs exact test for Hardy-Weinberg equilibrium revealed that for all the studied individuals of the germplasm collection from Cameroon, most of the allele frequencies and genotype frequencies were detected to be highly significant ($p < 0.001$). In general, the genotype and allele frequency varied along the studied genome chromosomes. The homozygous dominant genotype frequency and the first allele frequency were highly detected on all the SNP genome compare to their respective heterozygous genotype frequency, recessive genotype frequency and alternative allele frequency (Table 1).

Table 1 Homozygous dominant, recessive and heterozygous genotype frequencies in some oil palm genome obtained from Hardy-Weinberg equilibrium SNP Exact test of wild oil palm population of Cameroon

Chromosome number / SNP Position	N11 (homozygous dominant genotype frequency)	N12 (heterozygous genotype frequency)	N22 (homozygous recessive genotype frequency)	N1 (first allele frequency)	N2 (Alternative allele frequency)	P-value
X1.25550629.SNV	137	16	9	290	34	<0.0001
X2.5848956.SNV	80	38	51	198	140	<0.0001
X3.401952.SNV	54	63	43	171	149	0.011
X4.1059856.SNV	97	17	67	211	151	<0.0001
X5.7001627.SNV	150	9	8	309	25	<0.0001
X6.7525983.SNV	129	25	9	283	43	0.00025
X7.1289600.SNV	99	32	31	230	94	<0.0001
X8.13879587.SNV	113	30	15	256	60	<0.0001
X9.16022531.SNV	67	34	66	168	166	<0.0001
X10.7110638.SNV	78	46	46	202	138	<0.0001
X11.23242805.SNV	133	16	6	282	28	0.00028
X12.4572801.SNV	139	16	6	294	28	0.00022
X13.11763236.SNV	145	6	10	296	26	<0.0001
X14.13904463.SNV	111	35	24	257	83	<0.0001
X15.13471149.SNV	78	58	40	214	138	<0.0001
X16.13787189.SNV	159	11	9	329	29	<0.0001

3.2. Genomic association mapping

The statistical association between oil palm leaflet width and leaf area traits, and particular alleles were investigated by screening the different loci across the entire genome of studied individuals. The result from the Enriched Compressed Mixed Linear Model (ECMLM) revealed that six single nucleotide polymorphism (SNP) markers were significantly associated with oil palm leaflet wild character. These SNP markers were found in chromosome 1, 4, 8 and 16 of the genome. However, the most significant SNP marker for this trait was located in chromosome number 8 (Figure 1).

The quantile-quantile (QQ) plots displayed the strong significant association distribution between important SNP candidate numbers and leaflet width trait $P < 0.05$ (Figure 2).

Regarding the oil palm leaf area character, this study revealed that nine SNP markers, located in chromosomes 1, 2, 4 and 10 were significantly associated with this trait. The most significant SNP marker associated with leaf area was located in chromosome 4 of the oil palm genome (Figure 3). The quantile-quantile (QQ) plots displayed the strong significant association distribution between the SNP candidates for this oil palm vegetative phenotype $P < 0.05$ (Figure 4).

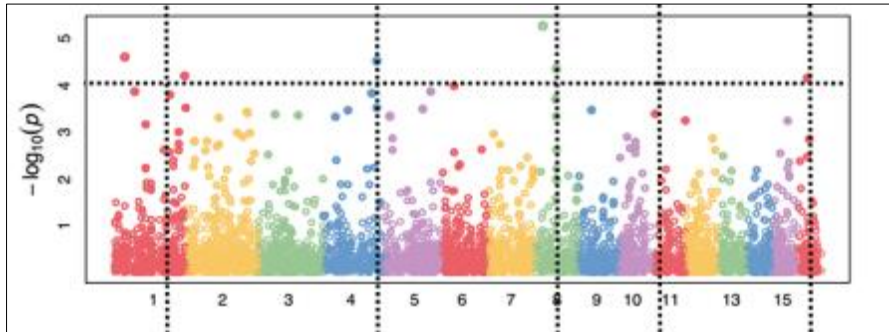


Figure 1 Manhattan plot of a genome-wide association with oil palm leaflet width study. Probability values are displayed in negative log scale with base of 10 ($-\log_{10} P$) against the physical map positions of genetic markers. Chromosomes are designated with different colors. Candidate genes and quantitative trait nucleotides (QTNs) are marked with black vertical dotted lines

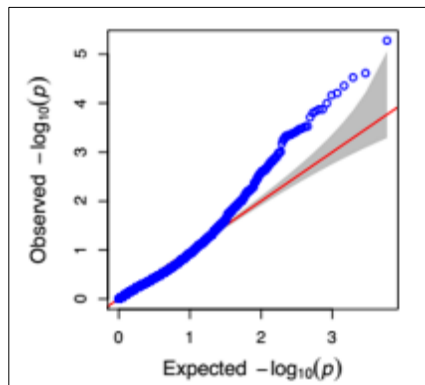


Figure 2 QQ plots illustrate how well the majority of genetic markers fit the null hypothesis (i.e., the SNP markers that are not associated with the leaflet width phenotype). A red line indicates the expectation. The area of the 95% confidence interval is filled in gray. The dots above the confidence interval on the right indicate the genetic markers that are associated with the oil palm leaflet width trait

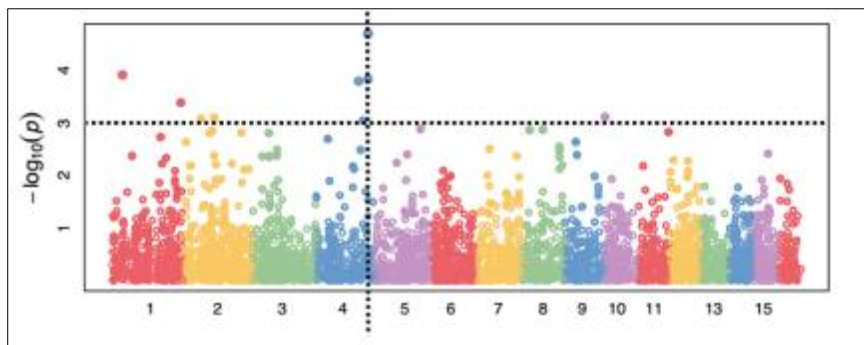


Figure 3 Manhattan plot of a genome-wide association with oil palm leaf area. Probability values are displayed in negative log scale with base of 10 ($-\log_{10} P$) against the physical map positions of genetic markers. Chromosomes are designated with different colors. Candidate genes and quantitative trait nucleotides (QTNs) are marked with black vertical dotted lines

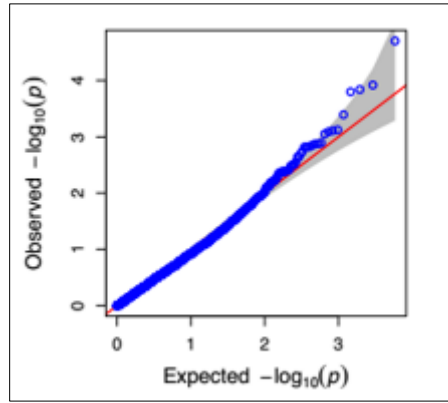


Figure 4 QQ plots illustrate how well the majority of genetic markers fit the null hypothesis (i.e., the SNP markers that are not associated with the leaf area phenotype). A red line indicates the expectation. The area of the 95% confidence interval is filled in gray. The dots above the confidence interval on the right indicate the genetic markers that are associated with the phenotype

4. Discussion

The SNP analysis revealed that the Hardy-Weinberg equilibrium was significant for this wild population. Thus, the alternative hypothesis should be considered, indicating that these allele and genotype frequencies were not in Hardy-Weinberg equilibrium. This result suggests that from generation to generation the proportion or frequency of alleles and genotypes will not be constant in that wild type population because of the presence of the genetic destructive action of the base wild population in their natural environment. This implies that we are in the presence of the great genes flow situation which is also confirmed by the presence of heterozygosity proportion detected in this natural population. Thus, important advises should be taken before selection of some accessions for breeding program. The Deviation from Hardy-Weinberg equilibrium in a population indicates possible inbreeding, population stratification and sometimes problems with the genotyping (genotyping error) or even actual association to the trait under study (Wittke-Thompson *et al.*, 2005).

The result of genome wide association mapping revealed that strong significant associations were found between oil palm leaflet width and genome region SNP markers of Chromosome 8, while the leaf area trait was associated with the SNP markers of chromosome 4. These suggested that the genes which are controlling all these two vegetative component traits of oil palm are located in these specific genome regions of chromosomes. Pootakham *et al.*, in 2013 identified three QTL affecting trunk height on Linkage Group (LG) 10, 14 and 15. According to Jin *et al.* (2016) a small set of SNPs covering the whole genome would be required to conduct genome-wide association studies (GWAS) for important traits for marker-assisted selection (MAS).

5. Conclusion

The study provide new genes associated region candidates for leaflet width and leaf expression trait phenotypes, useful for bioengineering. Oil palm molecular selection strategy currently associated with conventional breeding should use these results to accelerate the improvement of these important vegetative characters in oil palm breeding program. Thus, the current oil palm accession populations of Cameroon germplasm can also be introduced in some oil palm improvement selection program.

Compliance with ethical standards

Acknowledgments

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

- [1] Ajambang W., Mondjeli C., Ntsomboh Ntsefong G., Sudarsono S. (2016). RNA-seq analysis reveals influence of sugar level and photoperiod on seasonality in oil palm (*Elaeis guineensis* Jacq.) sex specific inflorescence emergence. *Journal of Applied Biology and Biotechnology* doi: 10.7324/JABB.2016.40203. 4 (02) 016 – 022.
- [2] Alvarado A., R. Escobar, F. Peralta. 2010. ASD's oil palm breeding program and its contributions to the oil palm industry. *ASD Oil Palm* 34: 1-16.
- [3] 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): 754-76 5.
- [4] 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.
- [5] Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler E.S. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633-2635.
- [6] Flint-Garcia S.A., A.C. Thuitet, J. Yu, G. Pressoir, S.M. Romero, S.E. Mitchell, J. Doebley, S. Kresovich, M.M. Goodman, E.S. Buckler. 2005. Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J.* 44: 1054-1064.
- [7] Ganai M.W., G. Durstewitz, A. Polley, A. Bérard, E.S. Buckler, A. Charcosset, J.D. Clarke, E-M. Graner, M. Hansen, J. Joets, M-C. Le Paslier, M.D. McMullen, P. Montalent, M. Rose, C-C. Schön, Q. Sun, H. Walter, O.C. Martin, M. Falque. 2011. A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS ONE*, 6: e28334
- [8] Goff S.A., D. Ricke, T.H. Lan, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science.* 296: 92-100.
- [9] Jeennor, 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.
- [10] 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.
- [11] Lipka, A.E. et al. 2012. GAPIT: genome association and prediction integrated tool. *Bioinformatics.* 28(18): 2397–2399.
- [12] Mangin B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C. 2012. Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* 108: 285–291.
- [13] Mariette et al., 2016, Genome wide association links candidate genes to resistance to Plum Pox Virus in apricot (*Prunus armeniaca*). *New Phytologist*, 209, 773-784. 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.
- [14] Mondjeli C., R. Sobir, M. Syukur, W.B. Suwarno, G. Ntsomboh-Ntsefong. 2017. Genetic diversity and interrelationship among some Dura x Tenera oil palm (*Elaeis guineensis* Jacq.) genotypes in Cameroon. *Journal of Agricultural Science and Technology A. DAVID PUBLISHING* 7: 81-90 doi: 10.17265/2161-6256/2017.02.002
- [15] Money, D., Gardner, K., Schwaninger, H., Zhong, G.Y., Myles, S. 2015. LinkImpute: fast and accurate genotype imputation for non-model organisms. *G3* 5(11):2383-2390. <https://doi.org/10.1534/g3.115.021667>
- [16] Morrell P.L., E.S. Buckler, J. Ross-Ibarra. 2011. Crop genomics: advances and applications. *Nature Reviews Genetics.* Vol. 13 (2): 85-96.
- [17] Paterson A.H., M. Freeling, T. Sasaki. 2005. Grains of knowledge: genomics of model cereals. *Genome Res.* 15: 1643-1650.
- [18] Pootakham W., P. Uthapaisanwong, D. Sangsrakru, T. Yoocha, S. Tragoonrung, S.

- [19] Tangphatsornruang. 2013. Development and characterization of single nucleotide polymorphism markers from 454 transcriptome sequences in oil palm (*Elaeis guineensis*). *Plant Breed.* 132: 711-717.
- [20] Prasanna B.M., K. Pixley, M.L. Warburton, C-X. Xie. 2010. Molecular marker-assisted breeding options for maize improvement in Asia. *Molecular Breeding* 26 (2): 339-356.
- [21] Singh, R., M.O. Abdullah, E.T.L. Low, M.A.A. Manaf, R. Rosli, R. Nookiah, L.Cheng-Li Ooi, S. Eng Ooi, K.L. Chan, M.A. Halim, N. Azizi, J.Y. Nagappan, B. Bacher, N. Lakey, S.W. Smith, D. He, M. Hogan, M.A. Budiman, E.K. Lee, R. DeSalle, D. Kudrna, J.L. Goicoechea, R.A. Wing, R.K. Wilson, R.S. Fulton. 2013. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature.* vol. 500, 335–339.
- [22] Wittke-Thompson J.K, A. Pluzhnikov, N.J. Cox. 2005. Rational inferences about departures from Hardy-Weinberg equilibrium. *Am J Hum Genet.* 76: 967-986.
- [23] Yan J., C.B. Kandianis, C.E. Harjes, L. Bai, E-H. Kim, X. Yang, D.J. Skinner, Z. Fu, S. Mitchell, Q. Li, M.G.S. Fernandez, M. Zaharieva, R. Rabu, Y. Fu, N. Palacios, J. Li, D. Dellapenna, T. Brutnell, E.S. Buckler, M.L. Warburton, T. Rocheford. 2010. Rare genetic variation at *Zea mays* crtRB1 increases β -carotene in maize grain. *Nature genetics.* 42 (4): 322-327.
- [24] Yang W., Z. Guo, C. Huang, L. Duan, G. Chen, N. Jiang., et al. 2014. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5: 5087.
- [25] Zila C.T., L.F. Samayoa, R. Santiago, A. Butron, J.B. Holland. A genome-wide association study reveals genes associated with *Fusarium* ear rot resistance in a maize core diversity panel G3-Genes. *Genom Genet* 2013, 3: 2095-2104.