Association Mapping Identifies Markers Linked with Yield Traits in an Oil Palm Breeding Population

Nipha Chuenpom[a]

Center for Agricultural Biotechnology, Kasetsart University, Kamphaeng Saen Campus, Kamphaeng Saen, Nakhon Pathom 73140

Hugo Volkaert[a]*

National Center for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Thailand Science Park, Khlong Nueng, Khlong Luang, Pathum Thani 12120

Abstract

Oil palm (*Elaeis guineensis*) is an important oil crop for human consumption worldwide. Oil production is an interested trait to get more value of palm oil. Oil yield traits in an oil palm breeding program is difficult as oil palm is a highly heterozygous out-crossing crop with a long life cycle and large size. This research aims to identify Single Nucleotide Polymorphism (SNP) markers linked with oil palm yield traits. Total 40 candidate SNP markers were tested with 499 individual palms from an unrelated set of 32 families using association testing for quantitative traits. 9 SSR markers were used to estimate population structure and kinship relationships. The results indicated that the genotypes at 36 markers associated with oil yield traits (BN, ABW, FFB and PH combined) at significant statistical P-value, corrected for multiple testing. Eg_Pairs05 and SSCP0020 were highly significant associated with traits that might be high potential markers for oil palm breeding in the future.

Keywords: GWAS; SNP marker; marker assisted selection (MAS)

1. คำนำ

Oil palm (*Elaeis guineensis* Jacq.) is an important oil crop with the highest oil production per area planted per year (Malaysian Palm Oil Council, 2012). Palm oil is obtained from the fleshy mesocarp while the palm kernel oil is extracted from the endosperm. Oil palm is still a relatively unimproved crop as it has gone through only a few cycles of intensive selection. Breeding for improved oil yield is a slow process as it is an outcrossed crop with high level of heterozygosity and the generation cycle is long. Since the plants are large, the number of plants established in progeny populations is generally few. Marker assisted selection might be useful for oil palm breeding as it can facilitate selection

[a]*2nd Address: Center of Excellence on Agricultural Biotechnology: (AG-BIO/PERDO-CHE), Bangkok 10900 Thailand*

**Corresponding author: hugo.v@ku.ac.th*

DOI: 10.14456/tjst.2017.1

for beneficial traits before establishment in the field. This research aimed to develop molecular markers for yield traits in oil palm and use them in operational breeding populations.

Oil yield is a quantitative trait, although the shell thickness locus which contributes substantially to yield is determined by a single Mendelian gene. Oil yield is broken down into several measurable traits such as bunch number (BN), average bunch weight (ABW), percentage of fruit per bunch (% F/B), mesocarp to fruit ratio (M/F), shell to fruit ratio (S/F), kernel to fruit ratio (K/F), oil content of the mesocarp and others (Rance *et al*., 2001; Singh *et al*., 2009; Corley *et al*., 1971). These are quantitative traits determined by multiple loci and thus complicated for selection. Up to now, very few markers linked to traits of interest have been found. Two RAPD markers were identified to be linked on both sides of the *sh*+ locus using an F_1 Tenera x Pisifera progeny and an AFLP marker closely linked to the *Sh* locus at 4.7 cM using 116 full-sibs derived from a cross between Tenera x Dura parents (Moretzsohn *et al*., 2000; Billotte *et al*., 2005). Eventually, the *SHELL* gene has been identified by mapping using high-throughput sequencing technology (Singh *et al*., 2013).

Since oil yield traits are controlled by multiple loci and oil palm is highly heterozygous and outcrossing therefore oil palm breeding should be improved using association markers for selection of the complicated traits. Association mapping is based on the detection of linkage disequilibrium (LD) which is defined as the non-random association of alleles at individual loci usually located closely together on the same chromosome. Association mapping is a powerful tool to identify markers linked to complex traits and has been used to identify the alleles that contribute to traits of interest in rice, maize, *Arabidopsis* and other plants (Courtois *et al*., 2013; Yang *et al*., 2010; Zhao *et al*., 2005).

Association mapping model were famous to combined population structure because population structure is one important evolutionary factors to influence LD across genomes (Dudley *et al*., 2002). Population structure occurs from the unequal distribution of alleles among subpopulations of different ancestries. When these subgroups are sampled to construct a panel of lines for LD, the intentional or unintentional mixing of individuals with different allele frequencies creates LD. Significant LD between unlinked loci results in false-positive associations between a marker and a trait (Slatkin, 2008). However, when individuals are mixed, in the newly created artificial single population, false polymorphisms and consequently significant but spurious LD is observed.

2. Materials and methods

2.1 Plant materials and phenotyping

Oil palm population for marker development, the highest oil producing palm trees, having high bunch number with bunches of average weight, were identified in each of 15 full-sib dura x pisifera progenies planted in the breeding population of Univanich Palm Oil Public Company Ltd., Krabi, Thailand. For each selected tree, a control tree was chosen among

the palm trees planted immediately surrounding it. The control tree had a bunch weight and bunch number close to the average for that particular progeny.

Total 499 trees from 32 full-sib families of oil palm (Dura x Pisifera) were collected from the breeding plantation of Univanich Palm Oil Public Company Ltd., Krabi, Thailand. Only those palm trees surrounded on all sides by healthy trees planted at the same time, excluding those where one or more surrounding plants had been replanted because of death of the originally plant, were selected in order to limit the effect of competition environment on phenotype measurements. All palm trees in the progenies were measured for four vegetative traits in 2012 viz. rachis length (RL), palm height (PH), petiole cross section (PX) and leaf area (LA). Data on four bunch yield traits were collected from 2008-2012 including bunch number (BN), fresh fruit bunch (FFB), average bunch weight (ABW) and percent oil per bunch (% O/B).

Statistical analysis of phenotype data, R program was used (R Core Team, 2003) to test normality and calculate correlation values between traits on the basis of the individual variables.

2.2 Genotyping

Total 40 candidate SNPs primers including CAPSs, dCAPSs, and SSCPs were used for genotyping. The candidate SNPs were obtained from 24 reduced representation libraries in oil palm (Dura x Pisifera) which developed by Chuenpom *et al*., (Submitted). The SNPs database had been available on http://www.it.cab.kps.ku.ac.th/breedserve/oilpalsnp. In addition, 9 SSR primers (Billotte *et al*., 2005) and 3 genes expression (Jeennor and Volkaert, 2013) were genotyped with 499 oil palm tree from 32 families.

PCR amplification of SSRs, ESTs, SSCPs, CAPSs and dCAPSs were performed using approximately 30 ng of genomic DNA template, 1X PCR buffer with $(NH_4)_2SO_4$, 0.25 mM of $MgCl_2$, 0.15 mM of each dNTP, 0.3 µM of each primer, and 0.2 U of *Taq* DNA polymerase (Fermentas, Lithuania) in a total volume of 10 µl. Reactions were incubated for Pre-denature at 94 ^oC for 3 min, followed by 40 cycles of denature at 94 ^oC for 40 s, primer-specific annealing temperature at 55-60 ^oC for 40 s, extension at 72 ^oC for 30 s and a final extension of 72 ^oC for 5 min. Then the PCR products were checked on 1 % agarose gel electrophoresis before separated the polymorphic bands by polyacrylamide gel electrophoresis.

Analysis of polymorphism in SSR-PCR fragments was done by mixing 2 µl of PCR product with an appropriate volume (6-10 volumes, depending on concentration of the PCR fragment) of loading dye (95 % (v/v) formamide, 0.025 % bromophenol blue, 0.025 % xylene cyanol and 10 mM EDTA). The electrophoresis was performed on 4.5 % denaturing polyacrylamide gels (acrylamide : bris-acrylamide 19 : 1) containing 8 M urea and 1X TBE. Sample aliquots (2 µl) were loaded and Pre-run in 1x TBE buffer at constant 120 watt for 30 min then run at 75 watt for 1.5-2 hrs at room

temperature. DNA bands in polyacrylamide gel were visualized with silver staining (Bassam *et al*., 1991).

Polymorphism of the SNP-SSCPs candidate markers were screened on non-denaturing polyacrylamide gels using single strand conformation polymorphism (SSCP) conditions (Orita *et al*., 1989). Dilution of the PCR product : SSCP dye (1 : 4) (SSCP-loading dye (95 % (v/v) formamide, 0.025 % bromphenol blue, 0.025 % xylene cyanol and 10 mM NaOH), then denatured at 95 ^{o}C for 5 min and immediately placed in ice-cold water to stabilize single strands. Next 2 µl aliquots were loaded on the 5.6 % polyacrylamide gels (acrylamide : bris-acrylamide 99 : 1) with 1X TBE as electrophoresis buffer at constant 10-12 Watt for 10-12 hrs in a 4 ^{o}C refrigerator. DNA bands in the polyacrylamide gel were visualized with silver staining

SNP-CAPS and dCAPS markers genotyping by CAPS assay, the remaining of PCR product after checking on agarose gel about 5-7 µl were digested with the required restriction enzyme. Each digestion reaction contained 0.5-1 Unit of the restriction enzyme and 1X compatible buffer in a total volume of 5 µl. Enzyme digestions were incubated at the appropriate temperature for at least 2 hrs. Digestion products were separated on 3 % agarose gel electrophoresis.

2.3 Population clustering

Population structure was analyzed using STRUCTURE program (Pritchard *et al*., 2007) to identify the appropriate number of distinct populations (K). The STRUCTURE program was run five times for each K, ranging from 1 to 30, using the admixture and correlated allele frequency model with 250,000 steps for burn-in and 750,000 steps of the MCMC sampling after burn-in. The LOC_PRIOR option was used, with each progeny assigned to a distinct assumed population. The final population subgroups were determined based on Ln P(D) and Var[Ln P(D)] for the model choice criterion to refer as true number of populations (K). STRUCTURE HARVESTER (Earl and vonHoldt, 2012) was used for collating the results from STRUCTURE program to assess and visualize likelihood values across multiple values of K for easier detection of the number of genetic groups that best fit the data according to the Evanno method (Evanno *et al*., 2005). The data from STRUCTURE were made consistent across runs by the CLUMPP program (Jakobsson and Rosenberg, 2007). DISTRUCT (Rosenberg, 2004) was used to redraw the bar plots.

Population structure were analyzed the best K for all traits. However, we attempted to describe “What K was appropriated for each trait?”. Phylogenetic tree were used for clustering of oil palm population using DARwin6 program (Perrier *et al*., 2003). The program calculates a genetic distance based on the number of shared alleles among diploid organisms. The weighted neighbor joining method was used to calculate based on 492 oil palm trees after exclude sample with high missing data. Moreover, principal component analysis (PCA) was calculated to confirm probability of clustering using R program

'FactoMineR' version 1.31.3 (R Core Team, 2003).

2.4 Association analysis

Association analysis between genotypes and phenotypes among 499 samples were analyzed by Tassel program version 3.0 (Bradbury *et al*., 2007). The general linear model (GLM) and mixed linear model (MLM) approaches were explored for association analysis combined with population structure and kinship data. Kinship coefficients between individuals were estimated with a subset of genetic markers across the entire genome using SPAGeDi program (Hardy and Vekemans, 2002) following Loiselle *et al*., (1995).

3. Results

3.1 Phenotyping of test progenies

Eight traits including four vegetative traits were measured: rachis length (RL), palm height (PH), petiole cross section (PX) and leaf area (LA) with ranges 2.97-7.90 m, 1.33-4.96 m, 12.22-65.49 cm^2 and 4.97-16.31 cm^2 respectively. Four bunch yield traits were collected over 4 years and analyzed including bunch number (BN), fresh fruit bunch (FFB) and average bunch weight (ABW) with ranges of values were 16-167 bunches, 191.60-1,505.60 kg and 5.03-17.24 kg/bunch. The percent oil per bunch (% O/B) was available only for some of the palm trees and ranged from 15-43 %.

Then measurements for eight agronomical traits were derived from 32 crosses using R program (Figure 1). Correlation coefficients and their significance between the 8 traits are given in table 5. Highly significant positive correlations were found among BN with FFB, PH with FFB, PX and LA with ABW, and PX with RL. A highly negative correlation was found for BN with ABW (-0.70395, P-value <0.01) (Table 1).

Table 1 Correlation coefficients among 8 phenotypes from 499 oil palm trees (below diagonal) and significance (above diagonal)

Traits	FFB	BN	ABW	RL	PH	PX	LA	% O/B
FFB	-	<0.0001	<0.0001	0.00418	<0.0001	<0.0001	<0.0001	0.15980
BN	0.44800	-	<0.0001	<0.0001	0.41845	<0.0001	<0.0001	0.05504
ABW	0.28747	-0.70395	-	<0.0001	<0.0001	<0.0001	<0.0001	0.00640
RL	0.12805	-0.24867	0.35301	-	0.77377	<0.0001	<0.0001	0.29852
PH	0.41603	0.03633	0.26384	0.01293	-	0.00039	0.46747	<0.0001
PX	0.22206	-0.33624	0.50842	0.53505	0.15830	-	<0.0001	0.0004
LA	0.20352	-0.32336	0.45346	0.33949	0.03262	0.52537	-	0.2528
% O/B	0.08667	-0.11890	0.16711	0.06413	0.43692	0.21420	0.07059	-

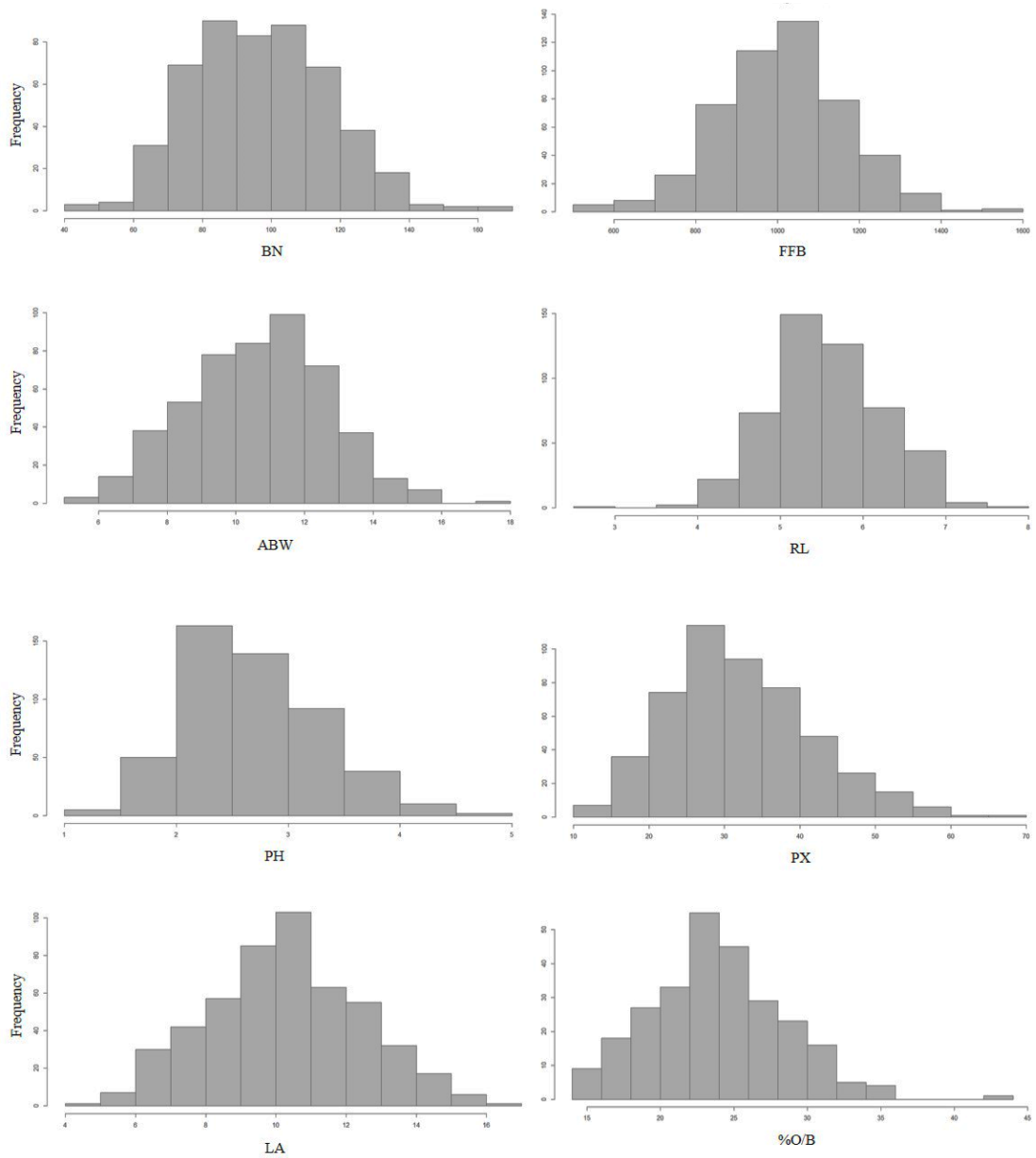

Figure 1 Distribution of 8 agronomical traits from 499 oil palm trees. (BN) Bunch number, (FFB) Fresh fruit bunch weight, (ABW) Average bunch weight, (RL) Rachis length, (PH) Palm height, (PX) Petiole cross section, (LA) Leaf area and (% O/B) Percent oil per bunch

3.2 Genotyping of candidate markers

Two or 3 alleles were found for the SSCP markers while the CAPS and dCAPS assays yielded binary markers. Then SNP genotypes were scored in co-dominant system for association analysis. The kinship relationships derived from the all marker data indicated that almost all trees belonging to the same progeny

Table 2 The results of population structure using STRUCTURE program and STRUCTURE HARVESTER program

K	Mean LnP (K)	Stdev LnP (K)	Ln' (K)	|Ln'' (K)|	Delta K
1	-39009.00	0.1225	NA	NA	NA
2	-37146.84	8.6685	1862.16	713.98	82.3648
3	-35998.66	18.3590	1148.18	231.18	12.5922
4	-35081.66	28.0616	917.00	199.46	7.1079
5	-34364.12	39.7229	717.54	89.66	2.2571
6	-33736.24	133.8077	627.88	283.96	2.1222
7	-33392.32	161.1004	343.92	64.02	0.3974
8	-32984.38	91.6090	407.94	288.32	3.1473
10	-32634.80	138.8186	229.96	81.80	0.5893
11	-32486.64	4.3627	148.16	146.16	33.5023
12	-32484.64	205.6816	2.00	296.58	1.4419
13	-32186.06	125.6153	298.58	174.82	1.3917
14	-32062.30	203.4774	123.76	788.18	3.8736
15	-32726.72	1756.9724	-664.42	1139.8	0.6487
16	-32251.34	1060.0661	475.38	205.18	0.1936
17	-31981.14	728.0451	270.20	22.60	0.0310
18	-31733.54	211.0661	247.60	517.56	2.4521
19	-32003.50	678.5439	-269.96	661.52	0.9749
20	-31611.94	68.1696	391.56	711.32	10.4346
21	-31931.70	401.0500	-319.76	530.80	1.3235
22	-31720.66	152.9331	211.04	903.40	5.9072
23	-32413.02	713.3988	-692.36	747.38	1.0476
24	-32358.00	960.7901	55.02	102.64	0.1068
25	-32405.62	1179.5456	-47.62	2557.08	2.1679
26	-35010.32	3420.8402	-2604.70	48.78	0.0143
27	-37663.80	6042.9010	-2653.48	1181.30	0.1955
28	-39135.98	2134.4083	-1472.18	5286.06	2.4766
29	-35322.10	4846.8240	3813.88	3375.90	0.6965
30	-34884.12	4172.5752	437.98	NA	NA

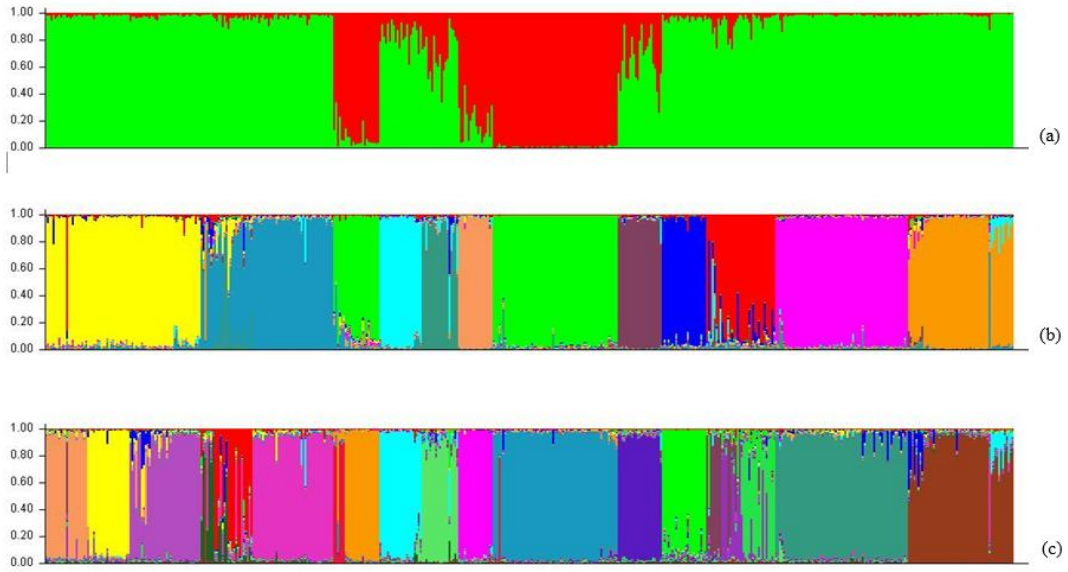

Figure 2 Population structure observed at K = 2 (a), K = 11 (b) and K = 20 (c)

were closely related with a few exceptions. The relationships between progeny sets did not differ substantially.

3.3 Population clustering

Five independent runs of the program STRUCTURE (Pritchard *et al*., 2007) were used to test different population genetic relationships assuming different numbers of ancestral populations (K = 1 to 30). The STRUCTURE results were further analysed in STRUCTURE HARVESTER (Earl and vonHoldt, 2012) to find the best K value following the approach as in Evanno *et al*., (2005). Based on this analysis a population structure with 2 clusters would be recommended with likelihood value mean LnP(K) = -37146.84 and delta K = 82.364825. As in the second derivative of K was observed at K = 11 with mean likelihood value -32486.64 and delta K = 33.502322 a population structure with 11 subpopulations was tested as well. Up to K = 20 all cluster assignments essentially were congruent with the half-sib progenies and thus K=20 was chosen as an appropriate K value for the association testing (Table 2, Figure 2 and Figure 3). Moreover, phylogenetic tree showed the best cluster at 11 clusters after using DARwin6 NJ-weight method (Figure 4). Also, the results from PCA analysis showed population structure when compare with phylogenetic tree as 11 groups at value of inertia 67.64 % (Table 3). From the third analyzations 11 clusters of population were suggested that might be perfected for marker association.

3.4 Association between genotypes and phenotypes

Association between genotypes and phenotypes were tested by Tassel program. A total of 36 markers showed significant association with oil production and/or vegetative traits from GLM and MLM model with and without

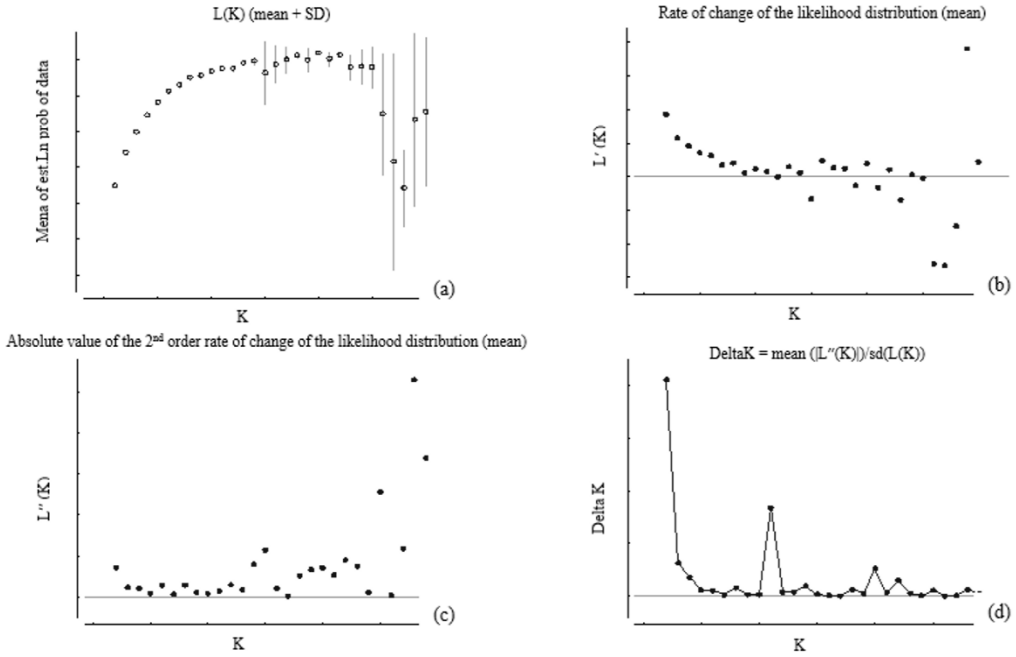

Figure 3 Analysis of the STRUCTURE output following the Evanno *et al*., (2005) approach as implemented in STRUCTURE HARVESTER to find the best K value. (a) graph of the calculated likelihoods and their first (b) and second (c) order derivatives. (d) graph of the resulting ΔK values

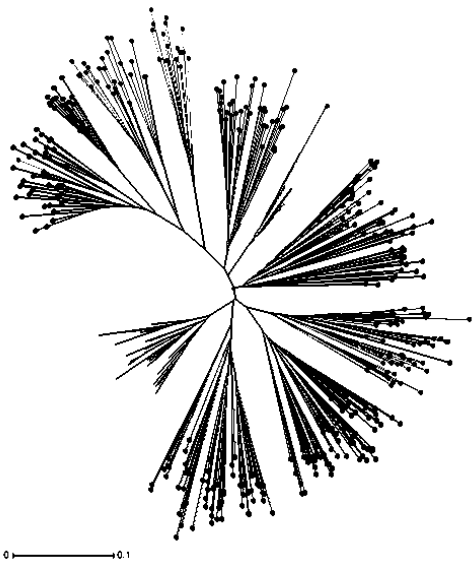

Figure 4 Phylogenetic tree of oil palm population using NJ-weight method

population structure. GLM analysis assuming 11 sub-populations (G+P+S11) showed 26 SNPs and 3 gene markers significantly associated with 7 traits with P-value after resampling correction 2.79E-08 to 4.81E-02. One SNP marker (Eg_Pairs05) was highly significant and associated with 5 traits (BN, ABW, LA, PX and RL). MLM model was combined with kinship data and without population structure (G+P+K) found that 22 markers significant with 8 traits. Eg_Pairs16 was significant with percent oil per bunch at 4.76E-02 and SSCP0020 marker was a common marker with (BN, ABW and LA). For MLM with population structure (G+P+K+S11) number of significant markers was decreased

to17 markers. However, Eg_pairs05 is still a high potential marker for all model.

Table 3 Summery of PCA analysis

PCA group	Eigenvalue	% inertia	Sum of % inertia
1	1.66E+01	1.73E+01	17.33392
2	9.54E+00	9.93E+00	27.26681
3	8.04E+00	8.38E+00	35.64700
4	5.90E+00	6.15E+00	41.79410
5	5.32E+00	5.54E+00	47.33086
6	4.21E+00	4.39E+00	51.72064
7	3.70E+00	3.85E+00	55.57402
8	3.41E+00	3.55E+00	59.12876
9	3.06E+00	3.19E+00	62.31907
10	2.85E+00	2.97E+00	65.28885
11	2.25E+00	2.35E+00	67.63504
12	1.95E+00	2.03E+00	69.66612
13	1.65E+00	1.72E+00	71.38158
14	1.45E+00	1.52E+00	72.89683
15	1.40E+00	1.46E+00	74.35627
16	1.33E+00	1.38E+00	75.73966
17	1.25E+00	1.30E+00	77.04237
18	1.22E+00	1.27E+00	78.31470
19	1.18E+00	1.23E+00	79.54589
20	1.06E+00	1.10E+00	80.64560

4. Discussion

4.1 Phenotype and environment of plant materials

Oil yield traits are controlled by several environmental factors in addition to genetics. Oil yield is affected by water status (drought which can be alleviated by irrigation), soil fertility and applied fertilizer, and other factors (Moretzsohn *et al*., 2000). Therefore multiple measurement of bunch production between 2008-2012 may have the effect when analyze the LD between genotypes and phenotypes. However, the association markers might be useful for wide population because we study and analyzed many alleles in oil palm at the same time and attempt to decrease the effect form population structure and relationship between individual.

4.2 Correlations among phenotype measurements

Highly significant correlations were observed among several of the measured variables in the 499 trees of the marker verification population. As generally observed (Ooi *et al*., 1973; Van der Vossen, 1974; Okoye *et al*., 2009), bunch number and average bunch weight were negatively correlated (-0.73). However, the high-yield and control trees of the marker discovery population were selected to have comparable bunch weights and differ mostly in bunch number, and the increased bunch harvest was solely due to a higher number of bunches. Thus it was expected that a more directed search for markers involved in bunch number could be done.

Some highly significant correlations were also noted among vegetative traits. Petiole cross-section was correlated with rachis length (0.53) and leaf area (0.56) as can be expected that larger leaves have a longer and stronger petiole. But there were also some interesting correlations among vegetative traits and yield

traits. Leaf area and petiole cross-section were correlated with average bunch weight (0.36 and 0.50 respectively) but less so with FFB (0.23 and 0.19, resp.). Palm height was correlated with total yield (FFB, 0.35) and oil per bunch (0.44). Many of these correlations can be interpreted that a more vigorously growing palm tree will produce larger leaves, which may make more biomass available for bunch production. A tree that grows slightly taller will also be able to extend its leaves above the leaves of neighbouring palms and thus be able to capture more solar energy, which then in turn is used mainly for higher oil and bunch production, but apparently less for vegetative growth as there was a very low and mostly non-significant correlation between palm height and rachis length, petiole cross-section and leaf area (Corley *et al*., 1971).

4.2 Clustering of oil palm population

Population structure can affect the association analysis tremendously. From the set-up of the breeding program, it is clear that the population structure has to be taken into account when performing association analysis. However, what is less clear is the best way to quantify this population structure. As the 499 samples were collected from 32 different full-sib progenies, one might expect that that would result in 32 clusters. However, some of the crosses might have shared a parent or grandparent, either as pollen donor or as seed parent, and thus one might expect some progenies to belong to the same cluster.

An approach to decide on the correct number of subpopulations that is followed quite widely is the calculation of the rates of change in likelihood for the different K-values and their comparison to the standard deviation of the likelihood as proposed by Evanno *et al*., (2005). However, the Evanno method has been validated on simulated datasets that assume equal size for each identifiable subpopulation and equal inter-population genetic distances. This is clearly not the case among the 499 samples, as some progeny families are smaller than most others (6-10 individuals only vs 18-20) and as some families share a common parent (or grandparent), while others are less related. Thus the Evanno analysis can be considered as a guide, but not necessarily will give the correct K value for the structure in a breeding population. The data were analysed assuming population structure with 2 clusters following by Evanno *et al*., (2005), with 11 clusters (2^{nd} highest ΔK) and with 20 clusters (highest likelihood overall, 3^{rd} highest ΔK). A neighbor-joining phylogenetic tree produced by DARwin to confirm oil palm population structure. From the third of analyzations 11 clusters of population were suggested that might be perfected for marker association. The markers will be useful in oil palm breeding program in the future.

4.3 Association analysis

The multiple alleles caused problems when used for association analysis using TASSEL program. Tassel was designed for homozygous line analysis, but oil palm is a highly heterozygous plant. The TASSEL results showed many SSR loci with highly significant

associations with traits. One factor contributing to the biased results is that the counts for each allele are smaller when there are many alleles in a population. As such, for some of the lower frequency alleles there could be spurious association with a phenotype. For a reliable X^2 test, it is recommended that one should have at least 5 observations for each marker genotype-phenotype combination. There is however an additional complication in that the Tassel program actually considers marker phenotypes, not genotypes. Thus a heterozygous locus is recoded as a third phenotype, different from either homozygote. For a multi-allelic marker that can result in a high number of marker phenotypes when there are 5 to 8 alleles in a population, limiting the statistical testing even more. It was also noted that partial genotypes, where the obtained data indicated the presence of one allele at a locus but could not be conclusive that the locus was homozygous or that the second allele might have been undetected, were recorded in Tassel as a separate marker phenotype class. Therefore, all partial genotypes had to be recoded as completely missing information.

However, SNP markers were found from GLM and MLM model. Eg_Pairs05 and SSCP0020 were highly associated with traits (BN, ABW, LA, PX and RL). Only Eg_Pairs16 was significant with percent oil per bunch at P-value 4.76E-02 from MLM model without population structure. Moreover, Eg_Pairs04 and SSCP0032 were high potential markers with PH and FFB traits, respectively.

5. Acknowledgements

This research was supported by Center of Excellence on Agricultural Biotechnology, Science and Technology Postgraduate Education and Research Development Office, Office of Higher Education Commission, Ministry of Education. (AG-BIO/PERDO-CHE), Thailand and National Research Council of Thailand (NRCT). We thank to the Univanich Palm Oil Public Company Limited, Krabi, Thailand for providing plant samples and phenotypic data. Thank you Prof. Kathleen Marchal, Department of Plant Biotechnology and Bioinformatics and Department of Information Technology, Ghent University, Belgium for help with population clustering analysis.

6. References

Bassam, B.J., Anollés, G.C. and Gresshoff, P.M., 1991, Fast and sensitive silver staining of DNA in polyacrylamide gels, Anal. Biochem. 196: 80-83.

Billotte, N., Marseillac, N., Risterucci, M.A., Adon, B., Brottier, P., Baurens, C.F., Singh, R., Herrán, A., Asmady, H., Billot, C., Amblard, P., Durand-Gasselin, T., Courtois, B., Asmono, D. and Cheah, S.C., 2005, Micro-satellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). Theor. Appl. Genet. 110: 754-765.

Bradbury, P.J., Zhang, Z., Kroon, E.D., Casstevens, M.T., Ramdoss, Y. and Buckler, S.E., 2007, Tassel: Software for association mapping of complex traits in diverse samples, Bioinformatics 23: 2633-5.

Corley, R.H.V., Hardon, J.J. and Tang, T.G., 1971, Analysis of growth of the oil palm (*Elaeis uineensis* Jacq.) I. estimation of growth parameters and application in breeding, Euphytica 20: 307-315.

Courtois, B., Audebert, A., Dardou, A., Roques, S., Ghneim-Herrera, T., Droc, G., Frouin, J., Rouan, L., Gozé, E., Kilian, A., Ahmadi, N. and Dingkuhn, M., 2013, Genome-wide association mapping of root traits in a Japonica rice panel, PLoS One 8(11): e78037.

Dudley, A.T., Ros, A.M. and Tabin, J.C., 2002, A Re-examination of proximodistal patterning during vertebrate limb development, Nature 418: 539-544.

Earl, D. and vonHoldt, B., 2012, STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method, Conserv. Genet. Resour. 4: 359-361.

Evanno, G., Regnaut, S. and Goudet, J., 2005, Detection the number of clusters of individuals using the software STRUCTURE: A simulation study, Mol. Ecol. 14: 2611-2620.

Hardy, O.J. and Vekemans, X., 2002, Spagedi: A versatile computer program to analyse spatial genetic structure at the individual or population levels, Mol. Ecol. Notes 2: 618-620.

Jakobsson, M. and Rosenberg, A.N., 2007, CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure, Bioinformatics 23: 1801-1806.

Jeennor, S. and Volkaert, H., 2013, Mapping of quantitative trait loci (QTLs) for oil yield using SSRs and gene-based markers in african oil palm (*Elaeis Guineensis* Jacq.), Tree Genet. Genom. 10: 1-14.

Loiselle, B.A., Sork, L.V., Nason, J. and Graham, C., 1995, Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). Am. J. Bot. 82: 1420-1425.

Malaysian Palm Oil Council, 2012, Market Statistics, Available Source: http://www.mpoc.org.my, April 25, 2015.

Moretzsohn, M.C., Nunes, M.D.C., Ferreira, E.M. and Grattapaglia, D., 2000, RAPD linkage mapping of the shell thickness locus in oil palm (*Elaeis Guineensis* Jacq.), Theor. Appl. Genet. 100: 63-70.

Okoye, M.N., Okwuagwu, O.C. and Uguru, I.M., 2009, Population improvement for fresh fruit bunch yield and yield components in oil palm (*Elaeis Guineensis* Jacq.), Am. Eur. J. Sci. Res. 4: 59-63.

Ooi, S.C., Hardon, J.J. and Phang, S.,1973, Variability in the Deli dura breeding population of the oil palm (*Elaeis guineensis* Jacq.) I. Components of bunch yield, Malaysian Agric. J. 49: 112-121.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T., 1989, Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms, Proc. Natl. Acad. Sci.

USA. 86: 2766-2770.

Perrier, X., Flori, A. and Bonnot, F., 2003, Methods of data analysis, pp. 43-76, In Hamon, P., Seguin, M., Perrier, X., Glaszmann, J.C., Enfield, N.H. (Eds.), Genetic Diversity of Cultivated Tropical Plants, Science Publishers and CIRAD, USA.

Pritchard, J.K., Xiaoquan, W., and Daniel, F., 2007, Documentation for structure software: Version 2.2, Available Source: http://Pritch.bsd.uchicago.edu/software.

R Core Team R, 2003, A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, Available Source: http://www.R-project.org.

Rance, K.A., Mayes, S., Price, Z., Jack, L.P. and Corley, V.H.R., 2001, Quantitative trait loci for yield components in oil palm (*Elaeis Guineensis* Jacq.), Theor. Appl. Genet. 103: 1302-1310.

Rosenberg, N.A., 2004, Distruct: A program for the graphical display of population structure, Available Source: http://rosenberglab.bioinformatics.med.umich.edu/distruct.html.

Singh, R., Tan, G.S., Panandam, M.J., Rahman, A.R., Ooi, C.L., Low, L.T.E., Sharma, M., Jansen, J. and Cheah, S.C., 2009, Mapping quantitative trait loci (QTLs) for fatty acid composition in an interspecific cross of oil palm, BMC Plant Biol. 9: 114.

Singh, R., Ong-Abdullah, M., Low, L.T.E., Manaf, A.A.M, Rosli, R., Nookiah, R., Ooi, L.C., Ooi, S.E., Chan, K.L., Halim, M.A., Azizi, N., Nagappan, J., Bacher, B., Lakey, N., Smith, S.W., He, D., Hogan, M., Budiman, M.A., Lee, E.K., DeSalle, R., Kudrna, D., Goicoechea, J.L., Wing, R.A., Wilson, R.K., Fulton, R.S., Ordway, J.M., Martienssen, R.A. and Sambanthamurthi, R., 2013, Oil palm genome sequence reveals divergence of interfertile species in old and new worlds, Nature 500: 335-339.

Slatkin, M., 2008, Linkage disequilibrium-understanding the evolutionary past and mapping the medical future, Nature Rev. Genet. 9: 477-485.

van der Vossen, H.A.M., 1974, Towards more efficient selection for oil yield in the oil palm (*Elaeis guineensis* Jacquin), Doctoral Thesis, Wageningen University, Nederlands.

Yang, X., Yan, J., Shah, T., Warburton, L.M., Li, Q., Li, L., Gao, Y., Chai, Y., Fu, Z., Zhou, Y., Xu, S., Bai, G., Meng, Y., Zheng, Y. and Li, J., 2010, Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection, Theor. Appl. Genet. 121: 417-431.

Zhao, K., Aranzana, J.M., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P. and Nordborg, M., 2005, An arabidopsis example of association mapping in structured samples, PLoS Genet. 3(1): e4.