Intron length polymorphism markers of flower developmental genes for genetic diversity and QTL mapping for oil yield in African oil palm

Sonicha U-thoomporn[1], Anak Limsriwilai[2], Kittipat Uksokit[1*]

[1]Department of Biotechnology Faculty of Science and Technology, Thammasat University, Thailand
[2]Golden Tenera Company Limited, ad, Muang, Krabi, Thailand

**Corresponding author: ku@tu.ac.th*

ABSTRACT

The yield of oil from the palm (*Elaeis guineensis* Jacq.) is mainly proportional to the number of bunches of fruit and their weights. This is determined by the sex ratio and the ratio of mesocarp to the total dry mass of the bunch. In this study, 139 intron length polymorphism (ILP) markers were developed from 47 expressed sequence tags (ESTs) of genes involved in flower development. All ILP primers successfully amplified genomic DNA with products of the expected sizes and 79 markers were found to be polymorphic in a diversity panel of 41 palms. The efficiency of the ILP markers was demonstrated by the average polymorphic information content of 0.34 and the mean observed heterozygosity of 0.42. Genotypes were found to cluster into three major groups, based on pedigree and geographical origin. Thirteen polymorphic ILP markers developed from 11 candidate genes were mapped onto six published oil palm linkage groups corresponding to six oil palm chromosomes. Quantitative trait loci mapping for traits related to oil yield identified the ILP marker developed from AGAMOUS of a *MAS*-box gene. ILP markers developed from genes involved in flower development in the oil palm provide a powerful tool for the identification of functional markers that are associated with traits determining oil yield.

Keywords: oil palm; flower developmental genes; ILP; QTL mapping; genetic diversity

INTRODUCTION

The African oil palm (*Elaeis guineensis* Jacq.) has the highest potential oil yield per acre and is an important oil crop in Southeast Asia, Africa, and South America. The yield is mainly proportional to the number of bunches of fruit harvested per month and their weight (Corley and Tinker, 2007). The oil yield is determined by the sex ratio (i.e. the ratio of female inflorescences to total inflorescences), bunch failure resulting in failure of fruit setting, and the ratio of mesocarp to the total dry mass of the bunch. The sex ratio is affected by both sex determination and the preferential abortion of female or male inflorescences (Corley, 1976). Molecular genetic studies have revealed that genes involved in sex ratio regulation are homeotic in flower development (Adam *et al.*, 2006; Adam *et al.*, 2007). A large family of MADS-box genes plays roles in the regulatory pathways of flower formation (Coen and Meyerowitz, 1991). These genes encode transcription factors that control plant growth and development, including the formation of the flowering meristem and the male and female flower development organ (Adam *et al.*, 2007), and are responsible for the mantled phenotype of oil palm (Alwee *et al.*, 2006; Jaligot *et al.*, 2011). Several candidate genes in mantled oil palm have been investigated using transcriptome analysis (Shearman *et al.* 2013) and have been shown to be involved in the homeotic transformation of stamens and staminodes (rudimentary stamens) into pseudocarpels in the male and female flowers, respectively (Beulé *et al.*, 2011). Knowledge regarding genes involved in flower development and sex determination in oil palm is therefore necessary to support genetic improvement and increase oil yield.

QTL mapping supported by molecular markers can be used to identify markers that cosegregate with oil yield component traits (Ukoskit *et al.*, 2014; Pootakham *et al.* 2015). Recently, developments in structural and functional genomics have allowed mapping of gene-specific markers, derived from polymorphic sites within genes that influence phenotypic trait variation (Andersen and Lübberstedt, 2003). Introns are non-coding regions of a gene that are less strongly conserved and thus accumulate greater genetic variation than the exon regions (Choi *et al.*, 2004). Among the different types of polymorphism, intron length polymorphism (ILP) is the most sensitive

variation, and the sequences can be exploited to identify gene-specific markers (SaiSug and Ukoskit, 2013). A general method for identifying introns is to compare cDNA/EST sequences with the genomic sequences encoding them. Primers on the adjacent exon regions can be designed to amplify the genomic DNA across the intron using exon-primed intron-crossing (EPIC) markers (Li *et al.*, 2010). ILP has been exploited as a molecular marker across a range of plant species (Gupta *et al.*, 2012; Li *et al.*, 2012; Muthamilarasan *et al.*, 2014; SaiSug and Ukoskit, 2013). ILPs have been successfully used in diversity analyses of numerous crops (Gupta *et al.*, 2011; Poczai *et al.*, 2014; Shu *et al.*, 2010). Their use allows direct gene tagging for quantitative trait locus (QTL) mapping of traits with agronomic importance (Deng and Davis, 2001) increasing the efficiency of marker-assisted selection (Sheikh *et al.*, 2018).

In the present study, a set of genes were selected for the analyses. These genes are known to play roles in flower formation regulatory pathways such as genes associated with MADS-box, sex determination, and mantled oil palm inflorescence. Representative polymorphic ILP markers were evaluated for their utility in genetic diversity analysis. The candidate genes as ILPs were mapped into a published oil palm genetic map (Ukoskit *et al.*, 2014). The expanded map was used to identify the genome locations of the QTLs associated with oil yield. This approach demonstrates the integration of gene-specific makers such as ILP markers to improve our understanding of the complex traits underlying the palm oil yield.

MATERIALS AND METHODS

Plant materials

A segregating population of 208 progenies, derived from crossing of Clone B tenera and Clone D tenera at the Golden Tenera Limited Partnership, Krabi, Thailand, was field-planted in 2007 and used for linkage mapping as reported in a previous study (Ukoskit *et al.*, 2014). A subset of ten random individuals and their parent plants was used for preliminary screening of polymorphism before applying ILP markers to the entire population. Forty-one oil palm accessions were used for evaluation of ILP markers: AVROS (5), Calabar-AVROS (1), Ghana-AVROS (2), La Me-AVROS (2), Yangambi-AVROS (1), Calabar (4), DAMI (6) Deli Dura (8), Ghana (3), Ghana-Yangambi (1), Kazemba (1), Nigeria (5), Tanzania (1), and Yangambi (1). DNA was extracted from leaf samples taken from individual plants, using the method described by Gawel and Jarret (1991).

ILP Marker development and analysis

ILP markers were developed based on sequences of 26, 9, and 12 genes associated with MADS-box genes, flower/fruit development and sex determination, respectively (Table 1). The oil palm genome sequence data were downloaded from NCBI (Singh *et al.* 2013). The prediction of intron lengths and positions, and the design of ILP primers, were performed following the method described by Yang *et al.* (2007). Introns of a manageable size (<1000 bp) were extracted by aligning the EST sequences with their corresponding genome sequences using GeneSeqer (Brendel *et al.*, 2004), and were used to acquire information on intron position and length. A pair of specific primers flanking each of the predicted intron positions were designed using the Primer3 program. Amplification by PCR was performed in a 20 µl reaction mixture containing 4 ng template DNA, 1× PCR buffer (20 mM Tris pH 9.0, 100 mM KCl, 3.0 mM $MgCl_2$), 2 mM of each of the four dNTPs, 5 µM of each of the forward and reverse primers, and 2.5 units of *Taq* DNA polymerase. The following PCR conditions were used: 95 ^oC for 1 min, followed by 35 cycles of 94 ^oC for 30 s, the annealing temperature (45–57) ^oC appropriate to each primer pair for 30 s, 72 ^oC for 1 min, and a final extension at 72 ^oC for 5 min. The PCR products were separated on 4.5% denaturing polyacrylamide gel and silver-stained, following Benbouza *et al.* (2006). The genotypic profiles were used to determine the genetic diversity parameters, including the number of alleles (NA), observed heterozygosity (H), and polymorphic information content (PIC), using PowerMarker (Liu and Muse, 2005). Pairwise genetic distances between accessions were calculated based on similarity coefficients, following Nei and Li (1979). The similarity matrices were used to construct dendrograms using the Unweighted Pair Group Method with Arithmetic Averages (UPGMA). Cluster analysis was performed using NTSYS-pc version 2.02 (Rohlf, 1998).

Linkage Map Construction and QTL Analyses

Segregating marker data were added to an existing dataset (Ukoskit *et al.*, 2014) and used to construct the oil palm genetic map. The final map comprised four types of molecular markers: amplified fragment length polymorphism (AFLP), genomic simple sequence repeat (gSSR), EST-SSR, and ILP markers. Linkage analysis was performed as reported by Ukoskit *et al.* (2014) using JoinMap 3.0 (Van Ooijen and Voorrips, 2000) with the 'cp' option for a cross between two heterozygous parents. Linkage groups (LGs) were determined using a minimum LOD threshold of 4.0 and a recombination fraction threshold of 0.35. The Kosambi

mapping function was applied to convert recombination fractions to genetic distances. The chromosome locations of the mapped ILP markers were identified using the BLAST program with the marker sequences as query and oil palm reference genome (Singh *et al.*, 2013) as subject.

Table 1. Flower developmental genes used for developing the ILP markers.

Genes	Abbreviation	Size (bp)	Accession no.	References	No. intron
MADS-box genes					
GLOBOSA1	*GLO1*	10,985	AF227195.1	Adam *et al.*, 2007	5
GLOBOSA2	*GLO2*	24,104	AF411848.1	Adam *et al.*, 2007	5
AGAMOUS1	*AG1*	28,546	AY739698.1	Adam *et al.*, 2007	1
AGAMOUS2	*AG2*	13,468	AY739699.1	Adam *et al.*, 2007	1
AGAMOUS-Like2-1	*AGL2-1*	34,002	AF411843.1	Adam *et al.*, 2007	4
AGAMOUS-Like2-2	*AGL2-2*	24,479	AF411844.1	Adam *et al.*, 2006	3
AGAMOUS-Like2-3	*AGL2-3*	37,063	AF411845.1	Adam *et al.*, 2006	5
AGAMOUS-Like2-4	*AGL2-4*	21,243	AF411846.1	Adam *et al.*, 2006	5
AGAMOUS-Like2-5	*AGL2-5*	7,621	AF411847.1	Adam *et al.*, 2006	3
AGAMOUS-Like6-1	*AGL6-1*	3,388	AY739701.1	Kim *et al., 2013*	1
SQUAMOSA2	*SQUA2*	16,789	AF411841.1	Adam *et al.*, 2007	2
SQUAMOSA3	*SQUA3*	45,214	AF411842.1	Adam *et al.*, 2007	4
DEFICIENS1	*DEF1*	6,675	AY739700.1	Adam *et al.*, 2007	1
sativa MADS	*STMADS11*	12,408	AY739702.1	Adam *et al.*, 2006	1
Elaeis guineensis MADs3	*EgMADS3*	23,699	AJ581468.1	Alwee *et al.*, 2006	4
Elaeis guineensis MADs4	*EgMADS4*	10,014	AJ581469.1	Alwee *et al.*, 2006	4
Elaeis guineensis MADs5	*EgMADS5*	21,243	AJ581470.1	Alwee *et al.*, 2006	4
Elaeis guineensis MADs8	*EgMADS8*	34,002	AJ581461.1	Alwee *et al.*, 2006	5
Elaeis guineensis MADs9	*EgMADS9*	32,775	AJ581462.1	Alwee *et al.*, 2006	5
Elaeis guineensis MADs11	*EgMADS11*	12,587	AJ581465.1	Alwee *et al.*, 2006	5
Elaeis guineensis MADs12	*EgMADS12*	10,403	AJ581463.1	Alwee *et al.*, 2006	5
Elaeis guineensis MADs14	*EgMADS14*	45,214	AJ581466.1	Alwee *et al.*, 2006	4
Elaeis guineensis AD1	*EgAD1*	5,898	AF322914.1	Beulé *et al.*, 2011	1
Elaeis guineensis RING1	*EgRING1*	13,034	GQ914934.1	Beulé *et al.*, 2011	1
Elaeis guineensis FB1	*EgFB1*	9,794	GQ914933.1	Beulé *et al.*, 2011	3
Elaeis guineensis PHOS1	*EgPHOS1*	34,909	GT119837.1	Beulé *et al.*, 2011	1
Flower/fruit development					
GT119161	*GT119161*	643	GT119161.1	Beulé *et al.*, 2011	3
GT119339	*GT119339*	519	GT119339.1	Beulé *et al.*, 2011	3
GT119493	*GT119493*	750	GT119493.1	Beulé *et al.*, 2011	1
GT119578	*GT119578*	1,619	GT119578.1	Beulé *et al.*, 2011	2
GT119987	*GT119987*	570	GT119987.1	Beulé *et al.*, 2011	2
GT120040	*GT120040*	4,339	GT120040.1	Beulé *et al.*, 2011	2
GT120266	*GT120266*	1,642	GT120266.1	Beulé *et al.*, 2011	5
GT120324	*GT120324*	1,538	GT120324.1	Beulé *et al.*, 2011	2
GT120393	*GT120393*	1,020	GT120393.1	Beulé *et al.*, 2011	2
Sex-determination genes					
Tasselseed 1	*TS1*	15,448	FJ360855.1	*Kim et al.*, 2007	8
Anther ear 1	*An1*	26,774	NM_001111859	*Kim et al.*, 2007	8
Alcohol Dehydrogenase1	*Adh1*	19,655	NM_001111939	Kinney *et al.*, 2003	2
indeterminate spikelet1	*Ids1*	19,234	AF048900.1	Chuck *et al.*, 2007	7
APETALA2	*AP2*	6,097	NM_001153337	Chuck *et al.*, 2007	4
Glossy15	*Glossy15*	82,351	U41466.1	Lauter *et al.*, 2005	5
Dwarf1	*D1*	31,277	JX307638.1	*Kim et al.*, 2007	1
Dwarf2	*D2*	31,142	HQ619956.1	*Kim et al.*, 2007	8
Dwarf3	*D3*	37,946	U32579.1	*Kim et al.*, 2007	4
Dwarf8	*D8*	20,442	NM_001137157	*Kim et al.*, 2007	1
Cucumber sativus-1-aminocyclopropane-1-carboxylate	*CS-ACS1*	56,902	AB006803.1	Mathooko *et al.*, 1999	1
Cucumber sativus-1-aminocyclopropane-1-carboxylate	*CS-ACS2*	56,902	D89732.1	Mathooko *et al.*, 1999	3

Phenotypic data including sex ratio (SR), female inflorescence number (FN), male inflorescence number (MN), total inflorescence number (TI), and fresh fruit bunch yield (FFB) described by Ukoskit *et al.* (2014) were used in this study. QTL mapping was performed as reported in Ukoskit *et al.* (2014). Briefly, single marker analysis was performed by using the Kruskal–Wallis nonparametric rank-sum test (KW) and the multiple QTL analysis model (MQM) with MapQTL version 4.0 (Van Ooijen *et al.*, 2002). When determining the genome-wide (αG) and chromosome-wide (αC) LOD threshold for significance ($P < 0.05$), 1,000 permutations were performed to declare the significance level. QTLs were considered significant when the LOD score was above the genome-wide threshold.

RESULTS

ILP marker analysis

A total of 47 EST sequences were aligned with the oil palm genomic sequences. In total, 157 introns were predicted (Table 1): 83 from MADS-box genes, 22 from flower/fruit development genes, and 52 from sex determination genes, with an average of 3.50 introns per EST. For these introns, 139 ILP primers were designed. All ILP primers successfully amplified genomic DNA with products of the expected size, 79 of which were polymorphic in a diversity panel of 41 palms (Table 2). One marker, D8-1, generated more than two bands after PCR optimization, suggesting the possibility of duplicated loci for this genomic region. A total of 254 alleles were detected, ranging from two to seven alleles per locus, with a mean of 3.3. The estimated PIC value ranged from 0.05 to 0.78 with a mean of 0.34, and the observed heterozygosity from 0.02 to 0.97 with a mean of 0.42.

To gain further insight into the utilization of ILP markers, clustering analysis of 41 palms was performed based on the 79 polymorphic ILP primers. The 41 palms were clustered into three major groups (Figure 1) with similarity coefficients ranging from 0.07 to 0.31 and an average of 0.22. Group A comprised Calaba, AVROS, DAMI, and hybrids of Calabar-AVROS, Ghana-AVROS, and La Me-AVROS. Group B comprised Ghana, Nigeria, Tanzania, Yangambi, and hybrids of Ghana-AVROS, La Me-AVROS, and Yangambi-AVROS. Group C comprised of Deli Dura and one accession of AVROS.

Table 2 Characteristics of the 79 polymorphic ILP markers.

Marker	Forward primer (5'-3')	Reverse primer (3'-5')	Tm (◦C)	Size (bp)	NA	H	PIC
GLO1-1	AGCTCCGGCAAGATGTC	AGTTATGCTGGTACCTCTCG	56	157	6	0.37	0.38
GLO1-3	CTGGCAAGAAGCTCTG	TTTCTTGATCCGGTCG	56	160	4	0.48	0.58
GLO1-4	TTTCTTGATCCGGTCG	AAGAAAGAGAATGACAACAT	52	183	2	0.33	0.36
GLO2-1	CAGTGAGTTCAGATCCTC	CCTCCAGAATGGTCTCAT	55.8	162	3	0.30	0.53
GLO2-2	CCTCCAGAATGGTCTCAT	TCTTCTTGAGCTTCTTCA	56	456	2	0.12	0.10
GLO2-3	TCTTCTTGAGCTTCTTCA	AGAGGAGAACAAGCATCTGAC	56	159	3	0.02	0.11
GLO2-4	AGAGGAGAACAAGCATCTGAC	CCAGTTCCCTTACATTTGCATC	58.1	163	2	0.33	0.36
GLO2-5	CCAGTTCCCTTACATTTGCATC	GGGAAGATGTCCGAGTA	56	624	2	0.23	0.18
AG2	CTCCTCCACTTTCCCATCTT	GGCCGATCCAGCGTAG	58.1	256	2	0.51	0.35
AGL2-1-1	ATGCAACTATGGTGCTCCG	GCACGTGCTTTTAGTTTCAAG	56	820	2	0.12	0.11
AGL2-1-2	ACGTGCTGAAGCCTTAC	CTCCTTGCTGCTGAGTG	62	219	2	0.14	0.15
AGL2-1-3	TGCAGATCTTCAACGAAAG	TTTATTGGCCTCACATAGCA	61.2	414	4	0.35	0.29
AGL2-1-4	AGGCCAATAAAAGTCTAAGG	ACTTGCTGCTGACCA	55.8	242	3	0.12	0.25
AGL2-2-1-1	TAAAAGCACGTGTTGAAGC	TGCTGAGTGGACCCAA	58.1	231	2	0.23	0.21
AGL2-2-1-2	TGCTTGATCAGCTTGCAG	TTTTGTTGGCCTCACATAGC	60	414	4	0.15	0.22
AGL2-2-3	ATGTGAGGCCAACAAAAG	ACTTGCTGCTGACCA	58.1	162	7	0.83	0.59
AGL2-3-1	GCACGTGTTGAAGCATTAC	TGCTGAGTGGCCCTA	56	202	5	0.17	0.15
AGL2-4-1	GAAGCTGCTGCTCCG	CTCTTGCCTTCAGCCTCA	56	932	2	0.12	0.11
AGL2-4-3	GCTCGATCAGCTTTGTG	TTTGTTTCCTGCATCTCT	52.7	935	2	0.02	0.07
AGL2-4-4	ATGCAGGAAACAAACAGAT	CAATTGGAGGGGATTTTG	53.9	153	4	0.98	0.44
AGL2-4-5	CAGCCATTGGCATGT	CCTCCGTTTGGCTGG	53.9	165	3	0.84	0.40
SQUA3-1	TATTCAAATGCGGAAAAAGCTC	CAACCTTAGATTTCAGTTCACC	56	186	3	0.09	0.09
SQUA3-2	ATCTAAGGTTGAGGCTCT	GCGCTAAGGAGTCAAGT	56	146	3	0.07	0.07
SQUA3-3	GGCTTGAGAGTGCTTTGAGA	CCGAAGCTCAGCAATTGAATC	56	292	4	0.49	0.34
SQUA3-4	TGCAAGAGCAGAATTGC	TTCGCAGTGTGGATGT	56	166	2	0.19	0.26
EgMADS4-3	TCCATGCAGCCATCTCAAT	AACCCAGCCTAGCATGAA	60	307	3	0.52	0.34
EgMADS5-1	AAGGCAAGAGTTGAGTTTCTG	CTTAGTGGGTCCAAGTCCT	56	228	2	0.28	0.21
EgMADS5-3	CTGCTGGTTCATCAAATGGCCC	GGAAGGTAGCCATTGGAATCGT	56	263	2	0.05	0.05

Table 2 Characteristics of the 79 polymorphic ILP markers. (continued)

Marker	Forward primer (5'-3')	Reverse primer (3'-5')	Tm (◦C)	Size (bp)	NA	H	PIC
EgMADS8-2	ACGTGCTGAAGCCTTAC	AAGCTCCTTGCTGCTGA	56	209	4	0.57	0.52
EgMADS8-3	GGCAACTTGATGCATCGTTA	GCTGATCAAGCATGTATTGG	56	465	3	0.74	0.50
EgMADS8-4	TGTGAGGCCAATAAAAGTCT	GCTCTGTGGTTGAGGTG	56	121	4	0.86	0.49
EgMADS9-3	CATAATCATCAGAAACAGCAGG	GAGCTCCAGTTAGAGCAAAG	56	534	3	0.60	0.41
EgMADS11-3	TAAGTTGTGCAATCTGCTCTC	GAACTGCATGAAATAGAGGGTA	56	174	2	0.38	0.26
EgMADS12-3	TGCAAAGCAATCAGCATTAC	CCAGCTCTACACCTACAG	56	280	2	0.27	0.21
EgMADS14-1	CTTGAACGTTACCGGC	TCACCAAATTCATGTAACCA	56	200	5	0.23	0.21
EgMADS14-3	CTTGAGAGTGCTTTGAGACA	CGAAGCTCAGCAATTGAATC	56	299	4	0.12	0.11
EgMADS14-4	AGAAGTCATTGCAAGAGCAGA	GAAGTGGTGACGAAGAGCTTG	56	128	2	0.23	0.18
EgFB1-3	TTGGTGTCCAAGATTGTCCAT	ATCTCCCATATTTCCTTCTCTCC	56	110	3	0.55	0.40
EgPHOS1-1	TTGTCACAGATGAGCTGC	CCAAGCTATGTGGATGAG	56	338	2	0.10	0.09
GT119161-2	ATGCGAACAAAATCGGACA	CTTATTGAGAGCCTCTTCACAG	56	212	3	0.29	0.24
GT119161-3	TGCTTATAAGAGCAAGAGAACC	CCCTCGACCAATTCTTTTGA	56	747	3	0.71	0.44
GT119339-1	CTCCTTTTTCTGTGGGTCGG	TTCTCAAGCTCGGGCGG	56	226	3	0.49	0.36
GT119493-1	GGGGAAAAGAAAACACAGAG	TTCACGGGACGATTCGG	56	692	4	0.81	0.52
GT119578-1	GCCAACCTCCCTTCCC	TCACAGACATAAACCCGGA	60	236	2	0.19	0.16
GT119987-1	GGTCAGAAAGAGTAAAATCTGTAG	GCAGACACTTCCGGAATAA	56	220	6	0.53	0.41
GT119987-2	GGTAAAATCTTTTGAGGTCACAG	ATATCATCTGCCCCAGCC	56	160	2	0.33	0.24
GT120040-1	GAGCCGATCGTATGCCT	CGCGGCTGTCCTTTACT	56	209	2	0.32	0.23
GT120266-1	AGAAACTTCCATCACTGAGA	TGCATCAAGGCTCGTCA	56	360	4	0.36	0.30
GT120266-2	CCCAAGGTAATCCGATCCTC	GCAGTTCCTAGTGGAGCA	56	482	4	0.95	0.66
GT120266-3	AAAAGTATCAATCTGAGCCTGC	AGTATCATTGGACCAGCATTG	56	200	2	0.57	0.36
GT120266-4	CAGTCAATATTGCATTTGCACC	AACTTGATGGGACCTCTAATG	56	197	2	0.60	0.37
GT120324-1	AGCATTGAAGACAAGGAAGC	ACATTCCGCCAGCTCTT	56	289	3	0.45	0.33
GT120324-2	TCATCACCTCCACCAACA	TGCCTGCTGGGAACTT	56	954	4	0.67	0.49
GT120393-1	AGTCCGTCTTCGTTCTGTG	CTCGATCTCCCTCATCGG	56	200	2	0.74	0.36
GT120393-2	CGTCCTCAACATCTCCGT	TGCATCGCCTTTTCACC	56	738	7	0.77	0.76
Ts1-1	CGATCTCATCAGAAGGCAAG	AATTGGGTGATCATTTGAAGT	53.9	363	4	0.57	0.40
Ts1-2	TCTACGGTCCGCCTGAATCT	CCGGTCAACAAATGGGAGGT	53.9	319	3	0.38	0.48
Ts1-3	CCATGCCTAACATGGCACTT	AACTCATCGTCGCGCAGC	56	581	3	0.24	0.41
Ts1-4	ATGATACCCCTGACATCATA	AGCAGTTGCAATCACCC	56	278	5	0.80	0.54
Ts1-5	GATTGATCAAGTCTAATCCTTC	AATGGATCCTCCGCTC	56	252	2	0.21	0.17
Ts1-7	GGGTCCAACCCACCAG	GCCCCTTAGCTCCTTCAA	56	336	3	0.40	0.31
An1-1	GCGTCGTCGCGCTAA	GAAGGGAAGGCGATCTCA	56	200	2	0.20	0.37
An1-3	GAGCAACGAGAGGTGCCTCAA	CAGCCCACATGTGCTCGAA	56	186	2	0.26	0.20
An1-4	CCTTGGAATCTCCCGGTAT	CTAGAATTCCTTGCCCAACAA	56	184	7	0.64	0.78
An1-5	TGTGGATGACACGGC	TCCCAGTGACTGCTTG	52	251	2	0.70	0.35
An1-6	GCGGATGACGTTTGGATT	GGCAAGCTCCAGATACAC	56	159	4	0.55	0.41
An1-7	CAGGCACTCCATCAGGT	CGGCTGCCAGAAAATAGG	61	233	5	0.38	0.61
Adh1-1	GACCAGACGCTGCTACA	GCATTGGAAATGTTTCCGTC	56	407	4	0.21	0.19
Ids1-1	GCAAGTTAAGAAGAGCAGAAG	ATCGCAGCTCGATCATA	56	455	2	0.44	0.28
Ids1-2	GAGTTGACGCTGACATCAAT	GGTCACCCCTCTGTATTTTG	56	243	4	0.24	0.21
Ids1-3	CGCATGGGACAATTCCTT	GGGCTCAAAGTTGGTAAC	56	510	3	0.18	0.28
Ids1-4	AAGGAGAGTTGCTTACTGAG	GAATTGGATGCCCATTGA	56	217	5	0.55	0.42
AP2-1	ATCGCCGCTCGATCA	AATACCGTGGGGTCACT	62	514	2	0.57	0.32
AP2-2	CCTAGAGCTTCCTCTTGGAAAT	GATGATTATGAGGAGGATCTGAA	56	205	4	0.95	0.60
Glossy15-1	TCGAGAAGTTCCCAGTACC	CCCCACGAAACTTGATCG	56	588	3	0.50	0.45
Glossy15-2	TGGATGATTATGAGGAGGATCT	CGTCTAAGTACATGGACAAAT	56	181	4	0.76	0.53
D2-6	CCCATTTTTGTATCCTGAACC	CCTCCTCCACCAAACATCAT	56	207	4	0.39	0.31
D8-1 (1)	GAGCATGCTGTCCGAG	AGAGGAGGAAGAGGAGG	52	329	3	0.32	0.56
D8-1 (2)	GAGCATGCTGTCCGAG	AGAGGAGGAAGAGGAGG	52	329	2	0.44	0.29
Mean				322.9	3.3	0.42	0.34
Maximum				954	7	0.97	0.78
Minimum				110	2	0.02	0.05

N: Number of alleles, H: Observed heterozygosity, PIC: Polymorphism information content

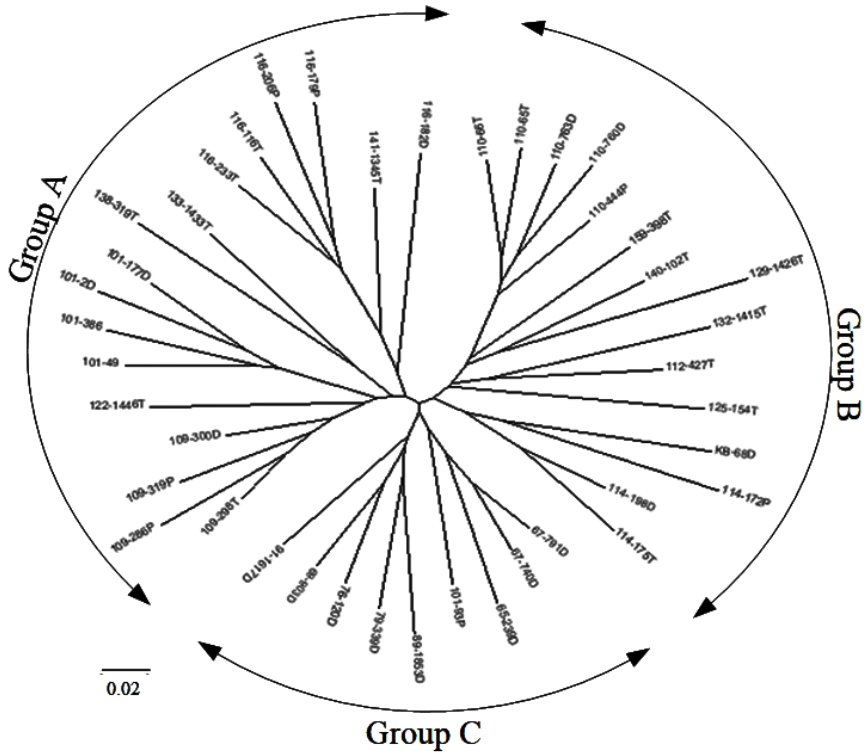

Figure 1 Dendrogram of the genetic similarities among 41 oil palm genotypes.

Linkage mapping and QTL analysis

For linkage analysis, the 79 polymorphic ILP primers were tested for polymorphisms between the two parents and ten random individuals from 208 progenies. Eleven ILP markers developed from MADS-box genes, three from flower/fruit development genes, and eight from sex determination genes were found to be polymorphic on the mapping population. Of these, two markers segregating with one allele with 1B and 1X cross-type configuration were identified for both configurations (Table 3). Eight, one, and eight markers were observed to segregate with the cross-type configurations 2B, 2D, and 2X, respectively. One marker segregating with a 3X cross-type configuration was detected. Thirteen of 22 ILP markers were placed on the framework map. The eight markers developed from MADS-box genes were distributed across four LGs: 1XAGL2-2-3 on LG3, 2BAGL2-2-1, 2XGLO1-4, and 2XGLO2-4 on LG4, 2XAGL2-4-2, and 2XEgMADS5-1 on LG8, and 2DSQUA3-1, and 1BEgMADS14-1 on LG16 (Figure 2). 2BGT119161-3 and 2XGT120266-1, two of the markers developed from flower/fruit development genes, were placed on LG8 and LG11, respectively. The three markers 2BAn1-5, 1XAn1-3, and 2XD8-1, developed from sex determination genes, were mapped onto LG3, LG11, and LG12, respectively.

Table 3 Cross-type configurations of ILP markers used for integrating in the published oil palm genetic map (Ukoskit *et al.*, 2014).

Segregating alleles	Cross-type configurations[a]	Segregation ratio	Number of segregating markers			Total
			MADS-box	Flower/fruit development	Sex-determination	
1 allele	1B	1:1	1	-	1	2
	1X	3:1	1	-	1	2
	2B	1:1	3	1	4	8
2 alleles	2D	1:1	-	1	-	1
	2X	1:2:1	6	-	2	8
3 alleles	3X	1:1:1:1	-	1	-	1

[a] Number 1, 2, or 3, represents the number of segregating alleles. Letter B or D indicates the parental polymorphism origin. Letter X indicates loci heterozygous in both parents

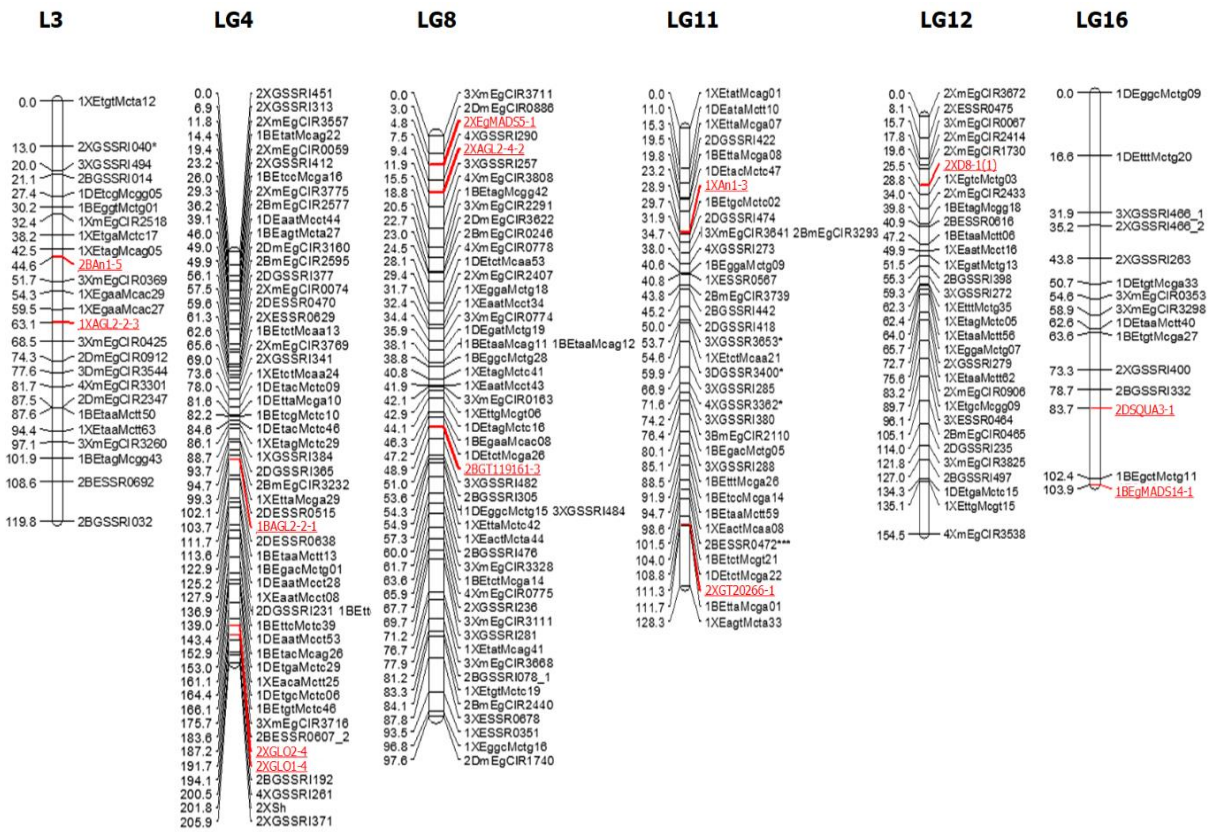

Figure 2 Integration of ILP markers into the published oil palm genetic map (Ukoskit *et al.*, 2014). The cumulative distances in centiMorgans (Kosambi) are given at the left of the bar. Newly-added ILP markers are underlined.

To determine the physical positions of the mapped ILP markers on the oil palm reference genome (Singh *et al.*, 2013), the chromosomal locations of the mapped ILP markers were identified using the BLAST program, with the marker sequences as query and the oil palm reference genome as subject. The ILP markers assigned to LGs in this work were consistent with chromosomes in the oil palm genome (Table 4). The order of the ILP markers was in broad agreement with their physical order along the corresponding chromosomes.

Table 4 The physical locations of the 13 mapped ILP markers in the oil palm genome (Singh *et al.*, 2013).

Marker	Linkage group	Chromosome	Physical position	NCBI accession
2BAn1-5	3	3	51,246,918	NC_025995.1
1XAGL2-2-3	3	3	43,760,713	NC_025995.1
2BAGL2-2-1	4	4	5,550,107	NC_025996.1
2XGLO2-4	4	4	1,472,867	NC_025996.1
2XGLO1-4	4	4	17,595,606	NC_025996.1
2XEgMADS5-1	8	8	20,541,062	NC_026000.1
2XAGL2-4-2	8	8	4,107,879	NC_026000.1
2BGT119161-3	8	8	50,804	NC_026000.1
1XAn1-3	11	11	9,085,887	NC_026003.1
2XGT120266-1	11	11	4,309,227	NC_026003.1
2XD8-1	12	12	22,667,273	NC_026004.1
2DSQUA3-1	16	16	8,156,546	NC_026008.1
1BEgMADS14-1	16	16	7,080,200	NC_026008.1

QTL analysis

As the goal of the study was to identify ILP markers linked to QTL for oil yield-related traits, only ILP markers were presented for QTL analysis. The other marker types related to QTLs were reported in a previous study (Ukoskit *et al.*, 2014). QTLs of all traits were detected using the Kruskal–Wallis test with a significance level of 0.01. In total, four ILP markers were detected for QTLs of the three traits studied. 2XD8-1, located on LG12 was associated with MN, 2BGT119161-3, and 2DSQUA3-1 located on LG8 and LG16 respectively, were associated with TI, and 1XAGL2-2-3, located on LG3, was associated with FFB (Table 5). MQM analysis confirmed 1XAGL2-2-3 genome-wide significance QTL for FFB. The LOD score value was 7.71, with phenotypic variation explaining 11.60%.

Table 5. Identified QTLs for sex ratio-related traits and fresh fruit bunch yield.

I. Results of the QTL analysis based on Kruskal–Wallis analysis test with a significance level of 0.01.

Traits[a]	LG[b]	Marker	Position (cM)	B x D	Genotype class means		
MN	12	2XD8-1	25.5	ab x ab	aa: 10.37	ab: 11.81	bb: 9.73
TI	8	2BGT119161-3	49.7	ab x aa	aa: 29.91	ab: 28.12	
	16	2DSQUA3-1	83.7	aa x ab	aa: 28.43	ab: 29.38	
FFB	3	1XAGL2-2-3	61.6	ao x ao	oo: 6.5	a-: 7.31	

II. Results of the QTL analysis based on Multiple-QTL Mapping.

Trait[a]	LG[b]	Marker	LOD	PVE[c]	LOD αC	LOD αG	Genotype class means			
							ac	ad	bc	bd
FFB	3	1XAGL2-2-3	7.71	11.60%	3.1	4.6	6.77	5.64	4.84	5.55

[a] MN: male inflorescence number, TI: total inflorescence number, FFB: fresh fruit bunch yield; [b] LG: linkage group; [c] PVE: phenotypic variation explained.

DISCUSSION

ILP has been widely used to amplify introns in conserved genes. When designing the PCR primers, the oil palm EST sequences of the *MAS* box genes and genes derived from transcriptome analysis of mantled inflorescence were chosen, while the EST sequences of sex determination genes were taken from two other species, maize, and cucumber (Table 1). We successfully amplified genomic DNA with the expected size products of between 110 and 954 bp, using all designed primers. The 100% PCR success rate of ILP primers in this study was greater than those reported for other plants: cowpea (89 %) (Gupta *et al.*, 2012), maize (88.4%) (He *et al.*, 2015), soybean (88.2 %) (Shu *et al.*, 2010), and rubber (70%) (Li *et al.*, 2012). PCR amplification based on primers developed from oil palm EST sequences was expected to have a high success rate. However, the 100% PCR success rate of sex determination ILP primers developed from the other two species suggested that the genes involved in sex determination had conserved intron positions. The conservation of genes related to the flower-formation regulatory pathways is assumed to reflect the functional importance of flowering.

The ILP markers developed in the present study have great potential for investigating genetic diversity in oil palm. Of the total 139 ILP markers tested on the 41 oil palm genotypes, 57.6% produced length polymorphisms greater than ILP markers in cowpea (36%) (Gupta *et al.*, 2012), maize (53.5%) (He *et al.*, 2015), and foxtail millet (45%) (Muthamilarasan *et al.*, 2014). The mean observed heterozygosity (0.42) was higher than that of ILP markers in foxtail millet (0.16) (Gupta *et al.*, 2011) and comparable with rice (0.43) (Wang *et al.*, 2005). The average PIC value (0.33) of the polymorphic markers was higher than that of ILP markers in foxtail millet (0.20) (Muthamilarasan *et al.*, 2014), and rubber (0.22) (Li *et al.*, 2012), and similar to that of cowpea (0.34) (Gupta *et al.*, 2012), but lower than that reported in rice (0.43) (Huang *et al.*, 2010). These variations in the detection rate of polymorphism may be due to differences in the number of genotypes and their genetic background, and the number of markers used. The clustering analysis of oil palm using the polymorphic ILP markers was consistent with their pedigree and geographical origin, which suggests that the ILP markers developed in this study are effective tools for the analysis of genetic relationships in oil palm. They may help oil palm breeders to identify similarities across their collections, develop strategies for field conservation, and select parents for hybridization.

The development of genic sequencing has allowed gene-associated sequences to be used as

genetic markers. This technique provides tools for the identification of the genes conferring traits of interest, and integrates our understanding of genes and QTL. PCR-based, co-dominant gene-associated markers have been used to improve the efficiency of genetic mapping (Deng and Davis, 2001). In the current study, mapping of candidate genes as intron length polymorphisms was made possible by the use of bi-parental crossing. The results integrated 13 new ILP markers into the existing linkage map (Ukoskit *et al*., 2014). Although these markers have not significantly increased the overall map length, the integration of gene-associated markers may open up new avenues for further saturation of the map, using functional makers not only for a wider range of oil palm developmental genes, but for other genes as well.

This study was based on a hypothesis that genes that play a role in flower-formation regulatory pathways, including *MADS*-box genes, sex determination genes, and genes derived from transcriptome analysis of mantled inflorescences, might influence palm oil yield. To test this hypothesis, 11 candidate genes were mapped on the basis of PCR-detectable intron length polymorphism. By mapping gene-associated markers rather than anonymous DNA markers, we were able to determine the genomic location of the *MADS*-box gene *AGAMOUS-Like2-2* for QTL related to oil palm yield within the oil palm genome. *AGAMOUS* determines stamen and carpel identity and later controls senescence of the flower (Jibran *et al.*, 2017). Senescence functions to transfer nutrients from dying parts of the plant to the flowers, fruit, and grains (Gan and Amasino, 1997), and is associated with yield.

To the best of our knowledge, this is the first study to report the development of gene-based markers as ILP, from conserved DNA sequences of flower developmental genes and to apply these markers to genetic diversity, linkage mapping and QTL mapping for oil palm yield. The markers have potential applications in the characterization of the oil palm germplasm. ILP markers representing flower developmental genes were mapped and localized on an existing oil palm genetic map. Broad agreement was found in the physical order along the chromosomes. The linkage map incorporating ILP markers was used to identify *AGAMOUS* of a *MAS* box gene associated with a QTL for FFB. We demonstrated that the identification of ILPs representing gene-specific markers on a genetic map can be used to enhance the efficiency of QTL mapping. Our finding will contribute to oil palm breeding programs and help increase palm oil production.

ACKNOWLEDGEMENTS

The authors thank the Faculty of Science and Technology, Thammasat University, for financial support.

REFERENCES

Adam H, Jouannic S, Morcillo F, Richaud F, Duval Y, Tregear JW (2006) *MADS* box genes in oil palm (*Elaeis guineensis*): the *SQUAMOSA*, *DEFICIENS*, *GLOBOSA*, *AGAMOUS*, and *SEPALLATA* subfamilies. J Mol Evol 62:15–31.

Adam H, Jouannic S, Orieux Y, Morcillo F, Richaud F, Duval Y, Tregear JW (2007) Functional characterization of MADS box genes involved in the determination of oil palm flower structure. J Exp Bot 58: 1245–1259.

Alwee SS, Van der Linden CG, Van der Schoot J, de Folter S, Angenent GC, Cheah S-C, Smulders MJM (2006) Characterization of oil palm MADS box genes in relation to the mantled flower abnormality. Plant Cell Tissue Organ Cult 85: 331-344.

Andersen JR, Lübberstedt T (2003) Functional markers in plants. Trends Plant Sci. 8: 554–560.

Benbouza H, Jacquemin JM, Baudoin JP, Mergeai G (2006) Optimization of a reliable, fast, cheap and sensitive silver staining method to detect SSR markers in polyacrylamide gels. BASE 10: 77–81

Beulé T, Camps C, Debiesse S, Tranchant C, Dussert S, Sabau X, Jaligot E, Alwee SSRS, Tregear JW (2011) Transcriptome analysis reveals differentially expressed genes associated with the mantled homeotic flowering abnormality in oil palm (*Elaeis guineensis*). Tree Genet Genomes 7: 169–182.

Brendel V, Xing L, Zhu W (2004) Gene structure prediction from consensus spliced alignment of multiple ESTs matching the same genomic locus. Bioinformatics 20: 1157–1169.

Choi H-K, Kim D, Uhm T, Limpens E, Lim H, Mun J-H, Kalo P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss GB, Cook DR (2004) A sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. Genetics 166: 1463–1502.

Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. Nature 353: 31.

Corley RHV (1976) Oil palm yield components and yield cycles. In: International Developments in Oil Palm. Incorporated Society of Planters,. Kuala Lumpur, Malaysia, pp 116–129

Corley RHV, Tinker PB (2007) The establishment of oil palms in the field. In: Corley RHV, Tinker, P.B.

(ed) The Oil Palm, 4 th Edition. Wiley-Blackwell, Oxford.

Deng C, Davis TM (2001) Molecular identification of the yellow fruit color (c) locus in diploid strawberry: a candidate gene approach. Theor Appl Genet 103: 316–322.

Gan S, Amasino RM (1997) Making sense of senescence (molecular genetic regulation and manipulation of leaf senescence). Plant Physiol 113: 313–319.

Gawel NJ, Jarret RL (1991) A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. Plant Mol Biol Rep 9: 262–266.

Gupta S, Kumari K, Das J, Lata C, Puranik S, Prasad M (2011) Development and utilization of novel intron length polymorphic markers in foxtail millet (*Setaria italica* (L.) P. Beauv.). Genome 54: 586–602.

Gupta SK, Bansal R, Gopalakrishna T (2012) Development of intron length polymorphism markers in cowpea [*Vigna unguiculata* (L.) Walp.] and their transferability to other *Vigna* species. Mol Breed 30: 1363–1370.

He C, Liu H, Su S, Lu Y, Luo B, Nie Z, Wu L, Liu D, Zhang X, Rong T, Gao S, Léon J (2015) Genome wide identification of candidate phosphate starvation responsive genes and the development of intron length polymorphism markers in maize. Plant Breed 134: 11–16.

Huang M, Xie F-m, Chen L-y, Zhao X-q, Jojee L, Madonna D (2010) Comparative analysis of genetic diversity and structure in rice using ILP and SSR markers. Rice Sci 17: 257–268.

Jaligot E, Adler S, Debladis É, Beulé T, Richaud F, Ilbert P, Finnegan EJ, Rival A (2011) Epigenetic imbalance and the floral developmental abnormality of the in vitro-regenerated oil palm *Elaeis guineensis*. Ann Bot 108: 1453–1462.

Jibran R, Tahir J, Cooney J, Hunter DA, Dijkwel PP (2017) Arabidopsis AGAMOUS regulates sepal senescence by driving jasmonate production. Front Plant Sci 8: 2101.

Li C, Riethoven J-JM, Ma L (2010) Exon-primed intron-crossing (EPIC) markers for non-model teleost fishes. BMC Evol Biol 10: 90.

Li D, Xia Z, Deng Z, Liu X, Dong J, Feng F (2012) Development and characterization of intron-flanking EST-PCR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). Mol Biotechnol 51: 148–159.

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21: 2128–2129.

Muthamilarasan M, Venkata Suresh B, Pandey G, Kumari K, Parida SK, Prasad M (2014) Development of 5123 intron-length polymorphic markers for large-scale genotyping applications in foxtail millet. DNA Res 21: 41–52.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A 76: 5269–5273.

Poczai P, Cernák I, Varga I, Hyvönen J (2014) Nuclear intron-targeting markers in genetic diversity analysis of black nightshade (Solanum sect. Solanum, Solanaceae) accessions. Genet Resour Crop Evol 61: 247–266.

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

Rohlf FJ (1998) Numerical Taxonomy and Multivariate Analysis System version 2.0. Exeter Software, Setauket, New York,

SaiSug W, Ukoskit K (2013) Comparative analysis of EST-derived markers for allelic variation in *Jatropha curcas* L. and cross transferability among economically important species of *Euphorbiaceae*. Genes Genomics 35: 1–12.

Sheikh I, Sharma P, Verma SK, Kumar S, Kumar N, Kumar S, Kumar R, Vyas P, Dhaliwal HS (2018) Development of intron targeted amplified polymorphic markers of metal homeostasis genes for monitoring their introgression from *Aegilops* species to wheat. Mol Breed 38: 47.

Shearman JR, Jantasuriyarat C, Sangsrakru D, Yoocha T, Vannavichit A, Tragoonrung S, Tangphatsornruang S (2013) Transcriptome analysis of normal and mantled developing oil palm flower and fruit. Genomics 101 : 306–312.

Shu Y, Li Y, Zhu Y, Zhu Z, Lv D, Bai X, Cai H, Ji W, Guo D (2010) Genome-wide identification of intron fragment insertion mutations and their potential use as SCAR molecular markers in the soybean. Theor Appl Genet 121: 1–8.

Singh R, Ong-Abdullah M, Low E-TL, Manaf MAA, Rosli R, Nookiah R, Ooi LC-L, Ooi SE, Chan K-L, Halim MA, Azizi N, Nagappan J, Bacher B, Lakey N, Smith SW, He D, Hogan M, Budiman MA, Lee EK, DeSalle R, Kudrna D, Goicoechea JL, Wing RA, Wilson RK, Fulton RS, Ordway JM, Martienssen RA, Sambanthamurthi R (2013) Oil palm genome

sequence reveals divergence of interfertile species in Old and New worlds. Nature 500: 335.

Ukoskit K, Chanroj V, Bhusudsawang G, Pipatchartlearnwong K, Tangphatsornruang S, Tragoonrung S (2014) Oil palm (*Elaeis guineensis* Jacq.) linkage map, and quantitative trait locus analysis for sex ratio and related traits. Mol Breed 33: 415–424.

Van Ooijen J, Voorrips R (2000) JoinMap v. 3, software in the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands,

Van Ooijen JW, Boer MP, Jansen RC, Maliepaard C (2002) MapQTL 4.0, software for the calculation of QTL positions on genetic maps. Plant Research International, Wageningen, The Netherlands,

Wang X, Zhao X, Zhu J, Wu W (2005) Genome-wide investigation of intron length polymorphisms and their potential as molecular markers in rice (*Oryza sativa* L.). DNA Res 12: 417–427.

Yang L, Jin G, Zhao X, Zheng Y, Xu Z, Wu W (2007) PIP: a database of potential intron polymorphism markers. Bioinformatics 23: 2174–2177.