BIOTROPIA Vol. 24 No. 2, 2017: 153 - 160 DOI: 10.11598/btb.2017.24.2.678

AN SNP MARKER POTENTIALLY LINKED TO SOMATIC EMBRYOGENESIS OF OIL PALM (*Elaeis guineensis*)

IRANG WAHYUNANTO[1,2*], DIANA E WATURANGI[1], NURITA TORUAN-MATHIUS[2] and ADI YULANDI[1]

[1]*Faculty of Biotechnology, Atma Jaya Catholic University of Indonesia, Jakarta 12930, Indonesia*
[2]*Plant Production and Biotechnology Division, PT SMART Tbk., Jakarta 10350, Indonesia*

Received 28 December 2016/Accepted 14 April 2017

ABSTRACT

Oil palm (*Elaeis guineensis*) is one of the most important oil-bearing crop in the world. This crop can be vegetatively propagated only using tissue culture technique. Oil palm tissue culture technique has low efficiency, with callogenesis and embryogenesis stages as the limiting factors. Genetic factor has a major role in determining the success rate of these two stages. The use of molecular markers which represent the rate of embryogenesis or callogenesis has the potential to improve the efficiency of oil palm tissue culture process. In this study, SNP mining was conducted on embryogenesis transcriptome data, oil palm cDNA database, oil palm genome database, and oil palm SNPs marker database in NCBI. The objective of this study was to obtain SNP marker which represents the embryogenesis potential, to be further used in marker assisted selection of oil palm ortets. One SNP (EMB6) showed significant association with embryogenesis rate. This SNP was found in one of Auxin Response Factor (ARF) family gene. Nucleotide replacement from Adenine to Guanine changed the 307th amino acid from Isoleucine to Methionine. Oil palms with Adenine homozygote (A/A) pattern on the EMB6 showed 8-fold higher chance to produce significantly higher embryogenesis rate than Adenine-Guanine heterozygote (A/G).

Keywords: Callogenesis, *Elaeis guineensis*, SNP, somatic embryogenesis

INTRODUCTION

Oil palm (*Elaeis guineensis*) is among the most important crop which produces vegetable oil. Vegetative propagation for oil palm can only be conducted using tissue culture methods to obtain highly productive oil palm seeds (Wong *et al.* 1997). Tissue culture technique is beneficial for individual propagation of oil palm having high productivity and/or certain trait of interest. The tissue culture technique can increase oil palm production by 30% compared to the commercial seed Dura X Pisifera (DXP) on a large scale field trials (Cochard *et al.* 1999; Wahid *et al.* 2005). Tissue explants of oil palm resulted from *in vitro* culture are developed through callus formation or normally termed as indirect embryogenesis (Te-chato & Hilae 2007). However, not all of the cultured explants have the potential to be developed into embryogenic callus. Those explants which are not successfully developed into embryogenic callus usually form a non-embryogenic callus, that will remain in the form of callus without any possibility to develop into ramet (Rohani *et al.* 2000).

Corley and Tinker (2003) reported that the level of callogenesis in oil palm is still low (around 19%), while the ability to form somatic embryoid is only 3 - 6% (Wooi 1995). One of the main factors that determine the ability of oil palm tissue culture is the genetic factor, which is indicated by the fact that some genotypes are more productive than the others (Wooi 1995).

Identification of Single Nucleotide Polymorphism (SNP) in oil palm genome had been attempted in several studies. Identification of Quantitative Trait Loci (QTLs) associated with callogenesis and embryogenesis in oil palm with broad range of markers from AFRLPs, RFLPs, and SSRs has been done (Ting *et al.* 2013), but studies to determine the process of de-differentiation and embryogenesis with

* Corresponding author: irangwahyunanto@hotmail.com

specific SNP markers in plant genomes are limited. In this study, the identification of candidate SNP associated with somatic embryogenesis in oil palm were done *in silico*, which was continued to the validation process based on the phenotype data of oil palm tissue culture. This study was aimed to obtain SNP marker candidates which represents the embryogenesis potential, to be further used in marker assisted selection of oil palm ortets, and thus increase the tissue culture process efficiency.

MATERIALS AND METHODS

In silico Selection of SNPs Marker

Target genes were derived from Expressed Sequence Tags (ESTs) from Low *et al.* (2008), and Lin *et al.* (2009) which were deposited in GenBank database EY396120-EY413718 and GH635901-GH637767. ESTs were subsequently assembled using CAP3 program (Huang & Madan 1999). Contigs and singletons were annotated to oil palm cDNA database from MPOB's research (genomsawit.mpob.gov.my) (Singh *et al.* 2013), using the BLAST tool (Altschul *et al.* 1990) to obtain a list of genes that was expressed during the process of embryogenesis.

The candidate genes were further aligned to the oil palm genome sequence (genomsawit.mpob.gov.my) (Singh *et al.* 2013) to get the full-length gene sequence. The alignment process was done using Sim4db program (Walenz & Florea 2011). The full-length gene sequence was generated using GetFastaBed tool in Galaxy platform (Quinlan & Hall 2010). Then, the full-length gene sequences were used as the database for BLAST alignment of oil palm SNP sequences obtained from MPOB (genomsawit.mpob.gov.my) (Ting *et al.* 2014), and NCBI (Teh *et al.* 2016) as the queries (Fig. 1).

Ten candidate SNP markers were selected based on the function, expression and sensitivity of the mutation position. DNA fragment sequences were aligned using Unipro UGENE (Okonechnikov *et al.* 2012). Primers of each SNP marker were designed using Primer-BLAST tool from NCBI, which combined the algorithm of Primer3 (Untergrasser *et al.* 2012) and NCBI BLAST. Primer quality was analyzed and selected using NetPrimer Tools from Premier Biosoft.

SNPs Marker Validation

Thirty ortets were kindly provided by the Tissue Culture Laboratory of PT SMART Tbk. These ortets were selected based on the tissue culture productivity data, including callogenesis and embryogenesis rate. Genomic DNA were isolated and purified from leaf tissue using NucleoSpin Plant II kit (Macherey-Nagel GmbH & Co KG, Duren, Germany). The quality of extracted DNA was measured using 1% agarose gel electrophoresis and nanodrop spectrophotometer (Thermoscientific, Massachusetts, USA). The DNA was amplified by

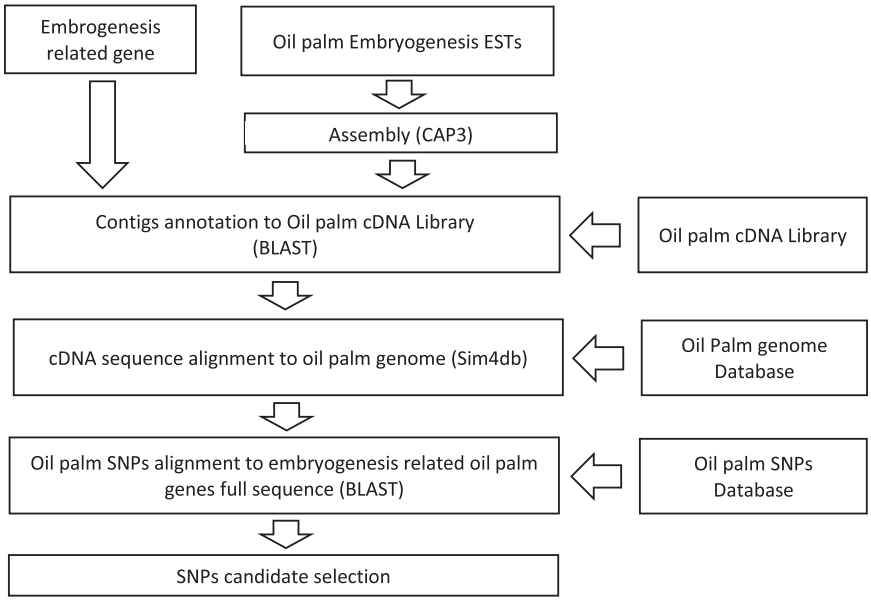

Figure 1 Oil palm embryogenesis SNP mining bioinformatics workflow

designed primers using PCR. The PCR products were confirmed using agarose gel electrophoresis.

PCR products were purified using QIAquick PCR Purification Kit (Catalog No.28104, QIAGEN, Hilden, Germany). Purified PCR products were sequenced, then statistically analyzed and compared with the callogenesis and embryogenesis productivities data using SPSS 20.0 (SPSS Inc., Chicago, USA). For odds ratio analysis, tissue culture productivity data (callogenesis and embryogenesis rate) were converted into categorical data (low callogenic, high callogenic, low embryogenic and high embryogenic), then compared with nucleotide variation using n Cochran's and Mantel-Haenszel cross tabulation statistics in SPSS. The tissue culture productivity data variation was analyzed with One-way ANOVA.

RESULTS AND DISCUSSION

A total of 19,471 embryogenesis related ESTs were obtained from Low *et al.* (2008) and Lin *et al.* (2009). The ESTs libraries were assembled using CAP3 program resulting to 13,020 sequences of contigs and singletons (Table 1).

Contigs and singletons from assembly process were aligned to two sets of MPOB's cDNA database (genomsawit.mpob.gov.my), namely V1 and V2 (Table 2).

The genes were selected based on its correlation to embryogenesis function via text mining methods. Selection of the embryogenesis related gene candidates' selection was according to Elhiti *et al.* (2010), resulting in 423 embryogenesis related genes. These genes only work as reference for this study. There might still be other genes involved in embryogenesis due to the complexity of embryogenesis process. The sequences were aligned to MPOB's genome database using Sim4db program to locate the index position of the intact gene (intron+exon). The index position in gff3 format was then used as input to generate the intact gene's FASTA sequence.

The whole genomic sequence of candidate genes was used as the database to align the oil palm SNP markers. From MPOB's 1,766 SNP positions, there were 12 SNPs with positive hit to target genes, 10 SNPs were in introns, while the other two were synonymous codon variant in exons. In NCBI, which has 112,360 SNP positions, there were 1,575 positive hits to target

Table 1 ESTs assembly results

Source	Subjects	ESTs	Contigs + singletons
Low *et al.* (2008)	NEC	6,498	3,760
	EC	2,717	2,130
Lin *et al.* (2009)	EMB	8,389	5,456
	Initiation	949	854
	Proliferation	918	820

Note: Source of ESTs: Non-Embryonic Callus (NEC), Embryogenic Callus (EC), Embryoid (EMB), Embryoid Initiation and Embryoid Proliferation

Table 2 Embryogenesis related candidate genes number

Subjects	Contigs + singletons	V1 (genes)	V2 (genes)
NEC	3,760	2,269	1,841
EC	2,130	1,085	821
EMB	5,456	3,640	3,074
Initiation	854	530	437
Proliferation	820	561	506

Note: Source of ESTs: Non-Embryonic Callus (NEC), Embryogenic Callus (EC), Embryoid (EMB), Embryoid Initiation and Embryoid Proliferation

Table 3 SNPs with positive hits to target genes

SNPs database	Functional Consequence	Amount (Hits)
MPOB	Intron variant	10
MPOB	Synonymous codon	2
NCBI	Intron variant	19
NCBI	Downstream variant	61
NCBI	Upstream variant	99
NCBI	3'UTR variant	36
NCBI	5'UTR variant	11
NCBI	Splice-donor variant	1
NCBI	Synonymous codon	34
NCBI	Missense variant	53
NCBI	Undefined	1,261

Table 4 SNPs marker for oil palm embryogenesis

SNPs	Gene	Amino acid variation	Used for Further Sequencing
EMB1	*Elaeis guineensis* E3 ubiquitin-protein ligase RING1-like, transcript variant X2, mRNA	Leu → Ser	Yes
EMB2	*Elaeis guineensis* auxin-responsive protein IAA10-like, mRNA	Arg → Ser	Yes
EMB3	*Elaeis guineensis* ethylene-responsive transcription factor 1-like, transcript variant X2, misc_RNA	Ile → Val	Yes
EMB4	*Elaeis guineensis* protein phosphatase 2C and cyclic nucleotide-binding/kinase domain-containing protein, transcript variant X5, misc_RNA	Thr → Ala	Yes
EMB5	*Elaeis guineensis* probable WRKY transcription factor 70, mRNA	Glu → Gly	Yes
EMB6	*Elaeis guineensis* auxin response factor-like, mRNA	Ile → Met	Yes
EMB7	*Elaeis guineensis* cytochrome P450 85A1-like, transcript variant X3, mRNA	Pro → Ser	Yes
EMB8	*Elaeis guineensis* coatomer subunit beta-1-like, mRNA	Ile → Leu → Val	No
EMB9	*Elaeis guineensis* coatomer subunit beta-1-like, mRNA	Pro → Ala → Ser	No
EMB10	*Elaeis guineensis* DNA-binding protein BIN4, transcript variant X4, mRNA	Thr → Pro → Ala	Yes
EMB11	*Elaeis guineensis* auxin response factor-like, transcript variant X1, mRNA	Met → Val → Leu	Yes
EMB12	*Elaeis guineensis* probable cellulose synthase A catalytic subunit 1 [UDP-forming], mRNA	Ile → Val	No
EMB13	*Elaeis guineensis* glutathione S-transferase zeta class-like, mRNA	Asp → His	Yes

Table 5 Statistical result of SNP variation and embryogenesis rate

Gene	Code	SNP*)	Indels**)	Cross Tab Chi Square	One Way ANOVA	Odds Ratio
E3 ubiquitin protein ligase RING 1	EMB1	0	No	Independent	Not significant	n/a
Auxin responsive protein IAA 10	EMB2	8	No	Independent	Not significant	n/a
Ethylene Responsive Transcription Factor	EMB3	1	No	Independent	Not significant	n/a
PP2C	EMB4	1	Yes	Independent	Not significant	n/a
WRKY transcription factor	EMB5	4	Yes	Independent	Not significant	n/a
Auxin response factor A	EMB6	3	Yes	Dependent (p = 0.01)	Significant (p = 0.015)	AA = 8x AG (p = 0.014)
Cytochrome P450 85 A 1 like	EMB7	0	No	Independent	Not significant	n/a
DNA binding protein BIN 4	EMB10	1	No	Dependent (p = 0.04)	Not significant	n/a
Auxin response factor B	EMB11	1	No	Independent	Not significant	n/a
Glutathione S Transferase	EMB13	1	Yes	Independent	Not significant	n/a

genes, which includes 1,261 undefined variants, 61 downstream variants, 19 intron variants, 1 splice-donor variants, 99 upstream variants, 36 3'UTR variants, 11 5'UTR variants, 34 synonymous codon variants and 53 missense variants (Table 3).

In this study, we focused on missense SNPs. However, the selection of SNPs marker had functional consequence only as a priority scale adjusting to the scope and research resources. SNP variations in other functional positions remain a potential determinant of embryogenesis rate (Ting *et al.* 2013). From 53 missense variant SNPs, 40 SNPs were eliminated because the cDNA annotation of the ESTs did not match with the gene locus defined in the SNP information. Thirteen SNPs position used for further process were described in Table 4.

From all thirteen selected SNPs, only 10 were selected for further DNA sequencing process. From DNA sequencing process, we obtained data of SNP variations, which were further statistically analyzed using Cross Tab Chi Square, odds ratio analysis and One-way ANOVA (SPSS 20.0) (Table 5). Cross Tab Chi Square was used to analyze the degree of dependency between SNP variations with embryogenesis rate. One-way ANOVA was used to observe the difference in embryogenesis rate between SNP variations. Odds ratio analysis was used to compare the relative odds of the SNP variations to the occurrence of high embryogenesis. From 10 SNP positions, only one SNP, in Auxin Response Factor family (EMB6), showed significant result in Cross Tab Chi-Square and in One-way ANOVA. Odds ratio analysis showed 8 times chance of higher embryogenesis when the SNP at EMB6 is Adenine homozygote (A/A) compared to Adenine-Guanine heterozygote (A/G).

Sample size of population in this study is still relatively small. Further observation with larger population is needed for marker revalidation. There is also a possibility of other genes influencing embryogenesis of oil palm, which needs further investigation.

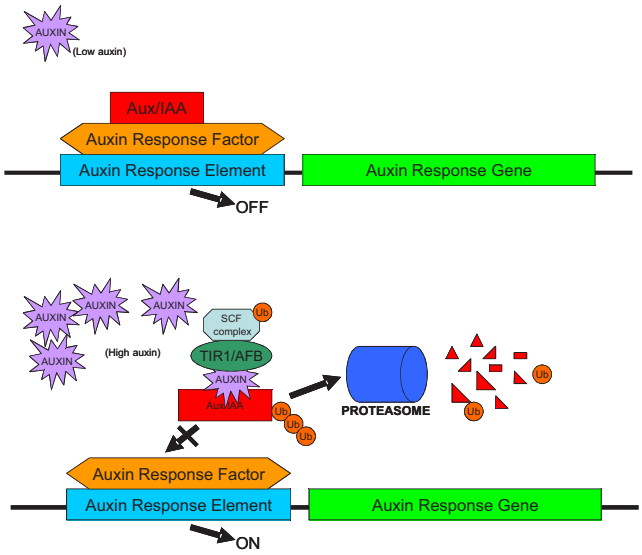

Figure 2 Auxin signalling pathway (adopted from da Costa *et al.* 2013; Salehin *et al.* 2015; Wang & Estelle 2014)

Auxin Response Factors (ARF) family has been suggested to play a key role in regulating the expression of auxin response genes (Liscum & Reed 2002). Gliwicka *et al.* (2013) found that the expression of over half of AUX/IAA and ARF genes were changed during somatic embryogenesis in *Arabidopsis*. Auxins play critical roles in most of the major growth responses throughout different developmental stages of plants; such as organogenesis, vascular tissue differentiation, apical dominance, root initiation, and tropism; as well as cellular level processes including extension, division, and differentiation (Guilfoyle & Hagen, 2007; Mockaitis & Estelle, 2008; Su *et al.* 2014). A large number of potentially auxin regulated candidate genes, which function in growth and developmental processes, have been identified in *Arabidopsis* and other plant species (Rosado *et al.* 2012; Liu *et al.* 2014; Di *et al.* 2015; Guilfoyle 2015).

ARF regulation is well-studied (Salehin *et al.* 2015). At low auxin levels, Aux/IAA proteins form dimers with ARFs to inhibit ARF activity resulting in the repression of auxin-responsive genes. At high auxin levels, Aux/IAAs bind to the $SCF^{TIR1/AFB}$ complex and subsequently become ubiquitinated and degraded by the 26S proteasome. The ARF is then released and regulate the transcription of its target auxin response genes (Wang & Estelle 2014) (Fig. 2).

Most of the ARF proteins consist of an N-terminal B3-type DNA binding domain (DBD), a variable middle region that functions as an activation domain (AD) or repression domain

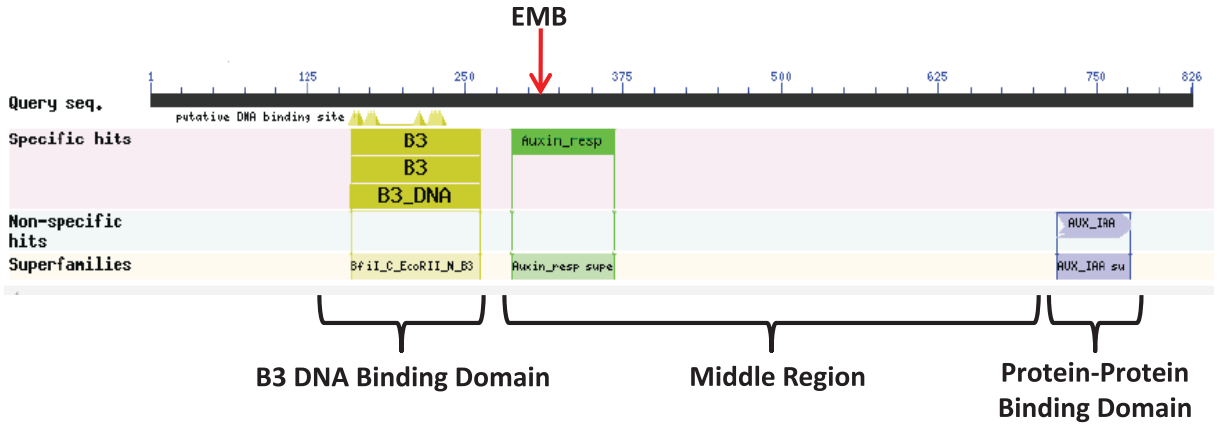

Figure 3 EMB6 Amino acid variation position

(RD) and a carboxy-terminal dimerization domain (CTD:domain III/IV), which is involved in protein– protein interactions by dimerizing with auxin/Indole-3-Acetic Acid (Aux/IAA) family genes or between ARFs (Kim *et al.* 1997; Guilfoyle & Hagen 2007; Piya *et al.* 2014). EMB6 was located at the 307th amino acid, inside the conserved domain of ARF family (Fig. 3). The position was in the middle region which is critical in determining ARF function. The variations found in this study were A/A and A/G, which change the amino acid from isoleucine to methionine. Although both amino acids are hydrophobic and have similar size, methionine has a unique structure as a sulphur-containing amino acid, which might change the protein folding due to its sulphuric bond.

CONCLUSIONS

In this study, bioinformatic analysis of SNP markers was performed by comparing the SNP database with the genes related to embryogenesis. One SNP located in an Auxin Respone Factor family gene showed a significant association with the embryogenesis rate of oil palm tissue culture. This SNP may be used as a marker to select sample source to increase the efficiency of oil palm tissue culture process in the near future. Further observation with a larger population is needed for marker revalidation.

ACKNOWLEDGEMENTS

This research was conducted at Biotechnology Department Laboratory; the samples and tissue culture productivity data were obtained from the Tissue Culture Department. Both places are part of the Plant Production and Biotechnology Division of PT SMART Tbk.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403-10.

Cochard B, Durand-Gasselin T, Amblard P, Konan EK, Gogor S. 1999. Performance of adult oil palm clones. In Ariffin D, Chan KW. Sharifah, editors. Emerging technologies and opportunities in the next millennium. agriculture conference. Proceedings of 1999 PORIM International Palm Oil Congress. 1 – 6 February 1999. Kuala Lumpur (MY). p. 53-64.

Corley RHV, Tinker PB. 2003. The oil palm. Fourth edition. Oxford (UK): Blackwell Science Ltd. p. 201-15.

Da Costa CT, de Almeida MR, Ruedell CM, Schwambach J, Maraschin FS, Fett-Neto AG. 2013. When stress and development go hand in hand: Main hormonal controls of adentitious rooting in cuttings. Front Plant Sci 4:133.

Di DW, Zhang C, Luo P,An CW, Guo GQ. 2015. The biosynthesis of auxin : How many paths truly lead to IAA? Plant Growth Regul:1–11.

Elhiti M, TahirM, Gulden RH, Khamiss K, Stasolla C. 2010. Modulation of embryo-forming capacity in culture through the expression of Brassica genes involved in the regulation of the shoot apical meristem. J Exp Bot 61:4069–85.

Gliwicka M, Nowak K, Balazadeh S, Mueller-Robber B, Gaj MD. 2013. Extensive modulation of the transcription factor transcriptome during somatic embryogenesis in *Arabidopsis thaliana*. PLoS ONE 8(7): e69261.

Guilfoyle TJ. 2015. The PB1 domain in auxin response factor and Aux/IAA proteins: A versatile protein interaction module in the auxin response. Plant Cell 27:33–43.

Guilfoyle TJ, Hagen G. 2007. Auxin response factors. Curr .Opin. Plant Biol 10:453–60.

Huang X, Madan A. 1999. CAP3: A DNA sequence assembly program. Genome Res 9**:**868-77.

Kim J, Harter K, Theologis A. 1997. Protein–protein interactions among the Aux/IAA proteins. Proc Natl Acad Sci USA 94:11786–91.

Lin HC, Morcillo F, Dussert S, Tranchant-Dubreuil C, Tregear JW, Tranbarger TJ. 2009. Transcriptome analysis during somatic embryogenesis of the tropical monocot *Elaeis guineensis*: Evidence for conserved gene functions in early development. Plant Mol Biol 70:173-92.

Liu X, Dinh TT, Li D, Shi B, Li Y, Cao X, Guo L, Pan Y, Jiao Y, Chen X. 2014. AUXIN RESPONSE FACTOR 3 integrates the functions of *AGAMOUS* and *APETALA 2* in floral meristem determinacy. Plant J 80:629–41.

Liscum E, Reed J. 2002. Genetics of Aux/IAA and ARF action in plant growth and development. Plant Mol Biol 49:387–400.

Low ETL, Alias H, Boon SH, Shariff EM, Tan CYA, Ooi LCL, ..., Singh R. 2008. Oil palm (*Elaeis guineensis*, Jacq.) tissue culture ESTs: Identifying genes associated with calogenesis and embryogenesis. BMC Plant Biol 8:62.

Mockaitis K, Estelle M. 2008. Auxin receptors and plant development: A new signaling paradigm. Annu Rev Cell Dev Biol 24:55–80.

Okonechnikov K, Golosova O, Fursov M, UGENE Team. 2012. Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics 28(8):1166-7.

Piya S, Shrestha SK, Binder B, Stewart CN. Jr, Hewezi T. 2014. Protein-protein interaction and gene co-expression maps of ARFs and Aux/IAAs in Arabidopsis. Front Plant Sci 5:744.

Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–2.

Rohani O, Sharifah SA, Rafii MY, Ong M, Tarmizi AH, Zamzuri I. 2000. Tissue culture of oil palm. In Basiron Y, Jalani BS, Chan KW, editors. Advances in oil palm research. Bangi (MY): Malaysian Palm Oil Board. p. 238-83.

Rosado A, Li R, VandeVen W, Hsu E, Raikhel NV. 2012. *Arabidopsis* ribosomal proteins control developmental programs through translational regulation of auxin response factors. Proc Natl Acad Sci USA 109:19537–44.

Salehin M, Bagchi R., Estelle M. 2015. SCFTIR1/AFB-based auxin perception: mechanism and role in plant growth and development. Plant Cell 27:9–19.

Singh R, Ong-Abdullah M, Low ETL, Manaf MAA, Rosli R, Nookiah R, ..., Sambanthamurthi R. 2013. Oil palm genome sequence reveals divergence of infertile species in old and new worlds. Nature 500:335-9.

Su YH, Liu YB, Bai B, Zhang XS. 2014. Establishment of embryonic shoot–root axis is involved in auxin and cytokinin response during *Arabidopsis* somatic embryogenesis. Front Plant Sci 5:792.

Te-Chato S, Hilae A. 2007. High-frequency plant regeneration through secondary somatic embryogenesis in oil palm (*Elaeis guineensis* Jaq. var. *tenera*). J Agr Technol 3(2):345-57.

The CK, Ong AL, Kwong QB, Apparow S, Chew FT, Mayes S, ..., Kulaveerasingam H. 2016. Genome-wide association study identifies three key loci for high mesokarp oil content in perrenial oil palm. Sci Rep 6:19075.

Ting NC, Jansen J, Nagappan J, Ishak Z, Chin CW, Tyan SG, ..., Singh R. 2013. Identification of QTLs associated with callogenesis and embryogenesis in oil palm using genetic linkage maps improved with SSR markers. PLoS ONE 8:e53076.

Ting NC, Jansen J, Mayes S, Massawe F, Sambanthamurthi R, Ooi LCL, ..., Singh R. 2014. High density SNP and SSR-based genetic maps of two independent oil palm hybrids. Bio Med Central 15:309.

Untergrasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3-new capabilities and interface. Nucleic Acid Res 40(15):e115.

Wahid MB, Abdullah SNA, Henson IE. 2005. Oil palm-achievements and potential. Plant Prod Sci 8:288-97.

Walenz B, Florea L. 2011. Sim4db and Leaff: Utilities for batch spliced alignment and sequence indexing. Bioinformatics 27(13):1869-70.

Wang R, Estelle M. 2014. Diversity and specificity: Auxin perception and signaling through the TIR1/AFB pathway. Curr Opin Plant Biol 21:51–8.

Wong G, Tan CC, Soh AC. 1997. Large scale propagation of oil palm clones – experiences to date. Acta Hort (ISHS) 447:649-58.

Wooi KC. 1995. Oil palm tissue culture-current practice and constraints. In Rao V, Henson IE, Rajanaidu N, editors. Proceedings of the 1993 ISOPB international symposium on recent developments in oil palm tissue culture and biotechnology: 24–25 September 1993. Kuala Lumpur (MY). p. 21-32.