Research article

Samsulrizal NH, Hazuki MD, Azmi NSA, Said ZSAM, Wahab NA, Ramli Z. 2020. Orthologous Revelation between *Elaeis guineensis*, *Arabidopsis thaliana* and *Solanum lycopersicum*. International Journal of Life Sciences and Biotechnology. 3(2):164- 179. DOI: 10.38001/ijlsb.717846

Orthologous Revelation between *Elaeis guineensis, Arabidopsis thaliana* and *Solanum lycopersicum*

Nurul Hidayah Samsulrizal[1,2]*, Muhammad Danial Hazuki1[2], Nur Sabrina Ahmad Azmi[1,2], Zatty Syamimi Adura Mat Said[3], Nurniwalis Abdul Wahab[3] and Zubaidah Ramli[3]

ABSTRACT

Oil palm is an important commodity crop in Malaysia as major contributor to agriculture sector. Thus, the need for better yield production is urgent to accommodate rising local and global demand while reducing the land expansion for oil palm plantation. This can be achieved by identifying the agronomical important traits in oil palm using comparative genomic approach. In this study, gene related to plant height, fruit development and fruit ripening in oil palm were predicted by comparing *Elaeis guineensis* genome sequence with *Arabidopsis thaliana* and *Solanum lycopersicum* genome sequence. The model plant chosen are based on its special attribute such as completely sequenced and fleshy fruit model. The analysis begun with orthology analysis using InParanoid, and SonicParanoid. There are 9,624 orthologous genes identified common among species selected. The reformatted orthologous genes were then annotated with Gene Ontology (GO) using Blast2GO program. The process of annotation includes blast with local database (DIAMOND), mapping, annotation and project statistical analysis. 100% of the orthologous gene has sequence of significant similarity but only 61.79% of the sequence has GO assignment. By using the annotated orthologous genes generated, only small amount of gene associated with trait of interest was predicted i.e. gibberellins (GAs) 10 genes, bassinosteroids (BRs) 14 genes, auxin (9 genes), fruit development (8 genes) and fruit ripeing (4 genes). These agronomical important genes could be utilized in genetic engineering and molecular breeding to improve the production of palm oils.

ARTICLE HISTORY
Received
11 April 2020
Accepted
29 June 2020

KEYWORDS
SonicParanoid, comparative genomic, ortholog,

Introduction

Palm oil is the major contributor to the production of world's oil and fat; and plays a significant role in accommodating the rising demand for global edible oils. About 23% of

[1] Department of Plant Science, Kulliyyah of Science, International Islamic University Malaysia.
[2] Research Unit for Bioinformatics and Computational Biology (RUBIC), Kulliyyah of Science, International Islamic University Malaysia, Kuantan, Pahang, Malaysia.
[3] Advanced Biotechnology and Breeding Centre (ABBC), Malaysian Palm Oil Board, 6 Persiaran Institusi, Bandar Baru Bangi, Kajang, Selangor, Malaysia.
*correspondence address: hidayahsamsulrizal@iium.edu.my

world fat and oil are contributed by palm oil production in which 85% of world palm oil productions are from Malaysia and Indonesia. The demand for palm oil increases constantly at the rate of 4% every year [1], thus the need for oil palm that has superior agronomic traits are more urgent than ever to accommodate high palm oil demands while keeping land exploration and expansion for agriculture purpose at minimum. This objective can be achieved in various ways such as breeding program and genetic modifications.

In 2013, oil palm genome has been completely sequenced and published by Singh and his team. It was estimated that the genome size of oil palm is 1800 Mbp with 34,800 numbers of genes has been predicted using a total of 1.535 Gb of assembled sequence and transcriptome data from 30 tissues types [3]. The genome information could be viewed via MyPalmViewer Genome Browser http://gbrowse.mpob.gov.my. From here, more research has been made to increase tolerance to biotic and abiotic stress with final goals is to increase the quantity and quality of the yield produce, achieve sustainable agriculture and reduce its carbon footprint [4, 5]

Comparative genomic is a way to predict genes function by comparing genes sequence from two or more species. Furthermore, comparative genomic study also involves an examination of gene loss, duplication, and horizontal gene transfer in evolutionary changes among organisms [21]. By identifying and manipulating the agronomical important genes, we can boost the quality and quantity of the yield produce from oil palm [2].In this research, we aim to infer orthologous relationship between oil palm (*E. guineensis*) with two dicot plants, which are model plants *A. thaliana* and tomato (*S. lycopersicum*) and to identify and characterize the agronomically related genes of interest especially in height, fruit development and fruit ripening.

Materials and Methods

Obtaining genomes sequences and bioinformatics tools

Genome sequence of *A. thaliana*, *E. guineensis* and *S. lycopersicum* are downloaded from their respective database which are The Arabidopsis Information Resource (TAIR), Genomsawit and Solanaceae Genomics Network. Blast package and its utility command are acquired from National Center for Biotechnology Information (NCBI). Stand-alone

InParanoid 4.1 and its extension MultiParanoid are acquired from Sonnhammer Bioinformatics Group website. SonicParanoid installation is done with Ubuntu terminal with manual instruction acquired from IWASAKI Lab/SonicParanoid page. Blast2GO software (OmicsBox) was locally installed on Windows and subscribed with all the functional annotation and analysis features. Genome sequence files which usually in compressed gz and tar.gz format are uncompressed using gunzip command in Linux.

Orthology analysis

Orthology analyses in this study consist of combination of reciprocal BLAST and ortholog group clustering InParanoid, and multi-species ortholog analysis (SonicParanoid).

Ortholog group clustering using inparanoid

The genome protein sequence for comparison (Arabidopsis, tomato and oil palm) which is in uncompressed fasta format was assembled in this directory. The output format of the analysis includes Human-readable text (Output.arab.fa-tomato.fa). Other than that, an HTML file serving as a slightly stylized representation of the previous (orthologs.arab.fa-tomato.fa.html) was generated. Furthermore, a tab-separated list listing each orthologous group on its own line (sqltable.arab.fa-tomato.fa) and a tab-separated list listing a given orthologous group across multiple lines (table.arab.fa-tomato.fa) can also be found. Summary of the analysis were viewed at file 'Human-readable text'.

Multi-species ortholog analysis using sonicparanoid

SonicParanoid program on the other hand, perform pairwise and multi-species ortholog analysis in one process. The genomes sequence was not formatted manually since SonicParanoid does not use blast search, but the file must be in fasta format. After the genome has been assembled, the virtual environment was created and activated. The Ubuntu terminal grep function are used to calculate the number of ortholog groups between species and extracted for annotation with Blast2GO. This output file contains number of total gene sequences, homolog sequences, in-paralogs in each species and groups of orthologs.

Extracting protein sequence from genome database to orthologous gene IDs

The list of orthologous gene IDs from SonicParanoid contain both the pairwise and orthologous of three species. Only the orthologous gene IDs between three species are

selected for annotation. The gene sequence of the selected orthologous gene IDs are then extracted from the *Arabidopsis* genome sequence to the newly created list using blastdb command in Linux. The *Arabidopsis* genome database must first be formatted for blast usage.

Blast search of orthologous *A. thaliana* protein sequence with NCBI database.

Three tools for Blast search used in this study includes Blast2GO high performance cloud server BLAST search, stand-alone NCBI-BLAST+ with local database and DIAMOND high performance analysis with local database. The local databases are acquired from NCBI ftp site from Non-redundant (nr) protein database and its md5 files. The md5 files are checksums to make sure the file is downloaded properly. The same files are used for Blast2GO BLAST search except the files are used directly from its private fast cloud server. For Blast2GO, the option for blast search and the parameters required can be chosen directly from the OMICSBOX software. Parameters which include number of hits, E-value, input types, output format and others are assign uniform among all blast tools.

Gene ontology annotation using BLAST2GO

The manually blast orthologous genes are loaded to the software for genes classification with GO terms by Gene Ontology mapping under functional analysis. Gene ontology annotation was then initiated with default set of annotation configuration. Successful annotation for each query sequence will result in table colour change for the respective sequence from light-green to blue at the Main Sequence Table, and only the annotated GOs will remain in the GO IDs column. The results are visualized and summarized using GO graph and charts.

Data mining for candidate genes

The GO IDs for plant height, fruit development and fruit ripening are identified by generating a GO graph using GO description search in Blast2GO. The GO IDs for plant height selected are GO:0009740 for gibberellic acid mediated signalling, GO:0009742 for brassinosteroid mediated signalling and GO:0009851 for auxin biosynthetic process. While, the GO IDs selected for fruit development and fruit ripening are GO:0010154 and GO:0009835. Annotated sequences assigned with these specific GO IDs were extracted, listed and discussed.

Result and Discussion

Comparative analysis was undertaken to detect orthologous relationships by using InParanoid program that applies *all-versus-all* sequence comparisons between two genomes, followed by merging these data using Sonicparanoid. Orthologous measurements can be divided into two categories; (1) cluster pairs of gene with the same biological function and (2) phylogenetic trees to identify functional divergence occurrence. The InParanoid program was used as an alternative to the phylogenetic method. In order to obtain the results, three well establish genomes have been selected i.e. *E. guineensis*, *A. thaliana* and *S. lycopersicum*. The results from the analysis were sorted and viewed manually using Microsoft Excel program. The total numbers of genes shared among all three species were obtained and calculated.

This research aims to identify gene that share similar function in three major group which plant height, fruit development and fruit ripening. *E. guineensis* genome (as our important crops) was compared to *A. thaliana* which is one of the established plant models and the genome was also completely sequenced and annotated. *S. lycopersicum* genome was selected as a representative of model plant for fleshy fruit specifically fruit development and ripening process which is part of traits of interest for this study.

The result show that, 10,163 genes was predicted orthologs to *S. lycopersicum*, and 9,819 of *E. guineensis* genes show to share similar function with *A. thaliana*. Apart from that, *A. thaliana* and *S. lycopersicum* seem to have the largest orthologous group compared to *E. guineensis*. Fig 1 on the intersections of the three pair sets shows that all three species share 9,624 orthologous genes in total contributing to estimately 37% of *E. guineensis* gene sequences (26,087). This latter set may reflect a basic gene(s) tool kit that important in adaptation to the environment [6]. These sets of genes also provide the information for this study which relating to the evolution of the species and the foundation set for selecting the genes in plant height, fruit development and fruit ripening.

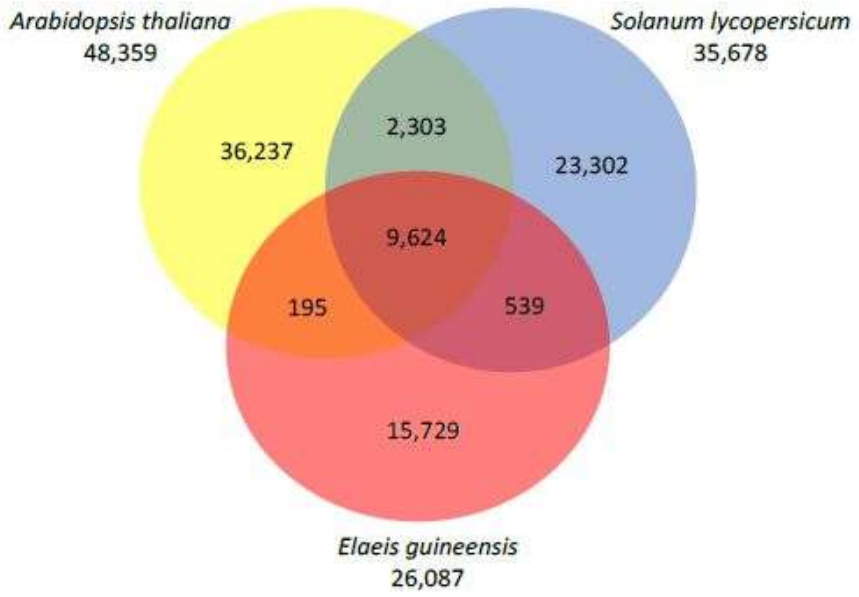

Fig 1 Venn diagram represents the result of comparative genomic analysis between oil palm and two dicot plants. Numbers in the area overlap indicate the number of orthologs predicted by InParanoid with E value = 10 x 10-5. Total of 9,624 protein sequences are identified as ortholog between all three species

Gene classıfıcatıon by gene ontology (GO)

Gene products are assigned to its respective class together with GO term, term that develop to describe its function. The 9,624 orthologous query sequences were blast using NCBI-BLAST+ tools, Blast2GO and DIAMOND to find high-scoring segment pairs (HSP). Majority of the sequences were blast with local database downloaded from NCBI using DIAMOND software since it has similar degree of sensitivity with NCBI-BLAST+ but 20,000 times faster [7]. General parameter set include E-value of 10x10-5 and 10 maximum hits. Successfully blast sequence (100%) was loaded into Blast2GO for annotation. Hits obtained by blast search contain the accession number of homolog gene that likely with same gene function. A total of 11,732 GO terms were loaded through mapping to Gene Ontology database. 5,947 or 61.79% of the query sequence (Table 1) are successfully assigned based on Gene Ontology Consortium in biological process (Fig 2A), molecular function (Fig 2B) and cellular components (Fig 2C). More than one term could be associated per gene [8] and each category can be split into other subcategories for high level classification [9].

Table 1 Gene Ontology classification analysis generated using Blast2GO suite

	Quantity	**Percentage (100%)**
Sequence has significant similarity (e-value ≤10^{-5})	9,624	100
Sequence has Gene Ontology assignment	5,947	61.79

GO classification consists of three major group i.e. biological process, molecular function and cellular component. Biological process term graph at second level of GO classification shows that gene for cellular process and metabolic process are the dominant conserved genes between genome of oil palm, *Arabidopsis* and tomato with 20% (4039 genes) and 18% (3744 genes) of the total gene sequence assigned by GO term followed by biological regulation (1854 genes), response to stimulus (1755 genes), regulation of biological process (1612 genes), cellular component organization or biogenesis (1174 genes), developmental process (1055 genes), multicellular organismal process (983 genes), localization (934 genes), reproduction (611 genes), reproductive process (608 genes), multi-organism process (497 genes), signalling (459 genes), negative regulation of biological process (419 genes), positive regulation of biological process (419 genes), growth (206 genes), immune system process (129 genes), rhythmic process (56 genes), detoxification (51 genes) and cell population proliferation (26 genes). Main plant cellular process and metabolic activity include cellular respiration and photosynthesis.

While for molecular function, binding (3499 genes) and catalytic activity (2836 genes) overwhelmingly dominate the graph with 46% are occupied only for binding category. These are followed by small percentage of transcription regulatory activity (476), transporter activity (360 genes), structural molecule activity (151 genes), molecular function regulator (148 genes), translation regulator activity (90 genes) and others. About 50% of the ortholog genes categorized in cellular component located at cellular anatomical entity. The locations of other genes are at intracellular, protein containing complex and only small ≈1.15% percentage of gene located at cell (123 genes).

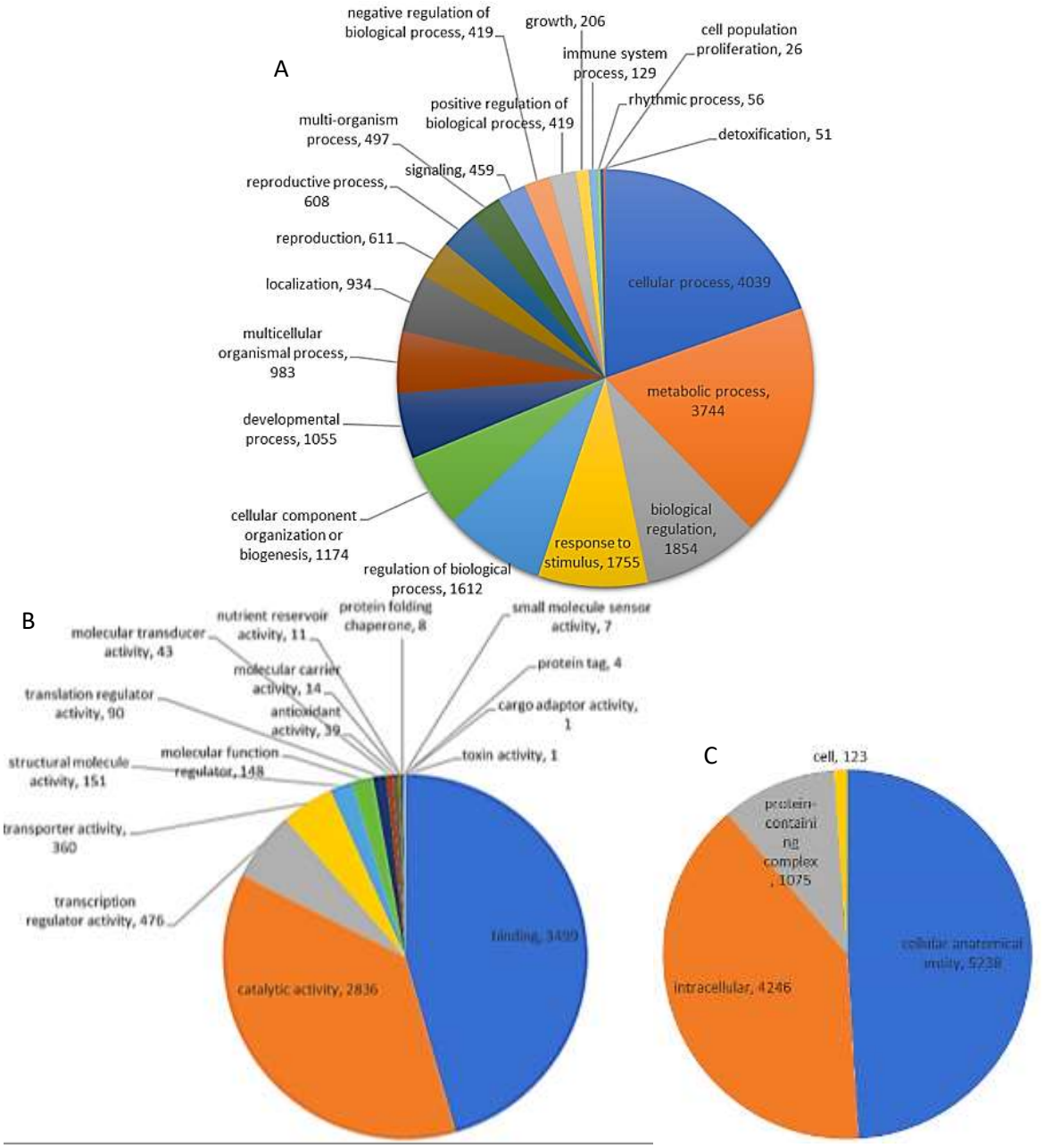

Fig 2 Gene Ontology Classifications. (a) Biological process term at second level of GO classification that was generated using BLAST2GO software. The metabolic process and cellular process were the most dominant second level, with0% and 18%. (b) Molecular function at second level of GO classification that generated using BLAST2GO software. There were 16 categories of second level molecular function terms that have been successfully assigned for all sequences in database with binding 46% as the dominant categories. (c) Cellular component at second level of GO classification that was generated using BLAST2GO software. There were 4 categories assigned cellular component term i.e. cell, intracellular, protein-containing complex and cellular anatomical entity

Data mining for candidate genes

Plant Height

Many of the earlier plant height studies have established gibberellin (GAs) and brassinosteroids (BRs) as significant stem elongation hormones [10]. Auxin biosynthesis and signalling also play an important role in controlling the length of the stem [11]. Thus, genes involve in regulating GAs, BRs and auxin are the target genes for regulating plant height in this study. The GO IDs identified for regulating 'GAs' is GO:0009740, gibberellic acid mediated signalling which has relationship with 28 GO terms within the hierarchical directed acyclic graph. Moreover, the GO IDs for regulating 'BRs' is GO:0009742 named brassinosteroid mediated signalling with 30 related GO terms and the GO IDs for 'auxin' hormone is GO:0009851 named auxin biosynthetic process with 12 GO terms. All three selected GO terms are annotated to biological process by gene ontology as visualized on their respective ancestor charts.

The ancestral charts represent the relationship between node (GO term). GO's structure can be represented as a graph, where each GO term is a node, and the connections between the nodes are edges. From the table generated by Blast2GO containing annotated orthologous genes, the genes of interest are identified using the selected GO terms (GO:0009740, GO:0009742 & GO:0009851) which involve in regulating plant height. There are 10 genes assigned to GO:0009740 which represent gibberellic acid mediated signalling in each species which is PTKs, ZNFs 8, AT1G68360, GASA14, LFY, NF-YC9, PIF3, GA2, GASA6 and AT1G22690 (Table 2). While for GO:0009742, that represent brassinosteroid mediated signalling were 14 genes from each species assigned to it such as transthyretin-like protein (TTL), rapid alkalinization factor 23 (RALF23), AT1G48270, AT2G42080, BSL2, AT1G63500, GSK1, VIK, scpl22, IWS1, BIM1, BEH4, BES1 and GF14 PHI (Table 3). For auxin biosynthetic process, GO:0009851, 9 annotated genes identified in each species (Table 4), such as Walls Are Thin 1 (WAT1), aldehyde oxidase 4 (AO4), TSA1, YUC1, YUC6, YUC6, YUC10, STY1 and LRP1. Each gene in Arabidopsis, oil palm and tomato carry the same gene function within their respective cluster but the level of expression during cell division and elongation for plant growth is not determined.

Table 2 Orthologous gene in *A. thaliana*, *E. guineensis* and *S. lycopersicum* that assigned to GO term gibberellic acid mediated signalling

Gene Description	Short Name	Arabidopsis Gene ID	Oil Palm Gene ID	Tomato Gene ID
Protein kinase family protein	PTKs	AT3G03940.1	p5.00_sc00008_p0027.1	Solyc03g119610.1.1
zinc finger protein 8	ZNFs 8	AT2G41940.1	p5.00_sc00133_p0046.1	Solyc03g058160.3.1
C2H2 and C2HC zinc fingers superfamily protein		AT1G68360.1	p5.00_sc00008_p0041.1	Solyc05g009170.2.1
GASA14		AT5G14920.1	p5.00_sc00005_p0101.1	Solyc03g113910.3.1
LEAFY transcription factor	LFY	AT5G61850.1	p5.00_sc00057_p0088.1	Solyc03g118160.2.1
nuclear factor Y, subunit C9	NF-YC9	AT1G08970.1	p5.00_sc00042_p0116.1	Solyc01g079870.3.1
phytochrome interacting factor 3	PIF3	AT1G09530.1	p5.00_sc00060_p0113.1	Solyc01g102300.3.1
Terpenoid cyclases/Protein prenyltransferases superfamily protein		AT1G79460.1	p5.00_sc00010_p0155.1	Solyc08g005640.3.1
Gibberellin-regulated family protein	GASA6	AT1G74670.1	p5.00_sc00035_p0062.1	Solyc03g116067.1.1
Gibberellin-regulated family protein		AT1G22690.1	p5.00_sc00045_p0180.1	Solyc11g017440.2.1

Table 3 Orthologous gene in *A. thaliana, E. guineensis* and *S. lycopersicum* that assigned to GO term bassinosteroid mediated signalling

Gene Description	Short Name	Arabidopsis Gene ID	Oil Palm Gene ID	Tomato Gene ID
transthyretin-like protein	TTL	AT5G58220.1	p5.00_sc00023_p0237.1	Solyc01g080940.3.1
protein-coupled receptor 1		AT1G48270.1	p5.00_sc00004_p0261.1	Solyc08g061260.3.1
Chaperone DnaJ-domain superfamily protein		AT2G42080.1	p5.00_sc01517_p0001.1	Solyc03g063350.3.1
rapid alkalinization factor 23	RALF23	AT3G16570.1	p5.00_sc00009_p0031.1	Solyc01g099520.3.1
BRI1 suppressor 1 (BSU1)-like 2	BSL2	AT1G08420.1	p5.00_sc00026_p0077.1	Solyc01g009280.3.1
kinase with tetratricopeptide repeat domain-containing protein		AT1G63500.1	p5.00_sc00197_p0020.1	Solyc11g064890.2.1
GSK3/SHAGGY-like protein kinase 1	GSK1	AT1G06390.1	p5.00_sc00065_p0093.1	Solyc07g055200.3.1
VH1-interacting kinase	VIK	AT1G14000.1	p5.00_sc00004_p0204.1	Solyc01g010950.3.1
serine carboxypeptidase-like 22	scpl22	AT2G24000.1	p5.00_sc00044_p0010.1	Solyc02g088820.3.1
Transcription elongation factor (TFIIS) family protein	IWS1	AT1G32130.1	p5.00_sc00200_p0018.1	Solyc06g066320.3.1
basic helix-loop-helix (bHLH) DNA-binding superfamily protein	BIM1	AT5G08130.1	p5.00_sc00079_p0054.1	Solyc01g080070.3.1
BES1/BZR1 homolog 4	BEH4	AT1G78700.1	p5.00_sc00059_p0105.1	Solyc02g071990.3.1
BR signaling positive regulator (BZR1) family protein	BES1	AT1G19350.1	p5.00_sc00106_p0015.1	Solyc02g063010.3.1
GF14 protein phi chain	GF14 PHI	AT1G35160.1	p5.00_sc00128_p0004.1	Solyc04g012120.3.1

Table 4 Orthologous gene in *A. thaliana, E. guineensis* and *S. lycopersicum* that assigned to GO term auxin biosynthetic process

Gene Description	Short Name	Arabidopsis Gene ID	Oil Palm Gene ID	Tomato Gene ID
tryptophan synthase alpha chain	TSA1	AT3G54640.1	p5.00_sc00072_p0034.1	Solyc01g098550.3.1
Walls Are Thin 1	WAT1	AT1G75500.1	p5.00_sc00134_p0082.1	Solyc04g080940.3.1
aldehyde oxidase 4	AO4	AT1G04580.1	p5.00_sc02118_p0001.1	Solyc01g009235.1.1
Flavin-binding monooxygenase family protein	YUC1	AT4G32540.1	p5.00_sc00105_p0039.1	Solyc06g065630.3.1
Flavin-binding monooxygenase family protein	YUC6	AT4G13260.1	p5.00_sc00059_p0124.1	Solyc08g068160.2.1
YUCCA 9	YUC9	AT1G04180.1	p5.00_sc00150_p0011.1	Solyc06g083700.3.1
Flavin-containing monooxygenase family protein	YUC10	AT1G48910.1	p5.00_sc00034_p0005.1	Solyc09g074430.3.1
Lateral root primordium (LRP) protein-like protein	STY1	AT3G51060.1	p5.00_sc00002_p0514.1	Solyc02g062400.3.1
Lateral root primordium (LRP) protein-like protein	LRP1	AT5G12330.1	p5.00_sc00032_p0044.1	Solyc11g064800.2.1

Fruit development and ripening

The development of fruit requires three fundamental stages. The first stage is the development of the ovary and the beginning of the division of cells, together called a set of fruit [12]. Cell division in the second phase and fruit grows in size during the third stage. Then, the ripening process will take the place once the fruit cells have fully expanded and the fruit has matured [13,14]. Each stage of fruit development and ripening involves specific gene activity as shown by transcriptomic analyses [15,16,17]. As predicted, several genes identified include transcription factors known to regulate major changes in fruit development and ripening, such as the Cnr locus [18] and several MADS-box genes expressed in the early stages of tomato fruit ripening [19, 20].

The GO IDs that was identified for fruit ripening is GO:0009835. In acyclic graph generated (Fig 3), fruit ripening and fruit development GO terms are within the same hierarchical directed acyclic graph. The GO IDs for fruit development is GO:0010154 which is the parent for GO term fruit ripening. The genes are annotated to biological process which involve 14 other GO terms. As expected, fruit development and ripening process involve heavily on the initiative for development of plant reproductive organ and structure (cell multiplication and differentiation). This can be observed by the identified GO terms in acyclic graph, which include anatomical structure development & reproductive system to fruit ripening.

From the table generated by Blast2GO containing annotated orthologous genes, our candidate genes are under GO terms (GO:0009835 & GO:0010154) which involve in development of fruit until ripening. There are 8 genes (CYP94B3, XET28, UBP15, AGB1, XIK, RPL, AFO, and POP2) assigned to GO:0009835 which represent fruit development in each species. While for GO:0010154 which represent fruit ripening has 4 genes from each species assigned such as aminocyclopropane-1-carboxylate synthase 4 (ACS4), NAC-like, activated by AP3/PI (NAP), ACS7 and ACS2. Each gene in *Arabidopsis*, oil palm and tomato carry the same gene function within their respective row yet the level of expression during cell division and elongation for plant growth still unknown.

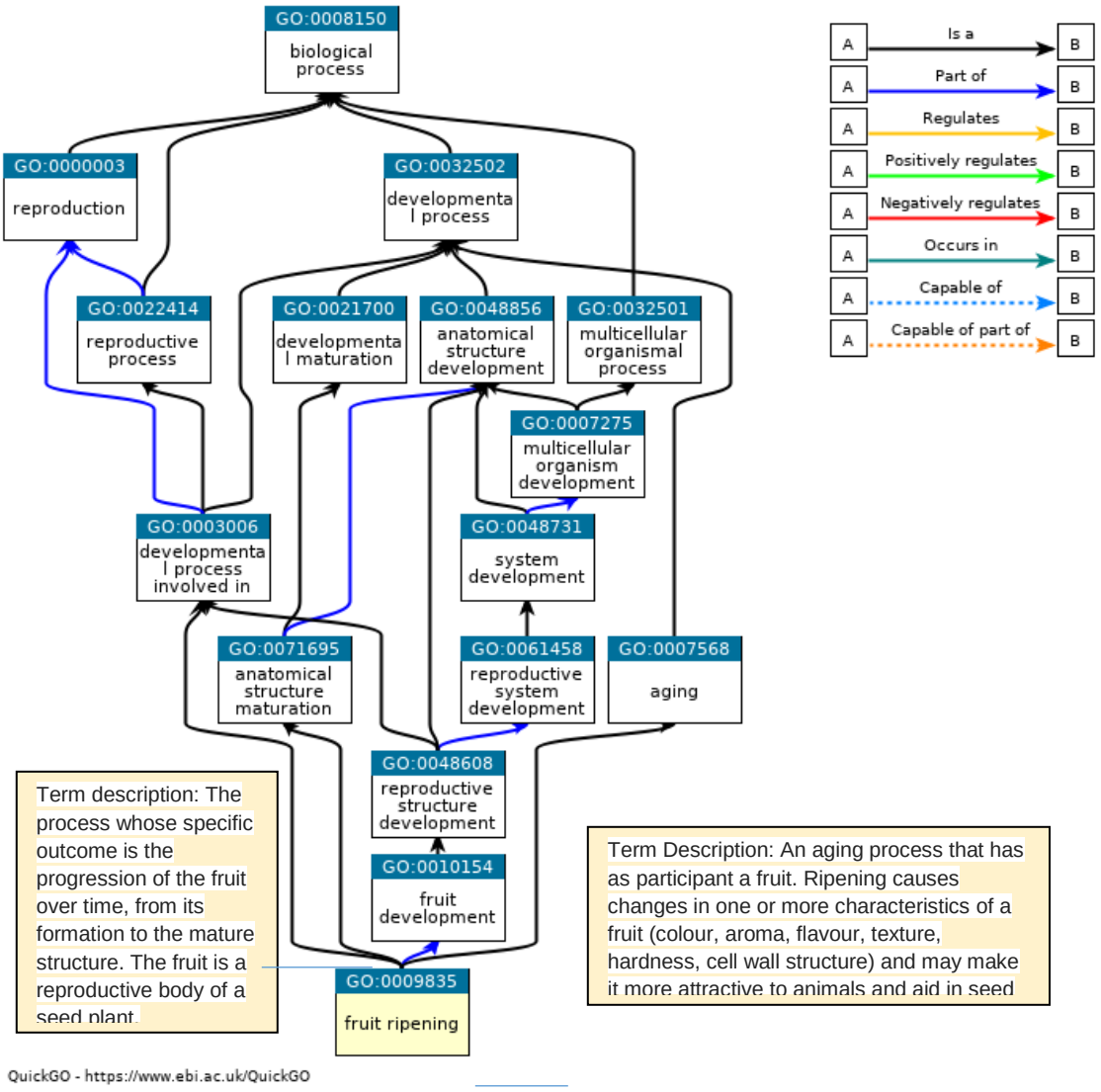

Fig 3 Acyclic graph for fruit ripening and fruit development

Conclusion

Comparative genomics approaches have provided connections between function and evolution in complex biological structures and systems. The aim of this study was to provide genes that were common emerge in group of plant height fruits development and ripening in *E. guineensis* and two dicot plants namely *A. thaliana* and *S. lycopersicum*. The objectives were successfully achieved using combination of bioinformatics tools i.e. InParanoid, SonicParanoid, Diamond BLAST and Blast2GO. After all, only a small

number of related trait genes are present as orthologous genes for each of the group(s). It was suggested in the future work, these agronomical important gene(s) can be utilised not only in molecular breeding but also in genetic engineering such as RNAi, TALENS, or CRISPR techniques. These approaches were then will lead to discovery of new line or variety that consists of desired genetic backgrounds.

Acknowledgments
We sincerely thank the staff and administration of International Islamic University Malaysia and The Malaysian Palm Oil Board for their assistance and support in this study. This study is funded by the International Islamic University Malaysia (IIUM) Research Initiative Grant Scheme (RIGS17-022-0597) for sponsored this work.

References

1. Abdullah, R., and Wahid, M. B., World palm oil supply, demand, price, and prospects: focus on Malaysian and Indonesian palm oil industry. Malaysian Palm Oil Board Press. Malaysia. 2010.
2. Samsulrizal, N. H., and Yusof, N. Y., In Silico Prediction of Cell Wall Remodeling Genes in Tomato, Banana, Melon and Grape. International Journal of Life Sciences and Biotechnology, 2019. 2(2), 108-121.
3. Singh, R., et.al., Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. Nature, 2013. 500(7462), pp.335-339.
4. Nur, S. Y., Suhaila, S., and Ping, L. Genome-wide comparative analysis in discovering potential evolved hypothetical genes in infected oil palm. Malaysian Applied Biology Journal, (2018). 47(5), 81-88.
5. Rosli, R., et al., Comparative genomic and transcriptomic analysis of selected fatty acid biosynthesis genes and CNL disease resistance genes in oil palm. PloS one, 2018. 13(4).
6. Somero, G. N., The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine 'winners' and 'losers'. Journal of Experimental Biology, 2010. 213(6), 912-920.
7. Buchfink, B., Xie, C., and Huson, D. H., Fast and sensitive protein alignment using DIAMOND. Nature methods, 2015. 12(1), 59.
8. The Gene Ontology Consortium., The gene ontology resource: 20 years and still GOing strong. Nucleic acids research, 2018. 47(D1), D330-D338.
9. Lewis, S. E., The Vision and Challenges of the Gene Ontology. The Gene Ontology Handbook, 2017. 291.
10. Gyawali, A., et al., Single-plant GWAS coupled with bulk segregant analysis allows rapid identification and corroboration of plant-height candidate SNPs. BMC plant biology, 2019. 19(1), 412.
11. Wang, Y., and Jiao, Y., Auxin and above-ground meristems. Journal of experimental botany, 2017. 69(2), 147-154.
12. Handa, A. K., et al., Fruit development and ripening: a molecular perspective. In Plant biotechnology and agriculture. Academic Press, 2012. pp. 405-424.
13. Gillaspy, G., Ben-David, H., and Gruissem, W., Fruits: a developmental perspective. The Plant Cell, 1993. 5(10), 1439.

14. Ezura, H., and Hiwasa-Tanase, K., Fruit development. In Plant Developmental Biology-Biotechnological Perspectives, Springer, Berlin, Heidelberg. 2010. pp. 301-318.
15. Carrari, F., et al., Integrated analysis of metabolite and transcript levels reveals the metabolic shifts that underlie tomato fruit development and highlight regulatory aspects of metabolic network behavior. Plant Physiology, 2006. 142(4), pp.1380-1396.
16. Mounet, F., et al., Gene and metabolite regulatory network analysis of early developing fruit tissues highlights new candidate genes for the control of tomato fruit composition and development. Plant Physiology, 2009. 149(3), 1505-1528.
17. Srivastava, A., et al., Maturity and ripening-stage specific modulation of tomato (Solanum lycopersicum) fruit transcriptome. GM crops, 2010. 1(4), 237-249.
18. Manning, K., et al., A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. Nature genetics, 2006. 38(8), 948.
19. Busi, M. V., et al., MADS-box genes expressed during tomato seed and fruit development. Plant molecular biology, 2003. 52(4), 801-815.
20. Giovannoni, J. J., Fruit ripening mutants yield insights into ripening control. Current opinion in plant biology, 2007. 10(3), 283-289.
21. Wickell, D.A. and Li, F.W., On the evolutionary significance of horizontal gene transfers in plants. New Phytologist, 2020. 225(1), pp.113-117.