Supplementary Information for

Oil palm genome sequence reveals divergence of interfertile species in old and new worlds

Rajinder Singh[1*], Meilina Ong-Abdullah[1*], Eng-Ti Leslie Low[1*], Mohamad Arif Abdul Manaf[1], Rozana Rosli[1], Rajanaidu Nookiah[1], Leslie Cheng-Li Ooi[1], Siew–Eng Ooi[1], Kuang-Lim Chan[1], Mohd Amin Halim[1], Norazah Azizi[1], Jayanthi Nagappan[1], Blaire Bacher[2], Nathan Lakey[2], Steven W Smith[2], Dong He[2], Michael Hogan[2], Muhammad A Budiman[2], Ernest K Lee[3], Rob DeSalle[3], David Kudrna[4], Jose Louis Goicoechea[4], Rod Wing[4], Richard K Wilson[5], Robert S Fulton[5], Jared M Ordway[2], Robert A Martienssen[6,7] and Ravigadevi Sambanthamurthi[1,7]

[1] Malaysian Palm Oil Board, 6, Persiaran Institusi Bandar Baru Bangi, 43000 Kajang Selangor, Malaysia
[2] Orion Genomics, 4041 Forest Park Avenue, St. Louis, MO 63108, USA
[3] Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY 10024, USA
[4] Arizona Genome Institute, University of Arizona, Tucson AZ 85705, USA
[5] The Genome Institute at Washington University, Washington University School of Medicine, Saint Louis, MO 63108, USA
[6] Howard Hughes Medical Institute-Gordon and Betty Moore Foundation, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA
[7]For correspondence: martiens@cshl.edu, raviga@mpob.gov.my
*These authors contributed equally to this work.

TABLE OF CONTENTS

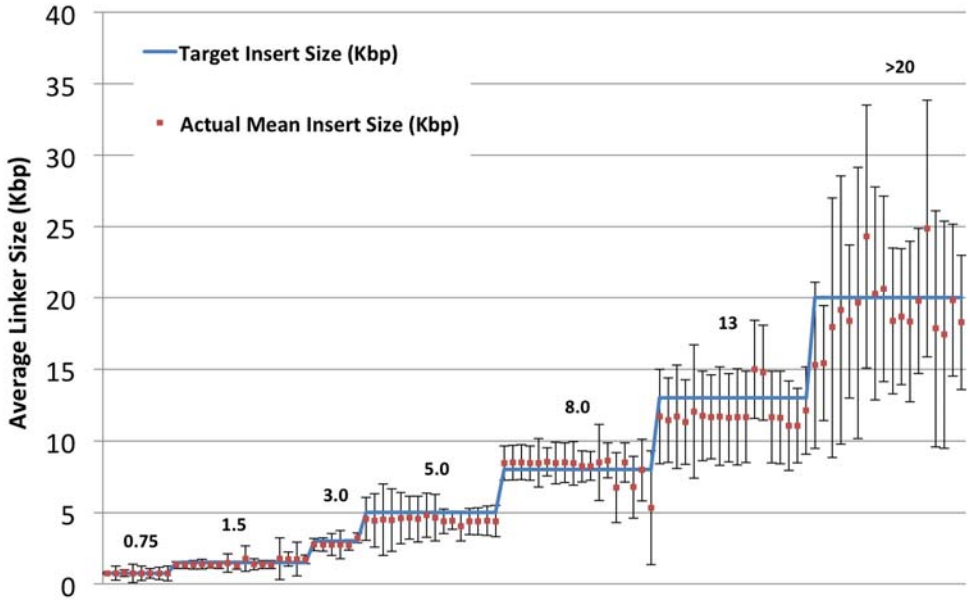
Supplementary Figures

Supplementary Tables

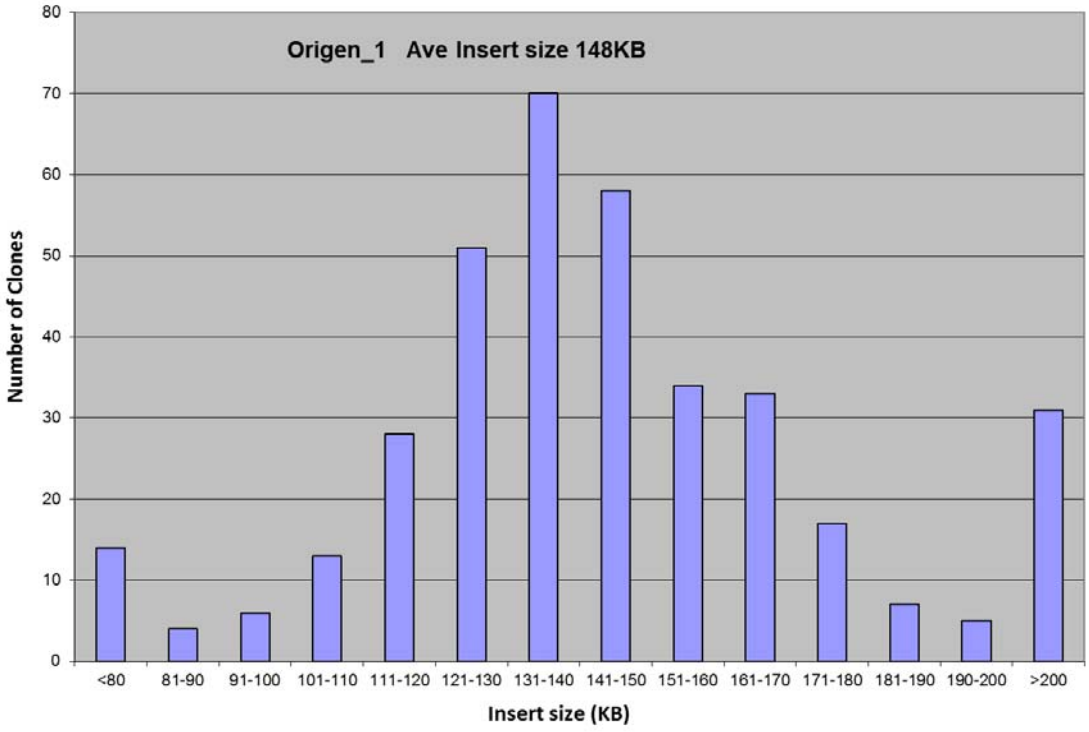
Supplementary Methods

Supplementary Notes

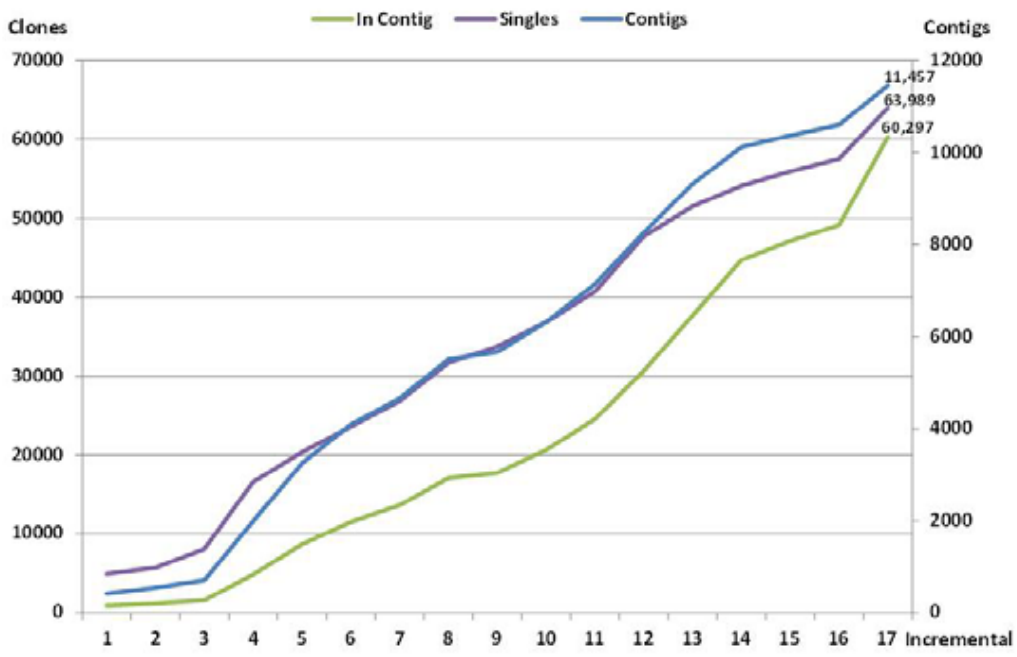
Supplementary Figures



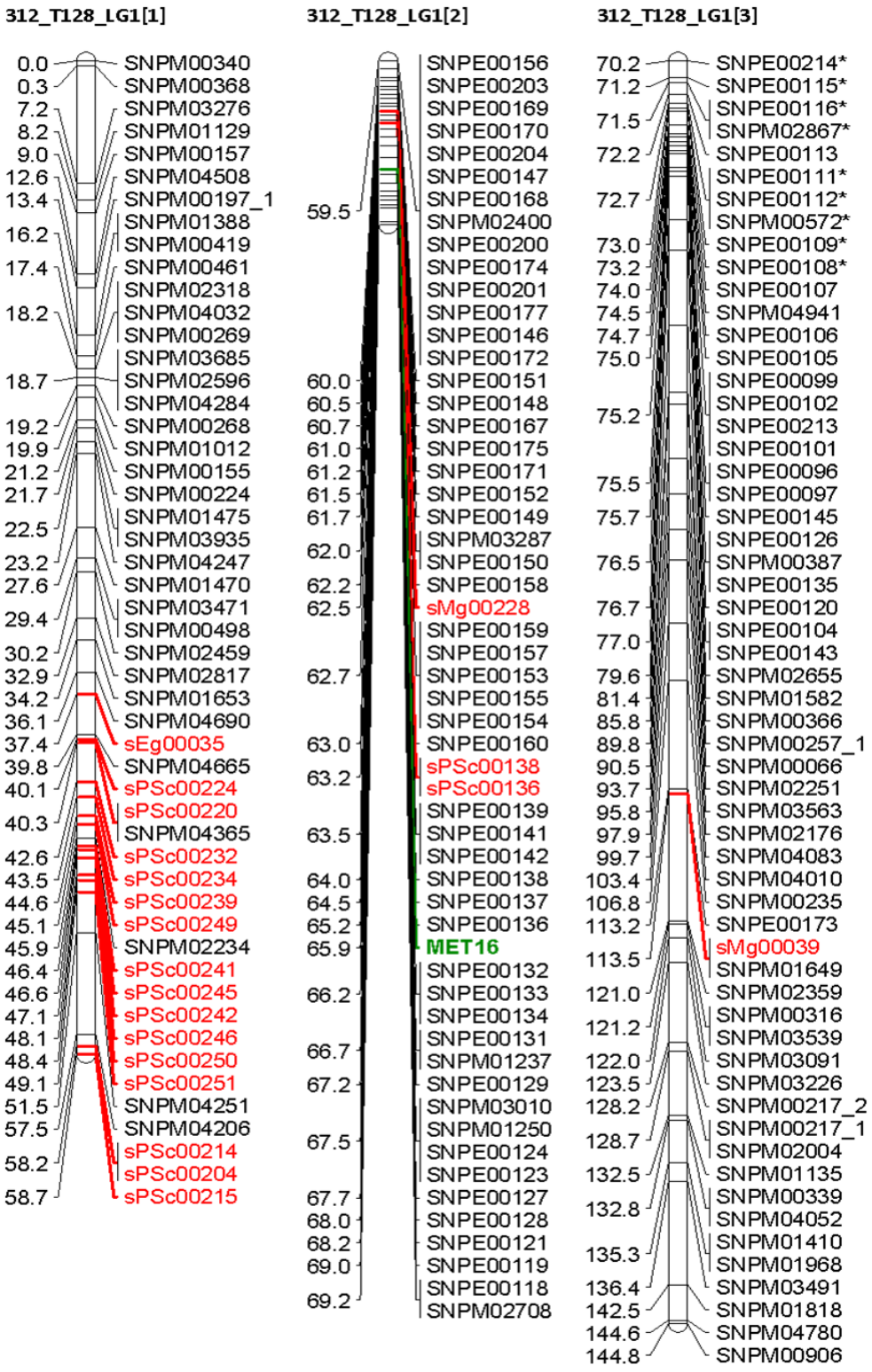
Supplementary Figure 1. 454 Linker Library Production. The fragment sizes going into each 454 linker library type were retrospectively monitored by BLAST comparison of linker reads from each half-plate sequencing run to an initial assembly of the *pisifera* reference genome. 50C50 linker library reads (including at least 50 bases up- and down-stream of the identified center adapter sequence) that mapped to a single contig or scaffold were identified. The DNA fragment size spanning the two paired linker read sequences was calculated based on the alignment to the initial reference sequence. The intended target insert sizes for each library type are represented by the solid blue line. The average calculated insert size for each half-plate run is indicated by a red box. Error bars represent +/- one standard deviation from the mean of all successfully measured insert sizes for the given half-plate 454 sequencing run. As shown in Supplementary Table 1, the intended 10 Kb target insert size linker libraries displayed a low 50C50 rate (27% overall). For this reason, too few reads generated successfully measured insert sizes to accurately calculate an actual average insert size, and therefore this library class is excluded from the graph.



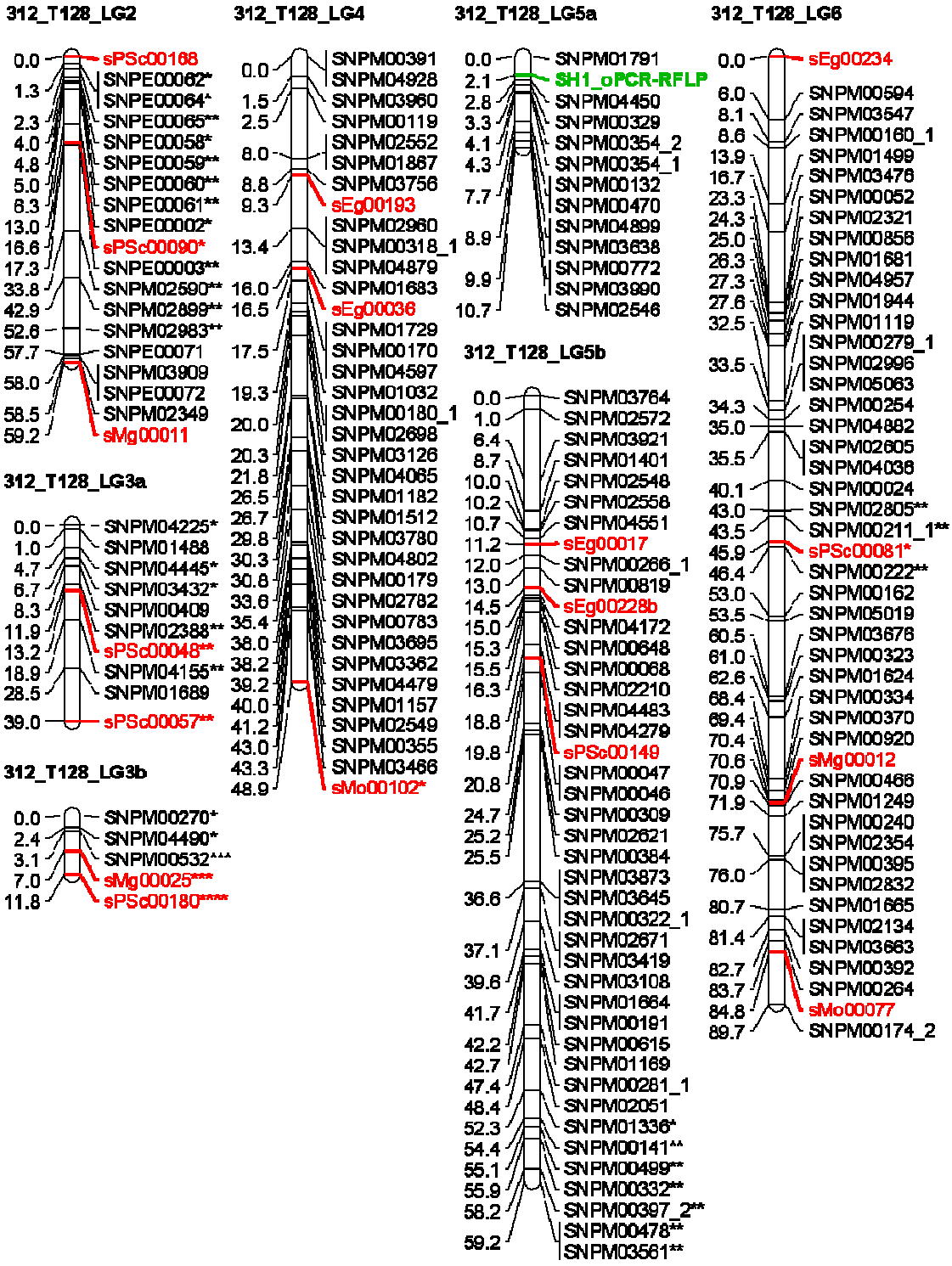
Supplementary Figure 2. Average insert size and distribution sizes of Origen_1 BAC library. Methods for BAC library construction were performed essentially as previously described[44,45]. The average insert size and size distribution of the library were determined by selecting random BAC clones and isolating BAC plasmid DNA followed by digestion with *Not*I restriction enzyme and separation on 1% agarose with pulse field gel electrophoresis (PFGE) (CHEF Mapper, BIO-RAD) at 1–50 sec linear ramp, 6 volts/cm, 14°C in 0.5X TBE buffer for 18–20 hours. The insert size of each clone was determined by comparison with molecular weight standards and averages and size distributions were calculated.



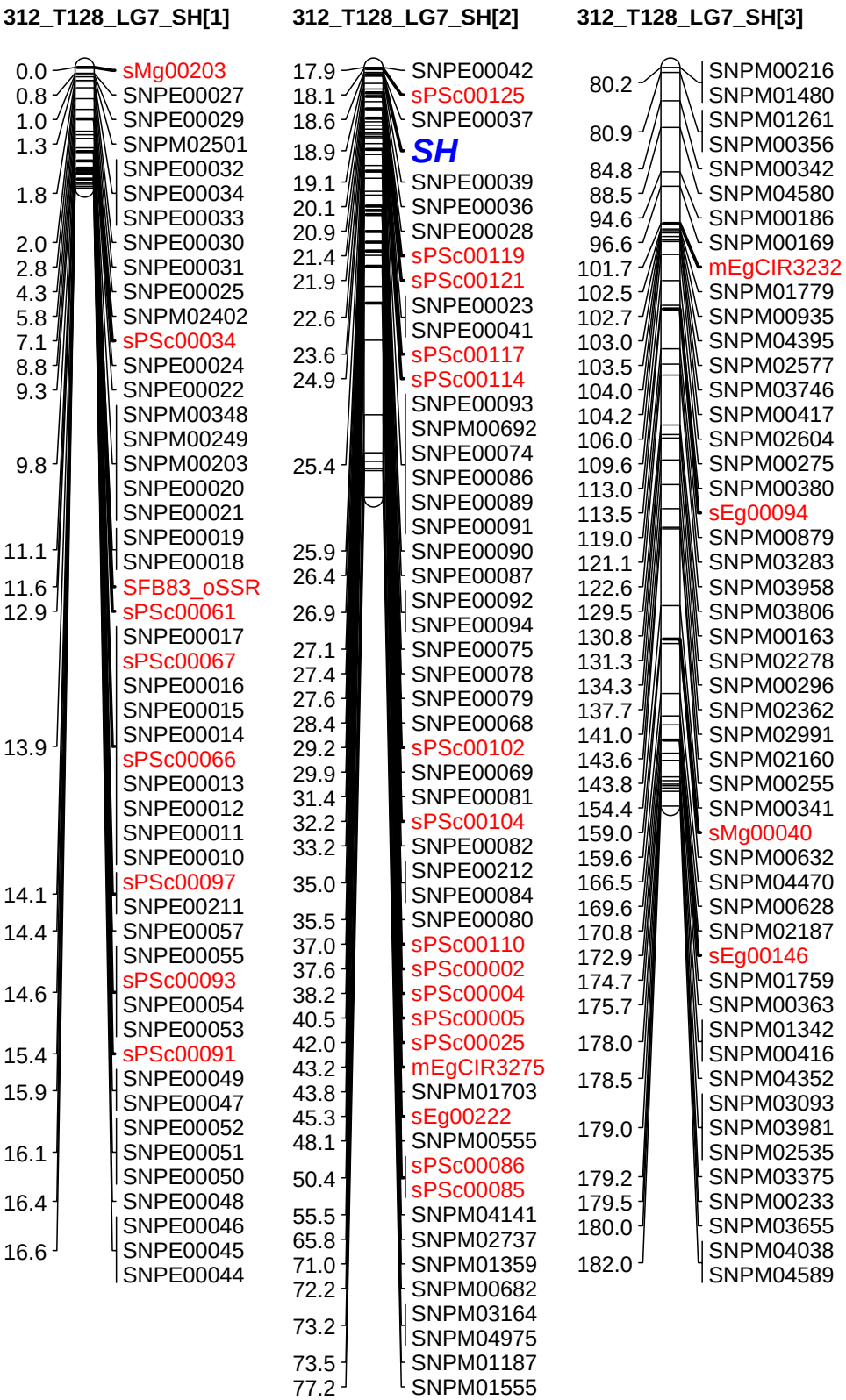
Supplementary Figure 3. Incremental assembly of fingerprinted BACs. Fingerprinted BACs were assembled incrementally into a total of 17 builds with FPC software (incremental builds on x-axis). The final build of the reference *pisifera* physical map resulted in 63,989 (51%) BACs which remained as singletons and 60,297 (49%) which formed 11,457 contigs. Of the 11,457 contigs, 3,579 (31%) had two BACs, 6,359 (56%) had 3-9 BACs each, 1,463 (13%) had 10-24 BACs each, and 56 (0.5%) contigs comprised 25 or more BAC clones each.



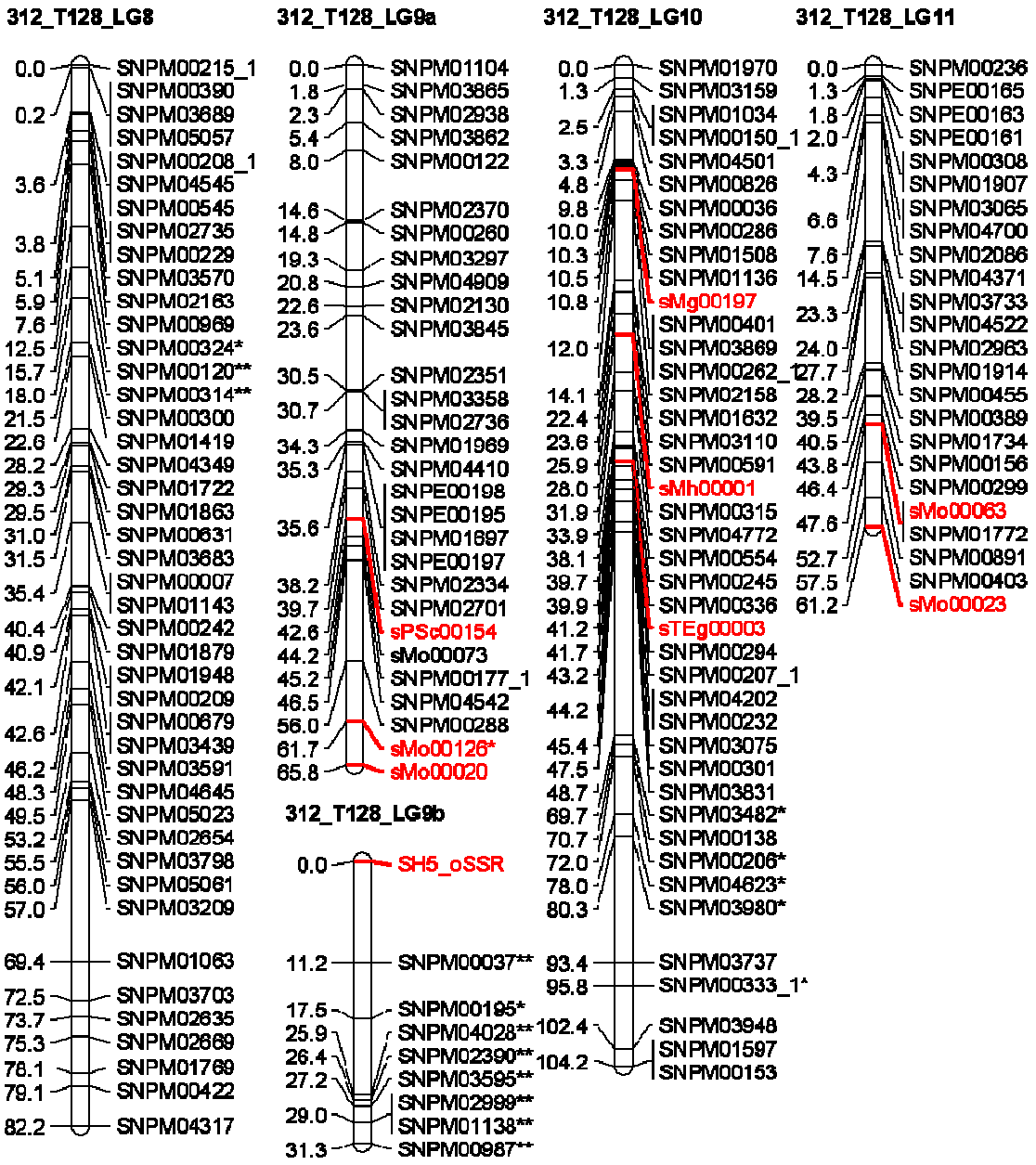
Supplementary Figure 4 continued on next page.



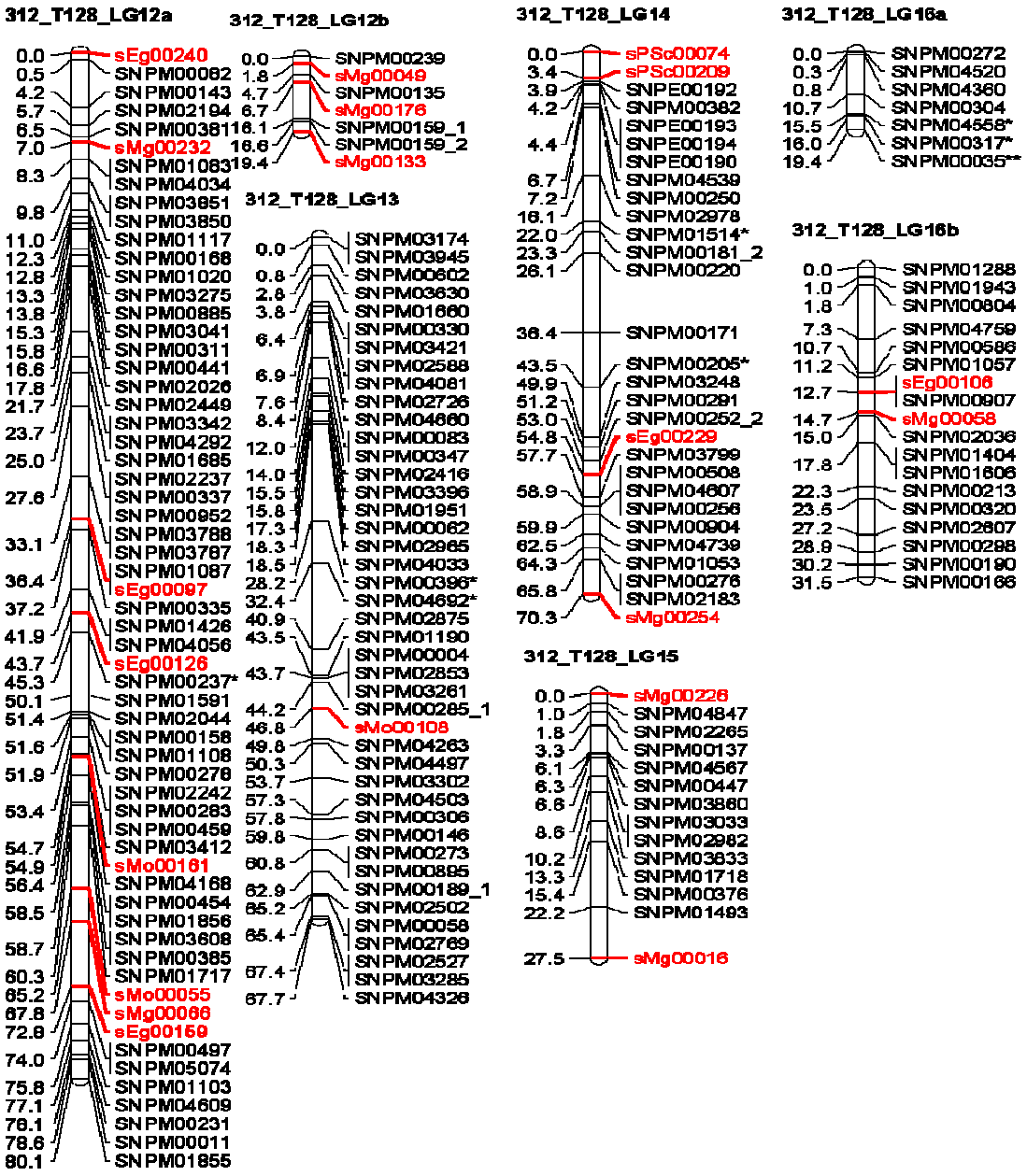
Supplementary Figure 4 continued on next page.



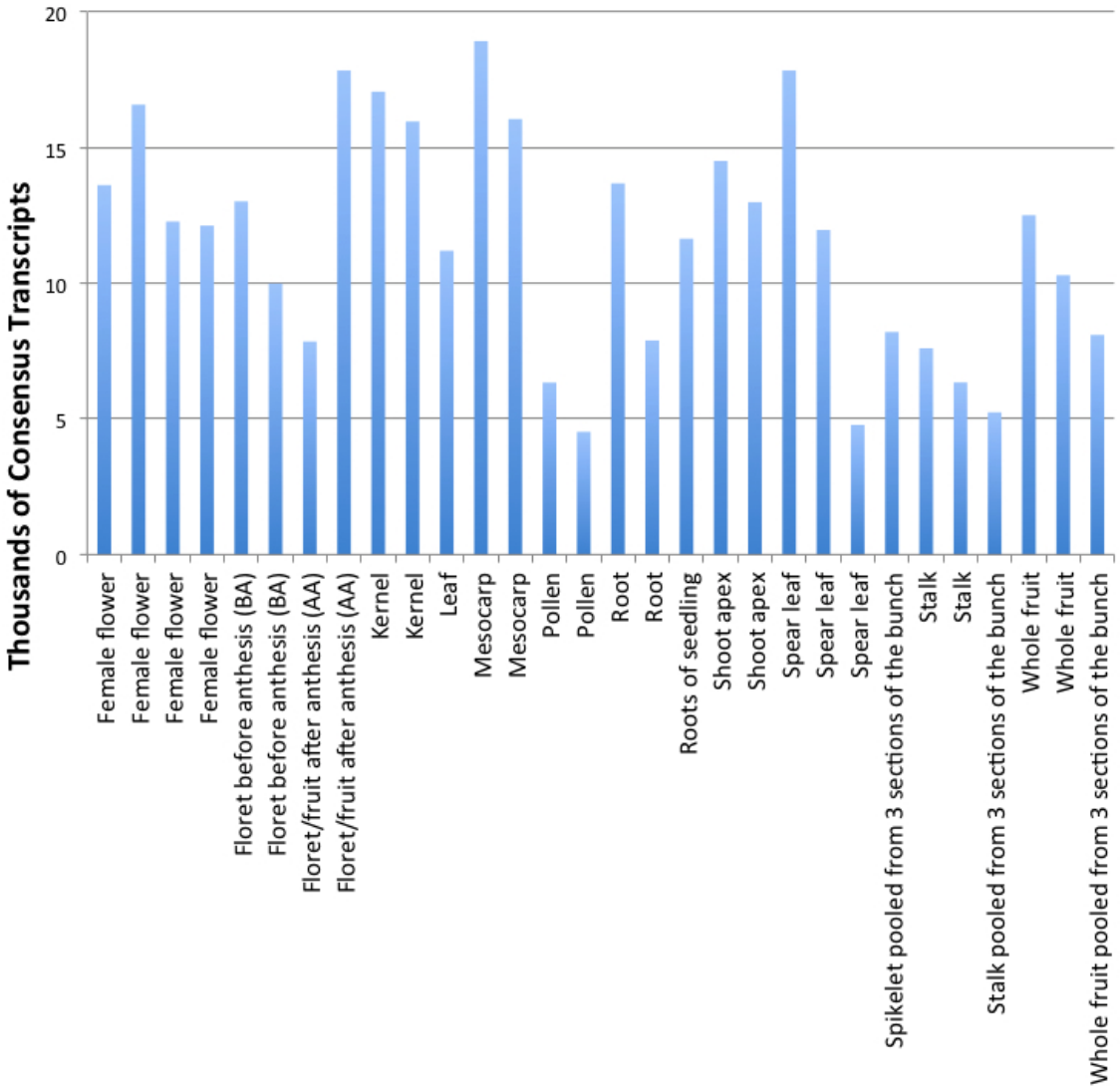
Supplementary Figure 4 continued on next page.



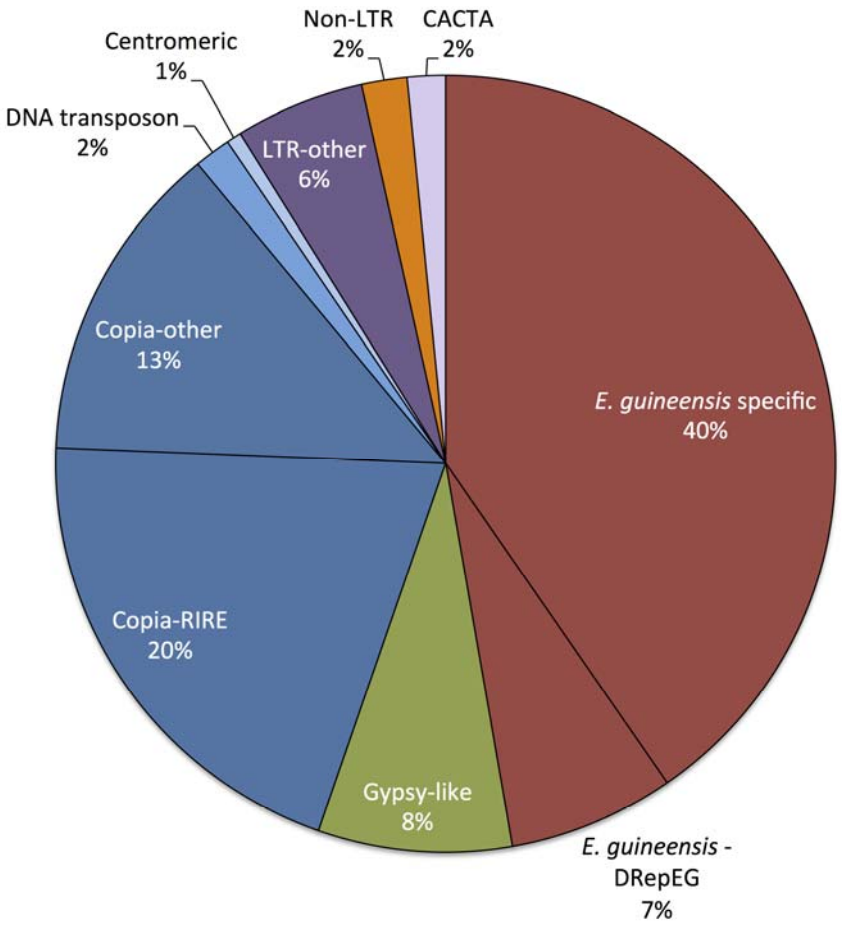
Supplementary Figure 4 continued on next page.



Supplementary Figure 4. The genetic map of the selfed Nigerian palm T128, constructed using a combination of markers namely SNP, SSR and RFLP. SSR markers are in red, RFLP markers are labeled green, the *Shell* gene is in blue colour while SNP markers are denoted in black. The single asterisk shows markers skewed at $P < 0.1$; double asterisk: skewed marker at $P < 0.05$; three asterisks: skewed marker at $P < 0.01$; four asterisks: skewed marker at $P < 0.005$; five asterisks: skewed marker at $P < 0.001$; six asterisks: skewed marker at $P < 0.0005$.



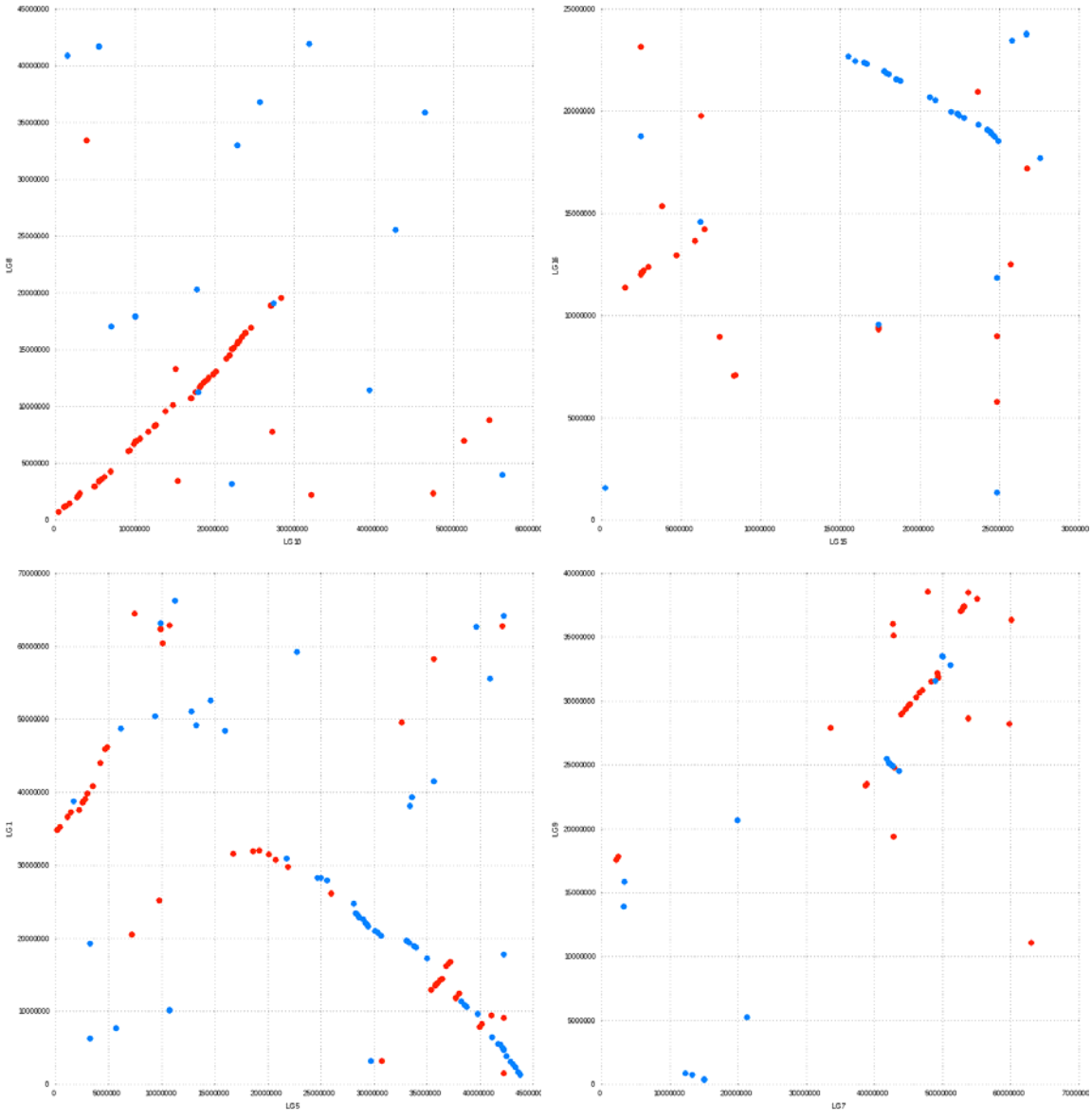
Supplementary Figure 5. Isotigs (unigene clusters) formed by assembly of transcriptome libraries. Thirty transcriptome libraries were constructed and sequenced by Roche/454 technology. In total, 15.4 million transcriptome sequences were generated comprising 5.2 Gb of sequence. Each of the 30 libraries was assembled, producing between 4,528 and 18,936 isotigs from fertile *pisifera* pollen (T124) and *pisifera* fruit, respectively.



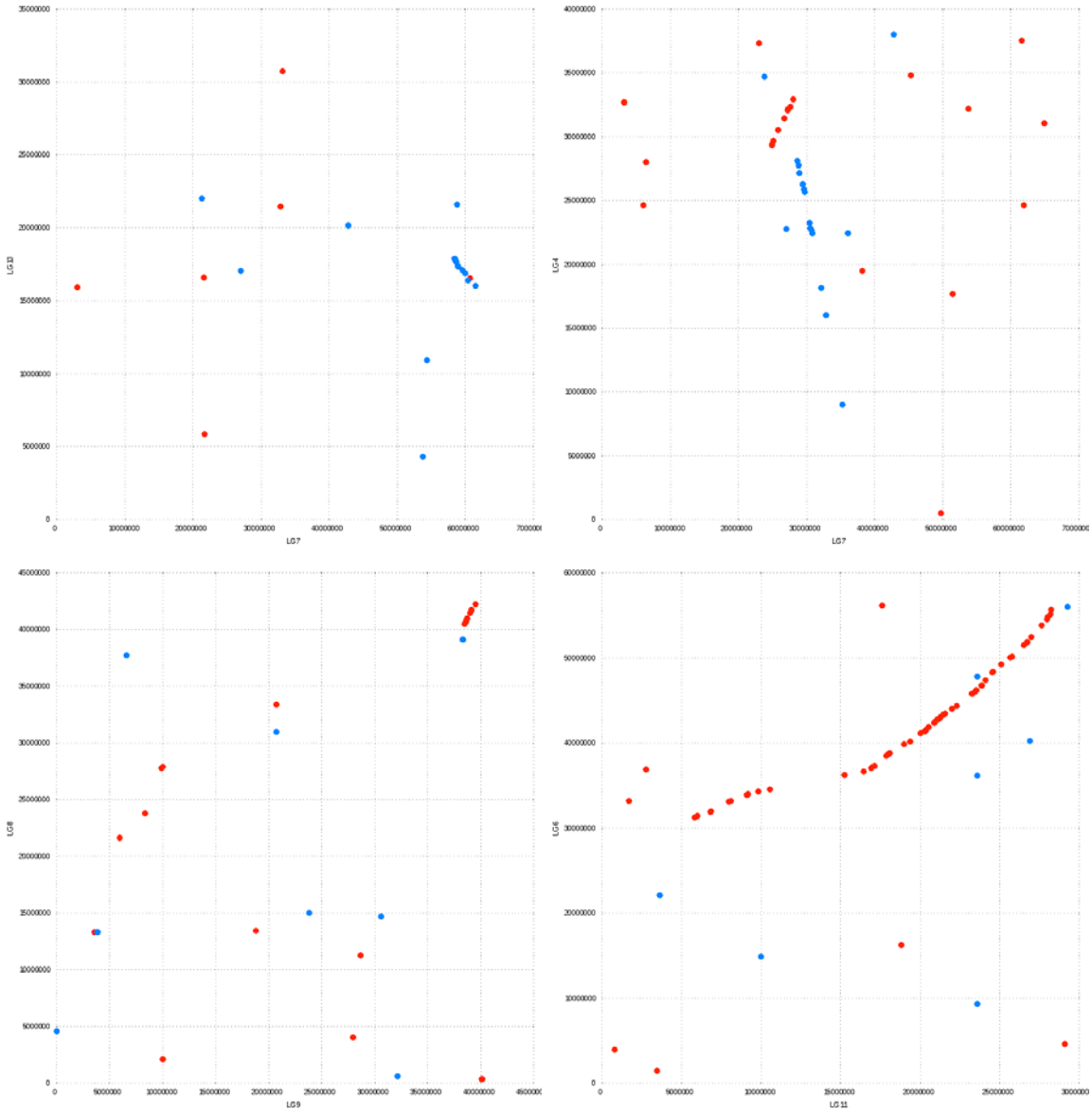
Supplementary Figure 6. Repeat content. Sequences from the P4 build were screened against known repeat databases from TIGR and RepBase[46] as well as against the previously constructed PisiferaRepeat and OleiferaRepeat databases using TBLASTN (Eval < 10^{-5}, low complexity filters disabled). Regions of the P4 build matching these repeat databases were extracted and annotated based on the repeat class to which they most closely matched. This resulted in genomic regions totaling 282.3 Mb (14 bp to 5106 bp in size, mean 363 bp, median 197).

Known repeats in oil palm are dominated by LTR (long terminal repeat) retrotransposons, especially the RIRE1-like subfamily which has undergone significant expansion, as have other *copia* elements. Among *copia* elements, RIRE1 was the most abundant subclass (20%), resembling certain species of rice in this respect[47], in agreement with the prevalence of *copia* elements in oil palm predicted by FISH[17]. There are very few non-LTR and DNA class transposons. Of the non-LTR retrotransposons, LINEs make up less than 1% of total repetitive elements, and SINEs are negligibly

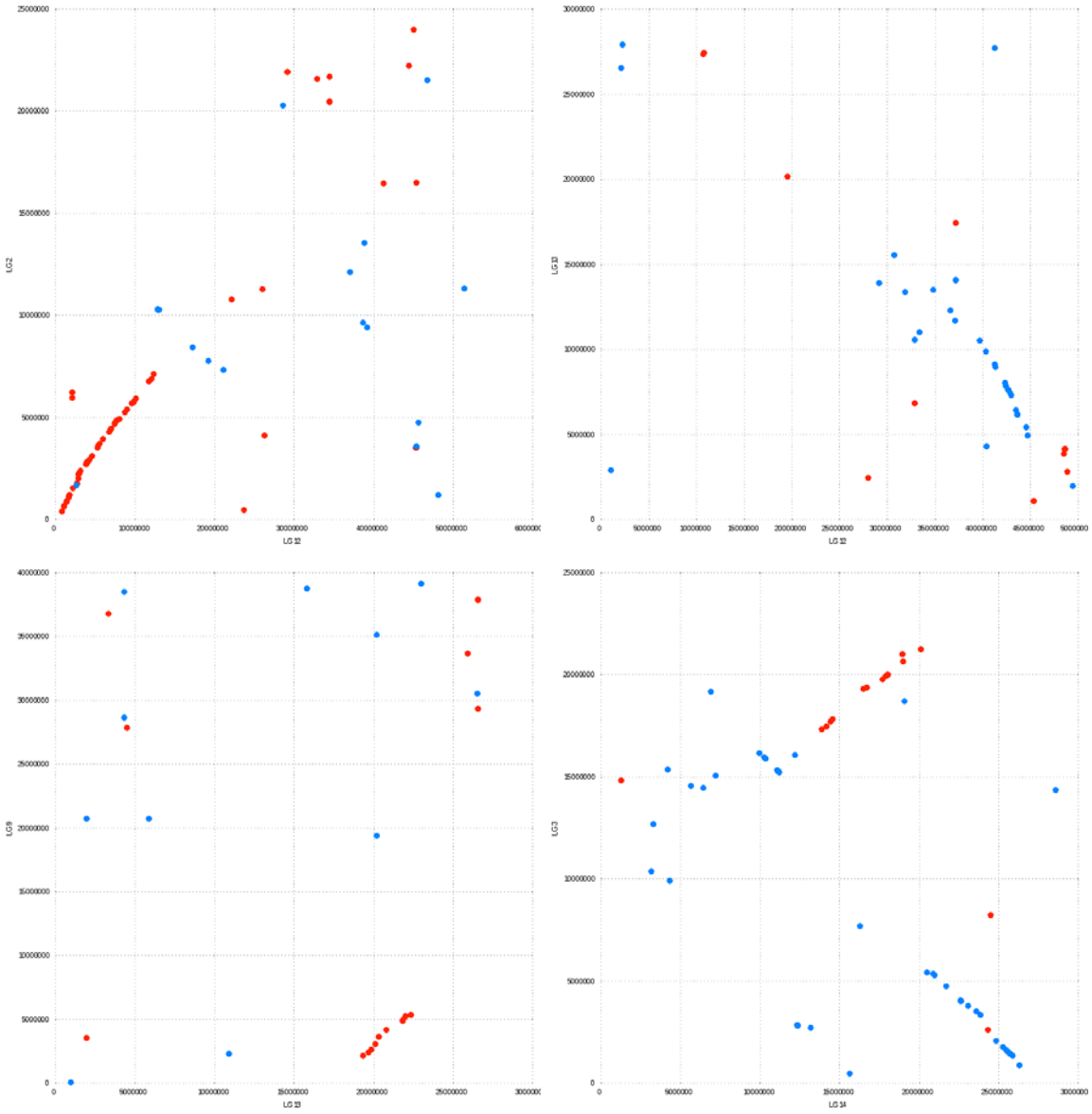
observed. Other classes of repetitive elements such as Class II DNA transposons, hAT-like transposons, MITEs and CACTA repeats make up relatively small proportions of the overall repeat content of the oil palm genome. About half of the repeats in the oil palm genome show no sequence similarity to other previously identified retro elements, although 7 percent are composed of a previously observed but uncharacterized dispersed repeat in *E. guineensis* DRepEG[17].



Supplementary Figure 7 continued on next page.

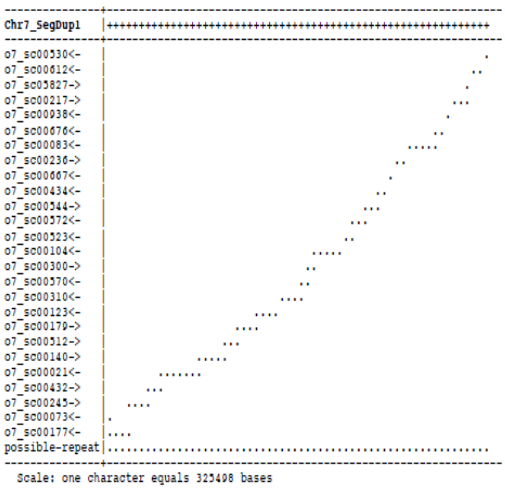


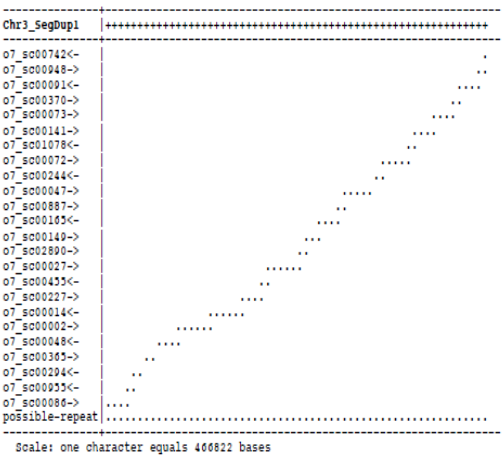
Supplementary Figure 7 continued on next page.

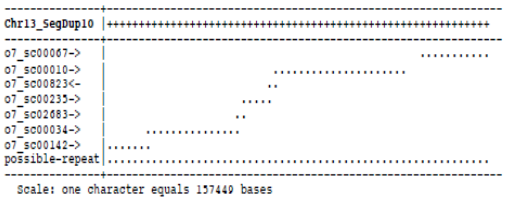


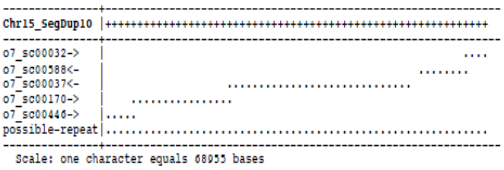
Supplementary Figure 7. Segmental duplications in *E. guineensis*. MUMmer plots generated by the PROmer program are presented for observed segmental duplications within oil palm. Dots represent windows of matching protein translations on the two chromosomes, matches in the same direction are shown in red, opposite strand in blue. Chromosome names are based on the nomenclature of the linkage groups in the T128 genetic map (Supplementary Table 7).

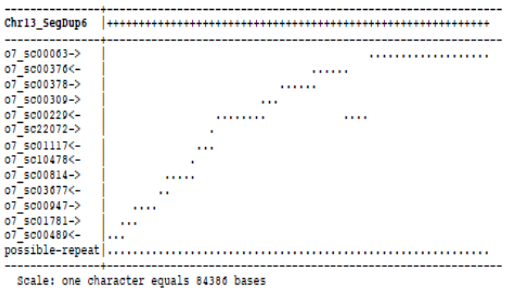
a *E.oleifera*.

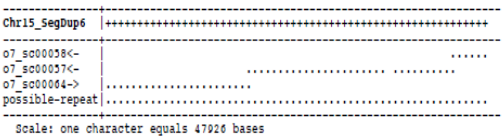




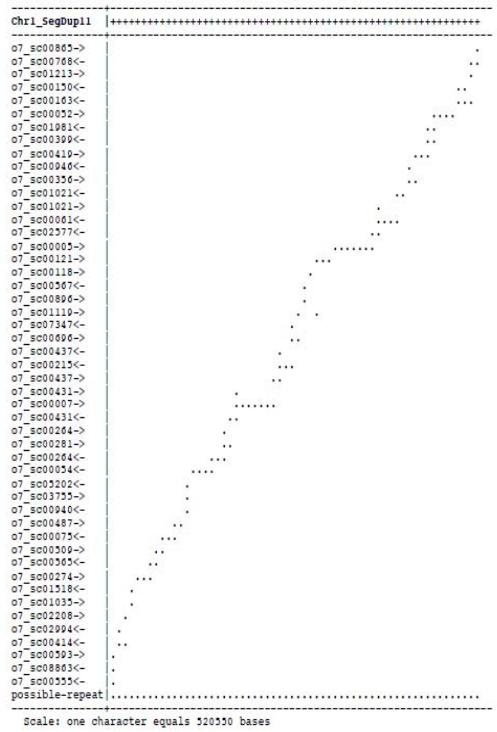


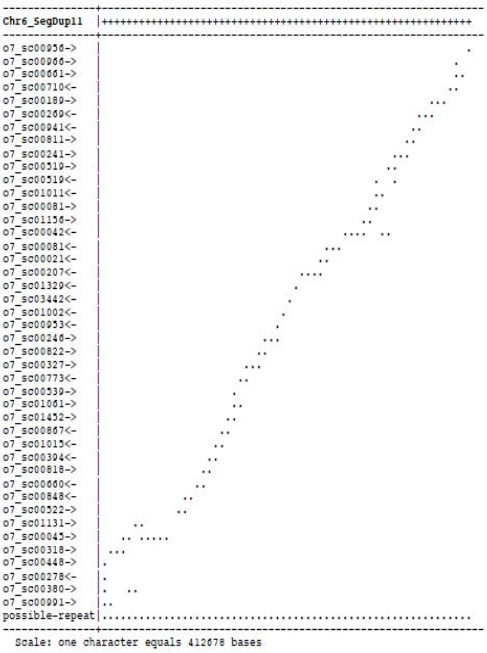


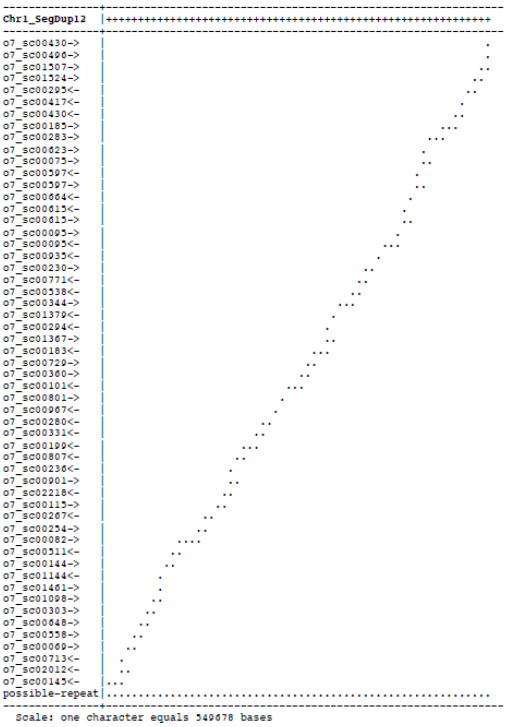


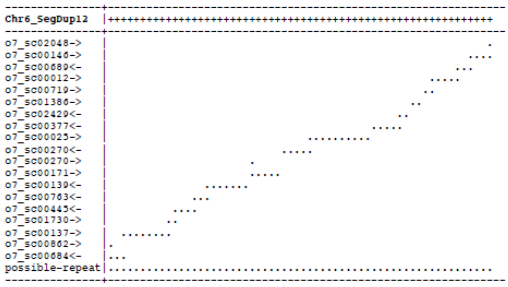


Supplementary Figure 8a continued on next page.

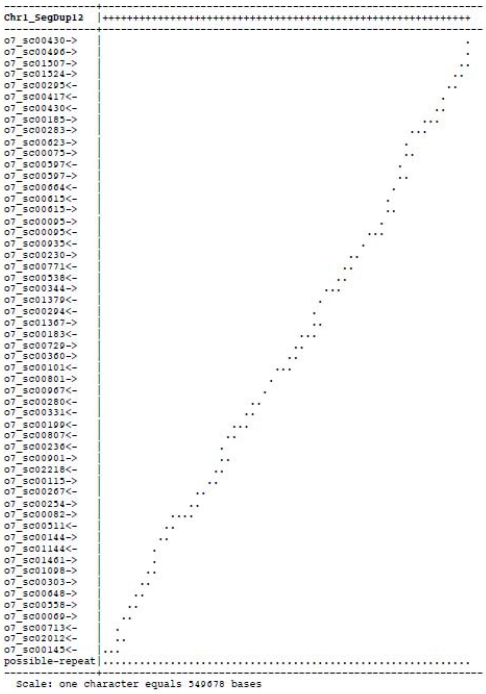


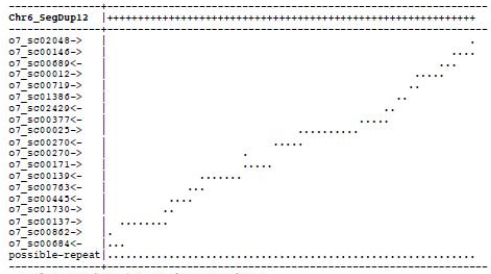


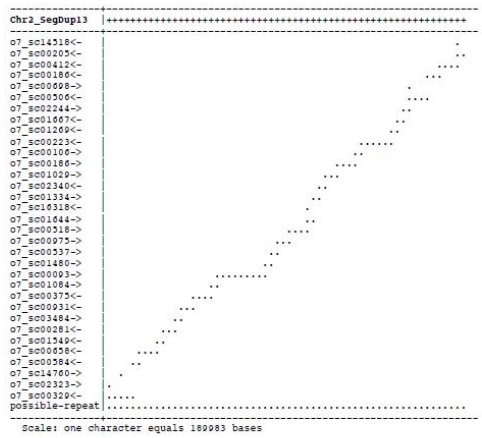


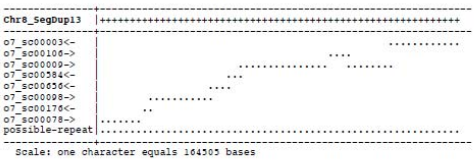


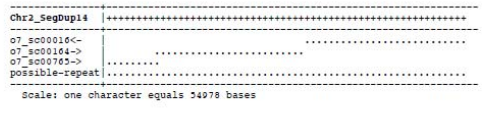
Supplementary Figure 8a continued on next page.

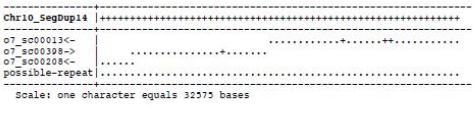


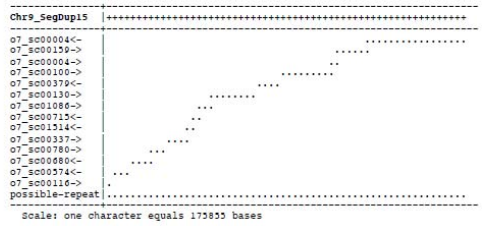


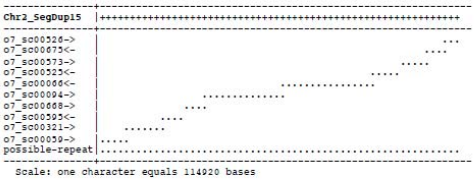




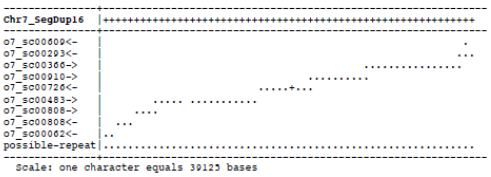


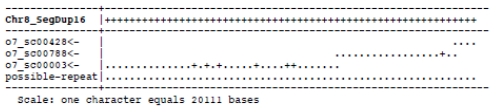


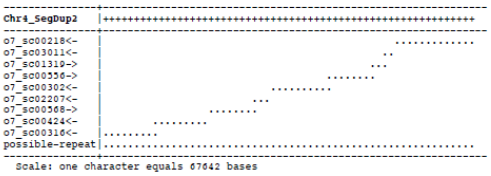


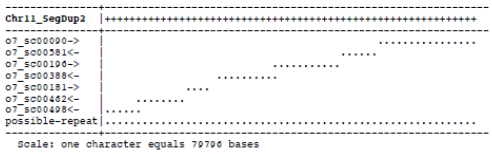


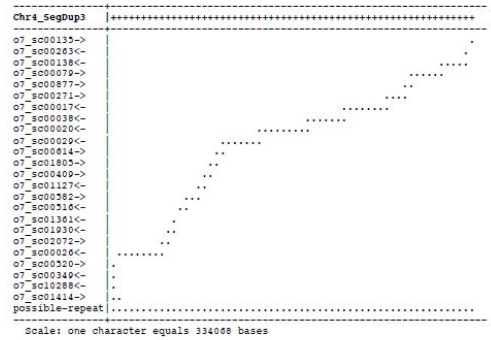
Supplementary Figure 8a continued on next page.

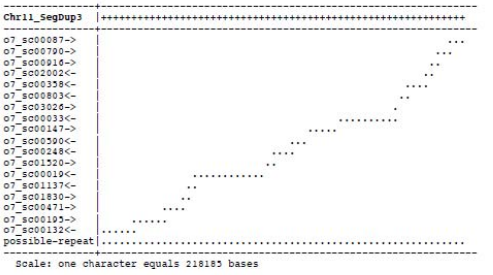


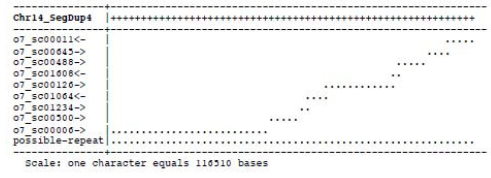


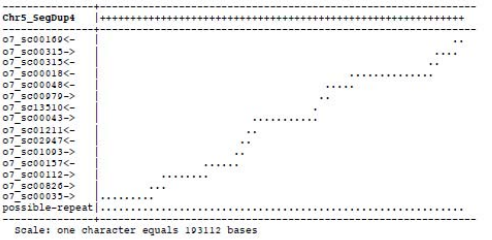




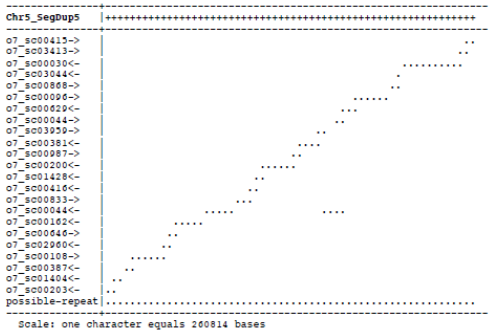


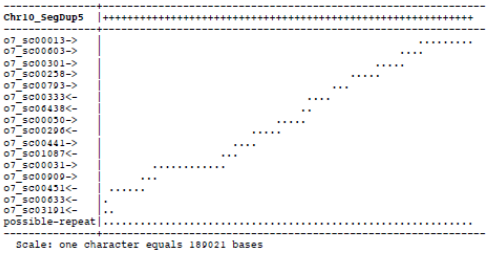


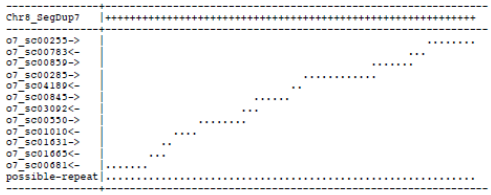


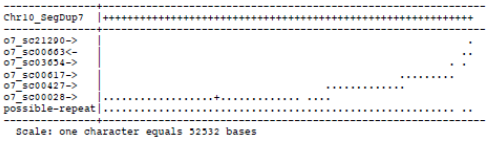


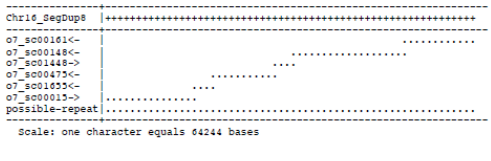
Supplementary Figure 8a continued on next page.

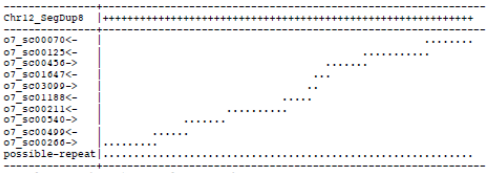


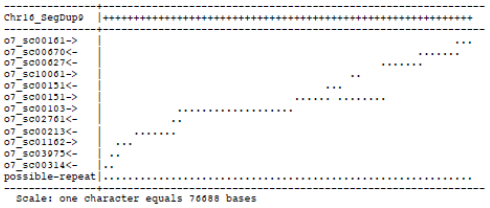


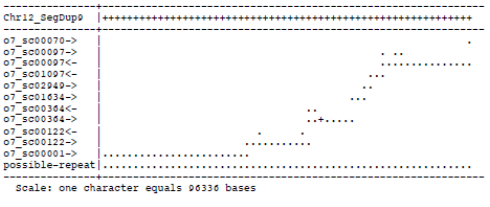






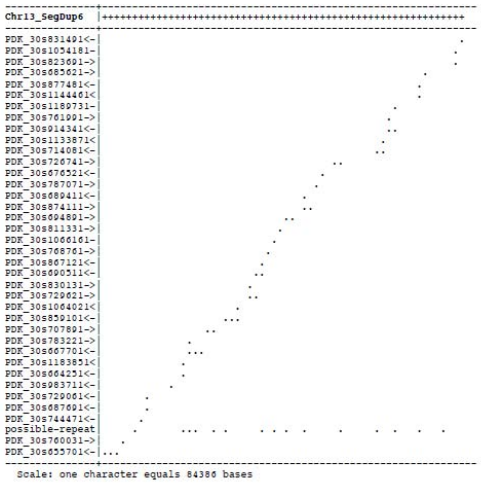


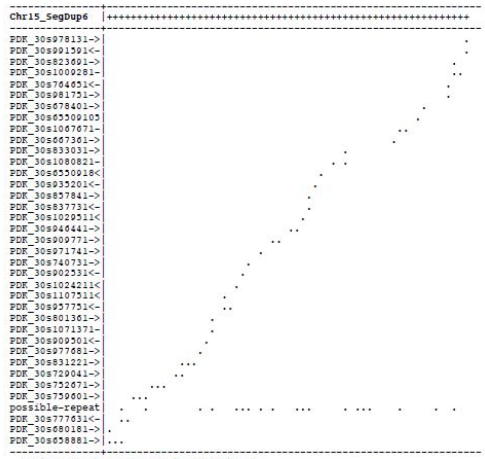


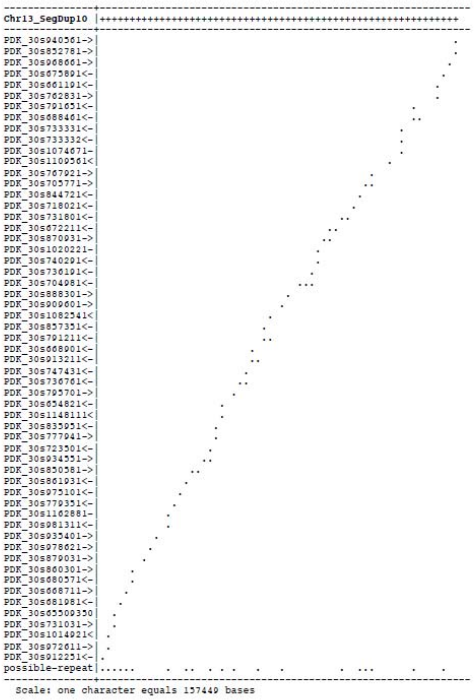


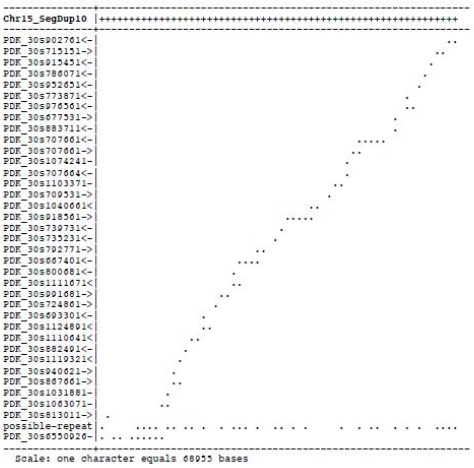
Supplementary Figure 8b continued on next page.

b *P.dactylifer*

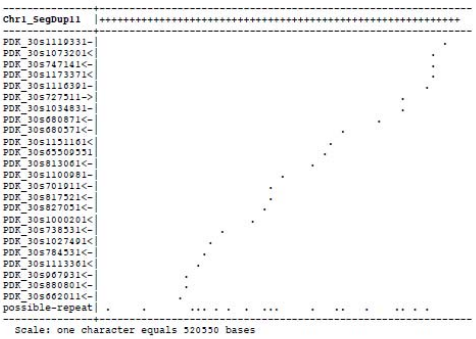


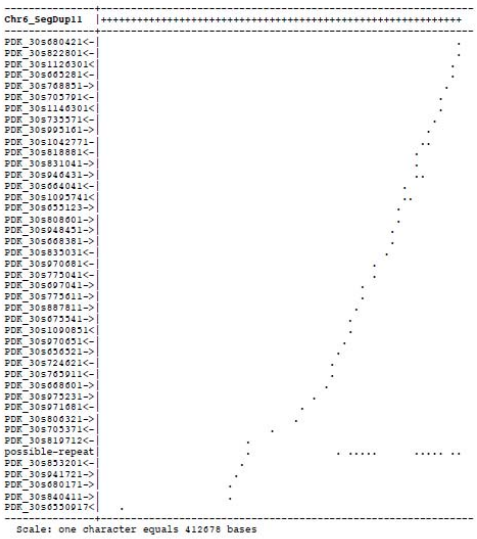


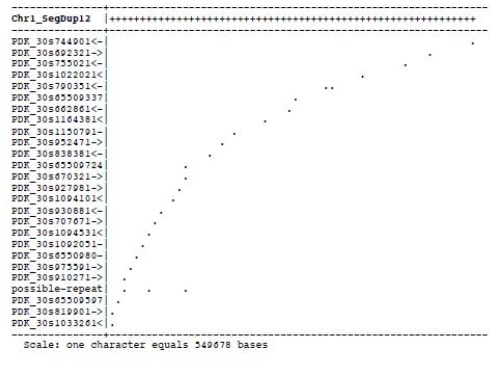


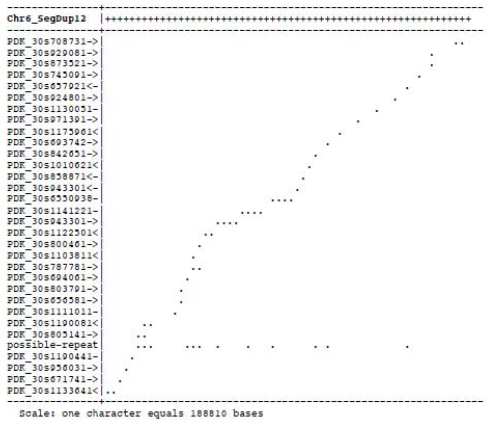


Supplementary Figure 8b continued on next page.

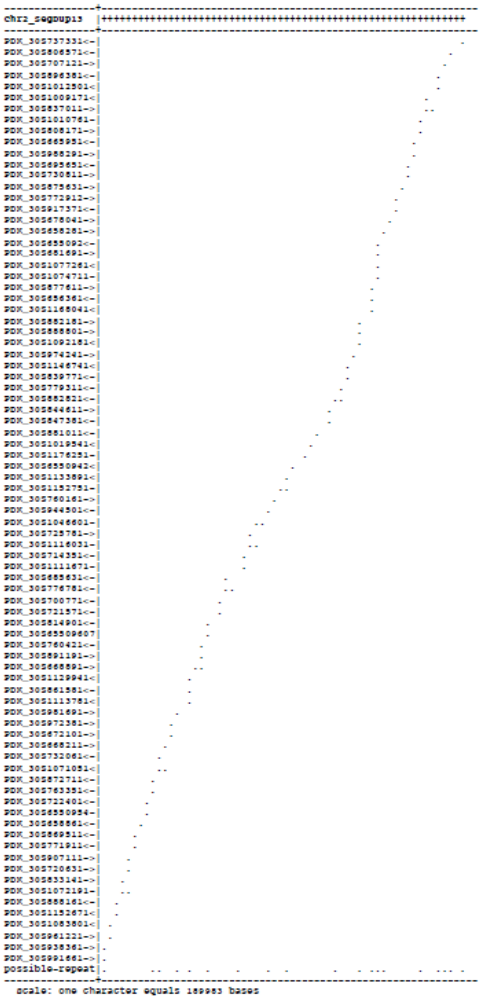


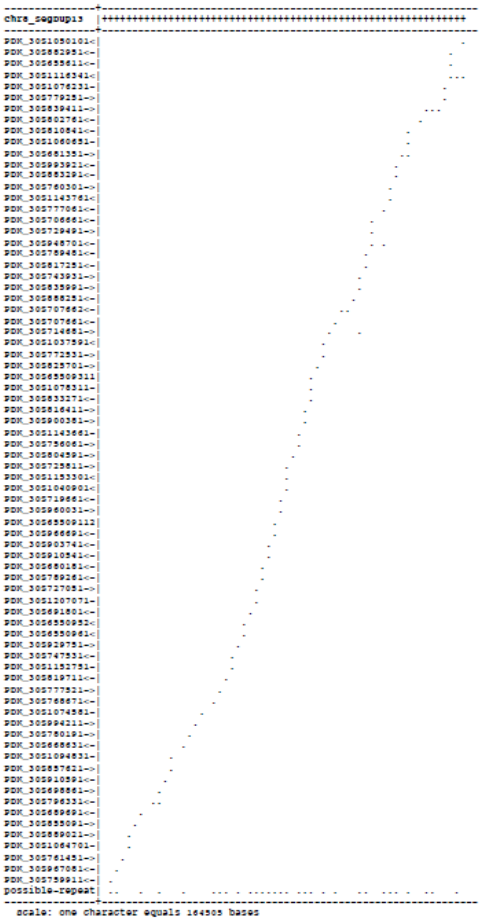




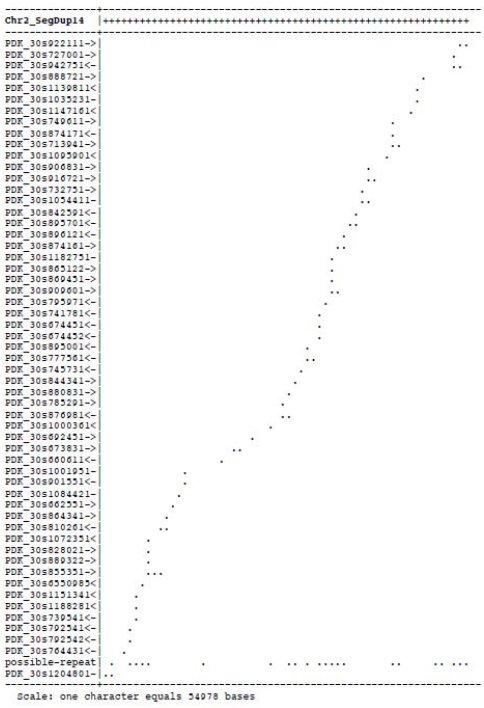


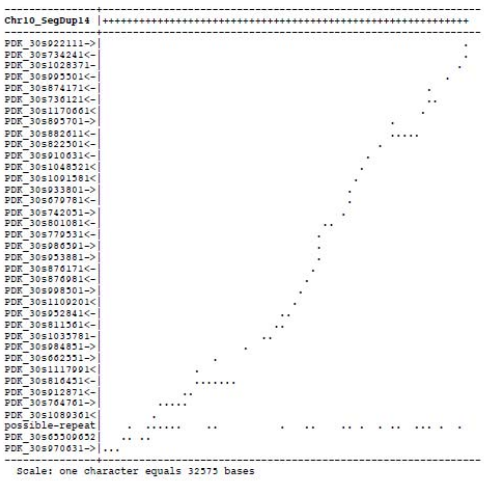
Supplementary Figure 8b continued on next page.



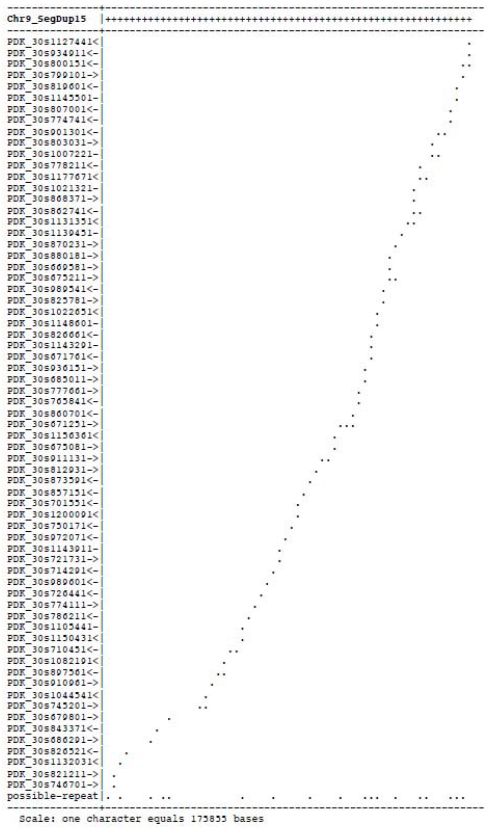


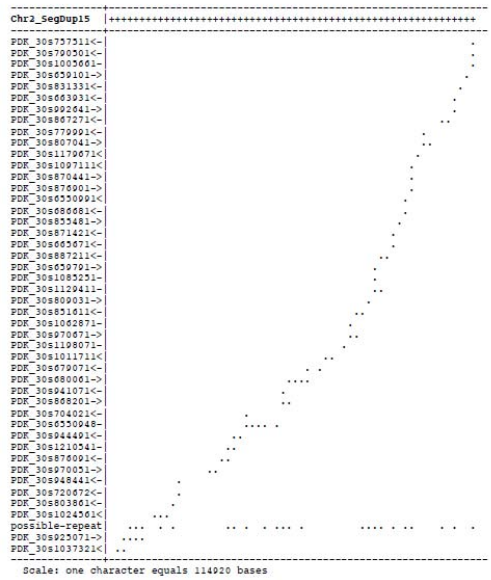
Supplementary Figure 8b continued on next page.



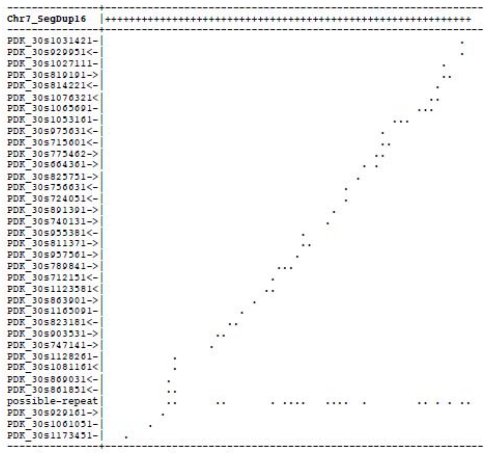


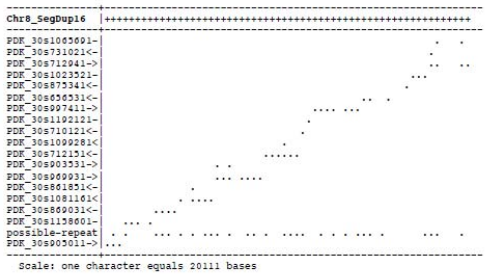
Supplementary Figure 8b continued on next page.

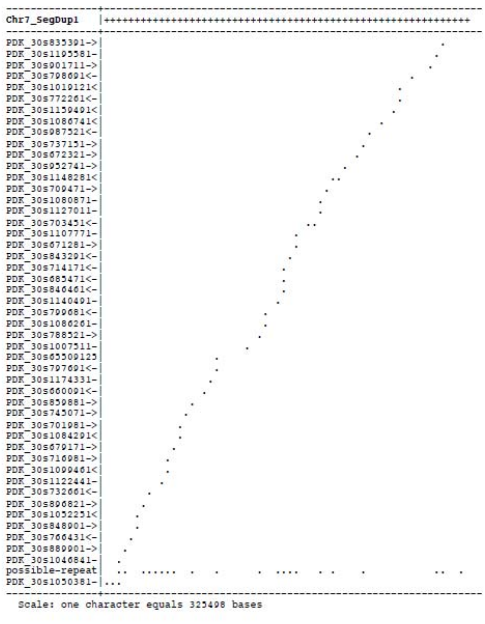


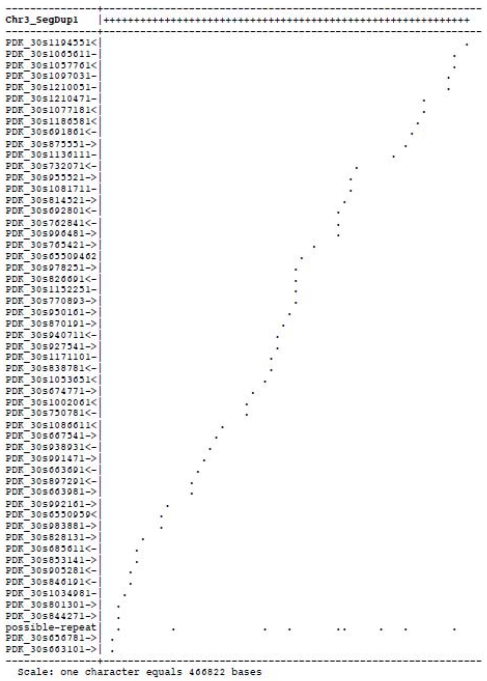


Supplementary Figure 8b continued on next page.

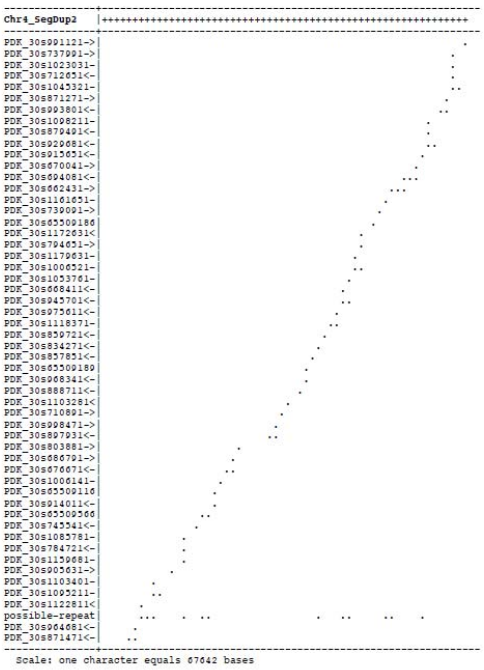


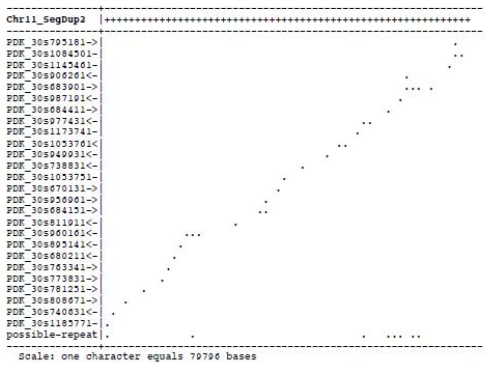




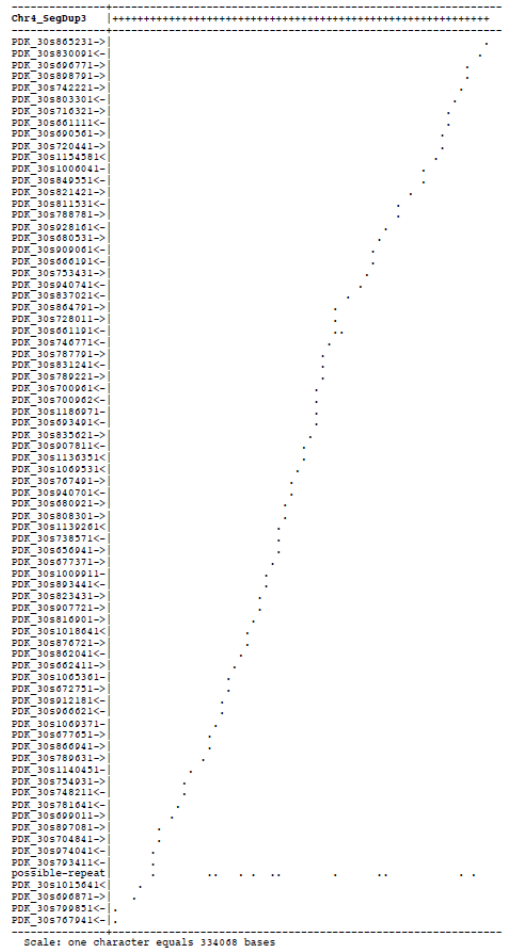


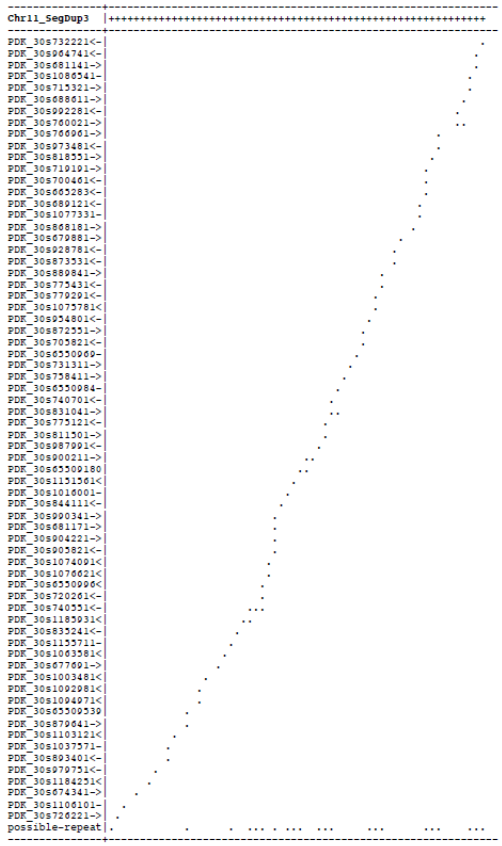
Supplementary Figure 8b continued on next page.



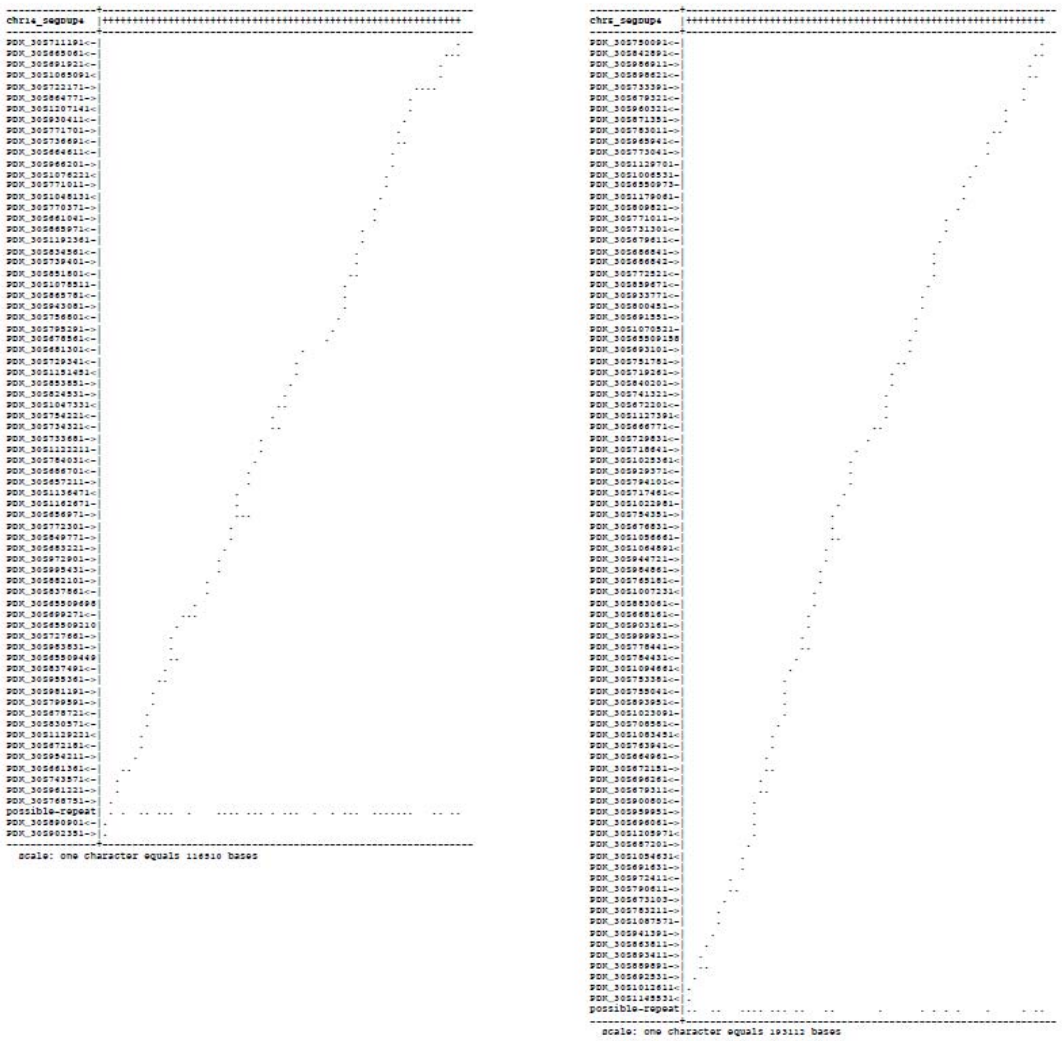


Supplementary Figure 8b continued on next page.

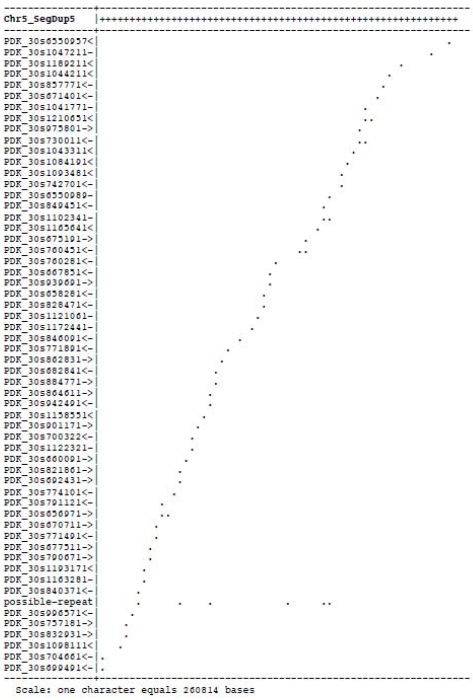


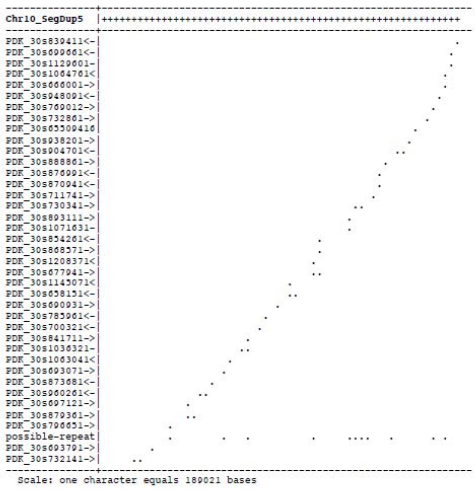


Supplementary Figure 8b continued on next page.

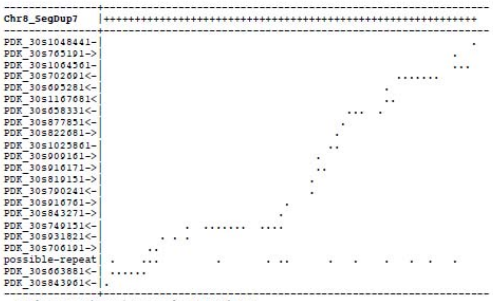


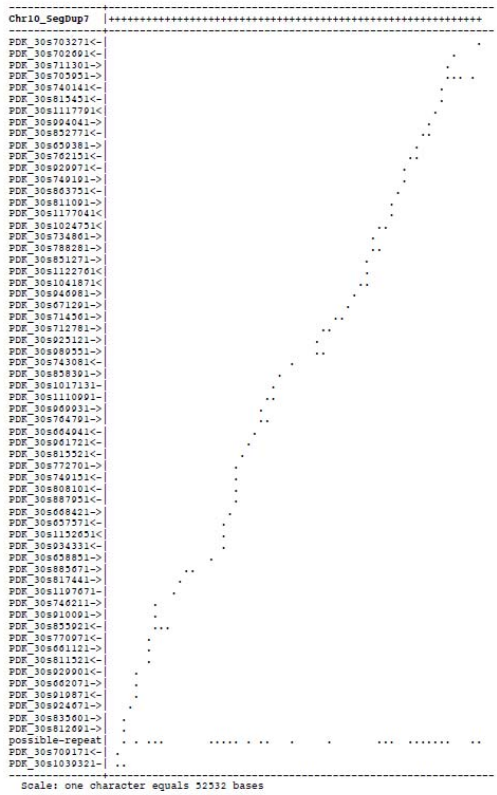
Supplementary Figure 8b continued on next page.



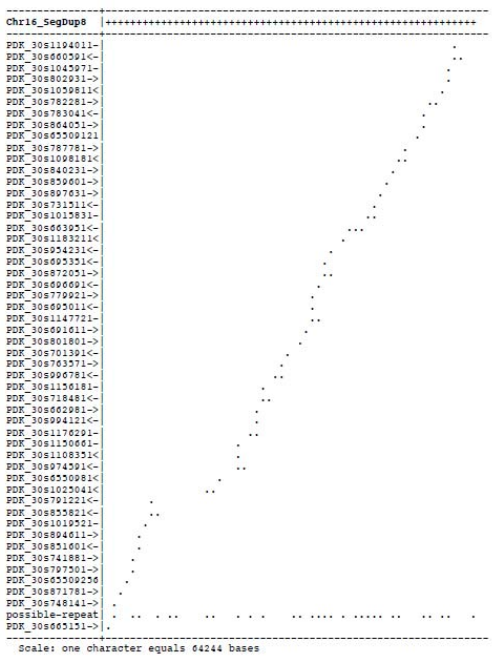


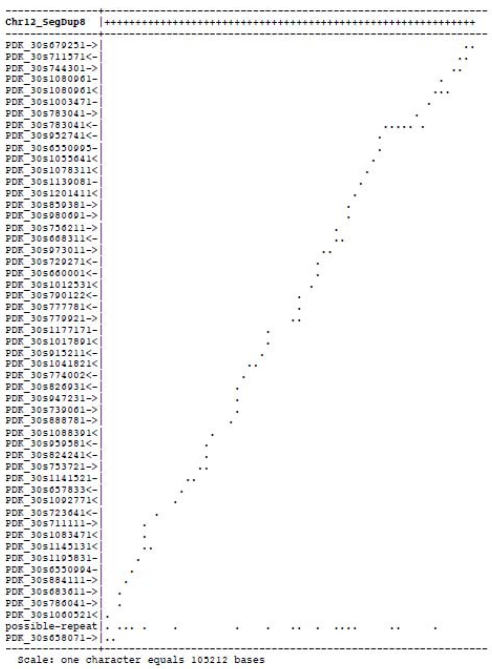
Supplementary Figure 8b continued on next page.



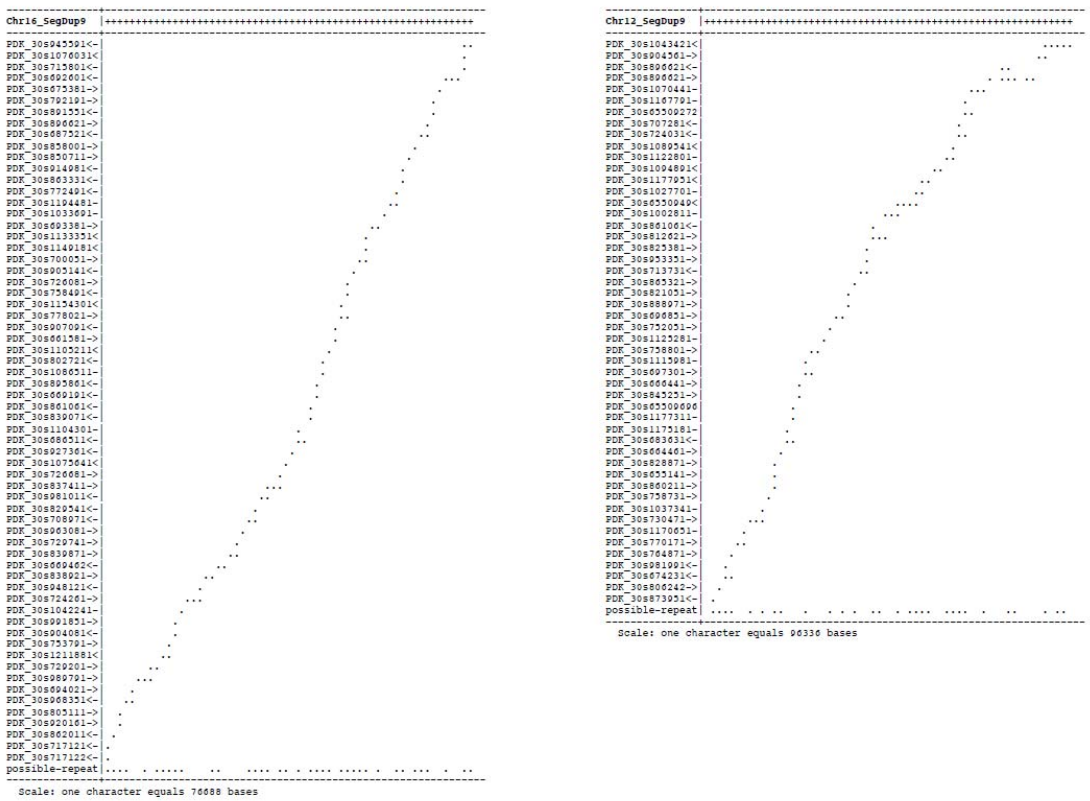


Supplementary Figure 8b continued on next page.

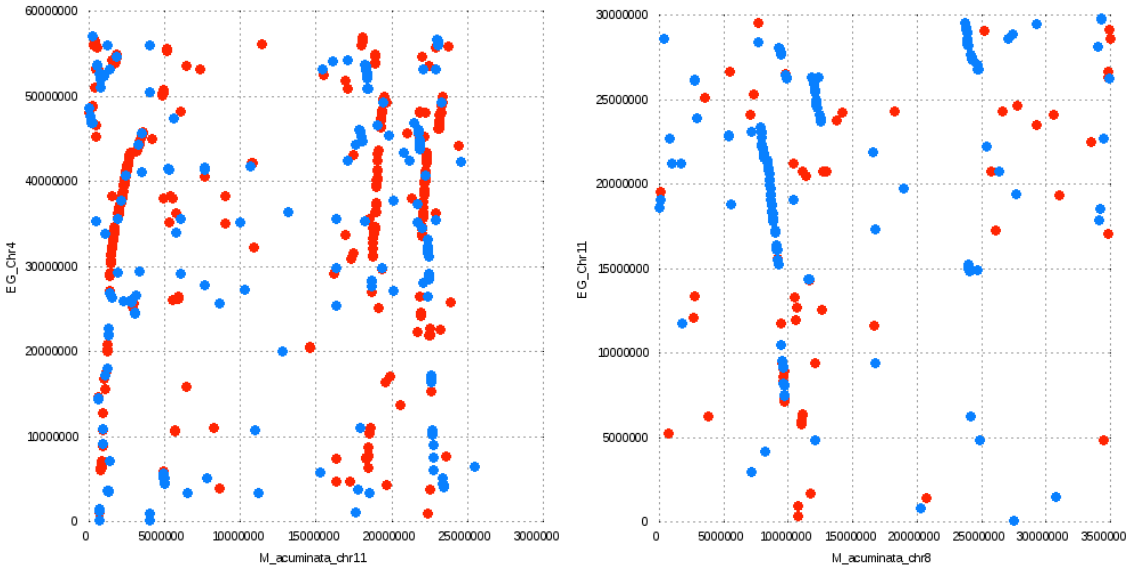




Supplementary Figure 8b continued on next page.



Supplementary Figure 8. **Segmental duplications. a,** Validating segmental duplications in *E.oleifera*. For each segmental duplication in *E. guineensis*, we verified the unique representation of the two corresponding EG-5 linkage groups in the O7 Build in *E.oleifera* using comparative genomics. The Segmental Duplication (SegDup) is plotted as a line of + signs each representing a number of bp (indicated below each plot). O7 scaffolds on the y-axis) correspond to each duplicated EG-5 region (x-axis). Plots for the two halves of the duplication in EG-5 are presented side-by-side. Note that some smaller overlapping scaffolds have been removed for clarity. **b,** Validating segmental duplications in *P.dactylifera*. For each EG-5 linkage group involved in a segmental duplication, we verified the unique representation of that region in the *P. dactylifera* scaffold assembly as above.



Supplementary Figure 9. Synteny between banana and oil palm. Two mummer plots showing synteny between *E. guineensis* chromosomes 4 and 11 with *M. acuminata* chromosomes 11 and 8. Plots show protein level conservation, and imply that several ancient internal rearrangements have occurred. Matches in the same direction are shown in red, opposite strand in blue.

Supplementary Tables.

Supplementary Table 1. *Pisifera* genome fragment and linker library statistics

		All Reads			Useful Reads	
Target Lib (Kb)	Plates	Reads (M)	Seq Cvg	50C50 Rate	Reads (M)	Seq Cvg
Fragments	41.0	53.5	11.6	-	99.6	20.0
0.75	4.0	5.4	1.0	63%	3.4	0.6
1.5	8.0	9.1	1.6	60%	5.4	1.0
3	3.0	2.5	0.5	54%	1.4	0.3
5	21.5	23.5	4.5	53%	12.5	2.4
8	8.5	11.8	2.3	51%	6.0	1.2
10	10.0	11.5	1.9	27%	3.1	0.6
13	8.0	10.1	1.9	46%	4.6	0.9
20	11.0	15.2	2.7	43%	6.6	1.2
Total	115.0	142.6	28.1		142.6	28.1

For fragment (non-linker) and linker libraries of the indicated insert sizes (Target Lib (Kb)), the number of 454 full plate runs (Plates), number of reads in millions (Reads), and fold sequence coverage generated (Seq Cvg) are provided. All Reads heading refers to total reads and raw sequence coverage. Useful Reads heading refers to linker library reads in which the center adapter, as well as at least 50 bases of sequence before and after the center adapter, were sequenced (50C50). Those reads which included genome sequence, but were not 50C50, were removed from linker library read groups and added to the fragment library read group. Total raw sequence fold coverage reported in Supplementary Table 1 is greater than that reported in the main text. Raw sequence coverage reported above (28.1x) is based on the total base pairs of sequence generated. Raw sequence coverage reported in the main text (26x) is based on quality filter passed reads used by Newbler in the genome assembly.

Supplementary Table 2. *Oleifera* genome fragment and linker library statistics

		All Reads			Useful Reads	
Target Lib (Kb)	Plates	Reads (M)	Seq Cvg	50C50 Rate	Reads (M)	Seq Cvg
Fragments	34.0	49.8	10.0	-	88.6	17.1
0.75	4.0	4.1	0.7	58%	2.4	0.4
1.5	4.5	5.3	0.9	63%	3.3	0.6
3	1.0	1.1	0.2	56%	0.6	0.1
5	27.5	31.3	5.5	57%	17.9	3.2
8	7.5	7.9	1.4	47%	3.7	0.7
10	14.0	17.3	3.3	42%	7.3	1.4
13	1.0	1.5	0.3	49%	0.7	0.1
20	8.5	12.1	2.4	48%	5.8	1.1
Total	102.0	130.3	24.7		130.3	24.7

Plate runs, reads and sequence coverage are provided as described in Supplementary Table 1.

Supplementary Table 3. BAC Pool Fragment and Linker Library Statistics

		All Reads			Useable Reads	
Target Lib (Kb)	Plates	Reads (M)	Seq Cvg	50C50 Rate	Reads (M)	Seq Cvg
Fragments	4.0	3.6	0.6	-	8.1	1.5
1.5	0.5	0.6	0.1	49%	0.3	0.1
3	4.5	4.2	0.8	47%	2.0	0.4
5	0.5	0.7	0.1	49%	0.4	0.1
8	1.5	2.0	0.4	47%	1.0	0.2
10	0.5	0.9	0.2	45%	0.4	0.1
Total	11.5	12.2	2.2		12.2	2.2

Plate runs, reads and sequence coverage are provided as described in Supplementary Table 1.

Supplementary Table 4. The oil palm genome sequence

Genome[a]	N	Size (Mb)	N50 (Kb)	Largest (Mb)	G+C content	% of P5 Scaffolds
P5 Build contigs	1,309,411	1,549	3.14	0.13	39%	-
P5 Build scaffolds	40,360	1,535	1,045	22.10	37%	100%
EG-5 Linked chromosomes	16	658	44,350	68.43	37%	43%
EG-5 Linked + P5 unlinked	40,072	1,535	1,270	68.43	37%	100%
Scaffold gaps	166,221	478	-	-	-	31%
Observed repetitive sequence[b]	-	282	-	-	39%	18%
Estimated repeat content	-	760	-	-	-	50%

Genes[c]	N	Size (Mb)	Mapped	Unmapped	G+C Content	% of P5 Scaffolds
Total predicted	158,946	92.1	-	-	50%	6%
With RefSeq support[d]	34,802	34.8	74%	26%	50%	2%
With RNASeq support[e]	15,311	6.8	63%	37%	52%	0%
Known retroelements[f]	67,169	38.2	33%	67%	50%	2%
ab initio only	41,664	12.3	60%	40%	51%	1%

[a]Characteristics of the *E. guineensis* genome assembly.

[b]Many partial, degraded and unknown retroelements included

[c]Characteristics of gene predictions for the *E. guineensis* genome.

[d]Gene predictions matching genes in RefSeq database version 35.

[e]Gene predictions supported by transcriptome sequencing, but without RefSeq match.

[f]Gene predictions matching elements in RepBase and/or TIGR Gramineae repeat database.

The assembly resulted in 1,309,411 contigs and 40,360 scaffolds. Scaffold sizes range from 2.0Kb to 22.1Mb. The scaffold N50 (minimal scaffold length of the smallest set of scaffolds that are able to span 50% of the genome) is 1.045Mb, and the combined total length of the assembly (P5 Build) is 1.535Gb, representing 85% of the estimated 1.8Gb genome size (Supplementary Table 5). Comparison of the P5 Build to genetic linkage maps (Supplementary Fig. 4, Methods) resulted in unambiguous positioning of 366 molecular markers within 304 of the longest P5 Build scaffolds, forming a new assembly, “EG-5 Linked”. This new assembly includes 16 genetic scaffolds (one scaffold for each of the 16 oil palm chromosomes) and 658Mb of sequence. EG-5 Linked scaffold sizes range from 21.3Mb to 68.4Mb, and the EG-5 Linked assembly has an N50 of

44.3Mb. Including the EG-5 Linked and P5 Build (unlinked) together, the combined assembly contains 40,072 scaffolds with an N50 of 1.26Mb.

Gene predictions from the P4 build were generated using Glimmer and SNAP resulting in 151,252 predictions from Glimmer and 158,897 predictions from SNAP. The two were screened for overlap in the P4 build, and overlapping predictions from SNAP were removed from the set resulting in 158,946 predicted transcripts in the P4 build. Gene predictions were then classified by BLAST searching (Eval < 10^{-5}) to reference genomes of *A. thaliana*, *O. sativa*, curated RefSeq35 (with retro-elements removed), and *E. guineensis* transcriptome libraries. In addition, gene predictions were screened against the TIGR Gramineae repeat database, as well as *pisifera* and *oleifera* repetitive sequence databases (see Repetitive sequence detection). Predictions were then grouped as follows: RefSeq support: 34,802 genes; retroelements: 67,169; Transcriptome only: 15,311; Gene predictions with no supporting evidence: 41,664.

Supplementary Table 5. Genome size as determined by flow cytometry

	E. guineensis		
Measurement	*dura* (Mb)	*pisifera* (Mb)	*E. oleifera* (Mb)
1	1801	1,806	1,820
2	1850	1,832	1,825
3	1825	1,836	1,801
4	1820	1,828	1,818
Average	1824	1,825	1,816
Std. Dev.	20	13	10
Average (*E. guineensis*)	1,825		
Std. Dev. (*E. guineensis*)	16		

The genome sizes of *E. guineensis* (*dura* and *pisifera*) and *E. oleifera* were determined by flow cytometry. Four independent measurements were made for each genome (personal communication, Dr. K. Arumuganathan, Virginia Mason Research Center).

The estimated genome size of *E. guineensis* is similar to previous flow cytometry studies[48,49]. Interestingly, the genome size of *E. oleifera* is similar to that of *E. guineensis*. However, the genome size of *E. oleifera* reported here is almost twice of that previously reported for an *E. oleifera* palm from Surinam[49]. The fibrous and lignified nature of the Surinam *E. oleifera* leaf samples most likely contributed to poor isolation of nuclei, and the subsequent lower genome size estimate for *E. oleifera*[49]

Supplementary Table 6. Transcriptome library sequencing and assembly details

Tissue	Species or cross	# Reads	Total Bases	Mean Length	Median Length	# Isotigs	Assembly Group: Dura (A)	Oleifera (B)	Pisifera (C)	Tenera (D)	Norm TC (E)	Abnorm TC (F)	All TC (G)	All G (H)
E. guineensis materials:														
spear leaf	Dura	611,301	234,730,367	741	790	17,849	1							1
Root	Dura	544,460	186,827,044	653	704	13,688	1							1
Floret before anthesis (BA)	Dura	430,617	136,972,388	644	701	13,025	1							1
Floret/fruit after anthesis (AA)	Dura	491,450	159,452,355	527	554	7,857	1							1
Spear leaf	Pisifera	353,102	126,429,118	566	583	11,969			1					1
Leaf	Pisifera	563,194	209,578,762	630	667	11,204			1					1
Root	Pisifera	471,740	156,149,475	619	662	7,897			1					1
Pollen	Pisifera	582,078	219,457,909	743	799	6,344			1					1
Pollen	Pisifera	435,176	128,684,750	522	558	4,528			1					1
Roots of seedling	Tenera	605,420	209,106,303	618	650	11,648				1				1
Floret before anthesis (BA)	Tenera	363,461	119,007,405	482	504	9,993				1				1
Floret/fruit after anthesis (AA)	Tenera	507,517	172,546,989	604	636	17,849				1				1
Mesocarp	Tenera	621,614	202,496,615	720	763	18,936				1				1
Kernel	Tenera	576,743	171,345,655	677	708	17,071				1				1
Mesocarp	Tenera	642,368	189,574,467	682	726	16,060				1				1
Kernel	Tenera	462,814	168,860,793	717	750	15,970				1				1
Total *E. guineensis:*		8,263,055	2,791,220,395			201,888								
Tissue cultured materials:														
Female flower	Normal DxP	424,077	122,673,087	606	631	13,625					1		1	1
Female flower	Normal DxP	535,423	201,686,040	722	762	16,591					1		1	1
Shoot apex	Normal DxP	423,331	151,604,419	673	704	14,513					1		1	1
Female flower	Abnormal DxP	393,127	138,049,299	589	604	12,281						1	1	1
Female flower	Abnormal DxP	456,919	159,486,885	574	601	12,130						1	1	1
Shoot apex	Abnormal DxP	406,206	136,110,784	708	748	12,996						1	1	1
Total tissue culture:		2,639,083	909,610,514			82,136								
E. oleifera materials:														
Spear leaf	Oleifera	512,619	177,817,256	702	761	4,785		1						
Whole fruit	Oleifera	617,890	211,555,425	671	726	12,513		1						
Whole fruit	Oleifera	583,884	182,911,966	598	642	10,302		1						
Whole fruit pooled from 3 sections of the bunch	Oleifera	663,995	223,986,124	469	494	8,100		1						
Stalk	Oleifera	405,170	123,182,197	604	663	7,606		1						
Stalk	Oleifera	549,553	188,401,707	637	692	6,349		1						
Stalk pooled from 3 sections of the bunch	Oleifera	569,983	180,314,388	479	502	5,249		1						
Spikelet pooled from 3 sections of the bunch	Oleifera	655,148	222,344,145	497	524	8,209		1						
Total *E. oliefera:*		4,558,242	1,510,513,208			63,113								
Total All:		15,460,380	5,211,344,117			347,137								

Tissue sources for each of the 30 RNA sequencing libraries is presented. Tissues and trees were selected to maximize library complexity for genome annotation purposes as well as to measure interesting expression variation for high value breeding traits such as oil production and clonal stability. Libraries were individually assembled using the Newbler program from Roche 454, resulting in transcriptome sets ranging from 4,258 to 18,936 isotigs per library. In addition to individual assemblies, group assemblies were generated for *dura* (A), *E. oleifera* (B), *pisifera* (C), and *tenera* tissue libraries (D), as well as normal clonal (E), abnormal (mantled) clonal (F), and all clonal tissues (G) derived from crosses of *dura* and *pisifera*. Validation of gene models for *E. guineensis* was based on homology to isotigs from all combined *E. guineensis* libraries (H).

Supplementary Table 7. Chromosomes and linkage groups in oil palm

Chromosome	T128_codominant	P2/Billotte	Size (bp)
CHR 1	LG1	lg_8	68,432,966
CHR 2	LG7	lg_4	65,556,141
CHR 3	LG10	lg_1	60,058,032
CHR 4	LG6	lg_11	57,248,047
CHR 5	LG12	lg_12	51,953,839
CHR 6	LG5	lg_10	44,354,769
CHR 7	LG8	lg_6	43,453,266
CHR 8	LG9	lg_2	40,192,799
CHR 9	LG4	lg_7	38,054,796
CHR 10	LG13	lg_15	31,889,635
CHR 11	LG11	lg_14	30,067,610
CHR 12	LG14	lg_13	28,799,275
CHR 13	LG15	lg_9	27,816,170
CHR 14	LG2	lg_3	24,378,543
CHR 15	LG16	lg_16	24,313,565
CHR 16	LG3	lg_5	21,370,583

Linkage groups in the T-128 Nigerian population[50] as well as the DxP (P2) pseudotestcross population which followed the nomenclature of Billotte et al. (2005)[51], were aligned using common SNP and SSR markers. Cytogenetically, individual oil palm chromosomes are difficult to distinguish but fall into 4 groups, comprising (I) the largest chromosome (which hybridizes to 5SrDNA), (II) 8 medium chromosomes; (III) 6 small chromosomes and (IV) a small acrocentric chromosome, which carried the 18S-25S rDNA[17, 52]. Attempts were made to identify these chromosomes by FISH using RFLP clones from early genetic maps, but the results were often ambiguous. We have re-numbered the chromosomes according to the size of sequence scaffolds in the EG5 linked build. The chromosomes correspond to Linkage Groups in each mapping population as shown.

Supplementary Table 8. Comparative genomics of oil palm

			Genome			
		# Gene Models	*A.thaliana*	*P.dactylifera*	*E. guineensis*	*M.acuminata*
Gene	*A. thaliana*	27,416	99.34%	79.90%	80.16%	80.49%
	P. dactylifera	28,889	81.88%	99.64%	94.38%	87.16%
	E. guineensis	34,802	86.00%	96.28%	99.93%	93.64%
	M. acuminata	36,549	79.94%	83.53%	83.54%	97.23%

Rows (Gene) represent previously reported gene models for *Arabidopsis* (*A. thaliana*), banana (*M. acuminata*) and date palm (*P. dactylifera*), as well as oil palm (*E. guineensis*) gene models reported here. Columns (Genome) represent the complete genome sequences for each species. The percentage of query genes matched to the target genome is shown for each pairwise comparison of gene models to genome.

Supplementary Table 9. Transcriptome sequencing of oil palm mesocarp and kernel at time points prior to and at the peak of oil accumulation.

	Arabidopsis homolog	Gene Description	Protein/Gene Abbreviation	15waa mesocarp	20waa mesocarp	10waa kernel	15waa kernel
Plastidial Fatty Acid Synthesis				214.84	117.62	266.24	138.86
		Biotin Carboxyl Carrier Protein of					
	AT5G15530	Heteromeric ACCase	BCCP2	-	-	-	-
	AT2G43710	Stearoyl-ACP Desaturase	fab2	1,773.88	1,656.33	1,176.32	621.84
	AT1G08510	Acyl-ACP Thioesterase B	FatB	170.90	197.34	2,204.94	917.18
	AT4G13050	Acyl-ACP Thioesterase A	FATA	60.41	74.40	100.75	15.59
	AT5G10160	Hydroxyacyl-ACP Dehydrase	HAD	191.95	71.93	248.28	94.00
		Pyruvate Dehydrogenase β subunit, E1b component of Pyruvate Dehydrogenase					
	AT2G34590	Complex	PDH (E1-β)	290.44	158.81	175.39	103.33
	AT5G46290	Ketoacyl-ACP Synthase I	KASI	288.19	93.62	387.27	194.96
	AT2G22230	Hydroxyacyl-ACP Dehydrase	HAD	22.46	14.31	25.75	4.56
	AT1G24360	Ketoacyl-ACP Reductase	KAR	1,004.04	265.72	596.18	261.49
	AT1G43800	Stearoyl-ACP Desaturase	DES6	190.10	1.27	78.63	14.07
		Carboxyltransferase α Subunit of					
	AT2G38040	Heteromeric ACCase	CT α	220.36	71.08	313.56	219.13
		Dihydrolipoamide Acetyltransferase, E2 component of Pyruvate Dehydrogenase					
	AT1G34430	Complex	EMB3003 (PDH-E2)	206.58	39.11	133.94	48.33
		Dihydrolipoamide Dehydrogenase, E3 component of Pyruvate Dehydrogenase					
	AT4G16155	Complex	LPD2 (PDH-E3)	112.52	26.83	157.21	76.23
	AT3G02630	Stearoyl-ACP Desaturase	DES5	3.45	1.57	293.95	267.08
	AT4G25050	Acyl Carrier Protein	ACP4	618.23	313.44	360.22	177.64
		Pyruvate Dehydrogenase α subunit, E1a component of Pyruvate Dehydrogenase					
	AT1G01090	Complex	PDH (E1-α)	215.69	94.70	162.47	92.72
	AT3G55290	Ketoacyl-ACP Reductase	KAR	151.05	85.26	372.77	321.22
		Pyruvate Dehydrogenase β subunit, E1b component of Pyruvate Dehydrogenase					
	AT1G30120	Complex	PDH (E1-β)	-	-	-	-
	AT2G05990	Enoyl-ACP Reductase	ENR1 (MOD1)	227.57	51.22	220.83	104.44
		Dihydrolipoamide Acetyltransferase, E2 component of Pyruvate Dehydrogenase					
	AT3G25860	Complex	LTA2 (PDH-E2)	85.49	17.32	65.75	38.66
	AT3G25110	Acyl-ACP Thioesterase A	FatA	-	-	-	-
	AT3G02610	Stearoyl-ACP Desaturase	DES2	-	-	-	-
	AT5G35360	Biotin Carboxylase of Heteromeric ACCase	BC	321.85	89.36	328.69	197.83
		Biotin Carboxyl Carrier Protein of					
	AT5G16390	Heteromeric ACCase	BCCP1	89.35	14.81	37.77	70.90
	AT1G74960	Ketoacyl-ACP Synthase II	KASII (fab1)	104.29	116.25	372.63	246.87
	AT5G16240	Stearoyl-ACP Desaturase	DES1	-	-	-	-
	AT1G62640	Ketoacyl-ACP Synthase III	KASIII	55.70	46.69	140.33	41.57
	AT1G62610	Ketoacyl-ACP Reductase	KAR	-	-	-	-
		Dihydrolipoamide Dehydrogenase, E3 component of Pyruvate Dehydrogenase					
	AT3G16950	Complex	LPD1 (PDH E3)	-	-	-	-
	AT2G30200	Malonyl CoA : ACP Malonyltransferase	MCMT	90.75	27.32	83.59	36.23

Supplementary Table 9 continued on next page.

Supplementary Table 9 continued.

Arabidopsis homolog	Protein Family	Protein/Gene Abbreviation	15waa mesocarp	20waa mesocarp	10waa kernel	15waa kernel
Other Fatty Acid and TAG Synthesis Related Genes			37.10	30.01	676.35	344.85
AT1G23240	Caleosin	CALO	-	-	-	-
AT1G36180	Acetyl-CoA Carboxylase (Plastidial, Homomeric)	ACC2	-	-	-	-
AT1G70670	Caleosin	CALO	2.87	2.82	0.14	1.10
AT1G70680	Caleosin	CALO	56.50	75.89	16.13	72.93
AT1G77590	Long-Chain Acyl-CoA Synthetase	LACS9	202.82	153.19	113.78	84.62
AT2G04350	Long-Chain Acyl-CoA Synthetase	LACS8	-	-	-	-
AT2G25890	Oil-Body Oleosin	OBO	0.04	-	324.38	122.92
AT3G11470	Holo-ACP Synthase (plastid)	HACPS	10.71	7.45	7.13	16.27
AT3G12120	Oleate Desaturase	FAD2	239.01	106.43	15.93	9.96
AT3G18570	Oil-Body Oleosin	OBO	1.04	1.86	289.25	315.13
AT3G27660	Oil-Body Oleosin	OBO	-	-	-	-
AT3G54320	Wrinkled1 (AP2/ERWEBP Transcription Factor)	WRI1	13.04	8.68	985.40	526.00
AT4G01900	PII protein	PII	72.65	104.25	7.91	11.20
AT4G10020	Steroleosin	STERO	0.50	0.53	1,808.14	864.62
AT4G14070	Acyl-ACP Synthetase	AAE15	-	-	-	-
AT4G25140	Oil-Body Oleosin	OBO	5.16	2.63	6,681.17	2,543.51
AT4G26740	Caleosin	CALO	2.29	1.29	424.91	425.66
AT5G36880	AcetylCoA synthase	ACS	134.24	133.22	24.48	19.60
AT5G40420	Oil-Body Oleosin	OBO	1.15	1.94	2,828.17	1,883.39
AT5G55240	Caleosin	CALO	-	-	-	-
Plastid Polar Lipids Glycero			41.02	48.86	15.26	15.92
AT5G66450	Phosphatidate Phosphatase	LPP-ε2	1.25	-	-	-
AT4G26770	CDP-DAG Synthase (plastidial)	CDP-DAGS	-	-	-	-
AT2G45150	CDP-DAG Synthase (plastidial)	CDP-DAGS	-	-	-	-
AT3G11170	Linoleate Desaturase	FAD7	5.80	5.10	1.54	2.26
AT4G30950	Oleate Desaturase	FAD6	16.96	8.83	7.67	34.28
AT4G27030	Phosphatidylglycerol Desaturase	FAD4	-	-	-	-
AT3G60620	CDP-DAG Synthase (plastidial)	CDP-DAGS	59.79	28.77	3.22	9.31
AT3G03530	Phospholipase C (Non specific)	NPC4	2.98	2.23	0.67	6.09
AT4G30580	1-acylglycerol-3-phosphate acyltransferase	LPAAT1	14.44	10.16	60.32	12.11
AT2G11810	Monogalactosyldiacylglycerol Synthase	MGD3	-	-	-	-
AT5G40610	NAD-dependent Glycerol-3-Phosphate Dehydrogenase (plastidial)	GPDH	0.05	0.07	60.08	73.75
AT5G20410	Monogalactosyldiacylglycerol Synthase	MGD2	2.01	29.78	0.36	1.94
AT3G50920	Phosphatidate Phosphatase	LPP-ε1	7.23	10.56	2.73	2.61
AT5G03080	Phosphatidate Phosphatase	LPP-γ	290.68	160.26	132.96	124.36
AT5G05580	Linoleate Desaturase	FAD8	361.44	423.71	1.38	1.54
AT3G05630	Phospholipase D, zeta	PLD--ζ2	-	-	-	-
AT3G11670	Digalactosyldiacylglycerol Synthase	DGD1	79.34	268.57	56.02	98.30
AT4G00550	Digalactosyldiacylglycerol Synthase	DGD2	17.70	22.98	2.53	5.15
AT1G32200	Glycerol-3-Phosphate Acyltransferase (plastidial)	ATS1/ACT1	23.08	39.42	0.59	0.73
AT3G15850	MGDG Desaturase (palmitate-specific)	FAD5	-	-	-	-
AT4G33030	UDP-Sulfoquinovose Synthase	SQD1	5.34	28.99	10.44	5.05
AT2G39290	Phosphatidylglycerol-Phosphate Synthase (plastidial)	PGP1	66.32	67.24	22.50	20.06
AT5G01220	UDP-sulfoquinovose:DAG sulfoquinovosyltransferase	SQD2	17.62	113.62	0.64	1.49
AT3G16785	Phospholipase D, zeta	PLD-ζ1	42.26	37.62	12.98	10.76
AT3G56040	UDP-Glucose Pyrophosphorylase	UGPase, UGP3	6.56	32.38	0.30	0.60
AT1G19800	Permease-like Protein of Inner Chloroplast Envelope	TGD1	13.74	10.10	11.95	7.58
AT4G31780	Monogalactosyldiacylglycerol Synthase	MGD1	72.90	18.92	23.02	11.86

Supplementary Table 9 continued on next page.

Supplementary Table 9 continued.

	Arabidopsis homolog	Protein Family	Protein/Gene Abbreviation	15waa mesocarp	20waa mesocarp	10waa kernel	15waa kernel
Mitochondrial Lipid Synthesis				53.86	46.25	25.49	15.59
	AT4G24160	Cardiolipin-Specific Deacylase	CLD	65.13	43.58	6.81	12.04
	AT1G06520	Glycerol-3-Phosphate Acyltransferase	GPAT1	1.02	0.62	-	-
	AT2G44620	Acyl Carrier Protein	ACP	126.25	139.44	121.82	41.00
	AT1G04640	Lipoyltransferase	LT	9.46	6.43	0.74	1.70
	AT3G17240	Mitochondrial Dihydrolipoamide Dehydrogenase, branched chain alpha-ketoacid dehydrogenase complex		-	-	-	-
	AT1G65290	Acyl Carrier Protein	ACP	212.05	329.20	99.35	53.94
	AT5G60340	Hydroxyacyl-ACP Dehydrase	HAD	112.07	64.07	100.67	56.47
	AT3G60510	Mitochondrial Enoyl-CoA Hydratase		0.43	-	-	-
	AT4G31810	Mitochondrial Enoyl-CoA Hydratase		36.13	31.73	10.94	10.22
	AT2G04540	Ketoacyl-ACP Synthase	KAS	9.11	17.29	5.10	3.51
	AT4G16700	Phosphatidylserine Decarboxylase	PSD	20.34	20.22	13.54	7.00
	AT4G04870	Cardiolipin Synthase	CLS	33.30	20.39	16.93	24.21
	AT3G02600	Phosphatidate Phosphatase (mitochondrial)	PP	1.84	1.74	10.27	2.06
	AT3G45770	Enoyl-ACP Reductase	ER	113.42	98.03	44.41	26.76
	AT1G78690	Cardiolipin Transacylase	TAZ	10.22	11.77	5.21	2.59
	AT2G39290	Phosphatidylglycerol-Phosphate Synthase	PGPS1	66.32	67.24	22.50	20.06
	AT5G47630	Acyl Carrier Protein	ACP	7.55	5.20	3.93	2.42
	AT3G10370	FAD-dependent Glycerol-3-Phosphate Dehydrogenase	SDP6	17.11	9.04	2.64	0.74
	AT3G05510	Cardiolipin Transacylase	TAZ	184.95	16.25	10.00	10.36
	AT1G48030	Mitochondrial Dihydrolipoamide Dehydrogenase, branched chain alpha-ketoacid dehydrogenase complex		85.82	72.12	54.04	46.50
	AT2G20860	Lipoate Synthase	LS or LIP1	18.58	16.96	6.29	5.90
Sphingolipid Synthesis				51.47	70.59	17.66	15.35
	AT2G37940	Inositolphosphorylceramide Synthase	ERH1	55.75	70.52	18.19	25.70
	AT3G25540	Ceramide Synthase	LAG1	-	-	-	-
	AT5G19200	Ketosphinganine Reductase	KSR	-	-	-	-
	AT2G34690	Sphingosine Transfer Protein	ACD11	94.44	78.85	4.86	4.37
	AT2G29525	Inositolphosphorylceramide Synthase	ERHL1	-	-	-	-
	AT5G03080	Long Chain Base 1-Phosphate Phosphatase	LCBPP	290.68	160.26	132.96	124.36
	AT3G06060	Ketosphinganine Reductase	KSR	75.49	155.86	40.53	24.48
	AT4G20870	Fatty Acid 2-hydroxylase	FAH2	-	-	1.02	5.79
	AT3G48780	Subunit of Serine Palmitoyltransferase (LCB2b)	LCB2b	-	-	-	-
	AT2G46210	Sphingobase-DB Desaturase	SLD2	3.35	5.78	0.17	2.97
	AT5G23670	Subunit of Serine Palmitoyltransferase (LCB2a)	LCB2a	233.04	483.10	23.04	36.52
	AT4G22330	Ceramidase	CES1	40.25	11.99	23.15	19.81
	AT1G13580	Ceramide Synthase	LAG13	33.54	35.41	8.14	20.86
	AT3G54020	Inositolphosphorylceramide Synthase	ERHL2	34.38	107.17	12.64	17.96
	AT4G39670	Sphingosine Transfer Protein	STP	11.58	60.24	-	1.43
	AT3G58490	Long Chain Base 1-Phosphate Phosphatase	LCBPP	57.36	186.33	17.61	6.63
	AT3G19260	Ceramide Synthase	LOH2	9.65	8.32	10.95	5.34
	AT4G04930	Dihydrosphingosine Delta-4 Desaturase	DSD1	15.13	27.31	17.26	13.65
	AT5G51290	Ceramide Kinase	CERK	25.75	27.09	3.92	5.02
	AT1G69640	Sphingobase C4-Hydroxylase	SBH1	46.94	117.05	40.76	24.28
	AT5G23450	Long Chain Base Kinase	LCBK	98.39	76.78	26.93	30.85
	AT4G21540	Long Chain Base Kinase	LCBK	33.01	61.18	15.23	7.92
	AT3G61580	Sphingobase-DB Desaturase	SLD1	2.98	0.31	9.09	0.14
	AT1G27980	Dihydrosphingosine Phosphate Lyase	DPL1	45.80	59.15	9.17	9.76
	AT2G34770	Fatty Acid 2-hydroxylase	FAH1	49.98	15.25	17.70	7.00
	AT1G14290	Sphingobase C4-Hydroxylase	SBH2	7.37	20.10	0.44	0.33
	AT4G36480	Subunit of Serine Palmitoyltransferase (LCB1)	LCB1	169.18	202.44	59.25	32.79
	AT2G19880	Glucosylceramide Synthase (UDP-glucose-dependent)	GCS	7.02	6.10	1.60	1.92

Supplementary Table 9 continued on next page.

Supplementary Table 9 continued.

	Arabidopsis homolog	Protein Family	Protein/Gene Abbreviation	15waa mesocarp	20waa mesocarp	10waa kernel	15waa kernel
Oxidation				76.22	66.54	16.80	15.07
	AT5G65940	Hydroxyisobutyryl-CoA Hydrolase	HIBCH	41.51	49.07	8.96	20.55
	AT2G30660	Hydroxyisobutyryl-CoA Hydrolase	HIBCH	-	-	-	-
	AT5G65110	Acyl-CoA Oxidase	ACX2	22.75	40.28	1.75	5.54
	AT2G33150	Ketoacyl-CoA Thiolase	KAT2/PED1	296.64	342.28	13.82	28.65
	AT1G76150	Peroxisomal Enoyl-CoA Hydratase 2	ECH2	51.05	63.33	13.93	15.67
	AT4G39850	ABC Acyl Transporter (peroxisomal)	CTS	41.35	57.43	11.57	14.02
	AT4G29010	Multifunctional Protein	AIM1	658.36	283.65	207.46	127.21
	AT4G16760	Acyl-CoA Oxidase	ACX1	153.12	183.15	2.06	5.98
	AT1G01710	Acyl-CoA Thioesterase	ACT2	52.45	61.84	0.40	0.73
	AT1G06310	Acyl-CoA Oxidase	ACX6	-	-	-	-
	AT5G27600	Long-Chain Acyl-CoA Synthetase (peroxisomal)	LACS7	5.63	10.76	-	-
	AT3G51840	Acyl-CoA Oxidase	ACX4	36.50	43.55	2.43	6.02
	AT5G48880	Ketoacyl-CoA Thiolase	PKT1	-	-	-	-
	AT4G14430	Enoyl CoA isomerase	ECI2	93.65	106.79	16.61	13.78
	AT5G43280	Dienoyl CoA Isomerase	DCI1	72.00	51.25	15.44	14.47
	AT2G30650	Hydroxyisobutyryl-CoA Hydrolase	HIBCH	-	-	-	-
	AT2G35690	Acyl-CoA Oxidase	ACX5	4.31	-	-	1.08
	AT1G04710	Ketoacyl-CoA Thiolase	KAT1	31.53	48.89	5.88	6.70
	AT3G05970	Long-Chain Acyl-CoA Synthetase (peroxisomal)	LACS6	55.05	41.01	2.86	0.82
	AT1G06290	Acyl-CoA Oxidase	ACX3	25.26	18.59	2.38	3.32
	AT4G00520	Acyl-CoA Thioesterase	ACT1	-	-	-	-
	AT1G65520	Enoyl CoA isomerase (peroxisomal)	ECI1	14.72	6.09	0.66	3.15
	AT3G12800	Dienoyl-CoA Reductase	Red	32.72	46.68	10.78	6.60
	AT3G06860	Multifunctional Protein	MFP2	140.60	142.33	86.32	87.38
Plastid Transporters				170.30	107.56	178.30	88.85
	AT1G15500	Nucleoside triphosphate transporter	NTT2	28.71	21.18	30.95	19.72
	AT1G61800	Glucose 6 phosphate:phosphate translocator	GPT2	374.70	122.60	53.07	38.50
	AT1G80300	Nucleoside triphosphate transporter	NTT1	383.92	420.01	337.14	331.24
	AT5G16150	Glucose transporter	GLT1	167.72	73.01	0.40	3.22
	AT5G33320	Phosphoenolpyruvate transporter	PPT1	220.56	95.48	817.19	228.11
	AT5G46110	Triose phosphate translocator	TPT	16.47	20.67	9.33	1.14
	AT5G54800	Glucose 6 phosphate:phosphate translocator	GPT1	-	-	-	-

Supplementary Table 9 continued on next page.

Supplementary Table 9 continued.

	Arabidopsis homolog	Protein Family	Protein/Gene Abbreviation	15waa mesocarp	20waa mesocarp	10waa kernel	15waa kernel
Glycolysis				231.46	110.10	323.97	97.47
	AT1G09780	Phosphoglyceromutase	i-PGM	727.71	400.46	590.34	137.84
	AT1G12000	Phosphofructokinase, pyrophosphate-dependant	PFP	94.08	70.88	61.53	54.65
	AT1G12900	Glyceraldehyde 3 phosphate dehydrogenase	GapA-2	-	-	-	-
	AT1G13440	Glyceraldehyde 3 phosphate dehydrogenase	GapC2	1,417.45	1,071.54	4,739.34	1,154.93
	AT1G16300	Glyceraldehyde 3 phosphate dehydrogenase	GapCp-2	312.61	65.62	190.99	89.45
	AT1G18270	Fructose 1,6 bis phosphate aldolase	FBA	20.70	33.46	1.57	1.74
	AT1G20950	Phosphofructokinase, pyrophosphate-dependant	PFP	-	-	-	-
	AT1G22170	Phosphoglyceromutase	d-PGM	10.20	8.15	56.55	56.00
	AT1G32440	Pyruvate kinase	pPK	66.67	29.82	72.20	54.22
	AT1G42970	Glyceraldehyde 3 phosphate dehydrogenase	Plastid-GapB	0.05	5.31	0.05	0.05
	AT1G56190	3-Phosphoglycerate kinase	PGK	19.08	15.69	22.12	9.72
	AT1G74030	Enolase	Plastid-ENO	64.00	33.72	54.12	22.24
	AT1G76550	Phosphofructokinase, pyrophosphate-dependant	PFP	116.82	33.59	587.02	164.52
	AT1G78050	Phosphoglyceromutase	d-PGM	-	-	-	-
	AT1G79530	Glyceraldehyde 3 phosphate dehydrogenase	GapCp-1	9.73	14.80	4.59	8.80
	AT1G79550	3-Phosphoglycerate kinase	PGK	822.66	278.75	1,159.64	618.23
	AT2G01140	Fructose 1,6 bis phosphate aldolase	FBA	277.81	143.63	153.27	148.67
	AT2G21170	Triose phosphate isomerase	Plastid-TPI	123.43	117.77	21.37	11.87
	AT2G21330	Fructose 1,6 bis phosphate aldolase	FBA	4.07	3.55	0.08	0.07
	AT2G22480	Phosphofructokinase, ATP-dependant	PFK5	131.15	117.08	81.00	41.65
	AT2G24270	Glyceraldehyde 3 phosphate dehydrogenase (non-phosphorylating)	Cytosol-np-GAP	41.01	55.54	2.69	16.99
	AT2G36460	Fructose 1,6 bis phosphate aldolase	FBA	2,428.65	858.77	1,342.62	414.02
	AT2G36530	Enolase	Cytosol-ENO	301.77	140.29	2,760.26	306.25
	AT2G36580	Pyruvate kinase	PK	0.72	2.70	5.56	4.10
	AT3G04120	Glyceraldehyde 3 phosphate dehydrogenase	GapC1	13.71	31.22	2.41	0.26
	AT3G08590	Phosphoglyceromutase	i-PGM	2.45	3.92	125.24	31.54
	AT3G12780	3-Phosphoglycerate kinase	PGK	25.71	32.89	26.99	15.10
	AT3G22960	Pyruvate kinase	pPK	198.60	35.90	747.84	282.36
	AT3G26650	Glyceraldehyde 3 phosphate dehydrogenase	GapA-1	8.66	5.55	14.82	1.15
	AT3G30841	Phosphoglyceromutase	i-PGM	16.74	20.81	82.28	39.32
	AT3G45090	3-Phosphoglycerate kinase	PGK	2.45	0.32	2.53	2.33
	AT3G49160	Pyruvate kinase	pPK	17.85	18.79	5.81	3.20
	AT3G52930	Fructose 1,6 bis phosphate aldolase	FBA	2,293.17	698.16	2,359.20	578.05
	AT3G52990	Pyruvate kinase	PK	594.17	153.73	144.05	152.99
	AT3G55440	Triose phosphate isomerase	Cytosol-TPI	381.99	360.38	511.10	136.62
	AT3G55650	Pyruvate kinase	PK	-	-	-	-
	AT3G55810	Pyruvate kinase	PK	-	-	-	-
	AT4G04040	Phosphofructokinase, pyrophosphate-dependant	PFP	-	-	-	-
	AT4G09520	Phosphoglyceromutase	i-PGM	0.37	2.60	-	0.65
	AT4G26270	Phosphofructokinase, ATP-dependant	PFK3	1,025.09	418.72	344.40	170.29
	AT4G26390	Pyruvate kinase	PK	-	-	-	-
	AT4G26530	Fructose 1,6 bis phosphate aldolase	FBA	1.37	-	4.02	-
	AT4G29220	Phosphofructokinase, ATP-dependant	PFK1	-	-	-	-
	AT4G32840	Phosphofructokinase, ATP-dependant	PFK6	-	-	-	-
	AT4G38970	Fructose 1,6 bis phosphate aldolase	FBA	26.49	32.49	7.89	26.51
	AT5G08570	Pyruvate kinase	PK	383.15	330.60	258.83	210.44
	AT5G47810	Phosphofructokinase, ATP-dependant	PFK2	103.11	44.46	64.79	57.35
	AT5G52920	Pyruvate kinase	pPK-β1	183.52	91.77	386.71	143.42
	AT5G56350	Pyruvate kinase	PK	118.89	45.75	388.24	22.61
	AT5G56630	Phosphofructokinase, ATP-dependant	PFK7	40.58	30.51	13.86	5.66
	AT5G60760	3-Phosphoglycerate kinase	PGK	53.56	69.13	71.75	51.47
	AT5G61450	3-Phosphoglycerate kinase	PGK	16.71	16.63	24.59	16.22
	AT5G61580	Phosphofructokinase, ATP-dependant	PFK4	-	-	-	-
	AT5G63680	Pyruvate kinase	PK	-	-	-	-

Supplementary Table 9 continued on next page.

Supplementary Table 9 continued.

	Arabidopsis homolog	Protein Family	Protein/Gene Abbreviation	15waa mesocarp	20waa mesocarp	10waa kernel	15waa kernel
Malate and Pyruvate Metabolism				104.00	66.10	47.18	29.56
	AT1G04410	Malate dehydrogenese	MDH	-	-	-	-
	AT1G53240	Malate dehydrogenese	MDH	176.23	102.88	65.30	25.59
	AT1G53310	Phosphoenolpyruvate carboxylase	PEPC1	274.57	302.73	51.93	60.54
	AT1G68750	Phosphoenolpyruvate carboxylase	PEPC4	20.02	12.59	2.36	1.32
	AT1G79750	Malic enzyme (NADP)	Plastid-ATNADP-ME4	-	-	-	-
	AT2G13560	Malic enzyme (NAD)	NAD-ME1	41.48	30.05	3.84	7.29
	AT2G19900	Malic enzyme (NADP)	ATNADP-ME1	-	-	-	-
	AT2G22780	Malate dehydrogenese	MDH	84.78	122.00	36.30	19.31
	AT2G42600	Phosphoenolpyruvate carboxylase	PEPC2	104.25	195.21	9.65	46.39
	AT3G04530	Phosphoenolpyruvate carboxykinase	PEPCK3	9.82	23.17	2.91	1.48
	AT3G14940	Phosphoenolpyruvate carboxylase	PEPC3	88.12	33.45	21.97	51.15
	AT3G15020	Malate dehydrogenese	MDH	3.19	0.40	2.67	0.23
	AT3G47520	Malate dehydrogenese	MDH	65.26	44.56	220.13	66.16
	AT4G00570	Malic enzyme (NAD)	NAD-ME2	51.50	29.40	1.53	1.05
	AT4G15530	Pyruvate orthophosphate dikinase	Plastid-PPDK	17.28	36.57	13.96	18.56
	AT4G37870	Phosphoenolpyruvate carboxykinase	PEPCK1	71.79	49.19	119.83	197.89
	AT5G09660	Malate dehydrogenese	MDH	-	-	-	-
	AT5G11670	Malic enzyme (NADP)	ATNADP-ME2	1,023.80	220.63	372.28	152.52
	AT5G25880	Malic enzyme (NADP)	ATNADP-ME3	91.34	54.78	111.45	40.22
	AT5G43330	Malate dehydrogenese	MDH	569.07	463.27	98.78	58.47
	AT5G56720	Malate dehydrogenese	MDH	-	-	-	-
	AT5G58330	Malate dehydrogenese	MDH	6.73	5.11	1.08	0.79
	AT5G65690	Phosphoenolpyruvate carboxykinase	PEPCK2	2.78	3.94	19.25	2.70
Calvin cycle				-	-	-	-
	AT1G32060	phosphoribulokinase	PRK	65.44	5.59	57.18	40.44
	AT1G67090	Ribulose bis phosphate carboxylase/oxygenase small subunit	RBCS-1A	0.96	29.01	67.74	12.93
	AT5G38410	Ribulose bis phosphate carboxylase/oxygenase small subunit	RBCS-3B	-	-	-	-
	AT5G38420	Ribulose bis phosphate carboxylase/oxygenase small subunit	RBCS-2B	247.57	152.35	88.04	52.30
	AT5G38430	Ribulose bis phosphate carboxylase/oxygenase small subunit	RBCS-1B	-	-	-	-
Sucrose Degradation				134.29	83.34	53.05	38.76
	AT1G12240	Invertase	BFRUCT4	22.08	3.05	-	0.16
	AT1G22650	Invertase	N-INV	-	-	-	-
	AT1G35580	Invertase	CINV1	-	-	-	-
	AT1G62660	Invertase	BFRUCT3	165.82	85.11	0.20	0.08
	AT1G72000	Invertase		-	-	-	-
	AT1G73370	Sucrose synthase	SUS6	0.33	0.18	0.06	0.66
	AT2G36190	Invertase	CWINV4	56.31	67.88	-	0.03
	AT3G05820	Invertase		5.48	1.62	1.54	1.42
	AT3G06500	Invertase		9.84	16.10	0.77	5.57
	AT3G13784	Invertase	CWINV5	-	-	-	-
	AT3G13790	Invertase	CWINV1	108.80	82.03	0.08	0.31
	AT3G43190	Sucrose synthase	SUS4	1,820.34	890.36	903.14	642.89
	AT3G52600	Invertase	CWINV2	43.05	12.15	0.36	-
	AT4G02280	Sucrose synthase	SUS3	227.92	187.42	53.91	66.20
	AT4G09510	Invertase	CINV2	199.59	226.49	5.42	8.80
	AT4G34860	Invertase		-	-	-	-
	AT5G20830	Sucrose synthase	SUS1	11.44	45.18	93.28	38.00
	AT5G22510	Invertase		13.82	47.43	2.23	10.39
	AT5G37180	Sucrose synthase	SUS5	0.97	1.78	-	0.15
	AT5G49190	Sucrose synthase	SUS2	-	-	-	0.61

Supplementary Table 9 continued on next page.

Supplementary Table 9 continued.

	Arabidopsis homolog	Protein Family	Protein/Gene Abbreviation	15waa mesocarp	20waa mesocarp	10waa kernel	15waa kernel
Sugar transport				114.72	116.87	9.63	20.47
	AT1G09960	Sucrose transporter	SUC4	89.48	42.08	28.78	32.45
	AT1G11260	Sugar transporter	STP1	48.32	51.43	2.69	59.80
	AT1G50310	Sugar transporter	STP9	10.68	0.72	3.95	1.41
	AT1G77210	Sugar transporter	STP14	7.40	17.18	8.01	21.99
	AT2G02860	Sucrose transporter	SUC3	426.89	180.59	26.58	92.14
	AT3G19930	Sugar transporter	STP4	40.28	81.15	3.72	1.31
	AT4G02050	Sugar transporter	STP7	28.94	5.05	5.36	6.81
	AT4G21480	Sugar transporter	STP12	-	-	-	-
	AT5G23270	Sugar transporter	STP11	-	-	-	-
	AT5G26340	Sugar transporter	STP13 MSS1	568.80	895.37	0.71	3.09
	AT5G61520	Sugar transporter	STP3	41.12	12.00	26.11	6.18
Starch metabolism				27.53	19.02	8.26	3.24
	AT1G05610	Glucose-1-phosphate adenylyltransferase	ApS2	6.87	1.66	0.50	0.42
	AT1G11720	Starch synthase	SS	10.79	4.99	4.86	5.86
	AT1G27680	Glucose-1-phosphate adenylyltransferase	ApL2	5.39	4.81	2.75	3.35
	AT1G32900	Starch synthase	SS	142.82	102.81	36.89	6.42
	AT4G18240	Starch synthase	SS	2.96	3.87	6.94	4.30
	AT5G19220	ADP-glucose pyrophosphorylase	ApL1	2.72	0.53	0.49	0.08
	AT5G48300	ADP-glucose pyrophosphorylase	ApS1	41.84	26.12	2.24	2.56
	AT5G65685	Starch synthase	SS	6.87	7.36	11.44	2.89
WRI1 related				6.26	3.55	235.02	149.67
	AT2G25170	PICKLE	PKL	24.43	11.74	129.27	82.26
	AT3G54320	Wrinkled1 (AP2/ERWEBP Transcription Factor)	WRI1	13.04	8.68	985.40	526.00
	AT1G28300	LEAFY COTYLEDON 2	LEC2	-	-	5.32	10.03
	AT1G21970	LEAFY COTYLEDON 1	LEC1	-	0.30	112.29	38.82
	AT3G24650	ABA INSENSITIVE 3	ABI3	0.08	0.55	149.73	205.04
	AT3G26790	Transcriptional factor with high similarity to the B3 region of the VP1/ABI3-like proteins	FUS3	-	-	28.08	35.86
Ripening				99.69	38.61	2.47	3.52
	AT1G05010	1-aminocyclopropane-1-carboxylate oxidase	ACO4	116.28	165.27	0.12	10.66
	AT1G06620	2-oxoglutarate-dependent dioxygenase		39.13	10.25	2.85	3.30
	AT2G43870	Pectin lyase-like superfamily protein		11.28	6.12	-	-
	AT3G48290	putative cytochrome P450	CYP71A24	-	-	-	-
	AT3G59850	Pectin lyase-like superfamily protein		8.14	2.54	-	-
	AT3G61510	1-aminocyclopropane-1-carboxylate (ACC) synthase (S-adenosyl-L-methionine methylthioadenosine-lyase	ACS1	3.40	6.46	-	-
	AT4G02290	glycosyl hydrolase	9B13	-	-	5.62	-
	AT4G37770	auxin inducible ACC synthase	ACS8	12.15	-	0.56	2.90
	AT5G47230	ERF/AP2 transcription factor family	ERF5	706.80	156.87	13.07	14.86

Supplementary Table 9 continued on next page.

Supplementary Table 9 continued.

	Arabidopsis homolog	Protein Family	Protein/Gene Abbreviation	15waa mesocarp	20waa mesocarp	10waa kernel	15waa kernel
Abscission				31.51	28.05	15.08	7.09
	AT1G02790	exopolygalacturonase	PGA4	-	-	-	-
	AT1G07640	DOF transcription factors	OBP2	0.13	4.95	0.55	3.02
	AT1G18350	MAP kinase kinase7	ATMKK7	-	-	-	-
	AT1G51660	mitogen-activated map kinase	MKK4	35.18	19.71	11.26	13.89
	AT1G51700	zinc finger protein	ADOF1	24.47	99.82	72.11	28.61
	AT1G60590	Pectin lyase-like superfamily protein		3.12	-	0.35	1.10
	AT2G01270	disulfide isomerase-like (PDIL) protein	QSOX2	29.80	15.28	13.07	17.19
	AT2G23760	BEL family of homeodomain proteins	BLH4	5.14	4.46	0.30	-
	AT2G27990	BEL1-like homeobox gene	BLH8	6.54	8.30	2.45	2.36
	AT2G31880	putative leucine rich repeat transmembrane protein	EVERSHED	26.01	179.89	0.27	0.22
	AT2G41370	a cytoplasmic and nuclear-localized NPR1 like protein with BTB/POZ domain and ankyrin repeats	BOP2	4.20	4.11	1.40	1.38
	AT2G43870	Pectin lyase-like superfamily protein		11.28	6.12	-	-
	AT3G21220	mitogen-activated kinase	KINASE 5	17.25	6.15	10.03	4.65
	AT3G55370	DNA binding transcription factor	OBP3	1.93	3.27	0.12	-
	AT4G34610	BEL1-like homeodomain 6	BLH6	60.34	38.64	7.66	14.69
	AT4G38000	DNA binding with one finger 4.7	DOF4.7	-	-	-	-
	AT5G39020	Malectin/receptor-like protein kinase family protein		-	-	-	-
	AT5G41410	Homeodomain protein	BEL1	404.86	170.36	182.12	54.72
	AT5G60850	zinc finger protein	OBP4	-	-	-	-
	AT5G66940	Dof-type zinc finger DNA-binding family protein		-	-	-	-

Supplementary Table 9 continued on next page.

Supplementary Table 9 continued.

	Arabidopsis homolog	Protein Family	Protein/Gene Abbreviation	15waa mesocarp	20waa mesocarp	10waa kernel	15waa kernel
Extra Plastidial Phospholipid Synthesis				48.26	56.18	43.08	45.53
	AT3G05420	Acyl CoA Binding Protein	ACBP4	109.04	58.93	101.40	72.36
	AT3G03310	Phospholipid : Acyl acceptor Acyltransferase	PLAT?	17.81	26.99	1.89	9.02
	AT3G07690	NAD-dependent Glycerol-3-Phosphate Dehydrogenase	GPDH	-	-	-	-
	AT1G63050	1-acylglycerol-3-phosphocholine Acyltransferase	LPCAT	-	-	-	-
	AT1G04010	Phospholipid : Acyl acceptor Acyltransferase	PLAT?	26.80	12.56	0.69	1.23
	AT2G45670	1-acylglycerol-3-phosphoethanolamine Acyltransferase	LPEAT	14.02	8.98	12.80	11.33
	AT4G23850	Long-Chain Acyl-CoA Synthetase	LACS4	107.13	105.93	460.61	430.59
	AT1G75020	1-acylglycerol-3-phosphate acyltransferase	LPAAT4	18.50	14.78	47.24	26.32
	AT4G38570	Phosphatidylinositol Synthase	PIS2	5.46	8.46	8.43	3.22
	AT5G27630	Acyl CoA Binding Protein	ACBP5	-	-	-	-
	AT1G74320	Choline Kinase	CK	84.36	185.05	20.19	49.36
	AT1G13560	Diacylglycerol Cholinephosphotransferase	DAG-CPT	38.85	15.13	2.00	2.82
	AT4G22340	CDP-DAG Synthase	CDP-DAGS	25.71	16.44	2.28	6.60
	AT1G48600	Phosphoethanolamine N-Methyltransferase	CPUORF31	-	-	-	-
	AT1G62430	CDP-DAG Synthase	CDS1	-	-	-	-
	AT4G24230	Acyl CoA Binding Protein	ACBP3	143.01	136.25	8.20	226.99
	AT2G41540	NAD-dependent Glycerol-3-Phosphate Dehydrogenase	GPDH	160.74	465.66	265.42	326.48
	AT4G25970	Phosphatidylserine Decarboxylase	PSD3	30.86	20.19	3.30	2.23
	AT3G18000	Phosphoethanolamine N-Methyltransferase	NMT1	107.12	16.41	19.46	6.27
	AT1G73600	Phosphoethanolamine N-Methyltransferase	PEAMT	0.33	0.21	26.72	10.29
	AT1G80860	Phosphatidylethanolamine methyltransferase	PLMT	3.44	1.37	7.88	5.32
	AT3G25585	Diacylglycerol Cholinephosphotransferase	DAG-CPT	-	-	-	-
	AT4G15130	Choline-Phosphate Cytidylyltransferase	CCT2	450.17	215.12	91.17	50.61
	AT1G68000	Phosphatidylinositol Synthase	PIS1	22.33	30.93	34.96	33.79
	AT5G60620	Glycerol-3-Phosphate Acyltransferase (mammalian homologue)	GPAT9	68.50	110.50	111.48	60.50
	AT3G18850	1-acylglycerol-3-phosphate acyltransferase	LPAAT5	-	-	-	-
	AT1G31812	Acyl CoA Binding Protein	ACBP6	144.17	455.56	75.39	36.71
	AT4G11030	Long-Chain Acyl-CoA Synthetase	LACS5	-	1.20	0.37	-
	AT1G12640	1-acylglycerol-3-phosphocholine Acyltransferase	LPCAT	62.68	24.42	112.32	32.22
	AT2G38670	CDP-Ethanoamine Synthase	PECT1	41.24	63.01	38.67	20.66
	AT4G16700	Phosphatidylserine Decarboxylase	PSD1	20.34	20.22	13.54	7.00
	AT3G57650	1-acylglycerol-3-phosphate acyltransferase	LPAAT2	51.54	39.72	80.34	133.68
	AT4G09760	Choline Kinase	CK	3.20	14.67	-	-
	AT3G58490	Phosphatidate Phosphatase	LPP-δ	57.36	186.33	17.61	6.63
	AT4G19860	Phospholipid : Acyl acceptor Acyltransferase	PLAT?	60.17	19.30	55.85	53.05
	AT4G27780	Acyl CoA Binding Protein	ACBP2	18.73	24.07	26.09	23.14
	AT4G22550	Phosphatidate Phosphatase	LPP-β	21.76	14.12	70.64	22.98
	AT5G42870	Phosphatidate Phosphatase (ER)	PAH2	36.46	35.72	35.60	67.16
	AT1G15110	Base-Exchange-type Phosphatidylserine Synthase	BE-PSS	37.18	34.25	17.86	12.72
	AT2G26830	Ethanolamine Kinase	EK	179.57	114.60	17.29	12.30
	AT1G71697	Choline Kinase	CK1	1.54	2.69	2.83	3.51
	AT5G57190	Phosphatidylserine Decarboxylase	PSD2	-	-	-	-
	AT3G15820	Phosphatidylcholine:diacylglycerol cholinephosphotransferase	PDCT/ROD1	20.10	27.31	0.25	0.90
	AT2G32260	Choline-Phosphate Cytidylyltransferase	CCT1	-	-	-	1.15
	AT3G09560	Phosphatidate Phosphatase (ER)	PAH1	16.10	24.12	13.90	12.74
	AT5G53470	Acyl CoA Binding Protein	ACBP1	13.41	33.20	176.86	312.43

Supplementary Table 9 continued on next page.

Supplementary Table 9 continued.

	Arabidopsis homolog	Protein Family	Protein/Gene Abbreviation	15waa mesocarp	20waa mesocarp	10waa kernel	15waa kernel
TAG Synthesis				37.93	76.56	56.92	63.07
	AT3G58490	Phosphatidate Phosphatase	LPP-δ	57.36	186.33	17.61	6.63
	AT3G07690	NAD-dependent Glycerol-3-Phosphate Dehydrogenase	GPDH	-	-	-	-
	AT4G22550	Phosphatidate Phosphatase	LPP-β	21.76	14.12	70.64	22.98
	AT3G51520	Acyl-CoA : Diacylglycerol Acyltransferase	DGAT2	9.47	10.92	8.64	12.25
	AT5G42870	Phosphatidate Phosphatase (ER)	PAH2	36.46	35.72	35.60	67.16
	AT3G18850	1-Acylglycerol-3-Phosphate Acyltransferase	LPAAT5	-	-	-	-
	AT5G60620	Glycerol-3-Phosphate Acyltransferase (mammalian homologue)	GPAT9	68.50	110.50	111.48	60.50
	AT5G13640	Phospholipid : Diacylglycerol Acyltransferase	PDAT1	21.23	28.34	31.31	105.48
	AT2G19450	Acyl-CoA : Diacylglycerol Acyltransferase	DGAT1	31.45	65.10	57.72	45.75
	AT1G75020	1-Acylglycerol-3-Phosphate Acyltransferase	LPAAT4	18.50	14.78	47.24	26.32
	AT3G09560	Phosphatidate Phosphatase (ER)	PAH1	16.10	24.12	13.90	12.74
	AT2G41540	NAD-dependent Glycerol-3-Phosphate Dehydrogenase (cytosolic)	GPDH	160.74	465.66	265.42	326.48
	AT3G57650	1-Acylglycerol-3-Phosphate Acyltransferase	LPAAT2	51.54	39.72	80.34	133.68

Supplementary Table 9 continued on next page.

Supplementary Table 9 continued.

	Arabidopsis homolog	Protein Family	Protein/Gene Abbreviation	15waa mesocarp	20waa mesocarp	10waa kernel	15waa kernel
Hexoses-P				92.70	61.97	162.84	62.46
	AT1G06020	Fructokinase	Probable FK-3	4.47	6.02	549.14	177.39
	AT1G06030	Fructokinase	Probable FK-2	34.38	38.07	1,213.50	267.61
	AT1G23190	Phosphoglucomutase	PGM	241.26	182.71	434.11	274.70
	AT1G47840	Hexokinase	Plastid-HXK3	11.88	3.81	1.12	3.91
	AT1G66430	Fructokinase	Plastid-pfkB-type	50.66	18.97	27.65	24.40
	AT1G70730	Phosphoglucomutase	PGM	78.12	29.46	108.42	42.50
	AT1G70820	Phosphoglucomutase	PGM2	-	-	-	-
	AT2G19860	Hexokinase	HXK2	53.11	30.07	10.66	24.57
	AT2G31390	Fructokinase	Probable FK-1	376.69	277.69	10.90	8.97
	AT3G03250	UDP-glucose pyrophosphorylase	UGP1	0.64	-	6.58	2.58
	AT3G56040	UDP-glucose pyrophosphorylase	Plastid-UGP3	6.56	32.38	0.30	0.60
	AT3G59480	Fructokinase	Probable FK-4	-	-	-	-
	AT4G10260	Fructokinase	Putative FK-5	183.34	81.78	38.33	29.71
	AT4G24620	Phosphoglucoisomerase	Plastid-PGI	51.65	43.82	31.07	11.65
	AT4G29130	Hexokinase	HXK1	384.58	256.28	31.11	46.19
	AT5G17310	UDP-glucose pyrophosphorylase	UGP2	178.68	118.59	395.54	189.27
	AT5G42740	Phosphoglucoisomerase	Cytosol-PGI	89.11	46.57	232.11	72.31
	AT5G51820	Phosphoglucomutase	PGM1	15.52	11.16	3.04	10.18
	AT5G51830	Fructokinase	pfkB-type	0.59	-	0.34	0.24
Oxidative Pentose Phosphate Pathway				130.32	114.83	34.15	33.17
	AT1G09420	Glucose-6-Phosphate dehydrogenase	G6PDH4	34.40	29.51	3.15	1.89
	AT1G12230	Transaldolase	TAL	968.04	999.76	119.14	29.61
	AT1G13700	Lactonase	PGL1	9.12	14.01	18.75	13.34
	AT1G24280	Glucose-6-Phosphate dehydrogenase	G6PDH3	17.11	12.61	1.54	-
	AT1G63290	Epimerase	EPI	-	-	-	-
	AT1G64190	6-phosphogluconate dehydrogenase	6PGDH	-	-	-	-
	AT1G71100	Isomerase	ISO	22.12	67.55	3.87	2.32
	AT2G01290	Isomerase	ISO	205.54	337.54	4.26	4.15
	AT2G45290	Transketolase	TKT	45.51	37.46	11.11	7.00
	AT3G01850	Epimerase	EPI	56.60	42.93	9.24	7.39
	AT3G02360	6-phosphogluconate dehydrogenase	Cytosol-6PGDH	523.14	147.45	70.50	86.73
	AT3G04790	Isomerase	ISO	45.28	31.32	17.24	14.96
	AT3G27300	Glucose-6-Phosphate dehydrogenase	G6PDH5	14.94	11.79	1.45	17.76
	AT3G60750	Transketolase	TKT	283.88	137.38	102.04	150.59
	AT5G13110	Glucose-6-Phosphate dehydrogenase	G6PDH2	48.36	23.10	17.91	10.69
	AT5G13420	Transaldolase	TAL	427.00	453.68	298.54	250.14
	AT5G24400	Lactonase	Plastid-PGL3	176.22	245.97	77.15	56.41
	AT5G24410	Lactonase	PGL4	-	-	-	-
	AT5G24420	Lactonase	PGL5	-	-	-	-
	AT5G35790	Glucose-6-Phosphate dehydrogenase	G6PDH1	23.64	4.37	1.00	0.52
	AT5G40760	Glucose-6-Phosphate dehydrogenase	G6PDH6	118.23	48.92	27.02	97.90
	AT5G41670	6-phosphogluconate dehydrogenase	6PGDH	84.81	76.10	28.10	41.46
	AT5G44520	Isomerase	ISO	0.48	1.02	-	-
	AT5G61410	Epimerase	Plastid-EPI	23.31	33.48	7.70	3.32

For each functional classification group of genes, the average transcripts per million tags sequenced is presented in the header for each group. For individual genes, expression levels are presented as transcripts per million tags sequenced. Data is presented for transcriptomes from mesocarp (15 and 20 WAA) and kernel (10 and 15 WAA).

Supplementary Methods

Repetitive sequence detection

To construct *Elaeis* specific repeat databases, repetitive sequences of *pisifera* and *oleifera* were determined by a genome wide self-self comparison based on 60mer windowed 15mer word frequencies. Each genome was decomposed into 60mer words with 40 base overlap, and scored for uniqueness using a SSAHA[31] based pipeline. The uniqueness score for an individual 60mer was calculated as:
60mer Uniqueness Score = Sum (i=1 to 45) [log(freq($15mer_i$))]. Perfectly unique 60mers receiving a score of 0, and highly repetitive 60mers scoring as high as 727 in the *pisifera* genome were calculated. Regions of each genome with uniqueness scores in the highest (most repetitive) 20% were collected and clustered using the PCAP algorithm to create the PisiferaRepeat and OleiferaRepeat databases with 5,426 and 440 sequences, respectively.

Public database sources for comparative genome sequences and gene models

Genome source:
A. thaliana:
ftp://ftp.arabidopsis.org/home/tair/Sequences/whole_chromosomes/TAIR10
O. sativa:
ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v8.0/Osativa/assembly/Osativa_193.fa.gz
P. dactylifera:
http://qatar-weill.cornell.edu/research/datepalmGenome/edition3/PdactyKAsm30_r20101206.fasta.gz

Gene model source:
A. thaliana:
ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/TAIR10_blastsets/TAIR10_pep_20110103_representative_gene_model_updated

O. sativa:

ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v8.0/Osativa/annotation/Osativa_193_peptide.fa.gz

P. dactylifera:

http://qatar-weill.cornell.edu/research/datepalmGenome/edition3/PDK30-pep.fsa.gz

Supplementary Notes

Triacylglycerol biosynthesis

TAG usually accumulates in oil bodies. Seeds (including oil palm kernel) maintain oil bodies as small individual units (0.5 to 2 µm) preventing their coalescence during seed desiccation and providing high surface area to volume ratio to facilitate access by lipases during germination. In contrast, the oleaginous fruits, oil palm, olive and avocado have large oil bodies (up to 20 µm) in the mesocarp owing to the lack of oleosins, structural proteins that stabilize and prevent the coalescence of the oil bodies. Unlike the date genome, oleosins are represented in the oil palm. Consistently, the transcriptome profile showed very high oleosin expression in the kernel and very low expression in the mesocarp (Figure 3c, Supplementary Table 9).

Fruit ripening and abscission

Interestingly the oil palm has two-fold more ripening and abscission genes than date palm (Fig. 2c). The oil palm has more *AMINOCYCLOPROPANE CARBOXYLATE OXIDASE* (*ACO*) genes responsible for the burst of ethylene and increased respiration on ripening. The oil palm genome is also enriched in *MAPKK* genes which are involved in ethylene signal transduction. *BRASSINOSTEROID INSENSITIVE1* Leucine-Rich Repeat (LRR) genes and *BLADE ON PETIOLE2* (*BOP2*) genes are similarly more highly represented in the oil palm genome. *BRI1* regulates transduction of steroid signals across membranes while *BOP2* controls abscission zone formation.

References

44. Luo, M. & Wing, R. A. An improved method for plant BAC library construction. *Methods Mol Biol* **236**, 3-20 (2003).
45. Ammiraju, J. S. et al. The Oryza bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus Oryza. *Genome Res* **16**, 140-7 (2006).
46. Jurka, J. et al. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**, 462-7 (2005).
47. Piegu, B. et al. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in Oryza australiensis, a wild relative of rice. *Genome Res* **16**, 1262-9 (2006).
48. Rival, A. et al. Comparative flow cytometric estimation of nuclear DNA content in oil palm (Elaeis guineensis, Jacq.) tissue cultures and seed derived plants. . *Plant Cell Reports* **16**, 884-887 (1997).
49. Madon, M., Phoon, L. Q., Clyde, M. M. & Mohd Din, A. Application of flow cytometry for estimation of nuclear DNA content in Elaeis. *Journal of Oil Palm Research* **20**, 447-452 (2008).
50. Rajanaidu, N., Rao, V., Abdul Halim, H. & A.S.H., O. Genetic resources: New developments in Oil Palm breeding. *Elaeis* **1**, 1-10 (1989).
51. Billotte, N. et al. Microsatellite-based high density linkage map in oil palm (Elaeis guineensis Jacq.). *Theor Appl Genet* **110**, 754-65 (2005).
52. Madon, M., Clyde, M. M. & Cheah, S. C. Cytological analysis of Elaeis guineensis (tenera) chromosomes. . *Elaeis* **7**, 122-134 (1995).