Metabolic Routes Affecting Rubber Biosynthesis in *Hevea brasiliensis* Latex

Keng-See Chow, Mohd.-Noor Mat- Isa, Azlina Bahari, Ahmad-Kamal Ghazali, Halimah Alias, Zainorlina Mohd.-Zainuddin, Chee-Choong Hoh, Kiew-Lian Wan

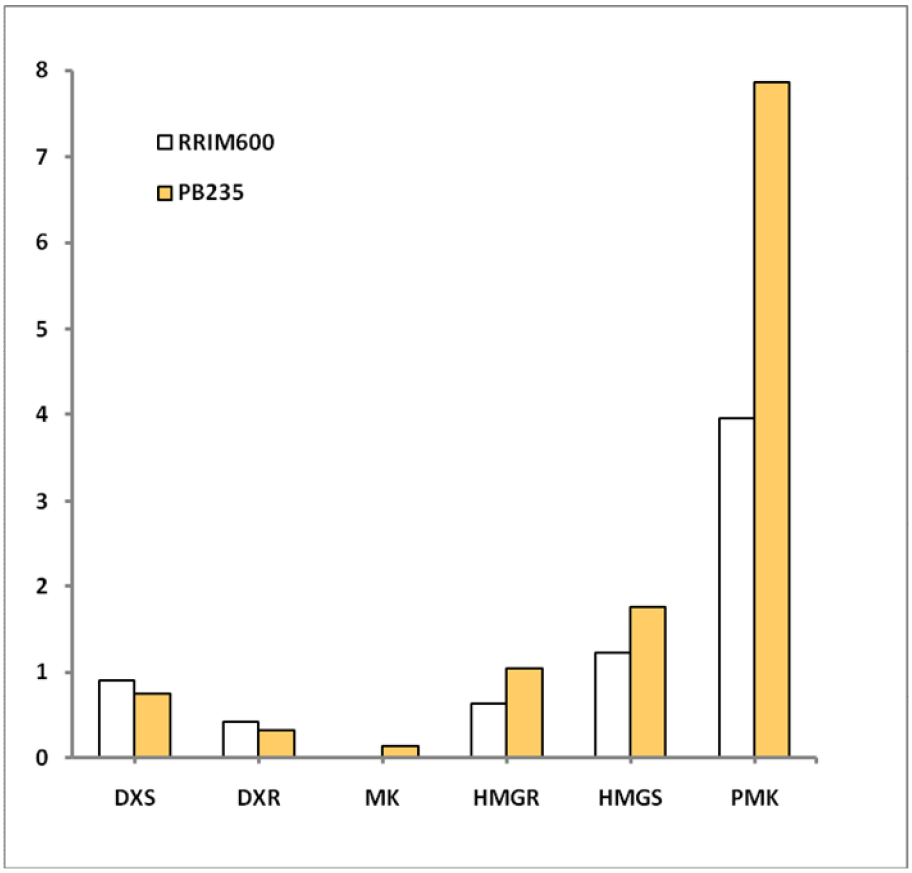
SUPPLEMENTARY DATA

Gene abbreviations

REF: rubber elongation factor; SRPP: small rubber particle protein; RBSP: rubber biosynthesis stimulator protein; RBIP: rubber biosynthesis inhibitor protein; FPPS: farnesyl diphosphate synthase; GGPPS: geranylgeranyl diphosphate synthase; IPPI: isopentenyl diphosphate isomerase; CPT: cis-prenyltransferase; HMGS: hydroxymethyglutaryl coenzyme A synthase; HMGR: hydroxymethyglutaryl coenzyme A reductase; MK: mevalonate kinase; PMK: phosphomevalonate kinase; DXS: 1-deoxy-D-xylulose 5-phosphate synthase; DXR: 1-deoxy-D-xylulose 5-phosphate reductoisomerase; 18S rRNA: 18S ribosomal RNA

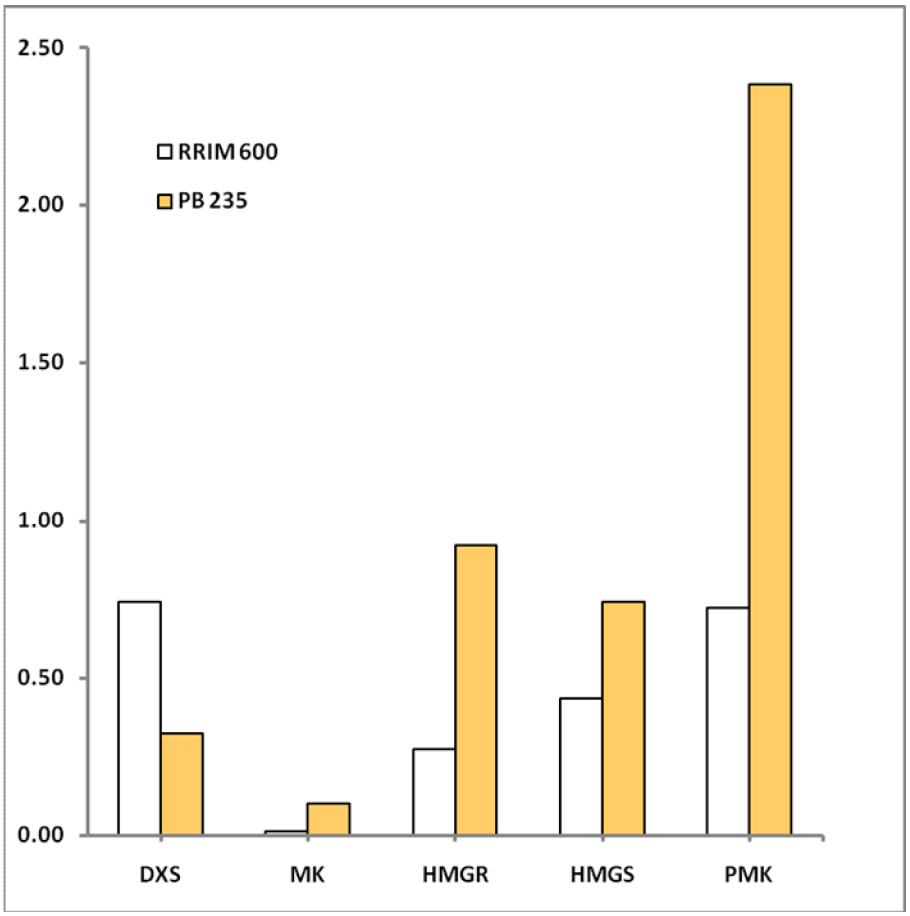
SUPPLEMENTARY FIGURE S1. qRT-PCR analysis of MVA and MEP pathway-specific gene transcript levels in RRIM 600 and PB 235. Trend of gene expression based on transcript level (Pfaffl ratio) was consistent in two biological replicates of the same experiment (**a** and **b**). DXR was not included in one of the replicates (**b**) because of insufficient template RNA. The stable expression of the 18S rRNA gene in clones RRIM 600 and PB 235 was assessed by Studentôs t-test (P=0.79). (CP=cycle threshold, Av: average; SD: standard deviation)

a



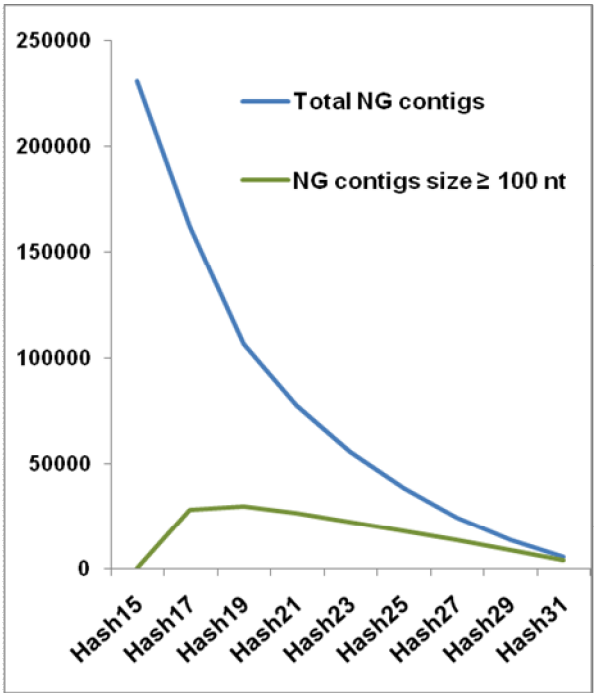
	RRIM 600		PB 235		P value (Av Pfaffl ratio RRIM 600 vs PB 235)
	Av CP±SD	Av Pfaffl ratio±SD	Av CP±SD	Av Pfaffl ratio±SD	
DXS	16.53±0.07	0.91±0.05	16.37±0.14	0.75±0.08	0.0376
DXR	17.53±0.49	0.42±0.13	18.23±0.59	0.32±0.11	0.3550
HMGR	17.21±0.46	0.63±0.21	16.04±0.52	1.05±0.34	0.1418
MK	24.47±0.08	0.005±0.001	19.14±0.31	0.14±0.03	0.0017
HMGS	16.46±0.59	1.22±0.53	15.84±0.07	1.76±0.09	0.1546
PMK	14.79±0.50	3.96±1.19	13.76±0.39	7.86±2.04	0.0461

b



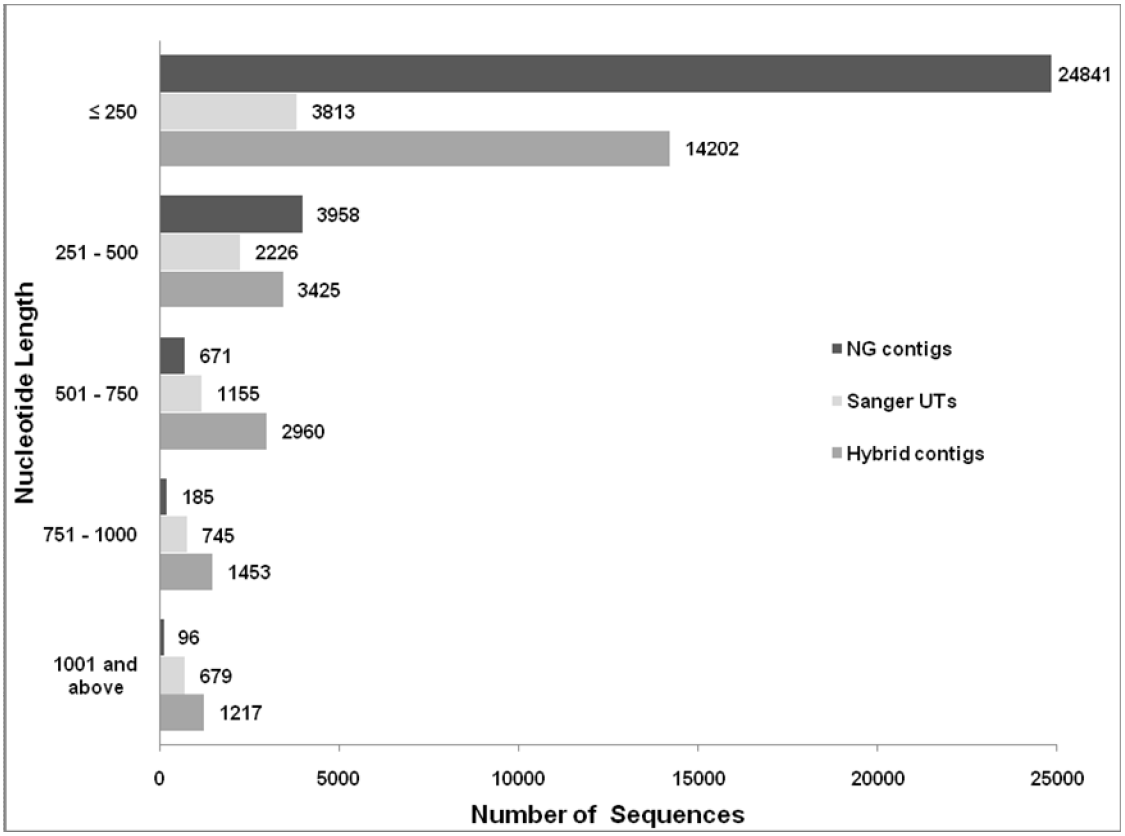
	RRIM 600		PB 235		P value (Av Pfaffl ratio RRIM 600 vs PB 235)
	Av CP±SD	Av Pfaffl ratio±SD	Av CP±SD	Av Pfaffl ratio±SD	
DXS	16.43±0.36	0.74±0.18	17.65±0.28	0.32±0.06	0.0201
HMGR	18.23±0.32	0.27±0.06	16.64±0.11	0.92±0.07	0.0003
MK	23.07±0.18	0.01±0.00	20.02±0.15	0.10±0.01	0.0001
HMGS	17.77±0.32	0.43±0.10	17.29±0.36	0.74±0.18	0.0831
PMK	17.11±0.07	0.72±0.04	15.65±0.11	2.38±0.18	0.0058

SUPPLEMENTARY FIGURE S2. NG contig assembly and validation. Illumina mRNAseq reads (35 nt) were first assembled using Velvet 0.7.63 with a hash value range of 15-31. Due to an inverse relationship between the hash values and total NG contigs, an arbitrary size limit of Ó100 nt was introduced for total NG contigs generated by each hash value. This resulted in a distribution of contigs where the maximum number (29,751 NG contigs) coincided with hash 19. As further validation, mapping of NG contigs (Ó100 nt) to 8,618 Sanger UTs also showed a distribution where the maximum number of hits (8,040) also peaked at hash 19. Therefore, this justified 29,751 NG contigs to be the optimal assembly of the Illumina mRNAseq reads.



	Total NG contigs	NG contigs (≥100 nt)	Hits by NG contigs (≥100 nt) to 8,618 Sanger UTs
Hash15	231,008	241	47
Hash17	162,103	28,118	7,215
Hash19	***106,669***	***29,751***	***8,040***
Hash21	77,462	26,278	7,468
Hash23	55,539	22,230	6,930
Hash25	38,268	17,971	6,251
Hash27	24,137	13,438	5,211
Hash29	13,413	8,568	3,882
Hash31	5,487	4,144	2,382

SUPPLEMENTARY FIGURE S3. Contig size distribution across three assembly data sets: Sanger UTs, NG contigs and hybrid contigs. Based on the size distribution of Sanger UTs, NG contigs and hybrid contigs, there was an increase in assembly length in all size categories where the number of hybrid contigs was more than at least one of the initial data sets (Sanger UTs and NG contigs). Therefore, co-assembly resulted in overall contig size improvement in the hybrid contig population.



SUPPLEMENTARY TABLE S1. qRT-PCR analysis of transcript levels of rubber biosynthesis pathway genes in ethylene-treated trees. (**a**) Gene-specific primer pairs for qRT-PCR were designed based on publicly available *Hevea* cDNAs encoding candidate pathway genes (accession numbers indicated). Genes for post-IPP steps are italicised. (**b**) qRT-PCR analysis of latex samples collected on day 1, 3 and 8 after ethylene treatment. Gene transcript level was expressed as the Pfaffl ratio. DXR transcript level was not available in this data set because of unacceptable qRT-PCR results. The stable expression of the 18S rRNA as a housekeeping gene was assessed by Studentôs t-test which showed that the CP difference between latex sampled from different days (P=0.43) and between ethylene treated and control untreated trees (P=0.11) was not significant. (**c**) To account for day-to-day variation, comparison was made between transcript level of each gene in ethylene treated and control latex samples collected from the same day. In this comparison, a positive fold change indicated induction of transcription by ethylene while a negative fold change indicated repression. Four trends of fold change over three days are shown in Fig. 2. (CP=cycle threshold, Av: average; SD: standard deviation)

a

Gene (Accession no.)	Amplicon (bp)	Forward (FW) and Reverse (RV) primer sequences	Primer pair efficiency	Annealing temperature (°C)
18S rRNA (AY496880)	110	FW 5ôCCATAAACGATGCCGACCAG3ô	1.98	56.1
		RV 5ôCAGCCTTGCGACCATACTC3ô		55.1
GGPPS (AB055496)	156	FW 5ôAAGGCAATTCCATTAACCAAGC3ô	2.00	55.2
		RV 5ôAGTCATTCCCACCAACAAGC3ô		55.3
SRPP (AF051317)	104	FW 5ôGCTGGAGTTTATGCTGTAGATTC3ô	2.08	55.0
		RV 5ôTTCACCACATTCTCAATAGTATCG3ô		54.5
RBIP (AF113546)	179	FW 5ôTGATGCGAAATGTAATGCTCTG3ô	2.00	55.0
		RV 5ôTTAGTGCTAGTAGTGTAGGATTGG3ô		55.2
MK (AF429384)	132	FW 5ôATCAAGTTCAAGTCTGGTAATCTG3ô	2.00	54.6
		RV 5ôGTGCCGTAAGGTTCTCTCTG3ô		55.0
PMK (AF429385)	138	FW 5ôGCTGCTTTACTTCATCACCTTG3ô	1.97	55.4
		RV 5ôAAATCCACTGCCGACTTTCC3ô		55.4
RBSP (AF516353)	132	FW 5ôGCTGATGGTGAAACAGTCTTTATG3ô	1.96	56.3
		RV 5ôCCTCCCTCATTACAAGCCTTATC3ô		56.0
HMGS (AY534617)	146	FW 5ôAGGTAAGCGGGTGATACTATTC3ô	1.98	54.6
		RV 5ôCGTGTCTTGTCTTCAACTTTCC3ô		55.3
IPPI (AF111842)	192	FW 5ôAATGTAAATCCAAACCCTGATGAG3ô	1.98	55.1
		RV 5ôAACTGCTTCCTTGAGTGTCC3ô		54.7
HMGR (X54659)	105	FW 5ôGCTGTTATATGAAGTATGGAGATAGG3ô	2.02	55.1
		RV 5ôAAGGGTAGAGAGAGAAGTAGAGG3ô		55.2
REF (X56535)	173	FW 5ôCAACTTATGCTGTGACTACCTTC3ô	2.00	54.9
		RV 5ôCGTGCTGTCTACAAACTTGAG3ô		54.5
FPPS (Z49786)	169	FW 5ôACACCAGATTCCCGTCAATG3ô	1.99	54.9
		RV 5ôATTCAATACACCAACCAAGAGC3ô		54.4
CPT (AB061234)	126	FW 5ôTCAAGACCGCAGCAGATAAG3ô	2.03	54.7
		RV 5ôTCAATTCAGAGGATTCTTCAACAG3ô		54.8
DXS (EC609693)	130	FW 5ôAGTCTGCGTCCGTTATCCAAGG3ô	2.03	54.2
		RV 5ôCCCATACCCAAGCAAAGCAACATC3ô		54.1
DXR (DQ437514)	78	FW 5ôATTGGACATCGTGGCAGAG3ô	2.00	54.2
		RV 5ôGCAAGAAGAGTAACATTTGAACC3ô		54.1

b

Day 1 (control)			Day 1 (ethylene treated)		
Gene	**Av CP ± SD**	**Av Pfaffl ratio ± SD**	**Gene**	**Av CP ± SD**	**Av Pfaffl ratio ± SD**
SRPP	5.46± 0.28	2848.86± 553.47	SRPP	4.98± 0.31	1778.57± 373.28
REF	8.40± 0.10	237.17± 15.64	REF	6.50± 0.12	690.39± 59.95
PMK	12.80± 0.57	20.35± 7.71	RBIP	11.66± 0.12	19.27± 1.66
CPT	13.83± 0.18	10.41±1.28	CPT	11.72± 0.18	15.60± 2.08
RBSP	14.11± 0.24	5.21±0.79	FPPS	12.98± 0.17	7.72± 0.90
DXS	14.84±0.49	2.16± 0.73	PMK	13.27± 0.22	7.08± 1.01
FPPS	16.26± 0.34	1.67± 0.38	RBSP	13.92± 0.06	5.77± 0.24
HMGS	17.03± 0.35	1.51± 0.34	GGPPS	13.48± 0.65	5.65± 2.74
HMGR	17.36± 0.04	0.69± 0.02	DXS	14.30± 0.19	3.01± 0.42
GGPPS	18.89± 0.46	0.37± 0.12	HMGS	14.72± 0.18	2.52± 0.31
RBIP	18.45± 0.44	0.23± 0.07	HMGR	15.73± 0.20	0.99± 0.14
IPPI	19.14± 0.20	0.16± 0.02	IPPI	16.18± 0.05	0.96± 0.03
MK	20.07± 0.28	0.12± 0.02	MK	17.06± 0.20	0.45± 0.06

Day 3 (control)			Day 3 (ethylene treated)		
Gene	**Av CP ± SD**	**Av Pfaffl ratio ± SD**	**Gene**	**Av CP ± SD**	**Av Pfaffl ratio ± SD**
SRPP	7.66± 0.81	438.44± 207.60	SRPP	5.87± 0.22	2583.47± 420.65
REF	8.46± 0.22	395.70± 56.80	REF	7.27± 0.43	774.84± 208.48
CPT	12.21± 0.12	22.36± 1.92	DXS	14.03± 0.40	10.63± 3.20
PMK	13.42± 0.27	10.21± 1.79	PMK	14.51± 0.08	9.86±0.57
DXS	13.53± 0.27	9.25± 1.72	RBIP	13.92± 0.40	7.71± 2.08
FPPS	14.70± 0.05	3.79± 0.13	CPT	14.13± 0.36	7.06± 1.75
RBIP	15.60± 0.13	2.79± 0.25	FPPS	15.42± 0.21	4.78± 0.66
RBSP	16.21± 0.50	2.22± 0.79	HMGS	15.92± 0.10	2.64± 0.19
HMGS	16.37± 0.19	1.61± 0.22	RBSP	17.16± 0.49	1.82± 0.56
GGPPS	16.71± 0.43	1.15± 0.36	IPPI	16.53± 0.22	1.42± 0.23
HMGR	17.43± 0.24	0.67± 0.11	GGPPS	17.03± 0.18	1.08± 0.13
IPPI	19.62± 0.03	0.20± 0.00	HMGR	17.59± 0.34	0.73± 0.19
MK	21.53± 0.30	0.04± 0.01	MK	19.75± 0.61	0.20± 0.09

Day 8 (control)			Day 8 (ethylene treated)		
Gene	**Av CP ± SD**	**Av Pfaffl ratio ± SD**	**Gene**	**Av CP ± SD**	**Av Pfaffl ratio ± SD**
SRPP	5.42± 0.39	8295.24± 2196.73	SRPP	5.64± 0.16	5502.96± 625.60
REF	8.18± 0.32	1064.62± 251.72	REF	5.82± 0.14	3278.35± 306.67

PMK	11.97± 0.64	63.62± 26.62	RBIP	11.63± 0.38	59.43± 15.06
CPT	12.84± 0.14	48.76± 5.01	PMK	13.54± 0.51	30.05± 11.01
FPPS	12.57± 0.50	38.68± 11.91	CPT	12.47± 0.01	29.40± 0.24
DXS	14.19± 0.69	19.83± 9.82	FPPS	13.73± 0.10	23.16± 1.58
RBIP	14.09± 0.30	17.64± 3.54	DXS	15.21± 0.74	13.94± 4.33
RBSP	15.63± 0.29	9.61± 1.91	HMGS	15.85± 0.74	3.83± 1.85
HMGS	16.45± 0.07	4.98± 0.25	IPPI	16.00± 0.24	3.21± 0.52
HMGR	16.18± 0.16	4.26±0.45	RBSP	17.59± 0.62	3.11± 1.23
GGPPS	16.98± 0.44	3.19± 1.03	GGPPS	16.38± 0.37	2.27± 0.59
IPPI	18.23± 0.24	1.14± 0.19	HMGR	17.78± 0.41	1.89± 0.54
MK	19.12± 0.19	0.62± 0.08	MK	22.55± 0.44	0.08± 0.03

c

Trend	Gene	Fold change in gene expression based on Pfaffl ratio of ethylene treated vs. control latex samples (P value)		
		Day 1	Day 3	Day 8
I	GGPPS	15.3 (0.0290)	-0.9 (0.7673)	-0.7 (0.2506)
	CPT	1.5 (0.0212)	-0.3 (0.0005)	-0.6 (0.0026)
	RBSP	1.1 (0.3052)	-0.8 (0.5139)	-0.3 (0.0077)
II	HMGR	1.4 (0.0213)	1.1 (0.7212)	-0.4 (0.0125)
	MK	3.8 (0.0008)	5.0 (0.0376)	-0.1 (0.0031)
	FPPS	4.6 (0.0128)	1.3 (0.1402)	-0.6 (0.1801)
	DXS	1.4 (0.1851)	1.1 (0.5465)	-0.7 (0.3956)
	HMGS	1.7 (0.0191)	1.6 (0.0036)	-0.8 (0.3305)
III	SRPP	-0.6 (0.0500)	5.9 (0.0014)	-0.7 (0.1932)
	PMK	-0.3 (0.0489)	1.0 (0.8146)	-0.5 (0.1308)
IV	RBIP	83.8 (0.0006)	2.8 (0.0153)	3.4 (0.0095)
	IPPI	6.0 (0.0001)	7.1 (0.0057)	2.8 (0.0029)
	REF	2.9 (0.0002)	2.0 (0.0384)	3.1 (0.0006)

SUPPLEMENTARY TABLE S2. Interrogation of Sanger UTs, NG contigs and hybrid contigs with 195 annotated rubber cDNAs to obtain a measure of complete protein coding regions. Each cross (+) indicates a match with a rubber cDNA. Our results showed that 80.5% of 195 complete rubber protein coding regions matched with the hybrid contigs, at the same time having a hit to at least one of the other two assembly data sets. This suggested that a large percentage of complete coding regions exist within the 23,257 hybrid contigs and thus provided evidence for biologically correct transcripts.

Number of rubber cDNAs with hits	Assembly data set			Total (%)
	NG contigs	Sanger UTs	Hybrid contigs	
37	-	-	-	37 (19.0)
24	-	+	+	
12	+	-	+	157 (80.5)
121	+	+	+	
1	+	+	-	1 (0.5)
				195 (100.0)

SUPPLEMENTARY TABLE S3. List of 195 rubber cDNAs with complete protein coding regions, functional description and tissue of origin. *(Provided as separate file)* Unlike high-throughput sequences which are analysed by bulk statistics, the 195 rubber cDNAs in the public databases represent more reliable protein coding regions as they had been curated manually. The reason why some rubber cDNAs did not match with any hybrid contig could be because they are sequence isoforms from non-latex tissues.

SUPPLEMENTARY TABLE S4. Hybrid contig sequences in fasta format. *Provided as separate file.*

SUPPLEMENTARY TABLE S5. BLAST2GO annotations of hybrid contigs. *(Provided as separate file.)* Putative gene functions and pathways of hybrid contigs are based on BLAST2GO and KEGG Pathway.