Supporting information 1

Table S1 BLAST output filtering with query coverage and cumulative identity. Qcovs (query coverage); ID (cumulative identity); NR plant, NCBI non-redundant plant protein database; NT, NCBI non-redundant nucleotide database; #sequences, number of sequences.

Table S2 Coffee long read sequencing isoforms annotated by Rfam (rfam.xfam.org).

Figure S1 Species distribution of coffee long read sequencing isoforms according to the result of BLAST against NCBI non-redundant database.

Figure S2 InterProScan families distribution of coffee long read sequencing isoforms.

Figure S3 Distribution of InterProScan families from coffee long read sequencing dataset.

Figure S4 Pie chart and word cloud of coffee long read sequencing isoforms distribution to biological process, cellular component and molecular function

Figure S5 The KEGG pathway distribution of coffee long read sequencing isoforms.

Figure S6 Putative transcript variants from long-read sequencing aligned to reported genes encoding 7-methylxanthine N-methyltransferase. G-CaMXMT1, Arabica MXMT1 genomic DNA sequence; G-CaMXMT2, Arabica MXMT2 genomic DNA sequence; MXMT1, MXMT2, MXMT1, MXMT2 coding sequence; Ca, Arabica coffee; Cc, Robusta coffee; c20397/f5p1/1361, c10402/f2p3/1277, coffee long read sequencing isoforms.

Figure S7 Pie chart and word cloud of long sequences (coffee long read sequencing isoforms >10kb) distribution to biological process, cellular component and molecular function.

BLASTx output (NR plant)									
High	300-1,000 bp	1,000-3,000 bp	3,000-5,000 bp	>5,000 bp					
Qcovs ($\%$, \geq)	60	50	40	30					
ID (%, ≥)	70	70	60	60					
#sequences	3.								
Medium	300-1,000 bp	1,000-3,000 bp	3,000-5,000 bp	>5,000 bp					
Qcovs (%, \geq)	50	40	20	10					
ID (%, ≥)	60	60	50	50					
#sequences	13,655								
Low	300-1,000 bp	1,000-3,000 bp	3,000-5,000 bp	>5,000 bp					
Qcovs ($\%$, \geq)	9	4	2	1					
ID (%, ≥)	22.69	22.71	25.50	24.01					
HSP length range (bp)	88-961	92-2,844	88-4,938	97-14,424					
	40,341								
#sequences with hit				40,341					
#sequences with hit #sequences without hit to NR plant				<u>40,341</u> 7,280					
#sequences with hit #sequences without hit to NR plant BLASTn out	put (NT) with 7	,280 sequences		<u>40,341</u> 7,280					
#sequences with hit #sequences without hit to NR plant BLASTn out	put (NT) with 7 300-1,000 bp	,280 sequences 1,000-3,000 bp	3,000-5,000 bp	40,341 7,280					
#sequences with hit #sequences without hit to NR plant BLASTn out Qcovs (%, ≥)	put (NT) with 7 300-1,000 bp 5	,280 sequences 1,000-3,000 bp 2	3,000-5,000 bp 1	40,341 7,280 >5,000 bp 1					
#sequences with hit #sequences without hit to NR plant BLASTn out Qcovs ($\%$, \geq) ID ($\%$, \geq)	put (NT) with 7 300-1,000 bp 5 73.32	,280 sequences 1,000-3,000 bp 2 71.83	3,000-5,000 bp 1 75.51	40,341 7,280 >5,000 bp 1 76.09					
#sequences with hit#sequences without hit to NR plantBLASTn outQcovs ($\%$, \geq)ID ($\%$, \geq)HSP length range (bp)	put (NT) with 7 300-1,000 bp 5 73.32 41-962	,280 sequences 1,000-3,000 bp 2 71.83 37-1559	3,000-5,000 bp 1 75.51 32-4219	40,341 7,280 >5,000 bp 1 76.09 50-5578					
#sequences with hit#sequences without hit to NR plantBLASTn outQcovs ($\%$, \geq)ID ($\%$, \geq)HSP length range (bp)#sequences with hit	put (NT) with 7 300-1,000 bp 5 73.32 41-962	,280 sequences 1,000-3,000 bp 2 71.83 37-1559	3,000-5,000 bp 1 75.51 32-4219	40,341 7,280 >5,000 bp 1 76.09 50-5578 1,981					
#sequences with hit #sequences without hit to NR plant BLASTn out Qcovs (%, ≥) ID (%, ≥) HSP length range (bp) #sequences with hit #sequences without hit to NR plant and NT	put (NT) with 7 300-1,000 bp 5 73.32 41-962	,280 sequences 1,000-3,000 bp 2 71.83 37-1559	3,000-5,000 bp 1 75.51 32-4219	40,341 7,280 >5,000 bp 1 76.09 50-5578 1,981 5,299					
#sequences with hit #sequences without hit to NR plant BLASTn out Qcovs (%, ≥) ID (%, ≥) HSP length range (bp) #sequences with hit #sequences without hit to NR plant and NT BLASTn output (C.canephora CDS v)	put (NT) with 7 300-1,000 bp 5 73.32 41-962 with UTR and C	,280 sequences 1,000-3,000 bp 2 71.83 37-1559 .arabica contigs)	3,000-5,000 bp 1 75.51 32-4219 with 5,299 seque	40,341 7,280 >5,000 bp 1 76.09 50-5578 1,981 5,299 nces					
#sequences with hit #sequences without hit to NR plant BLASTn out Qcovs (%, ≥) ID (%, ≥) HSP length range (bp) #sequences with hit #sequences without hit to NR plant and NT BLASTn output (C.canephora CDS v) #sequences without hit	put (NT) with 7 300-1,000 bp 5 73.32 41-962 vith UTR and C	,280 sequences 1,000-3,000 bp 2 71.83 37-1559 .arabica contigs)	3,000-5,000 bp 1 75.51 32-4219 with 5,299 seque	40,341 7,280 >5,000 bp 1 76.09 50-5578 1,981 5,299 nces 1,217					
#sequences with hit#sequences without hit to NR plantBLASTn outQcovs ($\%$, \geq)ID ($\%$, \geq)HSP length range (bp)#sequences with hit#sequences without hit to NR plant and NTBLASTn output (C.canephora CDS v#sequences without hitPutative novel gen	put (NT) with 7 300-1,000 bp 5 73.32 41-962 with UTR and C	,280 sequences 1,000-3,000 bp 2 71.83 37-1559 .arabica contigs) ter filter with Rfa	3,000-5,000 bp 1 75.51 32-4219 with 5,299 seque m)	40,341 7,280 >5,000 bp 1 76.09 50-5578 1,981 5,299 nces 1,217					

Table S1 BLAST output filtering with query coverage and cumulative identity. Qcovs (query coverage); ID (cumulative identity); NR plant, NCBI non-redundant plant protein database; NT, NCBI non-redundant nucleotide database; #sequences, number of sequences.

	Sequence Name	Family Acc	Family ID	Score	E-Value	GC%	Start	End	Strand	Biotype	GOs
	a63645/f1p0/1000 PI	PE00075	mir 166	92.7	8.50E-24	0.42	298	452	+	Gene; miRNA	GO:0035068
	c03043/11p0/1090	KI'00075	1111-100								GO:0035195
	026778/f1p0/867	DE00267	anoD61	60.7	1 50E 14	0.26	115	520	+	Gene; snRNA;	GO:0005730
	c30778/11p0/807	KI'00207	5110IX04	09.7	1.JUL-14	0.30	445	559		snoRNA; CD-box	GO:0006396
	~42517/f1=0/1215	DE00260	amo 7107 D97	97.0	0.500.20	0.40	276	201		Gene; snRNA;	GO:0005730
C4251//11p0/1215	KF00300	SHOZIU/_K8/	87.9	9.30E-20	0.49	270	364	+	snoRNA; CD-box	GO:0006396	
c25730/f1p0/842	25720/f1=0/942 DE	DE01227	DE01007 moD92	016	1 00E 19	0.45	102	201		Gene; snRNA;	GO:0005730
	КГ01227	SHOK85	91.0	1.90E-18	0.45	185	521	+	snoRNA; HACA-box	GO:0006396	

Table S2 Coffee long read sequencing isoforms annotated by Rfam (rfam.xfam.org).



Figure S1 Species distribution of coffee long read sequencing isoforms according to the result of BLASTx against NCBI non-redundant plant proteins (NR-plant). #, number.



Figure S2 Top-Hit Species distribution of coffee long read sequencing isoforms according to the result of BLASTx against NCBI non-redundant plant proteins (NR-plant). #, number.



Figure S3 Distribution of InterProScan families from coffee long read sequencing dataset



Figure S4 Pie chart and word cloud of coffee long read sequencing isoforms distribution to biological process, cellular component and molecular function



Figure S5 The KEGG pathway distribution of coffee long read sequencing isoforms.



Figure S6 Putative transcript variants from long-read sequencing aligned to reported genes encoding 7-methylxanthine N-methyltransferase. G-CaMXMT1, Arabica MXMT1 genomic DNA sequence; G-CaMXMT2, Arabica MXMT2 genomic DNA sequence; MXMT1, MXMT2, MXMT1, MXMT2 coding sequence; Ca, Arabica coffee; Cc, Robusta coffee; c20397/f5p1/1361, c10402/f2p3/1277, coffee long read sequencing isoforms.

Note: Black colour in the aligned area (correspond to G-CaMXMT reference genes) indicates sequences different to the reference.



Figure S7 Pie chart and word cloud of long sequences (coffee long read sequencing isoforms >10kb) distribution to biological process, cellular component and molecular function.