

Supplementary data 1

RNA-Seq statistics

Section 3.1

Illumina sequencing, assembly and
annotation

Table S1. Summary statistics of transcriptome sequencing data before and after quality trimming. Raw and clean sequencing data metrics, including total read bases, total number of reads, GC content (%), and quality scores (Q20 and Q30) across the three analyzed tissue samples.

<i>Before trimming</i>					
Tissue	Total read bases	Total reads	GC (%)	Q20 (%)	Q30 (%)
Posterior tentacle	7,360,321,066	72,874,466	39.82	97.34	95.36
Anterior tentacle	7,364,316,626	72,914,026	41.36	97.30	95.27
Mantle tissue	8,341,441,328	82,588,528	43.22	96.88	94.61
Total	23,066,079,020	228,377,020			
<i>After trimming</i>					
Tissue	Total read bases	Total reads	GC (%)	Q20 (%)	Q30 (%)
Posterior tentacle	7,313,773,613	72,692,756	39.78	97.49	95.57
Anterior tentacle	7,322,282,394	72,711,988	41.34	97.46	95.50
Mantle tissue	8,273,652,876	82,215,318	43.18	97.11	94.93
Total	22,909,708,883	227,620,062			
Recovery rate (%)		99.67			

- Total read bases: Total number of bases sequenced (= Total reads x Read length)
- Total reads: Total number of reads
- GC (%): GC content
- Q20 (%): Ratio of bases that have phred quality score greater than or equal to 20
- Q30 (%): Ratio of bases that have phred quality score greater than or equal to 30

Table S2. Overall mapping rates of clean reads to the reference genome. Summary of mapped reads, unmapped reads, and overall mapping percentages for each tissue sample.

Tissue	Total Reads	Mapped Reads	Unmapped Reads	Mapping rate
Posterior tentacle	72,692,756	45,384,966	27,307,790	62.43%
Anterior tentacle	72,711,988	44,847,546	27,864,442	61.68%
Mantle tissue	82,215,318	49,500,806	32,714,512	60.21%

- Total Reads: Number of cleaned reads after trimming
- Mapped reads: Number of reads with at least one reported alignment
- Unmapped reads: Number of unmapped reads

Table S3. Summary statistics of Trinity *de novo* transcriptome assembly and unigene clustering. Assembly metrics for total contigs and clustered unigenes. Metrics include gene and transcript counts, GC content (%), length distribution statistics (N90- N10), length extremes (maximum, minimum, median, and average in bp), and total assembled bases.

Category	Contigs	Unigene
Total trinity genes	142,638	142,526
Total trinity transcripts	177,507	142,526
GC content (%)	40.66	40.52
N90	287	269
N80	399	358
N70	540	471
N60	724	619
N50	966	825
N40	1,278	1,108
N30	1,695	1,514
N20	2,275	2,098
N10	3,289	3,067
Maximum contig length	25,081	25,081
Minimum contig length	201	201
Median contig length	411	380
Average contig length	675.5	614.2
Total assembled bases	119,899,779	87,544,653

- Total trinity genes: Total number of assembled genes by Trinity
- Total trinity transcripts: Total number of assembled transcripts by Trinity
- Unigenes: The longest contigs of the assembled contigs are filtered and clustered into the non-redundant transcripts using CD-HIT-EST program.
- Percent GC : GC percent
- N[x] length statistic: At least x% of the assembled transcript nucleotides are found in contigs that are at least of Nx length. The traditional method is computing N50, such that at least half of all assembled bases are in transcript contigs of at least the N50 length value. The contig N50 values can often be exaggerated due to an assembly program generating too many transcript isoforms, especially for the longer transcripts. To mitigate this effect, Nx values are also calculated based on the longest isoform per gene.
- Maximum/Minimum/Median/Average contig length:
Maximum/Minimum/Median/Averaged length of assembled transcripts
- Total assembled bases: The left column is total bases of assembled contigs and right column is total bases of the longest isoform.

Table S4. Functional annotation summary of assembled unigenes across public databases. Number and percentage of unigenes annotated using automated homology searches against major databases (GO, UniProt, NR, Pfam, EggNOG, and NT) alongside manual annotation of GPCR and other chemosensory target genes. Percentage values are calculated relative to the total 142,526 assembled unigenes.

Annotation method	No. Unigenes annotated	Annotation ratio (%)
GO	19,074	13.4
UniProt	17,995	12.6
NR	27,522	19.3
Pfam	17,468	12.3
EggNOG	21,073	14.8
NT	5,266	3.7
Manual annotation	179	0.1
Overall	28,310	19.9

- GO: Gene Ontology terms
- UniProt: Protein sequence and functional information including Swiss-Prot and TrEMBL
- NR: NCBI Non-Redundant protein database
- Pfam: Database of Protein families and domains
- EggNOG: Database of Non-supervised Orthologous Groups (orthologs)
- NT: NCBI Nucleotide database
- Manual annotation: Manually annotated by authors mostly for GPCRs.