Supplemental Table I. Distribution of percent length coverage for the top matching UniProt database entries.

Percentage length coverage bin	Count in bin	Cumulative count
100	3071	3071
90	1205	4276
80	851	5127
70	806	5933
60	752	6685
50	697	7382
40	653	8035
30	532	8567
20	362	8929
10	0	8929
0	0	8929

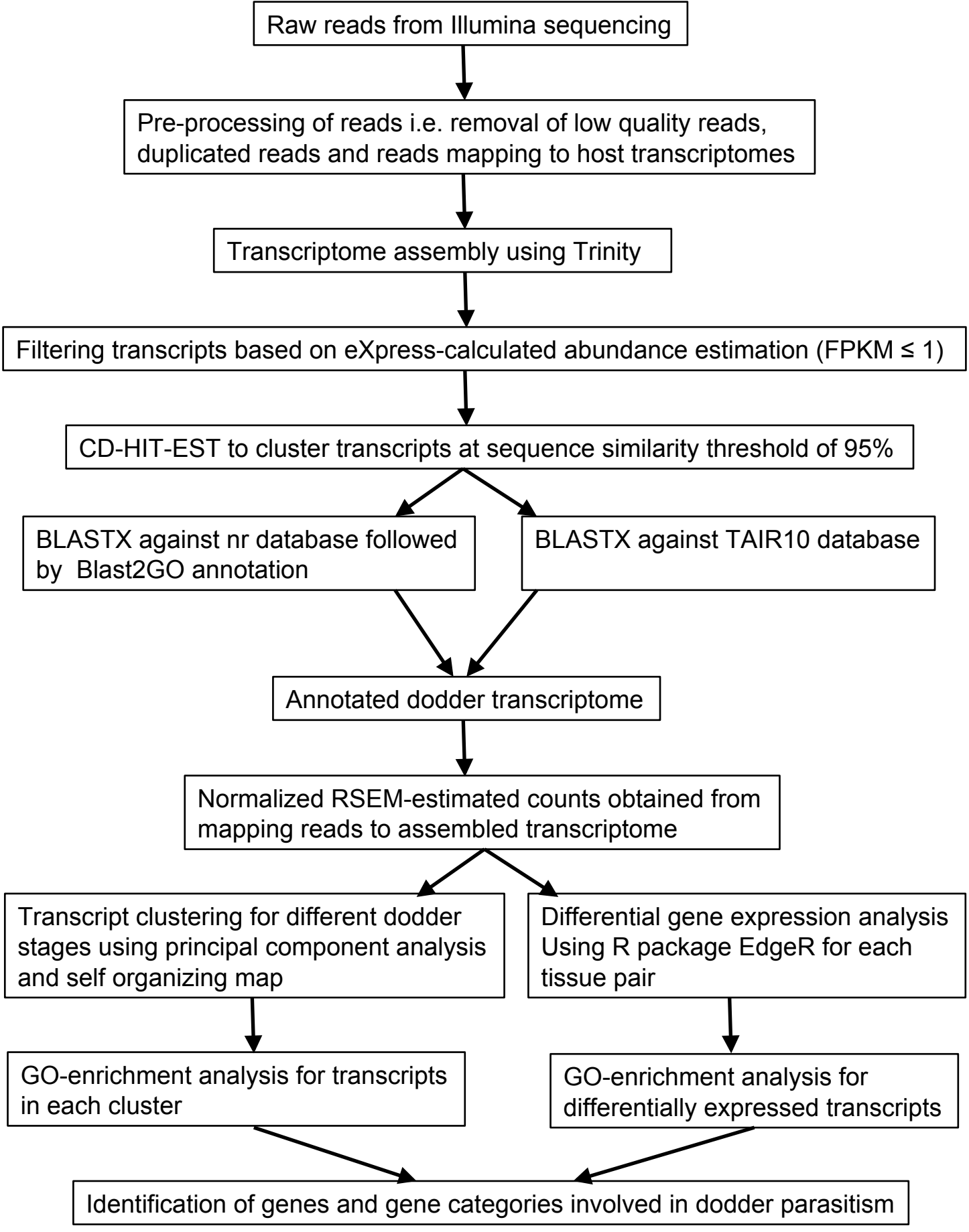
Examples of statements that can be made based on above table:
- 3071 database proteins matches to a Trinity transcript by > 90% alignment coverage.
- 1205 database proteins matches to a Trinity transcript with an alignment coverage
  between 80% and 90%.
- 4276 database proteins matches to a Trinity transcript by > 80% alignment coverage.

Supplemental Table II. Size statistics and percentage of read mapping to annotated and un-annotated transcripts.

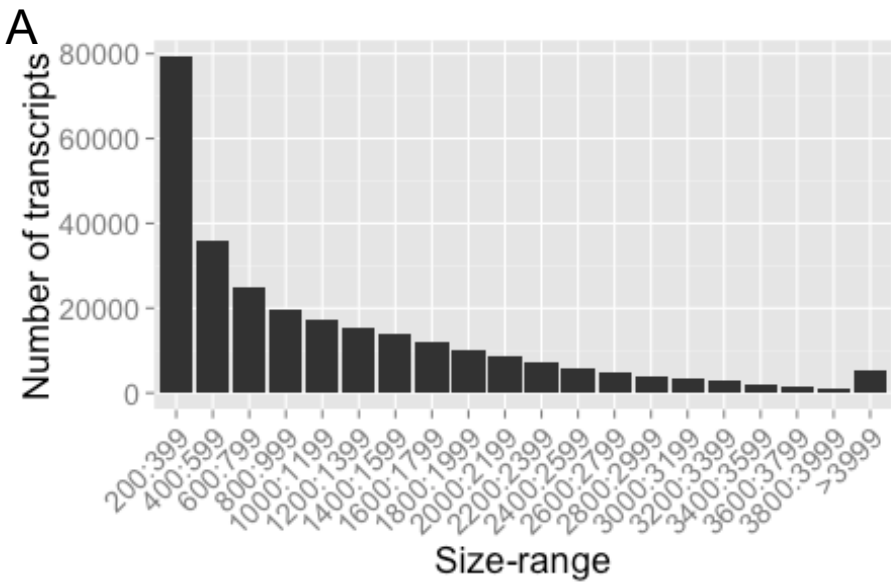
	Annotated_transcripts	Un-annotated transcripts
N50(bp)	1879	649
Average_transcript_length(bp)	1478	524
Median_transcript_length(bp)	1278	387
%total_mapped_reads	72.90	6.50
%Uniquely_mapped_reads	49.25	4.86

Supplemental Table III. Primers used in RT-PCR analysis.

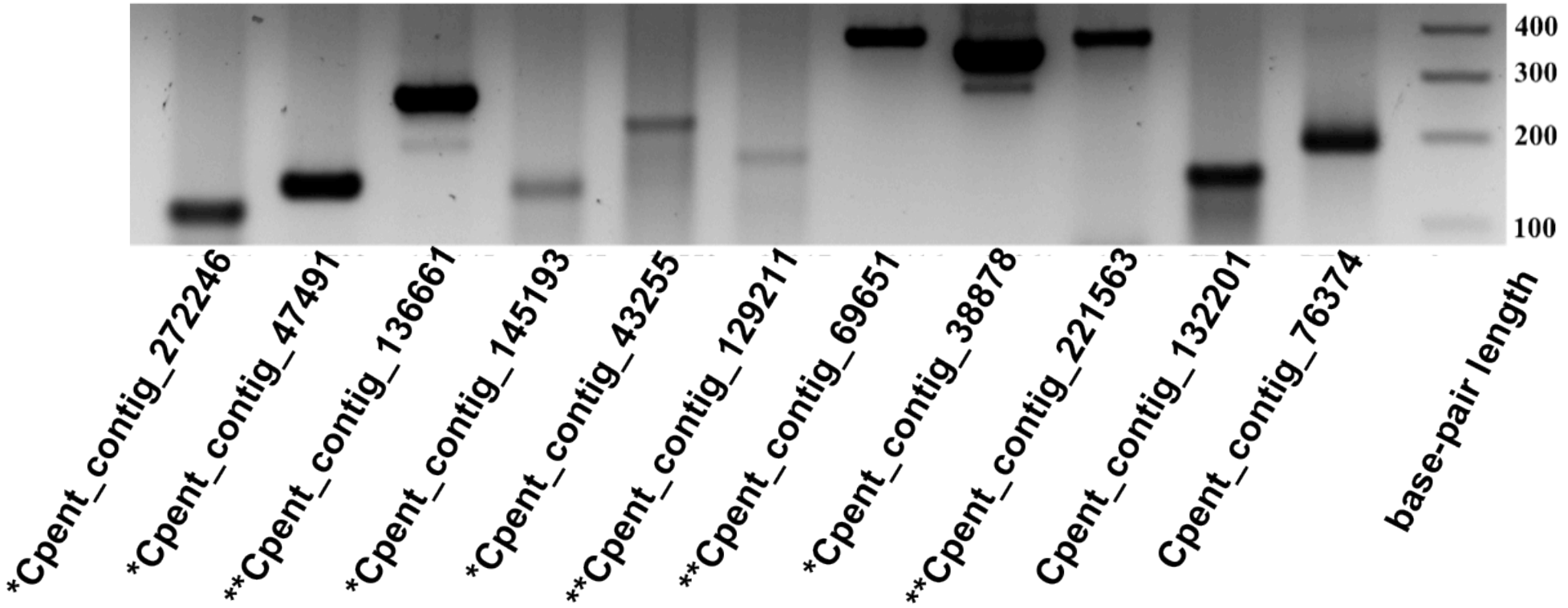
Transcript number (Cpent_contig)		Sequence (5' to 3')
272246	Sense	AAGGCCAACAAGCAAGATTATTA
	Anti-sense	ATTTCTTCTCCCCCACCATATAA
47491	Sense	GAAAGAAGAAACGAAATTTGTGC
	Anti-sense	TCAAAGTGTCAACCTTGCATTC
136661	Sense	TGATGATCCTTTTGAAGAATACGA
	Anti-sense	CAATCCCAAGATGATTCCTTACA
145193	Sense	TCTTGATCTGTGTCAGGACCTTT
	Anti-sense	ACTTCTCCCAATCATTACCCAAC
43255	Sense	TGATGATAGTGCATTTGTGAAGG
	Anti-sense	GGATGATACACAGTGAACCTTGG
129211	Sense	TGCCATCATCATTATTAAGCATT
	Anti-sense	CGGCTTCTTCTTGTTATTCTTCC
69651	Sense	ATGATTTGCCTAGCTTCACTCAA
	Anti-sense	TGATTCTCCTCCTTGTCATTTGT
38878	Sense	AATCGCAATGATGAGTTCTTGTT
	Anti-sense	CAGCTTCATGGAGTATTTGATCC
221563	Sense	CTCGAATAACAGAAGCAGCAAGT
	Anti-sense	GTCCAGATATGACCTTGAAGCAG
132201	Sense	ATGGAAAGTTGGTCCAGAAAGA
	Anti-sense	ATCTCTGCCCTGTCGATAAGAA
76374	Sense	CCATTTGAAGGTGGAGTTTTTC
	Anti-sense	AATAATTGCCCGACAGACAGAT



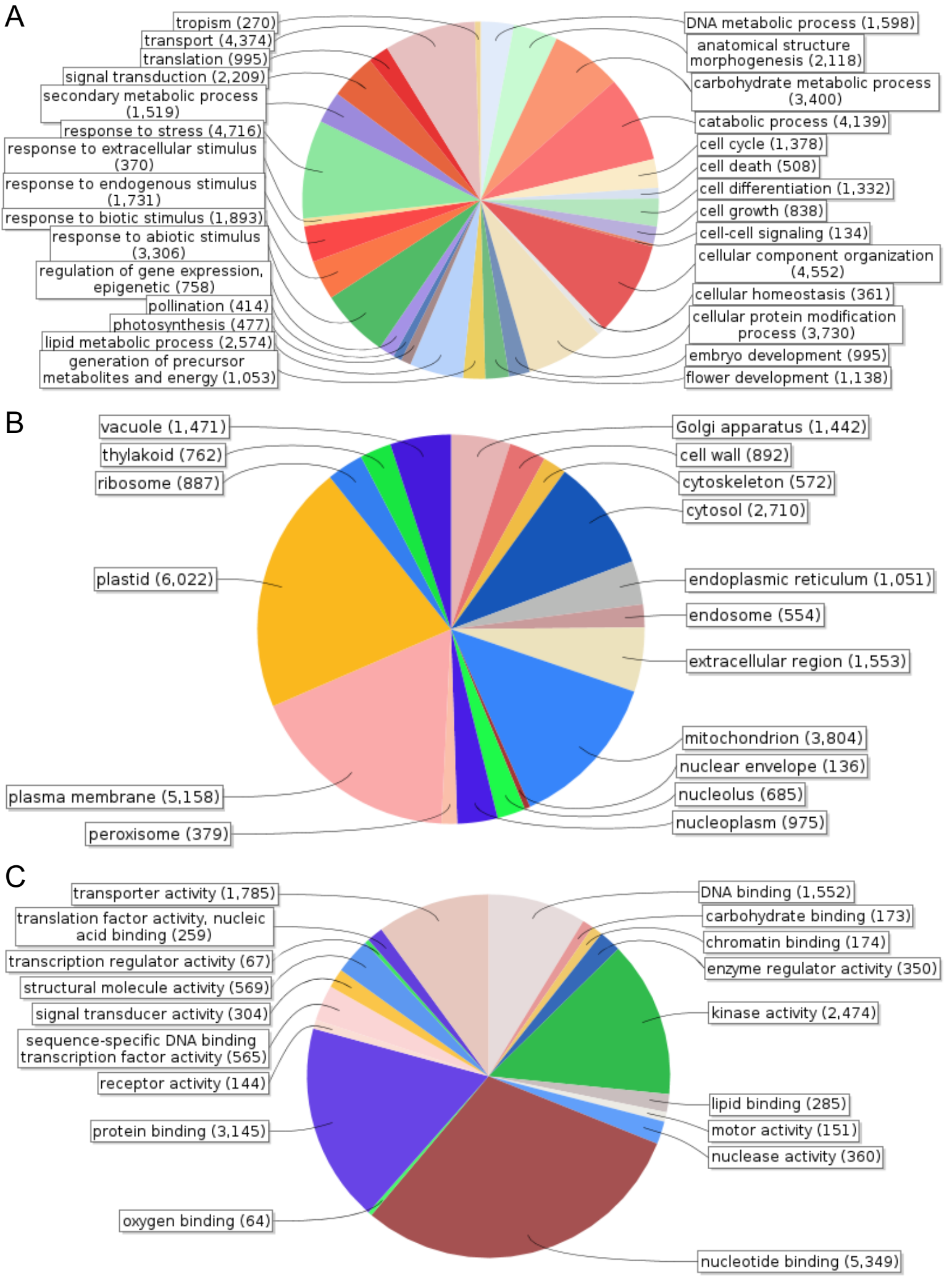
Supplemental Figure S1. Flow-chart showing steps in dodder transcriptome assembly and annotation, and downstream transcript clustering and differential expression analysis



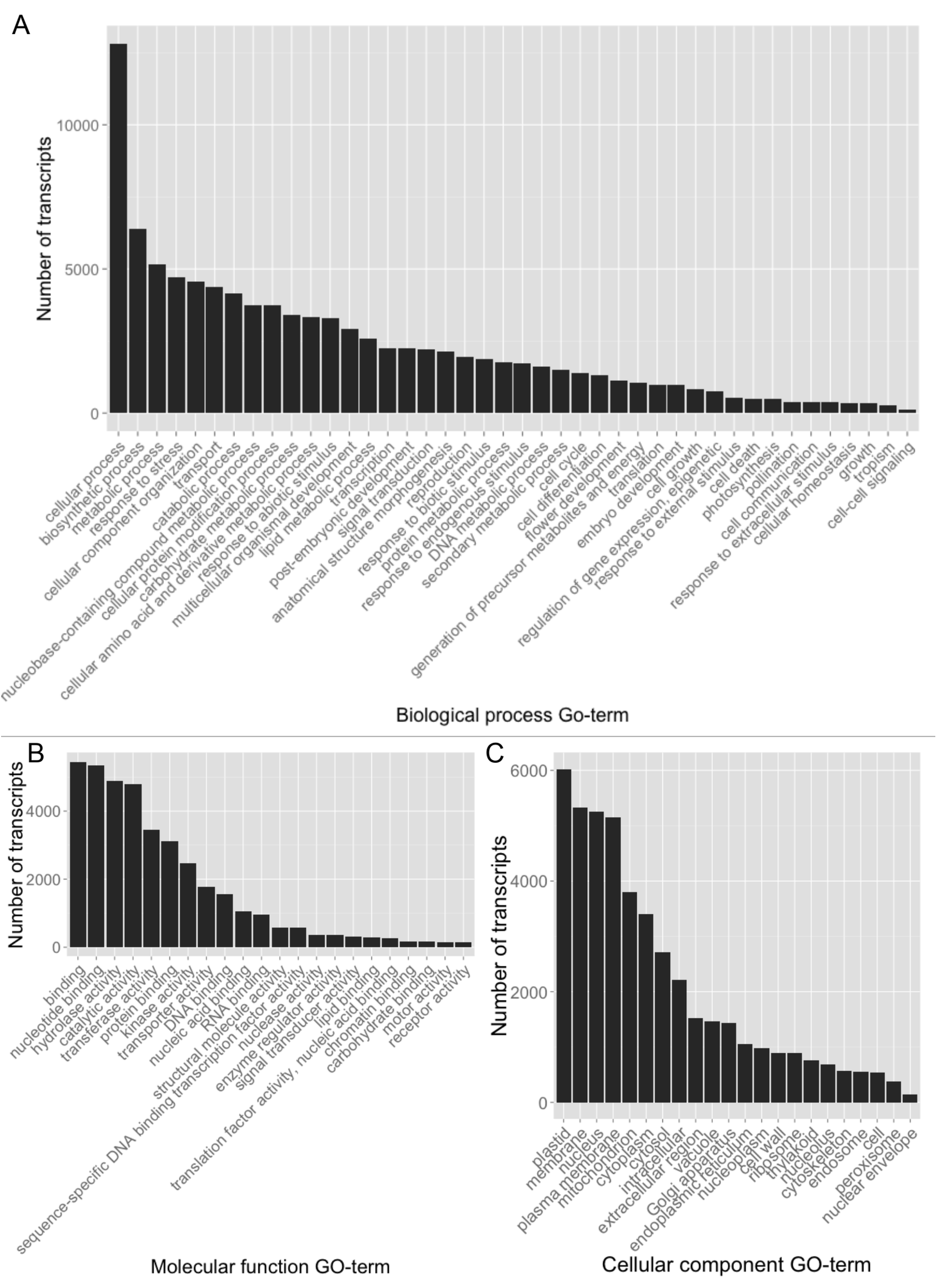
Supplemental Figure S2. Transcript size distribution for Dodder_all_transcriptome



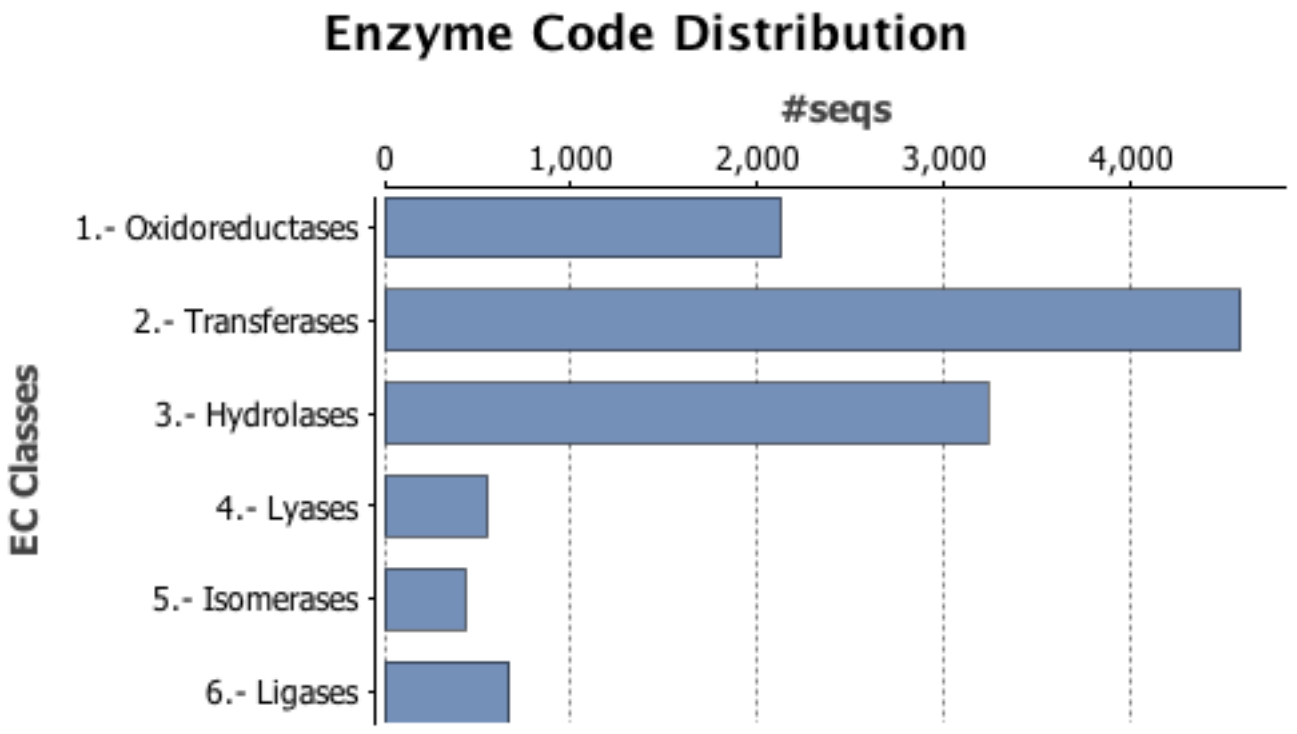
Supplemental Figure S3. Expression of non-annotated transcripts as detected by RT-PCR in dodder stems. Contigs with single asterisk are non-annotated transcripts with a predicted ORF, whereas with double asterisks are non-annotated transcripts without a predicted ORF. Cpent_contig_132201 and Cpent_contig_76374 were used as controls which are annotated as *CAP-BINDING PROTEIN 20* and *UBIQUITIN-CONJUGATING ENZYME 21*, respectively and found to be highly expressed in all tissues from expression data.



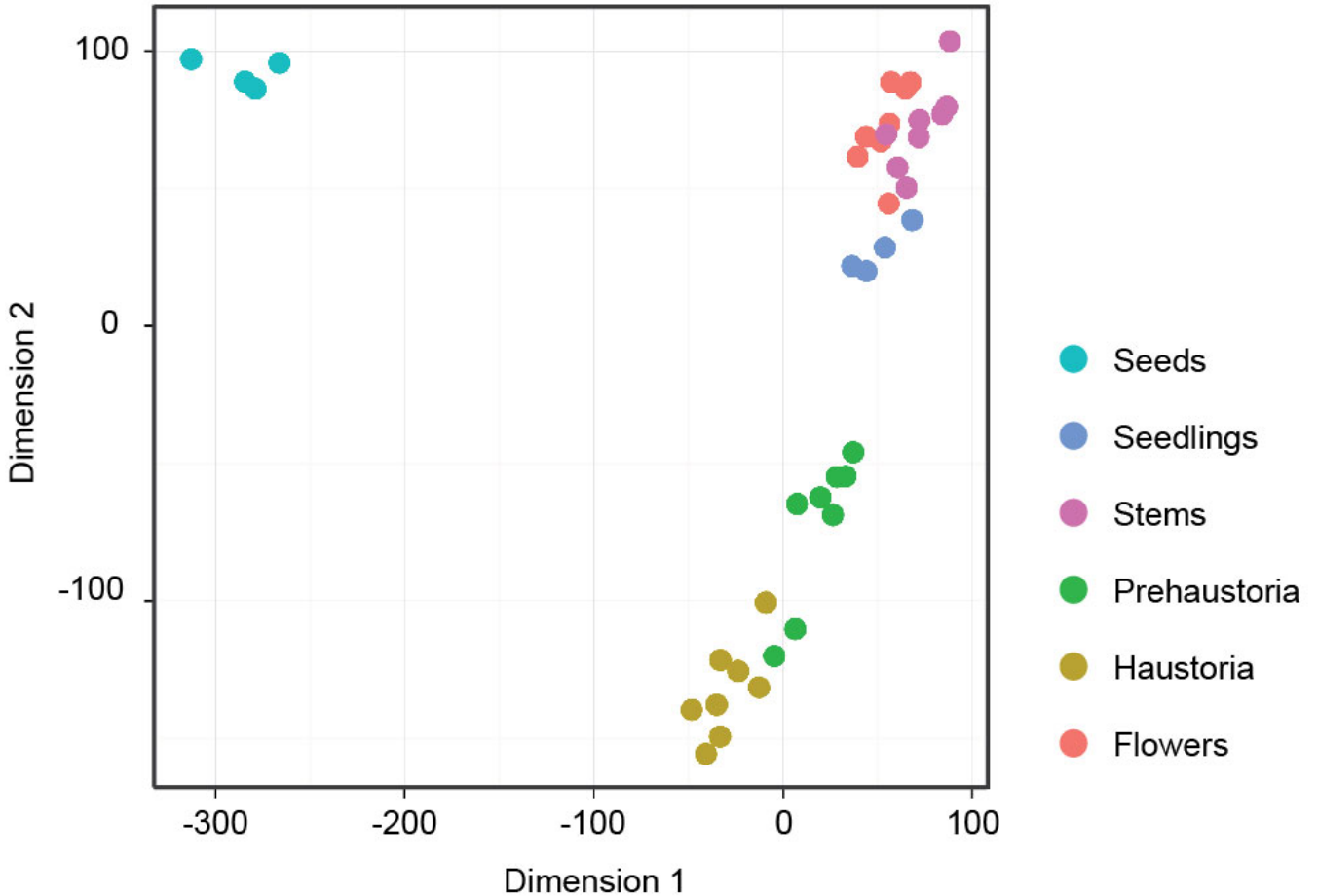
Supplemental Figure S4. Pie charts for multilevel GO distribution of annotated transcripts in three categories: biological processes (A), cellular components (B) and molecular function (C) .



Supplemental Figure S5. Histogram representation of GOslim classification in three categories: biological processes (A), molecular function (B) and cellular components (C).



Supplemental Figure S6. Distribution of transcripts annotated as enzymes among different enzyme classes.



Supplemental Figure S7. Multidimensional scaling (MDS) plot of all replicates of each dodder tissue used for transcriptome assembly and, subsequently, transcript clustering and differential expression analysis. There were eight replicates for stems, prehaustoria, haustoria and flowers (four replicates each from dodder grown on two host plants, tomato and tobacco) and four replicates for seeds and seedlings.