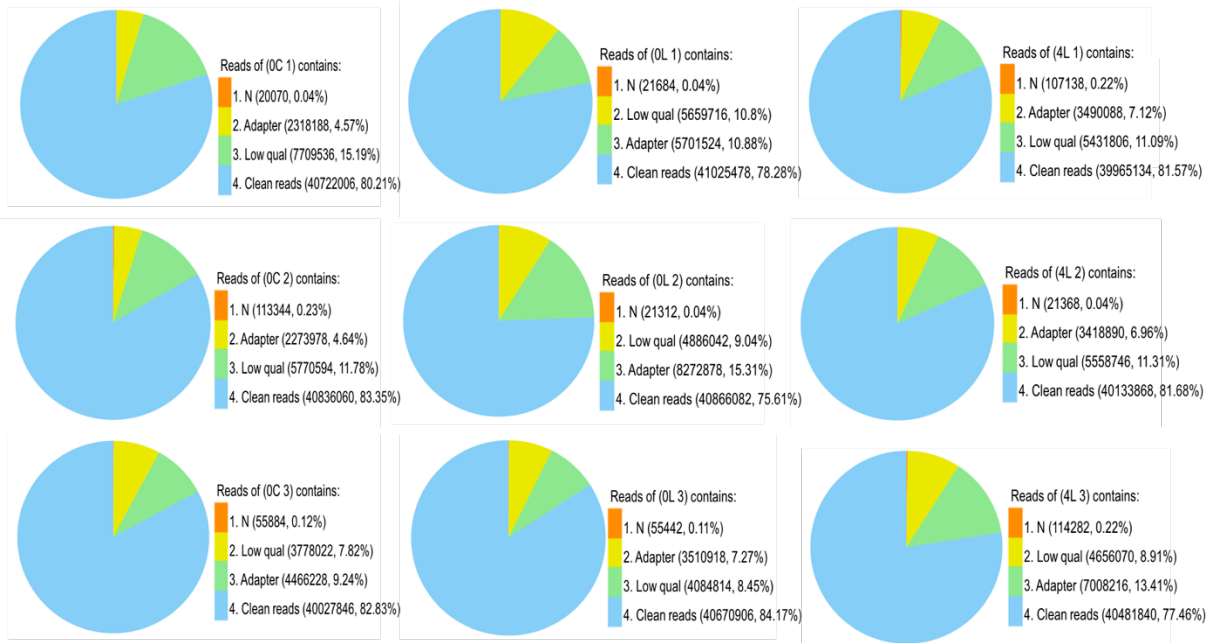


1 Title: **Large-scale transcriptome changes in the process of long-term visual memory**
 2 **formation in the bumblebee, *Bombus terrestris***

3 Li Li^{1*}, Songkun Su^{2*}, Clint J. Perry¹, Maurice R. Elphick¹, Lars Chittka¹, Eirik Søvik³

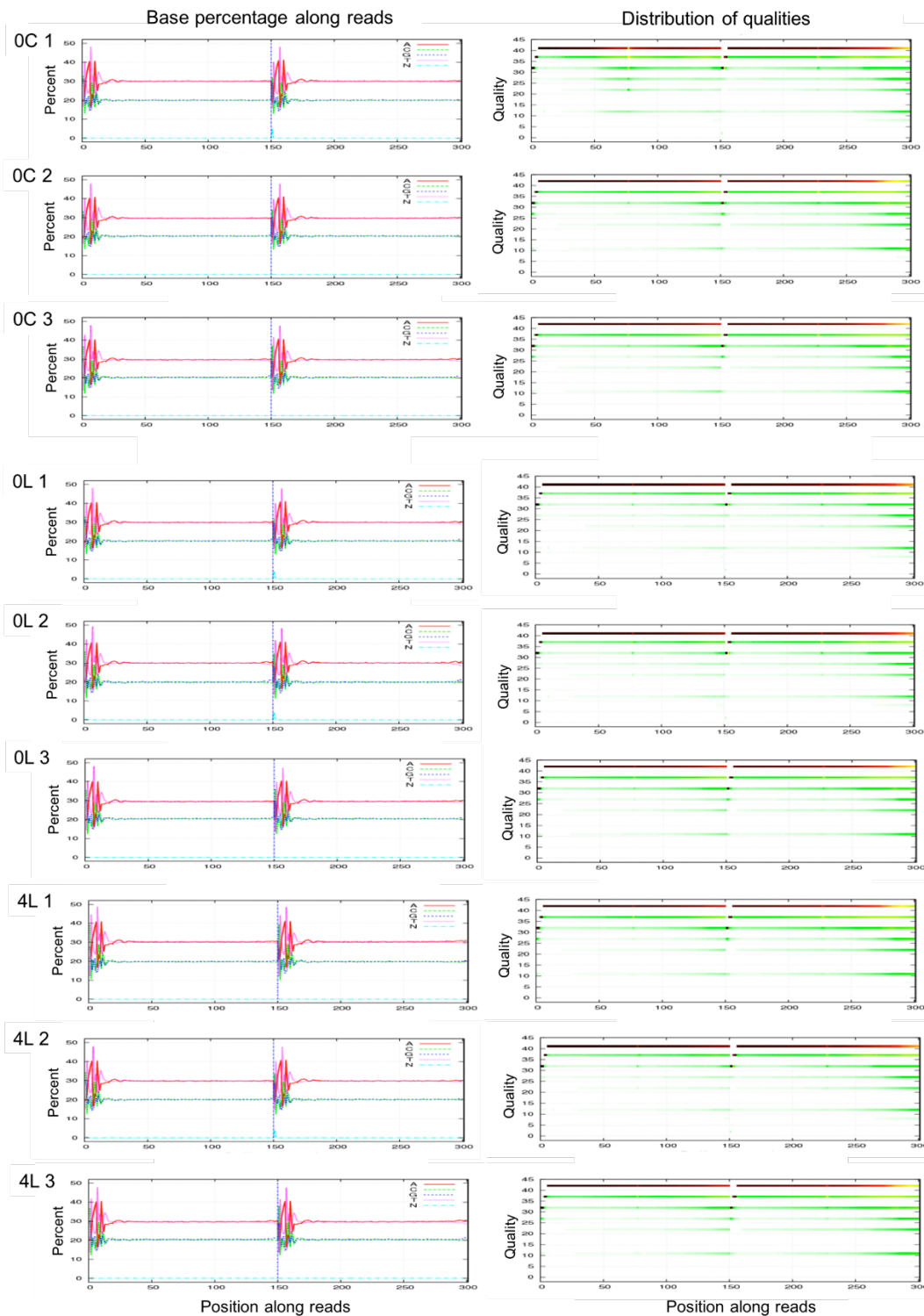
6 **Supplementary Information**

8 **Supplementary Figures**



10
 11 **Supplementary Figure 1 | Classification of raw reads for each sample.** Before data analysis, the reads
 12 with more than 10% unknown nucleotides (N), with adapters (Adapter), and low quality reads (Low qual)
 13 were removed from raw reads. The remaining reads were the clean reads. The value indicates reads
 14 number and its ratio on raw reads. OC: 0-hour Control; OL: 0-hour Learning; 4L: 4-hour Learning.

15

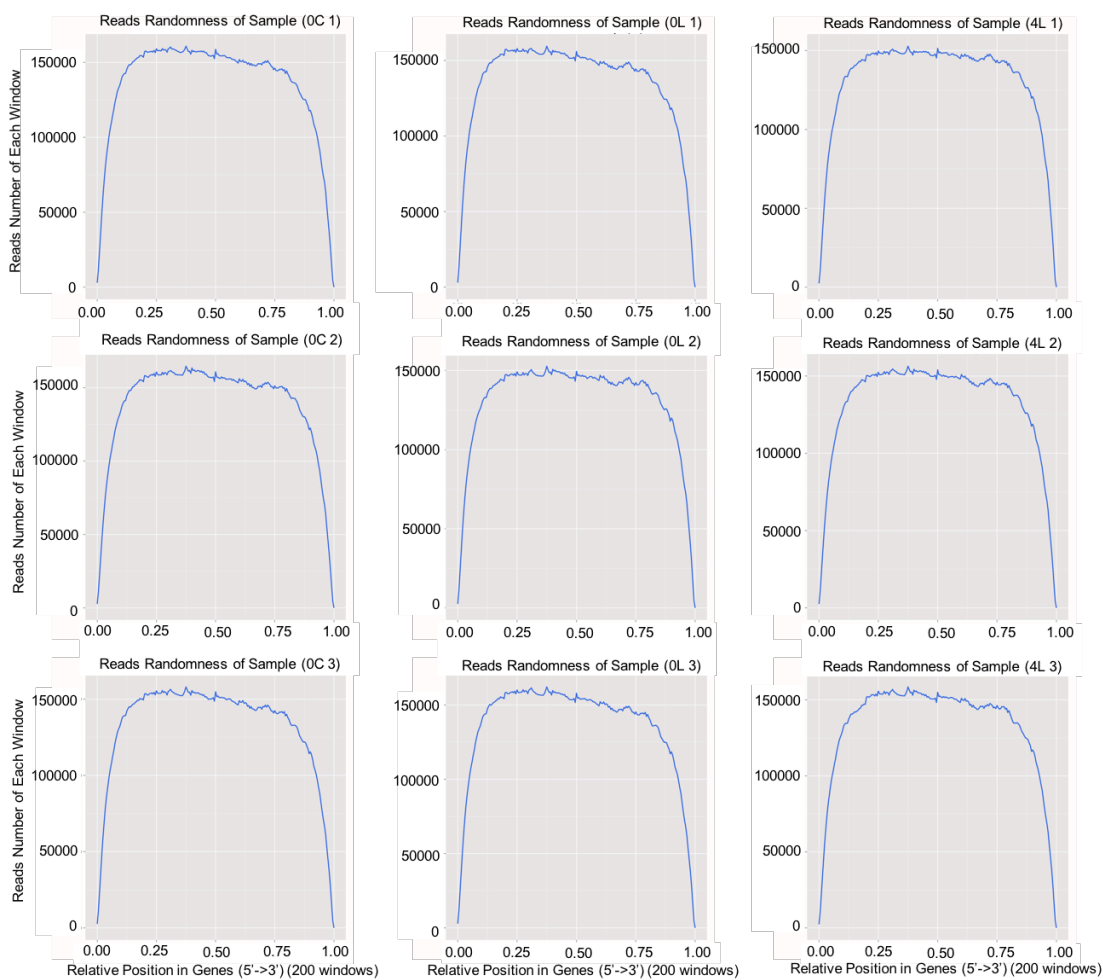


16

17 **Supplementary Figure 2 | Quality assessment of sequencing paired-end reads. (Left) The nucleotide**
 18 **composition of RNA-Seq reads.** The percentage of each nucleotide (Y-axis) is plotted against read length
 19 (Y-axis; 1-150bp and 151-300bp represent the reads from end 1 and end 2 separately). Colors indicate
 20 different nucleotides (A = red; C = Green; G = Blue; T = Magenta; N: Light Blue). A curve should be
 21 overlapped with T curve, while G curve be overlapped with C curve, which can be seen from all our

22 samples. If abnormal condition happens during sequencing, it may show an unbalanced composition. Note
 23 that the changing composition in the beginning of reads is common in RNA-Seq data. **(Right) Quality**
 24 **distribution of bases along RNA-Seq reads.** Each base quality (Y-axis) is plotted against read length (X-
 25 axis; 1-150bp and 151-300bp represent the reads from end 1 and end 2 separately). Quality score reflects
 26 the sequencing error rate and the relationship between them is sequencing error rate 1%, 0.1% and 0.01%
 27 corresponds to quality score 20, 30 and 40. Most of base positions in our sequencing showed good quality
 28 (score >20). 0C: 0-hour Control; 0L: 0-hour Learning; 4L: 4-hour Learning.

29

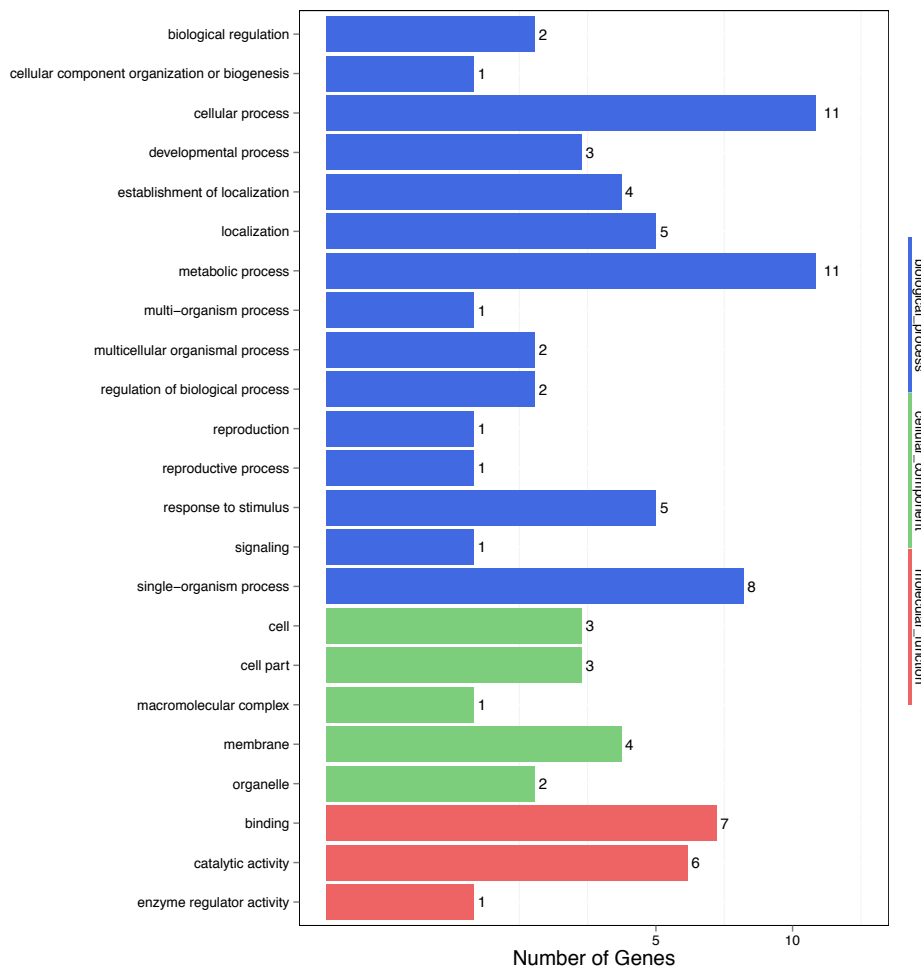


30

31 **Supplementary Figure 3 | The distributions of reads on bumblebee (*Bombus terrestris*) reference**
 32 **genes.** X-axis is the relative position in genes which is calculated as the ratio between read location and the
 33 gene length and Y-axis is the number of reads. Reads should be evenly distributed on reference genes,
 34 otherwise it means the randomness is poor (i.e. reads prefer to specific gene region) which will affect

35 following analysis. The read randomness in all our samples is good as shown in the figure. 0C: 0-hour
36 Control; 0L: 0-hour Learning; 4L: 4-hour Learning.

37
38
39



40

41 **Supplementary Figure 4 | GO functional classification of differentially expressed genes (DEGs).** The
42 DEGs were annotated into three main categories: biological process, cellular component and molecular
43 function. The number of genes in each GO terms were displayed and 58 of the 110 DEGs belonged to
44 biological process.

45

46

47

48

49 **Supplementary Table**

50

51 **Supplementary Table 1 | The quality of RNA samples used for sequencing.** RNA concentration and
 52 integrity were measured by Agilent 2100. Level A means the sample is qualified and the amount of sample
 53 satisfies two times library construction or more. 0C: 0-hour Control; 0L: 0-hour Learning; 4L: 4-hour
 54 Learning.

Sample Name	Concentration (ng/μl)	Volume (μl)	Total Mass (μg)	RIN	Library Type	Test Result
0C 1	456	55	25.08	7.6	HiSeq Eukaryotic Transcriptome	Level A
0L 1	474	55	26.07	7.7	HiSeq Eukaryotic Transcriptome	Level A
4L 1	393	56	22.01	7.9	HiSeq Eukaryotic Transcriptome	Level A
0C 2	480	57	27.36	7.9	HiSeq Eukaryotic Transcriptome	Level A
0L 2	516	57	29.41	8	HiSeq Eukaryotic Transcriptome	Level A
4L 2	627	57	35.74	7.9	HiSeq Eukaryotic Transcriptome	Level A
0C 3	528	57	30.1	8	HiSeq Eukaryotic Transcriptome	Level A
0L 3	549	54	29.65	8	HiSeq Eukaryotic Transcriptome	Level A
4L 3	639	57	36.42	7.9	HiSeq Eukaryotic Transcriptome	Level A

55

56

57