

Supplementary material for:

A comparative genomic analysis at the chromosomal-level reveals evolutionary patterns of aphid chromosomes

Chen Huang¹, Bingru Ji^{1,2}, Zhaohui Shi², Jiangyue Wang^{1,2}, Jiaqing Yuan¹, Peng Yang¹, Xiao Xu¹, Haohao Jing¹, Lulu Xu¹, Jing Fu¹, Le Zhao^{3,1}, Yandong Ren^{1*}, Kun Guo^{2*}, Gang Li^{1*}

¹College of Life Sciences, Shaanxi Normal University, Xi'an 710119, China.

²Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100193, China.

³QinLing-Bashan Mountains Bioresources Comprehensive Development C. I. C., School of Bioscience and Engineering, Shaanxi University of Technology, Hanzhong, 723000, P.R. China.

***Corresponding authors:** E-mails:

Yandong Ren, renyandong@snnu.edu.cn;

Kun Guo, kunguoimplad@foxmail.com;

Gang Li, gli@snnu.edu.cn.

Supplementary Texts

Gene expression pattern in *Semiaphis heraclei*

We analyzed the global gene expression patterns in both sexual females and males of *S. heraclei*, confirming dosage compensation across the X chromosome, despite minor differences in X-linked gene expression between sexes (Supplementary Figure 13a and Supplementary Figure 14). We observed that the X chromosome of *S. heraclei* exhibited significantly lower gene expression compared to autosomes, with a lower proportion of expressed genes (Supplementary Table 19, Mann-Whitney U test, p-value = 0.00011). This pattern was characterized by reduced gene expression density and levels, especially in the terminal regions of the X chromosome (Supplementary Figure 13c, d and Supplementary Figure 15). Interestingly, the decrease in X chromosome gene expression was inversely correlated with the density of DNA transposons, LTR retrotransposons, and LINE retrotransposons (Supplementary figs. 16-17 and Supplementary Tables 20-21). This correlation was not observed with SINE retrotransposons, likely due to their shorter length. These findings align with those of Mathers et al.¹.

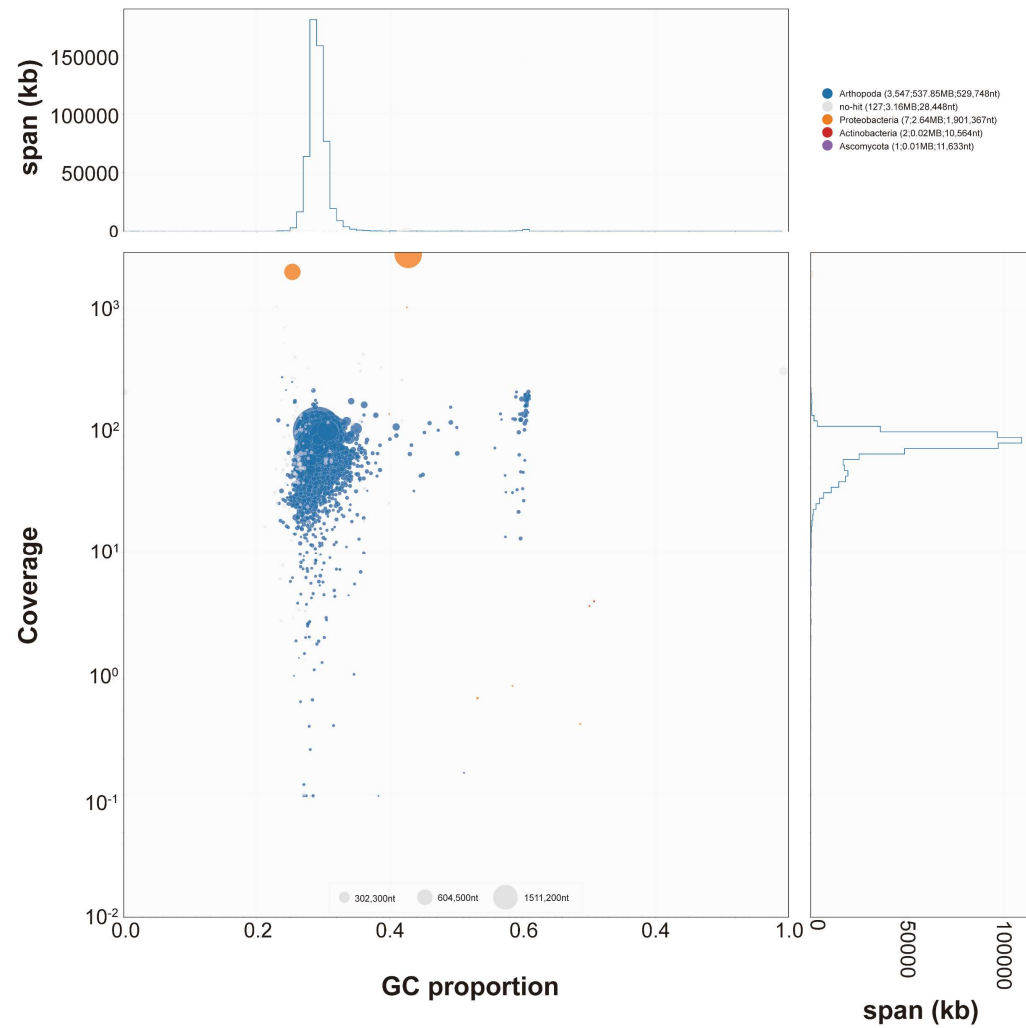
In our study, the gene expression of male and female of *S. heraclei* showed distinct patterns. Males exhibited upregulation of 2,350 genes (including 635 on the X chromosome and 1,719 on autosomes), which is higher than 915 upregulated genes of females (195 on the X chromosome and 715 on autosomes) (Supplementary Table 19, 22-23). The upregulated autosomal genes in males were enriched in functions related to olfactory sensation, response to foreign compounds, and metabolism, possibly aiding in their winged migration and environmental adaptation. The *fln* gene, crucial for insect flight² and located on chromosome SH4, was significantly upregulated in males (Supplementary Figure 13b). In contrast, male X-linked genes were enriched in functions related to cilia assembly, motility, and reproduction. Females showed a different profile with upregulated autosomal genes enriched in functions related to meiosis, chromosome organization, and double-strand break repair. Female X-linked genes were enriched in functions related to transposon regulation, oocyte development, and meiotic chromosome organization (Supplementary Figure 18 and table 24). These results suggested a divergence in genetic expression patterns tailored to the distinct reproductive and survival strategies of each sex.

Supplementary Methods

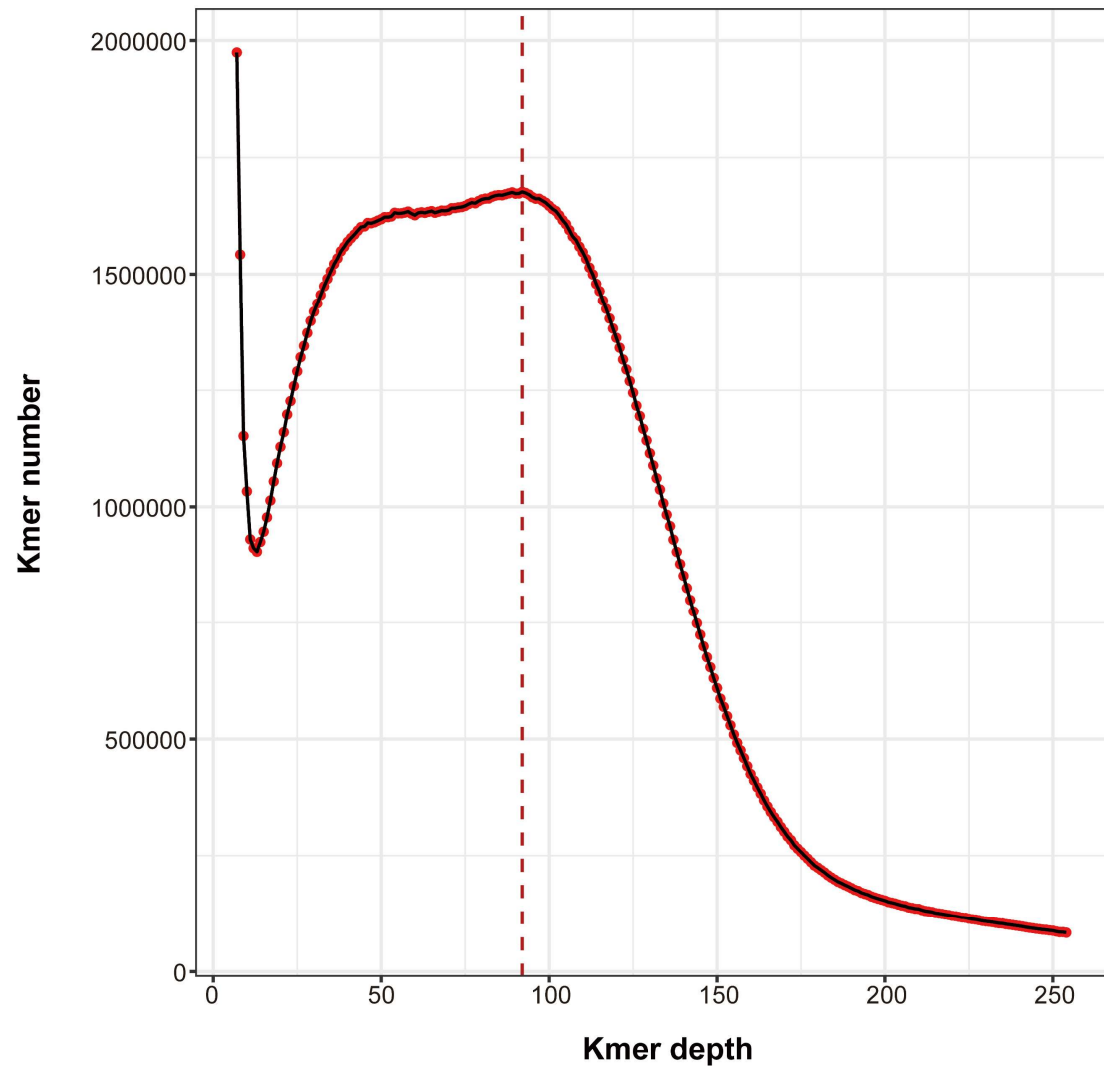
Gene expression analysis

We used newly generated RNA-seq data to explore gene expression patterns within the *S. heraclei* genome, including sexual females and winged males, with each group consisting of three biological replicates (Supplementary Table 31). Gene expression level for each sample was estimated using StringTie v2.2.1³. Subsequently, we identified differentially expressed genes between sexual reproduction females and winged males of *S. heraclei* using the DESeq2 v1.32.0 R package⁴. The analyzed genes needed to exhibit an average TPM greater than 1 in at least one of the morphs. To generate gene expression and expression level distribution on *S. heraclei* chromosomes (Supplementary fig 13ce), we used newly generated RNA-seq data from *S. heraclei* asexual female to calculate TPM for each protein-coding gene (Supplementary Table 31). Genes with TPM values greater than 1 were considered as expressed genes. We counted the number of expressed genes in 100 kb fixed windows and the distribution of log10 values for TPM of expressed genes.

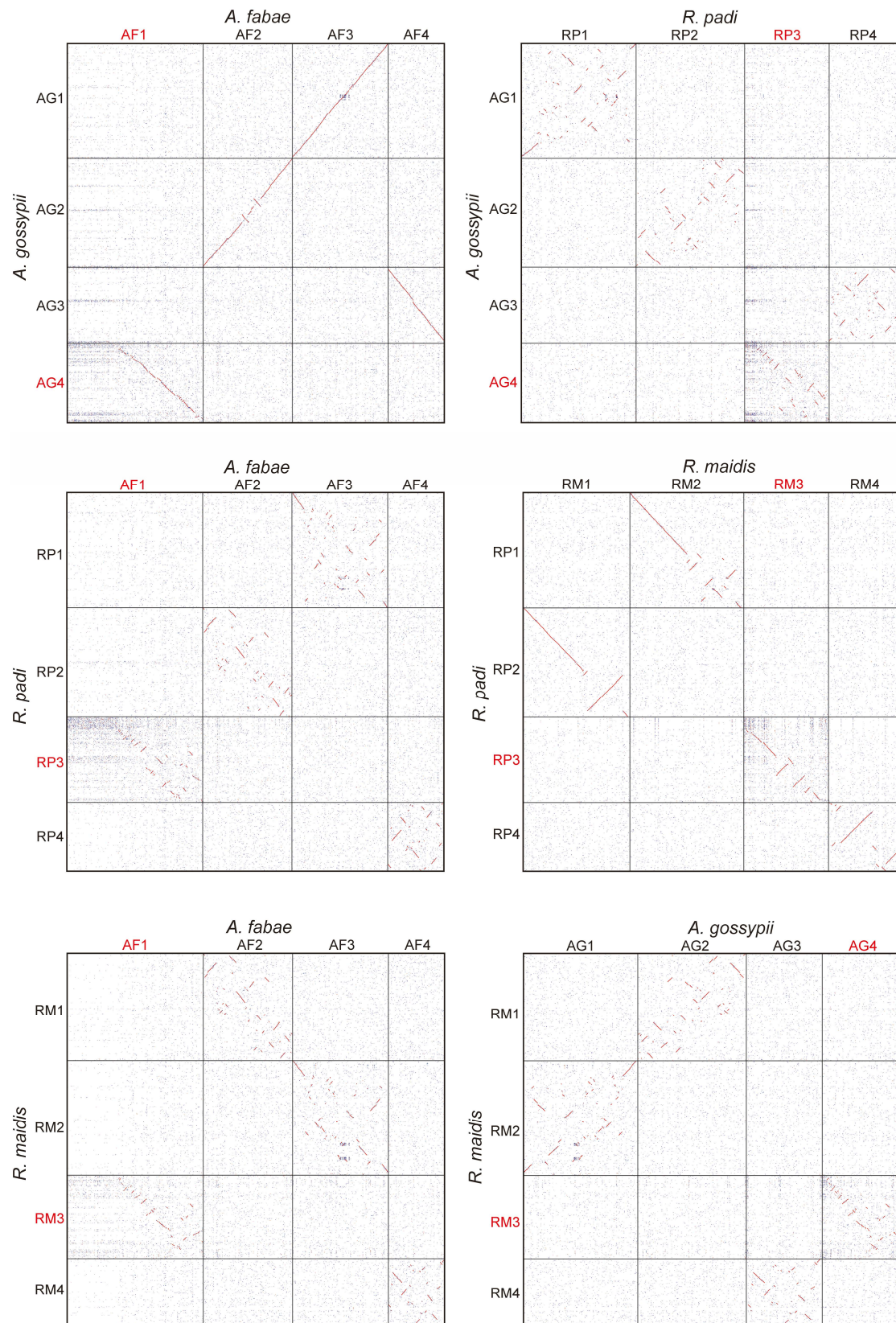
Supplementary Figures



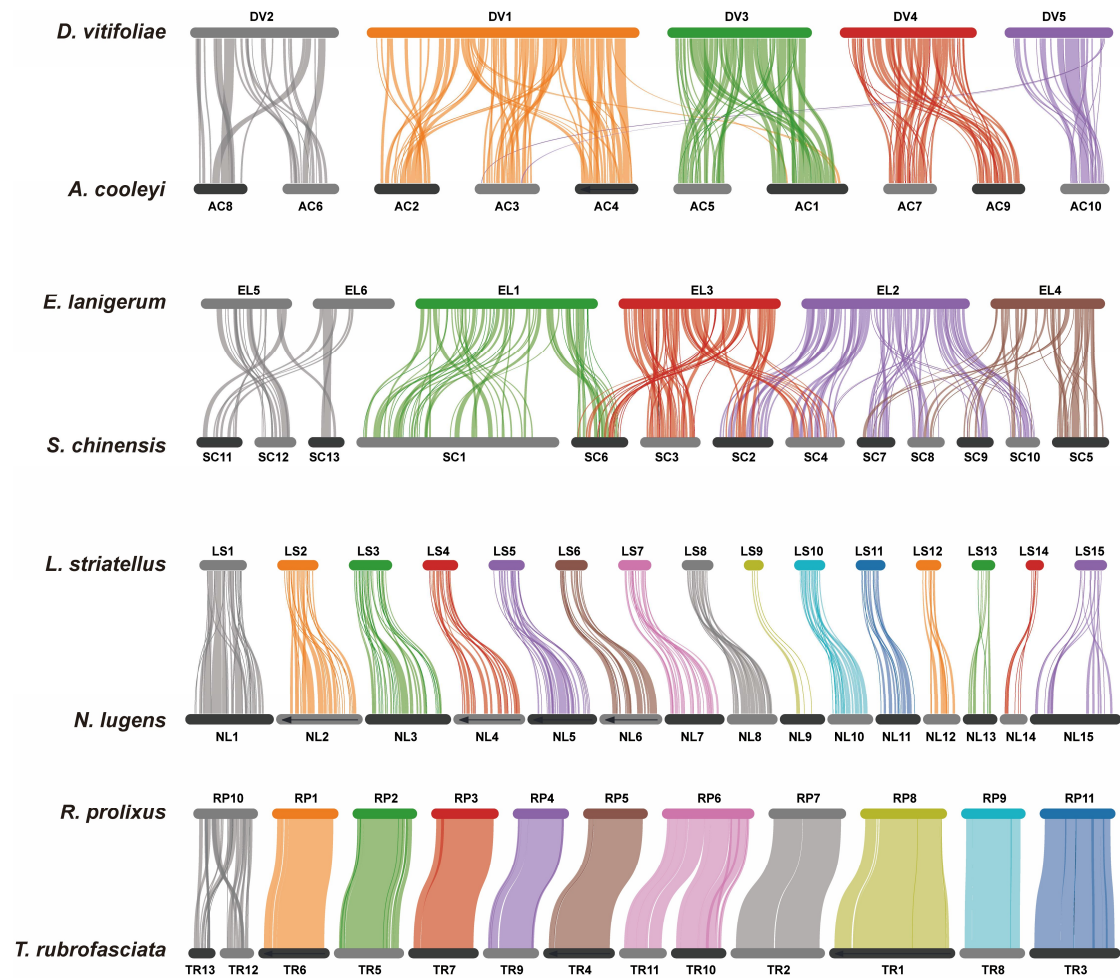
Supplementary Figure 1. The GC-content-coverage plot of the *S. heraclei* genome assembly. Each circle represents a sequence, with circle size scaled by sequence length, and colored according to the NCBI classification assigned by Blobtools. The x-axis represents the average GC content of each sequence, while the y-axis displays the average coverage calculated from the alignment of original Illumina reads. The marginal histograms depict the cumulative genome content (in kb) within GC content (x-axis) and coverage (y-axis) intervals.



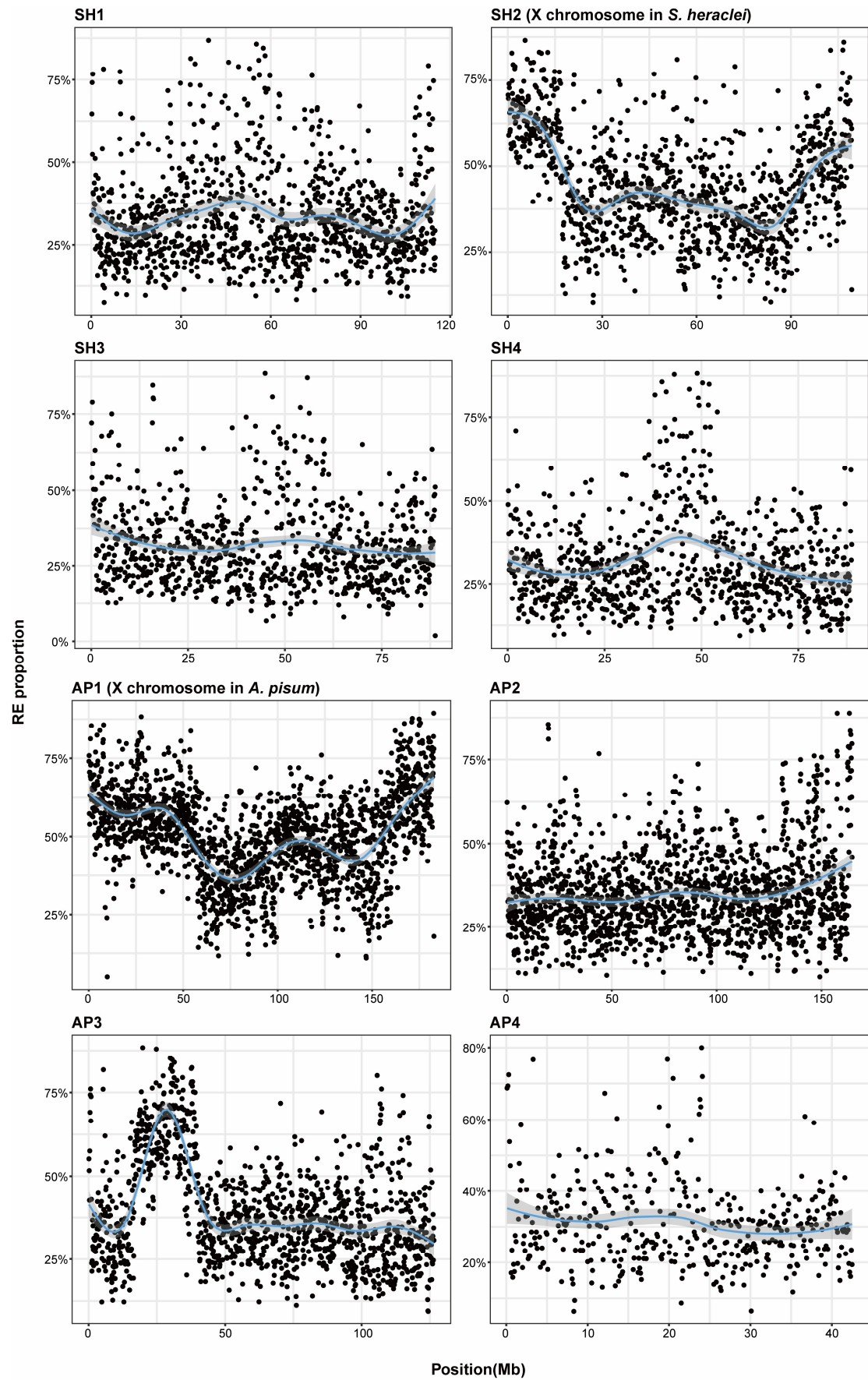
Supplementary Figure 2. K-mer frequency distribution plot. The x-axis represents k-mer coverage depth, while the y-axis indicates the count of corresponding k-mers. The red dashed line represents the peak depth of k-mer frequency distribution (kmer = 92).



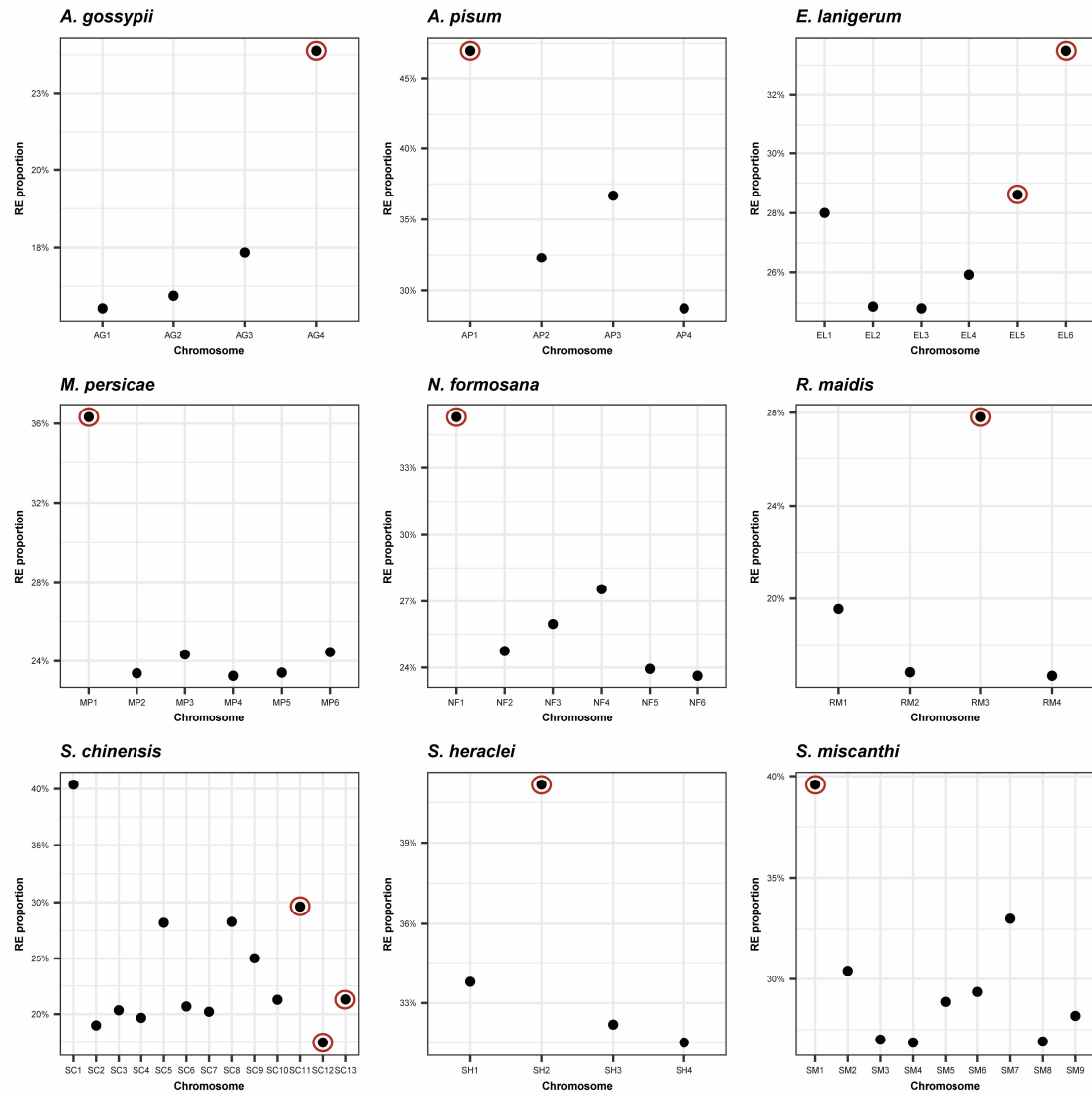
Supplementary Figure 3. Synteny dot plot between the genomes of four aphid species in Aphidini. AG4, AF1, RM3, and RP3 correspond to the X chromosome of the respective species, shown in red font.



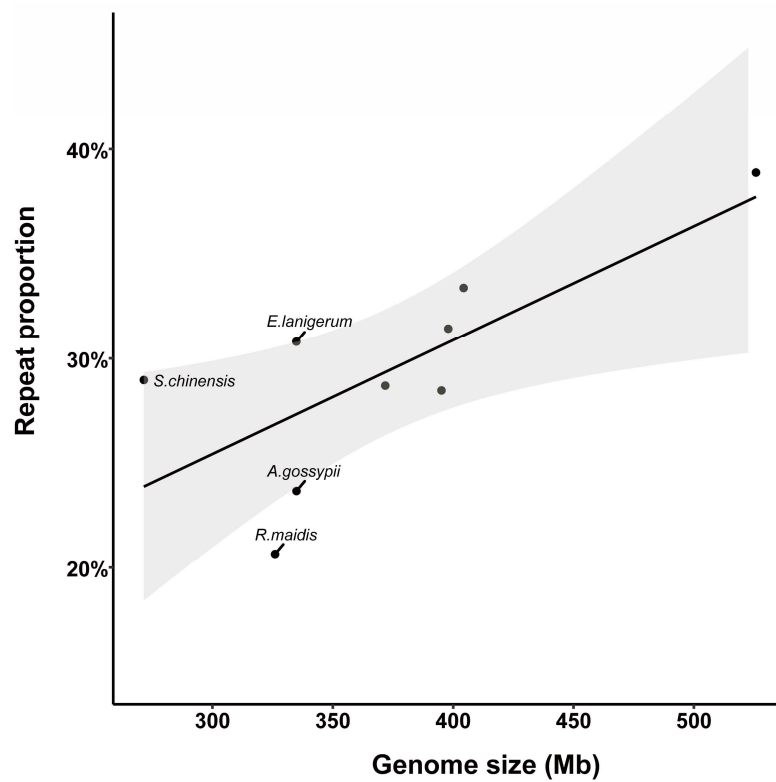
Supplementary Figure 4. The chromosomal Synteny of *E. lanigerum*, *S. chinensis* and other insects in Hemiptera. Synteny diagrams are generated based on the order of genes, with different chromosomes encoded with different colors of connecting lines, while the X chromosome of each species is positioned on the far left and represented in gray. Bars in *A. cooleyi* represent the top 10 largest scaffolds.



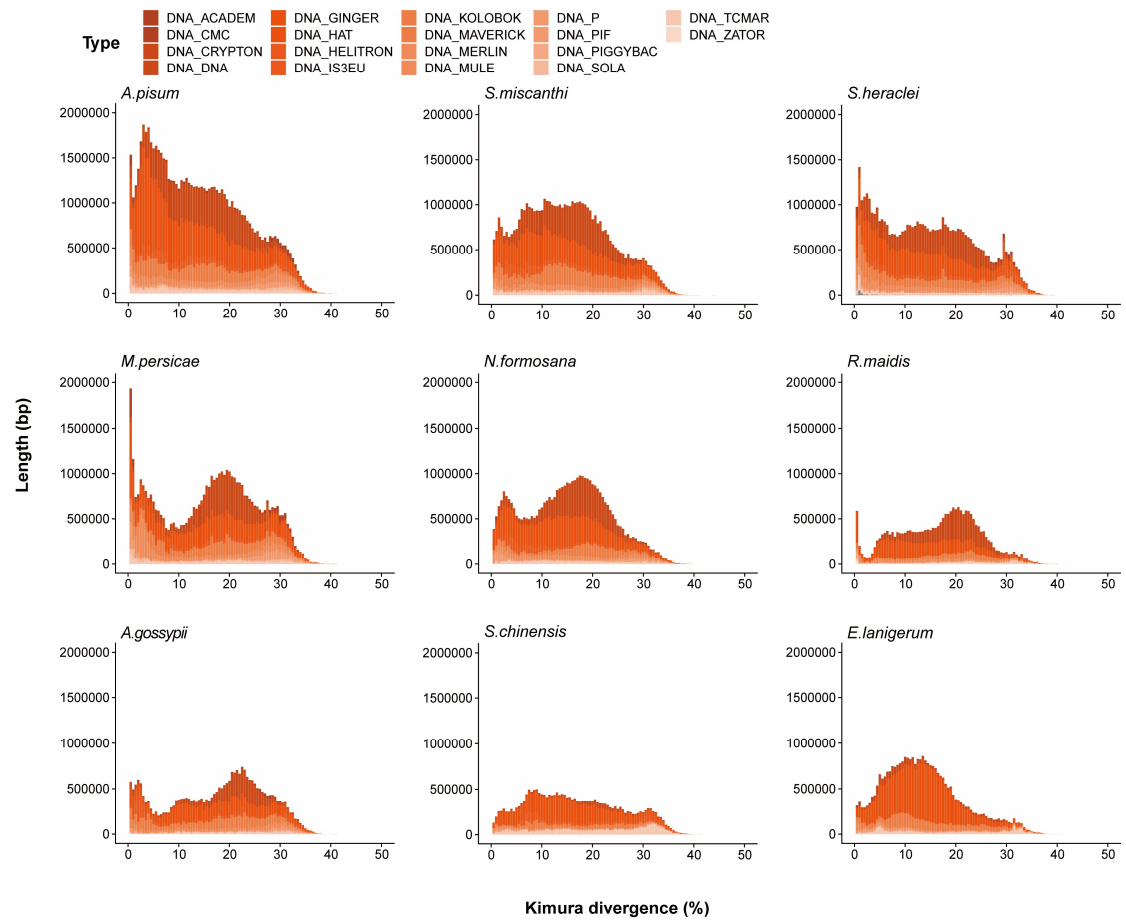
Supplementary Figure 5. Distribution of repetitive sequences on the chromosomes of *S.heraclei* (SH1, SH2, SH3, and SH4) and *A.pisum* (AP1, AP2, AP3, and AP4). The Y-axis represents the proportion of repetitive sequences in 100 kb windows, while the X-axis represents the starting position of the windows. The blue curve represents the fitted curve. SH2 and AP1 are X chromosomes.



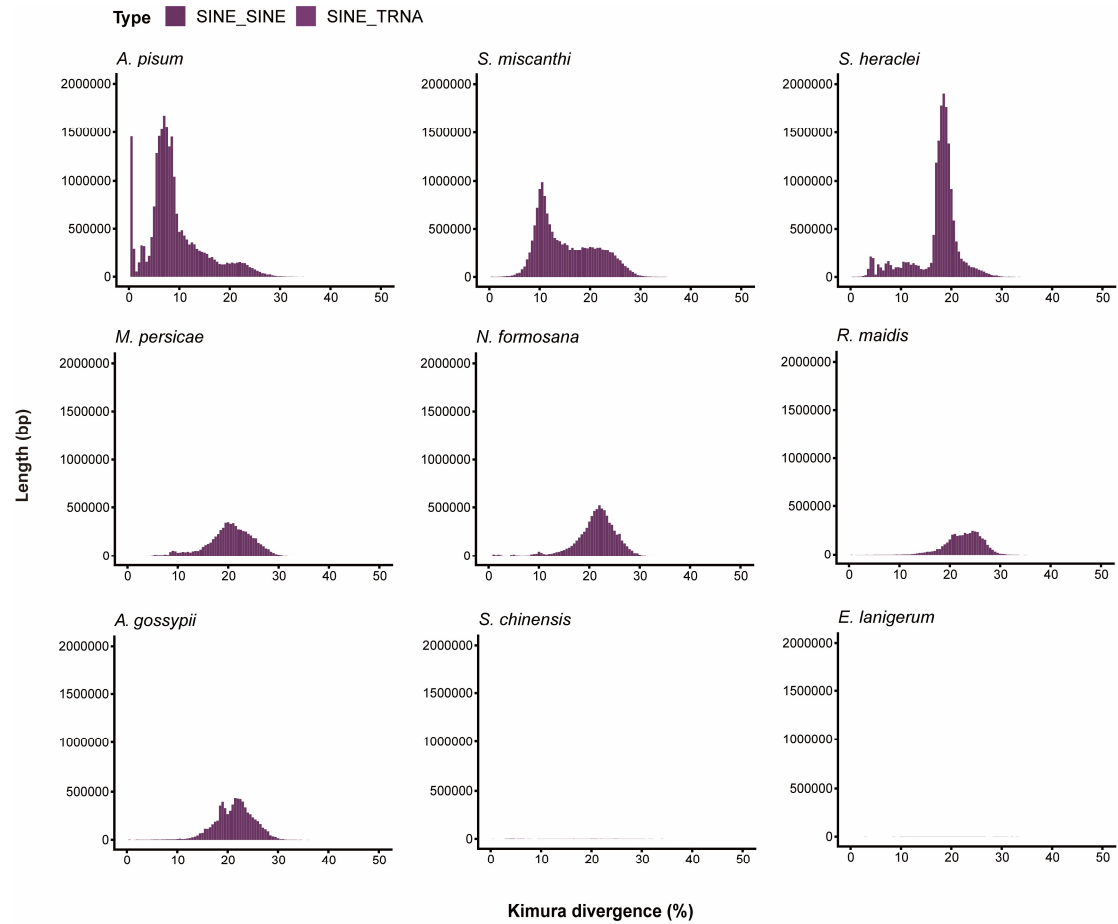
Supplementary Figure 6. Repetitive sequence content of chromosomes in the 9 aphid genomes. The Y-axis represents the proportion of repetitive sequences in the chromosomes, and the X-axis represents the ID of chromosomes. The X chromosomes are marked with red circles.



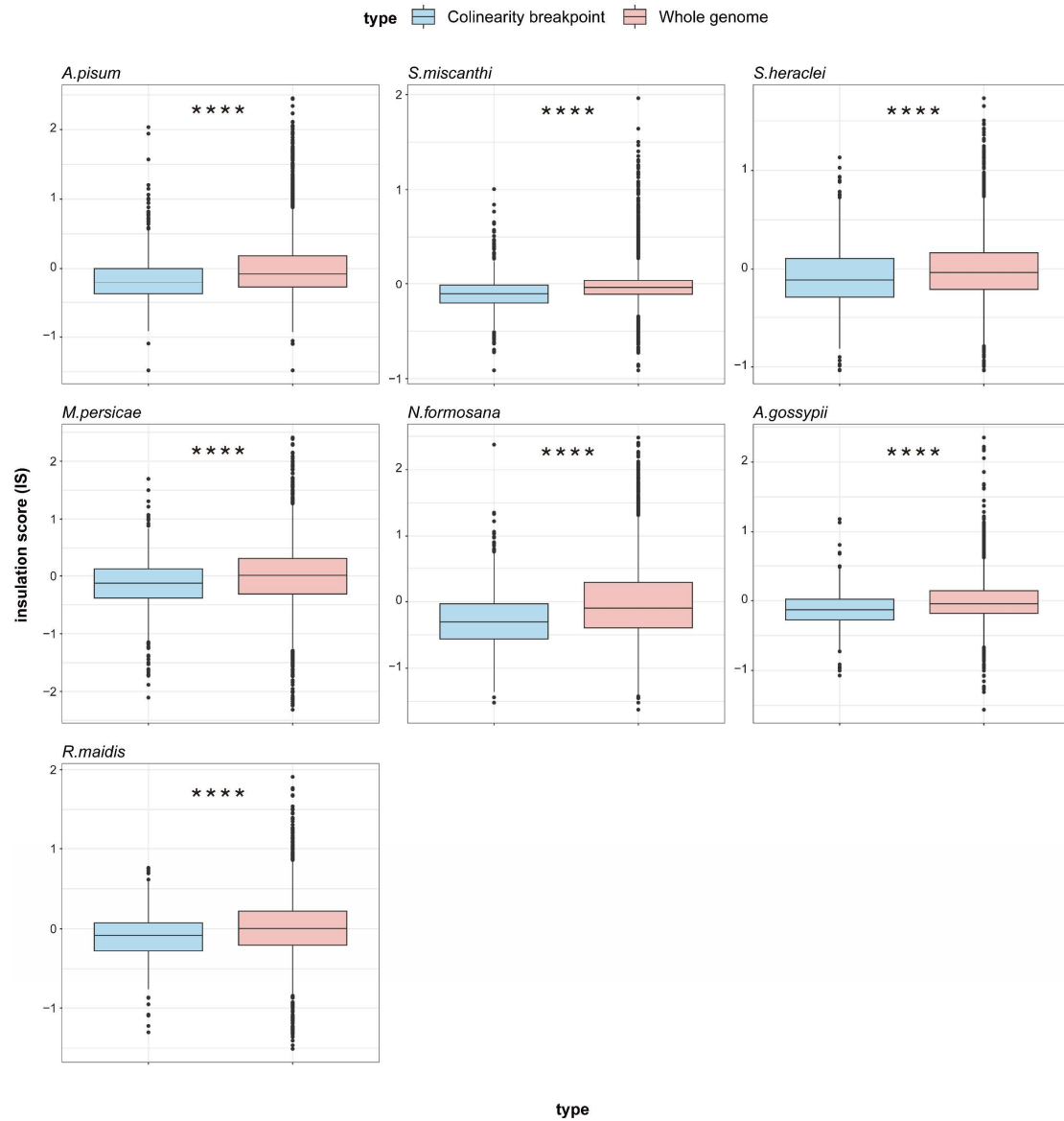
Supplementary Figure 7. Correlation between repetitive sequence content and genome size. Dots represent species, and the correlation analysis reveals a positive correlation between repetitive sequence content and genome assembly size in the seven aphid genomes from Aphidinae (Pearson correlation coefficient, $r = 0.74$, $p\text{-value} = 0.02334$). Two species from Aphidini (*R.maidis* and *A.gossypii*) and two species from Eriosomatinae (*E.lanigerum* and *S.chinensis*) are highlighted in the graph.



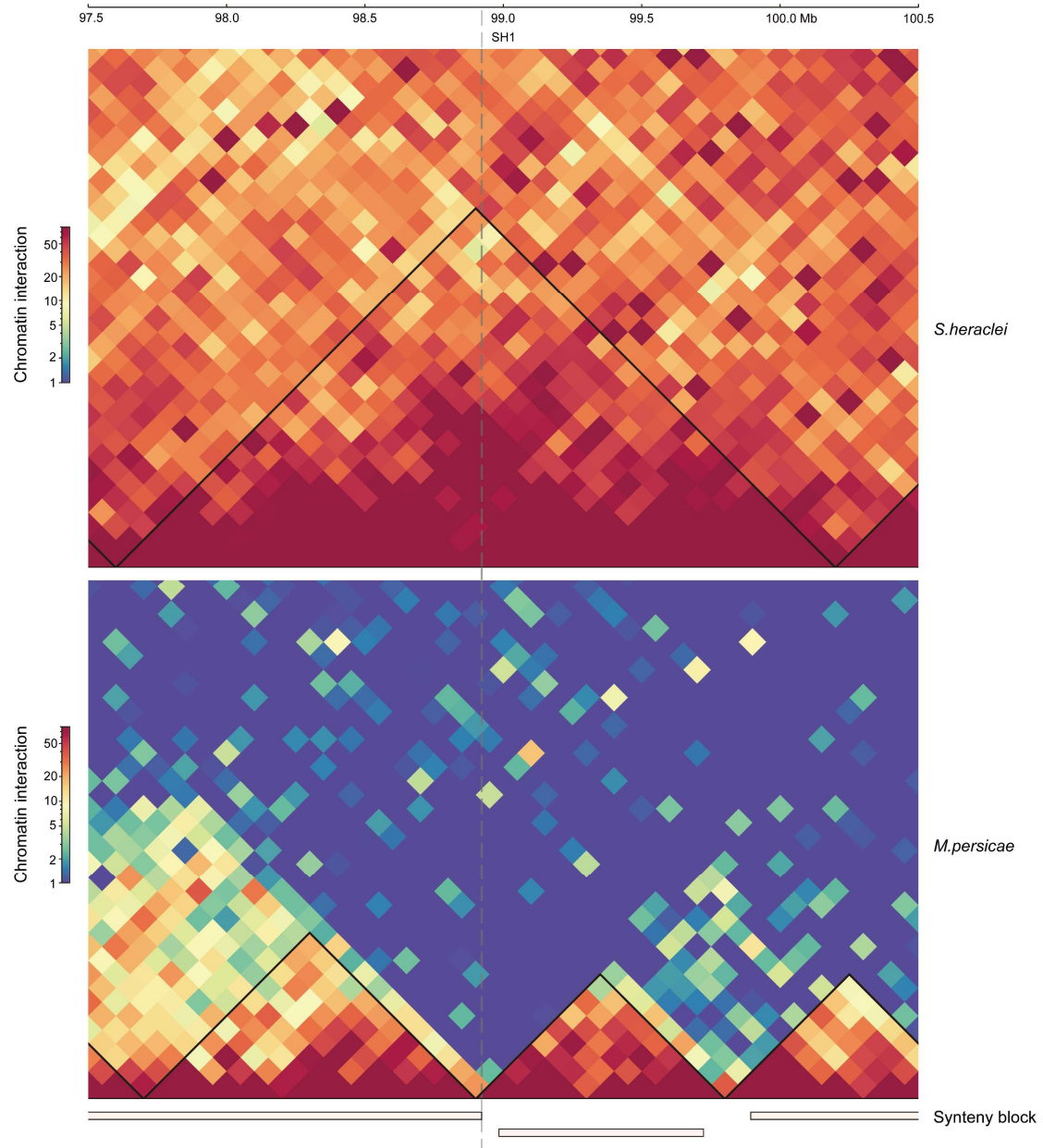
Supplementary Figure 8. Age distribution of DNA transposons in the genomes of the 9 aphid species. The dashed gray and black boxes represent two DNA transposon expansion events shared by seven aphids in Aphidinae.



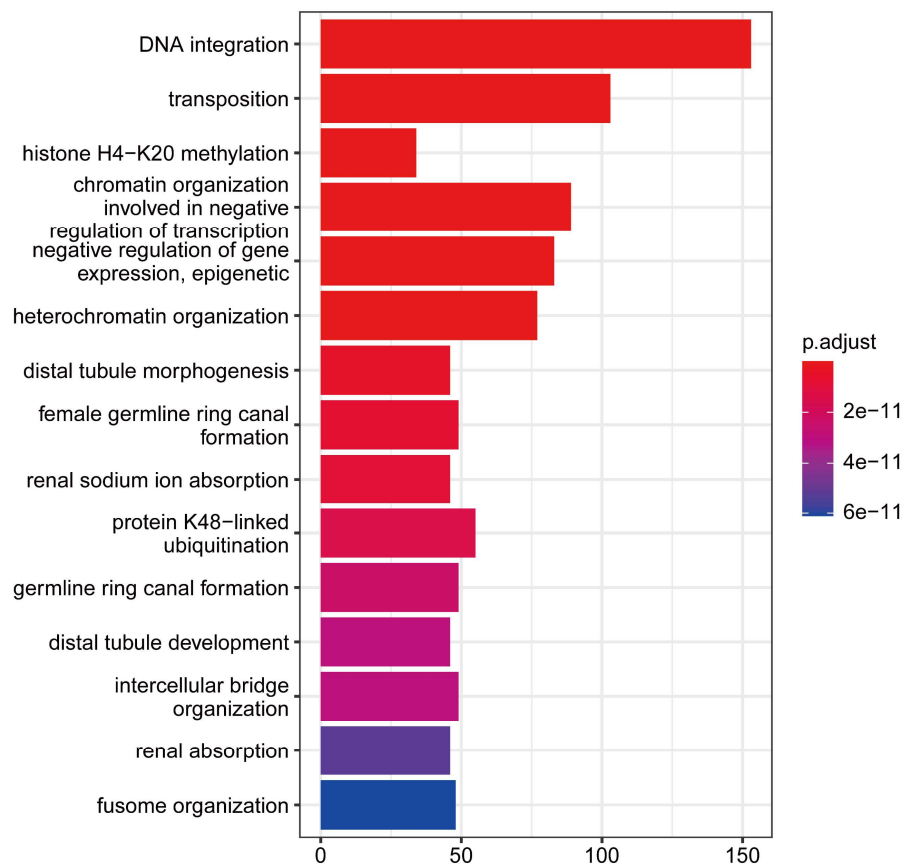
Supplementary Figure 9. Age distribution of SINE retrotransposons in the genomes of the 9 aphid species.



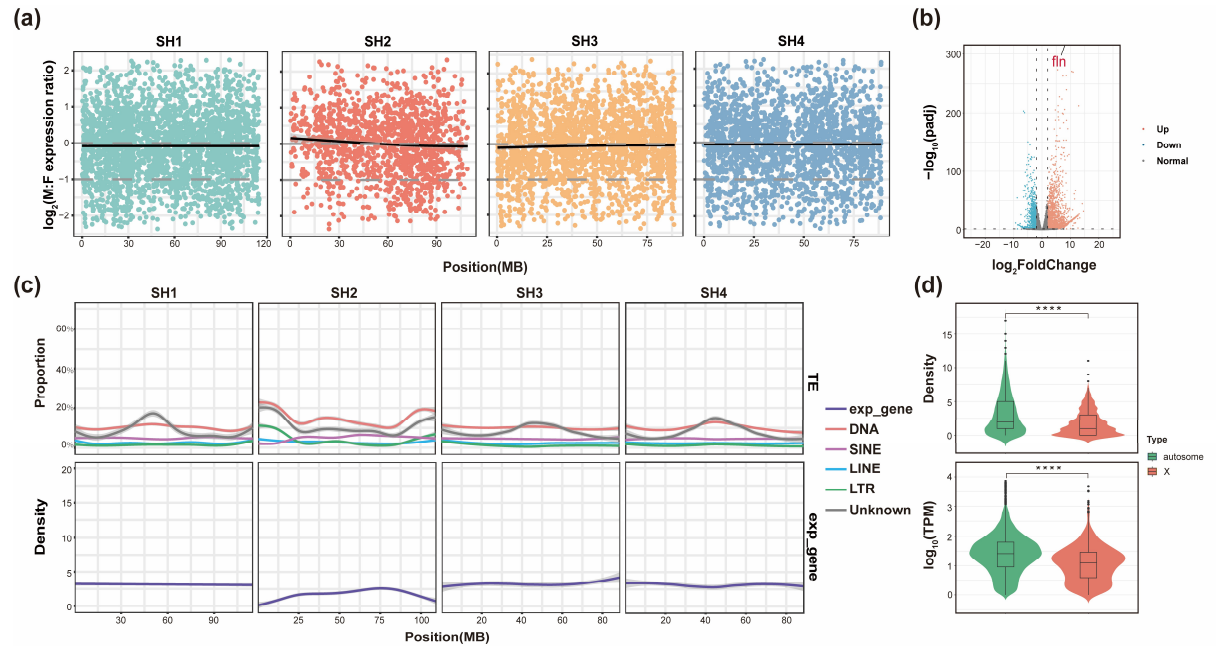
Supplementary Figure 10. Comparison of insulation scores between synteny breakpoint regions and the whole genome in the genomes of the seven aphids in Aphidinae (Mann-Whitney U test, **** = $p\text{-value} < 2.2 \times 10^{-16}$).



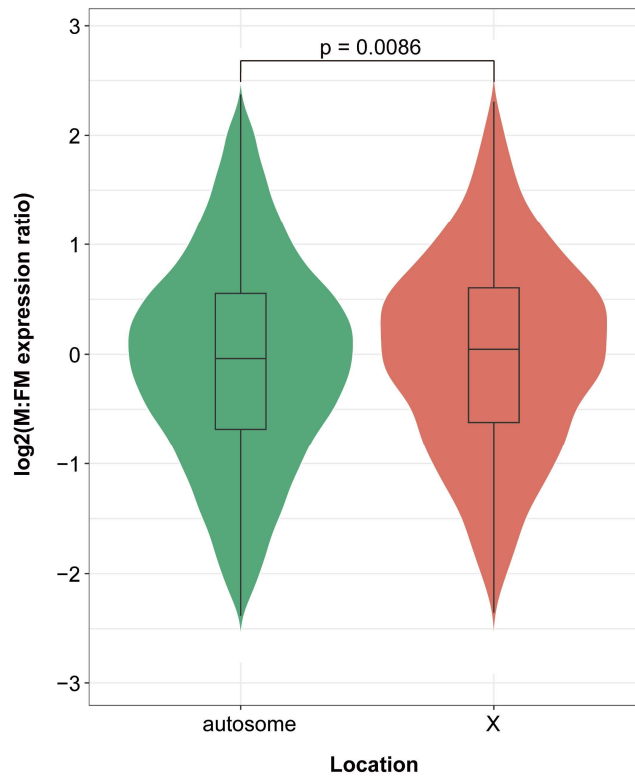
Supplementary Figure 11. Chromosomal fusion and rearrangement-induced changes in TAD structure. From top to bottom, they represent: TAD of *S.heraclei*, TAD of *M.persicae*, chromosomal syntenic blocks of *S.heraclei* and *M.persicae*. The gray dashed lines represent chromosomal fusion sites, and the black lines depict TAD domains.



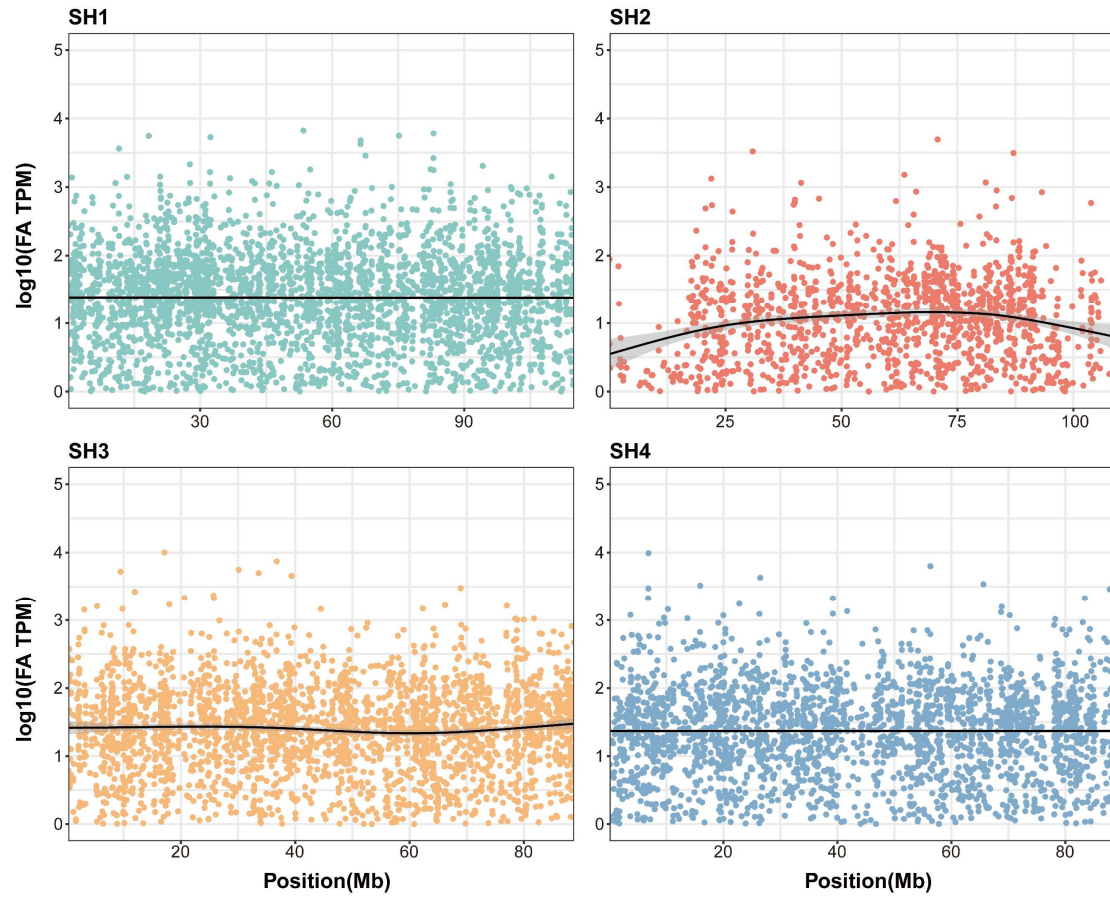
Supplementary Figure 12. GO functional enrichment bar chart of genes in the non-synteny region of the X chromosome in *A. pisum*. The X-axis represents the number of genes.



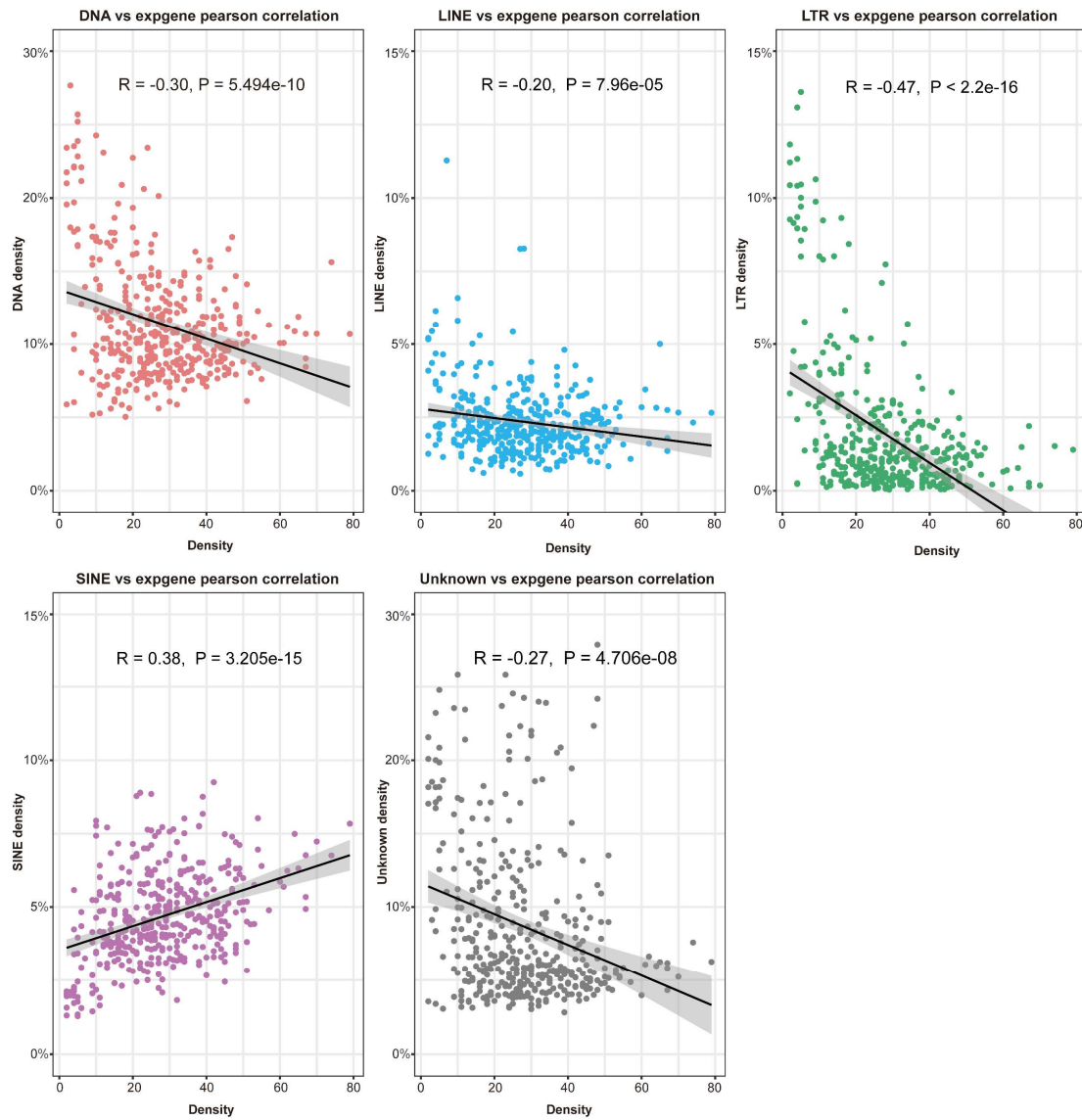
Supplementary Figure 13. X chromosome gene expression patterns. (a) Scatterplot depicting the distribution of $\log_2$ -transformed male-to-female gene expression ratios in *S. heraclei* along the chromosomes for genes with an average expression of at least 1 TPM in both males and females. Each point represents a gene, with a black line indicating the LOESS smoothed average. The gray dashed lines represent the expected range of male-to-female expression ratio under complete dosage compensation ($\log_2[\text{M:F expression ratio}] = 0$) and without dosage compensation ($\log_2[\text{M:F expression ratio}] = -1$). (b) Volcano plot illustrating the differential gene expression in male and female *S. heraclei* aphids. Each point represents a gene, with blue indicating genes downregulated in males and red indicating genes upregulated in males. The *fln* gene is highlighted in red. (c) Distribution of gene expression on the chromosomes and TE content in *S. heraclei*. Genes with an average TPM > 1 in all male and female samples were considered as expressed genes. Lines depict LOESS smoothed average curves for the percentage of various TE classes and the number of expressed genes in 100 kb windows. (d) Violin plot comparing gene expression density and expression level on *S. heraclei* autosomes and X chromosome (Mann-Whitney U test, $p\text{-value} < 2.2 \times 10^{-16}$) ($n = 3$).



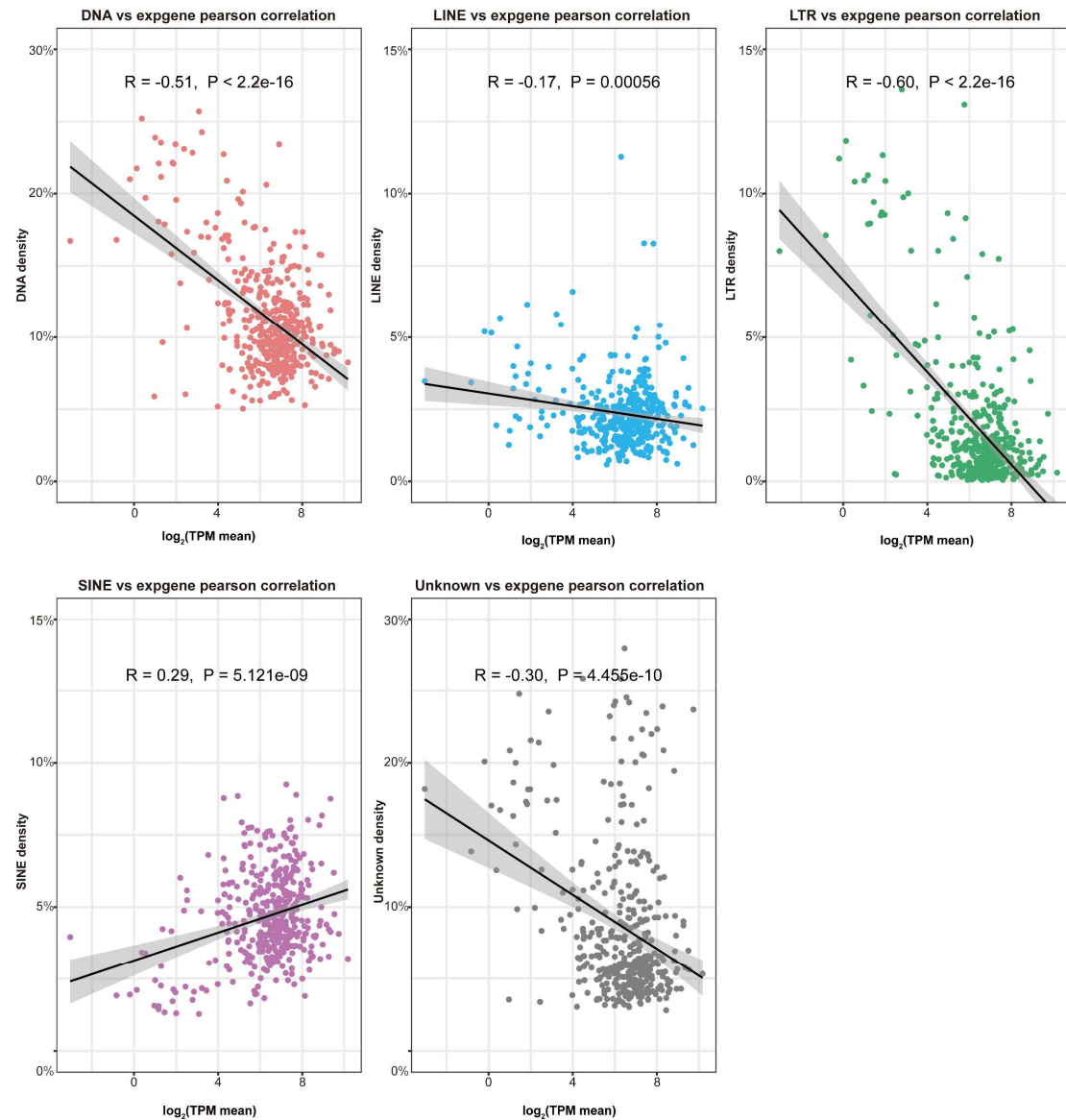
Supplementary Figure 14. Comparison of $\log_2$ values of the ratio of expression levels between male and female genes on autosomes and the X chromosome in *S. heraclei*. The violin plots depict the distribution of $\log_2$ M to FA gene expression ratios for autosomes (autosome) and the X chromosome (X) for genes with an average expression of at least 1 TPM in males and females. The horizontal lines inside the colored regions of the boxplots represent the median, and the top and bottom lines represent the interquartile range ($n = 3$).



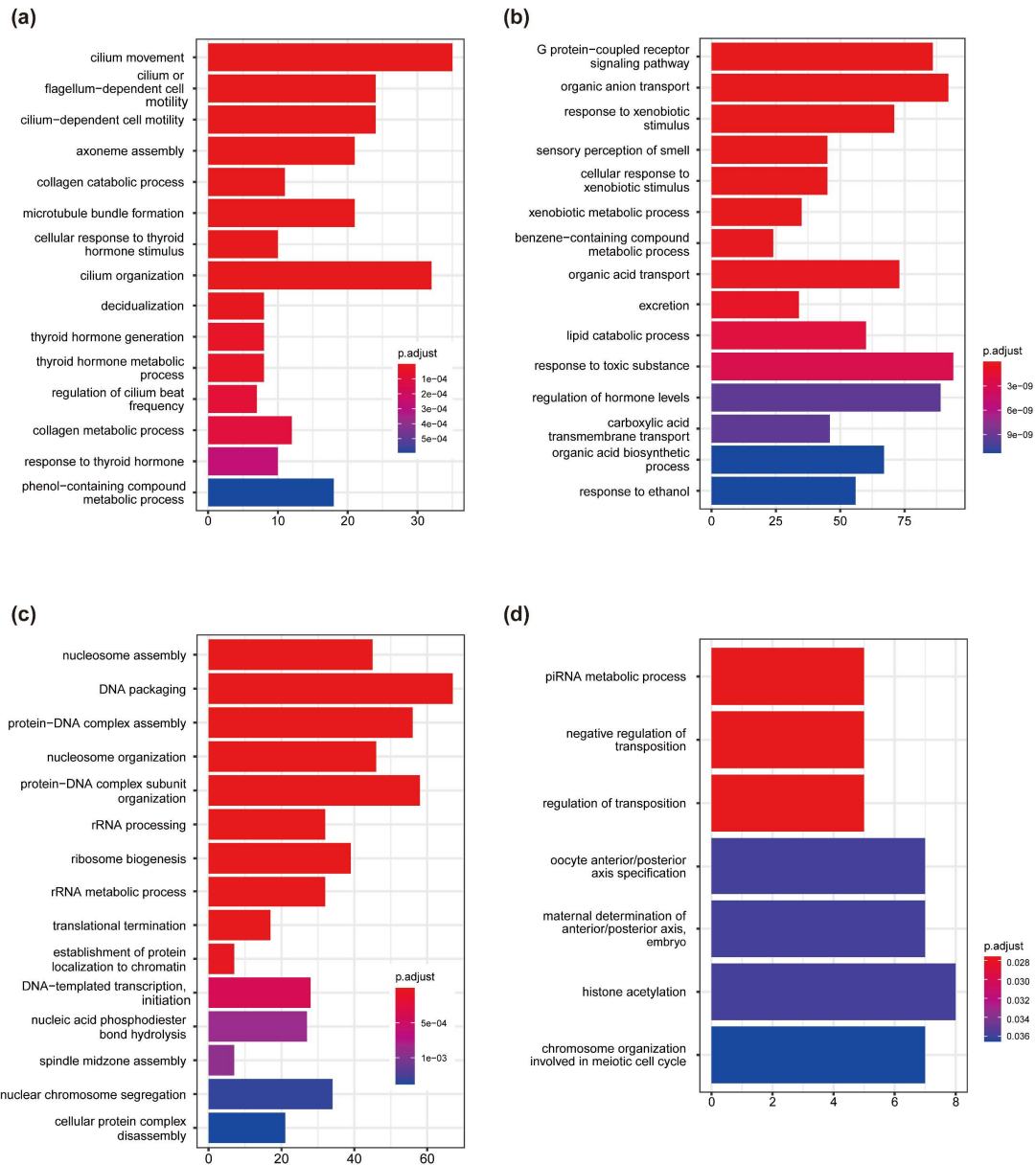
Supplementary Figure 15. Gene expression ratios in each chromosome of *S. heraclei* asexual females. Each point represents a gene, and the black line represents the LOESS smoothed average.



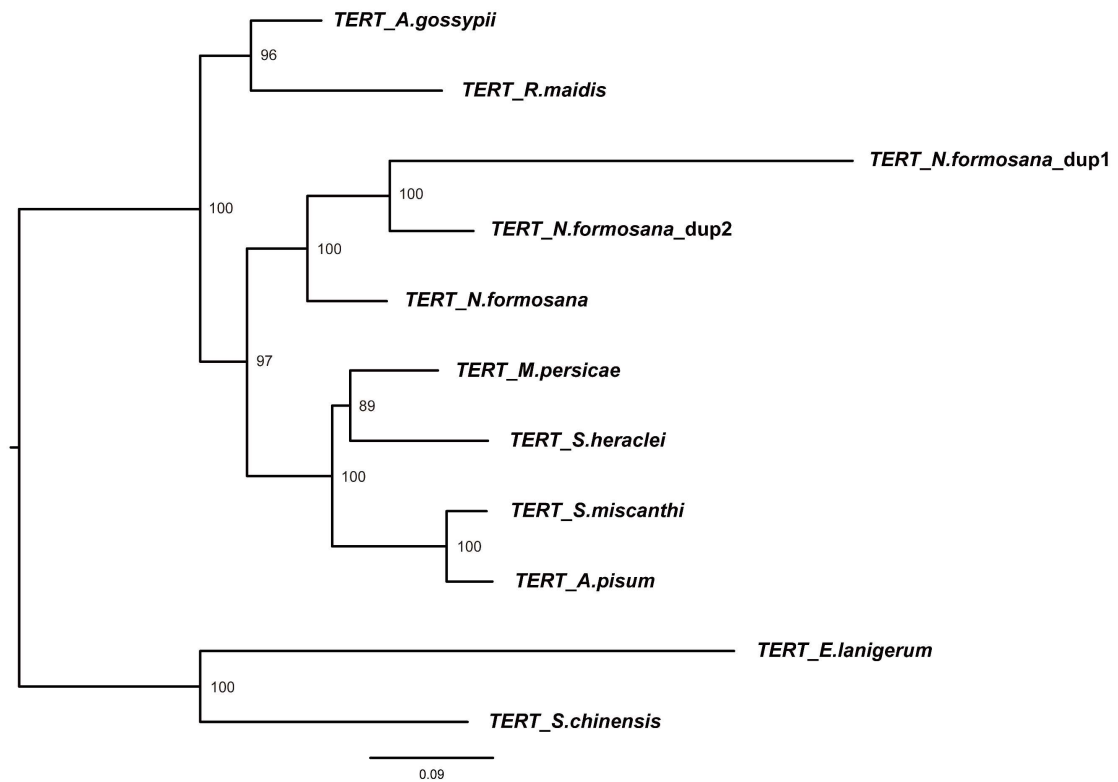
Supplementary Figure 16. Relationship between TE density and expressed gene density in the *S. heraclei* genome. Each point represents the feature density in individual 1 Mb windows of the genome.



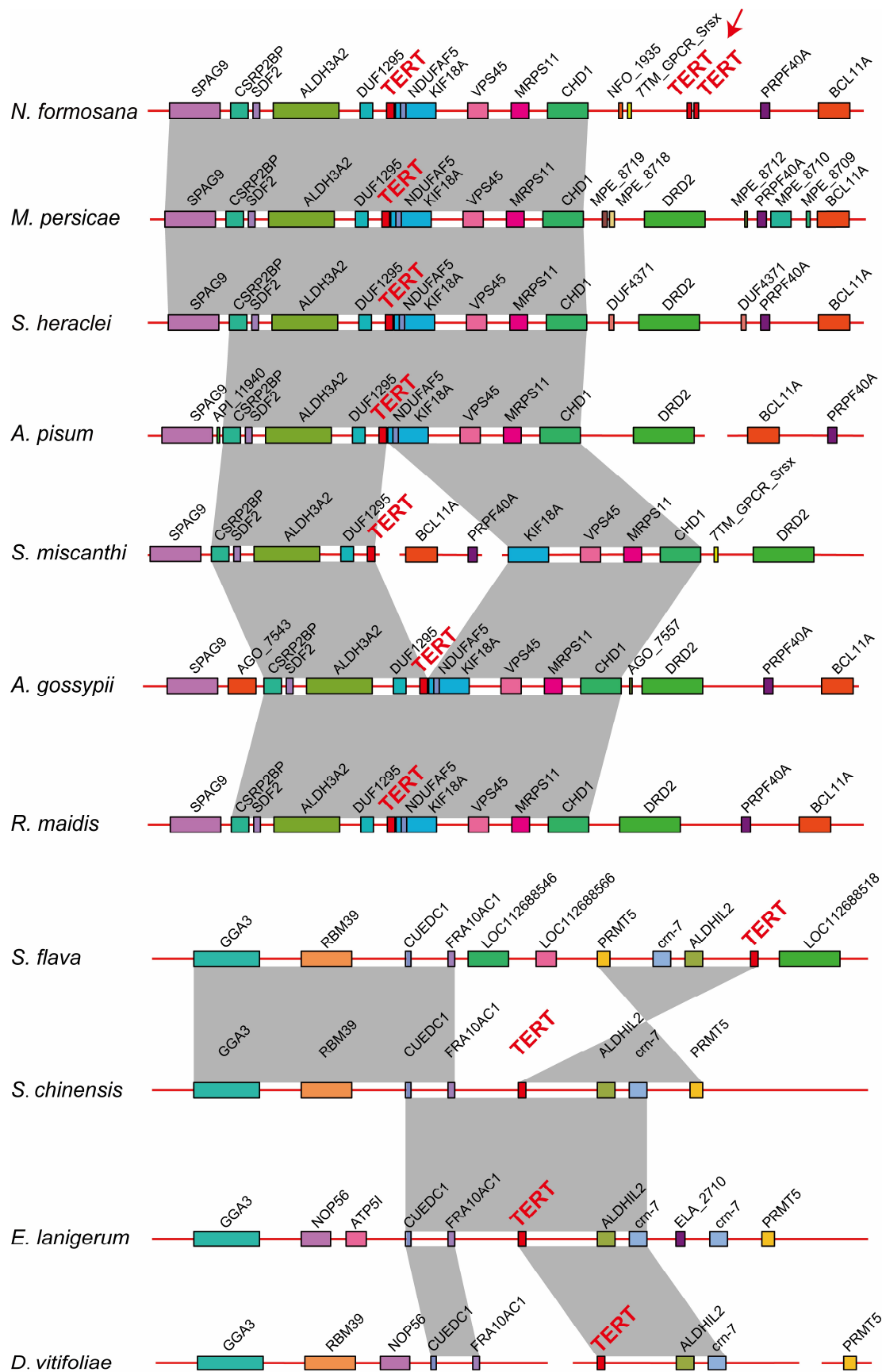
Supplementary Figure 17. Relationship between TE density and gene expression levels in the *S. heraclei* genome. Each point represents the feature density in individual 1 Mb windows of the genome. The X-axis represents the average TPM of all expressed genes in the 1 Mb window.



Supplementary Figure 18. GO enrichment bar charts for upregulated genes on the autosomes and X-chromosomes of female and male aphids in *S. heraclei*. The X-axis represents the number of genes. (a) Upregulated genes on the X-chromosome in males. (b) Upregulated genes on the autosomes in males. (c) Upregulated genes on the autosomes in females. (d) Upregulated genes on the X-chromosome in females.



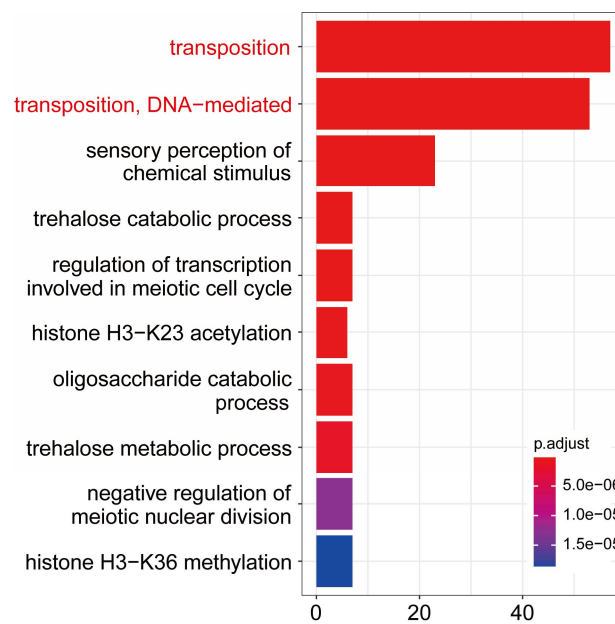
Supplementary Figure 19. Phylogenetic tree of *TERT* genes in the nine aphid species. The phylogenetic tree was constructed using nucleotide sequences and the RAXML v8.2.12 software. *TERT_N.formosana_dup1* and *TERT_N.formosana_dup2* represent two copies of the *TERT* gene in the *N. formosana* genome.



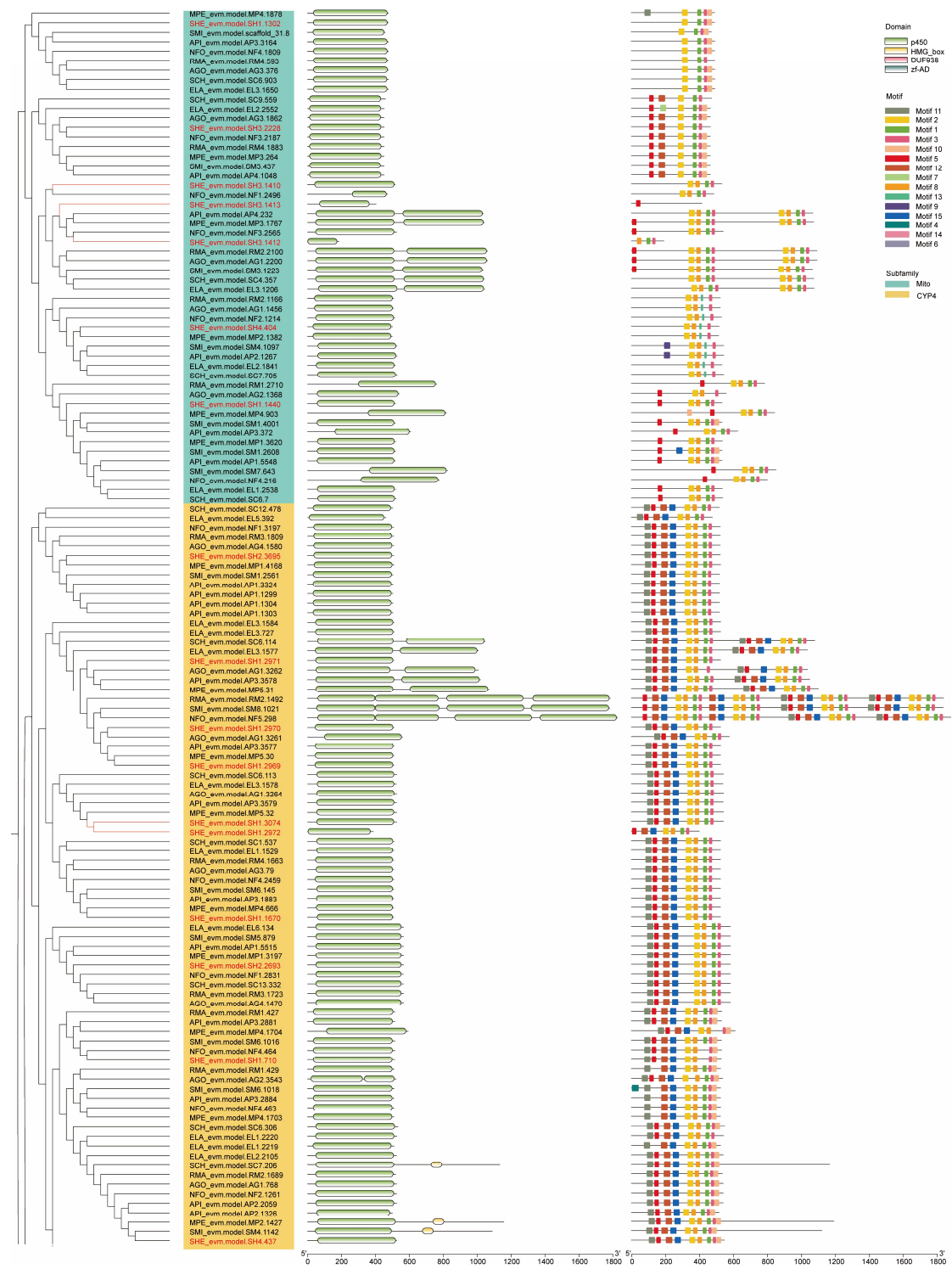
Supplementary Figure 20. Conservation of the genomic region near the *TERT* gene. Gray links represent genomic regions that are conserved in all species, with the *TERT* gene highlighted in red. Red arrows indicate the tandem duplication of the *TERT* gene.



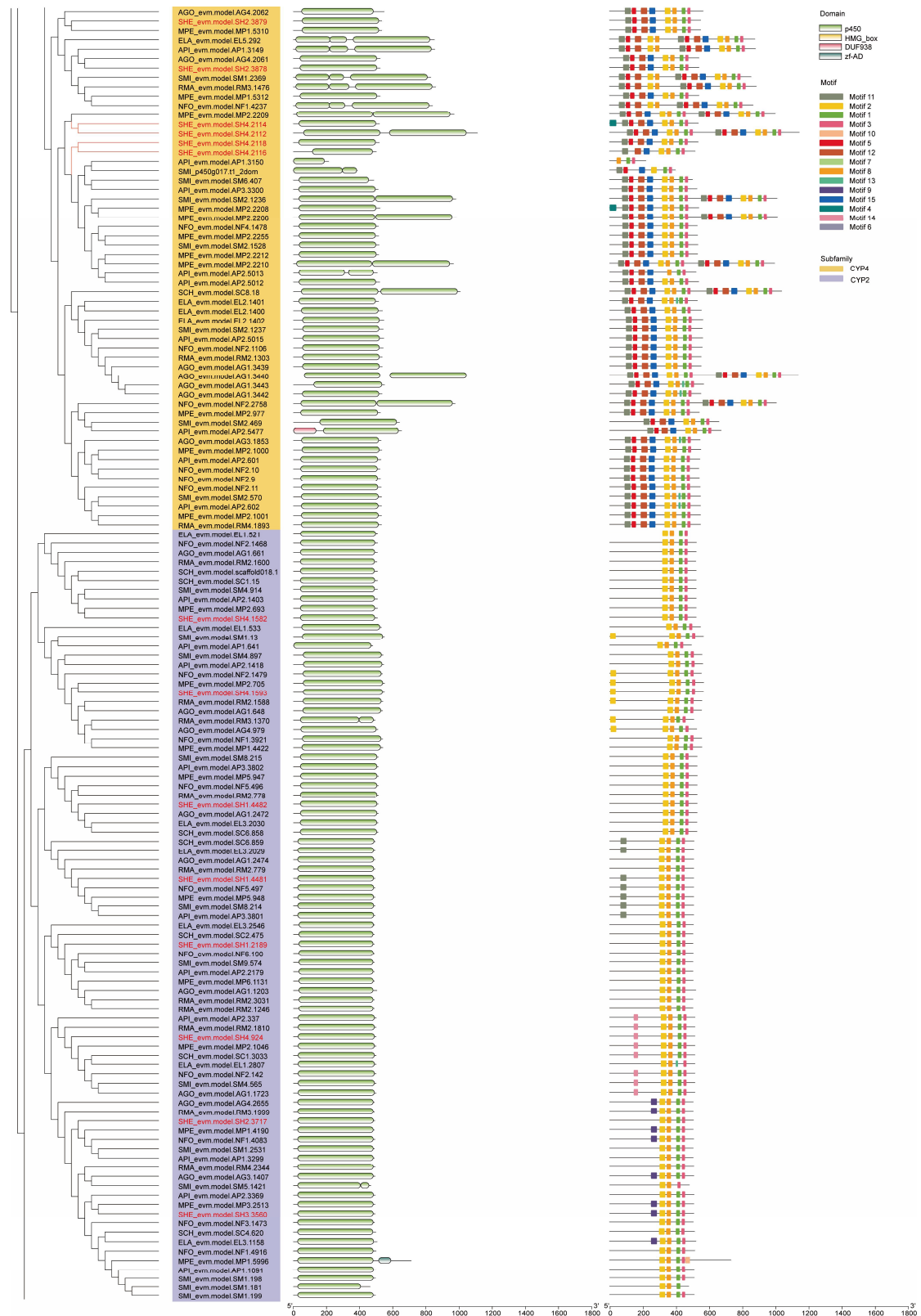
Supplementary Figure 21. Genomic region assembly continuity for the tandem duplication of *TERT* in the *N. formosana* genome. The annotations for the two protein-coding genes are tandem duplicate genes dup1 and dup2 of *TERT*.



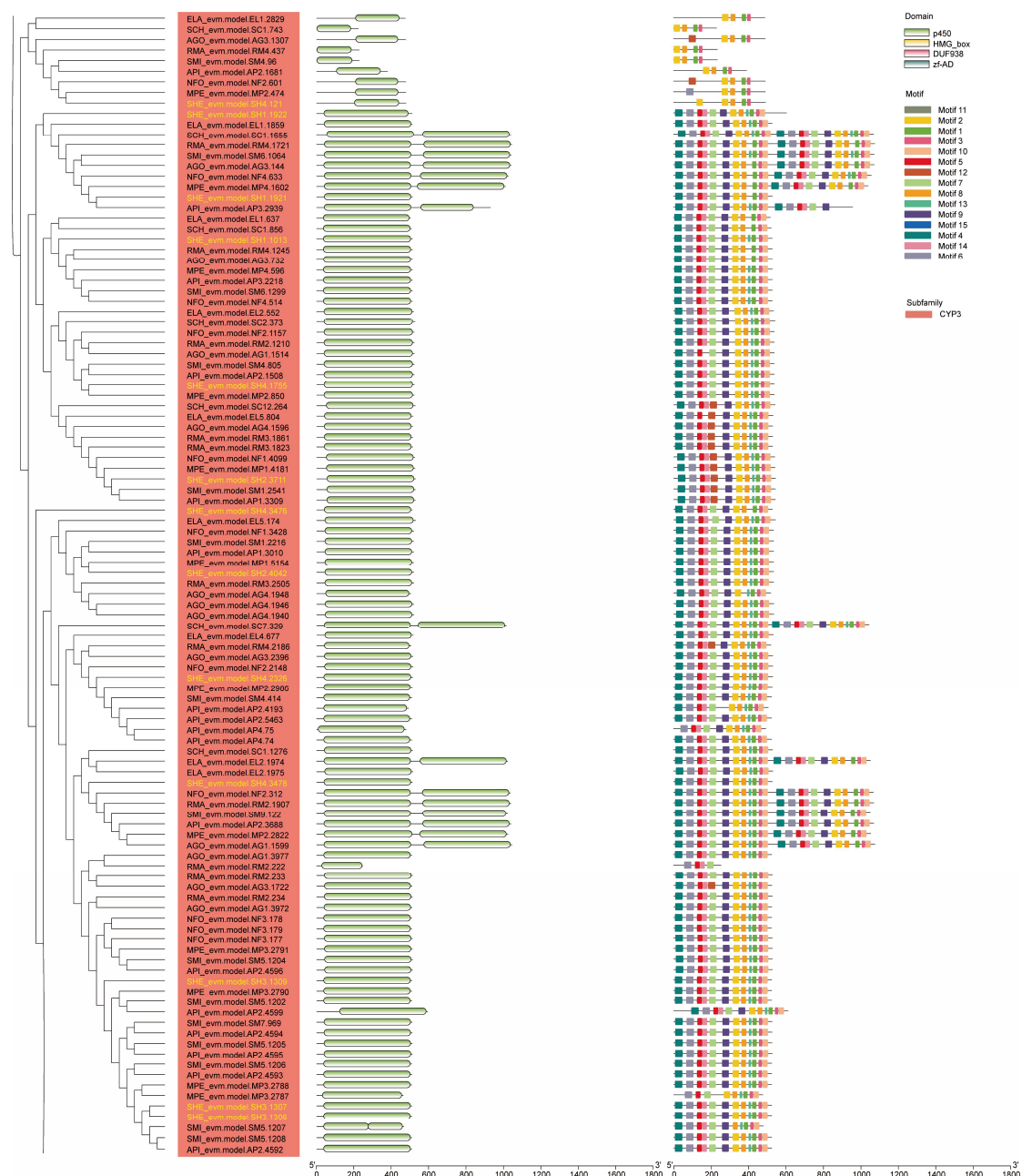
Supplementary Figure 22. GO biological process enrichment results of expanded gene families in five aphids within Macrosiphini.



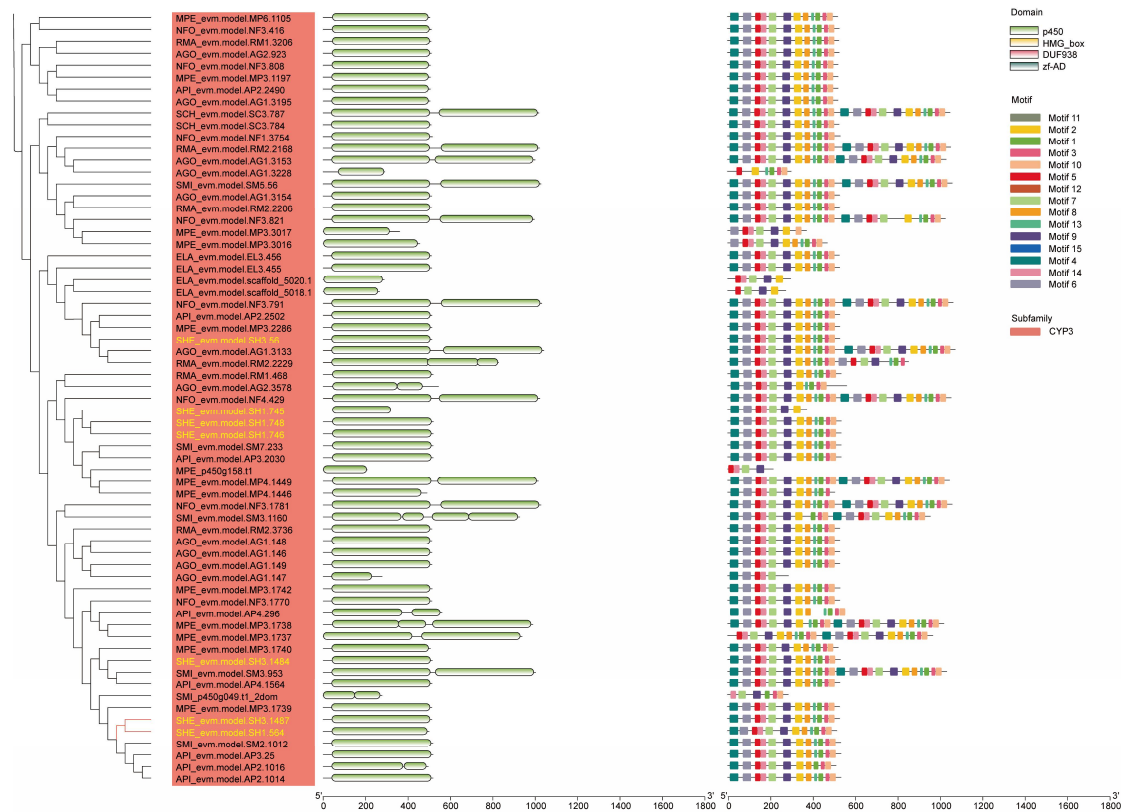
Supplementary Figure 23. Phylogenetic relationships, domain and motif between p450s from different aphids. Highlight the gene names belonging to *Semiaphis heraclei* with distinct colors. The gene branches in *Semiaphis heraclei* that underwent species-specific expansion are highlighted in red.



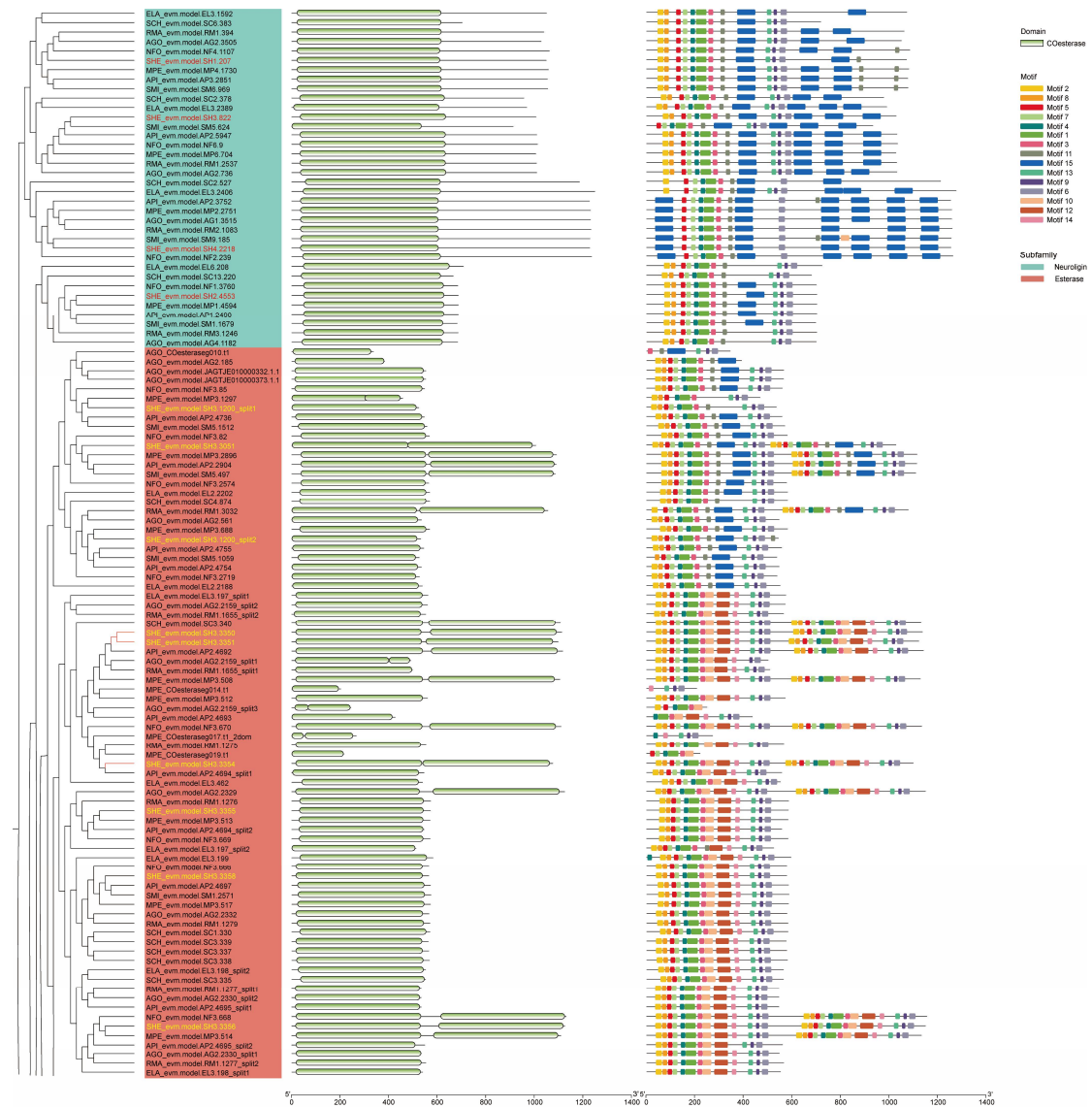
Continued Supplementary Figure 23. Phylogenetic relationships, domain and motif between p450s from different aphids. Highlight the gene names belonging to *Semiaphis heraclei* with distinct colors. The gene branches in *Semiaphis heraclei* that underwent species-specific expansion are highlighted in red.



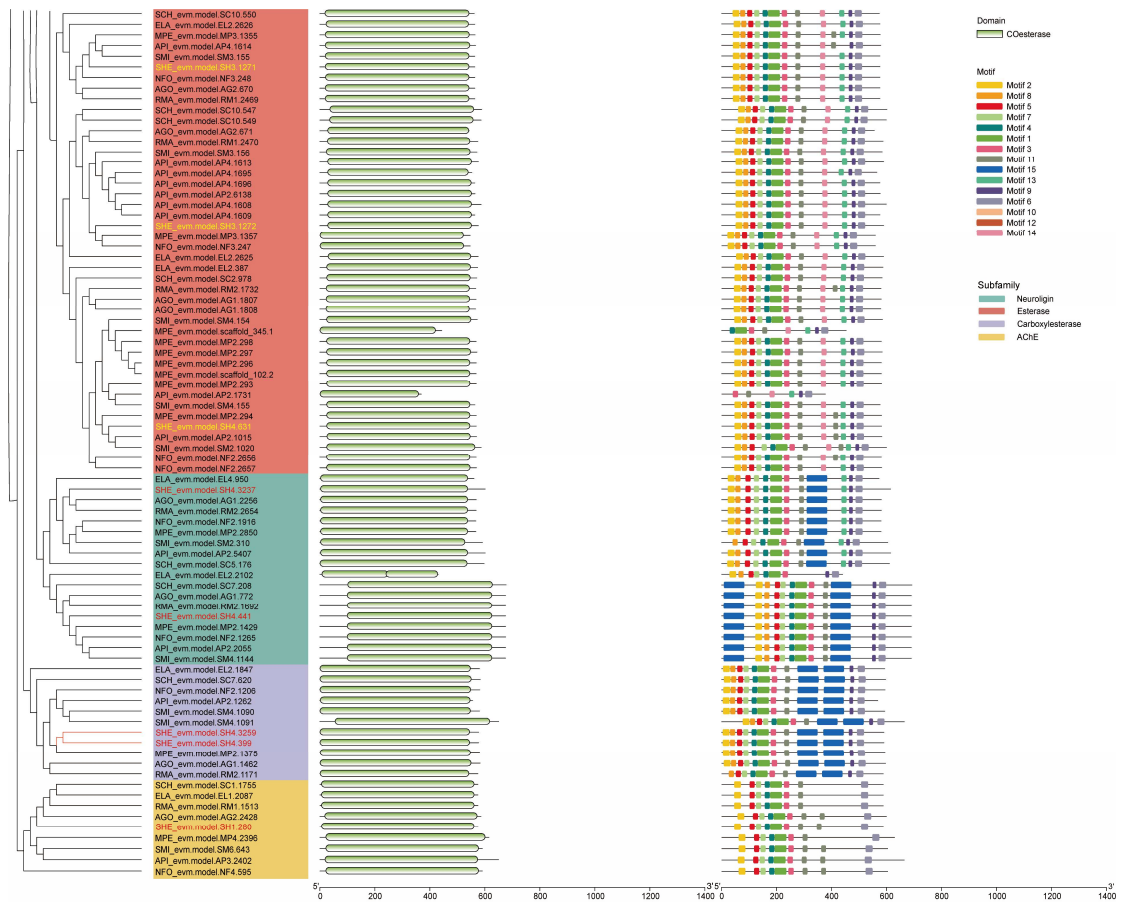
Continued Supplementary Figure 23. Phylogenetic relationships, domain and motif between p450s from different aphids. Highlight the gene names belonging to *Semiaphis heraclei* with distinct colors.

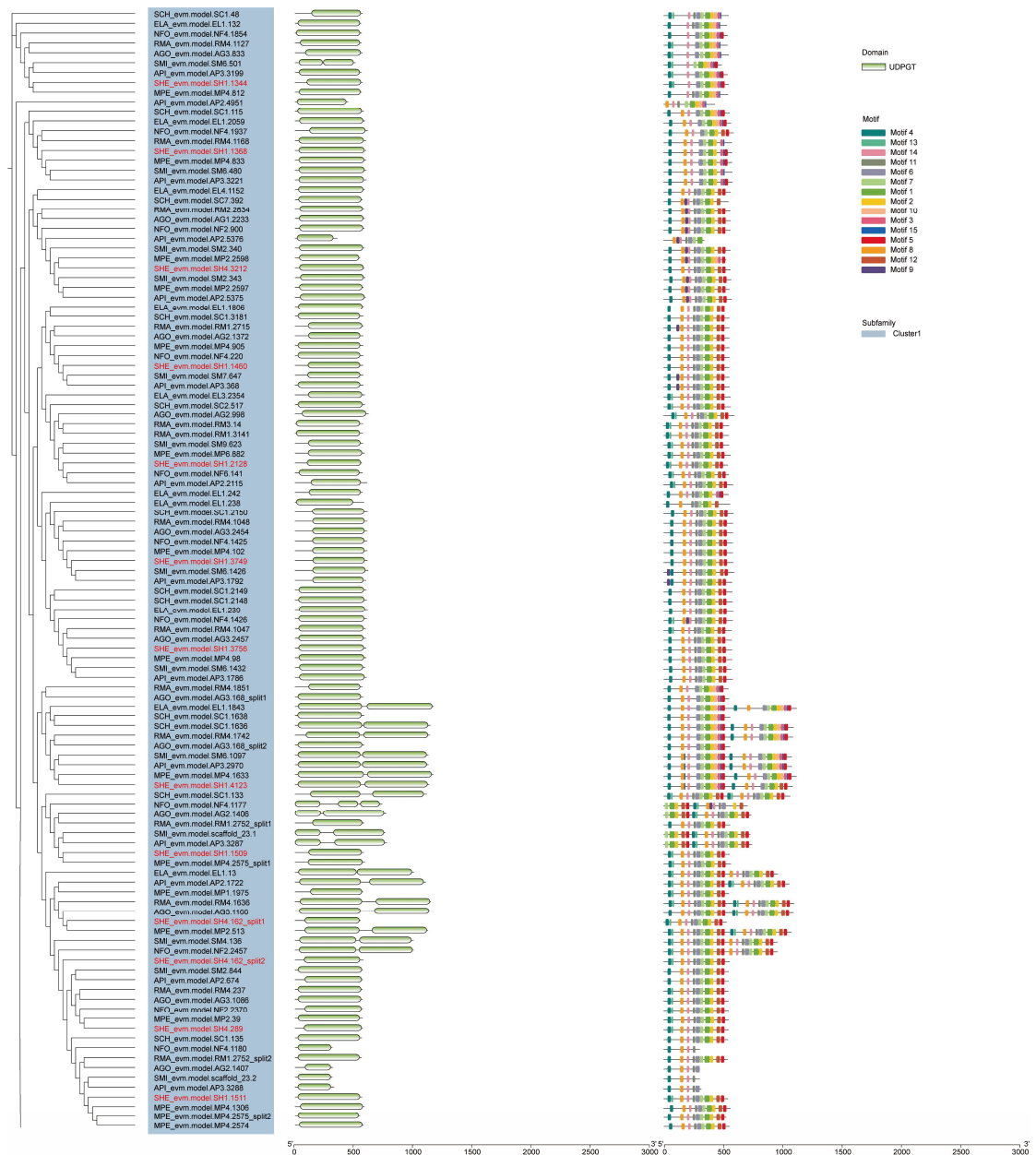


Continued Supplementary Figure 23. Phylogenetic relationships, domain and motif between p450s from different aphids. Highlight the gene names belonging to *Semiaphis heraclei* with distinct colors. The gene branches in *Semiaphis heraclei* that underwent species-specific expansion are highlighted in red.

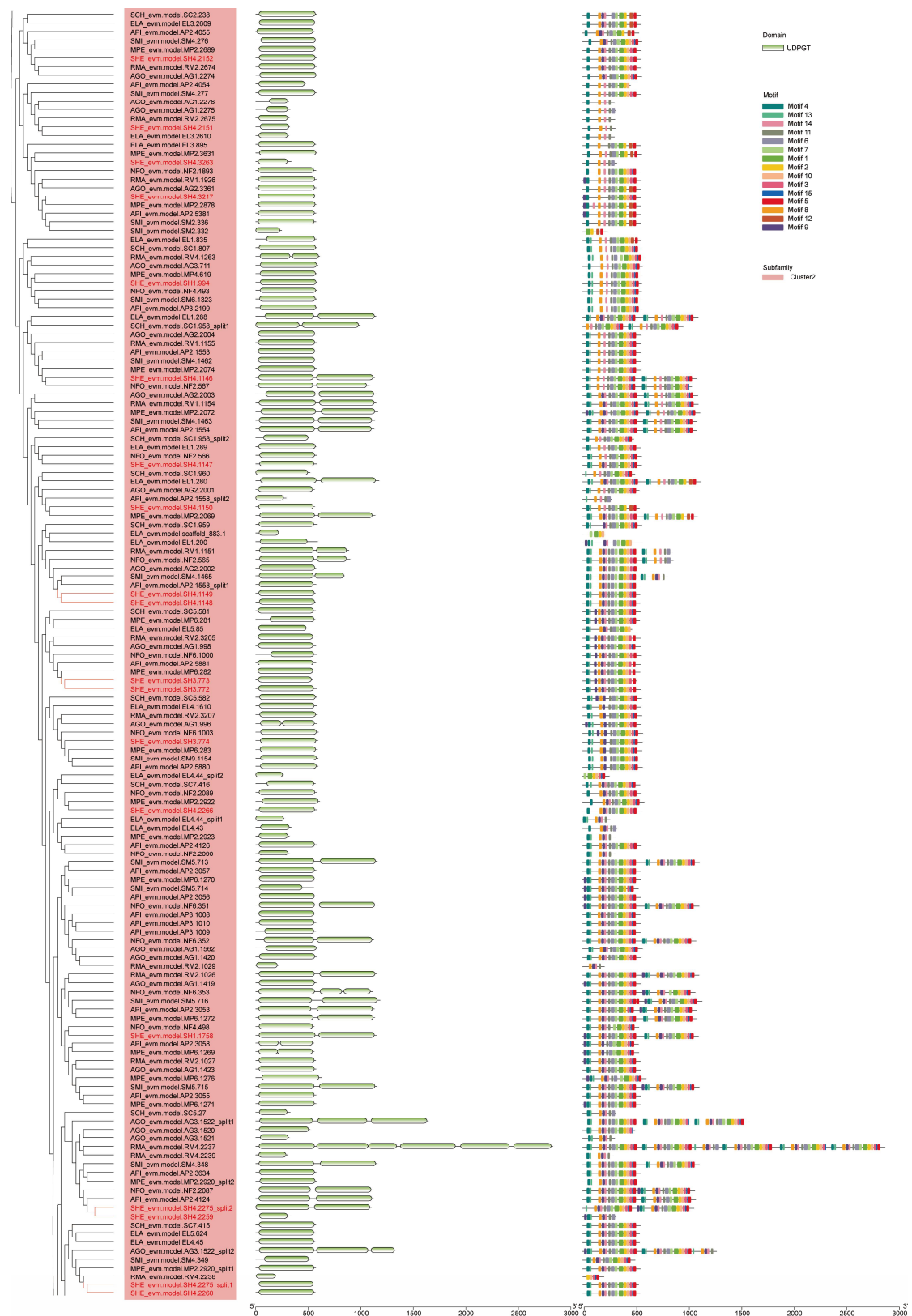


Continued Supplementary Figure 24. Phylogenetic relationships, domain and motif between CCEs from different aphids. Highlight the gene names belonging to *Semiaphis heraclei* with distinct colors. The gene branches in *Semiaphis heraclei* that underwent species-specific expansion are highlighted in red.

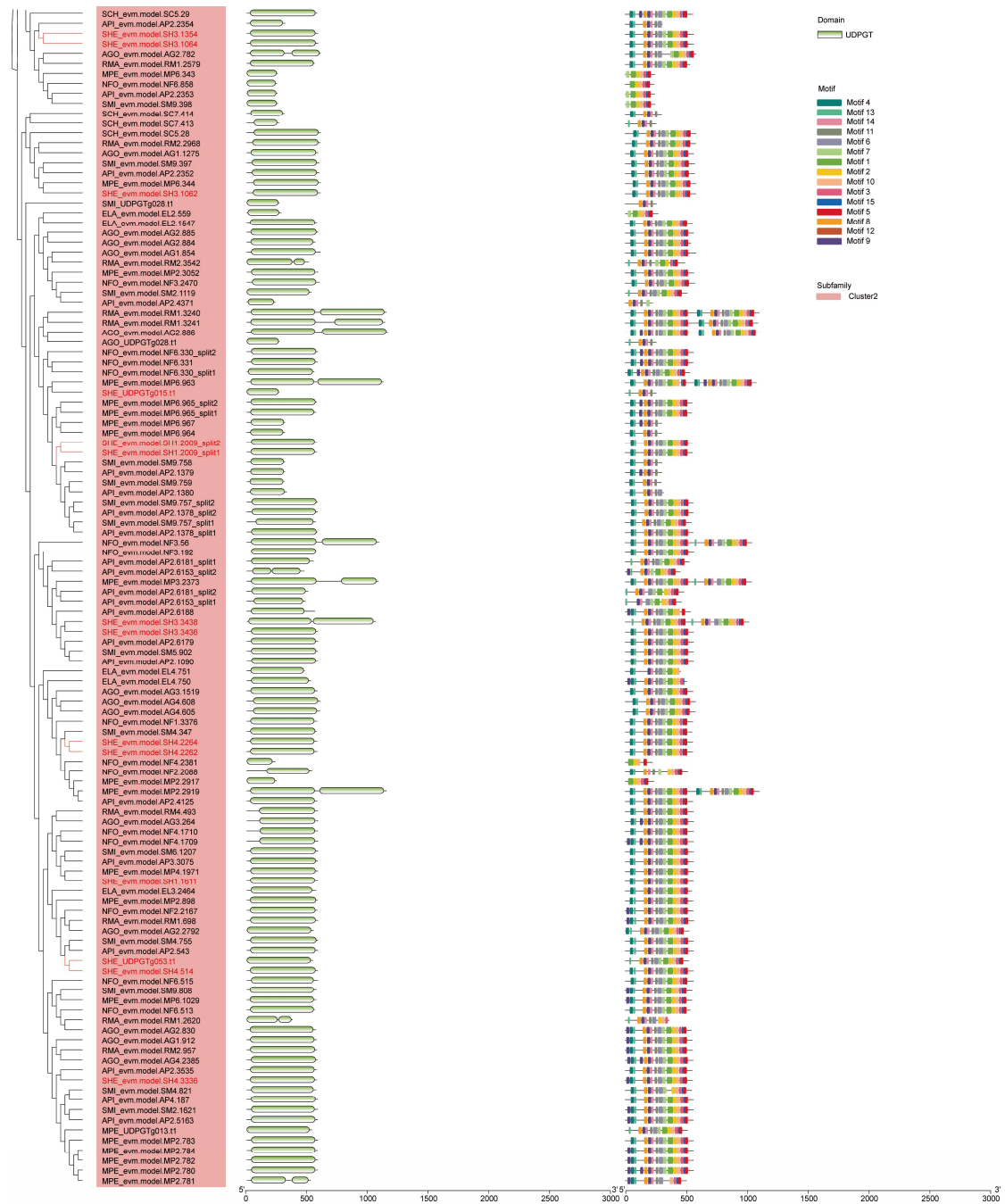




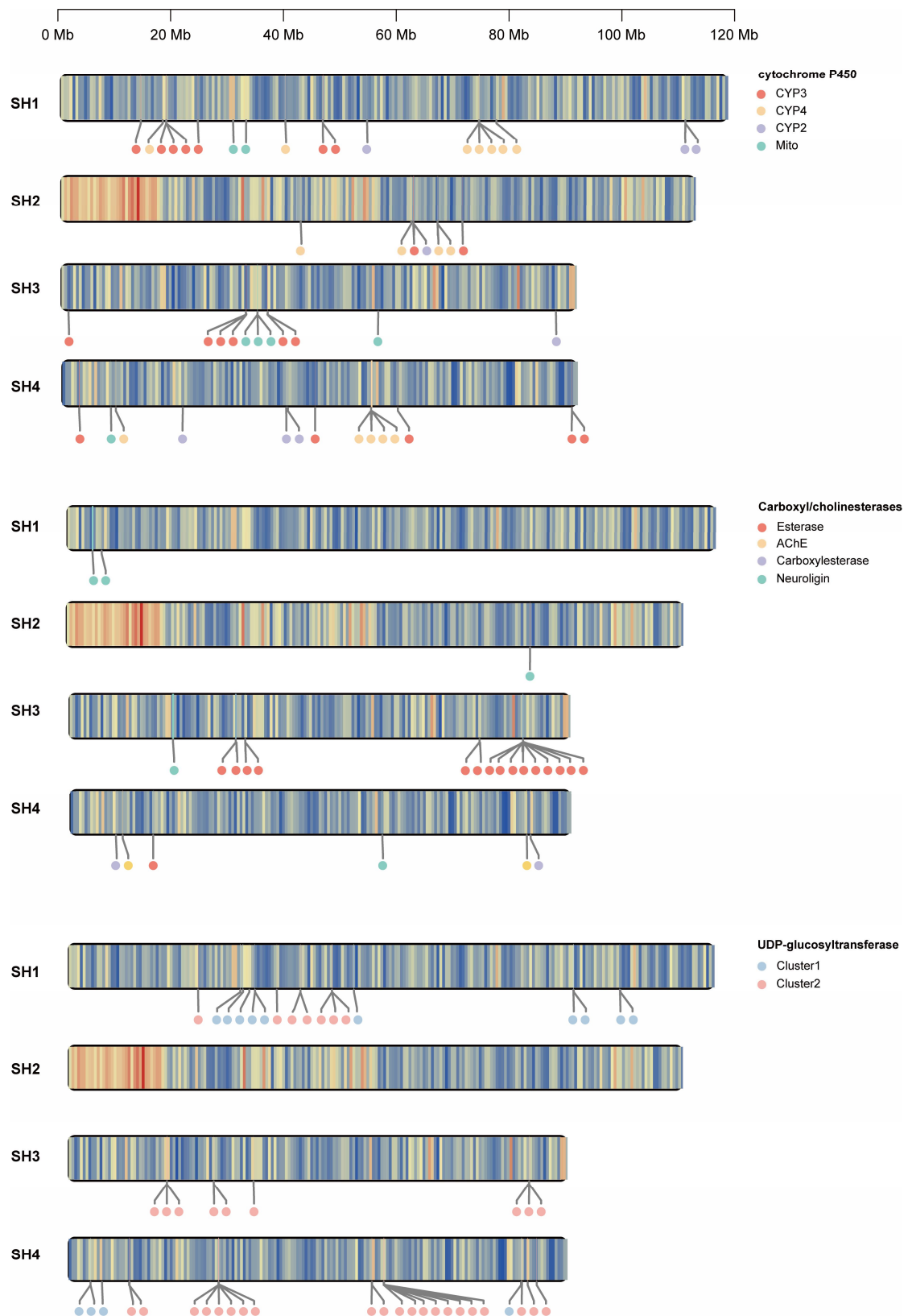
Supplementary Figure 25. Phylogenetic relationships, domain and motif between UGTs from different aphids. Highlight the gene names belonging to *Semiaphis heraclei* with distinct colors.



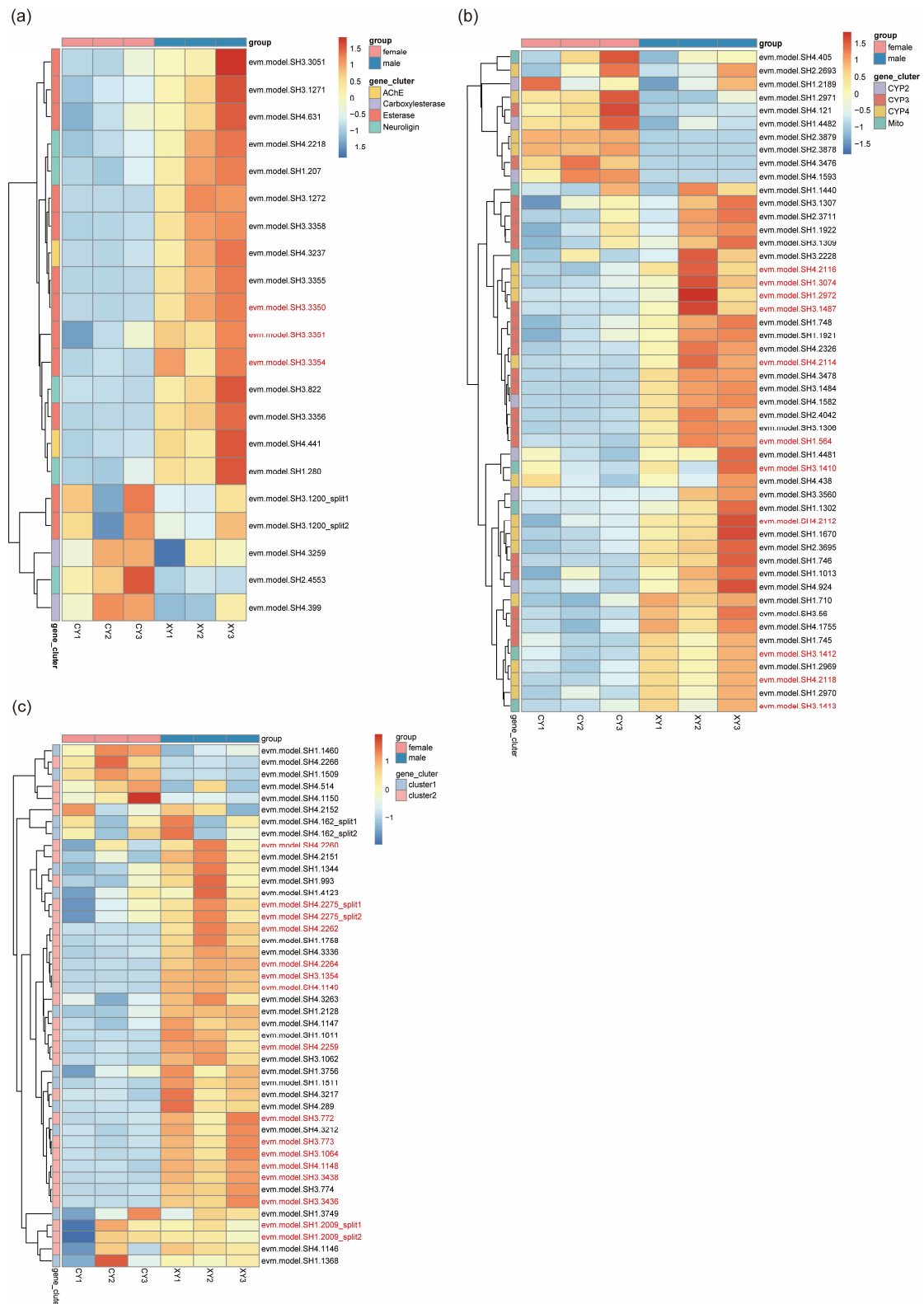
Continued Supplementary Figure 25. Phylogenetic relationships, domain and motif between UGTs from different aphids. Highlight the gene names belonging to *Semiaphis heraclei* with distinct colors. The gene branches in *Semiaphis heraclei* that underwent species-specific expansion are highlighted in red.



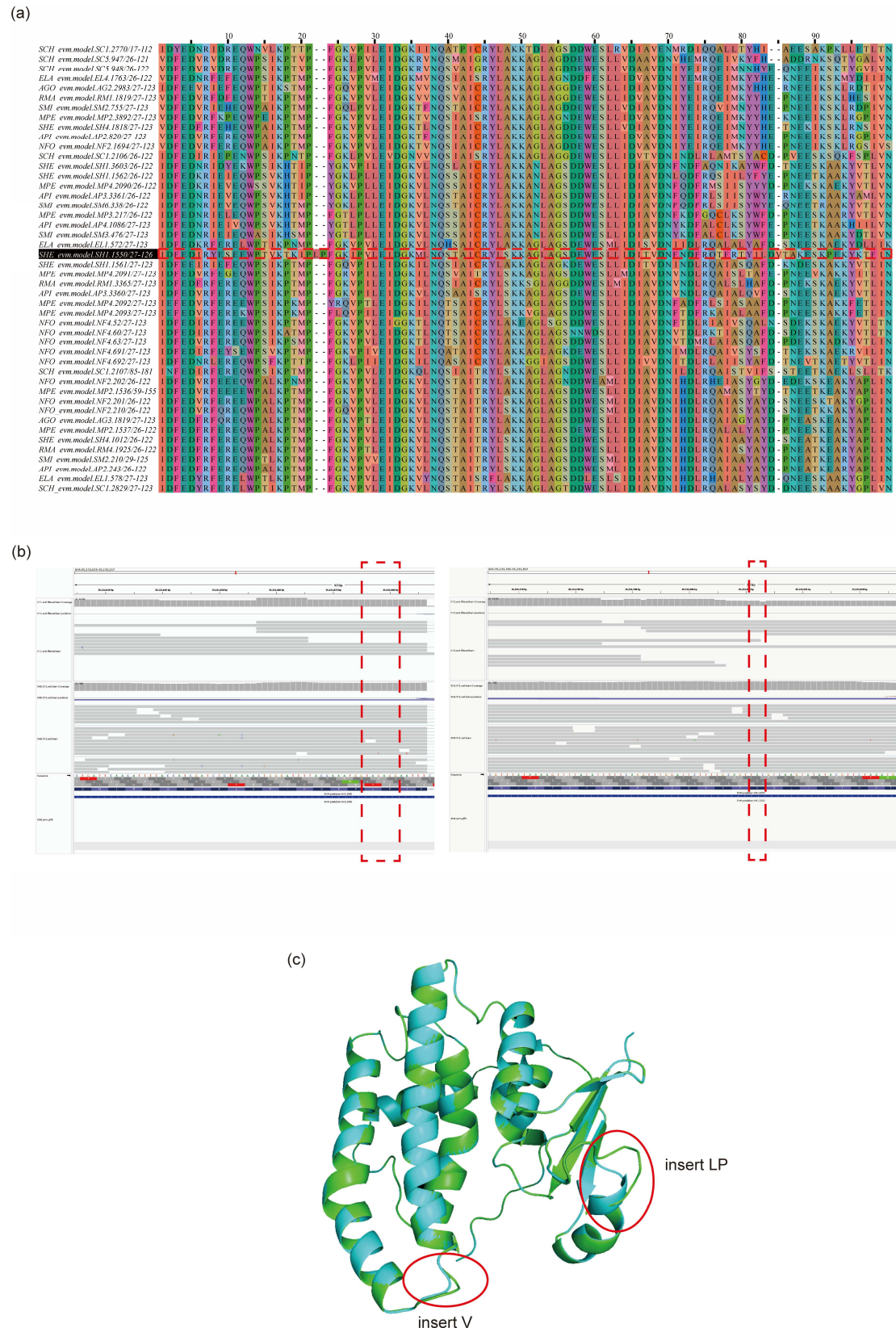
Continued Supplementary Figure 25. Phylogenetic relationships, domain and motif between UGTs from different aphids. Highlight the gene names belonging to *Semiaphis heraclei* with distinct colors. The gene branches in *Semiaphis heraclei* that underwent species-specific expansion are highlighted in red.



Supplementary Figure 26. Distribution of P450, CCEs, and UGTs genes in the genome of *Semiaphis heracleid*.



Supplementary Figure 27. Comparison of the expression levels of CCEs, P450s, and UGTs genes in sexual females and males aphids of *S. heraclei*. (a) Expression differences of CCEs genes between sexual females and males. (b) Expression differences of P450s genes between sexual females and males. (c) Expression differences of UGTs genes between sexual females and males.



Supplementary Figure 28. The insertion event in the gene SHE.evm.model.SH1.1550 may have impacted its function. (a) Alignment of GST sigma protein sequences from nine aphid species. (b) Verification of annotation and assembly errors using IGV. The upper track represents transcriptome reads from female aphids, while the lower track represents transcriptome reads from male aphids. The inserted sequences are circled in red with dashed lines. (c) Comparison of GST protein structures before and after

insertion. The green structure represents the protein structure after the insertion event, and the blue structure represents the protein structure before the insertion event. The structural differences caused by the two insertion events are highlighted with red circles.

SH1_plus_108895597_109372481.mRNA1/1-2276 1 AAGTTCCTAATAGGTTCTATTGTGATAAACCAAAACGAGTATTTCTGTGAAAAATGTTTTCAATATTATTTAAGTATAATATTACTTATAGTACTATGCATC 100
evm.model.SH1.4561/1-2034
SH1_plus_108895597_109372481.mRNA1/1-2276 101 GGAATGGAATTGATTTCATAATTTCGTCAACAACAATAACTGAAGCAGAAAGTTTAGACTCGATCGAGAAACGCCCAATGCATCTATATACATCAGCTTAC 200
evm.model.SH1.4561/1-2034 1 - - ATGGAATTGATTTCATAATTCTTCAACAACAATAACTGA - - AGAAAGTTAGACTCGATCGAGAAACGCCCAATGGATCTATACACATCAGCTTAC 94
SH1_plus_108895597_109372481.mRNA1/1-2276 201 GCTGTTCAATTGACAATCCGTCATGCAGATCTGTGTTGCTGACTGCCAAAGCTCTAGGTTTGGAAATTTCTTTACTACCTATAGATCTTTTACATGGTGA 300
evm.model.SH1.4561/1-2034 95 GATGTTCAATTGACAATCCGTCATGCAGATCTGTGTTGCTGACTGCCAAAGCTCTAGGTTTGGAAATTTCTTTACTACCTATAGATCTTTTACATGGTGA 194
SH1_plus_108895597_109372481.mRNA1/1-2276 301 ACAACTAGATCACGAAATCTTAGAGATACATATTGAACACACTTTACCGACAAAACGAAAGATACCCCTTGTTTTGTCAGATAGTCGGTAAATTAATGTT 400
evm.model.SH1.4561/1-2034 195 ACAACTAGATCACGAAATCTTAGAGATACATATTGAACACACTTTACCGACAAAACGAAAGATACCCCTTGTTTTGTCAGATAGTCGGTAAATTAATGTT 294
SH1_plus_108895597_109372481.mRNA1/1-2276 401 TACTTGGTTCAAAAATATGGAGGAGAAAAACAACCCGATGTACCCCAATGATCCTAAACACACAGCAAAAATAAACGTAATACTGGGTTTCAACTACGAAG 500
evm.model.SH1.4561/1-2034 295 TACTTGGTTCAAAAATATGGAGGAGAAAAACAACCCGATGTACCCCAATGATCCTAAACACACAGCAAAAATAAACGTAATACTGGGTTTCAACTACGAAG 394
SH1_plus_108895597_109372481.mRNA1/1-2276 501 ACTTGTACCCGCCTTTAAAAATCAATACAAGCCGTGGATGTTCAAAGGAGTACCGAAAAACCGAGGAAGCAGAGAAGGAAATCCACGTACAGCTTAGACA 600
evm.model.SH1.4561/1-2034 395 ACTTGTACCCGCCTTTAAAAATCAATACAAGCCGTGGATGTTCAAAGGAGTACCGAAAAACCGAGGAAGCAGAGAAGGAAATCCACGTACAGCTTAGACA 494
SH1_plus_108895597_109372481.mRNA1/1-2276 601 TTTGGAAAAGTATTAGAATCTTCCACTTTTGCTGCGCGCGAGTCAGTGACAGTGGCTGATATGGCTTTGGTTCATCTATTTGCACCTTTGAGGTCGCC 700
evm.model.SH1.4561/1-2034 495 TTTGGAAAAGTATTAGAATCTTCCACTTTTGCTGCGCGCGAGTCAGTGACAGTGGCTGATATGGCTTTGGTTCATCTATTTGCACCTTTGAGGTCGCC 594
SH1_plus_108895597_109372481.mRNA1/1-2276 701 GGTTGTTGCTTGAGAGAATACACCAAAATCAATGCATGGTTTGAAGGTGCAAGAAAGAATTAGTGGGATATGAAGAAGCCGATCAAATAGGAATCGATG 800
evm.model.SH1.4561/1-2034 595 GGTTGTTGCTTGAGAGAATACACCAAAATCAATGCATGGTTTGAAGGTGCAAGAAAGAATTAGTGGGATATGAAGAAGCCGATCAAATAGGAATCGATG 694
SH1_plus_108895597_109372481.mRNA1/1-2276 801 AAATCCAAAAGAAAATAGCCTTACGGAACCACTTGAATCTTCAACATCAGCTTATACTATTCAATTCTCAGTCCACCATGCAGATCTGTGTTGCTGAC 900
evm.model.SH1.4561/1-2034 695 AAATCCAAAAGAAAATAGCCTTACGGAACCACTTGAATCTTCAACATCAGCTTATACTATTCAATTCTCAGTCCACCATGCAGATCTGTGTTGCTGAC 794
SH1_plus_108895597_109372481.mRNA1/1-2276 901 TGCCCAAGCTTTACAGTTGGAATTTGATTTACAACCTATAGACCTCTTGACGGTAGACACACTCGATCCAGAAATTCATAAAGATAAATTTTCAACACACT 1000
evm.model.SH1.4561/1-2034 795 TGCCCAAGCTTTACAGTTGGAATTTGATTTACAACCTATAGACCTCTTGACGGTAGACACACTCGATCCAGAAATTCATAAAGATAAATTTTCAACACACT 894
SH1_plus_108895597_109372481.mRNA1/1-2276 1001 TTACCGACAATAAAGGAAGATAACTTTGTATTGTGGGATAGTCACGCAATTATCGTGTACTTGGTTAAACAATATGGAGGAGAAAAACAACACGCTGTACC 1100
evm.model.SH1.4561/1-2034 895 TTACCGACAATAAAGGAAGATAACTTTGTATTGTGGGATAGTCACGCAATTATCGTGTACTTGGTTAAACAATATGGAGGAGAAAAACAACACGCTGTACC 994
SH1_plus_108895597_109372481.mRNA1/1-2276 1101 CCAATGATCCTGAAAAACAATCAAGAATCAACGAAATACTGTATTTCAACAACGGACAATTGTACCCGCTCTTTAAAAATCAATACACCCGTGGGTGTT 1200
evm.model.SH1.4561/1-2034 995 CCAATGATCCTGAAAAACAATCAAGAATCAACGAAATACTGTATTTCAACAACGGACAATTGTACCCGCTCTTTAAAAATCAATACACCCGTGGGTGTT 1094
SH1_plus_108895597_109372481.mRNA1/1-2276 1201 CAATGGAGTACCGAAAAACCGACGAAGAGAGGATCAAATCCAGCTACAGTCGAGAAAGTTAGAAAAGTATTTAGAATCTTCCACTTTTGCTGCTGGCGTG 1300
evm.model.SH1.4561/1-2034 1095 CAATGGAGTACCGAAAAACCGACGAAGAGAGGATCAAATCCAGCTACAGTCGAGAAAGTTAGAAAAGTATTTAGAATCTTCCACTTTTGCTGCTGGCGTG 1194
SH1_plus_108895597_109372481.mRNA1/1-2276 1301 TCACTGACATTGGCTGATATGGCTTTGGTCGATCTATTTCAACTTTTGAGGTCACCTGGTGTGACCTTGACAGAATACACCAAAATCAATGAATGGCTTG 1400
evm.model.SH1.4561/1-2034 1195 TCACTGACATTGGCTGATATGGCTTTGGTCGATCTATTTCAACTTTTGAGGTCACCTGGTGTGACCTTGACAGAATACACCAAAATCAATGAATGGCTTG 1294
SH1_plus_108895597_109372481.mRNA1/1-2276 1401 AAAGATGCAAGAAAAAATAGTGGGATATGAAGAAGCCAATCAAATAGGAATCGATAAAATCAAAGGATAATAGCGCCACTAATAGTCACCGAATCTAC 1500
evm.model.SH1.4561/1-2034 1295 AAAGATGCAAGAAAAAATAGTGGGATATGAAGAAGCCAATCAAATAGGAATCGATAAAATCAAAGGATAATAGCGCCACTAATAGTCACCGAATCTAC 1394
SH1_plus_108895597_109372481.mRNA1/1-2276 1501 AAAAGTCAGCCCACTATGCAGATCTATTTTGCTGACTCTCAGTGTTTTAAATATTGGTTTGAATGAATAACCTATAGAATTCATAGAGCATGCAAGTGA 1600
evm.model.SH1.4561/1-2034 1395 AAAAGTCAGCCCACTATGCAGATCTATTTTGCTGACTCTCAGTGTTTTAAATATTGGTTTGAATGAATAACCTATAGAATTCATAGAGCATGCAAGTGA 1494
SH1_plus_108895597_109372481.mRNA1/1-2276 1601 GCAAAATCTTCCGAGTTAAGTTTTACACACACTTCACCGAGATTACAAATTGATGACAAATATTGGAGGACAGTCATGAAATCATCGAATCTTGATTA 1700
evm.model.SH1.4561/1-2034 1495 GCAAAATCTTCCGAGTTAAGTTTTACACACACTTCACCGAGATTACAAATTGATGACAAATATTGGAGGACAGTCATGAAATCATCGAATCTTGATTA 1594
SH1_plus_108895597_109372481.mRNA1/1-2276 1701 ATGAATACGGAAAAAATAAACTTGCTGTTTACCGATGATGTAATATGAAGAAGTAGTAAATGGTTTACTATTGTTCAACTACTCGCAATTGTACAA 1800
evm.model.SH1.4561/1-2034 1595 ATGAATACGGAAAAAATAAACTTGCTGTTTACCGATGATGTAATATGAAGAAGTAGTAAATGGTTTACTATTGTTCAACTACTCGCAATTGTACAA 1694
SH1_plus_108895597_109372481.mRNA1/1-2276 1801 GCCTTTTATAGATGAATATAACTCTTTGATTTTCTCTCGTCAAGATAAAATCGAAGGCGATGATGACTCTCCGATAAACTTTGAGAAGTGAATCGTAT 1900
evm.model.SH1.4561/1-2034 1695 GCCTTTTATAGATGAATATAACTCTTTGATTTTCTCTCGTCAAGATAAAATCGAAGGCGATGATGACTCTCCGATAAACTTTGAGAAGTGAATCGTAT 1794
SH1_plus_108895597_109372481.mRNA1/1-2276 1901 TTGAAGCAAAATAAAAAATAACACTTGGGCTGCGCGTGGCGATTGAGCGACAGCGGCTGATATTGCTATGGTCGCAACTATTTTCAACTATTTTGGAGGTTG 2000
evm.model.SH1.4561/1-2034 1795 TTGAAGCAAAATAAAAAATAACACTTGGGCTGCGCGTGGCGATTGAGCGACAGCGGCTGATATTGCTATGGTCGCAACTATTTTCAACTATTTTGGAGGTTG 1894
SH1_plus_108895597_109372481.mRNA1/1-2276 2001 CAAATGATTCCCTTGGAAAAATACCCCTCTATCAAGTCATGGCTAAACAAGTGCAAGAATGAATATTGGGATATGAAAAAGCCAAATCAAGAAGGAATCAA 2100
evm.model.SH1.4561/1-2034 1895 CAAATGATTCCCTTGGAAAAATACCCCTCTATCAAGTCATGGCTAAACAAGTGCAAGAATGAATATTGGGATATGAAAAAGCCAAATCAAGAAGGAATCAA 1994
SH1_plus_108895597_109372481.mRNA1/1-2276 2101 CATAATTAATAAATATGTTGACGGGCGTAAAAAATAAGGCATCTATATTATTACGCTTTGTTTAGATCAGAAGATTTTGTTCGATTTCTCAATGC 2200
evm.model.SH1.4561/1-2034 1995 CATAATTAATAAATATGTTGACGGGCGTAAAAAATAAG - - - - - 2034
SH1_plus_108895597_109372481.mRNA1/1-2276 2201 ACCATTTTATTTAAACACTGTCTAATGGTATAATACTAATATTACTGCATTTTATAAATTTTAGATACATAAAAAA 2276
evm.model.SH1.4561/1-2034

Supplementary Figure 29. Alignment of the SHE_evm.model.SH1.4561 gene sequence with the corresponding transcriptome-assembled gene sequence.

Reference

1. Mathers, T.C. *et al.* Chromosome-Scale Genome Assemblies of Aphids Reveal Extensively Rearranged Autosomes and Long-Term Conservation of the X Chromosome. *Molecular Biology and Evolution*. **38**, 856-875 (2021).
2. Chang, M. *et al.* miR-92a-1-p5 Modulated Expression of the flightin Gene Regulates Flight Muscle Formation and Wing Extension in the Pea Aphid, *Acyrtosiphon pisum* (Hemiptera: Aphidoidea). *Journal of Insect Science*. **22**(2022).
3. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology*. **33**, 290-295 (2015).
4. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*. **15**, 1-21 (2014).