

# MOLECULAR ECOLOGY RESOURCES

Supplemental Information for:

## ***Apolygus lucorum* genome provides insights into omnivorousness and mesophyll feeding**

Yang Liu<sup>1,#</sup>, Hangwei Liu<sup>2,#</sup>, Hengchao Wang<sup>2,#</sup>, Tianyu Huang<sup>1,#</sup>, Bo Liu<sup>2</sup>, Bin Yang<sup>1</sup>, Lijuan Yin<sup>2</sup>, Bin Li<sup>1</sup>, Yan Zhang<sup>2</sup>, Sai Zhang<sup>1</sup>, Fan Jiang<sup>2</sup>, Xiaxuan Zhang<sup>1</sup>, Yuwei Ren<sup>2</sup>, Bing Wang<sup>1</sup>, Sen Wang<sup>2</sup>, Yanhui Lu<sup>1</sup>, Kongming Wu<sup>1,\*</sup>, Wei Fan<sup>2,\*</sup>, Guirong Wang<sup>1,2,\*</sup>

### Table of Contents:

Title, Authors	Page 1
Supplemental text	Page 2-3
Figure S1, S2	Page 4
Figure S3, S4	Page 5
Figure S5, S6	Page 6
Figure S7, S8	Page 7
Figure S9	Page 8
Figure S10	Page 9
Figure S11	Page 10
Figure S12	Page 11
Figure S13	Page 12
Figure S14	Page 13
Figure S15	Page 14
Table S1, S2, S3	Page 15
Table S4, S5	Page 16
Table S6	Page 17
Table S7	Page 18
Table S8, S9	Page 19
Table S10	Page 20
Table S11	Page 21

# MOLECULAR ECOLOGY

## RESOURCES

### Supplemental text

#### 1 Genome assembly

##### 1.1 Estimate the *A. lucorum* genome size with K-mer

A K-mer refers to sequence with length of K nucleotides. A raw sequencing read with L bp contains (L-K+1) K-mers if the length of each K-mer is K bp. The frequency of each K-mer can be calculated from the genomic sequencing reads. The K-mer frequencies along the sequence depth gradient follow a poisson distribution in a given data set. During deduction, the genome size is  $G = K\_num / K\_depth$ , where the K\_num is the total number of K-mer, and K\_depth is the frequency occurring more frequently than the others (Li et al., 2010). We estimate the size of genome based on k-mer frequencies in Illumina reads. As the Illumina sequencing has base bias, genome size estimation through K-mer usually have 5% error rate (Liu et al., 2013). The genome size of *A. lucorum* is estimated to be 1.03G  $\pm$  52M (Figure S2). The first peak shows that the genome is highly heterozygous and has a heterozygosity rate about 1%.

##### 1.2 Lachesis configuration file

```
SPECIES = Alucorum
OUTPUT_DIR = /data1/lachesis_out
DRAFT_ASSEMBLY_FASTA = ../12.clean_final/canu_final.fa
SAM_DIR = /bowtie_results/bwt2/Apolygus/
SAM_FILES = LMC_Lib1_Lane1_Apolygus_canu_final.bwt2pairs.bam
RE_SITE_SEQ = GATC
USE_REFERENCE = 0
SIM_BIN_SIZE = 0
REF_ASSEMBLY_FASTA = -
BLAST_FILE_HEAD = -
DO_CLUSTERING = 1
DO_ORDERING = 1
DO_REPORTING = 1
OVERWRITE_GLM = 1
OVERWRITE_CLMS = 1
CLUSTER_N = 17
CLUSTER_CONTIGS_WITH_CENS = -1
CLUSTER_MIN_RE_SITES = 35
CLUSTER_MAX_LINK_DENSITY = 1.5
CLUSTER_NONINFORMATIVE_RATIO = 0
CLUSTER_DRAW_HEATMAP = 1
CLUSTER_DRAW_DOTPLOT = 1
ORDER_MIN_N_RES_IN_TRUNK = 25
ORDER_MIN_N_RES_IN_SHREDS = 15
ORDER_DRAW_DOTPLOTS = 1
REPORT_EXCLUDED_GROUPS = -1
REPORT_QUALITY_FILTER = 1
REPORT_DRAW_HEATMAP = 1
```

##### 2 EVM gene prediction pipeline

The gene models in *A. lucorum* were also predicted by EvidenceModeler (Haas et al., 2008) (v1.1.1), integrating evidences from *de novo* prediction, homology-based search and transcriptome sequencing assembly. Based on repeat soft masked genome, we used AUGUSTUS (Stanke et al., 2006) (v3.3.2) (--

# MOLECULAR ECOLOGY

## RESOURCES

species=Aluc\_busco --gff3=on --softmasking=on) for *de novo* gene prediction. Aluc\_busco is the training parameters that is generated by BUSCO assessment on chromosome genome, as suggested by BUSCO paper (Simao, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015). A total of 29,034 genes were predicted in this step.

For homology-based annotation, we used GenBlastA (She, Chu, Wang, Pei, & Chen, 2009) and Genewise (Birney, Clamp, & Durbin, 2004) to predict genes. Firstly, all Hemiptera proteins in NCBI RefSeq were download. Then, we aligned the proteins with the genome by tblastn (v2.7.1+) using 1e-5 as cutoff and filtered those that have less than 50% identity. Secondly, GenBlastA (with parameter: -e 1e-2 -g T -f F -a 0.5 -d 100000 -r 10 -c 0.5 -s 0) was used to define gene blocks on the genome and Genewise (with parameter: -tfor/-trev -gff -silent -kbyte 4000000 -quiet) was used to search those blocks against the remaining proteins to predict genes. A total of 13,805 were predicted through this process.

For the RNA-seq annotation, firstly, we filtered out adaptor, low quality and unpaired sequences in raw Illumina reads from 20 RNA-seq libraries from different life stages and tissues of *A. lucorum* by our in-house clean\_adapter (-s 15 -r 75), clean\_lowqual (-e 0.01 -r 75) and filter\_unpaired\_reads.pl. Then, the clean data were mapped to the genome by hisat2 (Kim, Langmead, & Salzberg, 2015) (v2.1.0), followed transcripts assembly by stringtie (Pertea et al., 2015) (v1.3.5). In total, 28,574 genes were predicted. For Iso-seq analysis, we used IsoSeq-3.0 (<https://github.com/ben-lerch/IsoSeq-3.0>) to generate full-length cDNA sequences and used seqclean in PASAPipeline<sup>3</sup> (v2.3.3) to screen vector based on UniVec database. After that, we used Launch\_PASA\_pipeline.pl to predict genes, resulting in 31,996 genes.

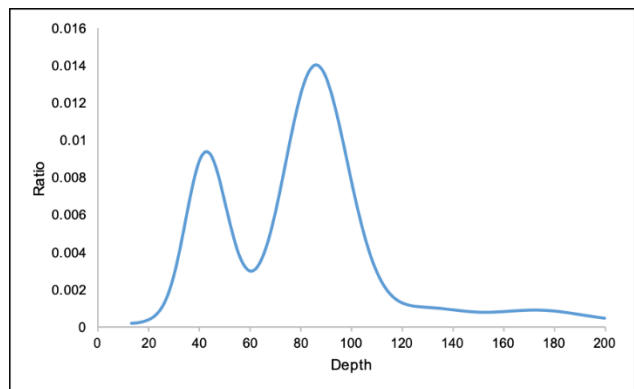
Finally, EVM was used to integrate the predicted genes with the *de novo*, homologous, and transcriptome data to produce a gene set, including 24,430 genes.

### References

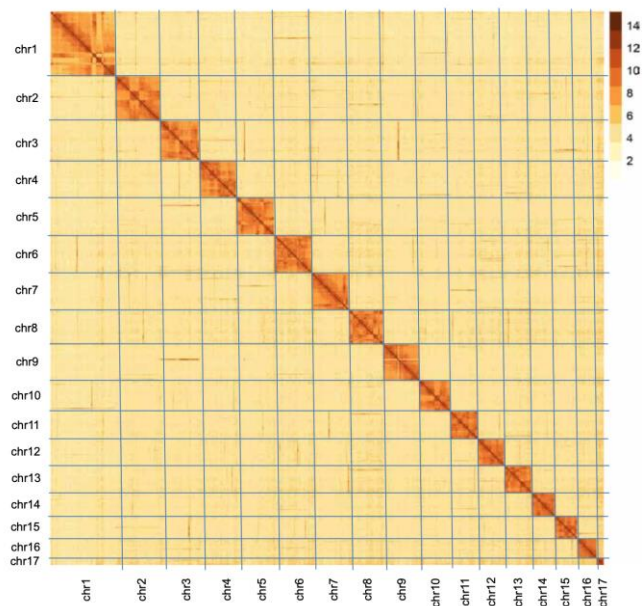
- Birney, E., Clamp, M., & Durbin, R. (2004). GeneWise and Genomewise. *Genome Research*, 14(5), 988-995. doi:10.1101/gr.1865504
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., . . . Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biology*, 9(1), R7. doi:10.1186/gb-2008-9-1-r7
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357-360. doi:10.1038/nmeth.3317
- Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., . . . Wang, J. (2010). The sequence and de novo assembly of the giant panda genome. *Nature*, 463(7279), 311-317. doi:10.1038/nature08696
- Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., . . . Fan, W. (2013). Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *arXiv e-prints*. Retrieved from <https://ui.adsabs.harvard.edu/#abs/2013arXiv1308.2012L>
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnol.*, 33(3), 290-295. doi:10.1038/nbt.3122
- She, R., Chu, J. S., Wang, K., Pei, J., & Chen, N. (2009). GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Research*, 19(1), 143-149. doi:10.1101/gr.082081.108
- Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210-3212. doi:10.1093/bioinformatics/btv351
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res*, 34 (Web Server issue), W435-439. doi:10.1093/nar/gkl200

# MOLECULAR ECOLOGY RESOURCES

## Supplemental figures

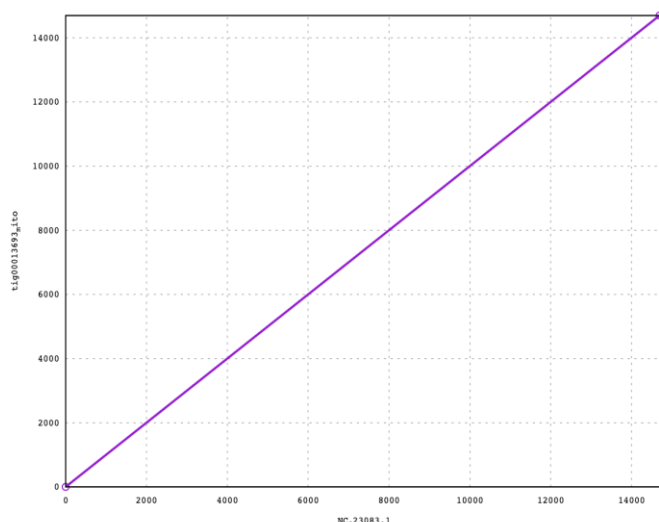


**Figure S1. Distribution of 17-mer frequency of *Apolygus lucorum* in the sequencing reads.** We used error corrected Illumina reads from the short insert-size libraries to calculate the k-mer frequency. The peak depth of this curve was 86. The estimated genome size of *A. lucorum* was  $1.03\text{G} \pm 52\text{M}$ ,  $1.082 \sim 0.978\text{G}$ .

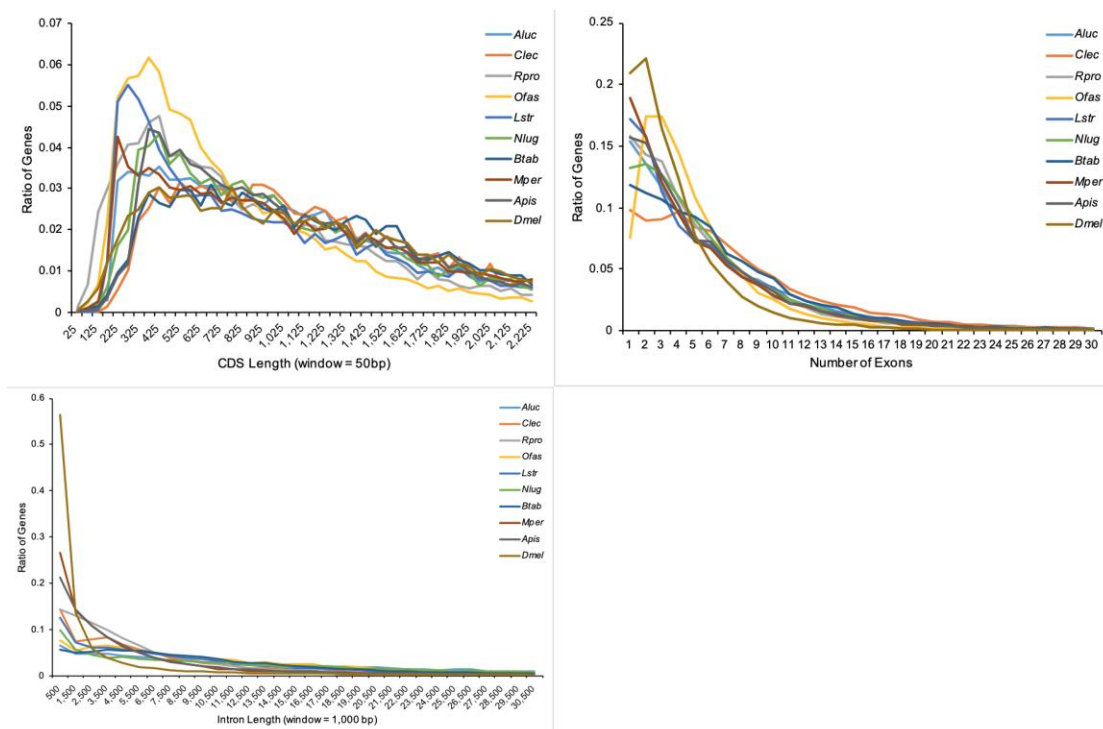


**Figure S2. Heatmap of Hi-C contact matrix.** A genome-wide contact matrix from Hi-C data between each pair of the 17 chromosomes using a 1 Mb window size. The color value indicates the base 2 logarithm of the number of valid reads ( $\log_2 [\text{valid reads}]$ ).

# MOLECULAR ECOLOGY RESOURCES

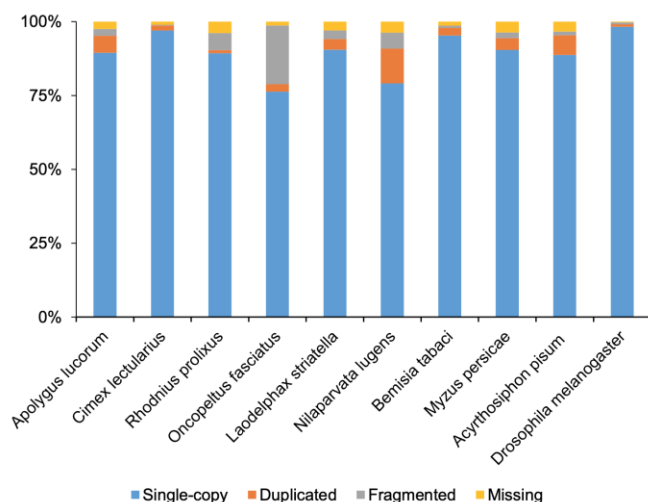


**Figure S3. The alignment between the reference mitochondrion genome and mitochondrion genome in this assembly.** The mitochondrion genome of this assembly has 99.7% identity with the reference, indicating that our genome is assembled successfully. Because of the circle architecture of mitochondrion genome, three times of alignment appeared between this contig and the reference and we extract the maximum alignment part of the contig as our mitochondrion genome. The length of mitochondrion in this genome is 14,694 bp, which is shorter than the reference mitochondrion genome length of 14,768.

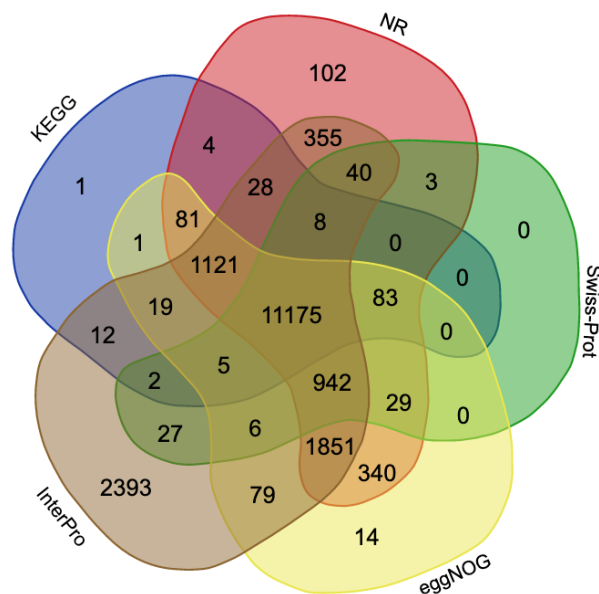


**Figure S4. Comparison of gene characteristics among the ten sequenced insect genomes.** CDS stands for coding sequences. Distribution of CDS length, number of exons and intron length among the gene sets of these insects.

# MOLECULAR ECOLOGY RESOURCES

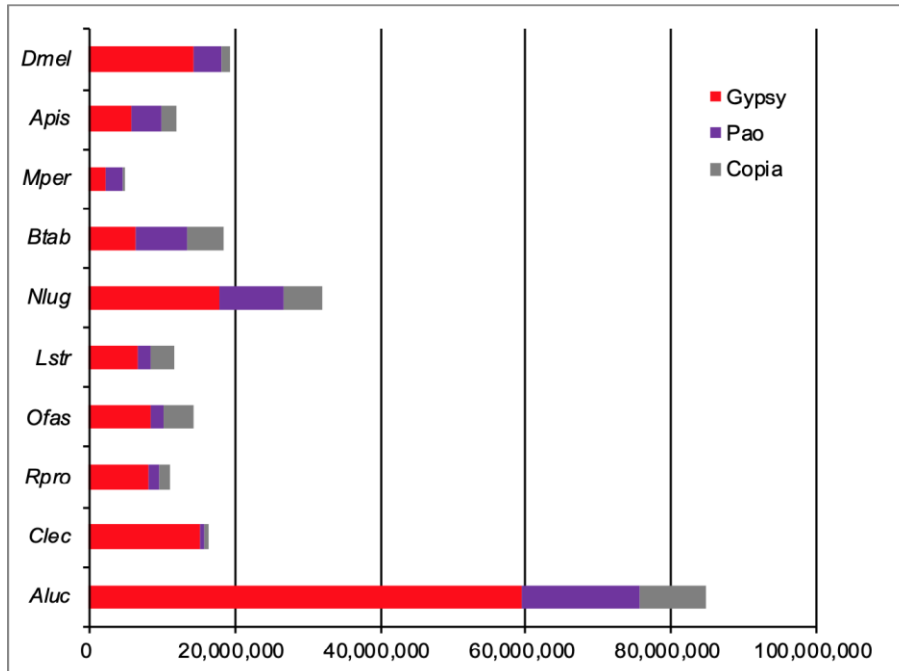


**Figure S5. BUSCO (Benchmarking Universal Single-Copy Orthologs) assessment of gene sets of *Apolygus lucorum* and other insects.** OrthoDB v9 of insecta was used (n = 1658). *A. lucorum* has 1579 (95.2%) complete BUSCOs, including 1483 (89.4%) complete and single-copy BUSCOs and 96 (5.8%) complete and duplicated BUSCOs.

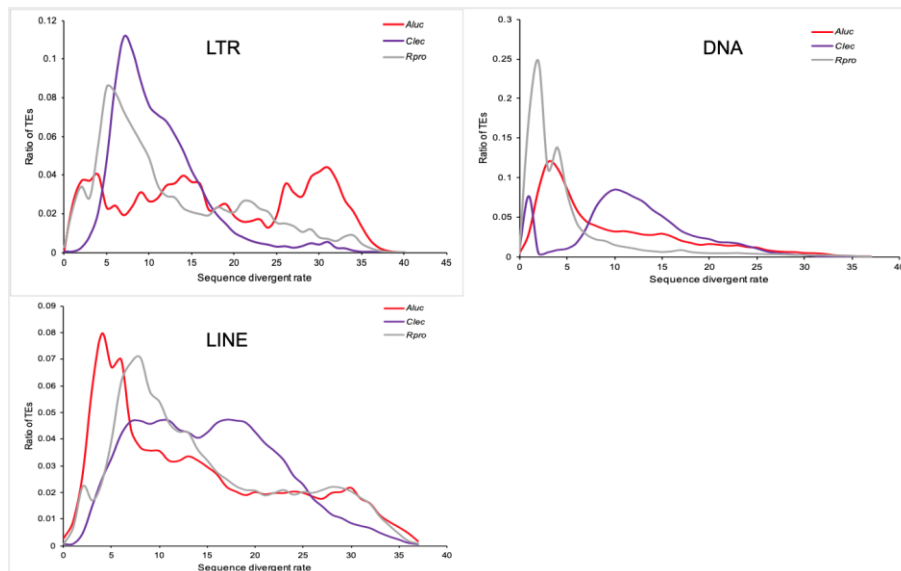


**Figure S6. Venn plot of functional annotation in gene set of *Apolygus lucorum*.** Figure is plotted by a tool (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). A total of 12,540 (61.61%), 16,162 (79.41%), 15,746 (77.36%), 18,063 (88.75%), 12,320 (60.53%) genes were functional annotated by KEGG, NR, eggNOG, Interpro and Swiss-Prot databases respectively. In total, 18,721 (91.98%) genes have homology information from either of these databases.

# MOLECULAR ECOLOGY RESOURCES

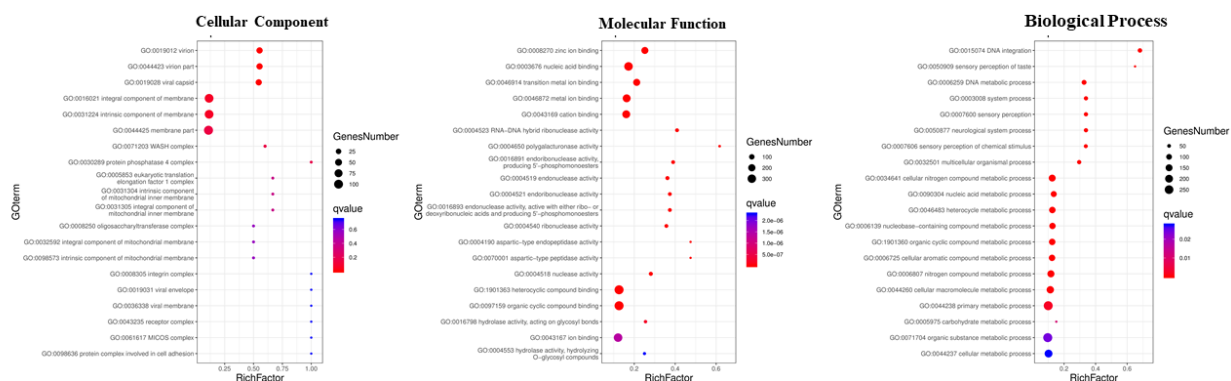


**Figure S7. Major subfamilies of LTR (long terminal repeats) composition comparison among ten insect genomes.** *A. lucorum* has much more LTRs than other genomes and the top three subfamilies are *Gypsy* (59,338,761), *Pao* (16,494,814) and *Copia* (9,038,478).



**Figure S8. Distribution of divergence rate of different types of TEs in three insect genomes.** Compared with two most relative species in Prosorrhyncha, DNA and LINE transposons in *A. lucorum* showed clear recent explosion.

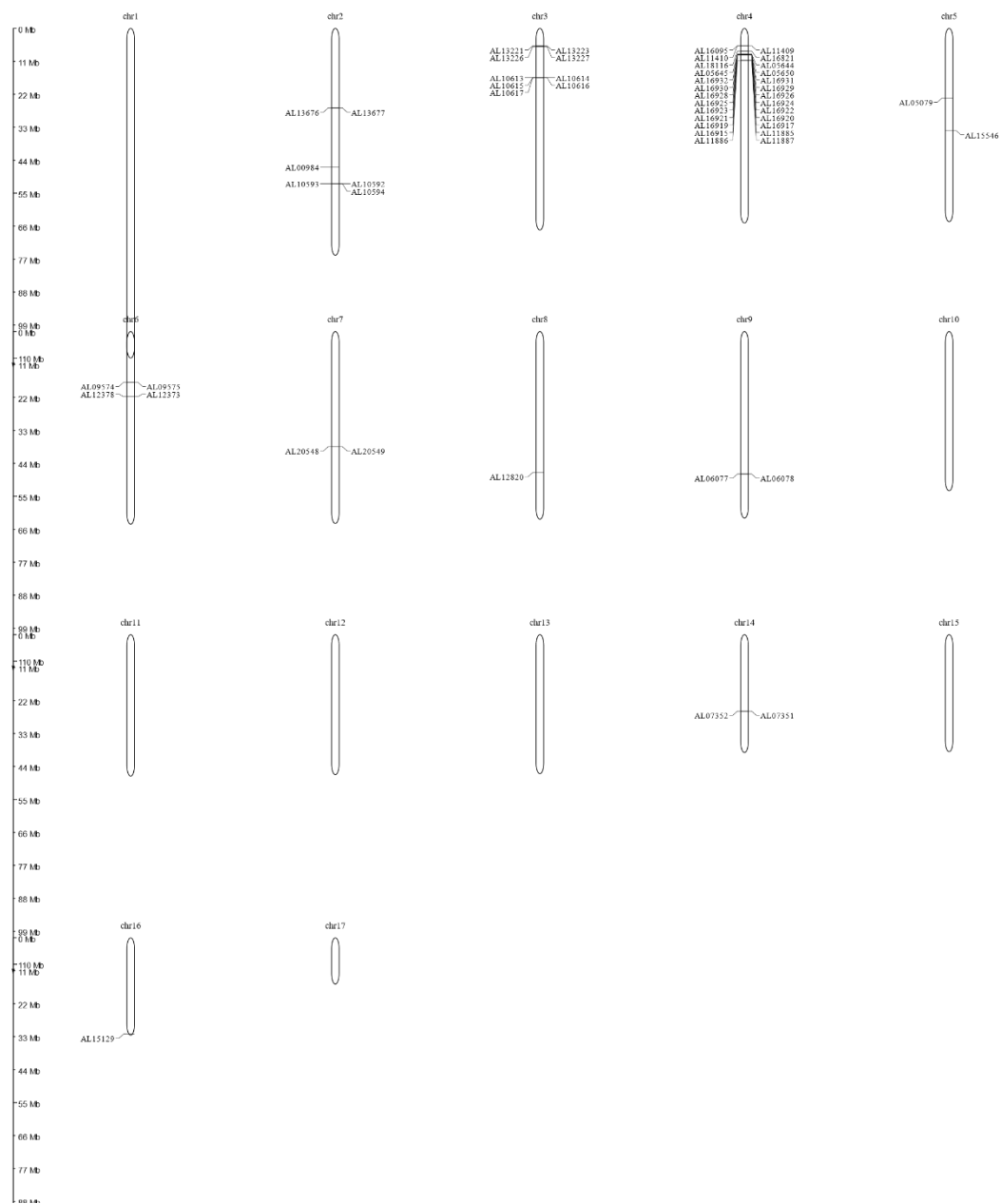
# MOLECULAR ECOLOGY RESOURCES



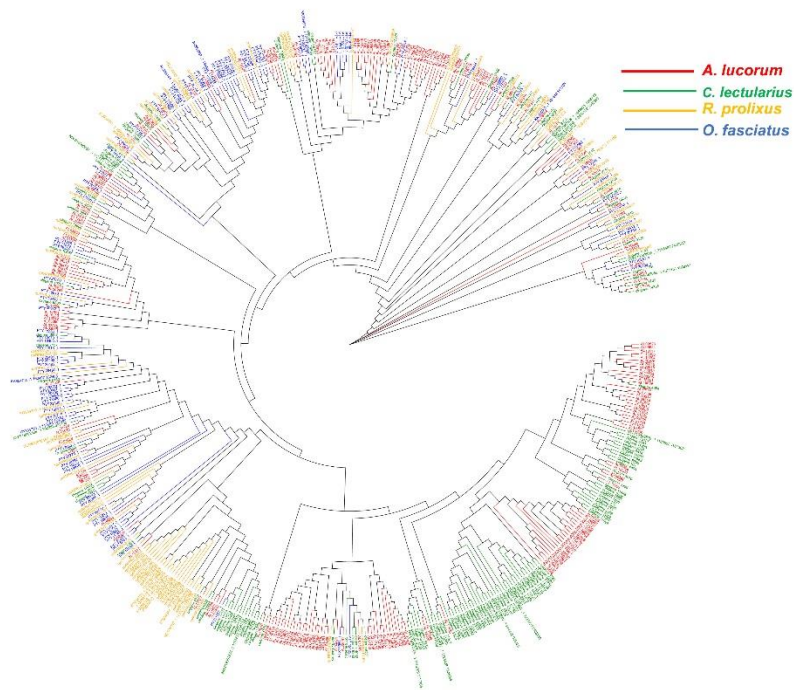
**Figure S9. GO enrichment of expanded genes in *Apolygus lucorum*.** a: GO enrichment in molecular function. b: GO enrichment in biological process. c: GO enrichment in cellular component. x-axis: RichFactor, GO term number in expanded genes/all genes; y-axis: pathway name.



# MOLECULAR ECOLOGY RESOURCES

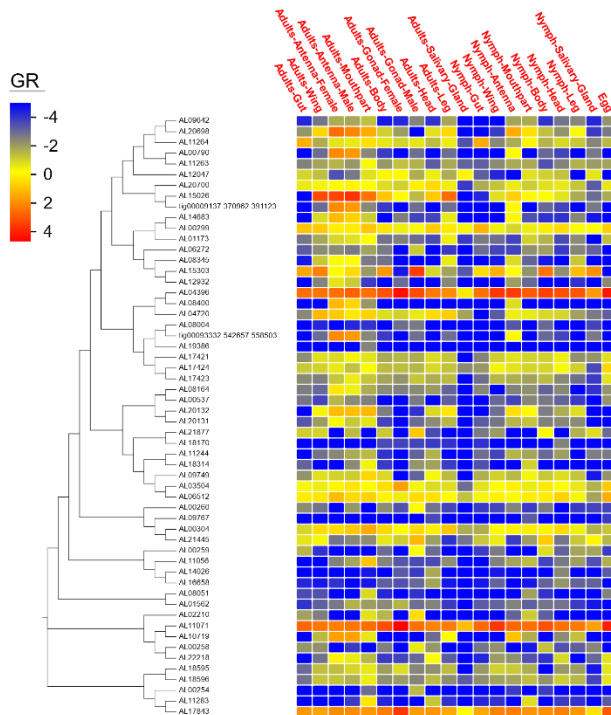


**Figure S10.** The distribution of the PG genes on the chromosome of *Apolygus lucorum*.



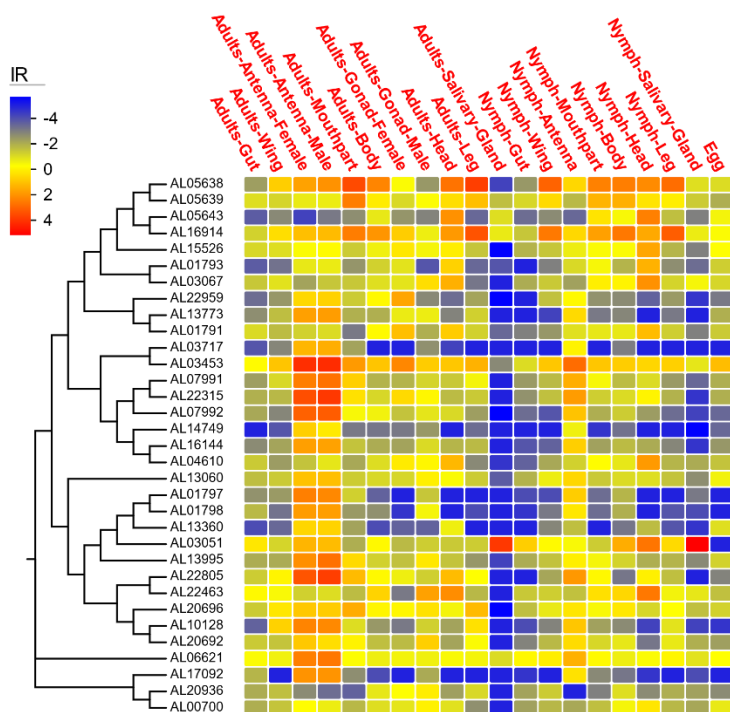
**Figure S11.** Phylogenetic tree of serine proteases in Prosorhyncha species.

# MOLECULAR ECOLOGY RESOURCES



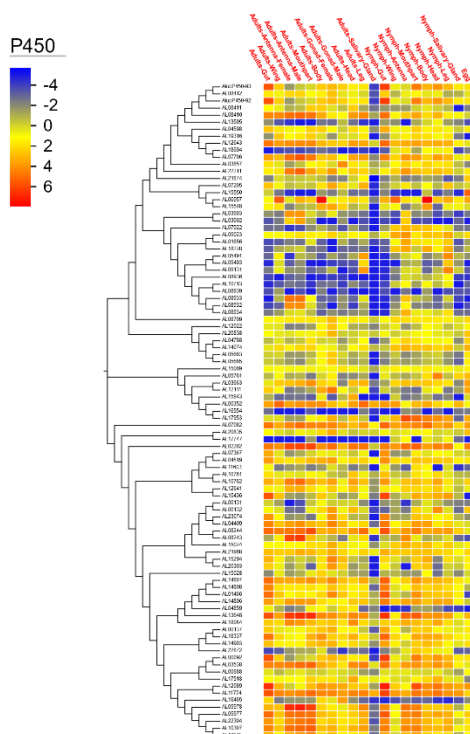
**Figure S12. Phylogenetic tree and expression profiling of GRs in *Apolygus lucorum*.** Each data block shows the base 10 logarithm of TPM ( $\log_{10}$ TPM) value of the corresponded tissue organ.

# MOLECULAR ECOLOGY RESOURCES



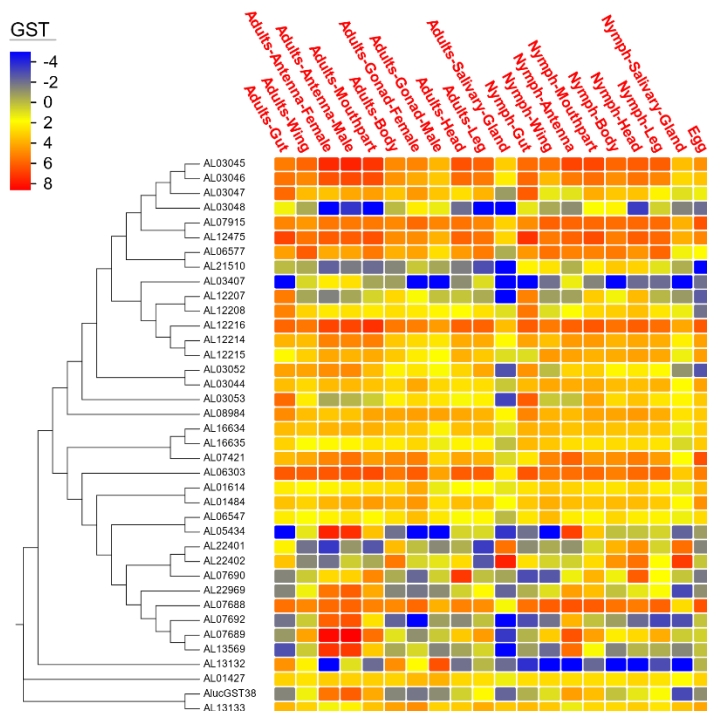
**Figure S13. Phylogenetic tree and expression profiling of IRs in *Apolygus lucorum*.** Each data block shows the base 10 logarithm of TPM (log<sub>10</sub>TPM) value of the corresponded tissue organ

# MOLECULAR ECOLOGY RESOURCES



**Figure S14. Phylogenetic tree and expression profiling of P450s in *Apolygus lucorum*.** Each data block shows the base 10 logarithm of TPM ( $\log_{10}\text{TPM}$ ) value of the corresponded tissue organ.

# MOLECULAR ECOLOGY RESOURCES



**Figure S15. Phylogenetic tree and expression profiling of GSTs in *Apolygus lucorum*.** Each data block shows the base 10 logarithm of TPM ( $\log_{10}$ TPM) value of the corresponded tissue organ.

# MOLECULAR ECOLOGY RESOURCES

## Supplemental tables

**Table S1. Sequencing statistics.**

Source	Libraries	Sequence Type	Insert Size	Mean Read Length	Total Raw Data (Gb)	Accession
Genomic DNA	Illumina	PE	400 bp	150 bp	140	SRR8767839, SRR8767840
	Illumina Hi-C	PE	400 bp	150 bp	199	SRR9673452
	PacBio-Sequel	SE	20 kb	9.7 kb	103	SRR8705016 ~ SRR8705026
Transcriptome Sequencing	Illumina	PE	400 bp	150 bp	633	SRR10411567 ~SRR10411626
	PacBio Iso-Seq	SE	20 kb	5.4 kb	17.6	SRR10417989 ~SRR10417991

**Table S2. Summary of polishing statistics.**

	Deletion	Insertion	Substitution	Total
Arrow polish	368,661	1,776,377	666,626	2,811,664
Pilon polish 1	418,833	393,967	3,048,053	3,860,853
Pilon polish 2	59,509	58,282	599,259	717,050

**Table S3. Summary of *Apolygus lucorum* assembly.**

	Contig assembly		Chromosome assembly	
	Count	Length (bp)	Count	Length (bp)
Total	3,818	1,023,255,709	191	1,023,618,409
Maximum	1	7,637,076	1	117,000,819
N10	21	3,197,054	1	117,000,819
N20	59	2,295,918	2	71,560,797
N30	115	1,521,203	3	69,083,832
N40	193	1,118,347	5	68,388,307
N50	302	785,215	6	68,132,828
N60	458	550,175	8	66,203,219
N70	689	351,741	9	56,526,986
N80	1,050	225,431	11	49,688,247
N90	1,696	109,307	13	41,843,727
Minimum	1	1,002	1	1,002

# MOLECULAR ECOLOGY RESOURCES

**Table S4. Summary of *Apolygus lucorum* chromosomes statistics.**

Chromosome	Contig number	Contig length	Chromosome length
chr1	255	116,975,419	117,000,819
chr2	250	80,503,971	80,528,871
chr3	206	71,540,297	71,560,797
chr4	310	69,052,932	69,083,832
chr5	173	68,599,956	68,617,156
chr6	378	68,350,607	68,388,307
chr7	208	68,112,128	68,132,828
chr8	508	66,560,408	66,611,108
chr9	110	66,192,319	66,203,219
chr10	129	56,514,186	56,526,986
chr11	79	50,190,941	50,198,741
chr12	307	49,657,647	49,688,247
chr13	272	49,310,982	49,338,082
chr14	93	41,834,527	41,843,727
chr15	122	41,534,163	41,546,263
chr16	114	34,535,072	34,546,372
chr17	130	16,416,141	16,429,041
Total anchored	3,644	1,015,881,696	1,016,244,396
Unanchored	174	7,402,966	NA

Note: NA: not available.

**Table S5. Summary of assembly and annotation of hemipteran genomes.**

	Assembled genome (Mb)	# scaffold	Scaffold N50 size (Mb)	Scaffold N90 size (Mb)	GC content (%)	No. of gene models	Average CDS size (Kb)	Average exon number	Average exon size (bp)	Average intron size (Kb)	BUSCO (%)	TEs (Mb)	Tandem repeat (Mb)
<i>Aluc</i>	1,023	191	68	42	39.39	20,353	1.4	6.6	204	3.8	97.5	656	73
<i>Clec</i>	511	1,574	1.64	0.49	34.82	11,936	1.6	8.0	196	3.4	98.9	218	18
<i>Rpro</i>	707	16,537	1.09	0.04	33.94	15,061	1.1	5.8	183	1.5	96.1	219	31
<i>Ofas</i>	1,099	17,222	0.34	0.04	32.42	19,489	0.9	5.1	177	3.0	98.7	319	29
<i>Lstr</i>	540	38,197	1.08	0.01	34.54	17,512	1.3	6.1	210	3.0	97.0	151	27
<i>Nlug</i>	1,141	46,559	0.36	0.05	34.57	19,806	1.3	6.2	215	4.2	96.3	489	73
<i>Btab</i>	615	19,751	3.23	0.38	39.64	13,901	1.6	6.9	233	4.9	98.6	286	17
<i>Mper</i>	347	4,022	0.44	0.07	30.03	18,529	1.3	5.8	230	1.5	96.3	78	4
<i>Apis</i>	541	21,920	1.8	0.04	29.76	18,270	1.3	6.0	223	2.2	96.6	164	11
<i>Dmel</i>	144	1,870	23.5	0.09	39.18	13,955	1.6	3.9	400	1.8	99.8	41	4

Note: BUSCOs using OrthoDB v9 of insecta (n = 1658) here included complete BUSCOs and fragmented BUSCOs, based on the gene sets of these insects.



# MOLECULAR ECOLOGY

## RESOURCES

**Table S6. Statistics of gene predictions from different gene prediction methods in *Apolygus lucorum*.**

Method	Augustus parameter	No. of Genes	Avg. CDS Length (bp)	Avg. No. of exon	Total No. of exon (K)	Total Length of CDS (Mb)	Avg. exon Length (bp)	BUSCO complete ratio (%)
August	pea_aphid	43,640	1,211	4.91	214	53	246	85.9%
August	busco	29,034	1,181	4.94	143	34	239	90.3%
GeneWise	NA	13,805	1,010	4.78	66	14	211	74.1%
RNA-seq	NA	28,574	2,703	4.70	134	77	574	82.4%
Iso-seq	NA	31,996	2665	2.82	90	85	946	68.3%
EVM	NA	24,430	1,252	5.445	133	31	229	91.0%
August + RNA hints	busco	22,297	1,313	6.35	142	29	206	95.2%
August + Iso-Seq hints	busco	20,643	1,171	5.54	114	24	211	90.9%
August + Protein hints	busco	21,417	1,246	4.80	103	27	259	85.2%
August + all hints (final)	busco	23,106	1,281	6.17	142	30	207	95.3%

Note: busco means parameters that was generated in BUSCO assessment of chromosome assembly, using OrthoDB v9 of insecta (n = 1658) database. We used the gene set from “August + all hints” as our final gene set, because it had the most complete BUSCO genes. “all hints” combined RNA, Iso-Seq and Protein hints. GeneWise stands for protein homology based prediction and RNA-seq stands for RNA-seq alignment and assembly based prediction. EVM integrates August with busco parameter, GeneWise and RNA-seq gene predictions. NA: not available.

# MOLECULAR ECOLOGY

## RESOURCES

**Table S7. The summary of tRNAs in *Apolygus lucorum*.**

tRNAs Type	Anti Codon	Total number
tRNA-Ala	AGC(18), CGC(3), TGC(4)	25
tRNA-Arg	ACG(4), CCT(6), TCG(5), TCT(3)	18
tRNA-Asn	GTT(12)	12
tRNA-Asp	GTC(28)	28
tRNA-Cys	GCA(5)	5
tRNA-Gln	CTG(8), TTG(7)	15
tRNA-Glu	CTC(8), TTC(15)	23
tRNA-Gly	CCC(3), GCC(19), TCC(12)	34
tRNA-His	GTG(13)	13
tRNA-Ile	AAT(13), TAT(4)	17
tRNA-Leu	AAG(13), CAA(6), CAG(6), TAA(1), TAG (2)	28
tRNA-Lys	CTT(23), TTT (9)	32
tRNA-Met	CAT(6)	6
tRNA-iMet	CAT (71)	71
tRNA-Phe	GAA(7)	7
tRNA-Pro	AGG(8), CGG(2), TGG(9)	19
tRNA-SeC	TCA(2)	2
tRNA-Ser	ACT(1), AGA(7), CGA(2), GCT(7), GGA(1), TGA(4)	22
Suppressor tRNAs	TTA(1)	1
tRNA-Thr	AGT(8), CGT(3), TGT(3)	14
tRNA-Trp	CCA(7)	7
tRNA-Tyr	GTA(7)	7
tRNA-Val	AAC(13), CAC(4), TAC(4)	21
Total	-	427

# MOLECULAR ECOLOGY RESOURCES

**Table S8. Repetitive elements identified in the *Apolygus lucorum* genome using different programs.**

Programs	Repeat Size (bp)	% of genome
TRF	72,964,816	7.13
RepeatMasker	101,964,198	9.96
RepeatProteinMask	66,739,710	6.52
RepeatModeler	562,564,777	54.98
Total TEs	655,921,134	64.10

Note: RepeatModeler used RepeatModeler to build *de novo* TE library and after that RepeatMasker was used to find TEs. Total TEs combined transposable elements identified by RepeatMasker, RepeatProteinMask and RepeatModeler.

**Table S9. Sources of genome data of 9 insect species.**

Species	Downloading sites
<i>Cimex lectularius</i>	<a href="https://www.ncbi.nlm.nih.gov/genome/?term=Cimex+lectularius">https://www.ncbi.nlm.nih.gov/genome/?term=Cimex+lectularius</a>
<i>Nilaparvata lugens</i>	<a href="https://www.ncbi.nlm.nih.gov/genome/2941">https://www.ncbi.nlm.nih.gov/genome/2941</a>
<i>Rhodnius prolixus</i>	<a href="https://www.vectorbase.org/organisms/rhodnius-prolixus">https://www.vectorbase.org/organisms/rhodnius-prolixus</a>
<i>Acyrtosiphon pisum</i>	<a href="https://www.ncbi.nlm.nih.gov/genome/448?genome_assembly_id=29162">https://www.ncbi.nlm.nih.gov/genome/448?genome_assembly_id=29162</a>
<i>Bemisia tabaci</i>	<a href="ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001_101000/100286/4.Q-type_B.tabaci.codinggenes.pep">ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001_101000/100286/4.Q-type_B.tabaci.codinggenes.pep</a>
<i>Myzus persicae</i>	<a href="http://bipaa.genouest.org/sp/myzus_persicae/">http://bipaa.genouest.org/sp/myzus_persicae/</a>
<i>Laodelphax striatella</i>	<a href="https://www.ncbi.nlm.nih.gov/genome/12198">https://www.ncbi.nlm.nih.gov/genome/12198</a>
<i>Oncopeltus fasciatus</i>	<a href="https://i5k.nal.usda.gov/Oncopeltus_fasciatus">https://i5k.nal.usda.gov/Oncopeltus_fasciatus</a>
<i>Drosophila melanogaster</i>	<a href="https://www.ncbi.nlm.nih.gov/genome/?term=Drosophila+melanogaster">https://www.ncbi.nlm.nih.gov/genome/?term=Drosophila+melanogaster</a>

# MOLECULAR ECOLOGY RESOURCES

**Table S10. Gene ID of digestive enzymes in *Apolygus lucorum***

	Gene IDs
Polygalacturonase	AL00984,AL05079,AL05644,AL05645,AL05650,AL06077,AL06078,AL07351,AL07352,AL09574,AL09575,AL10592,AL10593,A L10594,AL10613,AL10614,AL10615,AL10616,AL10617,AL11409,AL11410,AL11885,AL11886,AL11887,AL12373,AL12378,AL 12820,AL13221,AL13223,AL13226,AL13227,AL13676,AL13677,AL15129,AL15546,AL16095,AL16821,AL16915,AL16917,AL1 6919,AL16920,AL16921,AL16922,AL16923,AL16924,AL16925,AL16926,AL16928,AL16929,AL16930,AL16931,AL16932,AL18 116,AL20548,AL20549
Carboxypeptidase	AL01057,AL01618,AL01789,AL01970,AL02094,AL02095,AL02983,AL02984,AL04861,AL04923,AL05742,AL07725,AL08123,A L08124,AL08125,AL08126,AL08148,AL08397,AL09538,AL11468,AL11947,AL12389,AL12636,AL13892,AL13973,AL14276,AL 14689,AL15245,AL15922,AL15923,AL16047,AL16427,AL16429,AL16836,AL16983,AL16984,AL17134,AL17135,AL20763,AL2 1271,AL22917,AL22927,tig00002187_338709_340121,tig00012511_349520_350902,tig00014019_352546_354030,tig00016512_8 2360_83775,tig00090652_4413157_4414596,tig00091947_160175_161566,tig00472778_2305115_2306476
Cathepsin	AL02500,AL03008,AL04998,AL05942,AL05943,AL06326,AL06328,AL06329,AL06620,AL09394,AL09613,AL11757,AL11893,A L12286,AL12287,AL14465,AL15875,AL15878,AL15879,AL17738,AL19967,AL19968,AL19999,AL20000,AL20001,AL21372,AL 21373,AL21374,tig00006510_158903_163659,tig00007018_160378_169953,tig00007018_174347_182751,tig00007018_191031_19 8215,tig00007018_204624_232958,tig00020104_74551_79729,tig00020104_86553_97974,tig00090739_689891_696289,tig000923 19_943184_955386,tig00094168_6344_10557,tig00472802_394279_405813
Alpha-amylase	AL03478,AL04269,AL08857,AL08859,AL09026,AL16462,AL17635
Serine protease	AL00293,AL00312,AL00379,AL00380,AL00386,AL00776,AL01186,AL01229,AL01280,AL01281,AL01515,AL01519,AL01536,A L01662,AL01726,AL01728,AL01751,AL01809,AL01810,AL01811,AL01845,AL02030,AL02031,AL02182,AL02228,AL02398,AL 02602,AL02763,AL02764,AL03093,AL03482,AL03483,AL03490,AL03500,AL03501,AL03574,AL03878,AL03879,AL03880,AL0 4408,AL04471,AL04526,AL04635,AL05108,AL05219,AL05573,AL05850,AL05851,AL05865,AL06203,AL06225,AL06279,AL06 280,AL06505,AL06671,AL06805,AL07138,AL07488,AL07663,AL08026,AL08031,AL08394,AL08395,AL08412,AL08468,AL084 69,AL08470,AL08471,AL08511,AL08512,AL08513,AL08514,AL08515,AL08572,AL08574,AL08575,AL08705,AL08858,AL0894 7,AL08948,AL09035,AL09243,AL09486,AL09523,AL09524,AL10003,AL10127,AL10164,AL10165,AL10778,AL10816,AL10890 ,AL11058,AL11106,AL11215,AL11605,AL11609,AL11610,AL11611,AL11657,AL11673,AL11698,AL11867,AL11897,AL12013, AL12231,AL12232,AL12322,AL12503,AL12504,AL12505,AL12506,AL13117,AL13396,AL13510,AL13691,AL13728,AL13799,A L13803,AL13838,AL13839,AL13949,AL14027,AL14105,AL14116,AL14151,AL14152,AL14185,AL14332,AL14723,AL14994,AL 14995,AL14996,AL15154,AL15184,AL15267,AL15377,AL15551,AL15569,AL15604,AL15617,AL16054,AL16099,AL16100,AL1 6101,AL16135,AL16153,AL16154,AL16155,AL16312,AL16313,AL16317,AL16397,AL16409,AL16432,AL16433,AL16753,AL16 873,AL16943,AL16976,AL17064,AL17150,AL17200,AL17227,AL17302,AL17303,AL17455,AL17553,AL17554,AL17558,AL177 16,AL17718,AL17719,AL18125,AL18235,AL18501,AL18530,AL18774,AL18775,AL19137,AL19139,AL19220,AL19221,AL1929 4,AL19517,AL19552,AL19626,AL19735,AL19766,AL19846,AL20200,AL20492,AL20605,AL20693,AL20706,AL20708,AL20709 ,AL20745,AL20816,AL20818,AL20819,AL20822,AL20823,AL20872,AL21116,AL21117,AL21118,AL21337,AL21363,AL21678, AL21711,AL22146,AL22552,AL22556,AL22561,AL22562,AL22854,AL22855,AL23067,SP220,SP221,SP222,SP223,SP224
Aminopeptidase	AL01282,AL01977,AL03367,AL03625,AL04105,AL05838,AL06214,AL06270,AL06294,AL06536,AL06726,AL08407,AL08659,A L09255,AL09568,AL10878,AL11009,AL11013,AL12019,AL12648,AL12650,AL12725,AL13303,AL13304,AL13306,AL13329,AL 14383,AL14828,AL14829,AL14851,AL15472,AL15784,AL15990,AL16795,AL17095,AL17499,AL17669,AL18992,AL19032,AL1 9489,AL19829,AL20489,AL20716,AL22201,tig00031493_34393_51874
Phospholipase	AL00223,AL00284,AL00865,AL01886,AL01982,AL02059,AL02646,AL02648,AL02649,AL02652,AL02753,AL03254,AL03681,A L04180,AL05823,AL05825,AL05919,AL06772,AL07199,AL07308,AL08387,AL09497,AL11759,AL12301,AL14664,AL15679,AL 16179,AL17317,AL18735,AL18789,AL19224,AL19225,AL19621,AL19912,AL19913,AL19914,AL20225,AL20587,AL20943,AL2 0948,AL22254,AL22480,tig00002535_11434_50515,tig00002535_225012_248233,tig00002535_259525_293741,tig00002535_376 6_7653,tig00017424_152690_174782,tig00092812_391335_404252
Lipase	AL00350,AL00458,AL01669,AL01670,AL01671,AL01674,AL01905,AL01906,AL01982,AL02261,AL02263,AL02264,AL02267,A L02268,AL02269,AL02690,AL02934,AL02935,AL02957,AL02958,AL02959,AL03161,AL03698,AL04483,AL05006,AL05007,AL 05020,AL05382,AL05383,AL05458,AL05459,AL05462,AL05969,AL07032,AL07033,AL10235,AL10391,AL10392,AL10469,AL1 0808,AL10809,AL11288,AL12840,AL13568,AL14456,AL14458,AL14459,AL14701,AL14885,AL14897,AL15700,AL16483,AL16 541,AL16542,AL16598,AL16599,AL16600,AL16866,AL17519,AL18487,AL18673,AL18674,AL18675,AL18677,AL18732,AL196 94,AL21843,AL21875,AL22460,AL22461,AL22462,AL22852,tig00010510_350248_374399,tig00013472_44105_47665
Glucosidase	AL00196,AL02794,AL03339,AL04214,AL04215,AL04269,AL04493,AL05586,AL06584,AL06711,AL07110,AL07111,AL08773,A L08774,AL08776,AL08855,AL08857,AL08859,AL09026,AL09076,AL09239,AL10187,AL11065,AL11074,AL11739,AL11898,AL 11968,AL11990,AL13152,AL15930,AL17378,AL17527,AL17593,AL17687,AL18014,AL18015,AL18654,AL18659,AL19199,AL1 9997,AL20271,AL20332,AL21064,AL22496

# MOLECULAR ECOLOGY RESOURCES

**Table S11. Gene ID of detoxification genes in *Apolygus lucorum***

	Gene ID
CYP450	AL00092,AL00131,AL00132,AL00137,AL00352,AL00688,AL01466,AL01656,AL02262,AL03538,AL03602,AL03603,AL03663,AL03857,AL04409,AL04519,AL04598,AL04788,AL04859,AL05023,AL05131,AL05491,AL05493,AL05663,AL05665,AL05761,AL05977,AL05978,AL06957,AL07082,AL07205,AL07367,AL07706,AL07922,AL08243,AL08244,AL08410,AL08411,AL08412,AL08709,AL08932,AL08933,AL08934,AL08936,AL08939,AL10307,AL10436,AL10713,AL10761,AL10762,AL11603,AL11774,AL12022,AL12089,AL12111,AL12641,AL12643,AL13505,AL13648,AL14074,AL14806,AL14807,AL14808,AL14903,AL15024,AL15028,AL15069,AL15294,AL15548,AL15550,AL15943,AL16465,AL16554,AL17518,AL17747,AL17953,AL18084,AL18228,AL18337,AL18694,AL18945,AL19316,AL20303,AL20835,AL20838,AL21874,AL21968,AL22211,AL22304,AL22872,AL23074,tig00093624_129236_139984,tig00093624_166476_170648
GST	AL01427,AL01484,AL01614,AL03044,AL03045,AL03046,AL03047,AL03048,AL03052,AL03053,AL03407,AL05434,AL06303,AL06547,AL06577,AL07421,AL07688,AL07689,AL07690,AL07692,AL07915,AL08984,AL12207,AL12208,AL12214,AL12215,AL12216,AL12475,AL13132,AL13133,AL13569,AL16634,AL16635,AL21510,AL22401,AL22402,AL22969,tig00017592_104368_135869