


RESEARCH ARTICLE

Open Access



The draft genome of the specialist flea beetle *Altica viridicyanea* (Coleoptera: Chrysomelidae)

Huai-Jun Xue^{1,2*} , Yi-Wei Niu^{3,4†}, Kari A. Segraves^{5,6}, Rui-E Nie¹, Ya-Jing Hao^{3,4}, Li-Li Zhang^{3,4}, Xin-Chao Cheng⁷, Xue-Wen Zhang⁷, Wen-Zhu Li¹, Run-Sheng Chen^{3*} and Xing-Ke Yang^{1*}

Abstract

Background: *Altica* (Coleoptera: Chrysomelidae) is a highly diverse and taxonomically challenging flea beetle genus that has been used to address questions related to host plant specialization, reproductive isolation, and ecological speciation. To further evolutionary studies in this interesting group, here we present a draft genome of a representative specialist, *Altica viridicyanea*, the first Alticinae genome reported thus far.

Results: The genome is 864.8 Mb and consists of 4490 scaffolds with a N50 size of 557 kb, which covered 98.6% complete and 0.4% partial insect Benchmarking Universal Single-Copy Orthologs. Repetitive sequences accounted for 62.9% of the assembly, and a total of 17,730 protein-coding gene models and 2462 non-coding RNA models were predicted. To provide insight into host plant specialization of this monophagous species, we examined the key gene families involved in chemosensation, detoxification of plant secondary chemistry, and plant cell wall-degradation.

Conclusions: The genome assembled in this work provides an important resource for further studies on host plant adaptation and functionally affiliated genes. Moreover, this work also opens the way for comparative genomics studies among closely related *Altica* species, which may provide insight into the molecular evolutionary processes that occur during ecological speciation.

Keywords: *Altica*, Genome, Annotation, Host plant adaption, Chemosensory, Detoxification

Background

The high rate of diversification among host-specific herbivorous insects is thought to result from their shift and specialization to distinct host-plant species, creating conditions that promote reproductive isolation and contribute to the process of speciation [1–3]. One herbivore group that has been used to address questions about host plant specialization, reproductive isolation, and ecological speciation is the leaf beetle genus *Altica*

(Coleoptera: Chrysomelidae) [4–10]. This group has undergone rapid divergence that is largely associated with host plant use. For example, studies of three closely related species, *Altica viridicyanea*, *A. cirsiicola*, and *A. fragariae*, have demonstrated that although these species are broadly sympatric and quite similar in morphology, they feed on distantly related host plants from different plant families [6]. Consequently, their divergence is likely the result of dietary shifts to unrelated host plants [6]. Further, their close relationship is supported by crossing studies that show that interspecific hybrids can be generated under laboratory conditions [4–6, 8, 10], and phylogenetic analysis indicates that these species diverged

* Correspondence: xue@ioz.ac.cn; chenrs@ibp.ac.cn; yangxk@ioz.ac.cn

†Huai-Jun Xue and Yi-Wei Niu contributed equally to this work.

¹Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

³Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

over a relatively short period of time. Although *Altica* has been pivotal for understanding the linkages between host plant use and speciation [6, 11–13], the lack of a representative *Altica* genome hinders our ability to more thoroughly investigate the molecular mechanisms underlying the processes of ecological adaptation and diversification within this interesting group.

Key aspects of host plant adaptation and speciation among *Altica* beetles involve both behavioral adaptations to recognize and use the new host plant [14, 15] as well as physiological adaptations that allow them to feed on new plants containing different secondary compounds. These adaptations may involve changes in recognition cues to find the new host plants and new detoxification mechanisms that allow insects to avoid the deleterious effects of defensive chemistry [15, 16]. One aspect of host plant adaptation, then, is the prediction that the gene families involved in the detection of host chemical cues and those involved in xenobiotic detoxification will be key in facilitating successful shifts onto new host plant species. As a result, if we are to understand how host plant adaptation has played a role in speciation of *Altica* beetles, a reference genome would be helpful in making comparisons of candidate gene families involved in the diversification process.

The wealth of genetic and behavioral studies of *A. viridicyanea* makes it an excellent starting point for genomic investigations of host plant adaptation and speciation among *Altica* species. This species is an extreme host specialist of the plant *Geranium nepalense* (Sweet) (Geraniaceae) [4, 5, 7], and reproductive isolation is driven by the presence of species-specific cuticular hydrocarbons (CHC) that determine mating preferences [8, 9]. Studies of F₁ hybrids involving *A. viridicyanea* have shown that the CHC profiles can also be modified by the beetle's diet [8, 10]. Consequently, host plant use and mating preferences are intrinsically linked through chemistry.

Here we provide the genome assembly of *A. viridicyanea* (Fig. 1), the first genome of the subfamily Alticinae, and the fifth for the Chrysomelidae. Chrysomelids are a large and highly diverse family of beetles [17], many of which are economically important pests of agricultural crops [18]. The chrysomelid species for which genomes have been assembled exhibit intermediate preferences in host diet and are restricted to feeding on a single plant family (oligophagous). Consequently, the genome for *A. viridicyanea* will add to our genomic resources for a monophagous (restricted to a single host plant species) member of the Chrysomelidae. Furthermore, this assembly will also expand our knowledge of beetles in general as we currently have only 22 published beetle genome assemblies [19–31] (Table S1), a comparatively small number for such a diverse insect group [32–34] (as of



Fig. 1 Adult *Altica viridicyanea* (Photographed by Rui-E Nie and Qi-Long Lei)

May 1, 2020). Finally, the genome assembled here will provide an important resource for further studies on host plant adaptation and functionally affiliated genes.

Results and discussion

De novo genome assembly

Flow cytometry revealed that the genome size of *A. viridicyanea* ranged from 836.3 ± 13.8 Mb in females to 795.7 ± 8.3 Mb in males. We generated and assembled $187.3\times$ coverage from Illumina short reads and $72.7\times$ coverage via PacBio long reads from 157 female adults, thus creating a draft genome reference assembly of 864.8 Mb consisting of contig and scaffold N50s of 92.8 kb and 557.2 kb, respectively. The GC content was 31.67%. The size of the *A. viridicyanea* genome was larger than 85% of the currently published beetle genomes (Table S1). The draft genome assembly of *A. viridicyanea* was contained within 17,580 contigs that were assembled into 4490 scaffolds, with the longest scaffold size of 5.6 Mb. Using the reference set of 1658 insect BUSCOs, the genome contains 98.6% complete single-copy orthologs and multi-copy orthologs; using the reference set of 2442 Endopterygota BUSCOs, our genome contains 95.8% complete single-copy orthologs and multi-copy orthologs (Table 1). Together, the results of the above analyses indicate that the genome of *A. viridicyanea* is a robust assembly. Although robust, we note here that the annotated proteins of *A. viridicyanea* were consistently shorter than those from three beetles with relatively high N50 (*Tribolium castaneum*, *Dendroctonus ponderosae*, and *Anoplophora glabripennis*), indicating

Table 1 BUSCO results showing completeness of the *Altica viridicyanea* genome assembly and annotation

	Insect		Endopterygota	
	Counts	Percentage	Counts	Percentage
Complete BUSCOs	1635	98.6%	2340	95.8%
Complete and single-copy BUSCOs	1540	92.9%	2211	90.5%
Complete and duplicated BUSCOs	95	5.7%	129	5.3%
Fragmented BUSCOs	7	0.4%	52	2.1%
Missing BUSCOs	16	1.0%	50	2.1%
Total BUSCO groups searched	1658		2442	

that there is a potential for gene number inflation caused by the presence of partial genes in the current assembly. The estimated heterozygosity in the Illumina reads was about 0.70% ~ 0.96%, depending on k-mer size (k-mer 17, 19, 21, 23, 25 and 27).

We used these PacBio RNA-seq data to evaluate the genome assembly. Of the 13,550 polished reads, 60.46% could be successfully mapped to the genome. Transcripts that were unmapped or mapped with coverage or identity below the minimum threshold were partitioned into 1177 gene families based on k-mer similarity. Of the 1177 gene families, 121 could be mapped to the assembled genome, and hits to sequences from other species were found for 13 gene families. Blastn revealed that 1032 of the remaining gene families were similar to sequences from plants which may represent DNA contamination by plant material during the DNA isolation step, suggesting that 2 days is not long enough to complete gut clearing in *Altica*. We also analyzed the long reads discarded during QC to identify their origins. Of the 125,390 long reads, 93.27% could be classified, and over 50% of reads were assigned to bacteria. In addition, 37.8% of the reads were assigned to human, indicating contamination during sample and library preparation. These results highlight the importance of checking for genomic contamination during genome assembly.

Furthermore, the addition of HiC or optical mapping data would substantially improve the present genome assembly. Specifically, these additional data would improve the fragmented assembly and would also help resolve the assembly for the sex chromosomes. In the present study, we were limited by the availability of materials; however, future genomic studies in this system will bridge this gap.

Genome annotation

Prior to gene prediction using the assembled sequences, repeat sequences were identified in the genome of *A. viridicyanea*. The repetitive sequence content was about 62.91% of the assembly, which was similar to that of the cowpea weevil *Callosobruchus maculatus* (64%), lower than that of the ladybird *Propylea japonica* (71.33%),

and much higher than that of other beetle species (Table S1). Most of the repetitive sequences were transposable elements. According to a uniform classification system for eukaryotic transposable elements [35], retrotransposons (Class I) accounted for 41.27% whereas DNA transposons (Class II) accounted for 26.24% of the genome (Table 2).

To check whether the PacBio reads could span most of the repeats (transposons here), we aligned the PacBio reads to the assembled genome. Focusing on the primary alignment only, there were 89.01% (5,792,616/6,507,752) of reads successfully mapped to the genome. We found that 99.33% (1,913,673/1,926,492) annotated transposons are fully covered by at least one read. Of these regions fully spanned by PacBio reads, the longest one was 29,177 bp. The result shows the length distribution of repeats fully covered and repeats not fully covered by reads. It is clear that repeats fully covered by reads are significantly shorter than those repeats not fully covered. So, we suggested longer reads could help to resolve these regions.

The integration of de novo, RNA-seq-based and homology-based gene prediction methods identified 17,730 protein-coding genes in *A. viridicyanea* (Table 3, Fig. S1), a number slightly less than the average of beetle species with available genomes (~ 18,600 genes on average, Table S1). In total, 16,625 genes were assigned to putative functions, accounting for approximately 93.77% of the predicted genes (Table S2), and 750 putative pseudogenes were identified (Table S3). There were 2462 non-coding RNA models identified, including 45 miRNAs, 1093 rRNAs, and 1324 tRNAs, corresponding to 32, four and 24 gene families, respectively.

Phylogenetic analysis

We estimated the phylogenetic relationships of *A. viridicyanea* samples and an additional nine representative beetle species (*Anoplophora glabripennis*, *Aethina tumida*, *Agrilus planipennis*, *Dendroctonus ponderosae*, *Diabrotica virgifera virgifera*, *Leptinotarsa decemlineata*, *Nicrophorus vespilloides*, *Onthophagus taurus* and *Tribolium castaneum*). In total, 14,854 orthologs in *A.*

Table 2 Composition of repetitive sequences in the *Altica viridicyanea* genome assembly

Repeat type	Number of elements	Length (bp)	Rate (%)
Retrotransposons (transposable element class I)	1,033,903	356,902,348	41.27
DIRS	12,458	7,468,113	0.86
LINE	3,17,300	97,559,643	11.28
LTR/uncertain	45,604	28,406,497	3.28
LTR/Copia	15,362	7,752,670	0.9
LTR/Gypsy	143,535	83,332,210	9.64
LTR or DIRS	91	14,310	0
PLE or LARD	485,362	155,872,484	18.02
SINE	78	80,868	0.01
TRIM	13,553	6,401,760	0.74
Unknown	560	58,144	0.01
DNA transposons (transposable element class II)	803,681	226,943,053	26.24
TIR	550,895	150,574,911	17.41
MITE	10	635	0
Crypton	145,294	43,476,772	5.03
Helitron	29,008	8,403,268	0.97
Maverick	51,701	31,213,827	3.61
Unknown	26,773	5,423,361	0.63
SSR	583	165,863	0.02
Unknown	77,813	23,953,352	2.77
Total	1,848,679	544,002,544	62.91

Notes: DIRS dictyostelium intermediate repeat sequence, LINE long interspersed nuclear element, LTR long terminal repeat, PLE penelope-like elements, SINE short interspersed nuclear element, LARD large retrotransposon derivative elements, TIR terminal inverted repeat

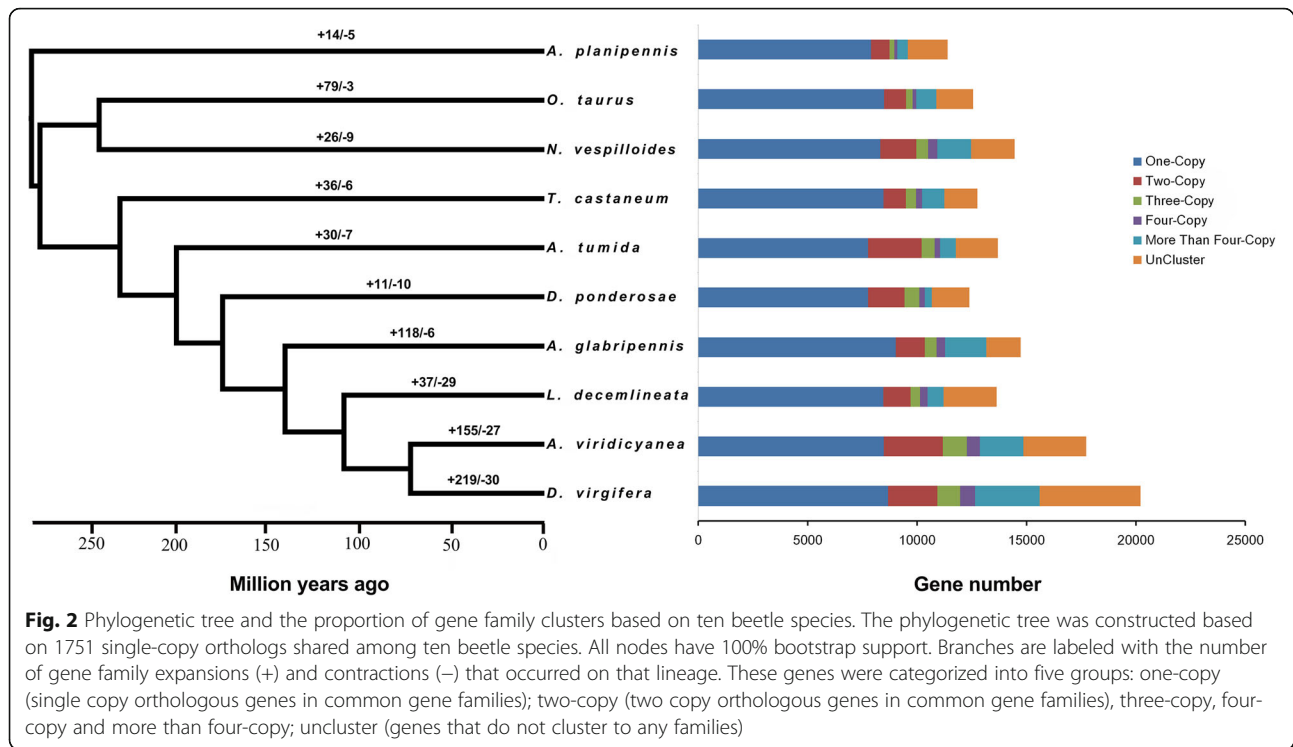
Table 3 Statistics of gene prediction of *Altica viridicyanea*

Method	Software	Gene number
<i>ab initio</i>	Genscan v1.1.0	15,170
	Augustus v2.4	31,813
	GlimmerHMM v3.0.4	58,872
	GeneID v1.4	14,970
	SNAP v2006-07-28	72,679
homology-based	GeMoMa v1.3.1	
	<i>Drosophila melanogaster</i>	9310
	<i>Tribolium castaneum</i>	14,234
	<i>Dendroctonus ponderosae</i>	13,066
	<i>Anoplophora glabripennis</i>	18,274
transcriptome-based	TransDecoder v2.0	73,432
	GeneMarkS-T v5.1	29,645
	PASA v2.0.2	23,200
Integration	EvidenceModeler v1.1.1	17,730

viridicyanea clustered with the other nine representative beetle species. We identified 1321 *A. viridicyanea* specific genes, corresponding to 470 gene families, and with the exception of *Diabrotica virgifera virgifera*, this number was much greater than the other representative beetle species included in this analysis (Table S4). The phylogenetic relationships were consistent with the results inferred from large datasets [36–38] based on 1751 conserved single copy orthologs. For example, *A. viridicyanea*, *Diabrotica virgifera virgifera* and *Leptinotarsa decemlineata*, all belonging to chrysomelids, formed a clade, and these species clustered with *Anoplophora glabripennis*, a member of the superfamily Chrysomeloidea (Fig. 2). The estimated divergence time between *A. viridicyanea* and *Diabrotica virgifera virgifera* was about 74.7 million years ago. From this analysis, we also identified 155 gene families that expanded and 27 gene families that contracted along the *A. viridicyanea* lineage (Fig. 2). Some of these gene families were related to chemosensory and detoxification functions.

Chemosensory gene families

In many herbivorous insects, feeding, mating and oviposition behaviors are mediated by chemical cues [39]. The chemosensory system may also play important roles



in speciation of some insects [40–42]. This is likely the case in *A. viridicyanea* as previous work has shown that this highly specialized beetle primarily uses chemical cues to achieve sexual isolation from its sibling species [8]. Furthermore, these contact chemicals also act as a mating signal to discriminate intraspecific variation in sexual maturity [9]. In addition, chemical cues are modified by and likely involved in host plant choice [8, 10]. Consequently, we investigated *A. viridicyanea* gene families known to be involved in chemosensory signaling in insects.

There are at least five gene families involved in the detection of chemicals, including three receptor families, odorant receptors (ORs), gustatory receptors (GRs) and ionotropic receptors (IRs), and two protein binding families, odorant binding proteins (OBPs) and chemosensory proteins (CSPs). These receptor families are usually expressed in insect olfactory sensory neurons and are involved in the detection of a suite of chemicals. For instance, volatile chemicals are detected by ORs [43–45], contact chemicals or carbon dioxide are detected by GRs [46], and nitrogen-containing compounds, acids, and aromatics are identified by IRs [47]. In contrast, the binding protein gene families are highly abundant in the sensillar lymph of insects and usually function as carriers of hydrophobic scent molecules to the receptors [48, 49].

In the genome of *A. viridicyanea*, we identified 173 putative chemosensory genes and two pseudogenes.

Perhaps not surprisingly, the gene repertoire of the monophagous *A. viridicyanea* was considerably reduced as compared to that of host generalist species such as *T. castaneum* (630 genes plus 103 pseudogenes) and *A. glabripennis* (451 genes plus 65 pseudogenes). Upholding this pattern, *A. viridicyanea* has fewer chemosensory genes than the oligophagous species such as *Dendroctonus ponderosae* (240 genes plus 10 pseudogenes) and *L. decemlineata* (> 300 genes) that specialize on a single family of host plants (Table S5). Yet there are outliers to this trend; *Agilus planipennis* (132 genes and two pseudogenes) and *Diabrotica virgifera virgifera* (135 genes, but the gene number for IRs is unavailable) are species that are intermediate in host range, but they have fewer chemosensory genes than *A. viridicyanea*. These findings are generally consistent with the hypothesis that chemosensory gene content and host specificity should correlate in phytophagous beetles [50], although there are clearly exceptions to this rule.

Insect ORs are proteins with seven transmembrane domains that are involved in the detection of volatile chemicals [44, 51, 52]. The number of ORs in beetle species varies widely from 30 to hundreds of ORs [53]. When we examined the *A. viridicyanea* genome for the presence of ORs, we found a diversity of gene families. There were 63 ORs and one pseudogene (Pseudo-Gene48) that were classified into eight subfamilies: Group 1, 2A, 2B, 3, 4, 5A, 5B and 7 (Fig. 3; Table S5). Following the new OR classification scheme [53], we also

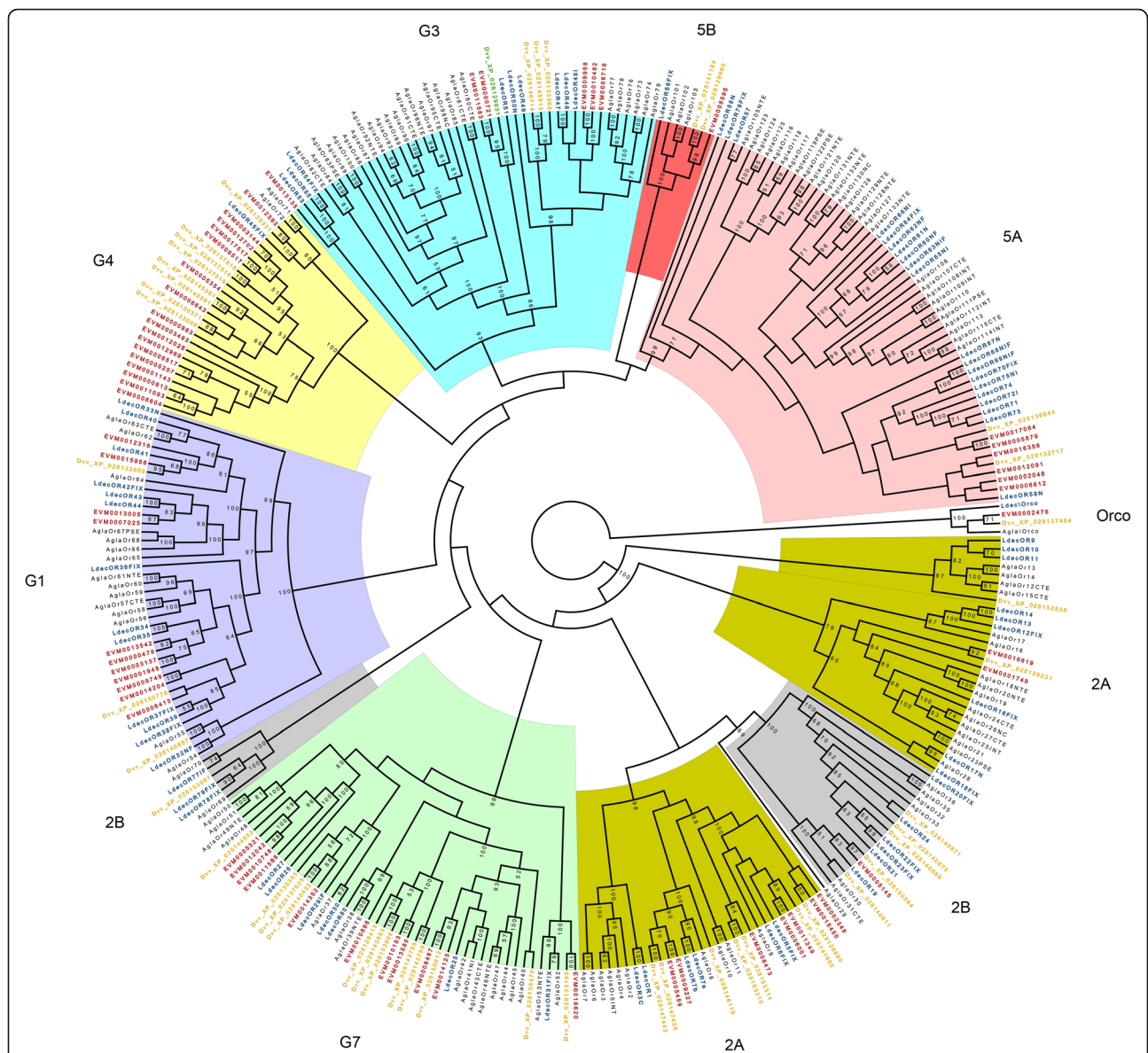


Fig. 3 Maximum likelihood cladogram of odorant receptor genes from four beetle species. *Altica viridicyanea* (red labels), *Leptinotarsa decemlineata* (blue labels), *Anoplophora glabripennis* (black labels) and *Diabrotica virgifera virgifera* (Dvv, yellow labels). Node support values lower than 50 are not shown

identified one highly conserved olfactory co-receptor, Orco, that has been found in other beetle species. Interestingly, we also found a large expansion in *A. viridicyanea* in Group 4 that contained 17 ORs (ten are full length). By comparison, eight Group 4 OR genes have been previously identified in *Diabrotica virgifera virgifera* and no more than four in any other surveyed beetle species [50, 53].

In addition to ORs, we also compared GRs across beetle taxa. Most GRs are expressed in gustatory receptor neurons in taste organs and are involved in contact chemoreception and detection of CO₂ [46]. We annotated 39 GRs in *A. viridicyanea*, including three

conserved candidate CO₂ receptors, nine candidate sugar receptors, and the remaining were candidate bitter taste receptors. Simple orthology of GRs is generally rare in beetles [50], and not surprisingly, no single-copy orthologs were revealed in the species that we compared. The phylogenetic analysis showed that 2–7 GRs from each of the seven species grouped within the clade of conserved sugar receptors. Additionally, two or three genes from five of the eight species, with the exception of nine genes from *Diabrotica virgifera virgifera*, formed a clade of CO₂ receptors (Fig. S2).

The number of GRs varied from 10 to 245 in the eleven surveyed beetles (Table S5). Comparisons with *A.*

viridicyanea identified as many as 147 GRs in an oligophagous chrysomelid species *Leptinotarsa decemlineata*, and 54 GRs in *Diabrotica virgifera virgifera*, whereas fewer than 20 GRs were annotated in four other chrysomelids (*Colaphellus bowringi*, *Ophraella communa*, *Pyrrhalta aenescens* and *Pyrrhalta maculicollis*). The extremely low numbers of GRs in the latter four species is likely the result of differences in data collection—those species only had transcriptomic data available, and that approach generally does not describe the full complement of chemosensory genes. For example, a study in the longhorn beetle *Anoplophora glabripennis* found 11 GRs when using transcriptomic data, however, genomic data revealed 234 GRs [34, 54].

The next chemosensory receptor group that we examined was the IRs, a conserved family that evolved from a family of synaptic ligand-gated ion channels, ionotropic glutamate receptors (iGluRs) [47, 55, 56]. In insects, the IRs include two groups: the conserved “antennal IRs” that have an olfactory function, and the species-specific “divergent IRs” which are candidate gustatory receptors [57]. Our genome annotations revealed 12 ionotropic receptors (IRs). Only the members of the conserved antennal IR21a group were identified in all eight of the beetle species that we surveyed, whereas the clades IR25a and IR76b were formed by single-copy orthologs from six species, excluding *P. aenescens* and *P. maculicollis* (transcriptomic data are available for both of these species); clade IR8a was also formed by single-copy orthologs from the same six species; however, there were four copies from *Diabrotica virgifera virgifera* (Fig. 4). Furthermore, IRs from all eight species fell within the well-supported non-single-copy IR75 clade (Fig. 4). Compared to other groups, IRs show a contraction in Galerucinae and Alticinae, two closely related subfamilies of Chrysomelidae [58] (*Colaphellus bowringi*, *Chrysomela lapponica*, *O. communa*, *P. aenescens*, *P. maculicollis*, *Diabrotica virgifera virgifera* and *A. viridicyanea*; Table S5).

Finally, we examined the protein binding gene families. OBPs and CSPs are generally regarded as carriers of pheromones and odorants in insect chemoreception, and a multitude of additional functions have also been suggested such as carrying semiochemicals and visual pigments, promoting development and regeneration, and digesting insoluble nutrients [59]. OBPs are small, soluble proteins with six conserved cysteines [48]. Although the detailed mechanisms remain unclear [60], it is believed that OBPs deliver hydrophobic molecules to the receptors [48]. In *A. viridicyanea*, we annotated 48 putative OBP genes and one pseudogene (PseudoGene855). Among these, 34 genes belonged to the Minus C OBPs. We found only one clade of classic OBPs, i.e., Classic VIII, which include single-copy orthologs from each of

the eight species in the analysis. In clade IX, two copies from *Diabrotica virgifera virgifera* clustered with single-copy orthologs from seven other species. In clade VII, two copies from *Dendroctonus ponderosae* clustered with single-copy orthologs from other seven species, whereas in Clade X two copies from *Dendroctonus ponderosae* and three copies from *Diabrotica virgifera virgifera* clustered with single-copy orthologs from six other species. Clades I and IV were formed by single-copy orthologs from seven species except for *Diabrotica virgifera virgifera*. The clades of Classic II, III, V and VI were formed by orthologs from 5 to 7 species (Fig. S3). Plus-C OBPs were not found in *A. viridicyanea*, and are also absent in the *Pyrrhalta* species and *Diabrotica virgifera virgifera* that belong to the “Galerucinae+Alticinae” taxonomic group.

CSPs are characterized by the presence of four cysteines that form two disulfide bridges [61]. We annotated 12 CSP genes in *A. viridicyanea*. The phylogenetic analysis revealed that only one clade (clade 1) was formed by single-copy orthologs from the eight species surveyed. Clades 2–7 were formed by single-copy orthologs from 5 to 7 beetle species. In these lineages, the absence of IRs from transcriptomic sources (e.g., *P. aenescens*, *P. maculicollis* and *O. communa*) was more common whereas the orthologs of *A. viridicyanea* also lacked members of clade 5 (Fig. S4).

Similar to previous work on GRs, transcriptomes often fail to describe the full set of chemosensory genes due to low expression, spatiotemporal variation in expression, or shallow sequencing depth. For instance, 106 chemosensory genes were detected in *Anoplophora glabripennis* using transcriptomic sequencing [54] whereas more than 500 chemosensory genes (65 pseudogenes included) were annotated from its genome [50].

Detoxification supergene families

Novel plant secondary compounds often present a challenge for herbivorous insects, and physiological adaptation to novel plant secondary metabolites is a key problem. The detoxification and metabolism of most xenobiotics occurs via a common set of detoxification-related enzymes, all of which belong to multigene families [62]. The cytochrome P450s (P450s), carboxyl/cholinesterases (CCEs), and glutathione S-transferases (GSTs) are widely regarded as the major insect gene/enzyme families involved in xenobiotic detoxification [63–65]. In addition, the UDP-glucuronosyltransferases (UGTs) and ATP binding cassette transporters (ABCs) can also play a role in detoxification [66–68]. This diversity of detoxification enzymes is critical for many herbivorous insects [16, 69] as their diets often contain a suite of plant chemicals that can be toxic, reduce palatability, or slow development time.

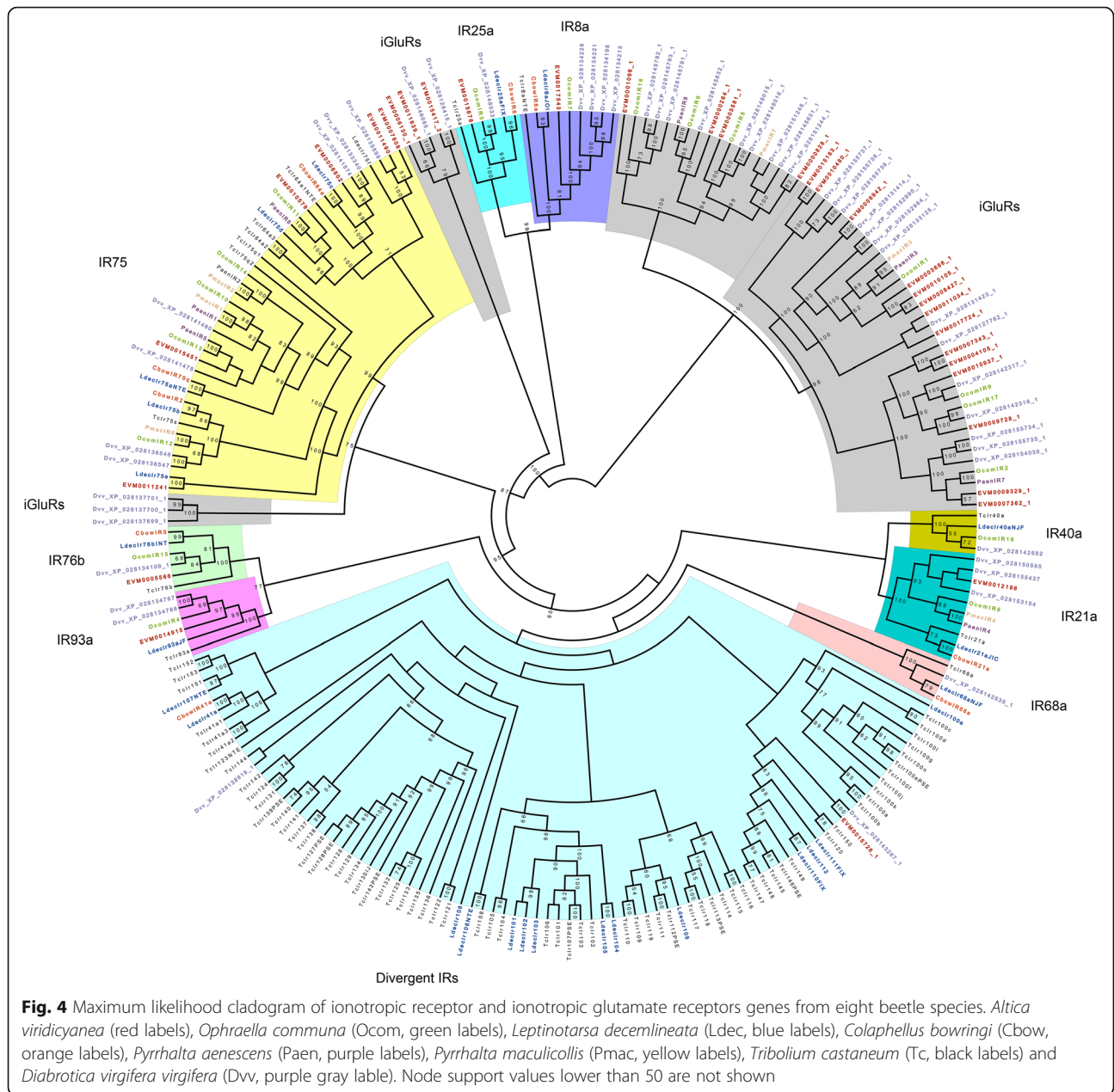
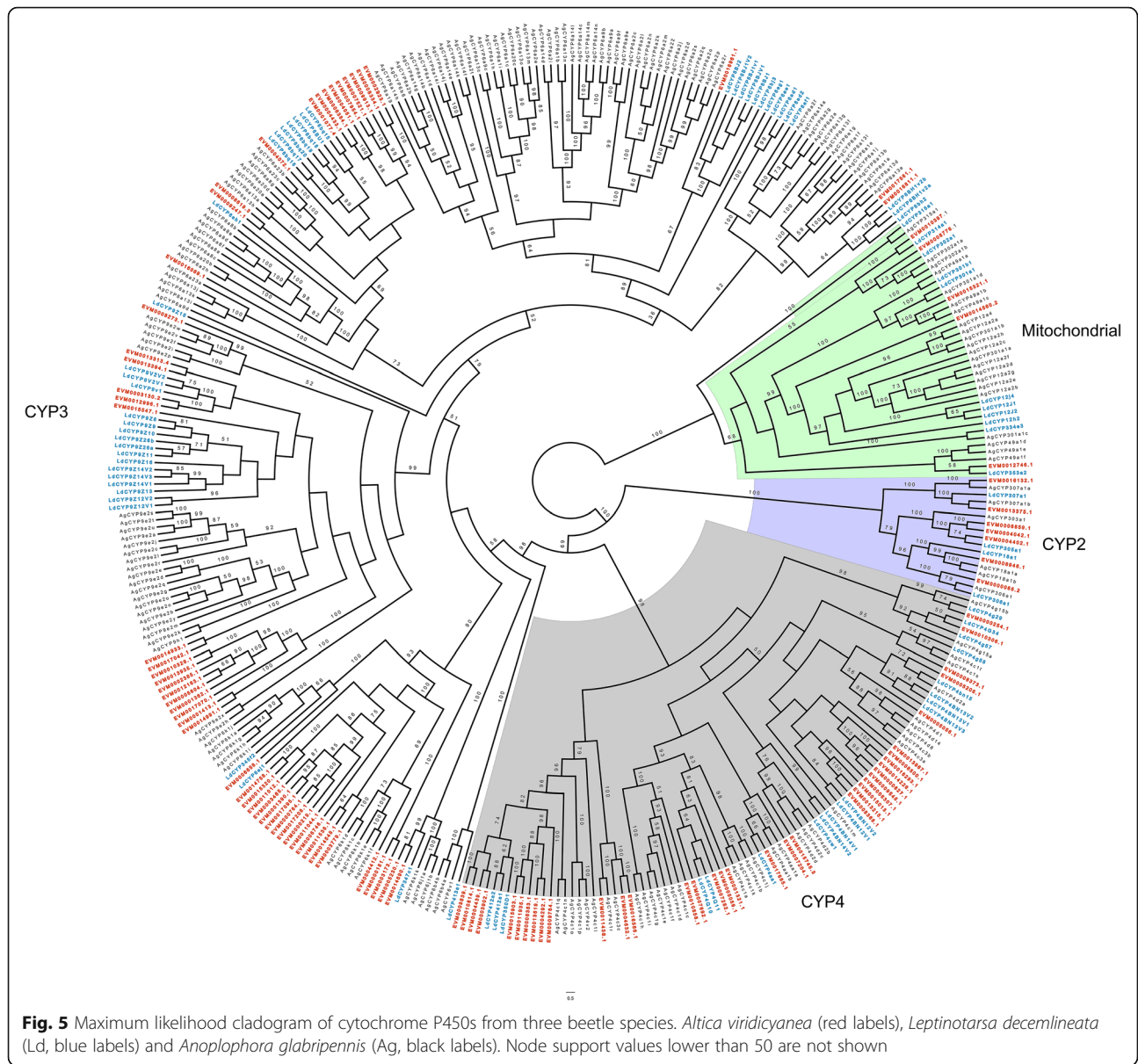


Fig. 4 Maximum likelihood cladogram of ionotropic receptor and ionotropic glutamate receptors genes from eight beetle species. *Altica viridicyanea* (red labels), *Ophraella communa* (Ocom, green labels), *Leptinotarsa decemlineata* (Ldec, blue labels), *Colaphellus bowringi* (Cbow, orange labels), *Pyrrhalta aenescens* (Paen, purple labels), *Pyrrhalta maculicollis* (Pmac, yellow labels), *Tribolium castaneum* (Tc, black labels) and *Diabrotica virgifera virgifera* (Dvv, purple gray label). Node support values lower than 50 are not shown

The host plant of *A. viridicyanea* is *Geranium nepalensis* which has a number of chemical defenses such as tannins, flavonoids and organic acids [70]. As a strict specialist, then, *A. viridicyanea* likely has adaptations that allow them to detoxify these chemicals. Indeed, we annotated 225 detoxification enzymes spanning all three families (101 P450s, 97 CCEs and 27 GSTs). Expansion and contraction of these gene families are considered important in adaptive phenotypic diversification [71]. Furthermore, meta-analyses have established that the size of the P450, CCE and GST gene families are correlated with insect diet breadth [64, 65, 72]. In contrast with these studies, we showed that although *A.*

viridicyanea has a greater number of detoxifying genes than that of the closely-related oligophagous *Leptinotarsa decemlineata* (197 genes) [19, 65], it has fewer detoxifying genes than generalist *T. castaneum* (275 genes) [24, 73].

Insect cytochrome P450 proteins are important in both xenobiotic detoxification and synthesis and degradation of endogenous molecules such as ecdysteroids and juvenile hormone [74–76]. In insects, the cytochrome P450 family is divided into four major clades: the mitochondrial P450 clade, the CYP2 clade, the CYP3 clade, and the CYP4 clade [77]. We found 101 P450s in *Altica viridicyanea* spanning all four clades: five in the



mitochondrial clade, seven in the CYP2 clade, 53 in CYP3 clade, and 36 in CYP4 clade (Fig. 5, Table S6). We found that a majority of these genes belonged to the CYP6 and CYP9 subfamilies of the CYP3 clade and the CYP4 subfamily of the CYP4 clade (Table S7). These P450 subfamilies are known to be involved in detoxification of plant allelochemicals as well as resistance to pesticides [19, 78–80].

In addition to the cytochrome P450s, the *A. viridicyanea* genome also contained 97 genes encoding putative CCEs (Fig. S5), which is slightly fewer than that of *L. decemlineata* (102), but more than that of the other eight beetle species that were included in the analysis (ranged from 44 to 82) [65]. The dietary/detoxification group included two clades: coleopteran xenobiotic

metabolizing CCE (clade A) and α -esterase type CCEs (clade B) [81]. In *A. viridicyanea*, there is a noteworthy expansion (62 genes) in clade A, whereas we did not identify any genes from Clade D (integument esterase), F (juvenile hormone esterase), or I (unknown function) (Fig. S5; Table S8).

Another group of detoxification enzymes that we examined are the GSTs. GSTs are involved in many cellular physiological activities, such as detoxification of endogenous and xenobiotic compounds, intracellular transport, biosynthesis of hormones and protection against oxidative stress [73, 81]. Insect GSTs are divided into two major groups, the cytosolic and the microsomal GST genes. The cytosolic group is further divided into six classes: Delta, Epsilon, Sigma, Omega, Theta, and

Zeta [82]. The Delta and Epsilon classes are thought to be insect-specific [73, 83, 84], and members of the Epsilon subfamily are commonly involved in detoxification of xenobiotics [85]. We detected a total of 27 GST genes in *A. viridicyanea* (Fig. S6; Table S9). Both the total number and the number of detoxification-related Epsilon subfamily in *A. viridicyanea* were lower than that of most beetles [65](Table S9).

UDP-glycosyltransferases (UGT) catalyze the conjugation of a range of diverse small lipophilic compounds with sugars to produce glycosides, playing an important role in the detoxification of xenobiotics and in the regulation of endobiotics in insects [66]. From 17 (*Oryctes borbonicus*) to 65 (*Anoplophora glabripennis*) UGTs were identified in the nine beetle species surveyed [3, 65]. Currently, the largest repertoire of UGTs in beetles was found in the polyphagous longhorn beetle *Anoplophora glabripennis*, with 65 putative UGT genes and 7 pseudogenes [34]. The expansion of UGTs in *A. glabripennis* is thought to be related to its ability to feed on a broad range of host plants [34]. In line with this, we annotated 32 UGTs in the *A. viridicyanea* genome. A number of UGT50s were identified in this species, which has been suggested as the most conserved UGT in insects [66], and we also observed a remarkable expansion in the UGT324 family (Fig. S7, Table S10).

Most ABC proteins engage in active transport of molecules across cell membranes. The ABC transporters are well-known components of various detoxification mechanisms across all phyla [86, 87]. In the present study, we identified 69 putative ABCs in *A. viridicyanea*, belonging to eight subfamilies (A to H). This is a similar number to two specialist species of chrysomelids, *Chrysomela populi* and *Diabrotica virgifera virgifera*, (65 in each based on transcriptomic data) (Table S11). The gene numbers of the conserved subfamilies D, E and F were consistent with other beetles analyzed (Table S11); however, the number of genes in subfamilies B and C in *A. viridicyanea* (46) are the highest among the five species with which we compared (Table S11; Fig. 6). These subfamilies are known to be involved in detoxification processes [62, 88].

Plant cell wall-degrading enzymes

Early views of insect digestion postulated that insects lack the endogenous enzymes required for plant cell wall (PCW) digestion, and that PCW digestion by insects depended on exogenous enzymes from symbiotic microorganisms [89]. Recent studies, however, have revealed that endogenous PCW degrading enzymes are present in many insects and are important in the digestion of cellulose, hemicelluloses, and pectin in PCW [38, 90]. In fact, these enzymes are likely a key innovation in the adaptive radiation of herbivorous beetles. Some insect PCW-

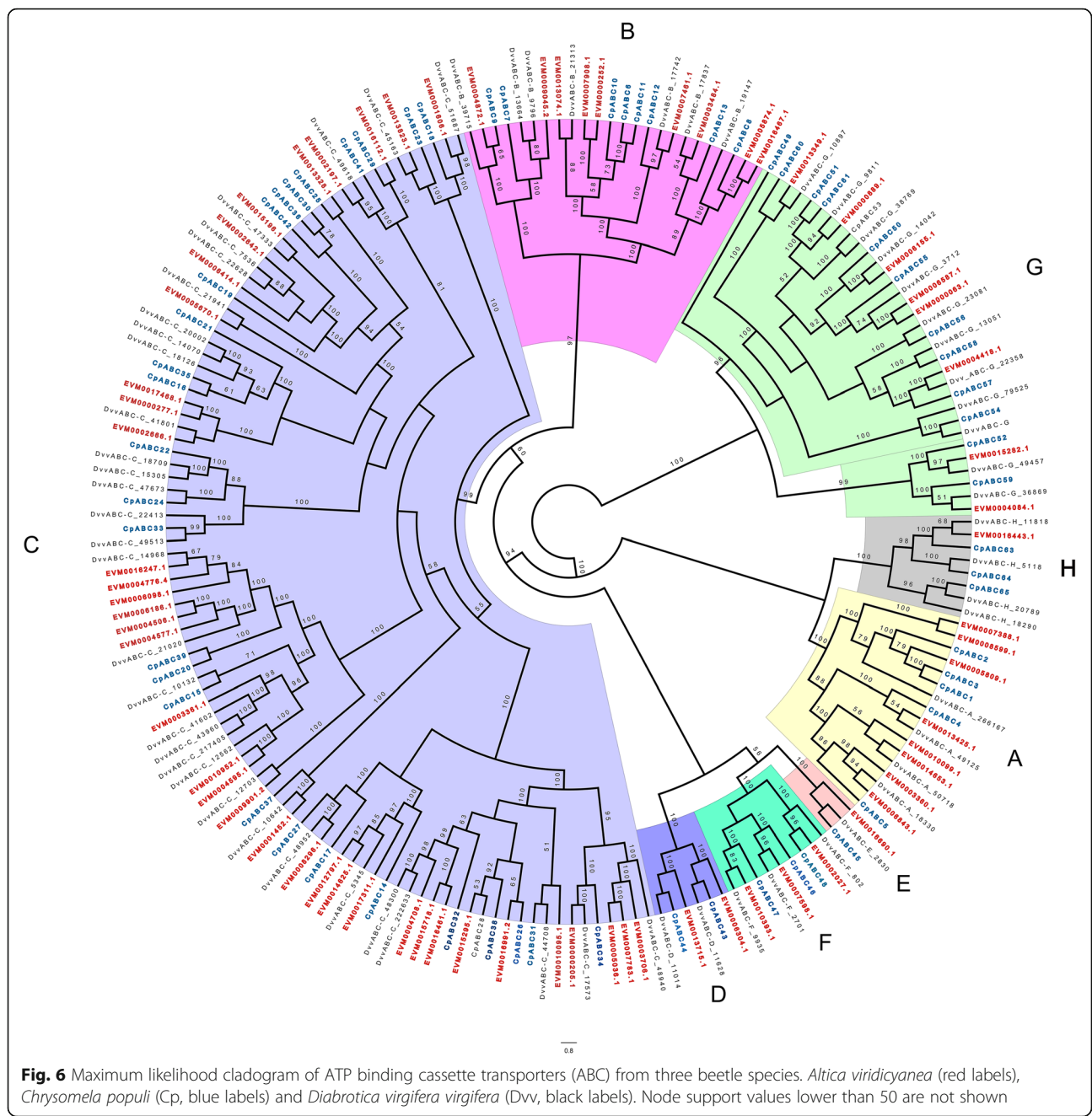
degrading enzymes are also involved in immune-defense responses and detoxification [38, 90].

Beetle-encoded plant cell wall-degrading enzymes are carbohydrate esterases (CE), polysaccharide lyases (PL), and mainly glycoside hydrolases (GH) [38]. In *A. viridicyanea*, we identified 65 putative glycoside hydrolases, including 35 GH1 genes, 10 GH45 genes, two GH48 genes and 18 GH28 genes (Table S12). Genes of GH1 originated anciently in animals and are ubiquitous in beetles. The species of Phytophaga (i.e., Chrysomeloidea and Curculionoidea) examined thus far have a greater number of GH1 genes than *A. viridicyanea*, for example, 228 were found in *Diabrotica undecimpunctata*, 135 in *Mastostethus salvini*, and 136 in *Rhynchitomacerinus kuscheli* [19, 34, 38]. For another ancient and ubiquitous gene family, GH9, there are at least a dozen independent losses in beetles [38]. GH9 was not detected in *A. viridicyanea*, along with 4 of 7 other chrysomelid species (*Callosobruchus maculatus*, *Donacia marginata*, *Diabrotica undecimpunctata* and *Leptinotarsa decemlineata*) [19, 38].

The other plant cell wall-degrading gene families (CE8, PL4, GH32, GH5, GH10, GH43, GH44, GH45, GH48 and GH28) are suggested to be obtained from bacteria and fungi via horizontal gene transfer, and are mainly found in Buprestoidea and Phytophaga, with scattered genes in a few other taxa [38]. In *A. viridicyanea*, in addition to the ubiquitous GH1 genes, three families of PCW-degrading enzymes were identified, including cellulose degrading GH45 and GH48, and pectin degrading GH28. These observed gene numbers are similar to that of closely related species in the Chrysomelinae (*Oreina cacaliae* and *Leptinotarsa decemlineata*) and Galerucinae (*Diabrotica undecimpunctata*) [19, 38].

Conclusions

In the present study, we combined long reads of PacBio with the higher fidelity of the short reads generated with Illumina sequencing. From the genome annotation we found that *A. viridicyanea*, a host specialist herbivore, has a reduced number of chemosensory and detoxification genes as compared to more generalist herbivorous beetles, consistent with the idea that diet breadth should positively correlate with chemosensory and detoxification gene content. Although *A. viridicyanea* had fewer chemosensory and detoxification genes than more polyphagous beetles, we did observe expansions in some gene families that may be related to host plant adaptation. As a result, the genome assembled here provides an important resource for further studies on host plant adaptation and functionally affiliated genes. Additionally, this work will also open the opportunity for comparative genomics studies among closely related *Altica* species



that may provide insights into the molecular evolutionary processes that occur during ecological speciation.

Methods

Beetles

Altica viridicyanea is a highly specialized herbivore of *Geranium nepalense*. They are elongate-ovate beetles with a length of 3–4 mm. The dorsal surface is black with a metallic blue reflection (Fig. 1). Equipped with dilated hind legs, these beetles typically jump in a flea-like fashion to escape predators.

To make genome sequencing easier, we created a laboratory colony by collecting adult *A. viridicyanea* in Changping (40.28°N, 116.05°E), Beijing, China. Adults were maintained in growth chambers held at 16:8 LD and 25 °C and fed with leaves of their host plant, *Geranium nepalense* (Sweet). A subset of these collected adults was used for genome size estimation (see below). We maintained the colony through successive single-pair sibling matings to create third generation lines for whole genome sequencing. When we were ready to sequence the beetles, beetles were starved for 2 days and

then killed in liquid nitrogen. The samples were stored at -80°C until DNA extraction. In total, we used 157 virgin females for sequencing.

Genome size estimation

In preparation for whole genome sequencing, we first determined the genome size of *A. viridicyanea*. Genome size was estimated via flow cytometry [91] on four adult males and four adult females. The thoracic muscle of living beetles was dissected with sterilized fine forceps, and cut into small pieces in Galbraith buffer. The ground suspension was filtered through 40-mm nylon mesh (Easystainer™) to remove cellular debris and the flow-through was collected in a 5-ml round-bottomed tube placed on ice. Propidium iodide was added to samples to a final concentration of 50 mg/ml and stained in the dark at 4°C for 2 h. Then fluorescence intensity was estimated for each beetle using a superfluid cell sorting system (MoFol XDP, Beckman Coulter Life Sciences). Genome size was calculated by comparing samples to an internal reference subsisting of chicken blood.

DNA extraction and sequencing

Genomic DNA (gDNA) was extracted from whole beetles using DNeasy Blood & Tissue Kit (Qiagen, Germany) following the manufacturer's instructions. Samples were first surface sterilized using 75% ethanol and sterile deionized water. The total amount of gDNA was measured using Qubit Fluorometer (Invitrogen), and the integrity of the gDNA was verified on an agarose gel that had reference lanes containing high molecular weight ladders (GeneRuler High Range DNA Ladder and D2000 DNA Ladder). From these extractions, seven Illumina sequencing libraries were prepared, with insert sizes and genome coverage of 270 bp (55.5 \times), 500 bp (29.4 \times), 800 bp (19.0 \times), 3 kb (12.5 \times), 5 kb (two libraries, 49.7 \times), to 10 kb (21.1 \times) (Table S13). The libraries were sequenced with 150 bp paired-end reads on the Illumina HiSeq 2500 platform. We also sequenced two PacBio libraries with 72.7 \times genome coverage, with N50 read lengths of 9.1 kb (Table S13). The first PacBio DNA library (20 kb) was constructed using the PacBio SMRTbell Template Prep Kit 3.0 (Pacific Biosciences, Menlo Park, CA, USA) and sequenced on a PacBio RS II sequencer with the P6 polymerase/C4 chemistry combination at 1GENE (Zhejiang, China). A total of 16 SMRT Cells were processed and the movie length was 6 h. The second PacBio DNA library (20 kb) was constructed using the SMRTBell template preparation kit 1.0 (PacBio, USA), for which six SMRT Cells were run on the PacBio Sequel instrument at BGI (Guangdong, China) with Sequel™ Sequencing Kit 2.1 (PacBio, USA). The movie length was 10 h.

RNA-seq library construction and sequencing

We produced transcriptomic data to facilitate gene prediction analyses. To obtain more full-length or near full-length gene sequences, we generated sequence data by combining PacBio full-length transcriptome sequencing (8 Gb clean data) and Illumina sequencing (10 Gb clean data). Whole bodies of four males and four females with three libraries (size-selection: 1–2 kb, 2–3 kb and > 3 kb lengths) were used for PacBio sequencing. The PacBio transcriptome libraries were constructed using PacBio SMRTbell Template Prep Kit 3.0 (Pacific Biosciences, Menlo Park, CA, USA) and sequenced on a PacBio RS II sequencer at 1GENE (Zhejiang, China). The sequencing chemistry was P6-C4, and the movie length was 4 h. Five SMRT Cells were processed: one for the 1–2 kb library, two for the 2–3 kb library and two for the > 3 kb library. To enrich for chemosensory genes, heads from 30 females and 30 males were also used for Illumina sequencing.

Data pre-processing

For Illumina Data, FastQC v0.11.6 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to check the quality of the raw reads under default settings, and then TrimGalore v0.4.5 (<https://github.com/FelixKrueger/TrimGalore>) was used to filter out the adapters and low-quality reads using the parameters “--length 36 -q 20 --trim-n”. Because *A. viridicyanea* are quite small, we used the whole body of adults for sequencing. Although we surface sterilized the insects, the likelihood of microbial contamination from gut contents was high; therefore, we used BBDuk in the BBDuk v37.80 (<https://sourceforge.net/projects/bbmap>) with parameters “ordered = t k = 31” to filter microbial contaminants. To do this, we first built a reference library which included all sequences of archaea, bacteria, fungi, protozoa, and viruses available in RefSeq (<https://ftp.ncbi.nlm.nih.gov/genomes/refseq/>) (accessed on August 15, 2019). The mitochondrion sequences of *A. viridicyanea* (GenBank accession numbers: MH477594, MH477596, MH477598 and MH477599) [92] were also included to remove reads originating from the mitochondrial genome. After these filtering steps, the remaining reads were used for downstream genome assembly and analysis.

To simplify the genome assembly process, we also excluded reads originating from microbial sources that were present in the PacBio data. To accomplish this, we mapped all the raw PacBio reads to the above Illumina library using minimap2 v2.12 [93], using the parameters “-x map-pb -a -Q --split-prefix”. To improve the mapping specificity, we included 11 available coleopteran genomes [26] in the reference library. Then we removed the reads that mapped to the microbial sequences.

For the PacBio Iso-Seq data processing, we used the new Iso-Seq3 pipeline (https://github.com/PacificBiosciences/IsoSeq_SA3nUP). First, we first converted the bax.h5 files of each SMRT Cell run to the BAM format using bax2bam v0.0.8 (<https://github.com/PacificBiosciences/bax2bam>) with default parameters. Circular Consensus Sequence (CCS) calling was then done by ccs v3.4.1 (<https://github.com/PacificBiosciences/ccs>) under default settings. To remove cDNA primers and obtain full-length reads, we used lima v2.0.0 (<https://github.com/pacificbiosciences/barcoding>) with parameters “--isoseq --peek-guess”. Next we used “isoseq3 refine --require-polya” to trim the poly (A) tails. Finally, to generate polished, high-quality reads, the SMRT cells were merged and we used “isoseq3 cluster” with parameters “--use-qvs”.

Heterozygosity estimation

Heterozygosity was estimated using gce v1.0.0 (parameters: -b 1 -H 1 -m 1 -D 8), based on the pooled Illumina reads of the sib-mated females. Before running gce, we first used kmer_freq_hash in the gce package to compute the histogram of kmer frequencies (k-mer from 17 to 25).

Genome assembly

A hybrid assembly strategy was used to assemble the PacBio and Illumina reads. PacBio reads were first assembled using Flye v2.3.5b [94] with the parameters set to “--genome-size 830 m” to obtain contigs. Then Redundans v0.14a [95], which uses an iterative scaffolding approach by calling SSPACE [96], was used to scaffold the PacBio contigs with the Illumina short reads, with parameters set to “--noreduction”. After that, SSPACE-Long v1.1 [97] was used to further scaffold with PacBio long reads using default settings (Table S14).

In addition, we used these PacBio RNA-seq data to evaluate the genome assembly. We mapped PacBio RNA-seq reads to the genome using minimap2 v2.12, with parameters “-ax splice -uf --secondary=no”. For those unmapped and poorly mapped transcripts (coverage < 0.99 or identity < 0.95), the Cogent pipeline (<https://github.com/Magdoll/Cogent>) was employed to explore their potential functions. Transcripts were partitioned into gene families based on k-mer similarity. For each gene family, contigs were reconstructed to represent the “union” of all coding bases. To check the potential functions of these transcripts, we used Blastn v2.7.1 [98] against the non-redundant sequence database (<ftp://ftp.ncbi.nih.gov/blast/db/FASTA>; accessed at November 2020).

To evaluate other potential sources of contamination, we examined the long reads discarded during QC to identify their origins. Kraken2 v2.1.1 was used for fast

classification of these long reads, with a standard database downloaded from <https://benlangmead.github.io/aws-indexes/k2>. The output was uploaded to the pavian (<https://breitwieser.shinyapps.io/pavian/>) web server to obtain a report.

Estimation of genome completeness

Benchmarking Universal Single-Copy Orthologs (BUSCO v3.0.1) [99] analyses against Insecta ($n = 1658$) and Endopterygota ($n = 2442$) were used to evaluate the integrity and quality of the genome assembly with parameters “-sp tribolium2012 -m geno”.

In addition, to characterize the partial genes in our assembly, we chose three beetle species to compare the length of proteins (*Tribolium castaneum*, *Dendroctonus ponderosae*, and *Anoplophora glabripennis*) since the N50 metric of these genomes was relatively high. We downloaded the latest genome annotations for these beetles from RefSeq (Tcas 5.2, DendPond male_1.0 and Agla 2.0), then Blastp v2.7.1 was employed to compare the sequences of the annotated *A. viridicyanea* proteins with proteins from these beetles. Focusing on only the top hit of each query sequence, we compared the lengths between query sequences and subject sequences.

Genome annotation

Prior to gene prediction using the assembled sequences, we identified repeat sequences in the genome of *A. viridicyanea*. LTR FINDER v1.05 [100], RepeatScout v1.0.5 [101] and PILER-DF v2.4 [102] were used to construct a library of repetitive sequences based on the *A. viridicyanea* genome. PASTEClassifier v1.0 [103] was used to classify these repeats, and the Repbase database [104] was used to merge them. Finally, RepeatMasker v4.0.5 [105] was used to identify and mask the genomic repeated sequences. All the parameters were set to “default” except RepeatMasker where the parameters were set to “-nolow -no_is -norna -engine wublast”. To check whether the PacBio reads could span most of the repeats (transposons here), we aligned the PacBio reads to the assembled genome using minimap2 v2.12 (parameters: -x map-pb). Then Bedtools v2.26.0 [106] was employed to get the read coverage of each repeat.

We combined the de novo, homology-based, and transcriptome-based predictions to identify protein-coding genes in the *A. viridicyanea* genome. The default settings for Genscan v1.1.0 [107], Augustus v2.4 [108], GlimmerHMM v3.0.4 [109], GeneID v1.4 [110] and SNAP (v2006-07-28) [111] were used to generate the de novo predictions. For the homology-based prediction, protein sequences of *Drosophila melanogaster*, *Tribolium castaneum*, *Dendroctonus ponderosae* and *Anoplophora glabripennis* were downloaded from NCBI and aligned to the assembled *A. viridicyanea* genome with

GeMoMa v1.3.1 [112, 113] using default parameters. For transcriptome sequencing-based prediction, we firstly assembled the Illumina short reads into unigenes using Hisat v2.0.4 with parameters “--max-intronlen 20000, --min-intronlen 20” [114] and StringTie v1.3.4 set to default parameters [114]. Second, we combined the unigenes with the PacBio full-length transcripts, then we predicted genes based on these combined sequences with PASA v2.0.2 using the parameters “-align_tools gma, -maxIntronLen 20000” [115]. EvidenceModeler v1.1.1 was used to integrate all of the gene predictions [116]. Finally, to obtain a final gene dataset, PASA v2.0.2 was used for further modification including adding 5'-UTR and 3'-UTRs, obtaining alternative splicing, and extending or shortening, adding or subtracting exons for the EVM-predicted results.

We predicted the non-coding RNAs based on the Rfam v12.1 [117] and miRBase v21.0 [118] databases. Putative microRNAs (miRNAs) and ribosomal RNAs (rRNAs) were predicted using Infernal v1.1 [119], and transfer RNAs (tRNAs) were predicted with tRNAscan-SE v1.3.1 [120]. Based on homology to known protein-coding genes, putative pseudogenes were first searched in the intergenic regions of the *A. viridicyanea* genome using genBlastA v1.0.4 [121]. Then GeneWise v2.4.1 [122] was used to search for premature stop codons or frameshift mutations in those sequences.

The predicted genes were annotated by aligning them to the NCBI non-redundant protein (nr) [123], non-redundant nucleotide (nt) [123], Swissprot [124], TrEMBL [124], KOG [125], and KEGG [126] databases using BLAST v2.2.31 [127] with a maximal e-value of $1e^{-5}$. We assigned gene ontology (GO) terms [128] to the genes using the BLAST2GO v2.5 pipeline [129].

Species phylogenetic analysis

We estimated the phylogenetic relationships of *A. viridicyanea* and an additional nine representative beetle species for which genomic data were available (*Anoplophora glabripennis*, *Aethina tumida*, *Agrilus planipennis*, *Dendroctonus ponderosae*, *Diabrotica virgifera virgifera*, *Leptinotarsa decemlineata*, *Nicrophorus vespilloides*, *Onthophagus taurus* and *Tribolium castaneum*). We used 1751 conserved single copy orthologs for this analysis. The analysis was implemented in PhyML v3.0 [130] with parameters “-gapRatio 0.5 -badRatio 0.25 -bootstrap 1000”.

We also estimated divergence times among the species using MCMCTree (PAML v4.8 package) (<http://abacus.gene.ucl.ac.uk/software/paml.html>) [131, 132] with parameters set to a burn-in time of 10,000, the sample number was 100,000, and the sampling frequency was 2. We used the time divergence data in timetree (<http://www.timetree.org/>) for calibration (*Nicrophorus*

vespilloides - *Onthophagus taurus* [245 ~ 296 MYA]; *Dendroctonus ponderosae* - *Anoplophora glabripennis* [149 ~ 240 MYA]). To examine gene family expansion and contraction among species, we used CAFE v4.1 [133] with “lambda -l 0.002” to automatically search for birth and death parameters (λ) of genes.

Gene family evolution

To provide insights into host plant specialization, we specifically focused on the evolution of chemosensory gene families, chemical detoxification supergene families, and plant cell wall-degrading enzymes. We first conducted multiple sequence alignment using Mafft (online version 7.305) [134] (<http://mafft.cbrc.jp/alignment/server/>), applying the L-INS-I algorithm. The best-fit models for amino acid sequence evolution were selected using the Akaike Information Criterion (AIC) in Prottest v3.4.2 [135]. (Table S15). Maximum likelihood trees were reconstructed using RaxML v8.2.9 [136] in the CIPRES Science Gateway v3.3 (<https://www.phylo.org/>) [137], with 1000 non-parametric bootstrap replicates. The resulting tree was viewed and edited with FigTree v1.3.1 [138] and Adobe Illustrator CS5.

Abbreviations

ABC: ATP binding cassette transporter; bp: Base pairs; BUSCO: Benchmarking Universal Single-Copy Orthologs; CCEs: Carboxyl/cholinesterase; CE: Carbohydrate esterase; CSP: Chemosensory protein; Gb: Gigabase pairs; GH: Glycoside hydrolase; GR: Gustatory receptor; GST: Glutathione S-transferase; IR: Ionotropic receptor; kb: Kilobase pairs; LINE: Long interspersed nuclear element; LTR: Long terminal repeat; Mb: Megabase pairs; ML: Maximum likelihood; OBP: Odorant binding protein; OR: Odorant receptor; P450: Cytochrome P450; PCW: Plant cell wall; PL: Polysaccharide lyase; RAXML: Randomized Axelerated Maximum Likelihood; RNAseq: RNA sequencing; SINE: Short interspersed nuclear element; TE: Transposable element; UGT: UDP-glucuronosyltransferase

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07558-6>.

Additional file 1: Figure S1. Venn diagram indicating the number of protein-coding genes of *A. viridicyanea* based on ab initio, RNA-seq-based and homology-based gene prediction methods. **Figure S2.** Maximum likelihood cladogram of gustatory receptor genes from eight beetle species: *Altica viridicyanea* (red labels), *Ophraella communa* (Ocom, dark blue labels), *Colaphellus bowringi* (Cbow, green labels), *Pyrrhalta aenescens* (Paen, purple labels), *Pyrrhalta maculicollis* (Pmac, yellow labels), *Dendroctonus ponderosae* (Dpon, pale blue labels), *Leptinotarsa decemlineata* (Ldec, black labels) and *Diabrotica virgifera virgifera* (Dvv, grey labels). Node support values lower than 50 are not shown. **Figure S3.** Maximum likelihood cladogram of odorant binding proteins from eight beetle species: *Altica viridicyanea* (red labels), *Ophraella communa* (Ocom, yellow labels), *Leptinotarsa decemlineata* (Ldec, green labels), *Colaphellus bowringi* (Cb, pale blue labels), *Pyrrhalta aenescens* (Paen, dark blue labels), *Pyrrhalta maculicollis* (Pmac, orange labels), *Dendroctonus ponderosae* (Dpon, black labels) and *Diabrotica virgifera virgifera* (Dvv, gray labels). Node support values lower than 50 are not shown. **Figure S4.** Maximum likelihood cladogram of chemosensory proteins from eight beetle species: *Altica viridicyanea* (red labels), *Ophraella communa* (Ocom, pale blue labels), *Colaphellus bowringi* (Cbow, purple labels), *Pyrrhalta aenescens* (Paen, orange labels), *Pyrrhalta maculicollis* (Pmac, dark blue labels),

Dendroctonus ponderosae (Dpon, yellow labels), *Anoplophora glabripennis* (Agl, green labels) and *Tribolium castaneum* (Tcas, black labels). Node support values lower than 50 are not shown. **Figure S5.** Maximum likelihood cladogram of carboxyl/cholinesterases from four beetle species. *Altica viridicyanea* (red labels), *Leptinotarsa decemlineata* (Ldec, yellow labels), *Aethina tumida* (At, blue labels) and *Tribolium castaneum* (Tc, black labels). Node support values lower than 50 are not shown. **Figure S6.** Maximum likelihood cladogram of glutathione S-transferases from four beetle species. *Altica viridicyanea* (red labels), *Leptinotarsa decemlineata* (Ld, blue labels), *Aethina tumida* (Atum, yellow labels) and *Tribolium castaneum* (Tc, black labels). Node support values lower than 50 are not shown. **Figure S7.** Maximum likelihood cladogram of UDP-glucuronosyltransferases from three beetle species. *Altica viridicyanea* (red labels), *Anoplophora glabripennis* (Agl, blue labels) and *Tribolium castaneum* (Tcas, black labels). Node support values lower than 50 are not shown.

Additional file 2: Table S1. Published beetle genomes. **Table S2.** Statistics of gene annotation for *Altica viridicyanea* sourced from different databases. **Table S3.** Pseudogenes identified in the genome of *Altica viridicyanea*. **Table S4.** Gene family classification for the beetle species used in the present study. The incomplete genes were excluded from the analysis. **Table S5.** Comparison of number of chemosensory functional genes (+pseudogenes) in *Altica viridicyanea* and other beetle species. **Table S6.** Comparison of the number of CYP450 genes in *Altica viridicyanea* and other beetle species. **Table S7.** Cytochrome P450 genes in *Altica viridicyanea*. **Table S8.** Comparison of number of carboxyl/cholinesterases (CCEs) in *Altica viridicyanea* and other beetle species. **Table S9.** Comparison of the number of glutathione S-transferases (GSTs) in *Altica viridicyanea* and other beetle species. **Table S10.** UDP-glucuronosyltransferases (UGTs) (+pseudogenes) in *Altica viridicyanea* and other beetle species. **Table S11.** Comparison of the number of ABC genes in *Altica viridicyanea* and other beetle species. **Table S12.** Plant cell wall degrading enzymes identified in *Altica viridicyanea* and other beetle species. **Table S13.** Summary of library construction and sequencing of *Altica viridicyanea*. **Table S14.** Summary of de novo assemblies of the *Altica viridicyanea* draft genome. **Table S15.** The best-fit models for amino acid sequence evolution were selected using the Akaike Information Criterion (AIC) in Prottest v3.4.2.

Acknowledgements

Not applicable.

Authors' contributions

H.J.X., X.K.Y. and R.S.C. designed and managed this project. H.J.X. and W.Z.L. maintained the beetle colony used for sequencing. Y.W.N., Y.J.H., L.L.Z., X.C.C., X.W.Z. and H.J.X. performed genome assembly and genome annotation. H.J.X., R.E.N. and W.Z.L. performed the phylogenetic analysis. H.J.X., K.A.S. and Y.W.N. wrote the paper. All authors read and approved the final manuscript.

Funding

This study was supported by National Natural Science Foundation of China (Grant Nos. 31472030, 31672334).

Availability of data and materials

The raw sequence data reported in this paper are available at the Genome Sequence Archive [139] in National Genomics Data Center [140] (<https://bigd.big.ac.cn/gsa>; accession number: CRA002741). And the whole genome sequence data have been deposited in the Genome Warehouse in the National Genomics Data Center [140] (<https://bigd.big.ac.cn/gwh>; accession number: GWHAMMQ00000000).

Declarations

Ethics approval and consent to participate

This study was carried out in full compliance of the laws of China. The study species is not included in the "List of Protected Animals in China" and no specific permits were required for sampling and experiments in this study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. ²College of Life Sciences, Nankai University, Tianjin 300071, China. ³Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. ⁴University of Chinese Academy of Sciences, Beijing 100049, China. ⁵Department of Biology, Syracuse University, 107 College Place, Syracuse, NY 13244, USA. ⁶Archbold Biological Station, 123 Main Drive, Venus, FL 33960, USA. ⁷Biomarker Technologies Corporation, Floor 8, Shunjie Building, 12 Fuqian Road, Nanfaxin Town, Shunyi District, Beijing 101300, China.

Received: 21 July 2020 Accepted: 25 March 2021

Published online: 07 April 2021

References

- Forbes AA, Devine SN, Hippee AC, Tvedte ES, Ward AKG, Widmayer HA, et al. Revisiting the particular role of host shifts in initiating insect speciation. *Evolution*. 2017;71(5):1126–37. <https://doi.org/10.1111/evo.13164>.
- Nosil P, Crespi BJ, Sandoval CP. Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature*. 2002;417(6887):440–3. <https://doi.org/10.1038/417440a>.
- Simon J-C, d'Alençon E, Guy E, Jacquin-Joly E, Jaquière J, Nouhaud P, et al. Genomics of adaptation to host-plants in herbivorous insects. *Brief Funct Genomics*. 2015;14(6):413–23. <https://doi.org/10.1093/bfgp/elt015>.
- Xue HJ, Li WZ, Yang XK. Genetic analysis of feeding preference in two related species of *Altica* (Coleoptera: Chrysomelidae: Alticinae). *Ecol Entomol*. 2009;34(1):74–80. <https://doi.org/10.1111/j.1365-2311.2008.01042.x>.
- Xue HJ, Magalhães S, Li WZ, Yang XK. Reproductive barriers between two sympatric beetle species specialized on different host plants. *J Evol Biol*. 2009;22(11):2258–66. <https://doi.org/10.1111/j.1420-9101.2009.01841.x>.
- Xue HJ, Li WZ, Nie RE, Yang XK. Recent speciation in three closely related sympatric specialists: inferences using multi-locus sequence, post-mating isolation and endosymbiont data. *PLoS One*. 2011;6(11):e27834. <https://doi.org/10.1371/journal.pone.0027834>.
- Xue HJ, Li WZ, Yang XK. Assortative mating between two sympatric closely-related specialists: inferred from molecular phylogenetic analysis and behavioral data. *Sci Rep*. 2014;4:5436. <https://doi.org/10.1038/srep05436>.
- Xue HJ, Wei JN, Magalhães S, Zhang B, Song KQ, Liu J, et al. Contact pheromones of 2 sympatric beetle species are modified by the host plant and affect mating. *Behav Ecol*. 2016;27(3):895–902. <https://doi.org/10.1093/beheco/arv238>.
- Xue HJ, Zhang B, Segraves KA, Wei JN, Nie RE, Song KQ, et al. Contact cuticular hydrocarbons act as a mating cue to discriminate intraspecific variation in *Altica* flea beetles. *Anim Behav*. 2016;111:217–24. <https://doi.org/10.1016/j.anbehav.2015.10.025>.
- Xue HJ, Segraves KA, Wei J, Zhang B, Nie RE, Li WZ, et al. Chemically mediated sexual signals restrict hybrid speciation in a flea beetle. *Behav Ecol*. 2018;29(6):1462–71. <https://doi.org/10.1093/beheco/ary105>.
- Laroche A, DeClerck-Floate RA, LeSage L, Floate KD, Demeke T, et al. Are *Altica carduorum* and *Altica cirsiicola* (Coleoptera: Chrysomelidae) different species? Implications for the release of *A. cirsiicola* for the biocontrol of Canada thistle in Canada. *Biol Control*. 1996;6(3):306–14. <https://doi.org/10.1006/bcon.1996.0039>.
- Jenkins TM, Braman SK, Chen Z, Eaton TD, Pettis GV, Boyd DW, et al. Insights into flea beetle (Coleoptera: Chrysomelidae: Galerucinae) host specificity from concordant mitochondrial and nuclear DNA phylogenies. *Ann Entomol Soc Am*. 2009;102(3):386–95. <https://doi.org/10.1603/008.102.0306>.
- Reid CA, Beatson M. Disentangling a taxonomic nightmare: a revision of the Australian, Indomalayan and Pacific species of *Altica* Geoffroy, 1762 (Coleoptera: Chrysomelidae: Galerucinae). *Zootaxa*. 2009;3918(4):503–51. <https://doi.org/10.11646/zootaxa.3918.4.3>.
- Bernays EA, Chapman RF. Host-plant selection by Phytophagous insects. New York: Chapman and Hall; 1994. <https://doi.org/10.1007/b102508>.
- Gassmann AJ, Levy A, Tran T, Futuyma DJ. Adaptations of an insect to a novel host plant: a phylogenetic approach. *Funct Ecol*. 2006;20(3):478–85. <https://doi.org/10.1111/j.1365-2435.2006.01118.x>.
- Heidel-Fischer HM, Vogel H. 2015. Molecular mechanisms of insect adaptation to plant secondary compounds. *Curr Opin Insect Sci*. 2015;8:8–14. <https://doi.org/10.1016/j.cois.2015.02.004>.

17. Leschen RAB, Beutel RG. Handbook of zoology, band 4: Arthropoda: Insecta, Teilband / part 40: Coleoptera, beetles, morphology and systematics (Phytophaga), vol. 3. Berlin: Walter de Gruyter; 2014.
18. Jolivet PH, Cox ML, Petitpierre E. Novel aspects of the biology of Chrysomelidae. Dordrecht/Boston/London: Kluwer Academic Publishers; 1994. <https://doi.org/10.1007/978-94-011-1781-4>.
19. Schoville SD, Chen YH, Andersson MN, Benoit JB, Bhandari A, Bowsher JH, et al. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). Sci Rep. 2018;8(1):1931. <https://doi.org/10.1038/s41598-018-20154-1>.
20. Sayadi A, Barrio AM, Immonen E, Dainat J, Berger D, Tellgren-Roth C, et al. The genomic footprint of sexual conflict. Nat Ecol Evol. 2019;3(12):1725–30. <https://doi.org/10.1038/s41559-019-1041-9>.
21. Sarah B, Laurent F, Heinz M-S. Genome assembly of the ragweed leaf beetle: a step forward to better predict rapid evolution of a weed biocontrol agent to environmental novelties. Genome Biol Evol. 2020;12(7):1167–73. <https://doi.org/10.1093/gbe/evaa102>.
22. Fu XH, Li JJ, Tian Y, Quan WP, Zhang S, Liu Q, et al. Long-read sequence assembly of the firefly *Pyrocoelia pectoralis* genome. GigaScience. 2017;6(12):1–7. <https://doi.org/10.1093/gigascience/gix112>.
23. Ando T, Matsuda T, Goto K, Hara K, Ito A, Hirata J, et al. Repeated inversions within a panner intron drive diversification of intraspecific colour patterns of ladybird beetles. Nat Commun. 2018;9(1):3843. <https://doi.org/10.1038/s41467-018-06116-1>.
24. Evans JD, McKenna D, Scully E, Cook SC, Dainat B, Egekwu N, et al. Genome of the small hive beetle (*Aethina tumida*, Coleoptera: Nitidulidae), a worldwide parasite of social bee colonies, provides insights into detoxification and herbivory. GigaScience. 2018;7:1–16. <https://doi.org/10.1093/gigascience/giy138>.
25. Fallon TR, Lower SE, Chang CH, Bessho-Uehara M, Martin GJ, Bewick AJ, et al. Firefly genomes illuminate parallel origins of bioluminescence in beetles. eLife. 2018;7:e36495. <https://doi.org/10.7554/eLife.36495>.
26. McKenna DD. Beetle genomes in the 21st century: prospects, progress and priorities. Curr Opin Insect Sci. 2018;25:76–82. <https://doi.org/10.1016/j.cois.2017.12.002>.
27. Wu YM, Li J, Chen XS. Draft genomes of two blister beetles *Hycleus cichorii* and *Hycleus phaleratus*. GigaScience. 2018;7(3):1–7. <https://doi.org/10.1093/gigascience/giy006>.
28. Kraaijeveld K, Neleman P, Mariën J, de Meijer E, Ellers J. Genomic resources for *Goniozus legneri*, *Aleochara bilineata* and *Paykullia maculata*, representing three independent origins of the parasitoid lifestyle in insects. G3-Genes Genom Genet. 2019;9:987–91. <https://doi.org/10.1534/g3.119.300584>.
29. Wang K, Li PP, Gao YY, Liu CQ, Wang QL, Yin J, et al. *De novo* genome assembly of the white-spotted flower chafer (*Protaetia brevitarsis*). GigaScience. 2019;8:1–9. <https://doi.org/10.1093/gigascience/giz019>.
30. Guan DL, Hao XQ, Mi D, Peng J, Li Y, Xie JY, et al. Draft genome of a blister beetle *Mylabris aulica*. Front Genet. 2020;10:1281. <https://doi.org/10.3389/fgene.2019.01281>.
31. Zhang LJ, Li S, Luo JY, Du P, Wu LK, Li YR, et al. Chromosome-level genome assembly of the predator *Propylea japonica* to understand its tolerance to insecticides and high temperatures. Mol Ecol Resour. 2020;20(1):292–307. <https://doi.org/10.1111/1755-0998.13100>.
32. Hammond PM. Species inventory. In: Groombridge B, editor. Global biodiversity, status of the Earth's living resources. London: Chapman and Hall; 1992. p. 17–39.
33. Slipinski SA, Leschen RAB, Lawrence JF. Order Coleoptera Linnaeus, 1758. In: Zhang Z-Q, editor. Animal biodiversity: an outline of higher-level classification and survey of taxonomic richness, Zootaxa, vol. 3148; 2011. p. 203–8.
34. McKenna DD, Scully ED, Pauchet Y, Hoover K, Kirsch R, Geib SM, et al. Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. Genome Biol. 2016;17(1):227. <https://doi.org/10.1186/s13059-016-1088-8>.
35. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8(12):973–82. <https://doi.org/10.1038/nrg2165>.
36. Hunt T, Bergsten J, Levkanicova Z, Papadopoulou A, John OS, Wild R, et al. A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. Science. 2007;318(5858):1913–6. <https://doi.org/10.1126/science.1146954>.
37. McKenna DD, Wild AL, Kanda K, Bellamy CL, Beutel RG, Caterino MS, et al. The beetle tree of life reveals that Coleoptera survived end-Permian mass extinction to diversify during the cretaceous terrestrial revolution. Syst Entomol. 2015;40(4):835–80. <https://doi.org/10.1111/syen.12132>.
38. McKenna DD, Shin S, Ahrens D, Balke M, Beza-Beza C, Clarke DJ, et al. The evolution and genomic basis of beetle diversity. Proc Natl Acad Sci U S A. 2019;116(49):24729–37. <https://doi.org/10.1073/pnas.1909655116>.
39. Andersson MN, Grosse-Wilde E, Keeling CI, Bengtsson JM, Yuen MMS, Li M, et al. Antennal transcriptome analysis of the chemosensory gene families in the tree killing bark beetles, *Ips typographus* and *Dendroctonus ponderosae* (Coleoptera: Curculionidae: Scolytinae). BMC Genomics. 2013;14(1):198. <https://doi.org/10.1186/1471-2164-14-198>.
40. Smadja C, Butlin RK. On the scent of speciation: the chemosensory system and its role in premating isolation. Heredity. 2009;102(1):77–97. <https://doi.org/10.1038/hdy.2008.55>.
41. Smadja CM, Canbäck B, Vitalis R, Gautier M, Ferrari J, Zhou JJ, et al. Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. Evolution. 2012;66(9):2723–38. <https://doi.org/10.1111/j.1558-5646.2012.01612.x>.
42. Zhang B, Zhang W, Nie RE, Li WZ, Segreaves KA, Yang XK, et al. Comparative transcriptome analysis of chemosensory genes in two sister leaf beetles provides insights into chemosensory speciation. Insect Biochem Mol Biol. 2016;79:108–18. <https://doi.org/10.1016/j.ibmb.2016.11.001>.
43. Hallem EA, Carlson JR. Coding of odors by a receptor repertoire. Cell. 2006;125(1):143–60. <https://doi.org/10.1016/j.cell.2006.01.050>.
44. Carey AF, Wang G, Su CY, Zwiebel LJ, Carlson JR. Odorant reception in the malaria mosquito *Anopheles gambiae*. Nature. 2010;464(7285):66–71. <https://doi.org/10.1038/nature08834>.
45. Stensmyr MC, Dweck HKM, Farhan A, Ibba I, Strutz A, Mukunda L, et al. A conserved dedicated olfactory circuit for detecting harmful microbes in *Drosophila*. Cell. 2012;151(6):1345–57. <https://doi.org/10.1016/j.cell.2012.09.046>.
46. Vosshall LB, Stocker RF. Molecular architecture of smell and taste in *Drosophila*. Annu Rev Neurosci. 2007;30(1):505–33. <https://doi.org/10.1146/annurev.neuro.30.051606.094306>.
47. Abuin L, Bargeton B, Ulbrich MH, Isacoff EY, Kellenberger S, Benton R, et al. Functional architecture of olfactory ionotropic glutamate receptors. Neuron. 2011;69(1):44–60. <https://doi.org/10.1016/j.neuron.2010.11.042>.
48. Sánchez-Gracia A, Vieira FG, Rozas J. 2009. Molecular evolution of the major chemosensory gene families in insects. Heredity. 2009;103(3):208–16. <https://doi.org/10.1038/hdy.2009.55>.
49. Zhou JJ. Odorant-binding proteins in insects. Vitam Horm Vitam Horm. 2010;83:241–72. [https://doi.org/10.1016/S0083-6729\(10\)83101-9](https://doi.org/10.1016/S0083-6729(10)83101-9).
50. Andersson MN, Keeling CI, Mitchell RF. Genomic content of chemosensory genes correlates with host range in wood-boring beetles (*Dendroctonus ponderosae*, *Agrilus planipennis*, and *Anoplophora glabripennis*). BMC Genomics. 2019;20(1):690. <https://doi.org/10.1186/s12864-019-6054-x>.
51. Benton R. On the ORigin of smell: odorant receptors in insects. Cell Mol Life Sci. 2006;63(14):1579–85. <https://doi.org/10.1007/s00018-006-6130-7>.
52. Nichols AS, Chen S, Luetje CW. Subunit contributions to insect olfactory receptor function: channel block and odorant recognition. Chem Senses. 2011;36(9):781–90. <https://doi.org/10.1093/chemse/bjr053>.
53. Mitchell RF, Schneider TM, Schwartz AM, Andersson MN, McKenna DD. The diversity and evolution of odorant receptors in beetles (Coleoptera). Insect Mol Biol. 2020;29(1):77–91. <https://doi.org/10.1111/imb.12611>.
54. Hu P, Wang JZ, Cui MM, Tao J, Luo YQ. Antennal transcriptome analysis of the Asian longhorned beetle *Anoplophora glabripennis*. Sci Rep. 2016;6(1):26652. <https://doi.org/10.1038/srep26652>.
55. Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB. Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. Cell. 2009;136(1):149–62. <https://doi.org/10.1016/j.cell.2008.12.001>.
56. Rytz R, Croset V, Benton R. 2013. Ionotropic receptors (IRs): chemosensory ionotropic glutamate receptors in *Drosophila* and beyond. Insect Biochem Mol Biol. 2013;43(9):888–97. <https://doi.org/10.1016/j.ibmb.2013.02.007>.
57. Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, et al. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. PLoS Genet. 2010;6(8):e1001064. <https://doi.org/10.1371/journal.pgen.1001064>.
58. Nie RE, Andújar C, Gómez-Rodríguez C, Bai M, Xue HJ, Tang M, et al. The phylogeny of leaf beetles (Chrysomelidae) inferred from mitochondrial

- genomes. *Syst Entomol.* 2020;45(1):188–204. <https://doi.org/10.1111/syen.12387>.
59. Pelosi P, Iovinella I, Zhu J, Wang GR, Dani FR. Beyond chemoreception: diverse tasks of soluble olfactory proteins in insects. *Biol Rev.* 2018;93(1):184–200. <https://doi.org/10.1111/brv.12339>.
 60. Leal WS. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Ann Rev Entomol.* 2013;58(1):373–91. <https://doi.org/10.1146/annurev-ento-120811-153635>.
 61. Pelosi P, Zhou JJ, Ban LP, Calvello M. Soluble proteins in insect chemical communication. *Cell Mol Life Sci.* 2006;63(14):1658–76. <https://doi.org/10.1007/s00018-005-5607-0>.
 62. Koenig C, Bretschneider A, Heckel DG, Grosse-Wilde E, Hansson BS, Vogel H. The plastic response of *Manduca sexta* to host and non-host plants. *Insect Biochem Mol Biol.* 2015;63:72–85. <https://doi.org/10.1016/j.ibmb.2015.06.001>.
 63. Despres L, David J-P, Gallet C. The evolutionary ecology of insect resistance to plant chemicals. *Trends Ecol Evol.* 2007;22(6):298–307. <https://doi.org/10.1016/j.tree.2007.02.010>.
 64. Rane RV, Walsh TK, Pearce SL, Jermiin LS, Gordon KHJ, Richards S, et al. Are feeding preferences and insecticide resistance associated with the size of detoxifying enzyme families in insect herbivores? *Curr Opin Insect Sci.* 2016;13:70–6. <https://doi.org/10.1016/j.cois.2015.12.001>.
 65. Rane RV, Ghodke AB, Hoffmann AA, Edwards OR, Walsh TK, Oakeshott JG. Detoxifying enzyme complements and host use phenotypes in 160 insect species. *Curr Opin Insect Sci.* 2019;31:131–8. <https://doi.org/10.1016/j.cois.2018.12.008>.
 66. Ahn SJ, Vogel H, Heckel DG. Comparative analysis of the UDP-glycosyltransferase multigene family in insects. *Insect Biochem Mol Biol.* 2012;42:133–47. <https://doi.org/10.1016/j.ibmb.2011.11.006>.
 67. Dermauw W, Van Leeuwen T. The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance. *Insect Biochem Mol Biol.* 2014;45:89–110. <https://doi.org/10.1016/j.ibmb.2013.11.001>.
 68. Merzendorfer H. ABC transporters and their role in protecting insects from pesticides and their metabolites. *Adv Insect Physiol.* 2014;46:1–72. <https://doi.org/10.1016/B978-0-12-417010-0.00001-X>.
 69. Ramsey JS, Rider DS, Walsh TK, De Vos M, Gordon KHJ, Ponnala L, et al. Comparative analysis of detoxification enzymes in *Acyrtosiphon pisum* and *Myzus persicae*. *Insect Mol Biol.* 2010;19(Suppl. 2):155–64. <https://doi.org/10.1111/j.1365-2583.2009.00973.x>.
 70. He WT, Jin ZX, Wang BQ. Research progress of *Geranium nepalense* sweet. *J Aerosp Med.* 2011;21:1200–2.
 71. Hahn MW, De Bie T, Stajich JE, Nguyen C, Cristianini N. Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res.* 2005;15(8):1153–60. <https://doi.org/10.1101/gr.3567505>.
 72. Kanost MR, Arrese EL, Cao XL, Chen YR, Chellapilla S, Goldsmith MR, et al. Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, *Manduca sexta*. *Insect Biochem Mol Biol.* 2016;76:118–47. <https://doi.org/10.1016/j.ibmb.2016.07.005>.
 73. Shi HX, Pei LH, Gu SS, Zhu SC, Wang YY, Zhang Y, et al. Glutathione S-transferase (GST) genes in the red flour beetle, *Tribolium castaneum*, and comparative analysis with five additional insects. *Genomics.* 2012;100(5):327–35. <https://doi.org/10.1016/j.ygeno.2012.07.010>.
 74. Scott JG. Cytochromes P450 and insecticide resistance. *Insect Biochem Mol Biol.* 1999;29(9):757–77. [https://doi.org/10.1016/S0965-1748\(99\)00038-7](https://doi.org/10.1016/S0965-1748(99)00038-7).
 75. Helvig C, Koener JF, Unnithan GC, Feyereisen R. CYP15A1, the cytochrome P450 that catalyzes epoxidation of methyl farnesoate to juvenile hormone III in cockroach *Corpora allata*. *Proc Natl Acad Sci U S A.* 2004;101(12):4024–9. <https://doi.org/10.1073/pnas.0306980101>.
 76. Rewitz KF, O'Connor MB, Gilbert LI. Molecular evolution of the insect Halloween family of cytochrome P450s: phylogeny, gene organization and functional conservation. *Insect Biochem Mol Biol.* 2007;37(8):741–53. <https://doi.org/10.1016/j.ibmb.2007.02.012>.
 77. Feyereisen R. Evolution of insect P450. *Biochem Soc Trans.* 2006;34(6):1252–5. <https://doi.org/10.1042/BST0341252>.
 78. Li XC, Berenbaum MR, Schuler MA. Plant allelochemicals differentially regulate *Helicoverpa zea* cytochrome P450 genes. *Insect Biochem Mol Biol.* 2002;11(4):343–51. <https://doi.org/10.1046/j.1365-2583.2002.00341.x>.
 79. Scully ED, Hoover K, Carlson JE, Tien M, Geib SM. Midgut transcriptome profiling of *Anoplophora glabripennis*, a lignocellulose degrading cerambycid beetle. *BMC Genomics.* 2013;14(1):850. <https://doi.org/10.1186/1471-2164-14-850>.
 80. Zhu F, Moural TW, Nelson DR, Palli SR. A specialist herbivore pest adaptation to xenobiotics through upregulation of multiple cytochrome P450s. *Sci Rep.* 2016;6(1):20421. <https://doi.org/10.1038/srep20421>.
 81. Lü FG, Fu K, Li Q, Guo WC, Ahmet T, Li GQ. Identification of carboxylesterase genes and their expression profiles in the Colorado potato beetle *Leptinotarsa decemlineata* treated with fipronil and cyhalothrin. *Pestic Biochem Physiol.* 2015;122:86–95. <https://doi.org/10.1016/j.pestbp.2014.12.015>.
 82. Friedman R. Genomic organization of the glutathione S-transferase family in insects. *Mol Phylogenet Evol.* 2011;61(3):924–32. <https://doi.org/10.1016/j.ympev.2011.08.027>.
 83. Enayati AA, Ranson H, Hemingway J. Insect glutathione transferases and insecticide resistance. *Insect Biochem Mol Biol.* 2005;14(1):3–8. <https://doi.org/10.1111/j.1365-2583.2004.00529.x>.
 84. Meyer JM, Markov GV, Baskaran P, Herrmann M, Sommer RJ, Rödelsperger C. Draft genome of the scarab beetle *Oryctes borbonicus* on La Réunion Island. *Genome Biol Evol.* 2016;8(7):2093–105. <https://doi.org/10.1093/gbe/evw133>.
 85. Han JB, Li GQ, Wan PJ, Zhu TT, Meng QW. Identification of glutathione S-transferase genes in *Leptinotarsa decemlineata* and their expression patterns under stress of three insecticides. *Pestic Biochem Physiol.* 2016;133:26–34. <https://doi.org/10.1016/j.pestbp.2016.03.008>.
 86. Broehan G, Kroeger T, Lorenzen M, Merzendorfer H. Functional analysis of the ATP-binding cassette (ABC) transporter gene family of *Tribolium castaneum*. *BMC Genomics.* 2013;14(1):6. <https://doi.org/10.1186/1471-2164-14-6>.
 87. Strauss AS, Peters S, Boland W, Gretscher RR, Groth M, Boland W, et al. ABC transporter functions as a pacemaker for sequestration of plant glucosides in leaf beetles. *eLife.* 2013;2:e01096. <https://doi.org/10.7554/eLife.01096>.
 88. Liu SM, Zhou S, Tian L, Guo EE, Luan YX, Zhang JZ, et al. Genome-wide identification and characterization of ATP-binding cassette transporters in the silkworm, *Bombyx mori*. *BMC Genomics.* 2011;12(1):491. <https://doi.org/10.1186/1471-2164-12-491>.
 89. Watanabe H, Tokuda G. Cellulolytic systems in insects. *Ann Rev Entomol.* 2010;55(1):609–32. <https://doi.org/10.1146/annurev-ento-112408-085319>.
 90. Calderón-Corfe N, Quesada M, Watanabe H, Cano-Camacho H, Oyama K. Endogenous plant cell wall digestion: A key mechanism in insect evolution. *Annu Rev Ecol Evol Syst.* 2012;43:45–71. <https://doi.org/10.1146/annurev-ecolsys-110411-160312>.
 91. Hare EE, Johnston JS. Genome size determination using flow cytometry of propidium iodide-stained nuclei. *Methods Mol Biol.* 2011;772:3–12.
 92. Nie RE, Wei J, Zhang SK, Vogler AP, Wu L, Konstantinov AS, et al. Diversification of mitogenomes in three sympatric *Altica* flea beetles (Insecta, Chrysomelidae). *Zool Scr.* 2019;48(5):657–66. <https://doi.org/10.1111/zsc.12371>.
 93. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018;34(18):3094–100. <https://doi.org/10.1093/bioinformatics/bty191>.
 94. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 2019;37(5):540–6. <https://doi.org/10.1038/s41587-019-0072-8>.
 95. Pryszyk LP, Gabaldón T. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* 2016;44(12):e113. <https://doi.org/10.1093/nar/gkw294>.
 96. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using. *Bioinformatics.* 2011;27(4):578–9. <https://doi.org/10.1093/bioinformatics/btq683>.
 97. Boetzer M, Pirovano W. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics.* 2014;15(1):211. <https://doi.org/10.1186/1471-2105-15-211>.
 98. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10(1):421. <https://doi.org/10.1186/1471-2105-10-421>.
 99. Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol.* 2018;35(3):543–8. <https://doi.org/10.1093/molbev/msx319>.
 100. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 2007;35(Web Server):W265–8. <https://doi.org/10.1093/nar/gkm286>.
 101. Price AL, Jones NC, Pevzner PA. *De novo* identification of repeat families in large genomes. *Bioinformatics.* 2005;21(Suppl 1):i351–8. <https://doi.org/10.1093/bioinformatics/bti1018>.

102. Edgar RC, Myers EW. PILER: identification and classification of genomic repeats. *Bioinformatics*. 2005;21(Suppl. 1):152–8.
103. Hoede C, Arnoux S, Moisset M, Chaumier T, Inizan O, Jamilloux V, et al. PAST EC: an automatic transposable element classification tool. *PLoS One*. 2014; 9(5):e91929. <https://doi.org/10.1371/journal.pone.0091929>.
104. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res*. 2005;110(1–4):462–7. <https://doi.org/10.1159/000084979>.
105. Tarailo-Graovac M, Chen NS. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*. 2009; 4(10). <https://doi.org/10.1002/0471250953.bi0410s25>.
106. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033>.
107. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol*. 1997;268(1):78–94. <https://doi.org/10.1006/jmbi.1997.0951>.
108. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*. 2003;19(Suppl. 2):215–25.
109. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics*. 2004;20(16):2878–9. <https://doi.org/10.1093/bioinformatics/bth315>.
110. Blanco E, Parra G, Guigo R. Using geneid to identify genes. *Current Protocols in Bioinformatics*. 2007;4(3). <https://doi.org/10.1002/0471250953.bi0403s18>.
111. Korf I. Gene finding in novel genomes. *BMC Bioinform*. 2004;5(1):59. <https://doi.org/10.1186/1471-2105-5-59>.
112. Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res*. 2016;44(9):e89. <https://doi.org/10.1093/nar/gkw092>.
113. Keilwagen J, Hartung F, Paulini M, Twardziok SO, Grau J. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinform*. 2018;19(1):189. <https://doi.org/10.1186/s12859-018-2203-5>.
114. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;11(9):1650–67. <https://doi.org/10.1038/nprot.2016.095>.
115. Campbell MA, Haas BJ, Hamilton JP, Mount SM, Buell CR. Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics*. 2006;7(1):327. <https://doi.org/10.1186/1471-2164-7-327>.
116. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. *Genome Biol*. 2008;9(1):R7. <https://doi.org/10.1186/gb-2008-9-1-r7>.
117. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res*. 2005;33(Database issue):121–4.
118. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*. 2006;34(Database issue):140–4.
119. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 2013;29(22):2933–5. <https://doi.org/10.1093/bioinformatics/btt509>.
120. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25(5): 955–64. <https://doi.org/10.1093/nar/25.5.955>.
121. She R, Chu SC, Uyar B, Wang J, Wang K, Chen NS. genBlastG: using BLAST searches to build homologous gene models. *Bioinformatics*. 2011;27(15): 2141–3. <https://doi.org/10.1093/bioinformatics/btr342>.
122. Birney E, Clamp M, Durbin R. GeneWise and Genomewise. *Genome Res*. 2004;14(5):988–95. <https://doi.org/10.1101/gr.1865504>.
123. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res*. 2011;39(Database issue):D225–9. <https://doi.org/10.1093/nar/gkq1189>.
124. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*. 2003;31(1):365–70. <https://doi.org/10.1093/nar/gkg095>.
125. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res*. 2001;29(1):22–8. <https://doi.org/10.1093/nar/29.1.22>.
126. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30. <https://doi.org/10.1093/nar/28.1.27>.
127. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10. [https://doi.org/10.1016/S0022-836\(05\)80360-2](https://doi.org/10.1016/S0022-836(05)80360-2).
128. Dimmer EC, Huntley RP, Alam-Faruque Y, Sawford T, O'Donovan C, Martin MJ, et al. The UniProt-GO annotation database in 2011. *Nucleic Acids Res*. 2012;40(D1):D565–70. <https://doi.org/10.1093/nar/gkr1048>.
129. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005;21(18):3674–6. <https://doi.org/10.1093/bioinformatics/bti610>.
130. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59(3):307–21. <https://doi.org/10.1093/sysbio/syq010>.
131. Yang ZH, Rannala B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol*. 2006;23(1):212–26. <https://doi.org/10.1093/molbev/msj024>.
132. Rannala B, Yang ZH. Inferring speciation times under an episodic molecular clock. *Syst Biol*. 2007;56(3):453–66. <https://doi.org/10.1080/10635150701420643>.
133. Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol*. 2013;30(8):1987–97. <https://doi.org/10.1093/molbev/mst100>.
134. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. 2017;20(4):1160–6. <https://doi.org/10.1093/bib/bbx108>.
135. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011;27(8):1164–5. <https://doi.org/10.1093/bioinformatics/btr088>.
136. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006; 22(21):2688–90. <https://doi.org/10.1093/bioinformatics/btl446>.
137. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES science gateway for inference of large phylogenetic trees. In: Gateway computing environments workshop (GCE). New Orleans: IEEE; 2010.
138. Rambaut A. FigTree v1.3.1: Tree figure drawing tool. 2009. Available: <http://tree.bio.ed.ac.uk/software/figtree/>.
139. Wang YQ, Song FH, Zhu JW, Zhang SS, Yang YD, Chen TT, et al. GSA: genome sequence archive. *Genom Proteom Bioinf*. 2017;15(1):14–8. <https://doi.org/10.1016/j.gpb.2017.01.001>.
140. National Genomics Data Center Members and Partners. Database resources of the National Genomics Data Center in 2020. *Nucleic Acids Res*. 2020;48: D24–33.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

