Genome assemblies of three closely related leaf beetle species

(Galerucella spp).

Xuyue Yang, Tanja Slotte, Jacques Dainat, Peter A. Hambäck

*Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm 10691, Sweden; †Department of Medical Biochemistry Microbiology and Genomics, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Uppsala 75237, Sweden

ORCID IDs: 0000-0003-2084-1651 (X.Y.Y); 0000-0001-6020-5102 (T.S); 0000-0001-6362-6199 (P.A.H); 0000-0002-6629-0173 (J.D)

Corresponding author: Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm 10691, Sweden. E-mail: peter.hambback@su.se

© The Author(s) (2021). Published by Oxford University Press on behalf of the Genetics Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Running title: Genome assemblies of *Galerucella* spp.

**KEYWORDS:** *Galerucella calmariensis, Galerucella pusilla, Galerucella tenella,* Coleoptera, leaf beetle

**Corresponding author:** Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm 10691, Sweden. Visiting address: Svante Arrhenius väg 20 A Telephone: 08-16 12 88 E-mail: peter.hamback@su.se
**Abstract**

_Galerucella_ (Coleoptera: Chrysomelidae) is a leaf beetle genus that has been extensively used for ecological and evolutionary studies. It has also been used as biological control agent against invading purple loosestrife in North America, with large effects on biodiversity. Here we report genome assembly and annotation of three closely related _Galerucella_ species: _G. calmariensis, G. pusilla_ and _G. tenella_. The three assemblies have a genome size ranging from 460 Mbp to 588 Mbp, with N50 from 31,588 kbp to 79.674 kbp, containing 29,202 to 40,929 scaffolds. Using an _ab initio_ evidence-driven approach, 30,302 to 33,794 protein-coding genes were identified and functionally annotated. These draft genomes will contribute to the understanding of host-parasitoid interactions, evolutionary comparisons of leaf beetle species and future population genomics studies.

**Introduction**

_Galerucella_ (Coleoptera: Chrysomelidae) is a leaf beetle genus that is distributed worldwide except in the Neotropics (Thomas et al. 2002). Some species have been used as biological control agents against invasive plants, and the host specificity and environmental impact of these species have attracted broad interest. The most common application is the introduction of _G. calmariensis_ and _G. pusilla_ from Europe to North America against the invasive wetland plant purple loosestrife (_Lythrum salicaria_). Since 1992, releases of _Galerucella_ populations have been made in many states in the USA and the colonization appears to have been
successful, leading to a dramatic decrease of *L. salicaria* populations (Blossey et al. 1994; Landis et al. 2003; McAvoy et al. 2016).

In addition to its application in biological control, *Galerucella* spp. has been widely investigated in both ecological (Pappers et al. 2002; Tanaka and Nakasuji 2002; Hori et al. 2006; Fors et al. 2016) and evolutionary (Ikonen et al. 2003; Stenberg and Axelsson 2008; Yang et al. 2020) studies. In particular, *Galerucella* spp. has been used to study ecological and evolutionary consequences of host-parasitoid interactions (Stenberg et al. 2007), mainly involving three closely related species (*G. calmariensis*, *G. pusilla* and *G. tenella*) with similar life cycles and their shared wasp parasitoid (*Asecodes parviclava*) (Hambäck et al. 2013; Fors et al. 2016). The divergence of these three species is fairly recent: *G. pusilla* and *G. calmariensis* diverged around 77,000 years ago while *G. tenella* diverged around 400,000 years ago (Hambäck et al. 2013). *G. pusilla* and *G. calmariensis* share an exclusive host plant (*L. salicaria*), whereas *G. tenella* feeds primarily on *Filipendula ulmaria* and occasionally on other Rosaceae species. In all three species, adults in the study area overwinter until mid-May and then lay eggs on leaves or stems of their host plant. Larvae hatch after 1-2 weeks, pupate in late June to early July, and the adults emerge from the pupae by the end of July. The three species are attacked by the same endoparasitoid wasp *Asecodes parviclava*, which lays one or more eggs in the beetle larvae. The successful wasp larvae kills the host, use it as food resources, and subsequently emerge during the following summer (Hansson and Hambäck 2013).

The demography, host searching behavior and immunology have previously been addressed in several *Galerucella* species (Zheng et al. 2008; Fors et al. 2015; Yang et al. 2020), but no genome assemblies of *Galerucella* species are currently available. The closest related species
that has an available genome assembly is ragweed leaf beetle (*Ophraella communa*), which also belongs to the leaf beetle (Chrysomelidae) family (Bouchemousse et al. 2020). Here we report de-novo genome assemblies for *G. calmariensis*, *G. pusilla* and *G. tenella*. We performed computational annotation, assigned gene ontology to functional proteins and performed ortholog cluster analysis between the three species. These draft genomes will be useful for understanding the mechanisms underlying beetle interactions with parasitoid and plant use, and for future population genomics studies (McKenna 2018).

**Materials and methods**

**DNA extraction and sequencing**

Larvae samples of *G. pusilla* and *G. calmariensis* were collected in mid-May 2018, from Iggön (59° 2'30.81"N, 17° 9'49.35"O), Sweden. *G. tenella* samples were collected in mid-May 2018, from Södersjön (59°51'8.72"N, 18° 6'26.59"O), Sweden. To reduce heterozygosity and bacterial contamination, we reared and inbred the beetles in the laboratory at room temperature for one generation and collected adults from the second generation for DNA extraction.

For each species, we extracted DNA from one individual using an adjusted version of the 10X Genomics sample preparation protocol “DNA Extraction from Single Insects” (https://assets.ctfassets.net/an68im79xiti/3oGwQ5kl6UyCocGgmoWQie/768ae48be4f99b1f9)
DNA extraction:

DNA concentrations were measured with a Qubit 3.0 Fluorometer using the dsDNA HS Assay Kit (Thermo Fisher Scientific) and DNA integrity was assessed on an agarose gel stained with 2% GelRed. 10X Genomics Chromium linked-read sequencing libraries were prepared and subsequently sequenced to yield paired-end 2x150 bp reads, on a HiSeq X platform at SciLifeLab (Stockholm, Sweden).

Genome assembly and scaffolding:

Raw 10X genomics reads were checked for sequencing quality using FastQC v0.11.5 (Andrews 2010), and de novo assembled using the Supernova v2.1.0 (Weisenfeld et al. 2017) assembler. We then polished the draft assembly using purge_dups v1.0.1 (Guan et al. 2020) to remove haplotigs and heterozygous overlaps based on sequence similarity and read depth. Subsequently, assemblies were scaffolded using arcs v1.0.6 (Yeo et al. 2018) and links v1.8.6, with the -a parameter, which controls maximum link ratio between two best contig pairs set to 0.7 (Warren et al. 2015). To remove sequence contamination from the assembly, we ran Kraken v2.0 (Wood and Salzberg 2014) against bacterial, archaeal, and viral domains, along with the human genome. We assessed the completeness of our polished genome assemblies assessed by Benchmarking Universal Single-Copy Orthologs v4.0.5 (BUSCO) (Simão et al. 2015) from OrthoDB v9.1 (Zdobnov et al. 2017) using Endopterygota as the taxonomic database.

Gene annotation and phylogenetic analysis:
We first assessed the repeat content of our genome assemblies and created a specific repeat library using RepeatModeler v1.0.11 (Smit et al. 2010b) for each genome assembly. Based on the repeat library, identification of repeat sequences in the genome was performed using RepeatMasker v3.0.9 (Smit et al. 2010a) and RepeatRunner (Yandell 2006) with default settings. RepeatRunner is a program that integrates RepeatMasker with BLASTX (Altschul et al. 1997), allowing the analysis of highly divergent repeats and identifications of divergent protein-coding portions of retro-elements and retroviruses.

Gene annotation was performed using the MAKER package v3.01.02 (Holt and Yandell 2011). First, for each genome, we generated one initial evidence-based annotation using both protein and transcriptome data sources. Protein databases came from the UniProt Swiss-Prot database (downloaded on 2019-11; 561,356 proteins) (Engler et al. 2020), as well a subset of manually selected proteins (uniport request: taxonomy: "Coleoptera [7041]", existence: "Inferred from homology [3]", 161,853 proteins). In addition to protein resources, transcriptome data containing 57,255 transcripts from G. pusilla were used as evidence for all three genomes (Yang et al. 2020). Next, we used the candidate genes from the initial annotation to train two different ab-initio gene predictors: Augustus v3.3.3 (Stanke et al. 2006) and Snap v2013_11_29 (Korf 2004). Finally, an ab-initio evidence-driven gene build was generated based on the initial evidence-based annotation and the ab-initio predictions. Additionally, we used EVidenceModeler v1.1.1 (Haas et al. 2008), which allows the construction of gene models based on the best possible set of exons produced by the ab-initio tools, and chooses those most consistent with the evidence. Functional inference for genes and transcripts was performed using the translated CDS features of each coding transcript.
Each predicted protein sequence was run against InterProscan (Jones et al. 2014) in order to retrieve functional information from 20 different sources. In addition, Blastp v2.9.0 (Altschul et al. 1990) was performed against the complete Swiss-Prot/UniProt database (downloaded 2019-11) with a maximum e-value cut-off of $1e^{-6}$ to assign putative functions to predicted proteins. tRNA have been predicted through tRNAscan v1.3.1 (Lowe and Eddy 1997).

To confirm the evolutionary relationships between the three *Galerucella* species and their position in the Chrysomelidae family, we reconstructed a species tree based on predicted protein sets from *G. calmariensis*, *G. pusilla* and *G. tenella*, *O. commun* and *L. decemlineata* using OrthoFinder v2.4.0, with default settings except using multiple sequence alignments (-M msa) to infer the species tree (Emms and Kelly 2015).

Ortholog cluster analysis

Identifying shared orthologous clusters allows the comparison of function and evolution of proteins across closely related species. An ortholog cluster analysis was performed by comparing the three complete *Galerucella* protein sets with each other via OrthoVenn2 (Xu et al. 2019) with default settings of $E = 1e^{-5}$ and an inflation value of 1.5.

**Results and discussion**

**Genome assemblies**
Sequencing of the 10X genomics libraries yielded a total of 683.34 million read pairs, resulting in a sequencing depth above 110X for each species. Due to the low molecular weight of the input DNA (average size < 20 kbp), the initial de novo assembly from Supernova was highly fragmented, with N50 values of 49.884 kbp, 19.764 kbp and 24.604 kbp for *G. calmaríensis*, *G. pusilla* and *G. tenella*. Redundancy removal by purge_dups and arcs+links scaffolding dramatically improved N50 values of assemblies (see supplementary materials Table 1 for the comparisons between assemblies). The decontamination process removed two contigs from *G. calmaríensis*, one contig from *G. pusilla* and two contigs from *G. tenella* which matched the human database with a kmer length >100 bp. Final assemblies for *G. calmaríensis* had a size = 588 Mbp, contained 39,255 scaffolds with a N50 = 79.674 kbp, final assemblies for *G. pusilla* had a size = 513 Mbp, 40,929 scaffolds with a N50 = 45,442 kbp whereas final assemblies for *G. tenella* has a size = 460 Mbp, 29,202 scaffolds with a N50 = 31,588 kbp (Table 1). Using 2,124 BUSCO groups with endopterygota_odb10 database, we found 91.3% complete orthologs and only 4.0% missing orthologs in *G. calmaríensis*, 85.3% complete orthologs and 6.5% missing orthologs for *G. pusilla* and 95.4% complete orthologs and 3.3% missing orthologs for *G. tenella*. Although the final assembly was still fragmented, the completeness of genome measured by BUSCO was satisfactory. The GC content of the three genomes ranged from 33.6% to 33.8%, which is slightly higher than the GC content of the ragweed leaf beetle genome assembly (Bouchemousse *et al.* 2020). The sizes of the genome assemblies of our three *Galerucella* species varied (460 Mbp to 588Mb) but is slightly smaller than the size of the colorado potato beetle (*Leptinotarsa decemlineata*) (642 Mbp) (Herndon *et al.* 2020) and the ragweed leaf beetle (774 Mbp) (Bouchemousse *et al.* 2020). Coleoptera is amongst the most diverse insect orders in terms of genome size, with an average genome size of 760 Mbp and ranges from 160 to 5,020 Mbp.
Within-genus variation in genome size is relatively small in these three assemblies compared with other Coleopteran species, possibly because of their close phylogenetic relationships and similarities in life cycle, food sources and wasp enemies.

**Table 1**

Gene annotation and phylogenetic analysis

RepeatMasker masked 48.55%, 46.65% and 40.84% of the *G. calmariensis*, *G. pusilla* and *G. tenella* genomes as repetitive elements. In addition, RepeatRunner further masked approximately 1% of each genome as repeats using MAKER TE as library (Supplementary table 2). The *ab initio* evidence-driven annotation using the MAKER pipeline revealed 32,294, 30,302 and 33,794 potential protein coding genes, accounting for 16.2%, 17.7% and 19.1% of the whole genome of *G. calmariensis*, *G. pusilla* and *G. tenella* respectively (Supplementary table 3). For each species, 84% to 86% of protein coding genes were assigned with a putative function, and 39% to 45% had a GO annotation (supplementary table 4, functional annotations using InterProscan from 20 different sources). Blast against the UniProt/Swiss-Prot database predicted 15,046, 14,404 and 17,958 hits with unique gene names for *G. calmariensis*, *G. pusilla* and *G. tenella*, respectively.

A maximum likelihood tree was built based on 1,242 orthogroups shared between three *Galerucella* species and two other leaf beetles (Figure 1). The phylogenetic relationship is in accordance with previous studies generated by mitochondrial and nuclear genetic markers with high bootstrap support values for each branch (Hambäck et al. 2013; Bouchemousse et al. 2020).
The three protein sets of *Galerucella* species were compared to identify shared orthologous clusters using OrthoVenn2 (Figure 2). The complete protein sets contain 40,031 sequences from *G. calamiensis*, 37,514 sequences from *G. pusilla* and 44,200 sequences from *G. tenella*, corresponding to 20,665, 19,730 and 19,106 ortholog clusters respectively.

Most annotated genes (12,372 orthogroups/48,398 proteins) were shared between the three species. Shared clusters between *G. calamiensis* and *G. pusilla* (16,594) account for 80.3% and 84.1% of ortholog clusters in *G. calamiensis* and *G. pusilla* respectively whereas the shared regions of either *G. calamiensis* and *G. pusilla* with *G. tenella* account for less than 75% of their clusters. Ortholog clusters unique to a single species account for 6.09%, 3.87% and 8.11% of the entire cluster set for *G. calamiensis*, *G. pusilla* and *G. tenella*, which indicates divergent regions between species (Ferguson *et al.* 2020). The inflated numbers of singleton clusters in *G. tenella* may be due to the high duplication levels in the genome, as BUSCO detected 33.1% complete duplicated BUSCOs in *G. tenella*. Whether this is due to gene duplication or assembly error should be further investigated. The duplication and fragmentation level detected by BUSCO are similar between *G. calamiensis* (1.8% duplicated and 5.1% fragmented) and *G. pusilla* (1.2% duplicated and 8.2% fragmented), however *G. calamiensis* harbours a higher level of singleton clusters than *G. pusilla*.

**Conclusion**
Galerucella species play an important role as biological control agents as well as for ecological and evolutionary research. Here we produced draft genome assemblies for three leaf beetles in the Galerucella genus, which are the first three genomes from the Galerucinae subfamily branch of the leaf beetle family. The genome sequencing of the three closely related beetles sharing a common wasp enemy also provide possibilities of understandings of food web and host-parasitoid interactions. In particular, comparing genomes of species with divergent immune resistance against parasitoid wasps may contribute to detecting essential genetic regions underlying host immunity and other potential traits participating in the arms race between host and parasitoids.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grant VR-2015-4232 (to PAH) from the Swedish Research Council.

Data availability

Raw read data and final assemblies are available at the EMBL-ENA database under BioProject PRJEB44256. Supplemental material including annotations is available at figshare: (https://doi.org/10.6084/m9.figshare.c.5470650). Command-line arguments and scripts for this study is available at: (https://github.com/Pikayy/Galerucella).
Acknowledgements

The authors would like to thank Remi-André Olsen at SciLifeLab Stockholm, Sweden for advice in genome assembly.

Funding

The authors would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, NGI, and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure. Computational analyses were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project snic2017-7-233. SNIC is partially funded by the Swedish Research Council through grant agreement no. 2018-05973”.

References

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 1990 Basic local alignment search tool. J. Mol. Biol. 215: 403–410.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang et al., 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389–3402.

Andrews, S., 2010 FastQC: a quality control tool for high throughput sequence data. Cambridge, UK Babraham Inst.

Blossey, B., D. Schroeder, S. D. Hight, and R. A. Malecki, 1994 Host specificity and environmental impact of two leaf beetles (Galerucella calmariensis and G. pusilla) for biological control of purple loosestrife (Lythrum salicaria). Weed Sci. 134–140.

Bouchemousse, S., L. Falquet, and H. Müller-Schärer, 2020 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. 12: 1167–1173.
Emms, D. M., and S. Kelly, 2015 OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16: 1–14.

Engler, J. O., Y. Lawrie, Y. Gansemans, F. van Nieuwerburgh, A. Suh et al., 2020 Genome Report: De novo genome assembly and annotation for the Taita white-eye (Zosterops silvanus). bioRxiv.

Ferguson, K. B., T. Kursch-Metz, E. C. Verhulst, and B. A. Pannebakker, 2020 Hybrid genome assembly and evidence-Based annotation of the egg parasitoid and biological control agent trichogramma brassicae. G3 Genes, Genomes, Genet. 10: 3533–3540.

Fors, L., I. Liblikas, and P. Andersson, 2015 Chemical communication and host search in Galerucella leaf beetles. 33–45.

Fors, L., R. Markus, U. Theopold, L. Ericson, P. A. Hambäck et al., 2016 Geographic variation and trade-offs in parasitoid virulence. J. Anim. Ecol. 85: 1595–1604.

Gregory, T. R., 2021 Animal genome size database. http://www.genomesize.com.

Guan, D., S. A. McCarthy, J. Wood, K. Howe, Y. Wang et al., 2020 Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics 36: 2896–2898.

Haas, B. J., S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen et al., 2008 Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9: 1–22.

Hambäck, P. A., E. Weingartner, L. Ericson, L. Fors, A. Cassel-Lundhagen et al., 2013 Bayesian species delimitation reveals generalist and specialist parasitic wasps on Galerucella beetles (Chrysomelidae): sorting by herbivore or plant host. BMC.

Hansson, C., and P. A. Hambäck, 2013 Three cryptic species in Asecodes ( Förster)(Hymenoptera, Eulophidae) parasitizing larvae of Galerucella spp.(Coleoptera, Chrysomelidae), including a new species. J. Hymenopt. Res. 64: 51–64.

Herndon, N., J. Shelton, L. Gerischer, P. Ioannidis, M. Ninova et al., 2020 Enhanced genome assembly and a new official gene set for Tribolium castaneum. BMC Genomics 21: 1–13.

Holt, C., and M. Yandell, 2011 MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12: 1–14.

Hori, M., K. Ohuchi, and K. Matsuda, 2006 Role of host plant volatile in the host-finding behavior of the strawberry leaf beetle, Galerucella vittaticollis Baly (Coleoptera: Chrysomelidae). Appl. Entomol. Zool. 41: 357–363.

Ikonen, A., M. Sipura, S. Miettinen, and J. Tahvanainen, 2003 Evidence for host race formation in the leaf beetle Galerucella lineola. Entomol. Exp. Appl. 108: 179–185.
Jones, P., D. Binns, H.-Y. Chang, M. Fraser, W. Li et al., 2014 InterProScan 5: genome-scale protein function classification. Bioinformatics 30: 1236–1240.

Korf, I., 2004 Gene finding in novel genomes. BMC Bioinformatics 5: 1–9.

Landis, D. A., D. C. Sebolt, M. J. Haas, and M. Klepinger, 2003 Establishment and impact of Galerucella calmariensis L. (Coleoptera: Chrysomelidae) on Lythrum salicaria L. and associated plant communities in Michigan. Biol. Control 28: 78–91.

Lowe, T. M., and S. R. Eddy, 1997 tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25: 955–964.

McAvoy, T. J., L. T. Kok, and N. Johnson, 2016 A multiyear year study of three plant communities with purple loosestrife and biological control agents in Virginia. Biol. Control 94: 62–73.

McKenna, D. D., 2018 Beetle genomes in the 21st century: prospects, progress and priorities. Curr. Opin. Insect Sci. 25: 76–82.

Pappers, S. M., G. Van der Velde, and J. N. Ouborg, 2002 Host preference and larval performance suggest host race formation in Galerucella nymphaeae. Oecologia 130: 433–440.

Schoville, S. D., Y. H. Chen, M. N. Andersson, J. B. Benoit, A. Bhandari et al., 2018 A model species for agricultural pest genomics: The genome of the Colorado potato beetle, Leptinotarsa decemlineata (Coleoptera: Chrysomelidae). Sci. Rep. 8: 1–18.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210–3212.

Smit, A., R. Hubley, and P. Green, 2010a RepeatMasker Open-3.0.

Smit, A., R. Hubley, and P. Green, 2010b RepeatModeler Open-1.0.

Stanke, M., O. Keller, I. Gunduz, A. Hayes, S. Waack et al., 2006 AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34: W435–W439.

Stenberg, J. A., and E. P. Axelsson, 2008 Host race formation in the meadowsweet and strawberry feeding leaf beetle Galerucella tenella. Basic Appl. Ecol. 9: 560–567.

Stenberg, J. A., J. Heijari, J. K. Holopainen, and L. Ericson, 2007 Presence of Lythrum salicaria enhances the bodyguard effects of the parasitoid Asecodes mento for Filipendula ulmaria. Oikos 116: 482–490.

Tanaka, M., and F. Nakasuji, 2002 Dynamic interaction between a leaf beetle, Galerucella nipponensis, and an aquatic plant, Trapa japonica. II. Dispersal behavior of larvae. Popul. Ecol. 44: 1–6.

Thomas, M. C., P. E. Skelley, and J. H. Frank, 2002 American Beetles, Volume II: Polyphaga:
Scarabaeoidea through Curculionoidea. CRC Press.

Warren, R. L., C. Yang, B. P. Vandervalk, B. Behsaz, A. Lagman et al., 2015 LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. Gigascience 4: s13742-015.

Weisenfeld, N. I., V. Kumar, P. Shah, D. M. Church, and D. B. Jaffe, 2017 Direct determination of diploid genome sequences. Genome Res. 27: 757–767.

Wood, D. E., and S. L. Salzberg, 2014 Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15: R46.

Xu, L., Z. Dong, L. Fang, Y. Luo, Z. Wei et al., 2019 OrthoVenn2: A web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 47: W52–W58.

Yandell, M., 2006 Comparative genomics library-RepeatRunner.

Yang, X., L. Fors, T. Slotte, U. Theopold, M. Binzer-Panchal et al., 2020 Differential expression of immune genes between two closely related beetle species with different immunocompetence following attack by asecodes parviclava. Genome Biol. Evol. 12: 522–534.

Yeo, S., L. Coombe, R. L. Warren, J. Chu, and I. Birol, 2018 ARCS: scaffolding genome drafts with linked reads. Bioinformatics 34: 725–731.

Zdobnov, E. M., F. Tegenfeldt, D. Kuznetsov, R. M. Waterhouse, F. A. Simao et al., 2017 OrthoDB v9. 1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. Nucleic Acids Res. 45: D744–D749.

Zheng, F., Y. Du, Z. Wang, and J. Xu, 2008 Effect of temperature on the demography of Galerucella birmanica (Coleoptera: Chrysomelidae). Insect Sci. 15: 375–380.

Figure 1 Rooted phylogenetic species tree of G. calmariensis, G. pusilla, G. tenella and two leaf beetle species, Ophraella communa and Leptinotarsa decemlineata (outgroup), derived from 1,242 orthogroups using OrthoFinder. Branch labels indicate support values based on 1,000 bootstrap replicates.

Figure 2 A Venn diagram of the orthologous gene clusters among the three Galerucella species: G. calmariensis, G. pusilla and G. tenella. The numbers of shared Ortholog clusters between species is indicated in the overlapping areas of the circles while the numbers of proteins corresponding to each cluster are underneath in parentheses.
Table 1 Summary of *G. calmariensis*, *G. pusilla* and *G. tenella* reference genomes. BUSCO score is based on the Endopterygota_db10 dataset.

| Species       | Assembly size | Number of scaffolds | scaffold N50 | Max scaffold length | Number of Ns | GC (%) | BUSCO (Complete %) |
|---------------|---------------|---------------------|---------------|---------------------|---------------|--------|--------------------|
| *G. calmariensis* | 588.27 Mbp    | 39,255              | 79.674 kbp    | 1.307 Mbp           | 37,346,600    | 33.8   | 1941/91.3%         |
| *G. pusilla*   | 513.24 Mbp    | 40,929              | 45.442 kbp    | 3.034 Mbp           | 36,426,200    | 33.7   | 1812/85.3%         |
| *G. tenella*   | 460.59 Mbp    | 29,202              | 31.588 kbp    | 0.234 Mbp           | 2,669,407     | 33.6   | 2027/95.4%         |
Galerucella calmariensis

1259 (3403)

Galerucella pusilla

764 (2336)

Galerucella tenella

2812 (6981)

12372 (48398)

2372 (5799)

1550 (5616)