Supporting Information

SI Results and Discussion

Comparison of cytochrome oxidase I (COI) sequences obtained for the *Psylliodes* species included in this study revealed less than 5% sequence divergence among several species. This was the case for a) *P. cuprea, P. marcida, P. isatidis, P. crambicola* and *P. luridipennis* (2.5-3.7%), b) *P. pyritosa*, *P. instabilis* and *P. hispana* (0.7-1.3%), c) *P. anatolica*, *P. wachsmanni* and *P. yalvacensis* (1.7-2.1%), d) *P. milleri* and *P. toelgi* (2.9%), and e) *P. chalcomera* and *P. hyoscyami* (4.9%). On the other hand, COI sequences of *P. chrysocephala* and *P. chrysocephala var. collaris* showed 6.8% sequence divergence.

To determine whether the low levels of COI sequence divergence between some *Psylliodes* species are consistent with publicly available COI data from other studies, we mined the NCBI database and analyzed the levels of sequence divergence between species. For *P. anatolica*, *P. wachsmanni* and *P. yalvacensis*, no sequences were available in NCBI for comparison. For *P. milleri*, one COI sequence was available (KF652933.1), which shared only 87.6% with our COI sequence (MW254873), but at least 99% sequence identity with COI sequences of *P. napi* (e.g., KM442200.1 and MW254874), suggesting this species is *P. napi* instead of *P. milleri*. We thus did not compare the level of sequence divergence between publicly available sequences of *P. milleri* and *P. toelgi*.

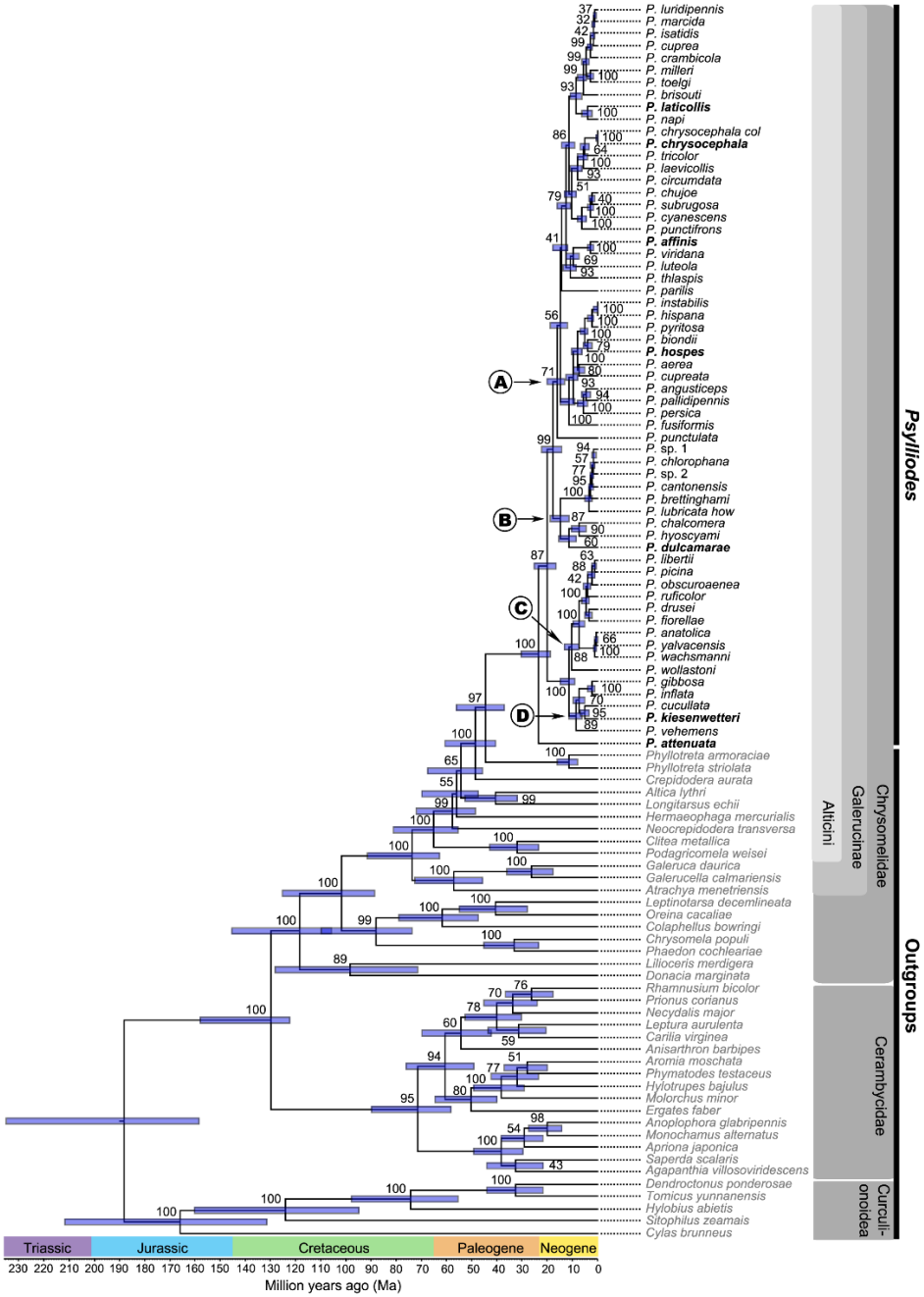
In agreement with our results, we found between 2.6 and 3.2% COI sequence divergence between *P. cuprea* (NCBI accession numbers: JF889836.1, KU910984.1), *P. marcida* (KJ962129.1, KJ961763.1)*, P. isatidis* (KM449184.1, KM448273.1)*,* and *P. crambicola* (KJ962171.1, KJ965632.1), and 4.9-5.1% sequence divergence between *P. chalcomera* (KF653547.1, KM452388.1) and *P. hyoscyami* (KM450066.1). *P. luridipennis* was not included in the comparison because the publicly available COI sequences for this species (EU110849.1, EU110848.1) do not overlap with the COI sequences mentioned above.

Our comparison of COI sequences from *P. pyritosa*, *P. instabilis* and *P. hispana* revealed confusing results. *P. hispana* was included in our study using a previously published COI sequence (KX943503.1). However, this sequence shared only 87.1% sequence identity with two other published COI sequences *for P. hispana* (EU110842.1, EU110843.1), which, in turn, shared 99.7% sequence identity with a published COI sequence of *P. cupreata*, and 99.4% sequence identity with our *P. cupreata* COI sequence (MW254853). Due to the confusing COI data for *P. hispana*, we did not compare COI sequences of *P. hispana* and other species. Instead, we focused on published COI sequences from *P. pyritosa* (MH323339.1, MH323338.1, MH323337.1) and *P. instabilis* (HQ953918.1, HQ953917.1, KU914115.1, KU911372.1), which showed 5.2%

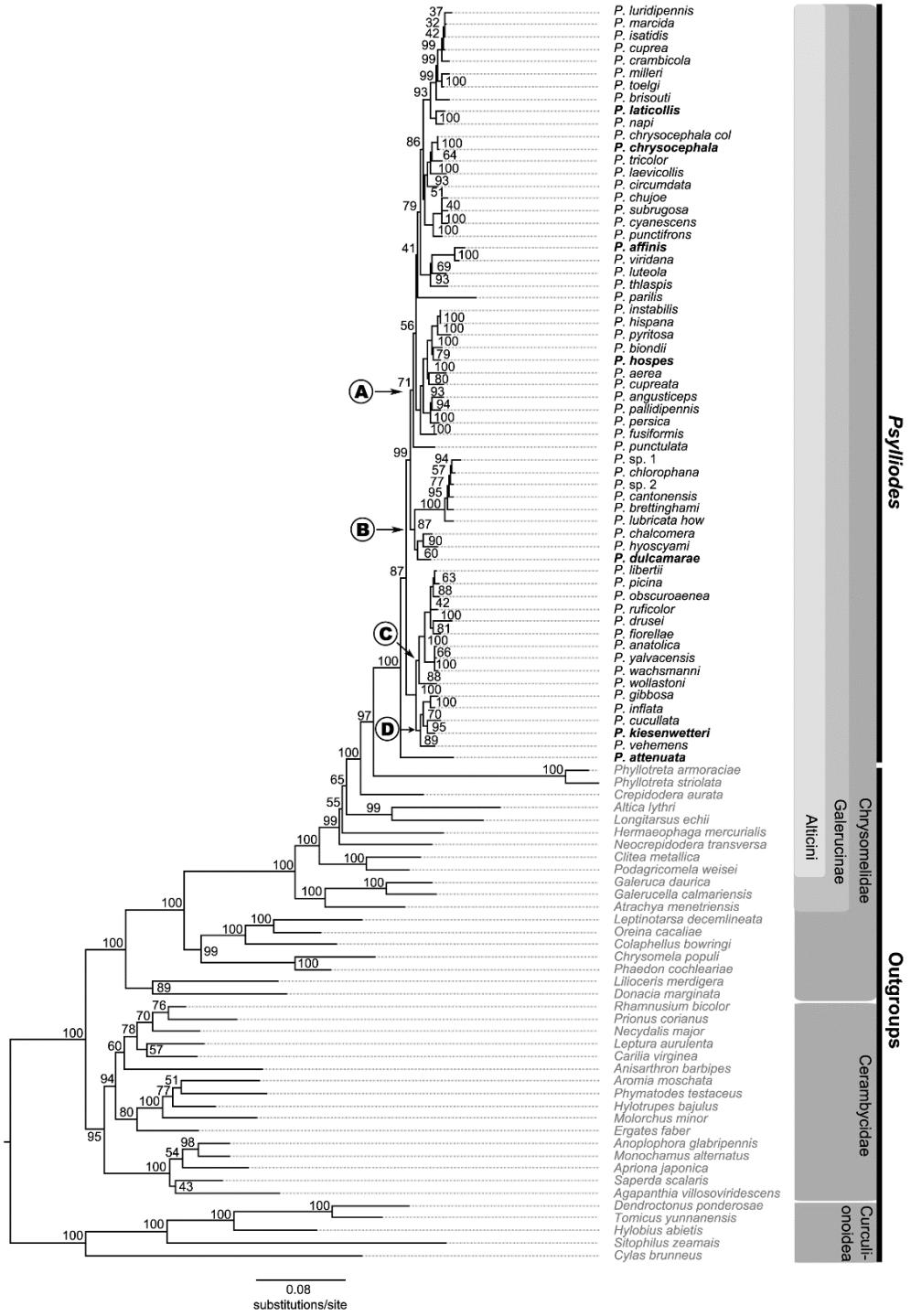
sequence divergence. This higher level of sequence divergence compared to our data is due to differences between our *P. pyritosa* COI sequence (MW254881), which shared 95.2% sequence identity with the published sequences from *P. pyritosa* (MH323339.1, MH323338.1, MH323337.1).

Several conclusions can be drawn from our analyses. First, there is strong evidence for low levels of COI sequence divergence (less than 5%) among several described *Psylliodes* species. Similar results have been reported for *Psylliodes* spp. and also other chrysomelid genera by Craven (2007) and Magoga et al. (2018), suggesting that COI sequences alone are sometimes not appropriate to reliably identify chrysomelid species. Second, our results provide first evidence for higher COI sequence divergence (more than 5%) between *P. chrysocephala* and *P. chrysocephala var. collaris*. We hope that our preliminary findings motivate comprehensive taxonomic comparisons of these and other difficult to differentiate *Psylliodes* species using genetic, morphological, and ecological data from different beetle populations. Third, efforts are needed to reconcile inconsistent COI sequence data (e.g., the case of *P. hispana* and *P. cupreata*) in publicly available databases with the help of specialists in *Psylliodes* taxonomy.

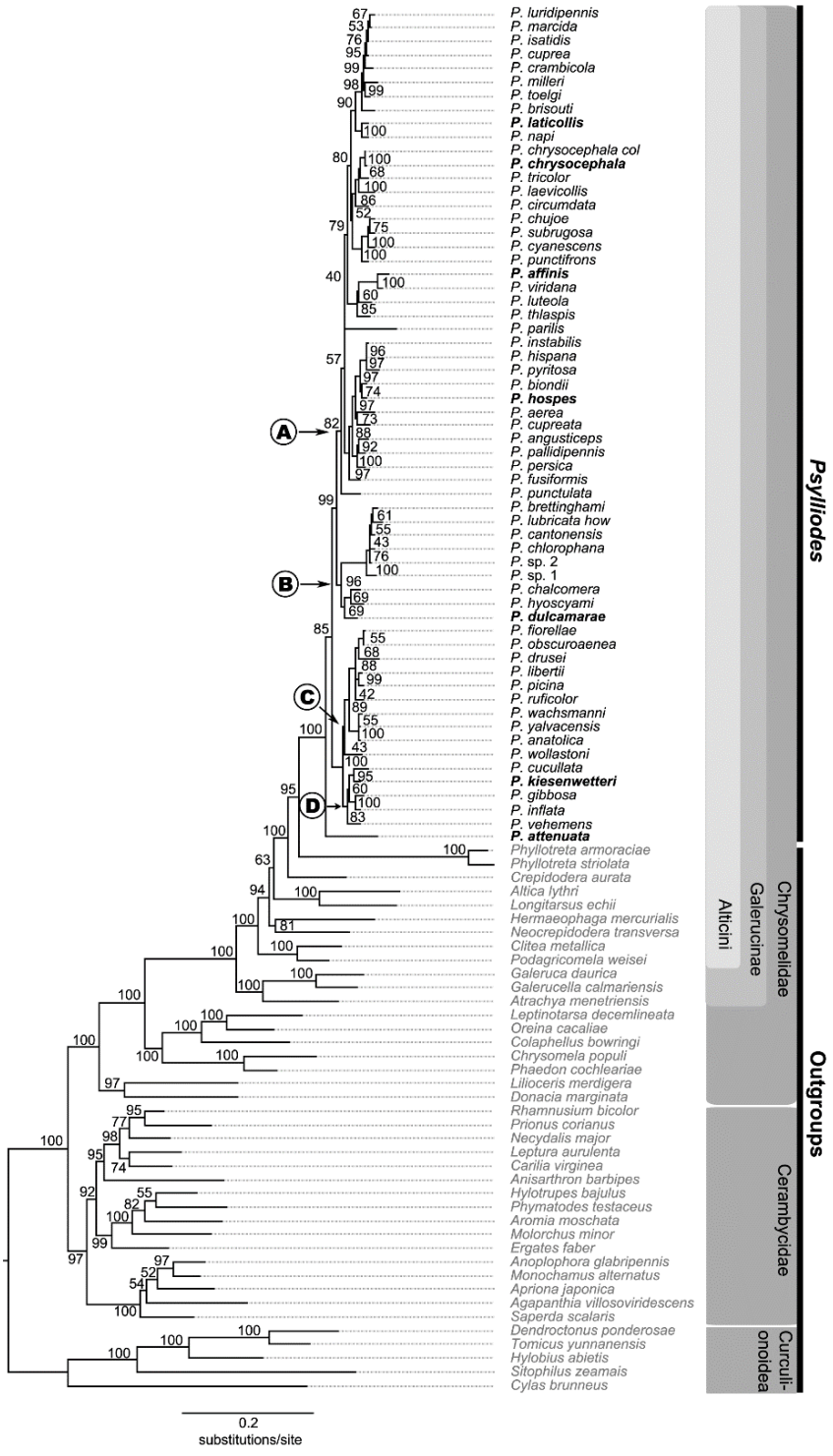
Supplementary Figures



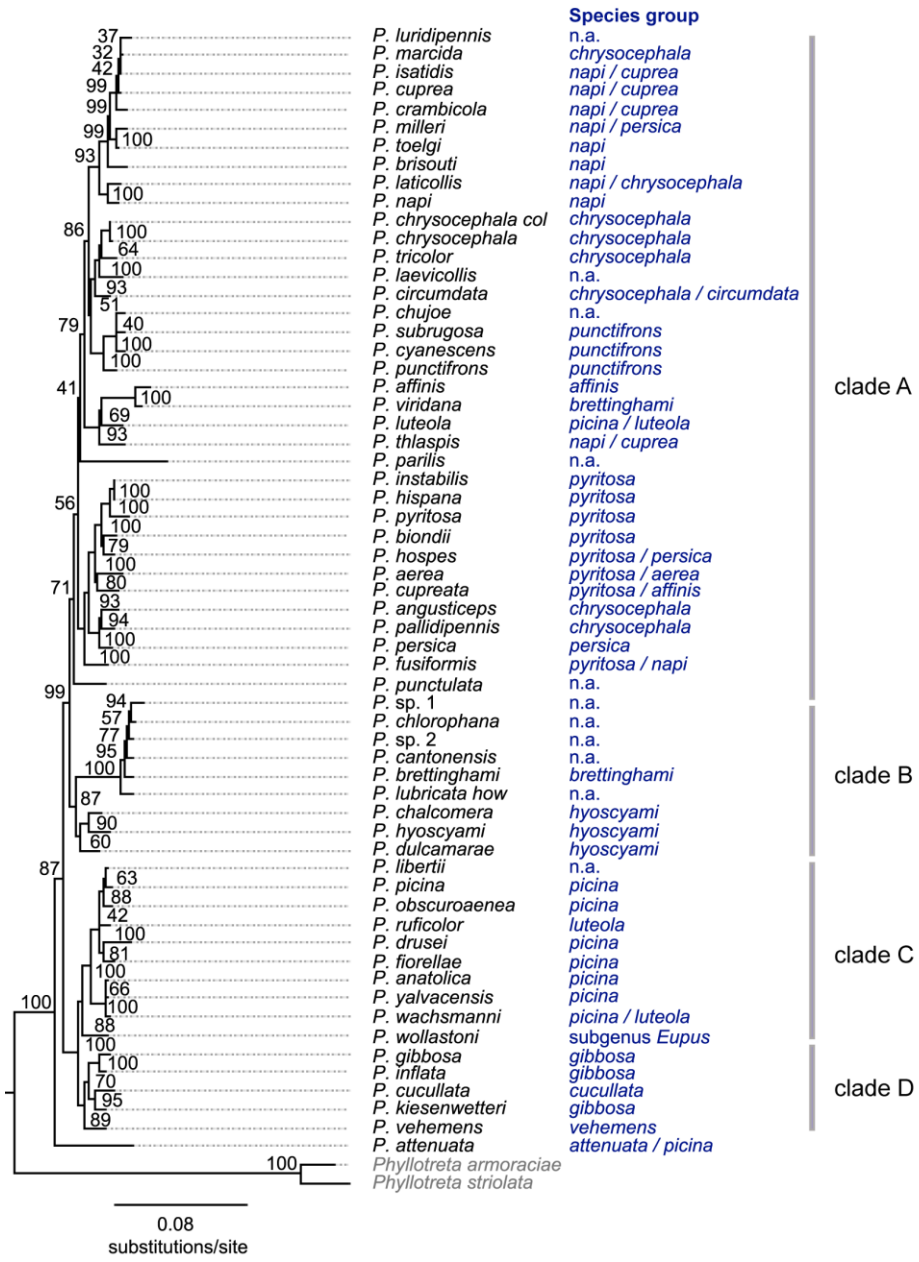
Supplementary Figure S1. Maximum clade credibility tree of the genus *Psylliodes* with median ages in million years ago (Ma) from Bayesian unlinked uniform analysis. The tree includes Chrysomelidae, Cerambycidae, and Curculionoidea species as outgroups. Ultrafast bootstrap values (1,000 replicates) are shown on each node. Letters A - D represent the major clades in *Psylliodes.* The node bars indicate 95% highest probability density (HPD) values on node height (age). *P. chrysocephala col*, *P. chrysocephala* var. *collaris; P. lubricata how, P. lubricata* var. *howensis*.



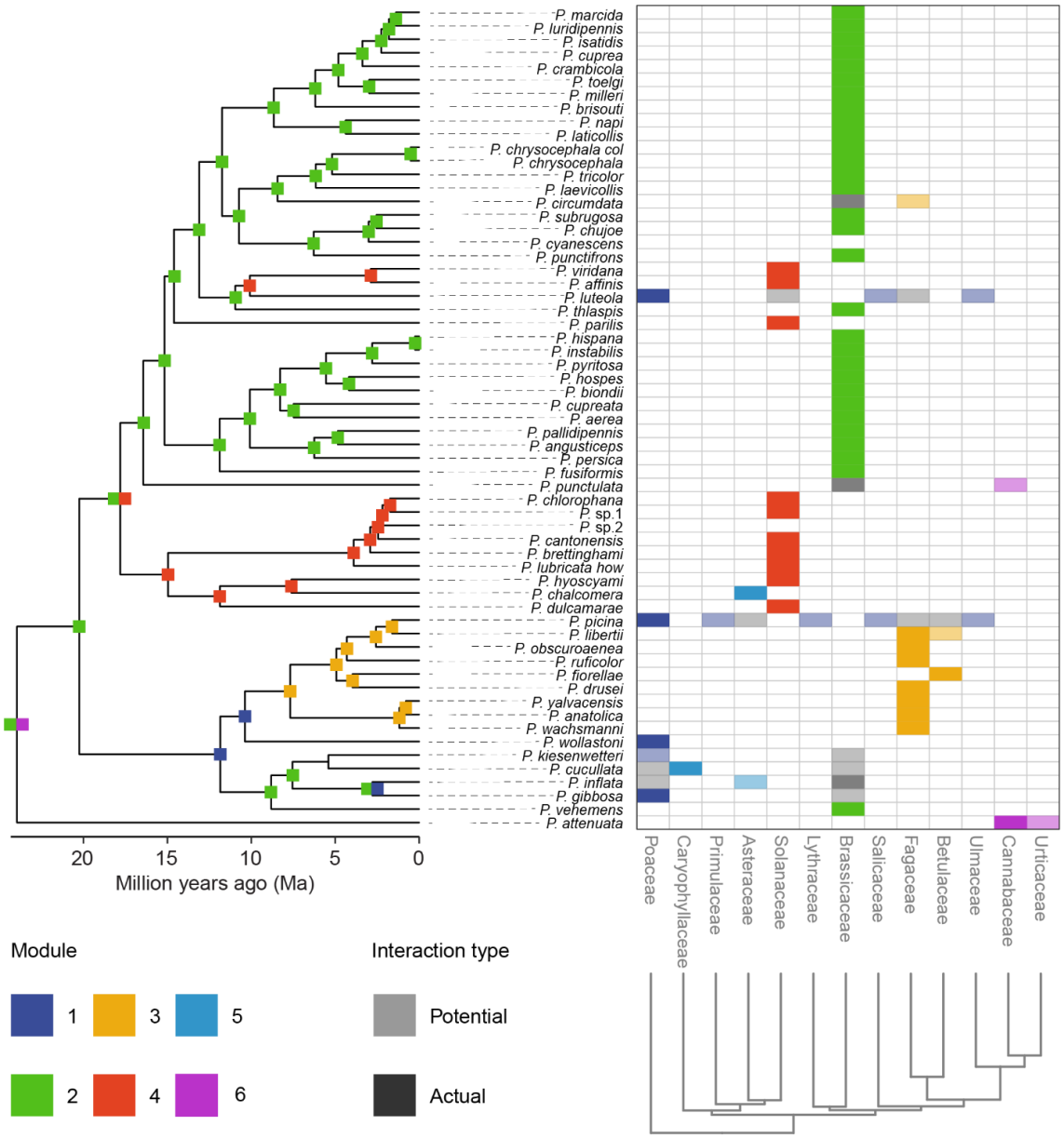
Supplementary Figure S2. Maximum likelihood phylogeny of the genus *Psylliodes* constructed from a combined nucleotide dataset using the first two codon positions of 8 single-copy nuclear genes and *cytochrome c oxidase I (COI)*. Ultrafast bootstrap support values (1,000 replicates) are shown on each node. Letters A - D represent the major clades in *Psylliodes.* Chrysomelidae, Cerambycidae, and Curculionoidea species were included as outgroups. *P. chrysocephala col*, *P. chrysocephala* var. *collaris; P. lubricata how, P. lubricata* var. *howensis*.



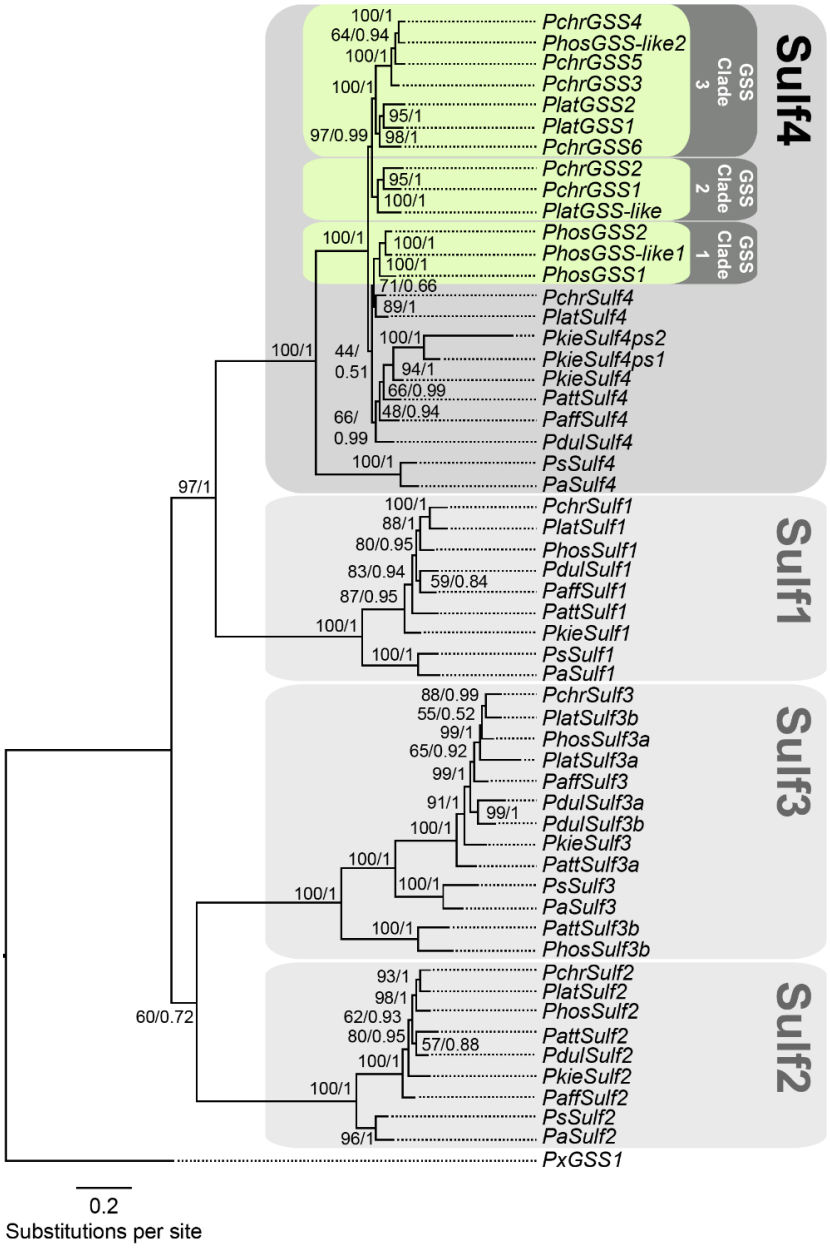
Supplementary Figure S3. Maximum likelihood phylogeny of the genus *Psylliodes* constructed from a combined amino acid dataset of 8 single-copy nuclear genes and *cytochrome c oxidase I (COI)*. Ultrafast bootstrap support values in 1,000 replicates are shown on each node. Letters A - D represent the major clades in *Psylliodes.* Chrysomelidae, Cerambycidae, and Curculionoidea species were included as outgroups. *P. chrysocephala col*, *P. chrysocephala* var*. collaris; P. lubricata how, P. lubricata* var*. howensis*.



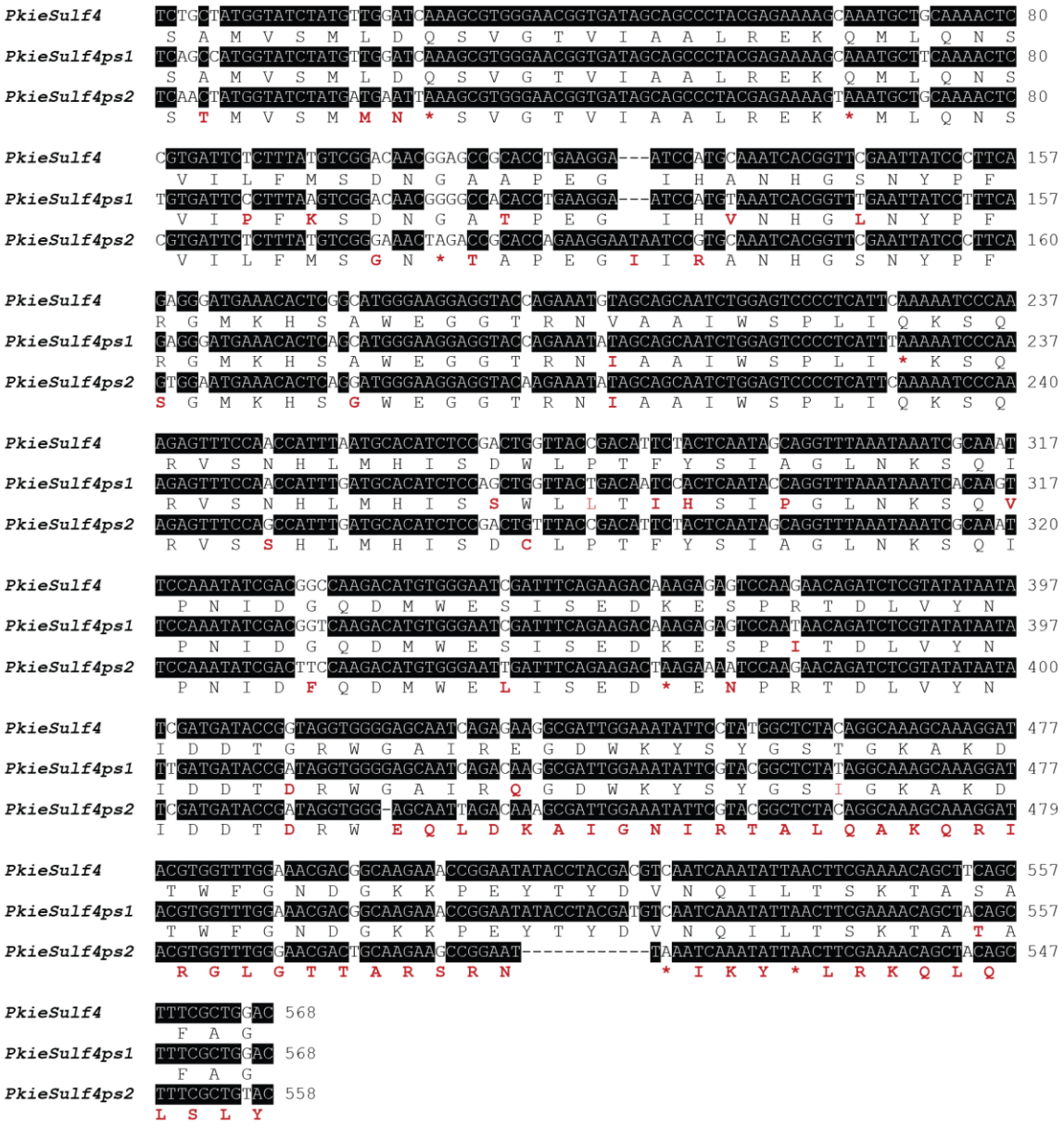
Supplementary Figure S4. Phylogeny and species groups of *Psylliodes* flea beetles. Maximum likelihood phylogeny of 60 *Psylliodes* species constructed from a combined nucleotide dataset using the first two codon positions of eight single-copy nuclear genes and *cytochrome c oxidase I* (corresponding to Supplementary Figure S2). Ultrafast bootstrap support values (1,000 replicates) are shown on each node. Except for *P. wollastoni*, which belongs to the subgenus *Eupus*, all species belong to the subgenus *Psylliodes* sensu stricto. Within *Psylliodes* s. str., several species groups have been proposed based on morphological characters (reviewed in Gikonyo et al. (2019). Depending on the study, some *Psylliodes* species have been assigned to different species groups (refer to Supplementary Table S1 for details). n.a., not assigned. *P. chrysocephala col*, *P. chrysocephala* var. *collaris; P. lubricata how, P. lubricata* var. *howensis*.



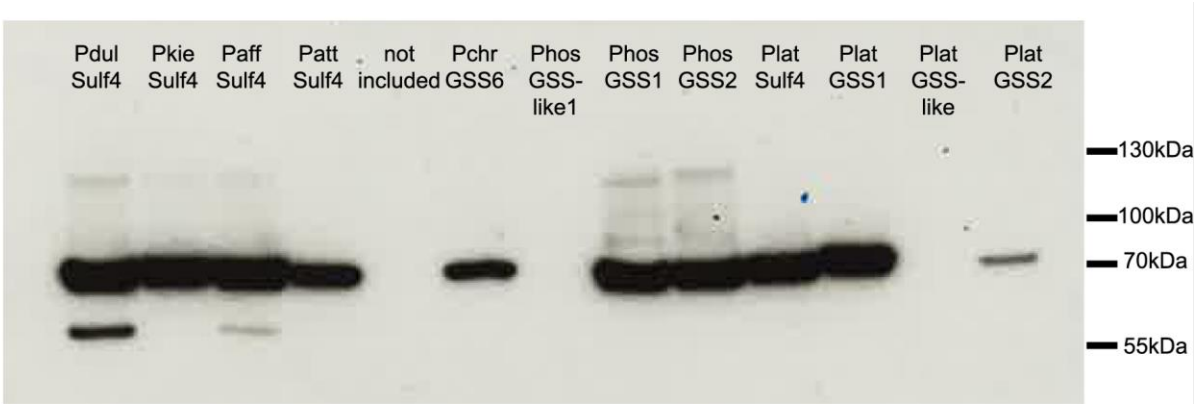
Supplementary Figure S5. Bayesian reconstruction of ancestral host repertoires of *Psylliodes* flea beetles showing interactions with marginal posterior probability ≥ 0.9. The evolution of host use is reconstructed along the phylogeny of *Psylliodes* flea beetles (left) based on observed beetle-plant interactions (right) in RevBayes. Each beetle species and host plant family is assigned to a module (M). Squares at the internal nodes of the *Psylliodes* phylogeny (left) are colored according to the module to which the host plant family belongs, i.e., Poaceae (M1), Brassicaeae (M2), Fagaceae/Betulaceae (M3), Solanaceae (M4), Asteraceae/Caryophyllaceae (M5), Cannabaceae/ Urticaceae (M6). Interactions between beetles and plants within modules are colored by module, whereas interactions between beetles and plants in different modules are shown in grey. Light colors and light grey indicate interactions with a potential host plant family or refugial plant family and dark colors indicate interactions with a major host plant family. *P. chrysocephala col*, *P. chrysocephala* var. *collaris; P. lubricata how, P. lubricata* var. *howensis*.



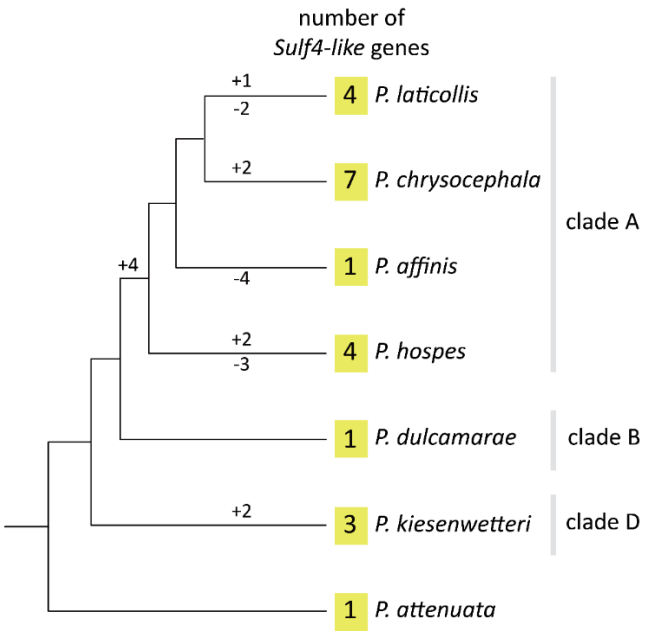
Supplementary Figure S6. Maximum likelihood and Bayesian inferred phylogeny constructed from arylsulfatase nucleotide sequences of *Psylliodes* and *Phyllotreta* spp. Ultrafast bootstrap support values (1,000 replicates) and Bayesian posterior probabilities are shown on each node. Sulf1, Sulf2, Sulf3, and Sulf4 designate the four major clades containing arylsulfatases from *Psylliodes* and are highlighted respectively. Clades of *Sulf4* genes, which encode enzymes with glucosinolate sulfatase activity (named GSS) are highlighted with a green background. The phylogeny was rooted using *PxGSS1* from *Plutella xylostella*. *Pchr, Psylliodes chrysocephala; Plat, P. laticollis; Phos, P. hospes; Paff, P. affinis; Pdul, P. dulcamarae; Pkie, P. kiesenwetteri; Patt, P. attenuata; Ps, Phyllotreta striolata; Pa, Phyllotreta armoraciae; Px, Plutella xylostella.*



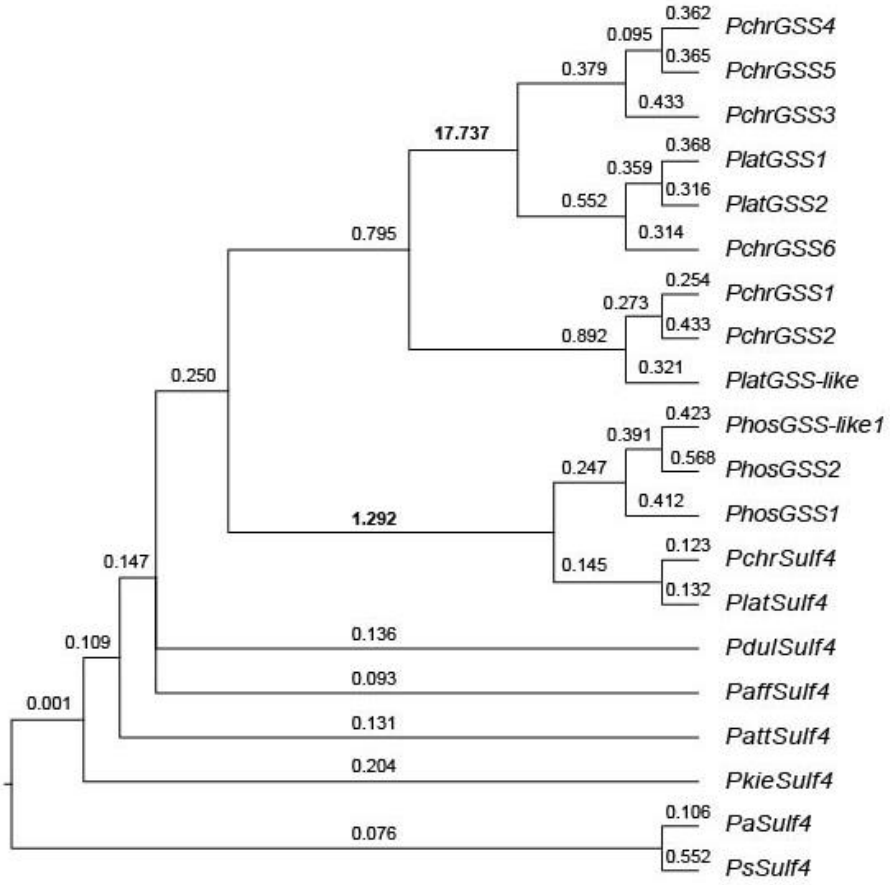
Supplementary Figure S7. Nucleotide alignment of partial *PkieSulf4*, *PkieSulf4ps1* and *PkieSulf4ps2* nucleotide sequences and corresponding translation. Identical nucleotides are highlighted in black. Amino acid residues of *PkieSulf4ps1* and *PkieSulf4ps2* differing from those of *PkiSulf4* are in bold red font. Substituted nucleotides and amino acids are highlighted in grey. Stop codons are represented by red asterisks (*). *Pkie, P. kiesenwetteri*.



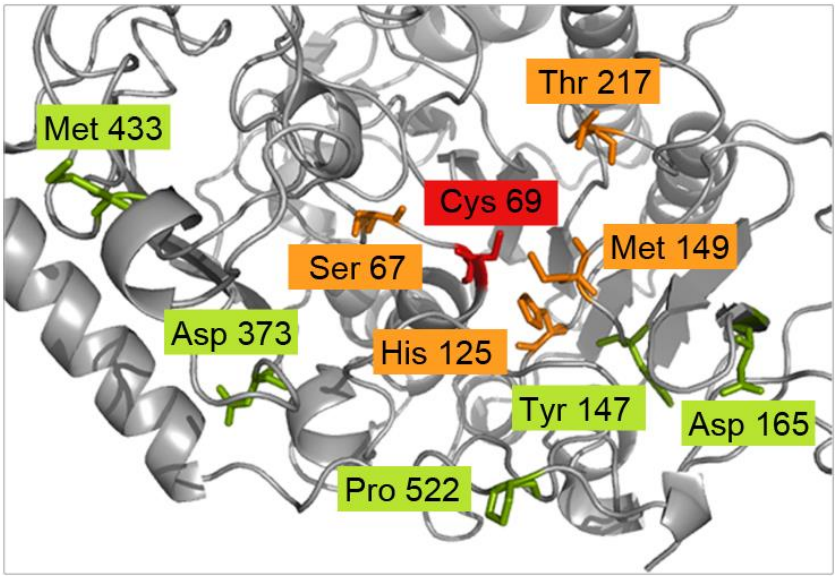
Supplementary Figure S8. Western blot of recombinant *Sulf4/GSS* genes from eight *Psylliodes* species. Recombinant *Sulf4/GSS* enzymes were heterologously expressed in *Sf*9cells. The species abbreviations are *Pdul, Psylliodes dulcamarae; Pkie, P. kiesenwetteri; Paff, P. affinis; Patt, P. attenuata; Pchr, P. chrysocephala; Phos, P. hospes; Plat, P. laticollis.*



Supplementary Figure S9. Estimation of gene duplication and loss events of *Sulf4-like* genes in the genus *Psylliodes* from Notung analysis. Numbers above branches indicate gene duplications, while those below branches indicate gene losses. The total number of *Sulf4-like* genes identified in the respective species is highlighted with a yellow background.



Supplementary Figure S10. Maximum likelihood tree of *Sulf4/GSS* genes from *Psylliodes* spp. and estimation of ω values (ratio of nonsynonymous (dN) to synonymous (dS) substitutions) for each branch. ω values > 1 indicating diversifying (positive) selection are written in bold font. Nodes with bootstrap support values ≤ 50 were collapsed for this analysis. The tree was rooted using *Sulf4* genes from *Phyllotreta striolata* (*PsSulf4*) and *Phyllotreta armoraciae* (*PaSulf4*). *Pchr, Psylliodes chrysocephala; Plat, P. laticollis; Phos, P. hospes; Paff, P. affinis; Pdul, P. dulcamarae; Pkie, P. kiesenwetteri; Patt, P. attenuata; Ps, Phyllotreta striolata; Pa, Phyllotreta armoraciae*.



Supplementary Figure S11. The predicted model of the three-dimensional structure of *Pchr*GSS2. The homology-based model was created using the crystal structure of N-acetylgalactosamine-4-sulfatase from *Homo sapiens* (pdb 1fsu.1.A) as template. The post-translationally modified cysteine essential for catalysis is shown in red. Active site residues under positive selection are marked in orange. Other amino acids under positive selection outside the active site are shown in green.

Literature Cited

Craven, J. C. (2007). *The evolution and conservation ecology of the Lundy cabbage and its beetles* [PhD Thesis, The University of Leeds]. Leeds, United Kingdom.

Magoga, G., Sahin, D. C., Fontaneto, D., & Montagna, M. (2018). Barcoding of Chrysomelidae of Euro-Mediterranean area: efficiency and problematic species. *Scientific Reports*, *8*(1), 13398. https://doi.org/10.1038/s41598-018-31545-9