

1 Rapid and independent evolution of ancestral and novel defenses in a genus of toxic plants
2 (*Erysimum*, Brassicaceae)

3

4 Tobias Züst^{*1}, Susan R. Strickler², Adrian F. Powell², Makenzie E. Mabry³, Hong An³, Mahdieh
5 Mirzaei², Thomas York², Cynthia K. Holland², Pavan Kumar^{2,8}, Matthias Erb¹, Georg Petschenka⁴,
6 José María Goméz⁵, Francisco Perfectti⁶, Caroline Müller⁷, J. Chris Pires³, Lukas A. Mueller², and
7 Georg Jander²

8

9 *correspondence to: tobias.zuest@ips.unibe.ch

10

11 ¹ Institute of Plant Sciences, University of Bern, Altenbergrain 21, CH-3013 Bern, Switzerland

12 ² Boyce Thompson Institute, 533 Tower Rd, Ithaca, NY 14853, USA

13 ³ Division of Biological Sciences, University of Missouri, Columbia, MO, 65211, USA

14 ⁴ Institut für Insektenbiotechnologie, Justus-Liebig-Universität Giessen, 35392 Giessen, Germany

15 ⁵ Department of Functional and Evolutionary Ecology, Estación Experimental de Zonas Áridas
16 (EEZA-CSIC), E-04120 Almería, Spain

17 ⁶ Department of Genetics, University of Granada, E-18071 Granada, Spain

18 ⁷ Department of Chemical Ecology, Bielefeld University, Universitätsstrasse 25, 33615 Bielefeld,
19 Germany

20 ⁸ Present address: Department of Entomology and Nematology, University of Florida, Gainesville, FL
21 32611, USA

22

23

24 **Abstract**

25 Phytochemical diversity is thought to result from coevolutionary cycles as specialization in herbivores
26 imposes diversifying selection on plant chemical defenses. Plants in the speciose genus *Erysimum*
27 (Brassicaceae) produce both ancestral glucosinolates and evolutionarily novel cardenolides as
28 defenses. Here we test macroevolutionary hypotheses on co-expression, co-regulation, and
29 diversification of these potentially redundant defenses across this genus. We sequenced and
30 assembled the genome of *E. cheiranthoides* and foliar transcriptomes of 47 additional *Erysimum*
31 species to construct a highly resolved phylogeny, revealing that cardenolide diversity increased
32 rapidly rather than gradually over evolutionary time. Concentrations, inducibility, and diversity of the
33 two defenses varied independently among species, with no evidence for trade-offs. Closely related
34 species shared similar cardenolide traits, but not glucosinolate traits, likely as a result of specific
35 selective pressures acting on distinct molecular diversification mechanisms. Ancestral and novel
36 chemical defenses in *Erysimum* thus appear to provide complementary rather than redundant
37 functions.

38

39 **Introduction**

40 Plant chemical defenses play a central role in the coevolutionary arms race with herbivorous insects.
41 In response to diverse environmental challenges, plants have evolved a plethora of structurally diverse
42 organic compounds with repellent, antinutritive, or toxic properties (Fraenkel 1959, Mithöfer and
43 Boland 2012). Chemical defenses can impose barriers to consumption by herbivores, but in parallel
44 may favor the evolution of specialized herbivores that can tolerate or disable these defenses (Cornell
45 and Hawkins 2003). Chemical diversity is likely evolving in response to a multitude of plant-
46 herbivore interactions (Salazar et al. 2018), and community-level phytochemical diversity may be a
47 key driver of niche segregation and insect community dynamics (Richards et al. 2015, Sedio et al.
48 2017).

49 For individual plants, the production of diverse mixtures of chemicals is often considered
50 advantageous (Romeo et al. 1996, Firn and Jones 2003, Gershenson et al. 2012, Forbey et al. 2013,
51 Richards et al. 2016). For example, different chemicals may target distinct herbivores (Iason et al.
52 2011, Richards et al. 2015), or may act synergistically to increase overall toxicity of a plant (Steppuhn
53 and Baldwin 2007). However, metabolic constraints can limit the extent of phytochemical diversity
54 within individual plants (Firn and Jones 2003). Most defensive metabolites originate from a small
55 group of precursor compounds and conserved biosynthetic pathways, which are modified in a
56 hierarchical process into diverse, species-specific end products (Moore et al. 2014). As constraints are
57 likely strongest for the early stages of these pathways, related plant species commonly share the same
58 functional ‘classes’ of defensive chemicals (Wink 2003), but vary considerably in the number of
59 compounds within each class (Fahey et al. 2001, Rasmann and Agrawal 2011).

60 Functional conservatism in defensive chemicals among related plants should facilitate host
61 expansion and the evolution of tolerance in herbivores (Cornell and Hawkins 2003), as specific
62 adaptations to deactivate or detoxify one compound are more likely to be effective against structurally
63 similar than structurally dissimilar compounds. This may result in a seemingly paradoxical scenario,
64 wherein well-defended plants are nonetheless attacked by a diverse community of specialized
65 herbivores (Agrawal 2005, Bidart-Bouzat and Kliebenstein 2008). For example, most plants in the
66 Brassicaceae produce glucosinolates as their primary defense, which upon activation by myrosinase
67 (thioglucoside glucohydrolase) enzymes at leaf damage become potent repellents of many herbivores
68 (Fahey et al. 2001). However, despite the potency of this defense system and the large diversity of
69 glucosinolates produced by the Brassicaceae, several specialized herbivores have evolved strategies to
70 overcome this defense, enabling them to consume most Brassicaceae and even to sequester
71 glucosinolates for their own defense against predators (Müller 2009, Winde and Wittstock 2011).

72 Plants may occasionally overcome the constraints on functional diversification and gain the
73 ability to produce new classes of defensive chemicals as a ‘second line of defense’ (Feeny 1977).
74 Although this phenomenon is likely widespread across the plant kingdom, it has most commonly been
75 reported from the well-studied Brassicaceae. In addition to producing evolutionarily ancestral

76 glucosinolates, plants in this family have gained the ability to produce saponins in *Barbarea vulgaris*
77 (Shinoda et al. 2002), alkaloids in *Cochlearia officinalis* (Brock et al. 2006), cucurbitacins in *Iberis*
78 spp. (Nielsen 1978b), alliarinoside in *Alliaria petiolata* (Frisch and Møller 2012), and cardenolides in
79 the genus *Erysimum* (Makarevich et al. 1994). These recently-evolved chemical defenses with modes
80 of action distinct from glucosinolates have likely allowed the plants to escape attack from specialized,
81 glucosinolate-adapted herbivores (Nielsen 1978b, Dimock et al. 1991, Haribal and Renwick 2001,
82 Shinoda et al. 2002). Gains of novel defenses are expected to result in a release from selective
83 pressures imposed by specialized antagonists, and thus may represent key steps in herbivore-plant
84 coevolution that lead to rapid phylogenetic diversification (Weber and Agrawal 2014).

85 The production of cardenolides by species in the genus *Erysimum* is one of the longest- and
86 best-studied examples of an evolutionarily recent gain of a novel chemical defense (Jaretzky and
87 Wilcke 1932, Nagata et al. 1957, Singh and Rastogi 1970, Makarevich et al. 1994). Cardenolides are a
88 type of cardiac glycoside, which act as allosteric inhibitors of Na^+/K^+ -ATPase, an essential membrane
89 ion transporter that is expressed ubiquitously in animal cells (Agrawal et al. 2012). Cardiac glycosides
90 are produced by plants in approximately sixty genera belonging to twelve plant families, and several
91 cardiac glycoside-producing plants are known for their toxicity or medicinal uses (Agrawal et al.
92 2012, Züst et al. 2018). *Erysimum* is a species-rich genus consisting of diploid and polyploid species
93 with diverse morphologies, growth habits, and ecological niches (Al-Shehbaz 1988, Polatschek and
94 Snogerup 2002, Al-Shehbaz 2010, Gómez et al. 2015). Of the *Erysimum* species evaluated to date, all
95 produced some of the novel cardenolide defenses (Makarevich et al. 1994). Previous phylogenetic
96 studies suggest a recent and rapid diversification of the genus, with estimates of the onset of radiation
97 ranging between 0.5 and 2 million years ago (Gómez et al. 2014, Moazzeni et al. 2014), and of 150 to
98 350 extant species (Polatschek and Snogerup 2002, Al-Shehbaz 2010). The large uncertainty in
99 species number reflects taxonomic challenges in this genus, which includes many species that readily
100 hybridize, as well as cryptic species with near-identical morphology (Abdelaziz et al. 2011).

101 In most *Erysimum* species, cardenolides appear to have enabled an escape from at least some
102 glucosinolate-adapted specialist herbivores. Cardenolides in *Erysimum* act as oviposition and feeding
103 deterrents for different pierid butterflies (Chew 1975, 1977, Wiklund and Åhrberg 1978, Renwick et
104 al. 1989, Dimock et al. 1991), and several glucosinolate-adapted beetles (*Phaedon* spp. and
105 *Phyllotreta* spp.) were deterred from feeding by dietary cardenolides at levels commonly found in
106 *Erysimum* (Nielsen 1978a, b). Nonetheless, *Erysimum* plants are still attacked by a range of
107 herbivores and seed predators, including some mammals and several glucosinolate-adapted aphids,
108 true bugs, and lepidopteran larvae (Gómez 2005, Züst et al. 2018). Despite their potency, cardenolides
109 thus do not provide a universal defense.

110 The gain of a novel chemical defense makes the genus *Erysimum* an excellent model system
111 to study the causes and consequences of phytochemical diversification (Züst et al. 2018). While an
112 increasing number of studies are beginning to describe taxon-wide patterns of chemical diversity in

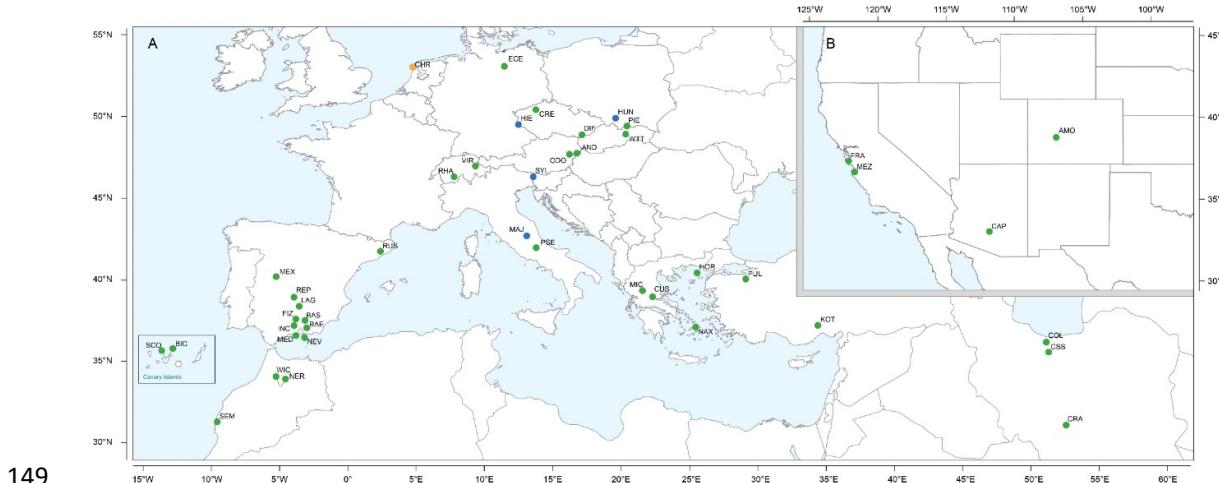
113 plants (e.g., Richards et al. 2015, Sedio et al. 2017, Salazar et al. 2018), the *Erysimum* system is
114 unique in combining two classes of plant metabolites with primarily defensive function – although a
115 broader role of glucosinolates is increasingly recognized (e.g., Katz et al. 2015). The system thus is
116 ideally suited to evaluate the evolutionary consequences of co-expressing two functionally distinct but
117 potentially redundant defenses. Here, we present a high-quality genome sequence assembly and
118 annotation for the short-lived annual *E. cheiranthoides* as an important resource for future molecular
119 studies in this system. Furthermore, we present a highly resolved phylogeny for 47 additional species
120 constructed from transcriptome sequences, corresponding to 10-30% of species in the genus
121 *Erysimum*. We combine this phylogeny with a characterization of the full diversity of glucosinolates
122 and cardenolides in leaves to evaluate macroevolutionary patterns in the evolution of phytochemical
123 diversity across the genus. We complemented the characterization of defensive phenotypes by
124 quantifying glucosinolate-activating myrosinase activity, inhibition of animal Na⁺/K⁺-ATPase by leaf
125 extracts, and defense inducibility in response to exogenous application of jasmonic acid (JA). By
126 assessing co-variation of diversity, abundance and inducibility of ancestral and novel defenses, we
127 provide evidence that the two defense metabolite classes evolved in response to different selective
128 pressures and appear to serve specific, non-redundant roles.

129

130 **Materials and Methods**

131 ***Plant material and growth conditions***

132 The genus *Erysimum* is distributed across the northern hemisphere, with the center of diversity
133 stretching from the Mediterranean Basin into Central Asia, and a smaller number of species centered
134 in western North America (Moazzeni et al. 2014). Seeds of *Erysimum* species spanning a range of
135 distributions in Europe and Western North America were collected in their native habitats or obtained
136 from botanical gardens and commercial seed suppliers (Figure 1, Table S1). Ploidy levels of species
137 were inferred from literature reports to test for the effect of ploidy on chemical diversity (Table S1).
138 For seeds obtained from botanical gardens, we mostly used species names as provided by the supplier.
139 As an exception, seeds of *E. collinum* (COL) had originally been designated as *E. passgalense*, but
140 these species names are now considered as synonymous (German 2014). Furthermore, plants of four
141 seed batches did not exhibit the expected phenotypes and likely were the result of seed mislabeling by
142 the suppliers; we nonetheless included these plants for transcriptome sequencing, but refer to them as
143 accessions ER1, ER2, ER3, and ER4 (Table S1). For genome sequencing of *E. cheiranthoides*, seeds
144 that were collected from a natural population in the Elbe River floodplain (Germany, Figure 1) were
145 planted in a greenhouse in one-liter pots in Cornell mix (by weight 56% peat moss, 35% vermiculite,
146 4% lime, 4% Osmocote slow-release fertilizer [Scotts, Marysville, OH], and 1% Unimix [Scotts]).
147 This lineage, which we have designated “Elbtalaue”, was propagated by self-pollination and single-
148 seed descent for six generations prior to further experiments.



149

150 **Figure 1.** Geographic location of *Erysimum* spp. source populations in Europe (A) and North America (B).
151 Inset: The Canary Islands (28°N , 16°W) are located further westward and southward than drawn in this map.
152 Green symbols are exact collection locations, while blue symbols indicate approximate locations based on
153 species distributions. Seeds of the originally Mediterranean species *E. cheiri* (CHR, orange symbol) were
154 collected from a naturalized population in the Netherlands. Five species/accessions (ALI, ER1, ER2, ER3, ER4)
155 could not be placed on the map due to uncertain species identity.

156

157 For transcriptome sequencing and metabolomic analyses of *Erysimum* species, subsets of the full
158 species pool were grown in three separate experiments in 2016 and 2017. While some species were
159 included in all three experiments, others could only be grown once due to limited seed availability or
160 germination. To maximize germination success, seeds were placed on water agar (1%) in Petri dishes
161 and cold-stratified for two weeks. After stratification, Petri dishes were moved to a growth chamber
162 set to 24°C day / 22°C night at a 16:8 h photoperiod. Viable seeds germinated within 3-10 days of
163 placement in the growth chamber. As soon as cotyledons had fully extended, we transplanted the
164 seedlings into 10×10 cm plastic pots filled with a mixture of peat-based germination soil
165 (Seedlingsubstrat, Klasmann-Deilmann GmbH, Geeste, Germany), field soil, sand, and vermiculite at
166 a ratio of 6:3:1:5. Plants were moved to a climate-controlled greenhouse set to 24°C day / 16°C night
167 and 60 % RH with natural light and supplemented artificial light set to a 14:10 h photoperiod. Plants
168 were watered as needed throughout the experiments, and fertilized with a single application of 0.1 L
169 of fertilizer solution (N:P:K 8:8:6, 160 ppm N) three weeks after transplanting.

170

171 ***Erysimum cheiranthoides* genome and transcriptome sequencing**

172 DNA sequencing for genome assembly and RNA sequencing for annotation were conducted with
173 samples prepared from sixth-generation inbred *E. cheiranthoides* var. *Elbtalaue*. High molecular
174 weight genomic DNA was extracted from the leaves of a single *E. cheiranthoides* plant using
175 Wizard® Genomic DNA Purification Kit (Promega, Madison WI, USA). The quantity and quality of
176 genomic DNA was assessed using a Qubit 3 fluorometer (Thermo Fisher, Waltham, MA, USA) and a

177 Bioanalyzer DNA12000 kit (Agilent, Santa Clara, CA, USA). Twelve µg of non-sheared DNA were
178 used to prepare the SMRTbell library, and the size-selection of 15-50 kb was performed on Sage
179 BluePippin (Sage Science, Beverly, MA, USA) following manufacturer's instructions (Pacific
180 Biosciences, Menlo Park, CA, USA) and as described previously (Chen et al. 2019). PacBio
181 sequencing was performed by the Sequencing and Genomic Technologies Core of the Duke Center
182 for Genomic and Computational Biology (Durham, NC, USA). For genome polishing, one DNA
183 library was prepared using the PCR-free TruSeq DNA sample preparation kit following the
184 manufacturer's instructions (Illumina, San Diego, CA), and sequenced on an Illumina MiSeq
185 instrument (paired-end 2×250bp) at the Cornell University Biotechnology Resource Center (Ithaca,
186 NY).

187 The transcriptome of sixth-generation inbred *E. cheiranthoides* var. *Elbtalaue* plants was
188 sequenced using both PacBio (Iso-Seq) and Illumina sequencing methods. Total RNA was isolated
189 from stems, flowers, buds, pods, young and mature leaves of five plants (siblings of the plant used for
190 genome sequencing) using the SV Total RNA Isolation Kit with on-column DNase I treatment
191 (Promega, Madison, WI, USA). The RNA quantity and quality were assessed by RIN (RNA Integrity
192 Number) using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). The samples with a RIN
193 value of >7 were pooled across all six tissue types. One µg of the pooled total RNA was used for the
194 Iso-Seq following the manufacturer's instructions (Iso-SeqTM). The library preparation and sequencing
195 were performed by Sequencing and Genomic Technologies Core of the Duke Center for Genomic and
196 Computational Biology (Durham, NC, USA). For Illumina sequencing, 2 µg of purified pooled total
197 RNA from three replicates was used for the preparation of strand-specific RNAseq libraries with 14
198 cycles of final amplification (Zhong et al 2012). The purified libraries were multiplexed and
199 sequenced with 101 bp paired-end read length in two-lanes on an Illumina HiSeq2500 instrument
200 (Illumina, San Diego, CA) at the Cornell University Biotechnology Resource Center (Ithaca, NY). For
201 Hi-C scaffolding, 500 mg of *E. cheiranthoides* leaf tissue was flash-frozen and sent to Phase
202 Genomics (Phase Genomics Inc. Seattle, WA, USA).

203

204 ***E. cheiranthoides* genome assembly and gene annotation**

205 PacBio sequences from the genome of *E. cheiranthoides* were assembled using Falcon (Chin et al.
206 2016). The assembly was polished using Arrow from SMRT Analysis v2.3.0
207 (<https://www.pacb.com/products-and-services/analytical-software/smrt-analysis/>) with PacBio reads,
208 and then assembled into chromosome-scale scaffolds using Hi-C methods by Phase Genomics
209 (Seattle, WA, USA). Scaffolding gaps were filled with PBJelly v13.10 (English et al. 2012) using
210 PacBio reads followed by three rounds of Pilon v1.23 correction (Walker et al. 2014) with 9 Gbp of
211 Illumina paired-end 2 x 150 reads. BUSCO v3 (Waterhouse et al. 2018) metrics were used to assess
212 the quality of the genome assemblies.

213 For gene model prediction, *de novo* repeats were predicted using RepeatModeler v1.0.11
214 (Smit, AFA, Hubley, R. *RepeatModeler Open-1.0*. 2008-2015 <http://www.repeatmasker.org>), known
215 protein domains were removed from this set based on identity to UniProt (Boutet et al. 2007) with the
216 ProtExcluder.pl script from the ProtExcluder v1.2 package (Campbell et al. 2014), and the output was
217 then used with RepeatMasker v4-0-8 (Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-4.0*.
218 2013-2015 <http://www.repeatmasker.org>) in conjunction with the Repbase library. For gene
219 prediction, RNA-seq reads were mapped to the genome with hisat2 v2.1.0 (Kim et al. 2015).
220 Portcullis v1.1.2 (Mapleson et al. 2018) and Mikado v1.2.2 (Venturini et al. 2018) were used to filter
221 the resulting bam files and make first-pass gene predictions. PacBio IsoSeq data were corrected using
222 the Iso-Seq classify + cluster pipeline (Gordon et al. 2015). Augustus v3.2 (Stanke et al. 2008) and
223 Snap v2.37.4ubuntu0.1 (Korf 2004) were trained and then implemented through the Maker pipeline
224 v2.31.10 (Cantarel et al. 2008) with Iso-Seq, proteins from Swiss-Prot, and processed RNA-seq added
225 as evidence. Functional annotation was performed with BLAST v2.7.1+ (Altschul et al. 1990) and
226 InterProScan v.5.36-75.0 (Jones et al. 2014).

227

228 *Repeat analysis*

229 The genome of *E. cheiranthoides* was analyzed for LTR retrotransposons using LTRharvest
230 (Ellinghaus et al. 2008), included in GenomeTools v1.5.10, with the parameters “-seqids yes -
231 minlenltr 100 -maxlenltr 5000 -mindistltr 1000 -motif TGCA -motifmis 1 -maxdistltr 15000 -similar
232 85 -mintsd 4 -maxtsd 6 -vic 10 -seed 20 -overlaps best”. The genome was also analyzed using
233 LTR_FINDER v1.07 (Xu and Wang 2007) with parameters “-D 15000 -d 1000 -L 5000 -l 100 -p 20 -
234 C -M 0.85 -w 0”. The results from LTRharvest and LTR_FINDER were then passed as inputs to
235 LTR_retriever v2.0 (Ou and Jiang 2018) using default parameters, including a neutral mutation rate
236 set at 1.3×10^{-8} .

237 Using the LTR_retriever repeat library, the genome was masked with RepeatMasker v4.0.7,
238 and additional repetitive elements were identified *de novo* in the genome using RepeatModeler. These
239 repeats were used with blastx v2.7.1+ (Altschul et al. 1990) against the Uniprot and Dfam libraries
240 and protein-coding sequences were excluded using the ProtExcluder.pl script from the ProtExcluder
241 v1.2 package (Campbell et al. 2014). The masked genome was then re-masked with RepeatMasker,
242 with the repeat library obtained from RepeatModeler. Coverage percentages for repeat types were
243 obtained using the fam_coverage.pl and fam_summary.pl scripts, which are included with
244 LTR_retriever. All percentages were calculated based on the total length of the assembly.

245

246 *Genome-wide plot of genic sequence and repeats*

247 A circular representation of the *E. cheiranthoides* genome was made with Circos v0.69-6 (Krzywinski
248 et al. 2009). Gene and repeat densities were calculated by generating 1Mb windows and by
249 calculating percent coverage for the features using bedtools coverage v2.26.0 (Quinlan and Hall

250 2010). The coverage values from the repeat library and the genome annotation were calculated
251 independently of each other. Similarly, the total percentage of genic sequence for the genome was
252 also calculated using bedtools genomecov v2.26.0. The gene and repeat percentages for the 1Mb
253 windows were then plotted as histogram tracks in Circos.

254 For analysis of synteny between *E. cheiranthoides* and *Arabidopsis thaliana* (Arabidopsis;
255 TAIR10; www.arabidopsis.org), the genome sequences were aligned using NUCmer, from MUMmer
256 v3.23 (Kurtz et al. 2004) with the parameters --maxgap=500 --mincluster=100. The alignments were
257 filtered with delta-filter -r -q -i 90 -l 1000 and coordinates of aligned segments were extracted with
258 show-coords. The extracted coordinates were then used as the links input for Circos.

259

260 ***Glucosinolate and myrosinase gene annotation in E. cheiranthoides***

261 Known glucosinolate biosynthetic genes were annotated in *E. cheiranthoides* based on homology to
262 Arabidopsis pathway genes in AraCyc (Rhee et al. 2006). Coding sequences for the Arabidopsis
263 glucosinolate and myrosinase biosynthetic genes were obtained from NCBI and used in a BLASTn
264 query against the *E. cheiranthoides* coding sequence dataset. A threshold percent identity of 80% and
265 70% or higher was set for glucosinolate and myrosinase genes, respectively. Protein sequences of
266 myrosinase gene homologs in Arabidopsis and *E. cheiranthoides* were aligned using MUSCLE with
267 UPGMA clustering and a phylogenetic tree was generated in MEGA X v10.0.4(Kumar et al. 2018)
268 using the neighbor-joining method by sampling 1000 bootstrap replicates.

269

270 ***Transcriptome sequencing of Erysimum species***

271 To generate a high number of gene sequences required for a well-resolved phylogeny, we sequenced
272 the foliar transcriptomes of 48 *Erysimum* species or accessions, including a first-generation inbred *E.*
273 *cheiranthoides* var. *Elbtalaue*. Transcriptomes were generated from pooled leaf material of several
274 individuals collected in the same experiment; five species were sequenced from plants in experiment
275 2016, 18 species from experiment 2017-1, and 25 species from experiment 2017-2 (Table S1). Leaf
276 material was harvested 5-7 weeks after plants were transplanted into soil. To average environmental
277 and individual effects on RNA expression, we pooled leaf material from 2-5 individual plants from
278 one or two time points (separated by 1-2 weeks) to create a single pooled RNA sample per species
279 (see Table S1 for details). For large-leaved species, we collected approximately 50 mg of fresh plant
280 material from each harvested plant using a heat-sterilized hole punch (0.5 cm diameter). For smaller-
281 leaved species, we collected an equivalent amount of material by harvesting multiple whole leaves.
282 All leaf tissue was immediately snap frozen in liquid nitrogen and stored at -80 °C until further
283 processing. For sample pooling, we combined leaf material of individual plants belonging to the same
284 species in a mortar under liquid nitrogen and ground all material to a fine powder. We then weighed
285 out 50-100 mg of frozen pooled powder for each species.

286 We extracted RNA from pooled leaf material using the RNeasy Plant Mini Kit (Qiagen AG,
287 Hombrechtikon, Switzerland), including a step for on-column DNase digestion, and following the
288 manufacturer's instructions. The purified total RNA was dissolved in 50 µL RNase-free water, split
289 into three aliquots, and stored at -80 °C until further processing. Assessment of RNA quality, library
290 preparation, and sequencing were all performed by the Next Generation Sequencing Platform of the
291 University of Bern (Bern, Switzerland). RNA quality was assessed in one aliquot per extract using a
292 Fragment Analyzer (Model CE12, Agilent Technologies, Santa Clara, USA), and samples with low
293 RIN scores (<7) were re-extracted and assessed again for quality. RNA libraries for TruSeq Stranded
294 mRNA (Illumina, San Diego, USA) were assembled for each species and multiplexed in groups of
295 eight, using unique index combinations (Illumina 2017). Groups of eight multiplexed libraries were
296 run individually on single lanes (for a total of six lanes) of an Illumina HiSeq 3000 sequencer using
297 150 bp paired-end reads.
298

299 ***De novo assembly of transcriptomes***

300 RNA-seq data were cleaned with fastq-mcf v1.04.636 (<https://github.com/ExpressionAnalysis/ea-utils/blob/wiki/FastqMcf.md>) using the following parameters: quality = 20, minimum read length =
301 50. Filtered reads were assembled using Trinity v2.4.0 (Haas et al. 2013). The longest ORF was
302 determined using TransDecoder v5.5.0 (<https://github.com/TransDecoder>). BUSCO v2 (Waterhouse
303 et al. 2018) was run with lineage Embryophyta to assess gene representation and Orthofinder v2.3.1
304 (Emms and Kelly 2015) was used to cluster proteins from all 48 transcriptomes into orthogroups.
305

307 ***Phylogenetic tree construction***

308 We constructed phylogenetic trees using two alternative methods. In the first approach, we translated
309 the assembled transcriptomes using TransDecoder v5.5.0. We then followed the Genome-Guided
310 Phylo-Transcriptomics Pipeline (Washburn et al. 2017) to infer orthologous genes using synteny
311 between genomes of *E. cheiranthoides* and Arabidopsis (TAIR10). Briefly, we obtained 26,830
312 orthologs between *E. cheiranthoides* and Arabidopsis through the syntenic blocks that were identified
313 by SynMap from CoGe (<https://genomevolution.org/coge/SynMap.pl>). Sequences for each of the 48
314 *Erysimum* species and Arabidopsis (TAIR10_pep_20101214; www.arabidopsis.org) (total of 49
315 samples) were annotated using protein sequences of the orthologs using blastp v2.7.1 with an e-value
316 < 10⁻⁴ and identity > 85%. After annotation, single copy genes and one copy of repetitive genes were
317 kept if they were present in more than 39 (>80% of 49) species. In total, we recovered 11,890 genes,
318 9,868 of which had orthologs with Arabidopsis. Each of these 9,868 genes was aligned using MAFFT
319 v7.394 (Katoh et al. 2002), and cleaned using Phyutility v2.2.6 (Smith and Dunn 2008) with the
320 parameter *-clean 0.3*. Maximum-likelihood tree estimation for each gene was constructed using
321 RAxML v8.2.8 (Stamatakis 2014) using the PROTCATWAG model with 100 bootstrap replicates.

322 Finally, coalescent species tree inference was performed with these 9,868 gene trees as input, using
323 ASTRAL-III v5.6.3 (Zhang et al. 2018).

324 In the second approach, we used transcriptome sequences translated by TransDecoder to
325 predict protein sequences, after which the longest predicted protein for each gene was retained. For *E.*
326 *cheiranthoides*, the genome rather than the transcriptome sequence was used. Next, we constructed
327 gene families by running OrthoFinder v1.1.10 (Emms and Kelly 2015) on a subset of 18 *Erysimum*
328 species ('E18'), seven other Brassicaceae species with published genomes (*A. thaliana*, *A. lyrata*,
329 *Boechera stricta*, *Capsella rubella*, *Eutrema salsugineum*, *Brassica rapa*, and *Schrenkia parvula*),
330 and three outgroup species (*Tarenaya hassleriana*, *Carica papaya*, and *Theobroma cacao*). The
331 *Tarenaya* and *Schrenkia* genomes were obtained from Plaza v4 Dicots (Van Bel et al. 2017), and
332 the remaining genomes from Phytozome v12.1 (Goodstein et al. 2011). We constructed gene trees for
333 each family using MAFFT v7.407 and FastTree v2.1.8 (Price et al. 2010) rooted with the three
334 outgroups, and retained 3,525 subtrees with single gene copies present in at least 17 of the E18
335 species and in at least 6 of the 7 other Brassicaceae species. The *Erysimum* protein sequences in the
336 3,525 subtrees were used to identify high quality matches against the full protein sequences of the
337 remaining 30 *Erysimum* species in the second set ('E30') by BLAST. High quality matches were
338 defined as matching at least 15 of the E18 species sequences in the subtree. We retained subtrees
339 having high quality matches in at least 24 of the E30 species, resulting in 3,098 subtrees. Finally, we
340 identified mutual best matches between both sets of *Erysimum* species by matching the E30 protein
341 sequences in the 3,098 subtrees against the full set of E18 protein sequences. For each subtree, we
342 required the matches in the second set to be mutual best matches to all of the E18 proteins in the
343 subtree, and that there be at least 24 of the E30 species in the second set. This resulted in a final set of
344 2,306 subtrees, from which we constructed protein sequence alignments for all *Erysimum* species and
345 *Arabidopsis* using GUIDANCE2 v2.0.2 (Sela et al. 2015) and MAFFT. We then eliminated all
346 alignment columns identified by GUIDANCE2 as low quality (column score < 0.93) and transformed
347 protein sequences to codons. From the codon alignments, we constructed trees using RAxML v8.2.8
348 with the GTRGAMMA model and treating the 1st, 2nd, and 3rd codon positions as three separate
349 partitions. We concatenated all 2,306 gene family alignments, inserting gaps where a species was
350 missing from an alignment. Additionally, we performed coalescent species tree inference with the
351 2,306 gene families as input, after deleting sequences with fewer than 100 non-ambiguous characters
352 in a gene family alignment.

353

354 ***Metabolite profiling of Erysimum leaves***

355 We harvested leaf material for targeted metabolomic analysis of defense compounds from the same
356 plants as used for transcriptome sequencing, one week after leaves for RNA extraction had been
357 harvested. In each of the three experiments, we collected several leaves from 1-5 plants per species,
358 and immediately snap froze the harvested leaves in liquid nitrogen. While most plant samples were

359 screened for constitutive levels of chemical defenses only, we quantified inducibility of chemical
360 defenses in a subset of 30 species with sufficient replication (eight or more plants) in the third
361 experiment (2017-2). For these species, half of all plants were randomly assigned to the induction
362 treatment and given a foliar spray of JA one week prior to harvest. Plants were sprayed with 2-3 mL
363 of a 0.5 mM JA solution (Cayman Chemical, MI, USA) in 2% ethanol until all leaves were evenly
364 covered in droplets on both sides. Control plants were sprayed with an equivalent amount of 2%
365 ethanol solution. Harvested frozen plant material was lyophilized to dryness and ground to a fine
366 powder. We weighed out 10 mg leaf powder per sample into a separate tube and added 1 mL of 70%
367 MeOH extraction solvent. Samples were extracted by adding three 3 mm ceramic beads to each tube
368 and shaking tubes on a Retsch MM400 ball mill three times for 3 min at 30 Hz. We centrifuged
369 samples at 18,000 x g and transferred 0.9 mL of the supernatant to a new tube. Samples were
370 centrifuged again, and 0.8 mL of the final supernatant was transferred to an HPLC vial for analysis by
371 high-resolution mass spectrometry.

372 We analyzed extracts of individual plants (experiments 2016, 2017-1) or of multiple pooled
373 individuals per species and induction treatment (experiment 2017-2) on an Acquity UHPLC system
374 coupled to a Xevo G2-XS QTOF mass spectrometer with electrospray ionization (Waters, Milford
375 MA, USA). Due to large differences in the physiochemical properties between glucosinolates and
376 cardenolides, each plant extract was analyzed in two different modes to optimize detection of each
377 compound class. For glucosinolates, extracts were separated on a Waters Acquity charged surface
378 hybrid (CSH) C18 100 × 2.1 mm column with 1.7 µm pore size, fitted with a CSH guard column. The
379 column was maintained at 40 °C and injections of 1 µl were eluted at a constant flow rate of 0.4
380 mL/min with a gradient of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B) as
381 follows: 0-6 min from 2% to 45 % B, 6-6.5 min from 45% to 100% B, followed by a 2 min wash
382 phase at 100% B, and 2 min reconditioning at 2% B. For cardenolides, extracts were separated on a
383 Waters Cortecs C18 150 × 2.1 mm column with 2.7 µm pore size, fitted with a Cortecs C18 guard
384 column. The column was maintained at 40 °C and injections were eluted at a constant flow rate of 0.4
385 mL/min with a gradient of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B) as
386 follows: 0-10 min from 5% to 40 % B, 10-15 min from 40% to 100% B, followed by a 2.5 min wash
387 phase at 100% B, and 2.5 min reconditioning at 5% B.

388 Compounds were ionized in negative mode for glucosinolate analysis and in positive mode
389 for cardenolide analysis. In both modes, ion data were acquired over an m/z range of 50 to 1200 Da in
390 MS^E mode using alternating scans of 0.15 s at low collision energy of 6 eV and 0.15 s at high
391 collision energy ramped from 10 to 40 eV. For both positive and negative modes, the electrospray
392 capillary voltage was set to 2 kV and the cone voltage was set to 20 V. The source temperature was
393 maintained at 140 °C and the desolvation gas temperature at 400 °C. The desolvation gas flow was set
394 to 1000 L/h, and argon was used as a collision gas. The mobile phase was diverted to waste during the
395 wash and reconditioning phase at the end of each gradient. Accurate mass measurements were

396 obtained by infusing a solution of leucine-enkephalin at 200 ng/mL at a flow rate of 10 μ L/min
397 through the LockSpray probe.

398

399 ***Identification and quantification of defense compounds***

400 Glucosinolates consist of a β -D-glucopyranose residue linked via a sulfur atom to a (Z)-N-
401 hydroximinosulfate ester and a variable R group (Halkier and Gershenzon 2006). We identified
402 candidate glucosinolate compounds from negative scan data by the exact molecular mass of
403 glucosinolates known to occur in *Erysimum* and related species (Huang et al. 1993, Fahey et al. 2001).
404 In addition, we screened all negative scan data for characteristic glucosinolate mass fragments to
405 identify additional candidate compounds (Cataldi et al. 2010). For mass features with multiple
406 possible identifications, we inferred the most likely compound identity from relative HPLC retention
407 times and the presence of biosynthetically related compounds in the same sample. We confirmed our
408 identifications using commercial standards for glucoiberin (3-methylsulfinylpropyl glucosinolate,
409 Phytolab GmbH, Germany), glucocheirolin (3-methylsulfonylpropyl glucosinolate, Phytolab GmbH),
410 and sinigrin (2-propenyl glucosinolate, Sigma-Aldrich), as well as by comparison to extracts of
411 *Arabidopsis* accessions with known glucosinolate profiles. Compound abundances of all
412 glucosinolates were quantified by integrating ion intensities of the [M-H]⁻ adducts using QuanLynx in
413 the MassLynx software (v4.1, Waters).

414 All cardenolides share a highly conserved structure consisting of a steroid core ($5\beta,14\beta$ -
415 androstan-3 β 14-diol) linked to a five-membered lactone ring, which as a unit (the genin) mediates
416 the specific binding of cardenolides to Na⁺/K⁺- ATPase (Dzimir et al. 1987). While cardenolide
417 genins are sufficient to inhibit Na⁺/K⁺- ATPase function, genins are commonly glycosylated or
418 modified by hydroxylation on the steroid moiety to change the physiochemical properties and binding
419 affinity of compounds (Dzimir et al. 1987, Petschenka et al. 2018). We obtained commercial
420 standards for the abundant *Erysimum* cardenolides erysimoside and helveticoside (Sigma-Aldrich),
421 allowing us to identify these compounds through comparison of retention times and mass
422 fragmentation patterns. Additional cardenolide compounds were tentatively identified from
423 characteristic LC-MS fragmentation patterns. Sachdev-Gupta et al. (1990, 1993) reported
424 fragmentation patterns for glycosides of strophanthidin, digitoxigenin, and canogenol from *E.*
425 *cheiranthoides*. Their results highlight the propensity of cardenolides to fragment at glycosidic bonds,
426 with genin masses in particular being a prominent feature of cardenolide mass spectra. Additionally,
427 cardenolide genins exhibit further fragmentation related to the loss of OH-groups from the steroidal
428 core structure. We confirmed these rules of fragmentation for our mass spectrometry system using
429 commercial standards of strophanthidin and digitoxigenin (Sigma-Aldrich). Importantly, while
430 fragments were most abundant under high-energy conditions (MS^E), they were still apparent under
431 standard MS conditions, likely due to in-source fragmentation.

432 Characteristic fragmentation allowed us to identify candidate cardenolide compounds in a
433 genin-guided approach, where the presence of characteristic genin fragments in a chromatographic
434 peak indicated the likely presence of a cardenolide molecule. We then identified the parental mass of
435 these chromatographic peaks from the presence of paired mass features separated by 21.98 m/z,
436 corresponding to the $[M+H]^+$ and $[M+Na]^+$ adducts of the intact molecule. For di-glycosidic
437 cardenolides, additional fragments corresponding to the loss of the outer sugar moiety allowed us to
438 determine the mass and order of sugar moieties in the linear glycoside chain of the molecule. We
439 screened our data for the presence of glycosides of strophanthidin, digitoxigenin, and cannogenol, and
440 additional genins known to occur in *Erysimum* species (Makarevich et al. 1994). Multiple cardenolide
441 genins can share the same molecular structure and may not be distinguished by mass spectrometry
442 alone. Thus, all genin identifications are tentative and based on previous literature reports. We
443 screened LC-MS data from all three experiments to generate a list of cardenolide compounds.
444 Compounds had to be consistently detectable in at least one *Erysimum* species in at least two out of
445 three experiments to be included in the final list. Relative compound abundances were quantified by
446 integrating the ion intensities of the $[M+H]^+$ or the $[M+Na]^+$ adduct, whichever was more abundant for
447 a given compound across all samples. In the third experiment, we added hydrocortisone (Sigma-
448 Aldrich) to each sample as an internal standard, but between-sample variation (technical noise) was
449 negligible compared to between-species variation.

450 For glucosinolate and cardenolide data separately, raw ion counts for each compound were
451 averaged across experiments to yield robust chemotype data. Raw ion counts were standardized by the
452 dry sample weight, possible dilution of samples, and internal standard concentrations (where
453 available). For pooled samples, ion counts were standardized by the average dry weight calculated
454 from all samples that contributed to a pool. The full set of standardized compound ion counts was
455 then analyzed using linear mixed effects models (package *nlme* v3.1-137 in R v3.5.3). Because
456 standardized ion counts still had a heavily skewed distribution, we applied a log(+0.1) transformation
457 to all values. Log-transformed ion counts were modelled treating experiment as a fixed effect, and a
458 species-by-compound identifier as the main random effect. Nested within the main random effect, we
459 fitted a species-by-compound-by-experiment identifier as a second random effect to account for the
460 difference of pooled or individual samples among experiments. The fixed effect of this model thus
461 captures the overall differences in compound ion counts between experiments, while the main random
462 effect captures the average deviation from an overall compound mean for each compound in each
463 species. We extracted the overall compound mean and the main random effects from these models,
464 providing us with average ion counts for each compound in each species on the log-scale. Negative
465 values on the log-scale were set to zero as they would correspond to values below the limit of reliable
466 detection of the LC-MS on the normal scale.

467

468 **Inhibition of mammal Na^+/K^+ -ATPase by leaf extracts**

469 Although all cardenolides target the same enzyme in animal cells, structural variation among different
470 cardenolides can significantly influence binding affinity and thus affect toxicity (Dzimiri et al. 1987,
471 Petschenka et al. 2018). Cardenolide quantification from LC-MS mass signal intensity does not
472 capture such differences in biological activity, and furthermore may be challenging due to compound-
473 specific response factors and narrow ranges of signal linearity. To evaluate whether total ion counts
474 are an appropriate and biologically relevant measure for between-species comparisons of defense
475 levels, we therefore quantified cardenolide concentrations by a separate method (Züst et al. 2019). For
476 the subset of plants in the 2016 experiment, we measured the biological activity of leaf extracts on the
477 Na^+/K^+ -ATPase from the cerebral cortex of pigs (*Sus scrofa*, Sigma-Aldrich, MO, USA) using an *in*
478 *vitro* assay introduced by Klauck and Luckner (1995) and adapted by Petschenka et al. (2013). This
479 colorimetric assay measures Na^+/K^+ -ATPase activity from phosphate released during ATP
480 consumption, and can be used to quantify relative enzymatic inhibition by cardenolide-containing
481 plant extracts. Briefly, we tested the inhibitory effect of each plant extract at four concentrations to
482 estimate the sigmoid enzyme inhibition function from which we could determine the cardenolide
483 content of the extract relative to a standard curve for ouabain (Sigma Aldrich, MO, USA). We dried a
484 100 μL aliquot of each extract used for metabolomic analyses at 45 °C on a vacuum concentrator
485 (SpeedVac, Labconco, MO, USA). Dried residues were dissolved in 200 μL 10% DMSO in water,
486 and further diluted 1:5, 1:50, and 1:500 using 10% DMSO. To quantify potential non-specific
487 enzymatic inhibition that could occur at high concentrations of plant extracts, we also included control
488 extracts from *Sinapis arvensis* leaves (a non-cardenolide producing species of the Brassicaceae) in
489 these assays.

490 Assays were carried out in 96-well microplate format. Reactions were started by adding 80 μL
491 of a reaction mix containing 0.0015 units of porcine Na^+/K^+ -ATPase to 20 μL of leaf extracts in 10%
492 DMSO, to achieve final well concentrations (in 100 μL) of 100 mM NaCl, 20 mM KCl, 4 mM MgCl₂,
493 50 mM imidazol, and 2.5 mM ATP at pH 7.4. To control for coloration of leaf extracts, we replicated
494 each reaction on the same 96-well plate using a buffered background mix with identical composition
495 as the reaction mix but lacking KCl, resulting in inactive Na^+/K^+ -ATPases. Plates were incubated at
496 37 °C for 20 minutes, after which enzymatic reactions were stopped by addition of 100 μL sodium
497 dodecyl sulfate (SDS, 10% plus 0.05% Antifoam A) to each well. Inorganic phosphate released from
498 enzymatically hydrolyzed ATP was quantified photometrically at 700 nm following the method
499 described by Taussky and Shorr (1953).

500 Absorbance values of reactions were corrected by their respective backgrounds, and sigmoid
501 dose-response curves were fitted to corrected absorbances using a non-linear mixed effects model
502 with a 4-parameter logistic function in the statistical software R (function *nlme* with *SSfpl* in package
503 *nlme* v3.1-137).

504

$$\text{Absorbance} = \frac{A + (B - A)}{1 + e^{((x_{mid} - x)/scal)}}$$

505 The absorbance values at four dilutions x are thus used to estimate the upper (A , fully active enzyme)
506 and lower (B , fully inhibited enzyme) asymptotes, the dilution value x_{mid} at which 50% inhibition is
507 achieved, and a shape parameter $scal$. In order to estimate four parameters from four absorbance
508 values per extract, the $scal$ parameter was fixed for all extracts and changed iteratively to optimize
509 overall model fit, judged by AIC. Individual plant extracts were treated as random effects to account
510 for lack of independence within extract dilution series. For each extract we estimated x_{mid} from the
511 average model fit and the extract-specific random deviate. Using a calibration curve made with
512 ouabain ranging from 10^{-3} to 10^{-8} M that was included on each 96-well plate, we then estimated the
513 concentration of the undiluted sample in ouabain equivalents, i.e., the amount of ouabain required to
514 achieve equivalent inhibition.

515

516 ***Quantification of myrosinase activity***

517 For the subset of plants in experiment 2017-2, we extracted the total amounts of soluble myrosinases
518 from leaf tissue and quantified their activity as an important component of the glucosinolate defense
519 system of these species. At time of harvest of metabolomic samples, we collected an additional set of
520 leaf disks from each plant, corresponding to approximately 50 mg fresh weight. After determination
521 of exact fresh weight, samples were flash frozen in liquid nitrogen and stored at -80 °C until enzyme
522 activity measurements. Following the protocol of Travers-Martin et al. (2008), frozen leaf material
523 was ground and extracted in Tris-EDTA buffer (200 mM Tris, 10 mM EDTA, pH 5.5) and internal
524 glucosinolates were removed by rinsing the extracts over a DEAE Sephadex A25 column (Sigma-
525 Aldrich). Myrosinase activities were determined by adding sinigrin to plant extracts and monitoring
526 the enzymatic release of glucose from its activation. Control reactions with sinigrin-free buffer were
527 used to correct for plant-derived glucose. All samples were measured in duplicate and mean values
528 related to a glucose calibration curve, measured also in duplicate. Reactions were carried out in 96-
529 well plates and concentrations of released glucose were measured by adding a mix of glucose oxidase,
530 peroxidase, 4-aminoantipyrine and phenol as color reagent to each well and measuring the kinetics for
531 45 min at room temperature in a microplate photometer (Multiskan EX, Thermo Electron, China) at
532 492 nm.

533

534 ***Similarity in defense profiles between Erysimum species***

535 To quantify chemical similarity among species, we performed separate cluster analyses on the
536 glucosinolate and cardenolide profile data averaged across the three experiments. For each species,
537 the log-transformed average ion counts of all compounds were converted to proportions (all
538 compounds produced by a species summing to 1). From this proportional data we then calculated
539 pairwise Bray-Curtis dissimilarities for all species pairs using function *vegdist* in the R package *vegan*
540 v2.5-4. We incorporated *vegdist* as a custom distance function for *pvclust* in the R package *pvclust*
541 v2.0 (Suzuki and Shimodaira 2014), which performs multiscale bootstrap resampling for cluster

542 analyses. We constructed dendograms of glucosinolate and cardenolide profile similarities by fitting
543 hierarchical clustering models (Ward's D) and estimated support for individual species clusters from
544 10,000 permutations. To compare chemical similarity to phylogenetic relatedness we performed
545 principal coordinate analyses (PCoA) on Bray-Curtis dissimilarity matrices of glucosinolate and
546 cardenolide data using function *pcoa* in R package *ape* v5.0 (Paradis and Schliep 2019), and extracted
547 the first two principal coordinates for each defense trait to test for phylogenetic signal.

548

549 *Relationship between plant traits and phylogenetic signal*

550 We evaluated a prevalence of phylogenetic signal in chemical defense traits, myrosinase activity, and
551 principal coordinates for both chemical similarity matrices using Blomberg's *K* (Blomberg et al.
552 2003). *K* is close to zero for traits lacking phylogenetic signal; it approaches 1 if trait similarity among
553 related species matches a Brownian motion model of evolution, and it can be >1 if similarity is even
554 higher than expected under a Brownian motion model. We estimated *K* for all traits using function
555 *phylosig* in the R package *phytools* v0.6-60 (Revell 2012). Additionally, we used the geographic
556 coordinates for all species with known collection locations to construct pairwise geographic distances,
557 calculated pairwise geographic dissimilarities, performed principal coordinate analyses on the
558 geographic dissimilarity matrix, and estimated *K* for the first two components.

559 To test for directional effects in the evolution of compound number and abundance for both
560 glucosinolates and cardenolides, we applied Pagel's method (Pagel 1999). Specifically, we compared
561 a Brownian motion model of trait evolution to a model in which additionally a directional trend is
562 assessed by regressing the path length (i.e., molecular branch length from root to tip) against trait
563 values. For this analysis we used the concatenated 2,306-gene tree for which branch lengths are an
564 estimate of substitutions per site. Models were fit using function *fitContinuous* in R package *geiger*
565 v2.0.6.2 (Harmon et al. 2008), where the default setting fits a Brownian motion model, whereas the
566 additional argument 'model=drift' specifies a directional trend model. Support for directional trends
567 in defense traits was evaluated using likelihood-ratio tests between the two models.

568

569 Results

570 *E. cheiranthoides* genome assembly

571 A total of 39.5 Gb of PacBio sequences with an average read length of 10,603 bp were assembled into
572 1,087 contigs with an N50 of 1.5 Mbp (Table 1). Hi-C scaffolding oriented 98.5 % of the assembly
573 into eight large scaffolds representing pseudomolecules (Table 1, Figure S1), while 216 small contigs
574 remained unanchored. The final assembly (v1.2) had a total length of 174.5 Mbp, representing 86% of
575 the estimated genome size of *E. cheiranthoides* and capturing 99% of the BUSCO gene set (Table 1,
576 Figure S2). Sequences were deposited under GenBank project ID PRJNA563696 and additionally are
577 provided at www.erysimum.org. A total of 29,947 gene models were predicted and captured 98% of
578 the BUSCO gene set (Figure S3). In the presumed centromere regions of each chromosome, genic

579 sequences were less abundant, whereas repeat sequences were more common (Fig 2A). Repetitive
580 sequences constituted approximately 29% of the genome (Table S2). Long terminal repeat
581 retrotransposons (LTR-RT) made up the largest proportion of the repeats identified (Figure S4).
582 Among these, repeats in the *Gypsy* superfamily constituted the largest fraction of the genome (Table
583 S2). The majority of the LTR elements appeared to be relatively young, with most having estimated
584 insertion times of less than 1 MYA (Figure S5). Synteny analysis showed evidence of several
585 chromosomal fusions and fissions between the eight chromosomes of *E. cheiranthoides* and the five
586 chromosomes of Arabidopsis (Figure 2B).

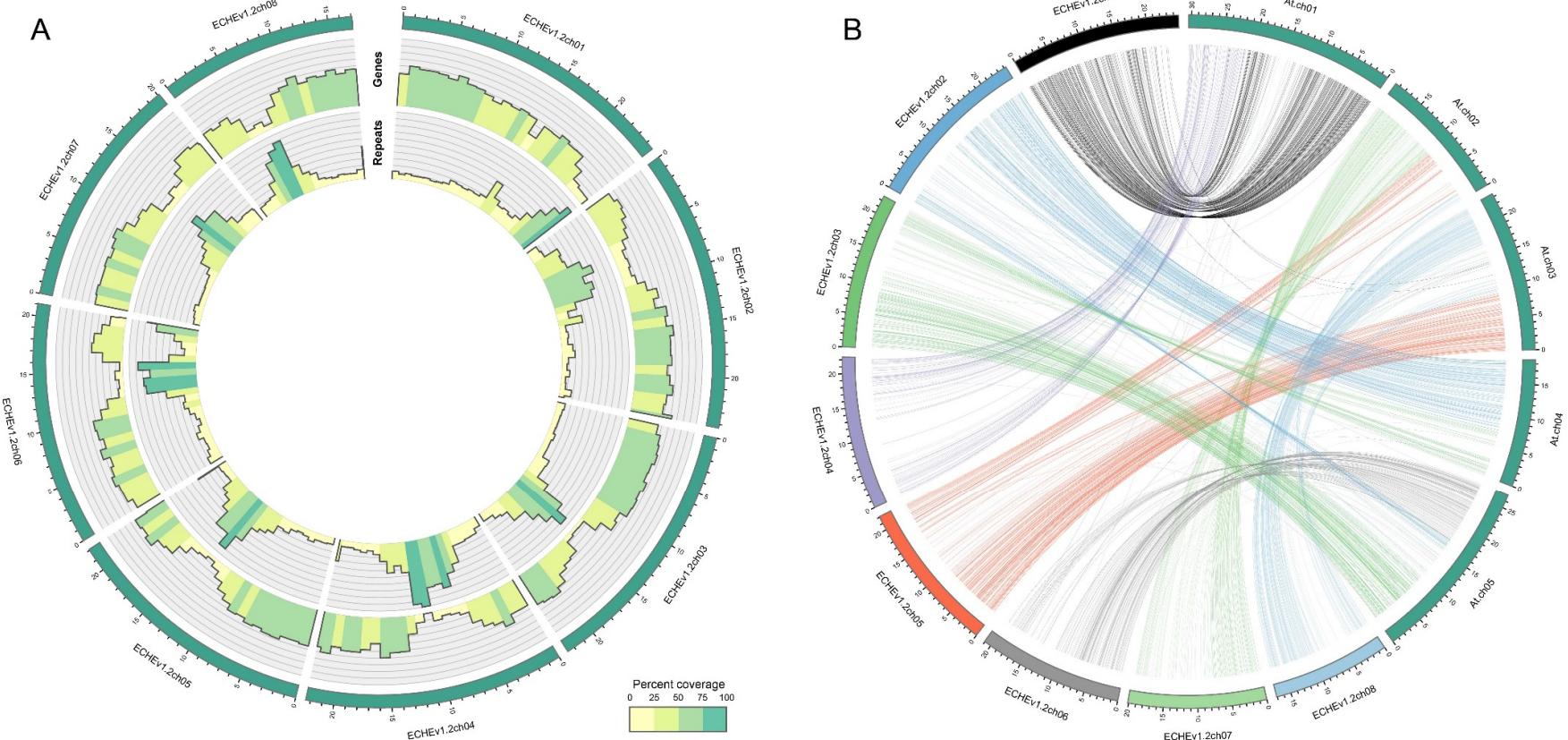
587

588 **Table 1.** Assembly metrics for the *E. cheiranthoides* genome: v0.9 = Falcon +Arrow assembly results, v1.2 =
589 genome assembly after Hi-C scaffolding and Pilon correction.

	v0.9	v1.2 pseudomolecules and contigs	v1.2 pseudomolecules only
total length (Mbp)	177.4	177.2	174.5
expected size (Mbp)	205	205	205
number of contigs	1087	224	8
N50 (Mbp)	1.5	22.4	22.4
complete BUSCOs (out of 1,375)	1359	1346	1356
complete and single copy BUSCOs (out of 1,375)	1271	1300	1306
complete and duplicated BUSCOs (out of 1,375)	88	46	50
fragmented BUSCOs (out of 1,375)	5	8	6
missing BUSCOs (out of 1,375)	11	21	13

590

591



592

593 **Figure 2.** (A) Circos plot of the *E. cheiranthoides* genome with gene densities (outer circle) and repeat densities (inner circle) shown as histogram tracks. Densities are
 594 calculated as percentages for 1 Mb windows. (B) Synteny plot of *E. cheiranthoides* and *A. thaliana*. Lines between chromosomes connect aligned sequences between the two
 595 genomes.

596 ***Glucosinolate and myrosinase genes in the E. cheiranthoides genome***

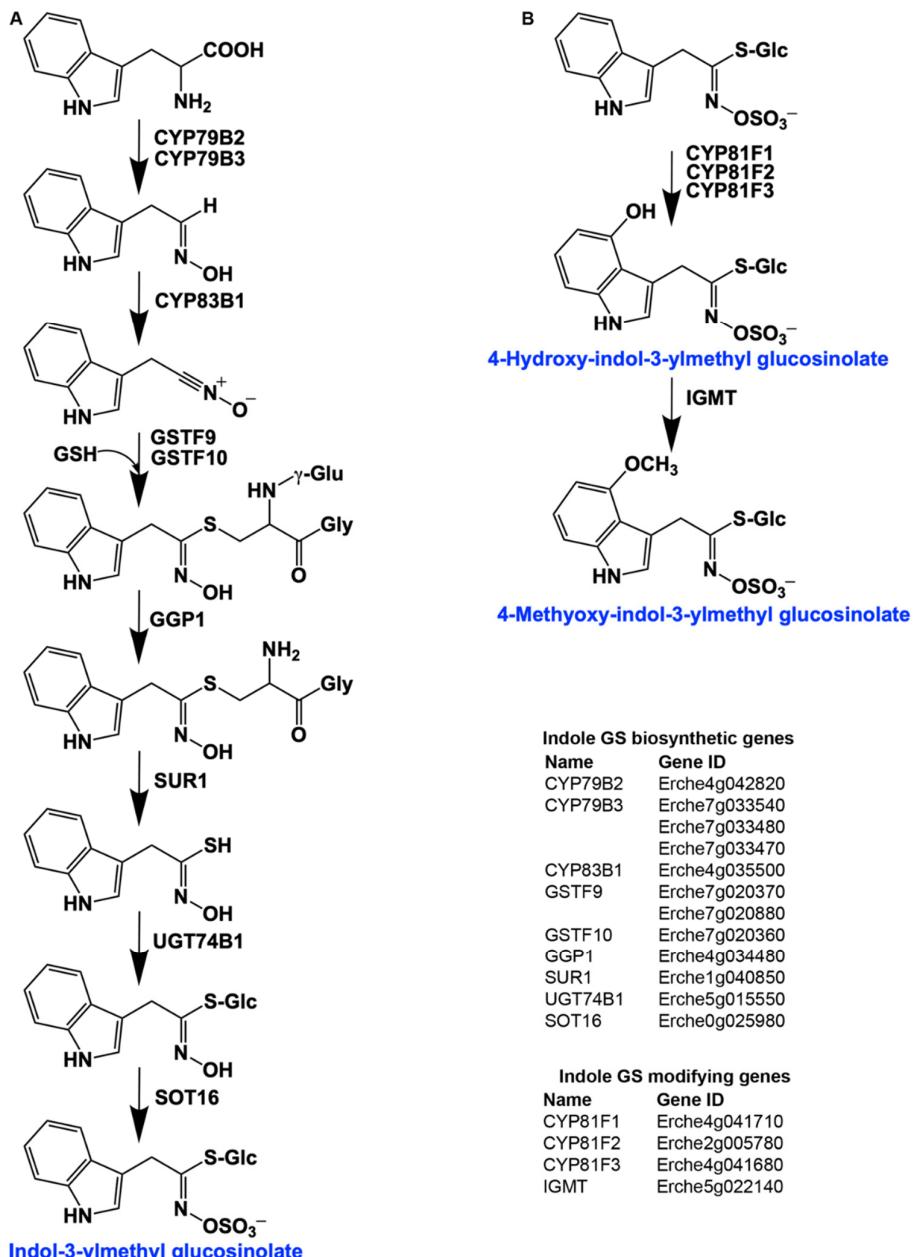
597 Three aliphatic glucosinolates – glucoiberverin (3-methylthiopropyl glucosinolate), glucoiberin, and
598 glucocheirolin – have been reported as the main glucosinolates in *E. cheiranthoides* (Cole 1976,
599 Huang et al. 1993). We confirmed their dominance in *E. cheiranthoides* var. *Elbtalaue*, but also
600 identified additional aliphatic and indole glucosinolates at lower concentrations. By making use of the
601 glucosinolate biosynthetic pathway for Arabidopsis (Halkier and Gershenson 2006) and comparing
602 nucleotide coding sequences of Arabidopsis and *E. cheiranthoides*, we identified homologs of genes
603 encoding both indole (Figure 3) and aliphatic (Figure 4) glucosinolate biosynthetic enzymes.
604 Homologs of all genes of the complete biosynthetic pathway for glucobrassicin (indol-3-ylmethyl
605 glucosinolate) and its 4-hydroxy and 4-methoxy derivatives were present in *E. cheiranthoides* (Figure
606 3). Consistent with the absence of neoglucobrassicin (1-methoxy-indol-3-ylmethyl glucosinolate) in
607 *E. cheiranthoides* var. *Elbtalaue*, we did not find homologs of the Arabidopsis genes encoding the
608 biosynthesis of this compound.

609 Genes encoding the complete biosynthetic pathway of the *E. cheiranthoides* aliphatic
610 glucosinolates glucoiberverin, glucoiberin, glucoerucin (4-methylthiobutyl glucosinolate), and
611 glucoraphanin (4-methylsulfinylbutyl glucosinolate) were present in the genome (Figure 4). Because
612 the *E. cheiranthoides* methylsulfonyl glucosinolates glucocheirolin, 4-methylsulfonylbutyl
613 glucosinolate, and 3-hydroxy-4-methylsulfonylbutyl glucosinolate are not present in Arabidopsis,
614 genes encoding their biosynthesis could not be identified. Consistent with the absence of an
615 Arabidopsis AOP2 homolog, we did not find alkenyl glucosinolates in *E. cheiranthoides*.

616 In response to insect feeding or pathogen infection, glucosinolates are activated by
617 myrosinase enzymes (Halkier and Gershenson 2006). Between-gene phylogenetic comparisons
618 revealed that homologs of known Arabidopsis myrosinases, the main foliar myrosinases *TGG1* and
619 *TGG2* (Barth and Jander 2006), root-expressed *TGG4* and *TGG5* (Andersson et al. 2009), and likely
620 pseudogenes *TGG3* and *TGG6* (Rask et al. 2000, Zhang et al. 2002), were also present in the *E.*
621 *cheiranthoides* genome (Figure S6). Additionally, we found homologs of the more distantly related
622 Arabidopsis myrosinases *PEN2* (Bednarek et al. 2009, Clay et al. 2009) and *PYK10* (Sherameti et al.
623 2008, Nakano et al. 2017). Thus, the pathway of glucosinolate activation appears to be largely
624 conserved between Arabidopsis and *E. cheiranthoides*.

625

626



Indole GS biosynthetic genes

Name	Gene ID
CYP79B2	Erche4g042820
CYP79B3	Erche7g033540
	Erche7g033480
	Erche7g033470
CYP83B1	Erche4g035500
GSTF9	Erche7g020370
	Erche7g020880
GSTF10	Erche7g020360
GGP1	Erche4g034480
SUR1	Erche1g040850
UGT74B1	Erche5g015550
SOT16	Erche0g025980

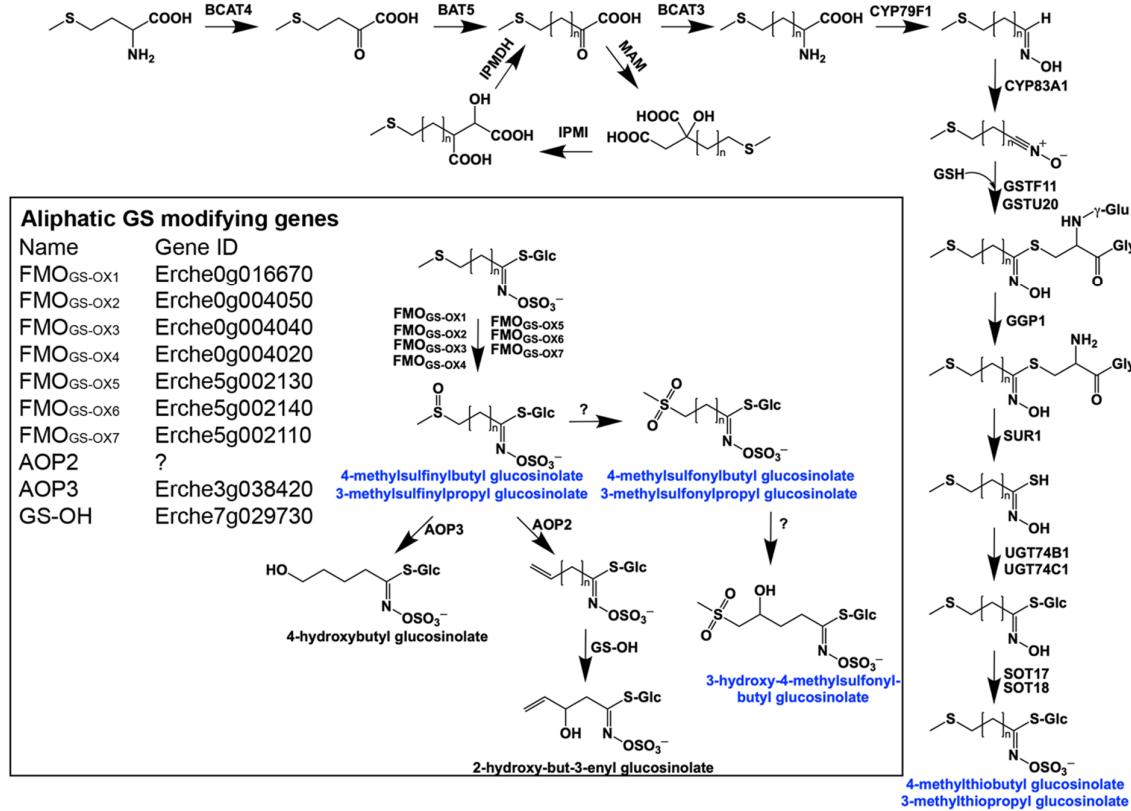
Indole GS modifying genes

Name	Gene ID
CYP81F1	Erche4g041710
CYP81F2	Erche2g005780
CYP81F3	Erche4g041680
IGMT	Erche5g022140

627

Indol-3-ylmethyl glucosinolate

628 **Figure 3.** Identification of indole glucosinolate biosynthetic genes and glucosinolate-modifying genes in
629 *Erysimum cheiranthoides*. (A) Starting with tryptophan, indole glucosinolates are synthesized using some
630 enzymes that also function in aliphatic glucosinolates biosynthesis (GGP1; SUR1; UGT74B1) while also using
631 indole glucosinolate-specific enzymes. (B) Indole glucosinolates can be modified by hydroxylation and
632 subsequent methylation. Glucosinolates with names highlighted in blue were identified in *Erysimum*
633 *cheiranthoides* var. *Elbtalaue*. Abbreviations: cytochrome P450 monooxygenase (CYP); glutathione S-
634 transferase F (GSTF); glutathione (GSH); γ-glutamyl peptidase 1 (GGP1); SUPERROOT 1 C-S lyase (SUR1);
635 UDP-dependent glycosyltransferase (UGT); sulfotransferase (SOT); glucosinolate (GS); indole glucosinolates
636 methyltransferase (IGMT).



Name	Gene ID	Name	Gene ID	Name	Gene ID
BCAT4	Erche1g022530	IPMDH1	Erche3g014980	GGP1	Erche4g034480
BAT5	Erche3g035820	IPMDH3	Erche0g032190	SUR1	Erche1g040850
BCAT3	Erche6g016740	CYP79F1	Erche5g017900	UGT74B1	Erche5g025290
	Erche2g013450	CYP83A1	Erche4g015050	UGT74C1	Erche7g019350
MAM	Erche3g023360	GSTF11	Erche1g003220	SOT17	Erche5g015550
IPMI-LSU1	Erche4g014340	GSTU20	Erche0g030320	SOT18	Erche0g025970
	Erche4g014330		Erche0g030330		
IPMI-SSU2	Erche7g006890		Erche0g030420		
	Erche7g006220				
IPMI-SSU3	Erche6g030770				

637

638 **Figure 4.** Identification of aliphatic glucosinolate biosynthetic genes in *Erysimum cheiranthoides* starting from
639 methionine and modifications of aliphatic glucosinolates (black box). Glucosinolates with names highlighted in
640 blue were identified in *Erysimum cheiranthoides* var. *Elbtalaue*. Abbreviations: branched-chain
641 aminotransferase (BCAT); bile acid transporter (BAT); methylthioalkylmalate synthase (MAM);
642 isopropylmalate isomerase (IPMI); large subunit (LSU); small subunit (SSU); isopropylmalate
643 dehydrogenase(IPMDH); cytochrome P450 monooxygenase (CYP); glutathione S-transferase F (GSTF);
644 glutathione S-transferase Tau (GSTU); glutathione (GSH); γ -glutamyl peptidase 1 (GGP1); SUPERROOT 1 C-
645 S lyase (SUR1); UDP-dependent glycosyltransferase (UGT); sulfotransferase (SOT); flavin monooxygenase
646 (FMO); glucosinolate oxoglutarate-dependent dioxygenase (AOP); 3-but enyl glucosinolate 2-hydroxylase (GS-
647 OH).

648

649 **Phylogenetic relationship of 48 *Erysimum* species**

650 Assemblies of transcriptomes from 48 *Erysimum* species (including *E. cheiranthoides*) had N50
651 values ranging from 574 - 2,160 bp (Table S3). Transcriptome assemblies contained completed genes
652 from 54% - 94% of the BUSCO set and coding sequence lengths were generally shorter on average
653 than the *E. cheiranthoides* coding sequence lengths (Table S3). Transcriptome sequences were
654 deposited under GenBank project ID PRJNA563696 and at www.erisium.org. The large number of
655 orthologous gene sequences identified among the *E. cheiranthoides* genome and the 48 transcriptomes
656 allowed us to infer phylogenetic relatedness with high confidence. We assume that the coalescent
657 phylogeny generated by the first approach of tree construction using 9,868 syntenic gene sequences
658 represents our best estimate of species relationships (Figure 5). However, both phylogenies generated
659 by the second approach using 2,306 orthologous genes had highly similar tree topologies (Figures S7,
660 S8), suggesting overall high reliability of our results regardless of the phylogenetic inference method.

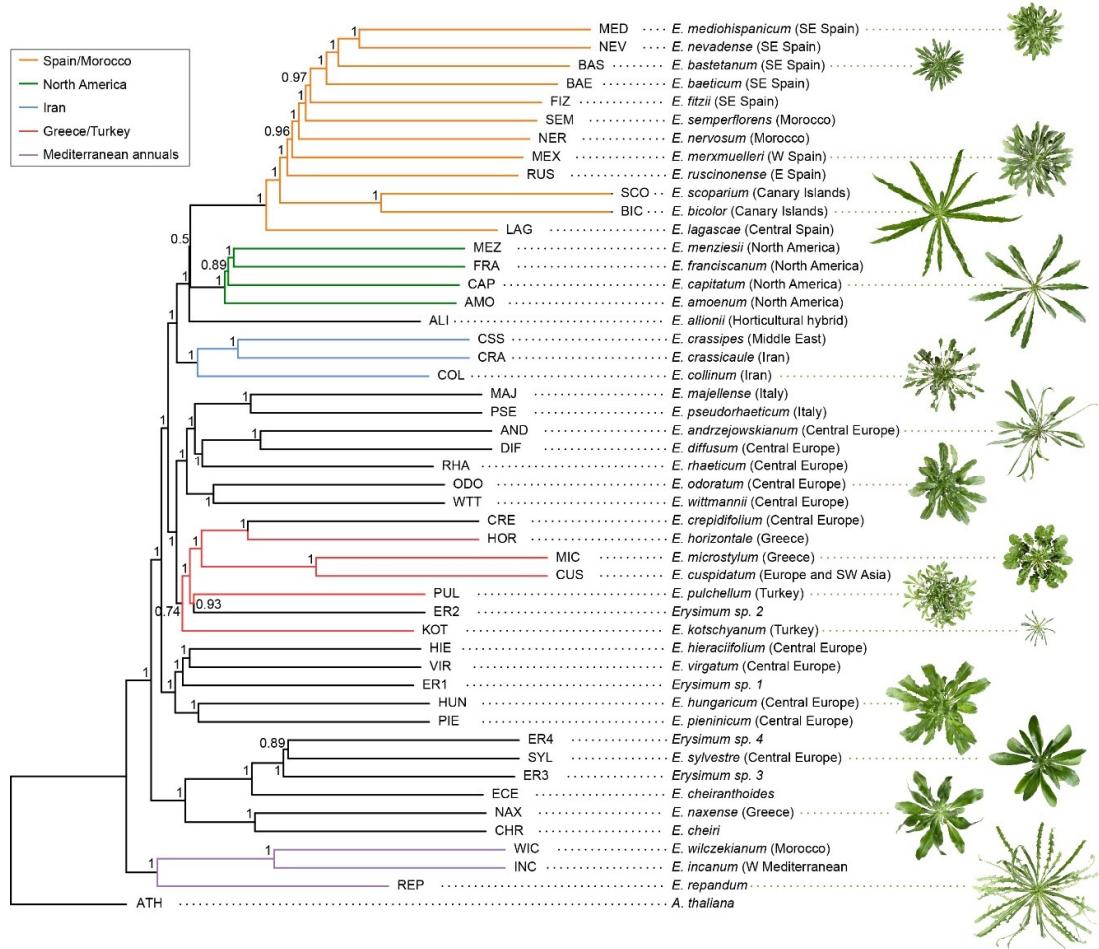
661 Virtually all phylogenetic nodes of the coalescent trees had high statistical support (all but
662 one node with local posterior probability > 0.7; Figure 5). The three Mediterranean annual species *E.*
663 *incanum* (INC), *E. repandum* (REP), and *E. wilczekianum* (WIC) formed a monophyletic sister clade
664 to all other sequenced species. The only other annual in the set of sampled species, *E. cheiranthoides*
665 (ECE), was part of a second early-diverging clade, together with several perennial species from
666 Greece and central Europe, including the widespread ornamental *E. cheiri* (CHR). Several other
667 geographic clades were apparent in the phylogeny, with species from the Iberian peninsula/Morocco,
668 North America, Iran, and Greece forming additional distinct geographic clades. The clear geographic
669 structure of the phylogeny was confirmed by a very strong phylogenetic signal for the first two
670 components of the geographic principal coordinate analysis (Table 2).

671

672 **Table 2.** Measure of phylogenetic signal for total defensive traits and principal coordinates of the cardenolide
673 and glucosinolate similarity matrices (PCO) using Blomberg's *K*. Significant values are highlighted in bold.

Plant trait	K statistics	P (10'000 simulations)
Glucosinolate PCO1 (18.8%)	0.89	0.176
Glucosinolate PCO2 (13.6%)	0.85	0.375
Total glucosinolate concentrations	0.88	0.216
Number of glucosinolate compounds	0.94	0.060
Myrosinase activity	1.02	0.019
Cardenolide PCO1 (16.5%)	1.55	<0.001
Cardenolide PCO2 (12.2%)	1.18	<0.001
Total cardenolide concentrations	1.08	0.016
Number of cardenolide compounds	1.16	0.001
Geographical PCO1 (15.8%) ¹	2.33	<0.001
Geographical PCO2 (12.8%) ¹	1.37	<0.001

674 ¹ results are for 43 species with reliable geographic information



675

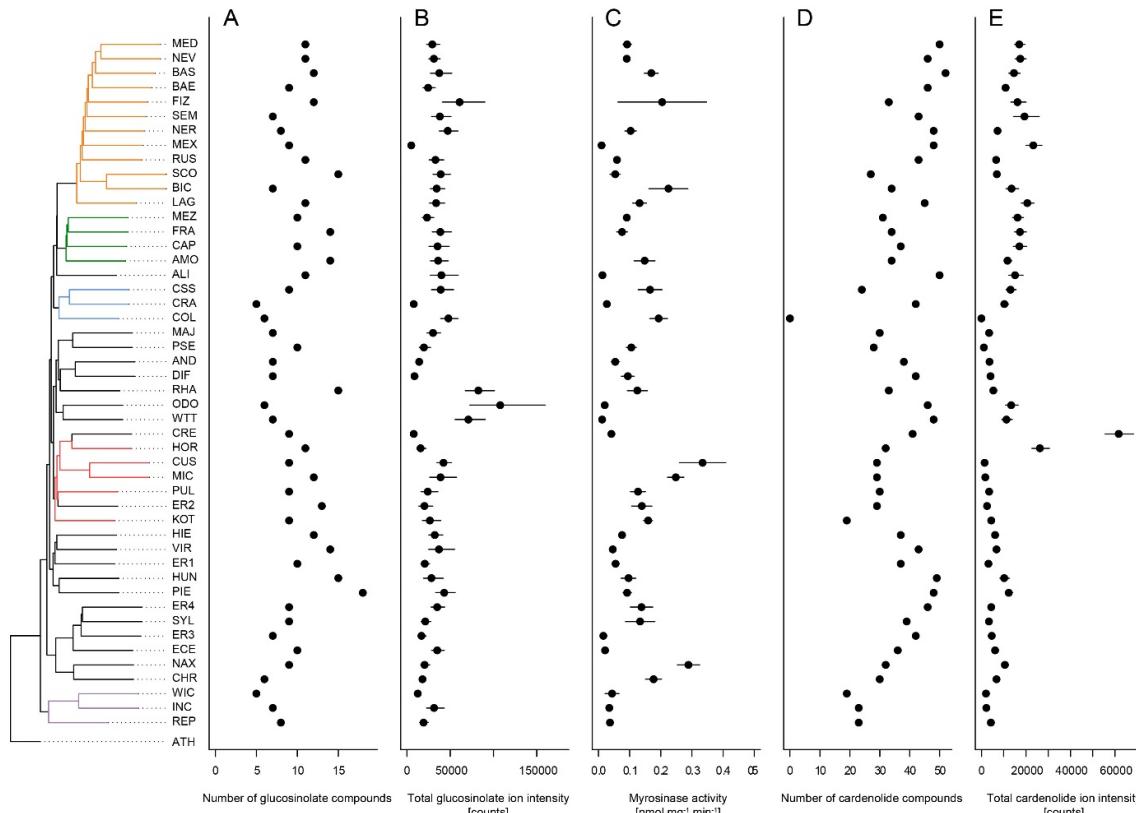
676 **Figure 5.** Genome-guided coalescent species tree of 48 *Erysimum* species. Phylogenetic relationships were
677 inferred from 9,868 orthologous genes using ASTRAL-III. Nodes are labelled with local posterior probability,
678 indicating level of support. Geographic range of species is provided in parentheses. The horticultural species *E.*
679 *cheiri* and the weedy species *E. cheiranthoides* and *E. repandum* are of European origin but are now widespread
680 across the Northern Hemisphere. Clades of species from shared geographic origins are highlighted in different
681 colors. On the right, pictures of rosettes of a representative subset of species is provided to highlight the
682 morphological diversity within this genus. Plants are of same age and relative size differences are conserved in
683 the pictures.

684

685 **Glucosinolate diversity and myrosinase activity**

686 Across the 48 *Erysimum* species, we identified 25 candidate glucosinolate compounds with distinct
687 molecular masses and HPLC retention times (Table S4). Of these, 24 compounds could be assigned to
688 known glucosinolate structures with high certainty. The remaining compound appeared to be an
689 unknown isomer of glucocheirolin. Individual *Erysimum* species produced between 5 and 18
690 glucosinolates (Figure 6A), and total glucosinolate concentrations were highly variable among species
691 (Figure 6B). The ploidy level of species explained a significant fraction of total variation in the

692 number of glucosinolates produced ($F_{4,38} = 4.63$, $p = 0.004$), with hexaploid species producing the
693 highest number of compounds (Figure S9). However, neither the numbers of distinct glucosinolates
694 nor the total concentrations exhibited a phylogenetic signal (Table 2). Similarly, there was no
695 evidence for a directional trend in either glucosinolate trait (Table 3), suggesting that glucosinolate
696 defenses varied among species independently of phylogenetic history.
697



698
699 **Figure 6.** Mean defense traits of 48 *Erysimum* species, grouped by phylogenetic relatedness. Not all traits could
700 be quantified for all species. (A) Total number of glucosinolate compounds detected in each species. (B) Total
701 glucosinolate concentration found in each species, quantified by total ion intensity in mass spectrometry
702 analyses. Values are means \pm 1SE. (C) Quantification of glucosinolate-activating myrosinase activity. Enzyme
703 kinetics were quantified against the standard glucosinolate sinigrin and are expressed per unit fresh plant tissue.
704 Values are means \pm 1SE. (D) Total number of cardenolide compounds detected in each species. (E) Total
705 cardenolide concentrations found in each species, quantified by total ion intensity in mass spectrometry
706 analyses. Values are means \pm 1SE.
707

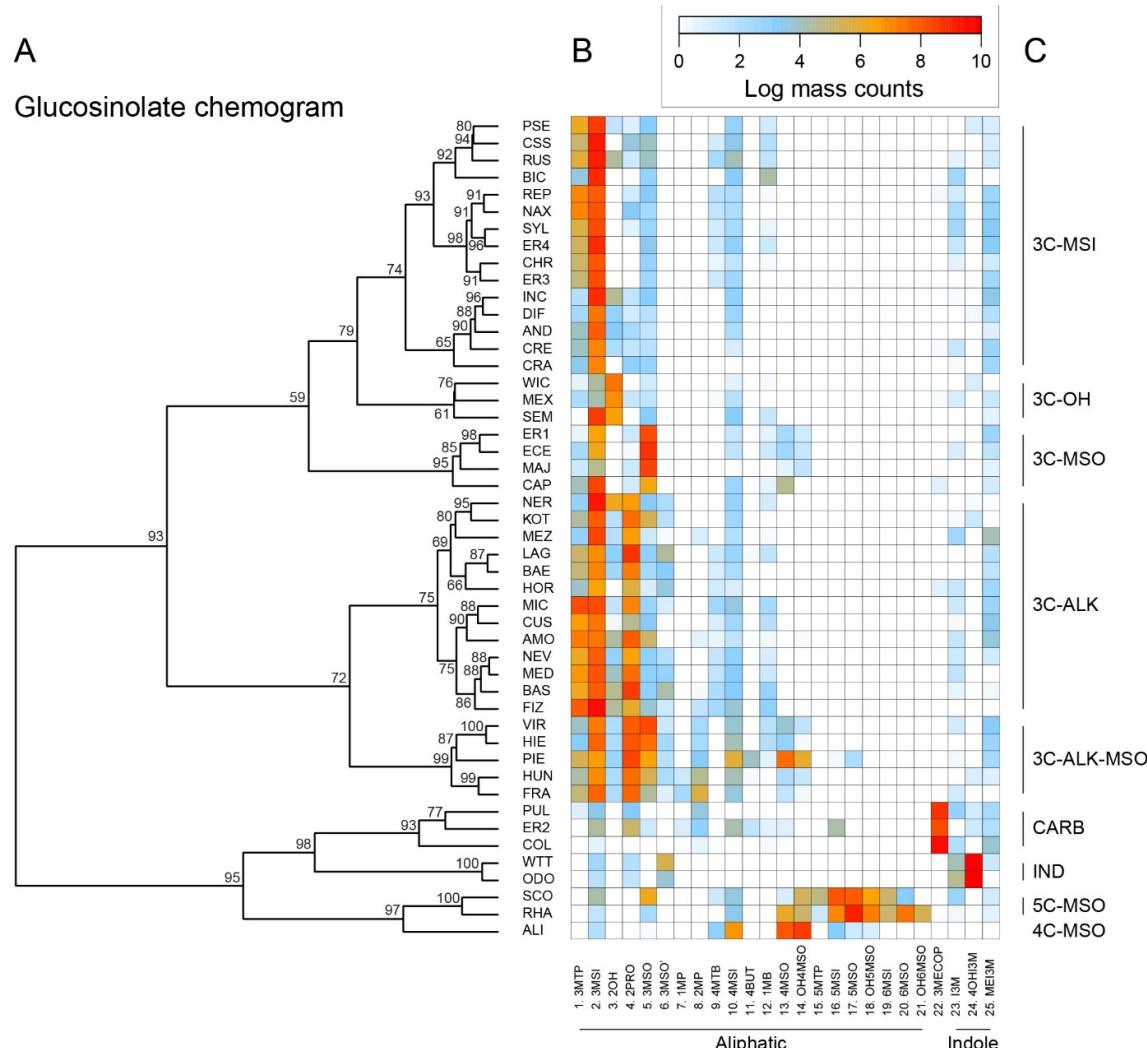
708 **Table 3.** Maximum-likelihood estimation of directional trends (β , root-to-tip regression) in cardenolide and
709 glucosinolate evolution. Directional trends are assessed for gradual models of evolution using the concatenated
710 2306-gene tree in which branch lengths are proportional to estimated substitutions per site. Each directional
711 model is assessed against a random walk model without a trend.

Plant traits	β	Likelihood ratio	p-value
Total glucosinolate concentrations	-37,160.0	0.001	0.974
Number of glucosinolate compounds	40.7	0.049	0.825
Total cardenolide concentrations	707,093.6	1.759	0.185
Number of cardenolide compounds	-218.4	0.173	0.677

712
713 Clustering species according to similarity in glucosinolate profiles mostly resulted in chemotype
714 groups corresponding to known underlying biosynthetic genes, although support for individual
715 species clusters was variable (Figure 7). The majority of species produced glucoiberin as the primary
716 glucosinolate. Of these, approximately half also produced sinigrin as a second dominant glucosinolate
717 compound. Further chemotypic subdivision, related to the production of glucocheirolin and 2-
718 hydroxypropyl glucosinolate, appeared to be present but only had relatively weak statistical support.
719 However, eight species clearly differed from these general patterns. The species *E. allionii* (ALI), *E.*
720 *rhaeticum* (RHA), and *E. scoroparium* (SCO) mostly lacked glucosinolates with 3-carbon side-chains,
721 but instead accumulated glucosinolates with 4-, 5- and 6-carbon side-chains. The two closely-related
722 species *E. odoratum* (ODO) and *E. wittmannii* (WIT) predominantly accumulated indole
723 glucosinolates, while *E. collinum* (COL), *E. pulchellum* (PUL), and accession ER2 predominantly
724 produced glucoerypestrin (3-methoxycarbonylpropyl glucosinolate), a glucosinolate that is
725 exclusively found within *Erysimum* (Fahey et al. 2001). As with total glucosinolate concentrations,
726 similarity in glucosinolate profiles of the species was again unrelated to phylogenetic relatedness
727 (Table 2).

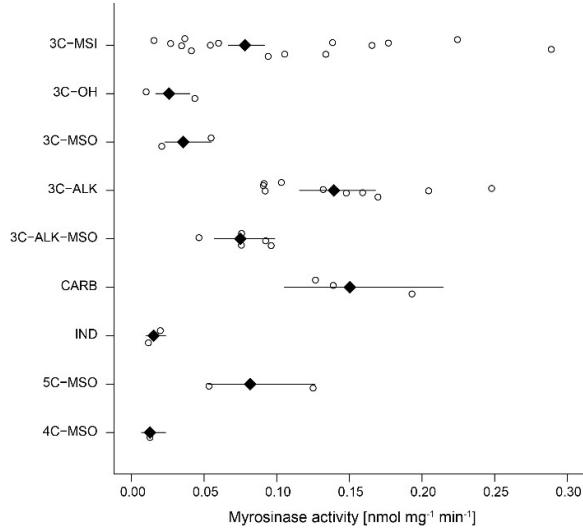
728 As glucosinolates require activation by myrosinase enzymes upon tissue damage by
729 herbivores, myrosinase activity in leaf tissue determines the rate at which toxins are released. We
730 quantified myrosinase activity of *Erysimum* leaf extracts and found it to be highly variable among
731 species (Figure 6C). After grouping species into nine chemotypes defined by chemical similarity and
732 the production of characteristic glucosinolate compounds (Figure 7C), we found that myrosinase
733 activity significantly differed among these chemotypes (Figure 8, $F_{8,33} = 7.06$, $p < 0.001$). Chemotypes
734 that predominantly accumulated methylsulfonyl glucosinolates, hydroxy glucosinolates, or indole
735 glucosinolates had low to negligible activity against the assayed glucosinolate sinigrin. It is important
736 to note that sinigrin is an alkenyl glucosinolate and activity with other, structurally dissimilar
737 glucosinolates may differ. After chemotype differences were accounted for, myrosinase activity was
738 related positively to total glucosinolate concentrations ($F_{1,33} = 5.92$, $p = 0.021$). Surprisingly,

739 uncorrected myrosinase activity was the only glucosinolate-related trait that showed a significant
740 phylogenetic signal (Table 2).
741



742

743 **Figure 7.** (A) Glucosinolate chemogram clustering species according to overlap in glucosinolate profiles.
744 Values at nodes are confidence estimates (approximately unbiased p-value, function *pvclust* in R) based on
745 10,000 iterations of multiscale bootstrap resampling. (B) Heatmap of glucosinolate profiles expressed by the 48
746 *Erysimum* species. Color intensity corresponds to log-transformed integrated ion counts recorded at the exact
747 parental mass ($[M-H]$) for each compound, averaged across samples from multiple independent experiments.
748 Compounds are grouped by major biosynthetic classes and labelled using systematic short names. See Table S4
749 for full glucosinolate names and additional compound information. (C) Classification of species chemotype
750 based on predominant glucosinolate compounds. 3C/4C/5C = length of carbon side chain, MSI = methylsulfinyl
751 glucosinolate, MSO = methylsulfonyl glucosinolate, OH = side chain with hydroxy group, ALK = side chain
752 with alkenyl group, CARB = carboxylic glucosinolate, IND = indole glucosinolate.
753



754

755 **Figure 8.** Myrosinase activity of *Erysimum* leaf extracts grouped by glucosinolate chemotype. Open circles are
756 species means and black diamonds are chemotype means ± 1 SE. See also Figure 7 for chemotype information.
757 3C/4C/5C = length of carbon side chain, MSI = methylsulfinyl glucosinolate, MSO = methylsulfonyl
758 glucosinolate, OH = side chain with hydroxy group, ALK = side chain with alkenyl group, CARB = carboxylic
759 glucosinolate, IND = indole glucosinolate.

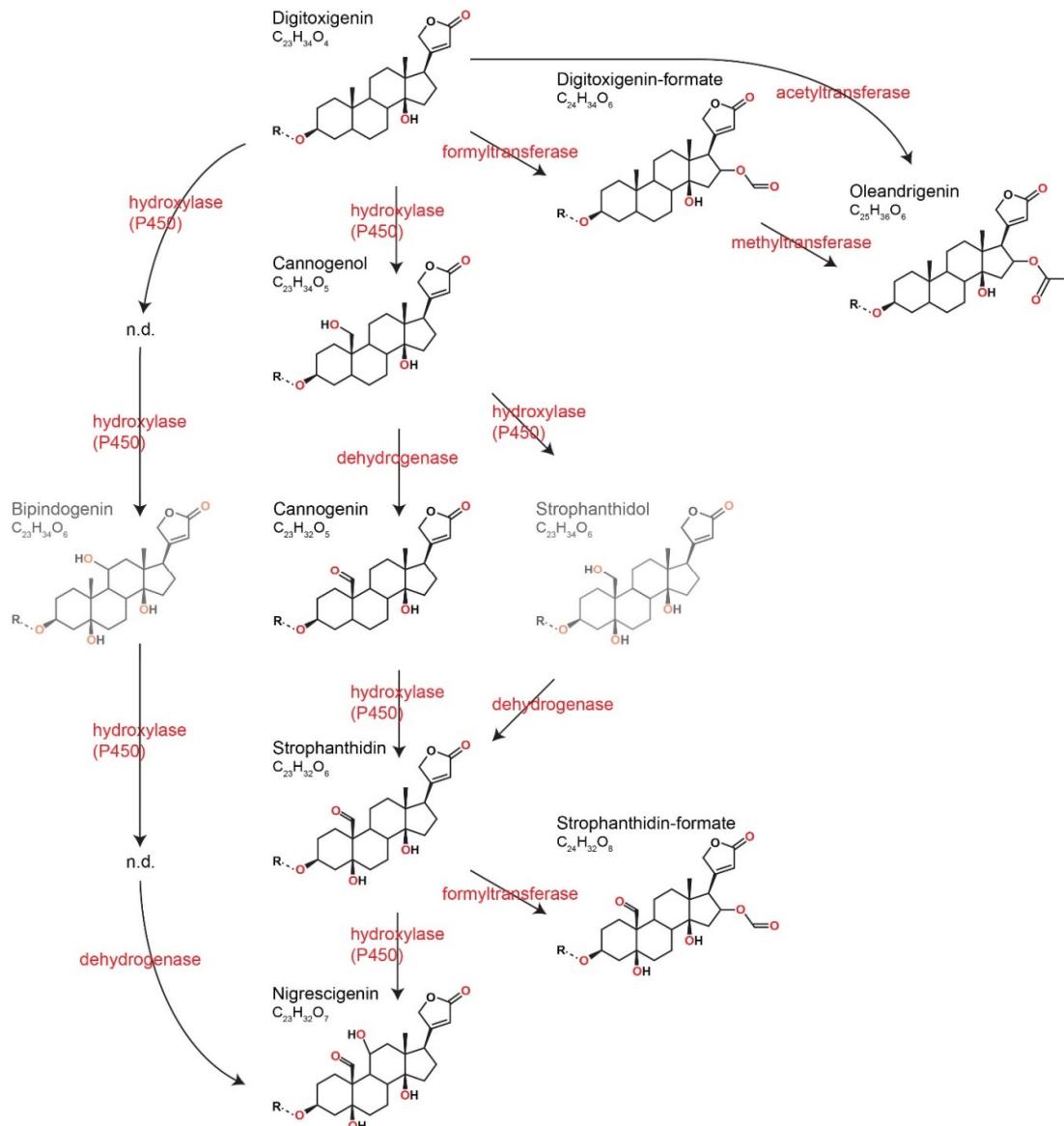
760

761 *Cardenolide diversity*

762 With the exception of *E. collinum* (COL), which only contained trace amounts of cardenolides in
763 leaves, all *Erysimum* species contained diverse mixtures of cardenolide compounds and accumulated
764 considerable amounts of cardenolides (Figure 6D-E). The ploidy level of species again explained a
765 significant fraction of the total variation in the number of cardenolides ($F_{4,38} = 3.47$, $p = 0.016$), with
766 hexaploid species producing the highest average number of compounds (Figure S9). To obtain an
767 estimate of biological activity and evaluate quantification from total MS ion counts, we used an
768 established assay that quantifies cardenolide concentrations from specific inhibition of animal Na^+/K^+ -
769 ATPase by crude *Erysimum* leaf extracts. We found generally strong enzymatic inhibition, with leaves
770 of *Erysimum* species containing an equivalent of $5.72 \pm 0.12 \mu\text{g mg}^{-1}$ (± 1 SE) of the reference
771 cardenolide ouabain on average. Despite only producing trace amounts of cardenolides, *E. collinum*
772 (COL) extracts caused significantly stronger inhibition than the Brassicaceae control, *S. arvensis*
773 (Figure S10). Overall, quantification of cardenolide concentrations by Na^+/K^+ -ATPase inhibition was
774 highly correlated with the total MS ion count (Fig S8, $r = 0.95$, $p < 0.001$). Thus, the use of ion count
775 data for cross-species comparisons was appropriate for this purpose. Both the total numbers of
776 compounds and the total abundances exhibited a strong phylogenetic signal (Table 2), indicating that
777 closely-related species were more similar in their cardenolide traits than expected by chance.
778 However, there was again no evidence for a directional trend in the evolution of either number or
779 abundance of cardenolides (Table 3), suggesting a rapid rather than a gradual gain of cardenolide

780 diversity, which is also evident from the considerable number of cardenolide compounds present in
781 the earliest-diverging species (Figure 6D).

782

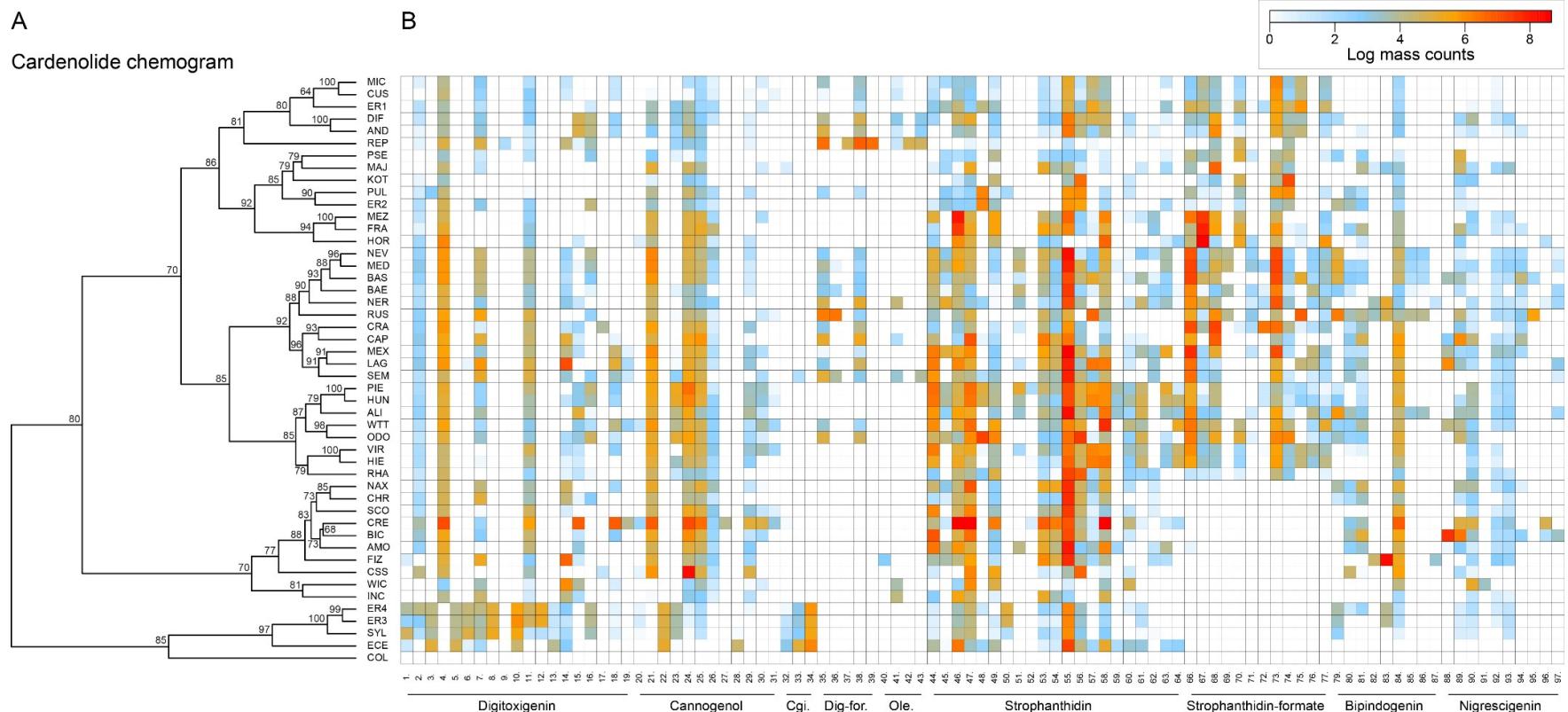


783

784 **Figure 9.** Predicted pathways of cardenolide genin modification in *Erysimum*. Genin diversity likely originates
785 from digitoxigenin, which by hydroxylases (P450-like enzymes), dehydrogenases, and formyl-, methyl-, or
786 acetyltransferases is transformed into structurally more complex cardenolides. Oleandrinogenin could be derived
787 from digitoxigenin or from digitoxigenin-formate, with the frequent co-occurrence of oleandrinogenin and
788 digitoxigenin-formate in leaf extracts suggesting the latter. According to their exact mass, frequently detected
789 dihydroxy-digitoxigenin compounds ($C_{23}H_{34}O_6$) could be either bipindogenin or strophanthidol. While
790 bipindogenin cardenolides have commonly been reported for *Erysimum* species in the literature, their structure
791 would require additional intermediate structures that were not detected here (n.d.). Thus, strophanthidol appears
792 to be the more likely isomer to occur in *Erysimum*. All cardenolide genins are further modified by glycosylation
793 at a conserved position in the molecule (R).

794 Cardenolide diversity was considerably higher than that of glucosinolates, with a total of 95
795 distinguishable candidate cardenolide compounds identified across the 48 *Erysimum* species (Table
796 S5). Of these, 46 compounds had distinct molecular masses and mass fragments, while the remaining
797 compounds likely were isomers, sharing a molecular mass with another compound but having a
798 distinct HPLC retention time. The 95 putative cardenolides comprised nine distinct genins (Figures 9,
799 S11), the majority of which were glycosylated with digitoxose, deoxy hexoses, xylose, or glucose
800 moieties. Only digitoxigenin and cannogenol accumulated as free genins, while all other compounds
801 occurred as either mono- or di-glycosides. A major source of isomeric cardenolide compounds was
802 thus likely the incorporation of different deoxy hexoses of equivalent mass, such as rhamnose, fucose,
803 or gulomethyllose. A subset of compounds had molecular masses that were heavier by 42.011 m/z than
804 known mono- or di-glycoside cardenolides. Such a gain in mass corresponds to the gain of an acetyl-
805 group, and mass fragmentation patterns indicated that these compounds were acetylated on the first
806 sugar moiety (Table S5). Out of the nine detected genins, six had previously been described from
807 *Erysimum* species (Makarevich et al. 1994). In addition, we identified three previously undescribed
808 mass features with fragmentation patterns characteristic of cardenolide genins (Figure S11). Of these
809 three, one matched an acetylated digitoxigenin (also known as oleandrigenin), a common cardenolide
810 in *Nerium oleander*. The other two matched molecular structures of digitoxigenin-formate (also
811 known as gitaloxigenin) and strophantidin-formate. Formate adducts can sometimes be formed
812 during LC-MS due to the addition of formic acid to solvents, although this is less common with
813 positive ionization. To exclude the possibility that these were technical artefacts, we analyzed a subset
814 of extracts by LC-MS without the addition of formic acid and found both genin-formates at
815 comparable concentrations (Figure S12). We therefore assume that all three novel structures are
816 natural variants of cardenolides produced by *Erysimum* plants, even though we currently lack final
817 structural elucidation.

818 Clustering of species based on similarity in cardenolide profiles revealed fewer obvious
819 species clusters than for glucosinolates, and particularly higher-level species clusters had only weak
820 statistical support (Figure 10). A clear exception to this was a species cluster that included *E.*
821 *cheiranthoides* (ECE) and *E. sylvestre* (SYL), which lacked several otherwise common cannogenol-
822 and strophantidin-glycosides, while accumulating unique digitoxigenin-glycosides. A second major
823 cluster that was visually apparent – yet not statistically significant – separated groups of species that
824 did or did not produce glycosides of the newly discovered putative strophantidin-formate (Figure
825 10). Similarity in cardenolide profiles among species quantified as the first and second principal
826 coordinate of the Bray-Curtis dissimilarity matrix exhibited a very strong phylogenetic signal (Table
827 2), suggesting that closely-related species not only were more similar in their total cardenolide
828 concentrations, but also had more similar cardenolide profiles than expected by chance.

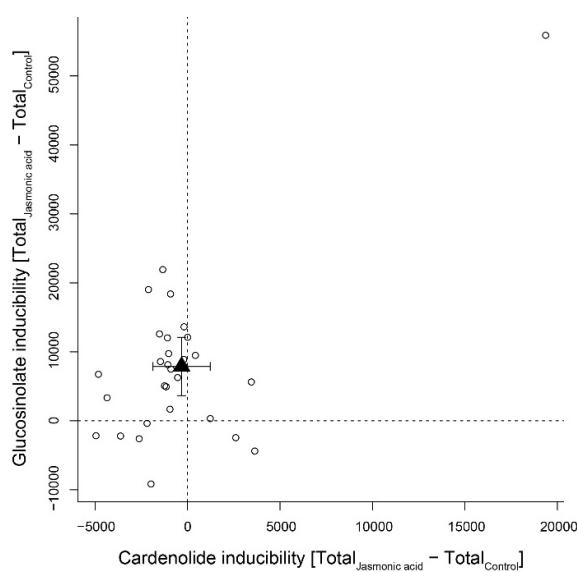


829

830 **Figure 10.** (A) Cardenolide chemogram clustering species according to similarity in cardenolide profiles. Values at nodes are confidence estimates based on 10,000 iterations
 831 of multiscale bootstrap resampling. (B) Heatmap of cardenolide profiles expressed by the 48 *Erysimum* species. Color intensity corresponds to log-transformed integrated ion
 832 counts recorded at the exact parental mass ($[M+H]^+$ or $[M+Na]^+$, whichever was more abundant) for each compound, averaged across samples from multiple independent
 833 experiments. The species *E. collinum* (COL) only expressed trace amounts of cardenolides, which are not visible on the color scale. Compounds are grouped by shared genin
 834 structures. Cgi. = Cannogenin, Dig-for. = Digitoxigenin-formate, Ole. = Oleandrigenicin. See Table S5 for additional compound information.

835 **Macroevolutionary patterns in defense and inducibility**

836 Given the very distinct patterns for glucosinolate and cardenolide diversity among *Erysimum* species,
837 it is unsurprising that concentrations of the two defense traits were not correlated (Pearson's
838 correlation: $r = -0.09$, $p = 0.534$). Foliar application of JA was expected to stimulate defense levels in
839 plant leaves, and among the 30 tested species, glucosinolate levels responded positively to JA, with
840 the majority of species increasing their foliar glucosinolate concentration (Figure 11). However, the
841 glucosinolate inducibility of a species was independent of constitutive glucosinolate levels ($r = -0.26$,
842 $p = 0.169$). By contrast, the majority of species exhibited lower cardenolide levels in response to JA,
843 resulting in lack of inducibility across species (Figure 11). The species *E. crepidifolium* (CRE)
844 heavily influenced inducibility patterns, as it not only had three times higher constitutive
845 concentrations of cardenolides than any other *Erysimum* species, but in addition pronouncedly
846 increased both glucosinolate and cardenolide concentrations in response to JA treatment (Figure 11).
847 If this outlier was removed, inducibility (or suppression) of foliar cardenolides was not correlated with
848 constitutive cardenolide levels ($r = -0.21$, $p = 0.284$), and inducibilities of glucosinolates and
849 cardenolides were likewise not correlated with each other ($r = 0.01$, $p = 0.995$).
850



851 **Figure 11.** Inducibility of foliar glucosinolates and cardenolides in response to exogenous application of
852 jasmonic acid (JA), expressed as absolute differences in total mass intensity between JA-treated and control
853 plants. Circles are species means, based on single pooled samples of multiple individual plants. The filled
854 triangle is the average inducibility of all measured species with 95% confidence interval. Non-overlap with zero
855 (dashed lines) corresponds to a significant effect. The species in the upper right corner is *E. crepidifolium*, an
856 outlier and strong inducer of both glucosinolates and cardenolides.
857
858

859 **Discussion**

860 The genus *Erysimum* is a fascinating model system of phytochemical diversification that combines
861 two potent classes of chemical defenses in the same plants. The assembled genome of the short-lived
862 annual plant *E. cheiranthoides* allowed us to identify almost the full set of genes involved in *E.*
863 *cheiranthoides* glucosinolate biosynthesis and myrosinase expression. This genome
864 (www.erysimum.org) will facilitate further identification of glucosinolate genes unique to *Erysimum*
865 and represents a central resource for the identification of cardenolide biosynthesis genes in this
866 emerging model system, as well as for future functional and evolutionary studies in the Brassicaceae.

867 The extant species diversity in the genus *Erysimum* is the result of an evolutionarily recent,
868 rapid radiation (Moazzeni et al. 2014). All but one species in our study produced evolutionary novel
869 cardenolides, while the likely closest relatives – the genera *Malcolmia*, *Physaria* or *Arabidopsis*
870 (Moazzeni et al. 2014, Huang et al. 2016) – almost certainly lack these defenses (Jaretzky and Wilcke
871 1932, Hegnauer 1964). The onset of diversification in *Erysimum* thus appears to coincide with the
872 gain of the cardenolide defense trait. However, even though most species co-expressed two different
873 classes of potentially costly defenses, there was no evidence for a trade-off between glucosinolates
874 and cardenolides. Furthermore, neither defense showed a directional trend from the root to the tip of
875 the phylogenetic tree, and both defenses were similarly diverse in early- and late-diverging species.

876 Potentially costly, obsolete defenses are expected to be selected against and should disappear
877 over evolutionary time. For example, cardenolides in the genus *Asclepias* and alkaloids across the
878 Apocynaceae decrease in concentration with speciation, consistent with co-evolutionary de-escalation
879 in response to specialized, sequestering herbivores (Agrawal and Fishbein 2008, Livshultz et al.
880 2018). Due to the relatively recent evolution of cardenolides in *Erysimum*, the system presumably
881 lacks cardenolide-specialized herbivores that could exert negative selection on cardenolides, while the
882 diversity of cardenolides may have evolved too rapidly to detect positive selection. It is therefore
883 likely that the two defenses serve distinct functions: glucosinolates are highly efficient in repelling
884 generalist herbivores (Kerwin et al. 2015), whereas cardenolides may be functionally relevant against
885 glucosinolate-specialized herbivores (Chew 1975, 1977, Wiklund and Åhrberg 1978, Renwick et al.
886 1989, Dimock et al. 1991).

887 As further evidence for the distinct roles of glucosinolates and cardenolides, the two defenses
888 responded differently to exogenous JA application. Glucosinolate concentrations were upregulated in
889 response to JA in the majority of species, with an average 52% increase relative to untreated controls.
890 This is similar to inducibility of glucosinolates reported for other Brassicaceae species (Textor and
891 Gershenzon 2009), suggesting that glucosinolate defense signaling remains unaffected by the
892 presence of cardenolides in *Erysimum* plants. In contrast, cardenolide levels were not inducible or
893 were even suppressed in response to exogenous application of JA in almost all tested species,
894 suggesting that inducibility of cardenolides is not a general strategy of *Erysimum*. In the more
895 commonly-studied milkweeds (*Asclepias* spp., Apocynaceae), cardenolides are usually inducible in

896 response to herbivore stimuli (Rasmann et al. 2009, Bingham and Agrawal 2010), but cardenolide
897 suppression is also common, particularly in plants with high constitutive cardenolide concentrations
898 (Bingham and Agrawal 2010, Rasmann and Agrawal 2011). It thus appears that cardenolides
899 accumulate constitutively in *Erysimum*, perhaps due to the presumed lack of cardenolide-specialized
900 herbivores that would use cardenolides as host-finding cues.

901

902 ***Phylogenetic relationships and phytochemical diversity***

903 The young evolutionary age of *Erysimum* and the potential ongoing gene flow among related species
904 may have limited genetic differentiation. In fact, both previous partial phylogenies of the genus
905 struggled to resolve polytomies among species using conventional approaches (Gómez et al. 2014,
906 Moazzeni et al. 2014). In contrast, we present a highly-resolved species tree for 10 – 30 % of the total
907 species diversity in the genus. Our tree, which was constructed from transcriptome sequences of 9,868
908 genes with syntenic genome locations, revealed a strong geographic signature in phylogenetic
909 relatedness, an observation shared by the previously published phylogenies (Gómez et al. 2014,
910 Moazzeni et al. 2014). In our species tree, all annual species belonged to early-diverging clades,
911 suggesting that the predominant perennial growth strategy in the genus is a derived state. Three of the
912 annual species, *E. repandum*, *E. incanum*, and *E. wilczekianum*, have widely different distribution
913 ranges but were all collected in Spain or Morocco for this study. Although they co-occur
914 geographically with several perennial *Erysimum* species, they are largely isolated by non-overlapping
915 flowering times.

916 Perennial species from Spain, Morocco, and the Canary Islands formed a monophyletic clade,
917 with species from southeastern Spain exhibiting closer relatedness to Moroccan species than to
918 species from northeastern or northwestern Spain. Among the species from southeastern Spain, our
919 phylogeny did not agree with a more fine-scale evaluation of species relatedness (Abdelaziz et al.
920 2014), although this may reflect the limitation of using single accessions of each species in our
921 approach, the possible hybridization occurring between these species (Abdelaziz et al. 2014), or the
922 different sensitivity of phylogenies based on internal transcribed spacer (ITS) sequences to incomplete
923 lineage sorting (Feliner and Rossello 2007). Additional geographic clades were recovered for species
924 from North America, Iran, and Greece/Turkey. Surprisingly, the North American clade was most
925 closely related to the Spanish clade, which does not fit traditional models of dispersal across land
926 bridges. However, the node connecting the two geographic clades had by far the weakest support
927 across the whole phylogeny (local posterior probability = 0.5) and, as our phylogeny did not include
928 East Asian *Erysimum* species, it is possible that the addition of further species would change this
929 grouping. The remaining species belonged to one of three distinct central European clades with no
930 obvious geographic separation. As a prominent exception, the central European *E. crepidifolium*
931 (CRE) was most closely related to Greek *Erysimum* species rather than to other central European
932 species.

933 Despite vast morphological differences among sampled *Erysimum* species, the diversity in
934 glucosinolate profiles was relatively limited compared to the diversity that is present within
935 *Arabidopsis* (Kliebenstein et al. 2001). However, broader comparative studies of glucosinolate
936 diversity in other Brassicaceae species would be needed to provide a more natural ‘baseline’ for
937 glucosinolate diversity. The majority of *Erysimum* species produced glucoiberin as their main
938 glucosinolate. Aliphatic glucosinolates such as glucoiberin are derived from methionine in a process
939 that involves elongation and modification of a variable side-chain (Halkier and Gershenson 2006),
940 and in this context the 3-carbon glucosinolate glucoiberin is one of the least biosynthetically complex
941 glucosinolates. However, the potential to produce additional aliphatic glucosinolates with longer side
942 chains clearly exists in the genus, as 4-, 5-, and 6-carbon glucosinolates with more complex
943 modifications were scattered across the phylogeny. A few species produced glucosinolates that are not
944 found in *Arabidopsis*, including a sub-class of aliphatic glucosinolates, the methylsulfonyl
945 glucosinolates. The homolog of 3-butetyl glucosinolate 2-hydroxylase (GS-OH), which in
946 *Arabidopsis* forms 2-hydroxy-but-3-enyl glucosinolate from 3-butetyl glucosinolate, does not have a
947 clear function in *E. cheiranthoides* due to the lack of alkenyl glucosinolates. However, it is possible
948 that the GS-OH homolog in *E. cheiranthoides* may code for the unknown enzyme that hydroxylates 4-
949 methylsulfonylbutyl glucosinolate to form 3-hydroxy-4-methylsulfonylbutyl glucosinolate (Figure 4).
950 Methylsulfonyl glucosinolates are found in several Brassicaceae genera (Fahey et al. 2001), and
951 glucocheirolin, the most abundant methylsulfonyl glucosinolate in *Erysimum* species, is only a weak
952 egg-laying stimulant for the cabbage white butterfly (*Pieris rapae*), compared to other glucosinolates
953 (Huang et al. 1993). Methylsulfonyl glucosinolates may thus represent a plant response to specialist
954 herbivores that use plant defenses as host-finding cues.

955 The species *E. pulchellum* (PUL) and *E. collinum* (COL) from Turkey and Iran, respectively,
956 accumulated glucoerypestrin as their main glucosinolate compound. This compound was first
957 described in *E. rupestre* [syn. *E. pulchellum*, (Polatschek 2011)] by Kjaer & Gmelin (1957) and to
958 date has been found exclusively in plants of the genus *Erysimum* (Fahey et al. 2001). Radioactive
959 labeling experiments indicated that glucoerypestrin is derived from a dicarboxylic amino acid,
960 possibly 2-amino-5-methoxycarbonyl-pentanoic acid (Chisholm 1973). Modification of the amino
961 acid side chain during methionine-derived aliphatic glucosinolate biosynthesis as a pathway to
962 glucoerypestrin is less likely, due to the lower specific incorporation of ¹⁴C-labeled methionine
963 compared to ¹⁴C-labeled dicarboxylic acids into this compound (Chisholm 1973). In any case, the gain
964 of glucoerypestrin represents yet another evolutionary novelty in the *Erysimum* genus, but its relative
965 toxicity and the adaptive benefits of its production have yet to be elucidated.

966 We found no phylogenetic signal of glucosinolate chemotype, as more closely related species
967 were not more likely to share the same glucosinolate profile. The pattern of more complex
968 glucosinolates scattered across the phylogenetic tree may be generated by horizontal gene transfer,
969 during hybridization, or by repeated gains and losses of biosynthesis genes as species diverge. The

970 latter may also be facilitated by changes in ploidy, as hexaploid species in particular accumulated
971 large numbers of both glucosinolate and cardenolide compounds. Alternatively, a full complement of
972 synthesis genes may be maintained in species' gene pools at low frequencies until they are favored by
973 a new environment, or they might be maintained in the genome but not expressed in leaves. More
974 extensive sampling within each species will be required to conclusively address this question,
975 although a preliminary screening of multiple *E. cheiranthoides* accessions suggests little to no
976 variation in glucosinolate profiles within this species (T. Züst, unpublished data).

977 Myrosinase activity levels differed among glucosinolate chemotypes, and activity was
978 positively correlated with glucosinolate abundance in plants when controlling for glucosinolate
979 chemotype. *Erysimum* species that predominantly produced indole glucosinolates or 4-methylsulfinyl
980 glucosinolates had negligible myrosinase activity against the assayed aliphatic glucosinolate sinigrin.
981 Indole glucosinolates can be activated by PEN2 – a thioglucosidase that is more specific for indole
982 glucosinolates (Bednarek et al. 2009, Clay et al. 2009) – or even break down in the absence of plant-
983 derived myrosinase (Kim et al. 2008). The negligible activity in these species could therefore indicate
984 the existence of selective pressures to tailor myrosinase expression to the type and concentrations of
985 glucosinolates that are produced. In contrast to glucosinolate defenses, myrosinase activity was more
986 similar among related species, suggesting that the two defense components are subject to different
987 selective regimes, with the potential for maladaptive combinations between glucosinolate defense and
988 myrosinase activity. In addition, myrosinase activity is highly dependent on the presence of other
989 proteins and cofactors (Halkier and Gershenson 2006), which may also differ between *Erysimum*
990 species.

991 We detected considerable amounts of the evolutionarily novel cardenolide defense in 47 out
992 of 48 *Erysimum* species or accessions. Among the 95 likely cardenolide compounds, there were
993 several structures that had not been described previously in *Erysimum*. This metabolic diversity had
994 three main sources: modification of the genin core structure, variation of the glycoside chain, or
995 isomeric variation (e.g., through the incorporation of different isomeric sugars). Structural variation in
996 cardenolides affects the relative inhibition of Na⁺/K⁺-ATPase (Dzimiri et al. 1987, Petschenka et al.
997 2018) and physiochemical properties such as lipophilicity, which play an important role in uptake and
998 metabolism of plant metabolites by insects (Duffey 1980). Individual *Erysimum* species produced
999 between 15 and 50 different cardenolide compounds, and the comparison of quantification by total
1000 mass ion counts vs. quantification by inhibition of Na⁺/K⁺-ATPase revealed highly similar results.
1001 While both methods of quantification are only approximate, this correspondence at least provides no
1002 obvious indication of vast differences in Na⁺/K⁺-ATPase inhibitory activity among *Erysimum*
1003 cardenolides.

1004 The metabolic pathways involved in the biosynthesis and modification of cardenolides have
1005 yet to be elucidated (Kreis and Müller-Uri 2010, Züst et al. 2018). Here, we propose a pathway for the
1006 modification of digitoxigenin, commonly assumed to be the least biosynthetically complex

1007 cardenolide (Kreis and Müller-Uri 2010), into the eight structurally more complex genins found
1008 within *Erysimum* (Figure 9). Variation in glycoside chains is likely mediated by glycosyltransferases
1009 that act on the different genins. In the Brassicaceae genus *Barbarea*, plants produce saponin
1010 glycosides as an evolutionary novel defense, and a significant proportion of glycoside diversity in this
1011 system has been linked to the action of a small set of UDP glycosyltransferases (Erthmann et al.
1012 2018). Similarly, through the joint action of genin-modifying enzymes and glycosyltransferases, a
1013 relatively small set of enzymes and corresponding genes could generate the vast cardenolide diversity
1014 found in the *Erysimum* genus. The identification and manipulation of these genes in different
1015 *Erysimum* species will make it possible to test the adaptive benefits of this structural diversity.

1016 On average, leaves of *Erysimum* species contained cardenolides equivalent to 6 µg ouabain
1017 per mg dry leaf weight (estimated from Na⁺/K⁺-ATPase inhibition), placing them slightly above most
1018 species of the well-studied cardenolide-producing genus *Asclepias* (Rasmann and Agrawal 2011).
1019 However, two species, *E. collinum* (COL) and *E. crepidifolium* (CRE), were clear outliers in terms of
1020 cardenolide content (Figure 6D-E). The almost complete absence of cardenolides in *E. collinum*
1021 (COL), which clustered phylogenetically with two other Middle Eastern species producing average
1022 concentrations of these compounds (*E. crassipes* [CSS] and *E. crassicaule* [CRA], Figure 5), likely
1023 represents a secondary loss of this trait in the course of evolution. This species also accumulated an
1024 evolutionary novel glucosinolate, glucoerypestrin (see above), which may have resulted in a shift in
1025 selective pressures that led to the loss of potentially costly cardenolide production. Conversely, *E.*
1026 *crepidifolium* (CRE) had cardenolide concentrations more than three times higher than any other
1027 tested *Erysimum* species. This is consistent with the highly toxic nature of this species, which has the
1028 German vernacular name ‘Gänsesterbe’ (geese death) and has been associated with mortality in geese
1029 that consume the plant.

1030 Whereas most species did not induce cardenolide accumulation in response to JA, *E.*
1031 *crepidifolium* (CRE) had a significant 48% increase. While not as extreme, this observation is similar
1032 to the results of Munkert et al. (2014), who reported a three-fold increase in cardenolide levels of *E.*
1033 *crepidifolium* in response to methyl jasmonate application. Plants use conserved transcriptional
1034 networks to continuously integrate signals from their environment and optimize allocation of
1035 resources to growth and defense (Havko et al. 2016). Thus, while these networks commonly govern
1036 hardwired responses (e.g., an attenuation of growth upon activation of JA signaling), they may
1037 nevertheless be altered by mutations at key nodes of the network (Campos et al. 2016). Given this
1038 relative flexibility in signaling networks, it is perhaps not surprising that the evolutionary novel
1039 cardenolides have been integrated into the defense signaling of *Erysimum* species to variable degrees.
1040 Investigating gene expression changes in the inducible *E. crepidifolium* as a contrast to the non-
1041 inducing *E. cheiranthoides* may therefore provide valuable insights into the molecular regulation of
1042 this defense.

1043 Cardenolide abundance and compound profiles of *Erysimum* species exhibited clear
1044 phylogenetic signals, with closely-related species being more phytochemically similar. However,
1045 similarities in cardenolide profiles changed more gradually between species than glucosinolate
1046 profiles, and distinct cardenolide chemotypes were less obvious. As the most distinct cardenolide
1047 cluster with underlying phylogenetic structure, the annual *E. cheiranthoides* (ECE) grouped together
1048 with *E. sylvestre* (SYL) and two accessions of commercial origin. These plants all shared a
1049 cardenolide chemotype defined by an unusually high proportion of digitoxigenin glycoside
1050 compounds, several of which were uniquely produced by plants of this cluster. This early-diverging
1051 species clade, which is defined by a chemotype of potentially lower biosynthetic complexity, could
1052 thus be an indication of a stepwise gain of structural complexity over the course of evolution.
1053

1054 **Conclusions**

1055 The study of the speciose genus *Erysimum* with two co-expressed chemical defense classes revealed
1056 largely independent evolution of the ancestral and the novel defense. With no evidence for trade-offs
1057 between the structurally and biosynthetically unrelated defenses, the diversity, abundance, and
1058 inducibility of each class of defenses appears to be evolving independently in response to the unique
1059 selective environment of each individual species. The evolutionarily recent gain of novel cardenolides
1060 has resulted in a system in which no known specific adaptations to cardenolides have evolved in
1061 insect herbivores, although general adaptations to toxic food may still allow herbivores to consume
1062 the plants. *Erysimum* is thus an ideal model system for phytochemical diversification, as it facilitates
1063 the study of coevolutionary adaptations in real time. Our current work provides the foundation for a
1064 more mechanistic evaluation of these processes, which promises to greatly improve our understanding
1065 of the role of phytochemical diversity for plant-insect interactions.
1066

1067 **Acknowledgments**

1068 We thank Hartmut Christier for collecting *E. cheiranthoides* seeds, Erik Poelman for collecting *E.*
1069 *cheiri* seeds, and the botanical gardens listed in Table S1 for providing seeds of additional species.
1070 Yvonne Künzi and Christoph Zwahlen assisted with the growing and maintenance of plants in the 48-
1071 species experiments, and with the harvesting and extraction of RNA and metabolomics samples.
1072 Sabrina Stiehler performed the Na⁺/K⁺-ATPase assay, and Jing Zhang helped with database
1073 maintenance. We thank Anurag Agrawal, Hamid Moazzeni, Gaurav Moghe, and Zephyr Züst for
1074 helpful advice and comments on the manuscript. This work was supported by Swiss National Science
1075 Foundation grant PZ00P3-161472 to TZ, a Triad Foundation grant to SRS and GJ, US National
1076 Science Foundation awards 1811965 to CKH and 1645256 to GJ, a fellowship from the Ministry of
1077 Science of Iran to MM, a German Research Foundation grant DFG-PE 2059/3-1 to GP, and a grant
1078 within the LOEWE program (Insect Biotechnology & Bioresources) of the State of Hesse, Germany
1079 to GP.

1080 **References**

- 1081 Abdelaziz, M., J. Lorite, A. J. Munoz-Pajares, M. B. Herrador, F. Perfectti, and J. M. Gomez. 2011.
1082 Using complementary techniques to distinguish cryptic species: a new *Erysimum*
1083 (Brassicaceae) species from North Africa. American Journal of Botany **98**:1049-1060.
1084 Abdelaziz, M., A. J. Muñoz-Pajares, J. Lorite, M. B. Herrador, F. Perfectti, and J. M. Gómez. 2014.
1085 Phylogenetic relationships of *Erysimum* (Brassicaceae) from the Baetic Mountains (SE
1086 Iberian Peninsula). Anales del Jardín Botánico de Madrid **71**:e005.
1087 Agrawal, A. A. 2005. Natural selection on common milkweed (*Asclepias syriaca*) by a community of
1088 specialized insect herbivores. Evolutionary Ecology Research **7**:651-667.
1089 Agrawal, A. A. and M. Fishbein. 2008. Phylogenetic escalation and decline of plant defense
1090 strategies. Proceedings of the National Academy of Sciences of the United States of America
1091 **105**:10057-10060.
1092 Agrawal, A. A., G. Petschenka, R. A. Bingham, M. G. Weber, and S. Rasmann. 2012. Toxic
1093 cardenolides: chemical ecology and coevolution of specialized plant-herbivore interactions.
1094 New Phytologist **194**:28-45.
1095 Al-Shehbaz, I. A. 1988. The genera of *Anchonieae* (Hesperideae) (Cruciferae; Brassicaceae) in the
1096 southeastern United States. Journal of the Arnold Arboretum **69**:193-212.
1097 Al-Shehbaz, I. A. 2010. *Erysimum* Linnaeus. Pages 534-545 in N. R. Morin, editor. Flora of North
1098 America North of Mexico. Oxford University Press, New York.
1099 Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment
1100 search tool. Journal of Molecular Biology **215**:403-410.
1101 Andersson, D., R. Chakrabarty, S. Bejai, J. M. Zhang, L. Rask, and J. Meijer. 2009. Myrosinases from
1102 root and leaves of *Arabidopsis thaliana* have different catalytic properties. Phytochemistry
1103 **70**:1345-1354.
1104 Barth, C. and G. Jander. 2006. *Arabidopsis* myrosinases TGG1 and TGG2 have redundant function in
1105 glucosinolate breakdown and insect defense. Plant Journal **46**:549-562.
1106 Bednarek, P., M. Pislewska-Bednarek, A. Svatos, B. Schneider, J. Doubsky, M. Mansurova, M.
1107 Humphry, C. Consonni, R. Panstruga, A. Sanchez-Vallet, A. Molina, and P. Schulze-Lefert.
1108 2009. A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum
1109 antifungal defense. Science **323**:101-106.
1110 Bidart-Bouzat, M. G. and D. J. Kliebenstein. 2008. Differential levels of insect herbivory in the field
1111 associated with genotypic variation in glucosinolates in *Arabidopsis thaliana*. Journal of
1112 Chemical Ecology **34**:1026-1037.
1113 Bingham, R. A. and A. A. Agrawal. 2010. Specificity and trade-offs in the induced plant defence of
1114 common milkweed *Asclepias syriaca* to two lepidopteran herbivores. Journal of Ecology
1115 **98**:1014-1022.
1116 Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for phylogenetic signal in comparative
1117 data: behavioral traits are more labile. Evolution **57**:717-745.
1118 Boutet, E., D. Lieberherr, M. Tognoli, M. Schneider, and A. Bairoch. 2007. UniProtKB/Swiss-Prot.
1119 Pages 89-112 in D. Edwards, editor. Plant Bioinformatics: Methods and Protocols. Humana
1120 Press.
1121 Brock, A., T. Herzfeld, R. Paschke, M. Koch, and B. Draeger. 2006. Brassicaceae contain nortropane
1122 alkaloids. Phytochemistry **67**:2050-2057.
1123 Campbell, M. S., M. Y. Law, C. Holt, J. C. Stein, G. D. Moghe, D. E. Hufnagel, J. K. Lei, R.
1124 Achawanantakun, D. Jiao, C. J. Lawrence, D. Ware, S. H. Shiu, K. L. Childs, Y. N. Sun, N.
1125 Jiang, and M. Yandell. 2014. MAKER-P: a tool kit for the rapid creation, management, and
1126 quality control of plant genome annotations. Plant Physiology **164**:513-524.
1127 Campos, M. L., Y. Yoshida, I. T. Major, D. D. Ferreira, S. M. Weraduwage, J. E. Froehlich, B. F.
1128 Johnson, D. M. Kramer, G. Jander, T. D. Sharkey, and G. A. Howe. 2016. Rewiring of
1129 jasmonate and phytochrome B signalling uncouples plant growth-defense tradeoffs. Nature
1130 Communications **7**.
1131 Cantarel, B. L., I. Korf, S. M. C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. S. Alvarado, and M.
1132 Yandell. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model
1133 organism genomes. Genome Research **18**:188-196.

- 1134 Cataldi, T. R. I., F. Lelario, D. Orlando, and S. A. Bufo. 2010. Collision-induced dissociation of the
1135 A+2 isotope ion facilitates glucosinolates structure elucidation by electrospray ionization-
1136 tandem mass spectrometry with a linear quadrupole ion trap. *Analitical Chemistry* **82**:5686-
1137 5696.
- 1138 Chen, W., S. Shakir, M. Bigham, A. Richter, Z. Feiz, and G. Jander. 2019. Genome sequence of the
1139 corn leaf aphid (*Rhopalosiphum maidis* Fitch). *GigaScience* **8**:giz033.
- 1140 Chew, F. S. 1975. Coevolution of pierid butterflies and their cruciferous foodplants. 1. Relative
1141 quality of available resources. *Oecologia* **20**:117-127.
- 1142 Chew, F. S. 1977. Coevolution of pierid butterflies and their cruciferous foodplants. 2. The
1143 distribution of eggs on potential foodplants. *Evolution* **31**:568-579.
- 1144 Chin, C. S., P. Peluso, F. J. Sedlazeck, M. Nattestad, G. T. Concepcion, A. Clum, C. Dunn, R.
1145 O'Malley, R. Figueroa-Balderas, A. Morales-Cruz, G. R. Cramer, M. Delledonne, C. Y. Luo,
1146 J. R. Ecker, D. Cantu, D. R. Rank, and M. C. Schatz. 2016. Phased diploid genome assembly
1147 with single-molecule real-time sequencing. *Nature Methods* **13**:1050-1054.
- 1148 Chisholm, M. 1973. Biosynthesis of 3-methoxycarbonylpropyl-glucosinolate in an *Erysimum* species.
1149 *Phytochemistry* **12**:605-608.
- 1150 Clay, N. K., A. M. Adio, C. Denoux, G. Jander, and F. M. Ausubel. 2009. Glucosinolate metabolites
1151 required for an *Arabidopsis* innate immune response. *Science* **323**:95-101.
- 1152 Cole, R. A. 1976. Isothiocyanates, nitriles and thiocyanates as products of autolysis of glucosinolates
1153 in *Cruciferae*. *Phytochemistry* **15**:759-762.
- 1154 Cornell, H. V. and B. A. Hawkins. 2003. Herbivore responses to plant secondary compounds: A test
1155 of phytochemical coevolution theory. *American Naturalist* **161**:507-522.
- 1156 Dimock, M. B., J. A. A. Renwick, C. D. Radke, and K. Sachdev-Gupta. 1991. Chemical constituents
1157 of an unacceptable crucifer, *Erysimum cheiranthoides*, deter feeding by *Pieris rapae*. *Journal
1158 of Chemical Ecology* **17**:525-533.
- 1159 Duffey, S. S. 1980. Sequestration of plant natural products by insects. *Annual Review of Entomology*
1160 **25**:447-477.
- 1161 Dzimir, N., U. Fricke, and W. Klaus. 1987. Influence of derivation on the lipophilicity and inhibitory
1162 actions of cardiac glycosides on myocardial Na⁺-K⁺-ATPase. *British journal of pharmacology*
1163 **91**:31-38.
- 1164 Ellinghaus, D., S. Kurtz, and U. Willhoeft. 2008. LTRharvest, an efficient and flexible software for de
1165 novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**:18.
- 1166 Emms, D. M. and S. Kelly. 2015. OrthoFinder: solving fundamental biases in whole genome
1167 comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**.
- 1168 English, A. C., S. Richards, Y. Han, M. Wang, V. Vee, J. X. Qu, X. Qin, D. M. Muzny, J. G. Reid, K.
1169 C. Worley, and R. A. Gibbs. 2012. Mind the gap: upgrading genomes with Pacific
1170 Biosciences RS long-read sequencing technology. *Plos One* **7**:e47768.
- 1171 Erthmann, P. O., N. Agerbirk, and S. Bak. 2018. A tandem array of UDP-glycosyltransferases from
1172 the UGT73C subfamily glycosylate saponins, forming a spectrum of mono- and
1173 bisdesmosidic saponins. *Plant Molecular Biology* **97**:37-55.
- 1174 Fahey, J. W., A. T. Zalcman, and P. Talalay. 2001. The chemical diversity and distribution of
1175 glucosinolates and isothiocyanates among plants. *Phytochemistry* **56**:5-51.
- 1176 Feeny, P. 1977. Defensive ecology of the Cruciferae. *Annals of the Missouri Botanical Garden*
1177 **64**:221-234.
- 1178 Feliner, G. N. and J. A. Rossello. 2007. Better the devil you know? Guidelines for insightful
1179 utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular
1180 Phylogenetics and Evolution* **44**:911-919.
- 1181 Firn, R. D. and C. G. Jones. 2003. Natural products - a simple model to explain chemical diversity.
1182 *Natural Product Reports* **20**:382-391.
- 1183 Forbey, J. S., M. D. Dearing, E. M. Gross, C. M. Orians, E. E. Sotka, and W. J. Foley. 2013. A
1184 pharm-ecological perspective of terrestrial and aquatic plant-herbivore interactions. *Journal of
1185 Chemical Ecology* **39**:465-480.
- 1186 Fraenkel, G. S. 1959. The raison d'être of secondary plant substances. *Science* **129**:1466-1470.
- 1187 Frisch, T. and B. L. Møller. 2012. Possible evolution of alliarinoside biosynthesis from the
1188 glucosinolate pathway in *Alliaria petiolata*. *Febs Journal* **279**:1545-1562.

- 1189 German, D. A. 2014. Notes on taxonomy of *Erysimum* (*Erysimeae*, Cruciferae) of Russia and adjacent
1190 states. 1. *Erysimum collinum* and *Erysimum hajastanicum*. *Turczaninowia* **17**:10-32.
- 1191 Gershenson, J., A. Fontana, M. Burow, U. Wittstock, and J. Degenhardt. 2012. Mixtures of plant
1192 secondary metabolites: metabolic origins and ecological benefits. In G. R. Iason, M. Dicke,
1193 and S. E. Hartley, editors. *The ecology of plant secondary metabolites*. Cambridge University
1194 Press, Cambridge.
- 1195 Gómez, J. M. 2005. Non-additive effects of herbivores and pollinators on *Erysimum mediohispanicum*
1196 (Cruciferae) fitness. *Oecologia* **143**:412-418.
- 1197 Gómez, J. M., F. Perfectti, and C. P. Klingenberg. 2014. The role of pollinator diversity in the
1198 evolution of corolla-shape integration in a pollination-generalist plant clade. *Philosophical
1199 Transactions of the Royal Society B* **369**:20130257.
- 1200 Gómez, J. M., F. Perfectti, and J. Lorite. 2015. The role of pollinators in floral diversification in a
1201 clade of generalist flowers. *Evolution* **69**:863-878.
- 1202 Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, T. Mitros, W. Dirks, U.
1203 Hellsten, N. Putnam, and D. S. Rokhsar. 2011. Phytozome: a comparative platform for green
1204 plant genomics. *Nucleic Acids Research* **40**:D1178-D1186.
- 1205 Gordon, S. P., E. Tseng, A. Salamov, J. W. Zhang, X. D. Meng, Z. Y. Zhao, D. W. Kang, J.
1206 Underwood, I. V. Grigoriev, M. Figueroa, J. S. Schilling, F. Chen, and Z. Wang. 2015.
1207 Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing.
1208 *Plos One* **10**:e0132628.
- 1209 Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D.
1210 Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks,
1211 R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. Leduc, N. Friedman, and A.
1212 Regev. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity
1213 platform for reference generation and analysis. *Nature Protocols* **8**:1494-1512.
- 1214 Halkier, B. A. and J. Gershenson. 2006. Biology and biochemistry of glucosinolates. *Annual Review
1215 of Plant Biology* **57**:303-333.
- 1216 Haribal, M. and J. A. A. Renwick. 2001. Seasonal and population variation in flavonoid and
1217 alliarinoside content of *Alliaria petiolata*. *Journal of Chemical Ecology* **27**:1585-1594.
- 1218 Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: investigating
1219 evolutionary radiations. *Bioinformatics* **24**:129-131.
- 1220 Havko, N. E., I. T. Major, J. B. Jewell, E. Attaran, J. Browse, and G. A. Howe. 2016. Control of
1221 carbon assimilation and partitioning by jasmonate: an accounting of growth–defense
1222 tradeoffs. *Plants* **5**:7.
- 1223 Hegnauer, R. 1964. *Chemotaxonomie der Pflanzen III*. Birkhäuser Verlag, Basel, Switzerland.
- 1224 Huang, C. H., R. R. Sun, Y. Hu, L. P. Zeng, N. Zhang, L. M. Cai, Q. Zhang, M. A. Koch, I. Al-
1225 Shehbaz, P. P. Edger, J. C. Pires, D. Y. Tan, Y. Zhong, and H. Ma. 2016. Resolution of
1226 Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports
1227 convergent morphological evolution. *Molecular Biology and Evolution* **33**:394-412.
- 1228 Huang, X., J. A. A. Renwick, and K. Sachdevgupta. 1993. A chemical basis for differential
1229 acceptance of *Erysimum cheiranthoides* by two *Pieris* species. *Journal of Chemical Ecology*
1230 **19**:195-210.
- 1231 Iason, G. R., J. M. O'Reilly-Wapstra, M. J. Brewer, R. W. Summers, and B. Moore. 2011. Do
1232 multiple herbivores maintain chemical diversity of Scots pine monoterpenes? *Philosophical
1233 Transactions of the Royal Society B: Biological Sciences* **366**:1337-1345.
- 1234 Illumina. 2017. Effects of index misassignment on multiplexing and downstream analysis. Illumina
1235 Whitepapers.
- 1236 Jaretzky, R. and M. Wilcke. 1932. Die herzwirksamen Glykoside von *Cheiranthes cheiri* und
1237 verwandten Arten. *Archiv der Pharmazie* **270**:81-94.
- 1238 Jones, P., D. Binns, H. Y. Chang, M. Fraser, W. Z. Li, C. McAnulla, H. McWilliam, J. Maslen, A.
1239 Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y.
1240 Yong, R. Lopez, and S. Hunter. 2014. InterProScan 5: genome-scale protein function
1241 classification. *Bioinformatics* **30**:1236-1240.
- 1242 Katoh, K., K. Misawa, K. Kuma, and T. Miyata. 2002. MAFFT: a novel method for rapid multiple
1243 sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**:3059-3066.

- 1244 Katz, E., S. Nisani, B. S. Yadav, M. G. Woldemariam, B. Shai, U. Obolski, M. Ehrlich, E. Shani, G.
1245 Jander, and D. A. Chamovitz. 2015. The glucosinolate breakdown product indole-3-carbinol
1246 acts as an auxin antagonist in roots of *Arabidopsis thaliana*. *The Plant Journal* **82**:547-555.
- 1247 Kerwin, R., J. Feusier, J. Corwin, M. Rubin, C. Lin, A. Muok, B. Larson, B. Li, B. Joseph, M.
1248 Francisco, D. Copeland, C. Weinig, and D. J. Kliebenstein. 2015. Natural genetic variation in
1249 *Arabidopsis thaliana* defense metabolism genes modulates field fitness. *eLife* **4**:1-28.
- 1250 Kim, D., B. Landmead, and S. L. Salzberg. 2015. HISAT: a fast spliced aligner with low memory
1251 requirements. *Nature Methods* **12**:357-360.
- 1252 Kim, J. H., B. W. Lee, F. C. Schroeder, and G. Jander. 2008. Identification of indole glucosinolate
1253 breakdown products with antifeedant effects on *Myzus persicae* (green peach aphid). *Plant
1254 Journal* **54**:1015-1026.
- 1255 Kjaer, A. and R. Gmelin. 1957. Isothiocyanates XXV: methyl 4-isothiocyanatobutyrate, a new
1256 mustard oil present as a glucoside (glucoerypestrin) in *Erysimum* species. *Acta Chemica
1257 Scandinavica* **11**:577-578.
- 1258 Klauck, D. and M. Luckner. 1995. In vitro measurement of digitalis-like compounds by inhibition of
1259 Na^+/K^+ -ATPase: determination of the inhibitory effect. *Pharmazie* **50**:395-399.
- 1260 Kliebenstein, D. J., J. Kroymann, P. Brown, A. Figuth, D. Pedersen, J. Gershenson, and T. Mitchell-
1261 Olds. 2001. Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation.
1262 *Plant Physiology* **126**:811-825.
- 1263 Korf, I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* **5**:59.
- 1264 Kreis, W. and F. Müller-Uri. 2010. Biochemistry of sterols, cardiac glycosides, brassinosteroids,
1265 phytoecdysteroids and steroid saponins. Pages 304-363 in M. Wink, editor. *Biochemistry of
1266 plant secondary metabolism*. CRC Press, Sheffield.
- 1267 Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A.
1268 Marra. 2009. Circos: an information aesthetic for comparative genomics. *Genome Research*
1269 **19**:1639-1645.
- 1270 Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura. 2018. MEGA X: molecular evolutionary
1271 genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**:1547-
1272 1549.
- 1273 Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. L. Salzberg.
1274 2004. Versatile and open software for comparing large genomes. *Genome Biology* **5**:R12.
- 1275 Livshultz, T., E. Kaltenegger, S. C. K. Straub, K. Weitemier, E. Hirsch, K. Koval, L. Mema, and A.
1276 Liston. 2018. Evolution of pyrrolizidine alkaloid biosynthesis in Apocynaceae: revisiting the
1277 defence de-escalation hypothesis. *New Phytologist* **218**:762-773.
- 1278 Makarevich, I. F., K. V. Zhernoklev, T. V. Slyusarskaya, and G. N. Yarmolenko. 1994. Cardenolide-
1279 containing plants of the family Cruciferae. *Chemistry of Natural Compounds* **30**:275-289.
- 1280 Mapleson, D., L. Venturini, G. Kaithakottil, and D. Swarbreck. 2018. Efficient and accurate detection
1281 of splice junctions from RNA-seq with Portcullis. *GigaScience* **7**:giy131.
- 1282 Mithöfer, A. and W. Boland. 2012. Plant defense against herbivores: chemical aspects. *Annual
1283 Review of Plant Biology* **63**:431-450.
- 1284 Moazzeni, H., S. Zarre, B. E. Pfeil, Y. J. K. Bertrand, D. A. German, I. A. Al-Shehbaz, K.
1285 Mummenhoff, and B. Oxelman. 2014. Phylogenetic perspectives on diversification and
1286 character evolution in the species-rich genus *Erysimum* (Erysimeae; Brassicaceae) based on a
1287 densely sampled ITS approach. *Botanical Journal of the Linnean Society* **175**:497-522.
- 1288 Moore, B., R. L. Andrew, C. Külheim, and W. J. Foley. 2014. Explaining intraspecific diversity in
1289 plant secondary metabolites in an ecological context. *New Phytologist* **201**:733-750.
- 1290 Müller, C. 2009. Interactions between glucosinolate- and myrosinase-containing plants and the sawfly
1291 *Athalia rosae*. *Phytochemistry Reviews* **8**:121-134.
- 1292 Munkert, J., M. Ernst, F. Müller-Uri, and W. Kreis. 2014. Identification and stress-induced expression
1293 of three β -hydroxysteroid dehydrogenases from *Erysimum crepidifolium* Rchb. and their
1294 putative role in cardenolide biosynthesis. *Phytochemistry* **100**:26-33.
- 1295 Nagata, W., C. Tamm, and T. Reichstein. 1957. Die Glykoside von *Erysimum crepidifolium* H. G. L.
1296 Reichenbach. *Helvetica Chimica Acta* **40**:41-61.
- 1297 Nakano, R. T., M. Pislewska-Bednarek, K. Yamada, P. P. Edger, M. Miyahara, M. Kondo, C.
1298 Bottcher, M. Mori, M. Nishimura, P. Schulze-Lefert, I. Hara-Nishimura, and P. Bednarek.

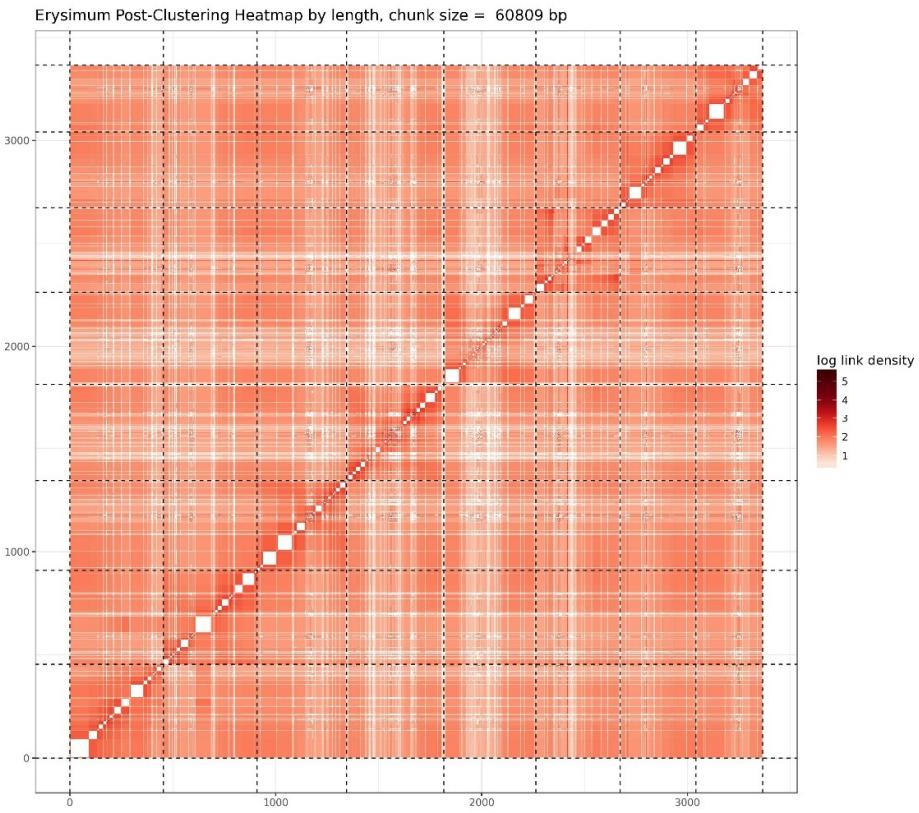
- 1299 2017. PYK10 myrosinase reveals a functional coordination between endoplasmic reticulum
1300 bodies and glucosinolates in *Arabidopsis thaliana*. *Plant Journal* **89**:204-220.
- 1301 Nielsen, J. K. 1978a. Host plant discrimination within Cruciferae: feeding responses of four leaf
1302 beetles (Coleoptera: Chrysomelidae) to glucosinolates, cucurbitacins, and cardenolides.
1303 *Entomologia Experimentalis Et Applicata* **24**:41-54.
- 1304 Nielsen, J. K. 1978b. Host plant selection of monophagous and oligophagous flea beetles feeding on
1305 crucifers. *Entomologia Experimentalis Et Applicata* **24**:562-569.
- 1306 Ou, S. J. and N. Jiang. 2018. LTR_retriever: a highly accurate and sensitive program for identification
1307 of long terminal repeat retrotransposons. *Plant Physiology* **176**:1410-1422.
- 1308 Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* **401**:877-884.
- 1309 Paradis, E. and K. Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary
1310 analyses in R. *Bioinformatics* **35**:526-528.
- 1311 Petschenka, G., S. Fandrich, N. Sander, V. Wagschal, M. Boppré, and S. Dobler. 2013. Stepwise
1312 evolution of resistance to toxic cardenolides via genetic substitutions in the Na⁺/K⁺-ATPase
1313 of milkweed butterflies (Lepidoptera: Danaini). *Evolution* **67**:2753-2761.
- 1314 Petschenka, G., C. S. Fei, J. J. Araya, S. Schröder, B. N. Timmermann, and A. A. Agrawal. 2018.
1315 Relative selectivity of plant cardenolides for Na⁺/K⁺-ATPases from the monarch butterfly and
1316 non-resistant insects. *Frontiers in Plant Science* **9**:1424.
- 1317 Polatschek, A. 2011. Revision der Gattung *Erysimum* (Cruciferae), Teil 2: Georgien, Armenien,
1318 Azerbaizjan, Türkei, Syrien, Libanon, Israel, Jordanien, Irak, Iran, Afghanistan. *Annalen des
1319 Naturhistorischen Museums in Wien – Serie B* **112**.
- 1320 Polatschek, A. and S. Snogerup. 2002. *Erysimum* in A. Strid and K. G. Tan, editors. *Flora Hellenica* 2.
1321 Koeltz Scientific Books, Koenigstein.
- 1322 Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. FastTree 2 - approximately maximum-likelihood
1323 trees for large alignments. *Plos One* **5**:e9490.
- 1324 Quinlan, A. R. and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic
1325 features. *Bioinformatics* **26**:841-842.
- 1326 Rask, L., E. Andreasson, B. Ekbom, S. Eriksson, B. Pontoppidan, and J. Meijer. 2000. Myrosinase:
1327 gene family evolution and herbivore defense in Brassicaceae. *Plant Molecular Biology* **42**:93-
1328 113.
- 1329 Rasemann, S. and A. A. Agrawal. 2011. Latitudinal patterns in plant defense: evolution of
1330 cardenolides, their toxicity and induction following herbivory. *Ecology Letters* **14**:476-483.
- 1331 Rasemann, S., M. D. Johnson, and A. A. Agrawal. 2009. Induced responses to herbivory and jasmonate
1332 in three milkweed species. *Journal of Chemical Ecology* **35**:1326-1334.
- 1333 Renwick, J. A. A., C. D. Radke, and K. Sachdevgupta. 1989. Chemical constituents of *Erysimum*
1334 *cheiranthoides* deterring oviposition by the cabbage butterfly, *Pieris rapae*. *Journal of
1335 Chemical Ecology* **15**:2161-2169.
- 1336 Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things).
1337 *Methods in Ecology and Evolution* **3**:217-223.
- 1338 Rhee, S. Y., P. Zhang, H. Foerster, and C. Tissier. 2006. AraCyc: Overview of an *Arabidopsis*
1339 metabolism database and its applications for plant research. Pages 141-154 in K. Saito, R. A.
1340 Dixon, and L. Willmitzer, editors. *Biotechnology in Agriculture and Forestry*. Springer,
1341 Berlin.
- 1342 Richards, L. A., L. A. Dyer, M. L. Forister, A. M. Smilanich, C. D. Dodson, M. D. Leonard, and C. S.
1343 Jeffrey. 2015. Phytochemical diversity drives plant-insect community diversity. *Proceedings
1344 of the National Academy of Sciences of the United States of America* **112**:10973-10978.
- 1345 Richards, L. A., A. E. Glassmire, K. M. Ochsenrider, A. M. Smilanich, C. D. Dodson, C. S. Jeffrey,
1346 and L. A. Dyer. 2016. Phytochemical diversity and synergistic effects on herbivores.
1347 *Phytochemistry Reviews* **15**:1153-1166.
- 1348 Romeo, J. T., J. A. Saunders, and P. Barbosa, editors. 1996. *Phytochemical diversity and redundancy
1349 in ecological interactions*. Plenum Press, New York.
- 1350 Sachdev-Gupta, K., C. D. Radke, J. A. A. Renwick, and M. B. Dimock. 1993. Cardenolides from
1351 *Erysimum cheiranthoides*: feeding deterrents to *Pieris rapae* larvae. *Journal of Chemical
1352 Ecology* **19**:1355-1369.

- 1353 Sachdev-Gupta, K., J. A. A. Renwick, and C. D. Radke. 1990. Isolation and identification of
1354 oviposition deterrents to cabbage butterfly, *Pieris rapae*, from *Erysimum cheiranthoides*.
1355 Journal of Chemical Ecology **16**:1059-1067.
- 1356 Salazar, D., J. Lokvam, I. Mesones, M. Vásquez Pilco, J. M. Ayarza Zuñiga, P. de Valpine, and P. V.
1357 A. Fine. 2018. Origin and maintenance of chemical diversity in a species-rich tropical tree
1358 lineage. Nature Ecology & Evolution **2**:983-990.
- 1359 Sedio, B. E., J. C. Rojas Echeverri, C. A. Boya P., and S. J. Wright. 2017. Sources of variation in
1360 foliar secondary chemistry in a tropical forest tree community. Ecology **98**:616-623.
- 1361 Sela, I., H. Ashkenazy, K. Katoh, and T. Pupko. 2015. GUIDANCE2: accurate detection of unreliable
1362 alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids
1363 Research **43**:W7-W14.
- 1364 Sherameti, I., Y. Venus, C. Drzewiecki, S. Tripathi, V. M. Dan, I. Nitz, A. Varma, F. M. Grundler,
1365 and R. Oelmuller. 2008. PYK10, a beta-glucosidase located in the endoplasmatic reticulum, is
1366 crucial for the beneficial interaction between *Arabidopsis thaliana* and the endophytic fungus
1367 *Piriformospora indica*. Plant Journal **54**:428-439.
- 1368 Shinoda, T., T. Nagao, M. Nakayama, H. Serizawa, M. Koshioka, H. Okabe, and A. Kawai. 2002.
1369 Identification of a triterpenoid saponin from a crucifer, *Barbarea vulgaris*, as a feeding
1370 deterrent to the diamondback moth, *Plutella xylostella*. Journal of Chemical Ecology **28**:587-
1371 599.
- 1372 Singh, B. and R. P. Rastogi. 1970. Cardenolides - glycosides and genins. Phytochemistry **9**:315-331.
- 1373 Smith, S. A. and C. W. Dunn. 2008. Phyutility: a phyloinformatics tool for trees, alignments and
1374 molecular data. Bioinformatics **24**:715-716.
- 1375 Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
1376 phylogenies. Bioinformatics **30**:1312-1313.
- 1377 Stanke, M., M. Diekhans, R. Baertsch, and D. Haussler. 2008. Using native and syntenically mapped
1378 cDNA alignments to improve de novo gene finding. Bioinformatics **24**:637-644.
- 1379 Steppuhn, A. and I. T. Baldwin. 2007. Resistance management in a native plant: nicotine prevents
1380 herbivores from compensating for plant protease inhibitors. Ecology Letters **10**:499-511.
- 1381 Suzuki, R. and H. Shimodaira. 2014. pvclust: hierarchical clustering with P-values via multiscale
1382 bootstrap resampling. Bioinformatics **22**:1540-1542.
- 1383 Taussky, H. H. and E. Shorr. 1953. A microcolorimetric method for the determination of inorganic
1384 phosphorus. Journal of Biological Chemistry **202**:675-685.
- 1385 Textor, S. and J. Gershenson. 2009. Herbivore induction of the glucosinolate-myrosinase defense
1386 system: major trends, biochemical bases and ecological significance. Phytochemistry Reviews
1387 **8**:149-170.
- 1388 Travers-Martin, N., F. Kuhlmann, and C. Müller. 2008. Revised determination of free and complexed
1389 myrosinase activities in plant extracts. Plant Physiology and Biochemistry **46**:506-516.
- 1390 Van Bel, M., T. Diels, E. Vancaester, L. Kreft, A. Botzki, Y. Van de Peer, F. Coppens, and K.
1391 Vandepoele. 2017. PLAZA 4.0: an integrative resource for functional, evolutionary and
1392 comparative plant genomics. Nucleic Acids Research **46**:D1190-D1196.
- 1393 Venturini, L., S. Caim, G. G. Kaithakottil, D. L. Mapleson, and D. Swarbreck. 2018. Leveraging
1394 multiple transcriptome assembly methods for improved gene structure annotation.
1395 GigaScience **7**:giy093.
- 1396 Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. D. Zeng,
1397 J. Wortman, S. K. Young, and A. M. Earl. 2014. Pilon: an integrated tool for comprehensive
1398 microbial variant detection and genome assembly improvement. Plos One **9**:e112963.
- 1399 Washburn, J. D., J. C. Schnable, G. C. Conant, T. P. Brutnell, Y. Shao, Y. Zhang, M. Ludwig, G.
1400 Davidse, and J. C. Pires. 2017. Genome-guided phylo-transcriptomic methods and the nuclear
1401 phylogenetic tree of the Paniceae grasses. Scientific Reports **7**:13528.
- 1402 Waterhouse, R. M., M. Seppey, F. A. Simao, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V.
1403 Kriventseva, and E. M. Zdobnov. 2018. BUSCO applications from quality assessments to
1404 gene prediction and phylogenomics. Molecular Biology and Evolution **35**:543-548.
- 1405 Weber, M. G. and A. A. Agrawal. 2014. Defense mutualisms enhance plant diversification.
1406 Proceedings of the National Academy of Sciences of the United States of America
1407 **111**:16442-16447.

- 1408 Wiklund, C. and C. Åhrberg. 1978. Host plants, nectar source plants, and habitat selection of males
1409 and females of *Anthocharis cardamines* (Lepidoptera). *Oikos* **31**:169-183.
- 1410 Winde, I. and U. Wittstock. 2011. Insect herbivore counteradaptations to the plant glucosinolate-
1411 myrosinase system. *Phytochemistry* **72**:1566-1575.
- 1412 Wink, M. 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic
1413 perspective. *Phytochemistry* **64**:3-19.
- 1414 Xu, Z. and H. Wang. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR
1415 retrotransposons. *Nucleic Acids Research* **35**:W265-W268.
- 1416 Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree
1417 reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**:153.
- 1418 Zhang, J. M., B. Pontoppidan, J. P. Xue, L. Rask, and J. Meijer. 2002. The third myrosinase gene
1419 TGG3 in *Arabidopsis thaliana* is a pseudogene specifically expressed in stamen and petal.
Physiologia Plantarum **115**:25-34.
- 1420 Züst, T., M. Mirzaei, and G. Jander. 2018. *Erysimum cheiranthoides*, an ecological research system
1421 with potential as a genetic and genomic model for studying cardiac glycoside biosynthesis.
Phytochemistry Reviews **17**:1239–1251.
- 1422 Züst, T., G. Petschenka, A. P. Hastings, and A. A. Agrawal. 2019. Toxicity of milkweed leaves and
1423 latex: chromatographic quantification versus biological activity of cardenolides in 16
1424 *Asclepias* species. *Journal of Chemical Ecology* **45**:50-60.
- 1425
- 1426
- 1427

1428 **Supplementary Figures and Tables**

1429

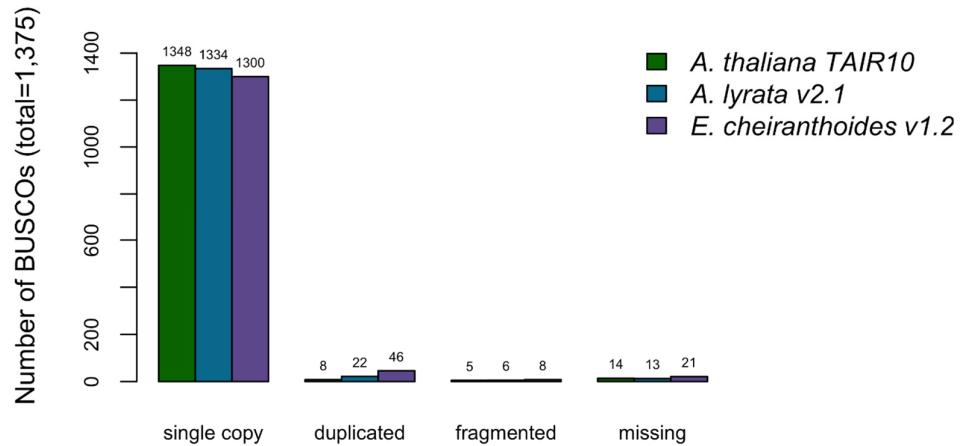


1430

1431 **Supplementary Figure S1.** Post-clustering heatmap showing the density of Hi-C interactions
1432 between scaffolds used in assembly. Intensity corresponds to the total number of reads per interaction.
1433 Dashed lines delimit the eight identified pseudomolecules.

1434

Genome BUSCO Results (pseudomolecules and contigs)



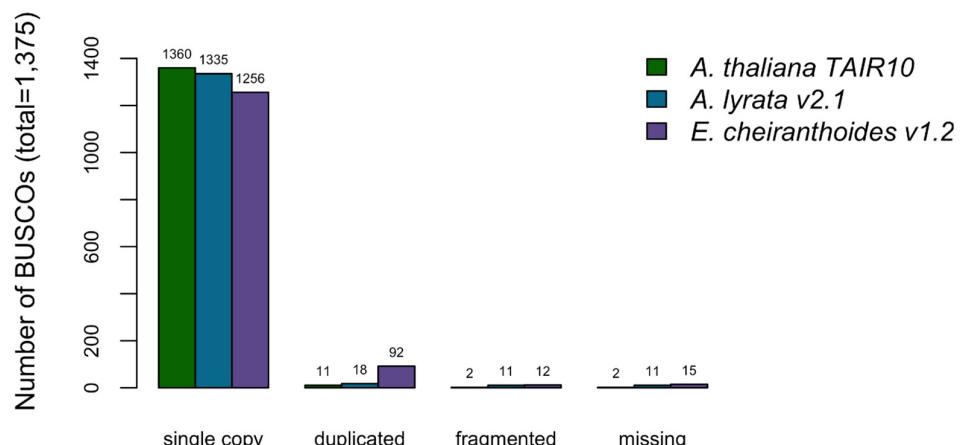
1435

1436 **Supplementary Figure S2.** BUSCO completeness assessment for the EC1.2 genome assembly.

1437 *A. thaliana* and *A. lyrata* results provided for comparison to EC1.2 genome assembly.

1438

Annotation BUSCO Results



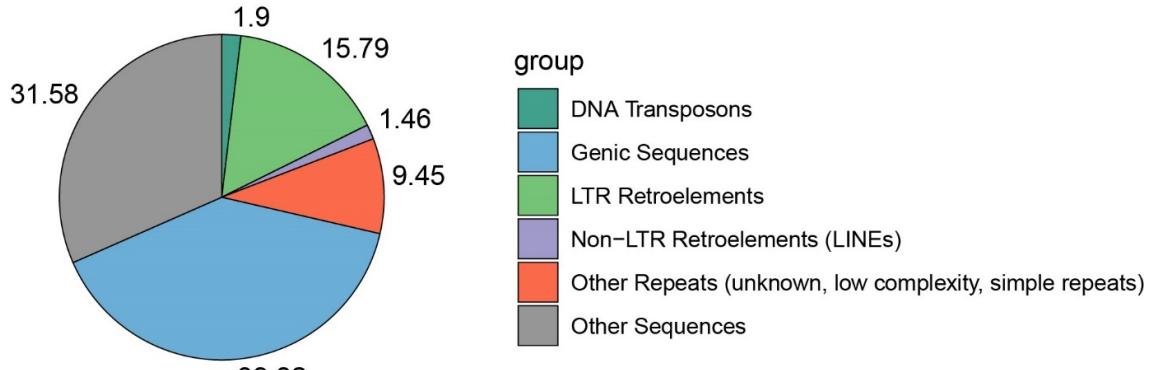
1439

1440 **Supplementary Figure S3.** BUSCO completeness assessment for the EC1.2 genome

1441 annotation. *A. thaliana* and *A. lyrata* results provided for comparison to the EC1.2 genome

1442 annotation.

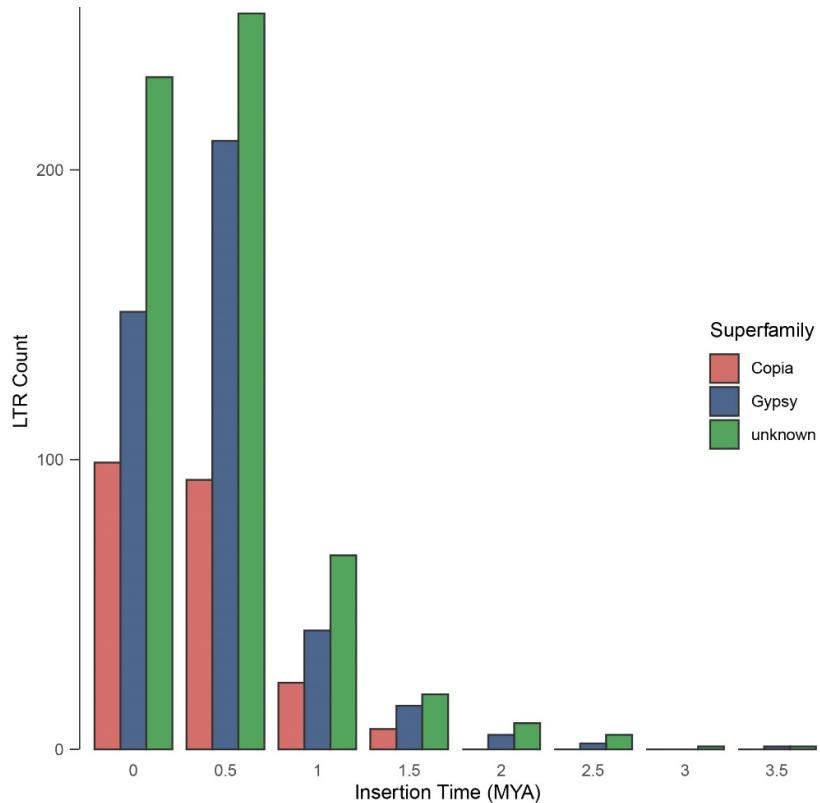
1443



1444

1445 **Supplementary Figure S4.** Proportional contributions of sequence classes to the genome
1446 sequence of *E. cheiranthoides*.

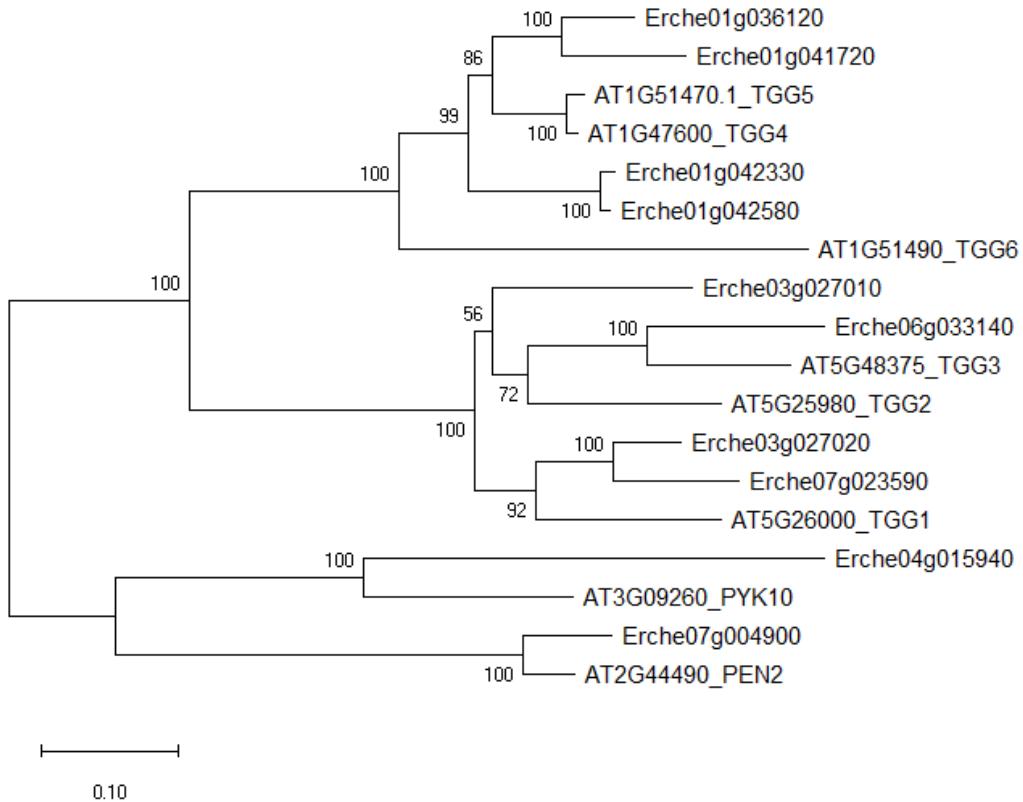
1447



1448

1449 **Supplementary Figure S5.** Age distributions of intact LTR-RTs identified by LTR_retriever
1450 in the genome of *E. cheiranthoides*.

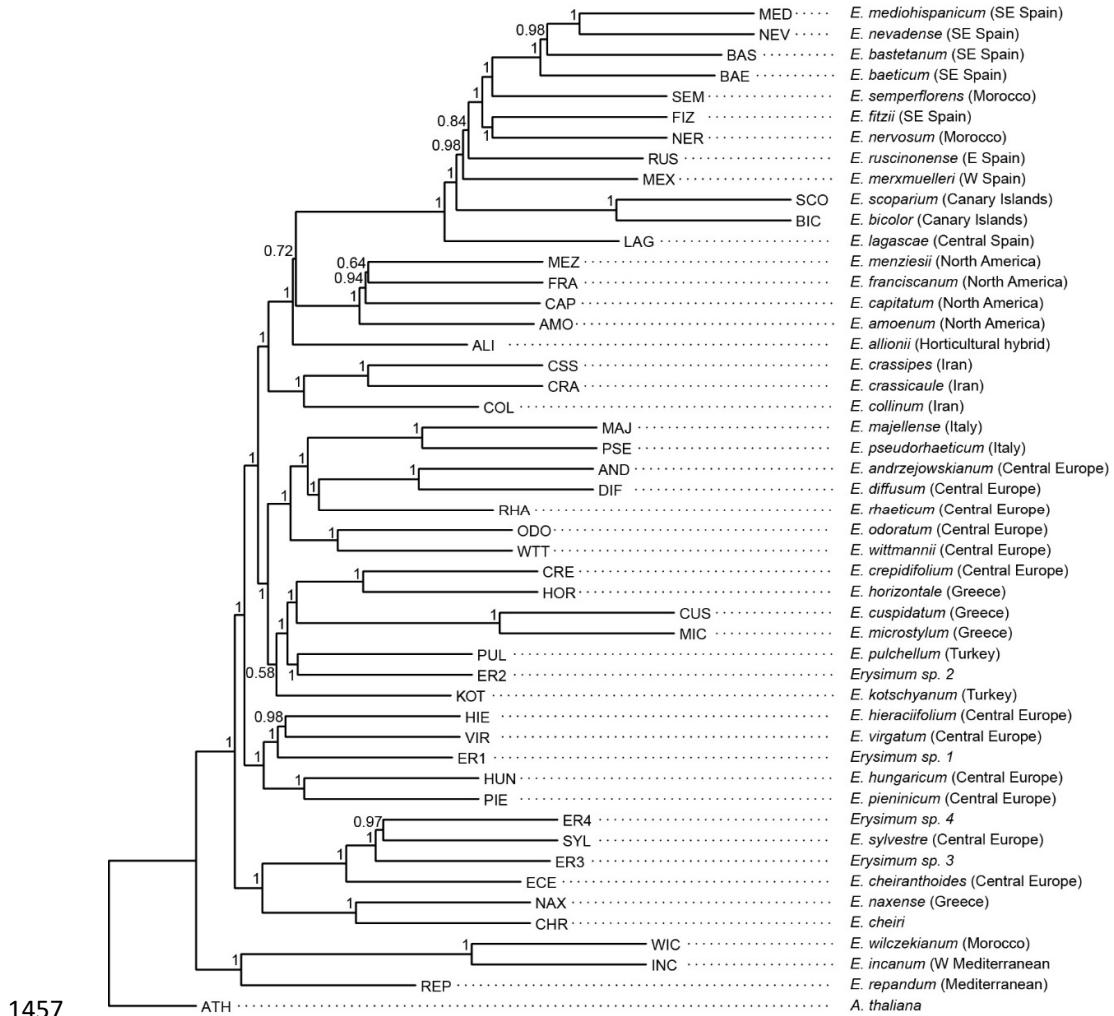
1451



1452

1453 **Supplementary Figure S6.** Phylogenetic analysis of myrosinase genes in *E. cheiranthoides* and *A.*
1454 *thaliana* using neighbor-joining methods. Nodes are labelled with bootstrap tests based on 1000
1455 replicates.

1456

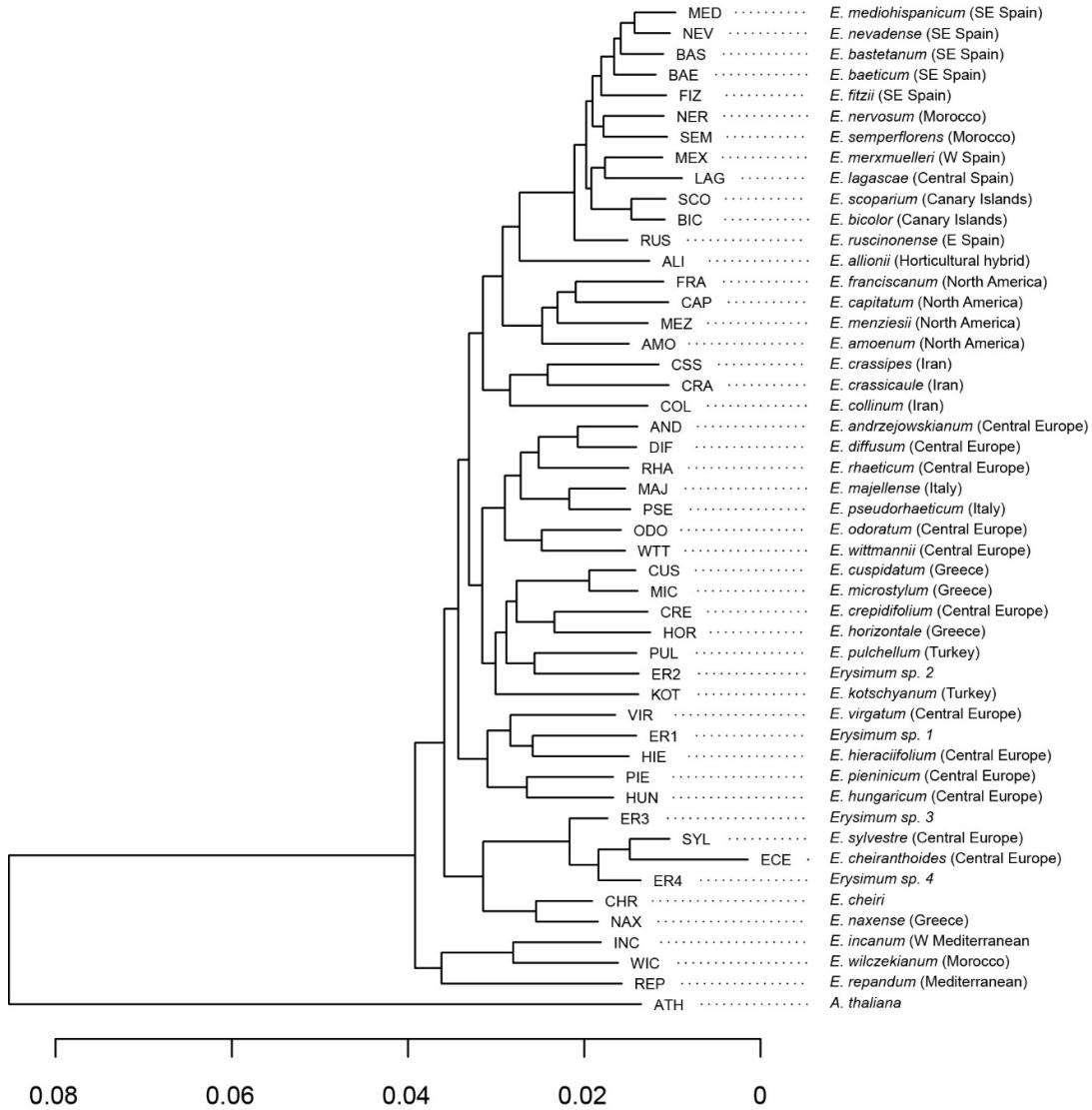


1457

1458 **Supplementary Figure S7.** Coalescent species tree inferred from 2,306 orthologous gene sequences.

1459 Nodes are labelled with local posterior probability, indicating level of support.

1460



1461

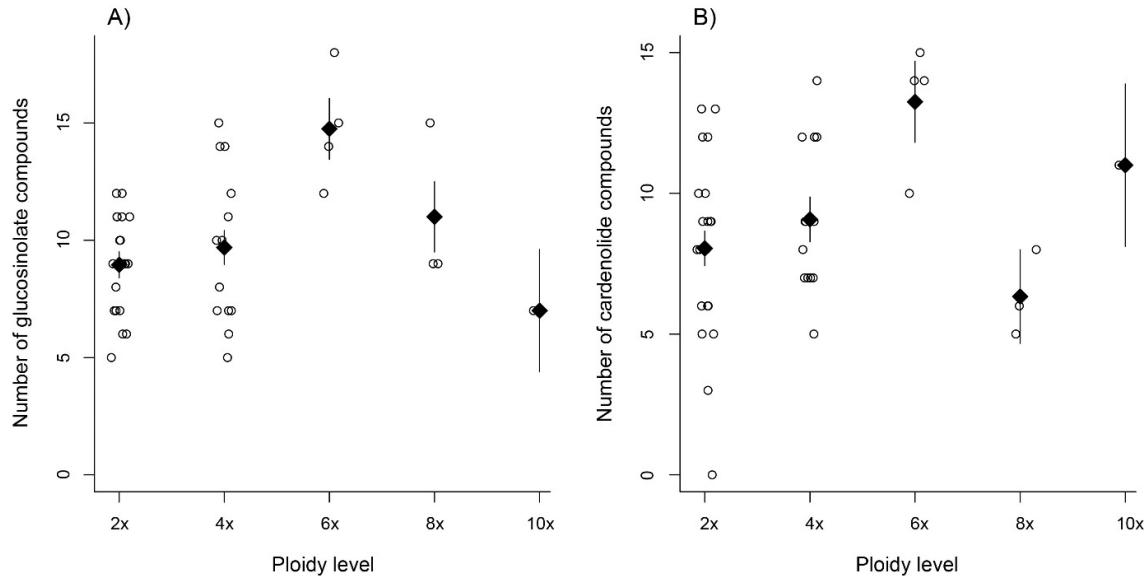
0.08 0.06 0.04 0.02 0

1462

Supplementary Figure S8. Concatenated species tree inferred from 2,306 orthologous gene sequences. Branch length corresponds to estimated number of substitutions per site.

1463

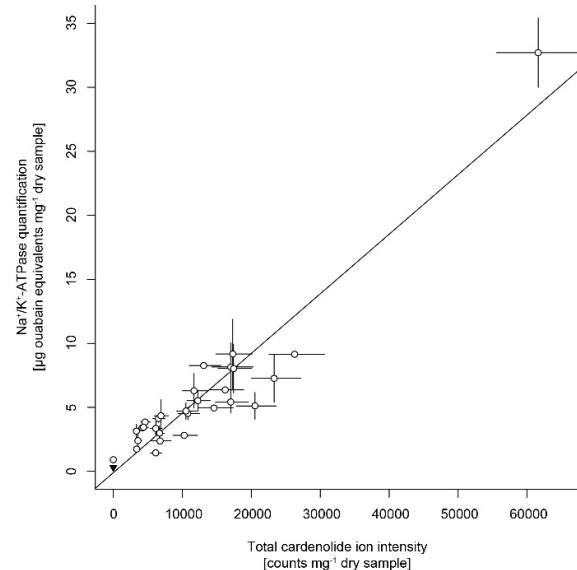
1464



1465

1466 **Supplementary Figure S9.** Effect of ploidy on compound diversity. Open circles correspond to
1467 species, with ploidy inferred from literature reports. Black triangles are mean values \pm 1 SE for each
1468 ploidy level. (A) Total number of glucosinolate compounds produced by each *Erysimum* species. (B)
1469 Number of cardenolide compounds which together constitute 80% of total cardenolide concentrations.
1470 As many cardenolide compounds were produced at very low concentrations, an effect of ploidy was
1471 obscured if the total number of cardenolide compounds was considered.

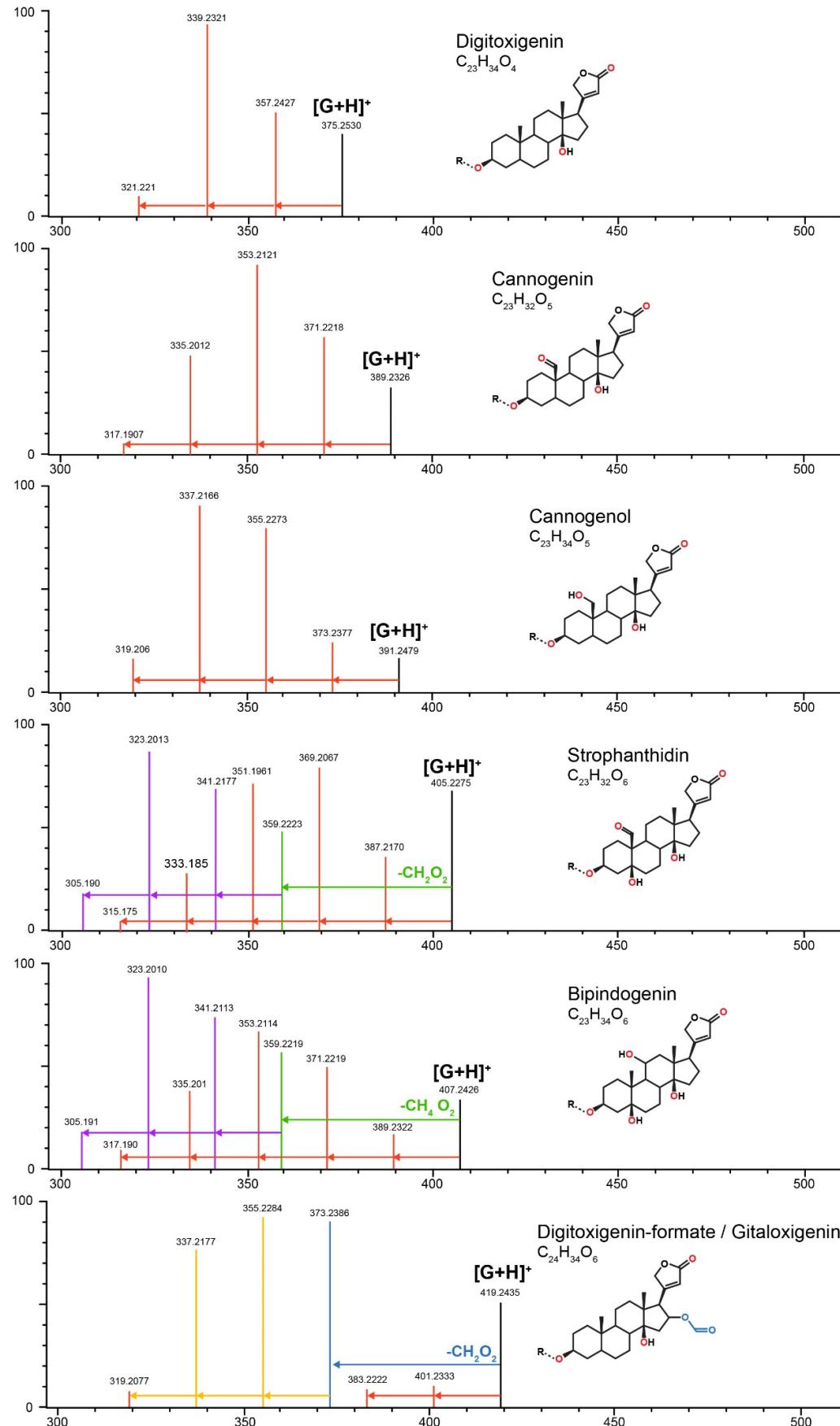
1472

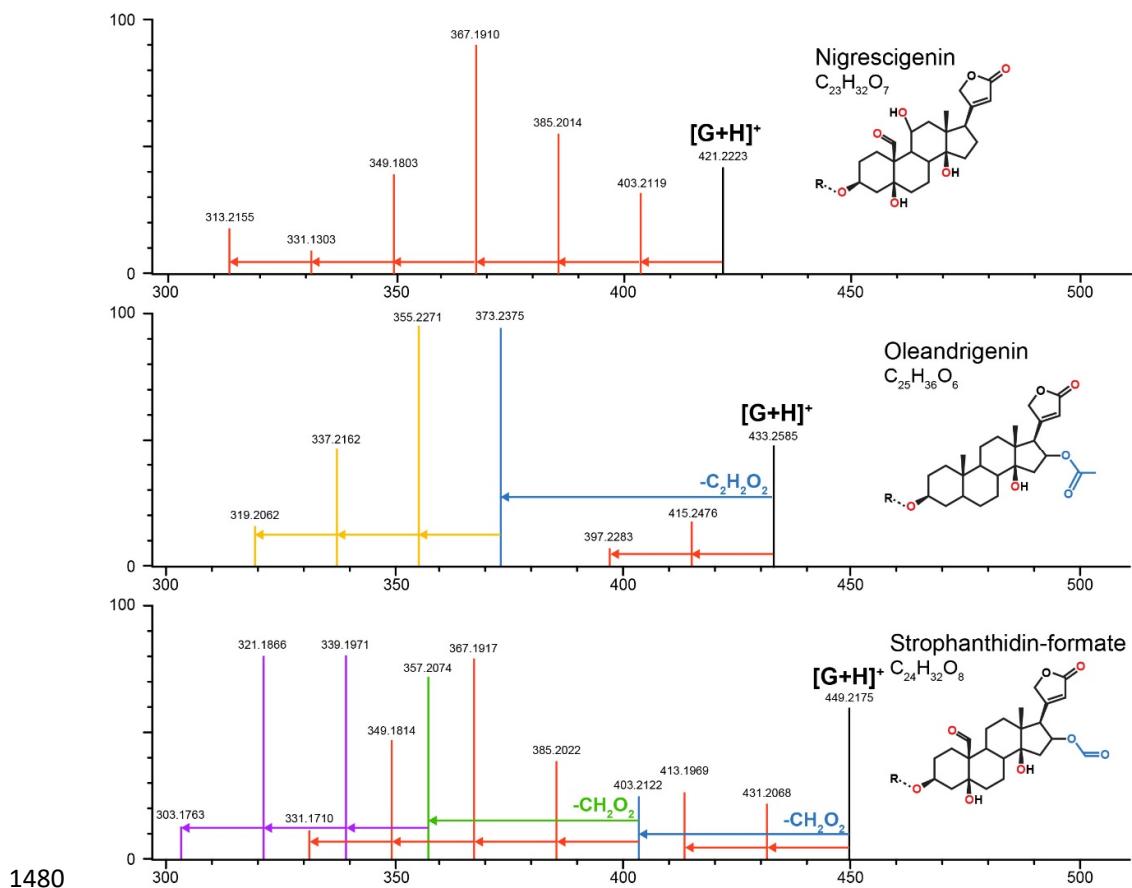


1473

1474 **Supplementary Figure S10.** Correlation between cardenolide concentrations approximated by total
1475 cardenolide ion intensity and by inhibition of animal Na⁺/K⁺-ATPase. Open circles are species means
1476 \pm 1 SE. The black triangle in the bottom left corner is the quantification of *Sinapis arvensis* tissue as a
1477 negative control. The solid line is the linear regression on species means.

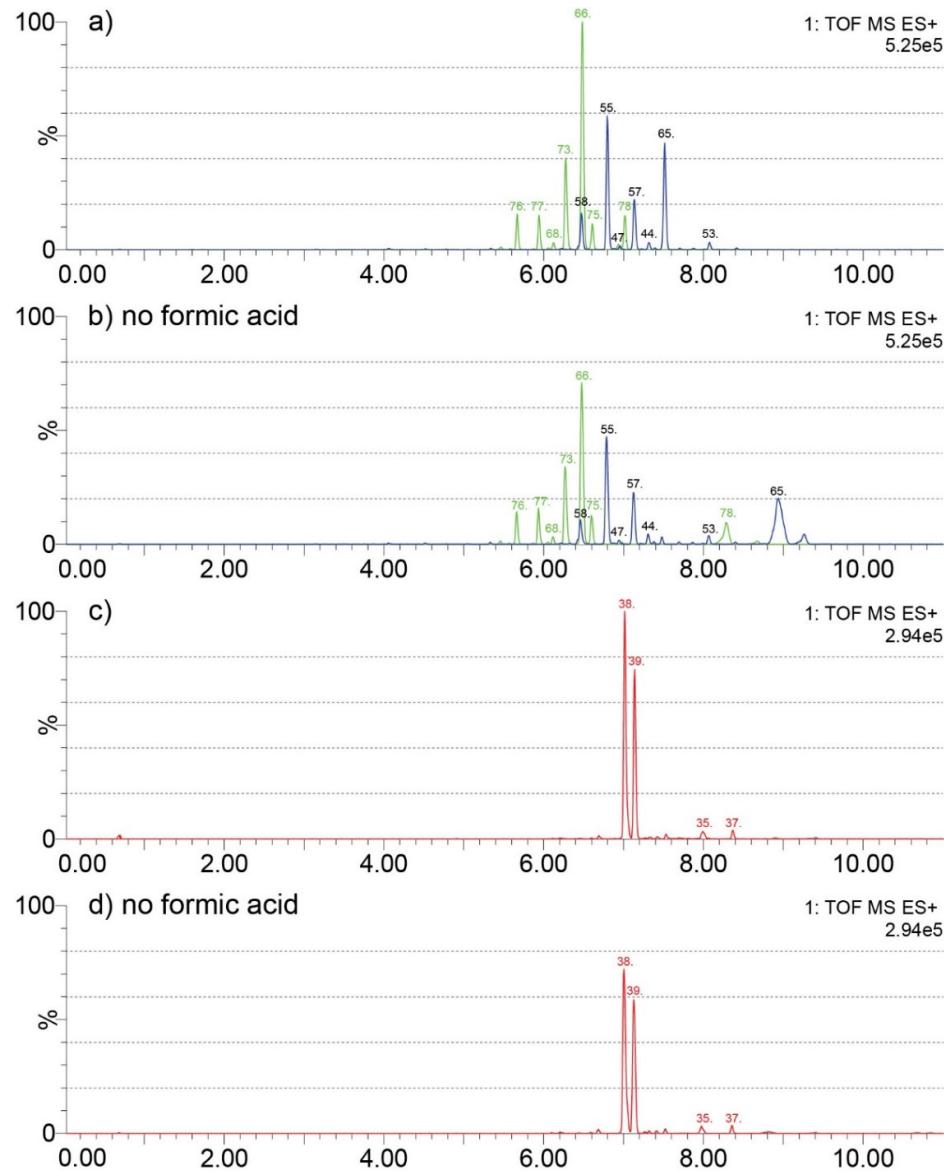
1478





1481 **Supplementary Figure S11.** Characteristic MS fragmentation patterns of cardenolide genins used for
1482 putative identification of compounds. Patterns are primarily generated by in-source fragmentation of
1483 compounds and are visible at standard MS conditions, but fragments are more abundant under MS^E
1484 (high energy fragmentation) conditions. Height of vertical lines indicates relative ion intensities (in
1485 MS^E) for representative compounds quantified in *Erysimum* spp. The mass and intensity of the intact
1486 genin ($[G+H]^+$) in each panel is indicated by a vertical black line. Genin fragments are colored to
1487 highlight fragmentation series: red vertical lines linked by arrows represent serial losses of water
1488 molecules (-18.01 m/z per molecule), corresponding to the number of exposed oxygen groups of that
1489 molecule (red symbols). Acetyl and formate groups are lost as intact units (blue lines/arrows and
1490 corresponding symbols), after which further loss of water molecules occurs (orange lines/arrows). In
1491 strophantidin and bipindogenin, additional larger fragments are lost, perhaps by reconfiguration of
1492 the genin molecule (green lines/arrows), and again, further loss of water molecules occurs (purple
1493 lines/arrows). Fragmentation patterns for strophantidin and digitoxigenin were confirmed by
1494 commercial standards. For remaining genins, likely identifications and structures were inferred from
1495 literature reports.

1496



1510 **Supplementary Table S1.** Origin of species and seed material, and year of original collection where available. Four species from mislabeled seed stocks are
 1511 referred to as accessions ER1-4, with false species name provided in quotation marks. Ploidy levels can be variable within species and are based on
 1512 measurements of the sampled populations where available (highlighted by *), or otherwise inferred from literature reports. Leaf material for each species or
 1513 accession was collected in one of three experiments, and RNA was extracted from pooled tissue of 2-5 individual plants that were sampled at 1-2 time points
 1514 (TP).

Species code	Species	Region/country of origin	Seed origin	Original collection	Ploidy	Source experiment	Pooled individuals
ALI	<i>E. allionii</i> (<i>E. x. marshallii</i> ¹)	n/a	Botanical Garden Kiel, Germany	1985	n/a	2017-2	5 / 2TP
AMO	<i>E. amoenum</i>	Colorado, USA	Alplains, CO, USA	2014	4x	2017-1	5 / 1 TP
AND	<i>E. andrzejowskianum</i>	Austria	Collected	2009	10x	2017-1	5 / 2TP
BAE	<i>E. baeticum</i>	Spain	Collected	2006	8x*	2016	5 / 1TP
BAS	<i>E. bastetanum</i>	Spain	Collected	2007	6x*	2017-2	5 / 2TP
BIC	<i>E. bicolor</i>	Canary Islands, Spain	Botanical Garden Konstanz, Germany	unknown	4x	2017-2	5 / 2TP
CAP	<i>E. capitatum</i>	USA	B&T World Seeds, France	unknown	4x	2017-1	4 / 2TP
CHR	<i>E. cheiri</i>	Netherlands	Collected	2016	2x	2017-2	5 / 2TP
COL	<i>E. collinum</i> ²	Iran	Botanical Garden Madrid, Spain	1993	2x	2017-1	3 / 1TP
CRA	<i>E. crassicaule</i>	Iran	Botanical Garden Madrid, Spain	unknown	2x	2017-2	2 / 2TP
CRE	<i>E. crepidifolium</i>	Czechia	Botanical Garden Berlin-Dahlem, Germany	1996	2x	2017-1	5 / 2TP
CSS	<i>E. crassipes</i>	Iran	Botanical Garden Madrid, Spain	1993	2x	2017-2	5 / 2TP
CUS	<i>E. cuspidatum</i>	Greece	Botanical Garden Berlin-Dahlem, Germany	1980	2x	2017-2	5 / 2TP
DIF	<i>E. diffusum</i>	Czechia	Botanical Garden Plzen, Czechia	2013	4x	2017-2	5 / 2TP
ECE	<i>E. cheiranthoides</i>	Germany	Collected	2015	2x*	2017-1	5 / 1TP
ER1	<i>Erysimum</i> sp. 1 ' <i>E. crepidifolium</i> '	n/a	Botanical Garden Berlin-Dahlem, Germany	unknown	n/a	2017-1	4 / 1TP
ER2	<i>Erysimum</i> sp. 2 ' <i>E. rhaeticum</i> '	n/a	Botanical Garden Nantes, France	unknown	n/a	2017-2	5 / 2TP
ER3	<i>Erysimum</i> sp. 3 ' <i>E. asperum</i> '	n/a	B&T World Seeds, France	unknown	n/a	2017-1	5 / 2TP
ER4	<i>Erysimum</i> sp. 4 ' <i>E. suffrutescens</i> '	n/a	B&T World Seeds, France	unknown	n/a	2017-2	5 / 2TP

FIZ	<i>E. fitzii</i>	Spain	Collected	2008	2x*	2017-2	2 / 2TP
FRA	<i>E. franciscanum</i>	California, USA	B&T World Seeds, France	unknown	4x	2017-1	5 / 2TP
HIE	<i>E. hieraciifolium</i>	Romania	Botanical Garden Jibou, Romania	unknown	4x	2017-2	5 / 2TP
HOR	<i>E. horizontale</i>	Greece	Botanical Garden Berlin-Dahlem, Germany	unknown	4x	2016	4 / 1TP
HUN	<i>E. hungaricum</i>	Romania	Botanical Garden Jibou, Romania	unknown	6x	2017-2	5 / 2TP
INC	<i>E. incanum</i>	Spain	Botanical Garden Madrid, Spain	unknown	2x	2017-2	4 / 2TP
KOT	<i>E. kotschyanum</i>	Turkey	Botanical Garden Tübingen, Germany	unknown	2x	2017-2	5 / 2TP
LAG	<i>E. lagascae</i>	Spain	Botanical Garden Madrid, Spain	unknown	2x*	2017-1	5 / 2TP
MAJ	<i>E. majellense</i>	Italy		unknown	4x	2016	5 / 1TP
MED	<i>E. mediohispanicum</i>	Spain	Collected	2007	2x*	2017-2	5 / 2TP
MEX	<i>E. merxmulleri</i>	Spain	Collected	2007	2x*	2017-2	5 / 2TP
MEZ	<i>E. menziesii</i>	California, USA	B&T World Seeds, France	unknown	4x	2017-1	5 / 2TP
MIC	<i>E. microstylum</i>	Greece	Botanical Garden Berlin-Dahlem, Germany	1980	2x	2017-1	5 / 2TP
NAX	<i>E. naxense</i>	Greece	Balkan Botanic Garden of Korussia, Greece	unknown	2x	2016	5 / 1TP
NER	<i>E. nervosum</i>	Morocco	Collected	2008	4x*	2017-1	5 / 1TP
NEV	<i>E. nevadense</i>	Spain	Collected	2007	2x*	2017-2	3 / 2TP
ODO	<i>E. odoratum</i>	Austria	Botanical Garden Bern, Switzerland	2015	4x	2017-1	5 / 2TP
PIE	<i>E. pieninicum</i>	Romania	Botanical Garden Jibou, Romania	unknown	6x	2017-2	5 / 2TP
PSE	<i>E. pseudorhaeticum</i>	Italy	Botanical Garden Nantes, France	unknown	2x	2017-2	5 / 2TP
PUL	<i>E. pulchellum</i>	Turkey	Botanical Garden Koursk, Russia	unknown	8x	2017-2	5 / 2TP
REP	<i>E. repandum</i>	Spain	Botanical Garden Madrid, Spain	unknown	2x	2017-2	5 / 2TP
RHA	<i>E. rhaeticum</i>	Switzerland	Collected	2016	8x	2017-1	5 / 2TP
RUS	<i>E. ruscinonense</i>	Spain	Collected	2007	2x	2016	5 / 1TP
SCO	<i>E. scoparium</i>	Canary Islands, Spain	B&T World Seeds, France	unknown	4x	2017-1	4 / 1TP
SEM	<i>E. semperflorens</i>	Morocco	Collected	2008	2x	2017-1	5 / 2TP
SYL	<i>E. sylvestre</i>	Slovenia	B&T World Seeds, France	unknown	2x	2017-1	5 / 2TP
VIR	<i>E. virgatum</i>	Switzerland	Botanical Garden St Gallen, Switzerland	1962	6x	2017-2	5 / 2TP
WIC	<i>E. wilczekianum</i>	Morocco	Collected	2008	4x*	2017-2	5 / 2TP
WTT	<i>E. wittmannii</i>	Slovenia	Botanical Garden Berlin-Dahlem, Germany	2010	2x	2017-1	5 / 2TP

1515 ¹horticultural hybrid; ² previously known as *E. passgalense*

1516

1517 **Supplementary Table S2.** Repetitive sequences and transposable elements in the *E. cheiranthoides*
1518 genome.

Classification	Order	Superfamily	No. of TEs	Coverage (Mb)	Fraction of genome (%)
Class I	LTR	<i>Copia</i>	3508	2.643	1.492
	LTR	<i>Gypsy</i>	13748	13.371	7.547
	LTR	Unknown/Other	24145	11.96	6.75
	LINE	<i>LI</i>	6105	2.585	1.459
Class II	TIR	<i>CMC-EnSpm</i>	1222	0.543	0.307
	TIR	<i>hAT</i>	1225	0.42	0.237
	Helitron	<i>Helitron</i>	1292	0.763	0.431
	TIR	<i>Mutator</i>	2880	1.406	0.794
	TIR	PIF-Harbinger	569	0.227	0.128
		Unknown	27	0.002	0.001
Other Simple					
Repeats			641	0.124	0.07
Other Unknown					
Repeats			45324	16.618	9.379
Total					28.59

1519

1520 **Supplementary Table S3.** Transcriptome assembly metrics, including number of sequences, N50 values, and recovered BUSCO gene number. Additionally,
 1521 transcript lengths were divided by the length of the top BLAST match to the *E. cheiranthoides* v1.1 gene model (EC1.1) to determine fragmentation of the
 1522 transcriptome assemblies (tophit average trinity_len/EC_len). RNA sequences from each of the 48 *Erysimum* species were mapped to the *E. cheiranthoides*
 1523 genome, and results are reported as the number of *E. cheiranthoides* genes represented and the mapping percentage.

Species	Sequence count	N50 [bp]	BUSCO genes				Tophit average trinity_len/EC_len	No. of EC1.1 genes represented	Mapping to EC1.1 (%)
			Complete (percent total)	Complete single-copy	Complete duplicated	Fragmented	Missing		
ALI	207422	595	1010 (70.1%)	276	734	250	180	0.74	19150 56.6
AMO	217430	888	903 (62.7%)	419	484	323	214	0.97	21450 59.4
AND	165687	1172	1043 (72.4%)	445	598	252	145	1.09	20995 58.5
AUC	164506	1367	1163 (80.8%)	439	724	183	94	1.21	20711 58.6
BAE	180487	1127	1041 (72.3%)	479	562	253	146	1.05	21187 59.1
BAS	135035	1420	1183 (82.2%)	442	741	147	110	0.91	20279 57.8
BIC	99234	1868	1314 (91.3%)	519	795	67	59	1.13	19925 57.8
CAP	260998	842	887 (61.6%)	406	481	366	187	0.97	22113 57.4
CHR	93525	1963	1350 (93.8%)	565	785	39	51	1.22	20213 57.3
CRA	143666	1582	1266 (87.9%)	417	849	113	61	1.01	20924 54.2
CRE	102472	1768	1321 (91.7%)	589	732	55	64	1.28	19686 59
CSS	118038	1588	1241 (86.2%)	469	773	90	108	1.01	20457 56.4
CUS	122476	1644	1263 (87.7%)	457	806	110	67	1.02	20077 58.1
DIF	139288	770	1189 (82.6%)	392	797	154	97	0.89	20472 56.7
ECE	81984	2139	1341 (93.1%)	658	683	34	65	1.41	19519 94.8
ER1	223508	900	951 (66.0%)	378	573	287	202	1.01	21335 66.5
ER2	258578	715	785 (54.5%)	307	478	379	276	0.57	23080 55.1
ER3	123653	1886	1338 (92.9%)	452	886	40	62	1.35	20130 77.5
ER4	89871	1844	1328 (92.2%)	469	859	49	63	1.13	23371 78.2
FIZ	94064	1956	1320 (91.7%)	551	769	50	70	1.21	21130 59.1
FRA	220291	1004	992 (68.9%)	384	608	284	164	1.04	21412 59.9
HIE	218565	955	960 (66.7%)	296	664	286	194	0.71	20878 66
HOR	109045	1784	1306 (90.7%)	501	805	77	57	1.29	19758 61.9
HUN	228679	881	903 (62.7%)	291	612	313	224	0.67	21558 61.4
INC	90295	2002	1337 (92.8%)	481	856	40	63	1.23	20671 61.4

KOT	104476	1755	1322 (91.8%)	531	791	56	62	1.07	21558	56.8
LAG	155771	1422	1196 (83.1%)	419	777	153	91	1.21	20777	59.1
MAJ	174858	1027	958 (66.5%)	462	496	286	196	1.01	20830	57.9
MED	106039	1869	1301 (90.3%)	506	795	69	70	1.41	20907	57.9
MEX	102500	1115	1284 (89.2%)	501	783	74	82	1.13	21693	57.4
MEZ	187227	1160	1074 (74.6%)	420	654	203	163	1.1	20811	56.9
MIC	115738	1696	1271 (88.3%)	513	758	94	75	1.29	19993	61.2
NAX	104553	2016	1351 (93.8%)	549	802	32	57	1.37	19589	60.5
NER	161135	1485	1228 (85.3%)	471	757	133	79	1.23	20833	59.4
NEV	115096	1710	1282 (89.0%)	450	832	84	74	1.08	19160	59.6
ODO	193451	1087	975 (67.7%)	421	554	289	176	1.07	21108	57.9
PIE	238425	878	887 (61.6%)	275	612	327	226	0.69	22492	59.5
PSE	129384	1513	1230 (85.4%)	428	802	134	76	0.95	22784	57.4
PUL	251705	741	773 (53.7%)	305	468	391	276	0.58	20633	54.6
REP	64315	2160	1340 (93.1%)	727	613	34	66	1.33	22932	57.3
RHA	220584	876	887 (61.6%)	440	447	335	218	0.97	21715	57.4
RUS	154276	1328	1125 (78.1%)	470	655	188	127	1.13	20616	60.7
SCO	92945	1897	1331 (92.4%)	625	706	48	61	1.34	19314	61.2
SEM	91502	1875	1317 (91.5%)	608	709	51	72	1.34	19479	59.9
SYL	113269	1918	1331 (92.4%)	449	882	44	65	1.37	20013	77.8
VIR	212293	574	990 (68.8%)	305	685	278	172	0.72	22901	66.1
WIC	91007	2039	1339 (93.0%)	586	753	36	65	1.28	23344	56.9
WTT	171465	1768	1284 (89.2%)	395	889	97	59	1.37	21116	57.8

1525 **Supplementary Table S4.** List of glucosinolate compounds, determined by exact mass, fragmentation patterns, and retention time. Asterisks (*) indicate
 1526 compounds confirmed by commercial standards.

#	Systematic short name	Systematic name (-glucosinolate)	Common name	Class	Molecular formula	Retention time	[M-H] ⁻	MS fragments
1	3MTP	3-methylthiopropyl	Glucoibererin	Aliphatic	C ₁₁ H ₂₁ NO ₉ S ₃	3.16	406.0300	259.013, 241.001, 195.033, 96.960
2	3MSI	3-methylsulfinylpropyl*	Glucoiberin	Aliphatic	C ₁₁ H ₂₁ NO ₁₀ S ₃	1.86	422.0249	358.0276 ¹ , 259.014, 195.034, 96.961
3	2OH	2-hydroxypropyl	-	Aliphatic	C ₁₀ H ₁₉ NO ₁₀ S ₂	1.76	376.0372	259.013, 195.034, 96.960
4	2PRO	2-propenyl*	Sinigrin	Aliphatic	C ₁₀ H ₁₇ NO ₉ S ₂	2.33	358.0266	259.013, 241.003, 195.033, 96.960
5	3MSO	3-methylsulfonylpropyl*	Glucocheirolin	Aliphatic	C ₁₁ H ₂₁ NO ₁₁ S ₃	2.06	438.0198	259.013, 241.002, 195.033, 96.960
6	3MSO'	3-methylsulfonylpropyl isomer		Aliphatic	C ₁₁ H ₂₁ NO ₁₁ S ₃	1.75	438.0202	96.961
7	1MP	1-methylpropyl	-	Aliphatic	C ₁₁ H ₂₁ NO ₉ S ₂	3.13	374.0579	96.961
8	2MP	2-methylpropyl	-	Aliphatic	C ₁₁ H ₂₁ NO ₉ S ₂	3.21	374.0579	96.960
9	4MTB	4-methylthiobutyl	Glucoerucin	Aliphatic	C ₁₂ H ₂₃ NO ₉ S ₃	3.69	420.0456	96.960
10	4MSI	4-methylsulfinylbutyl	Glucoraphanin	Aliphatic	C ₁₂ H ₂₃ NO ₁₀ S ₃	2.03	436.0405	372.043 ¹ , 259.013, 195.034, 96.961
11	4BUT	3-butenyl	Gluconapin	Aliphatic	C ₁₁ H ₁₉ NO ₉ S ₂	2.90	372.0423	96.960
12	1MB	1-methylbutyl	-	Aliphatic	C ₁₂ H ₂₃ NO ₉ S ₂	3.96	388.0736	96.960
13	4MSO	4-methylsulfonylbutyl	Glucoerysolin	Aliphatic	C ₁₂ H ₂₃ NO ₁₁ S ₃	2.24	452.0355	96.961
14	OH4MSO	3-hydroxy-4-methylsulfonylbutyl	-	Aliphatic	C ₁₂ H ₂₃ NO ₁₂ S ₃	1.99	468.0304	259.013, 195.033, 96.961
15	5MTP	5-methylthiopentyl	Glucoberteroин	Aliphatic	C ₁₃ H ₂₅ NO ₉ S ₃	4.31	434.0613	96.961
16	5MSI	5-methylsulfinylpentyl	Glucoalyssin	Aliphatic	C ₁₃ H ₂₅ NO ₁₀ S ₃	2.32	450.0562	386.059 ¹ , 259.013, 96.961
17	5MSO	5-methylsulfonylpentyl	-	Aliphatic	C ₁₃ H ₂₅ NO ₁₁ S ₃	2.55	466.0511	259.013, 241.004, 195.033, 96.961
18	OH5MSO	3-hydroxy-5-methylsulfonylpentyl	-	Aliphatic	C ₁₃ H ₂₅ NO ₁₂ S ₃	2.11	482.0460	259.013, 195.033, 96.961
19	6MSI	6-methylsulfinylhexyl	Glucohesperin	Aliphatic	C ₁₄ H ₂₇ NO ₁₀ S ₃	2.69	464.0719	400.075 ¹ , 259.013, 96.960
20	6MSO	6-methylsulfonylhexyl	-	Aliphatic	C ₁₄ H ₂₇ NO ₁₁ S ₃	2.99	480.0668	259.013, 195.033, 96.960
21	OH6MSO	3-hydroxy-6-methylsulfonylhexyl	-	Aliphatic	C ₁₄ H ₂₇ NO ₁₂ S ₃	2.31	496.0617	259.013, 96.960
22	3MECOP	3-methoxycarbonylpropyl	Glucoerypestrin	Carboxylic	C ₁₂ H ₂₁ NO ₁₁ S ₂	2.75	418.0477	259.013, 195.033, 96.961
23	I3M	indol-3-ylmethyl	Glucobrassicin	Indole	C ₁₆ H ₂₀ N ₂ O ₉ S ₂	4.00	447.0532	96.960

24	4OHI3M	4-hydroxy-indol-3-ylmethyl	4-Hydroxyglucobrassicin	Indole	C ₁₆ H ₂₀ N ₂ O ₁₀ S ₂	3.37	463.0481	259.013, 96.961
25	MEI3M	4-methoxy-indol-3-ylmethyl	4-Methoxyglucobrassicin	Indole	C ₁₇ H ₂₂ N ₂ O ₁₀ S ₂	4.39	477.0637	96.961

1527 ¹[M-CH₄OS-H]⁻

1528

1529

1530

1531

Supplementary Table S5. List of candidate cardenolide compounds, determined by exact mass and fragmentation patterns. Asterisks (*) indicate compounds confirmed by commercial standards. Compounds #65 and #78 were excluded due to potential artefact formation with formic acid (see Figure S12).

#	Compound name	Genin	Molecular formula	Retention time [min]	[M+H] ⁺	[M+Na] ⁺	1 st sugar	2 nd sugar	Additional fragments
1	Digitoxigenin*	Digitoxigenin	C ₂₃ H ₃₄ O ₄	9.96	375.254	397.234	-	-	
2		Digitoxigenin	C ₂₃ H ₃₄ O ₄	10.06	375.254	397.234	-		
3		Digitoxigenin	C ₂₉ H ₄₄ O ₈	8.81	521.312	543.293	Deoxyhexose	-	
4		Digitoxigenin	C ₂₉ H ₄₄ O ₈	9.01	521.312	543.293	Deoxyhexose	-	
5		Digitoxigenin	C ₂₉ H ₄₄ O ₈	9.06	521.312	543.293	Deoxyhexose	-	
6		Digitoxigenin	C ₂₉ H ₄₄ O ₈	9.12	521.312	543.293	Deoxyhexose	-	
7		Digitoxigenin	C ₂₉ H ₄₄ O ₈	9.18	521.312	543.293	Deoxyhexose	-	
8		Digitoxigenin	C ₂₉ H ₄₄ O ₈	9.35	521.312	543.293	Deoxyhexose	-	
9		Digitoxigenin	C ₂₉ H ₄₄ O ₈		521.312	543.293	Deoxyhexose	-	
10		Digitoxigenin	C ₂₉ H ₄₄ O ₉	7.18	537.305	559.288	Glucose	-	
11		Digitoxigenin	C ₂₉ H ₄₄ O ₉	7.33	537.305	559.288	Glucose	-	
12		Digitoxigenin	C ₂₉ H ₄₄ O ₉	7.76	537.305	559.288	Glucose	-	
13		Digitoxigenin	C ₃₅ H ₅₄ O ₁₃	7.84	683.369	705.346	Deoxyhexose	Glucose	521.311
14		Digitoxigenin	C ₃₅ H ₅₄ O ₁₃	8.04	683.369	705.346	Deoxyhexose	Glucose	521.311
15		Digitoxigenin	C ₃₅ H ₅₄ O ₁₃	8.1	683.369	705.346	Deoxyhexose	Glucose	521.311
16		Digitoxigenin	C ₃₅ H ₅₄ O ₁₃	8.16	683.369	705.346	Deoxyhexose	Glucose	521.311
17		Digitoxigenin	C ₃₅ H ₅₄ O ₁₃	8.46	683.369	705.346	Deoxyhexose	Glucose	521.311
18		Digitoxigenin	C ₃₇ H ₅₆ O ₁₄	8.68	725.373	747.356	Acetyl-Deoxyhexose	Glucose	563.322
19		Digitoxigenin	C ₃₇ H ₅₆ O ₁₄	9.44	725.373	747.356	Acetyl-Deoxyhexose	Glucose	563.322
20	Cannogenol	Cannogenol	C ₂₃ H ₃₅ O ₅	8.47	391.249	413.231	-	-	
21		Cannogenol	C ₂₉ H ₄₄ O ₉	7.02	537.304	559.288	Deoxyhexose	-	
22		Cannogenol	C ₂₉ H ₄₄ O ₉	7.43	537.304	559.288	Deoxyhexose	-	
23		Cannogenol	C ₂₉ H ₄₄ O ₉	7.7	537.304	559.288	Deoxyhexose	-	
24		Cannogenol	C ₂₉ H ₄₄ O ₉	7.78	537.304	559.288	Deoxyhexose	-	
25		Cannogenol	C ₂₉ H ₄₄ O ₉	7.98	537.304	559.288	Deoxyhexose	-	
26		Cannogenol	C ₂₉ H ₄₄ O ₉	8.04	537.304	559.288	Deoxyhexose	-	
27		Cannogenol	C ₃₁ H ₄₆ O ₁₀	9.02	579.319	601.3	Acetyl-Deoxyhexose	-	
28		Cannogenol	C ₃₅ H ₅₄ O ₁₄	6.82	699.359	721.341	Deoxyhexose	Glucose	537.307
29		Cannogenol	C ₃₅ H ₅₄ O ₁₄	7.01	699.359	721.341	Deoxyhexose	Glucose	537.307

30		Cannogenol	C ₃₇ H ₅₆ O ₁₅	7.45	741.374	763.352	Acetyl-Deoxyhexose	Glucose	579.319
31		Cannogenol	C ₃₇ H ₅₆ O ₁₅	8.06	741.374	763.352	Acetyl-Deoxyhexose	Glucose	579.319
32		Cannogenin	C ₃₄ H ₅₀ O ₁₂	7.78	651.337	673.32	Digitoxose	Xylose	519.296
33		Cannogenin	C ₃₅ H ₅₂ O ₁₃	7.34	681.348	703.33	Digitoxose	Glucose	519.296
34	Glucocheiranthoside	Cannogenin	C ₃₅ H ₅₂ O ₁₄	6.94	697.356	719.325	Deoxyhexose	Glucose	535.289
35		<i>Digitoxigenin-formate</i>	C ₃₀ H ₄₄ O ₁₀	8.01	565.302	587.283	Deoxyhexose	-	
36		<i>Digitoxigenin-formate</i>	C ₃₀ H ₄₄ O ₁₀	8.18	565.302	587.283	Deoxyhexose	-	
37		<i>Digitoxigenin-formate</i>	C ₃₆ H ₅₄ O ₁₄	8.37	711.360	733.342	Deoxyhexose	Deoxyhexose	565.302
38		<i>Digitoxigenin-formate</i>	C ₃₆ H ₅₄ O ₁₅	7.03	727.353	749.335	Deoxyhexose	Glucose	565.302
39		<i>Digitoxigenin-formate</i>	C ₃₆ H ₅₄ O ₁₅	7.15	727.353	749.335	Deoxyhexose	Glucose	565.302
40		<i>Oleandrigenin</i>	C ₃₇ H ₅₆ O ₁₅	7.16	741.374	763.352	Deoxyhexose	Glucose	579.317
41		<i>Oleandrigenin</i>	C ₃₇ H ₅₆ O ₁₅	7.34	741.374	763.352	Deoxyhexose	Glucose	579.317
42		<i>Oleandrigenin</i>	C ₃₇ H ₅₆ O ₁₅	7.41	741.374	763.352	Deoxyhexose	Glucose	579.317
43		<i>Oleandrigenin</i>	C ₃₇ H ₅₆ O ₁₅	7.59	741.374	763.352	Deoxyhexose	Glucose	579.317
44		Strophanthidin	C ₂₉ H ₄₂ O ₉	7.32	535.289	557.273	Digitoxose	-	
45		Strophanthidin	C ₂₉ H ₄₂ O ₉	7.4	535.289	557.273	Digitoxose	-	
46	Helveticoside*	Strophanthidin	C ₂₉ H ₄₂ O ₉	7.71	535.289	557.273	Digitoxose	-	
47		Strophanthidin	C ₂₉ H ₄₂ O ₁₀	6.96	551.286	573.267	Deoxyhexose	-	
48		Strophanthidin	C ₂₉ H ₄₂ O ₁₀	7.07	551.286	573.267	Deoxyhexose	-	
49		Strophanthidin	C ₂₉ H ₄₂ O ₁₀	7.29	551.286	573.267	Deoxyhexose	-	
50		Strophanthidin	C ₂₉ H ₄₂ O ₁₀	7.67	551.286	573.267	Deoxyhexose	-	
51		Strophanthidin	C ₃₁ H ₄₄ O ₁₀	8.91	577.301	599.283	Acetyl-Digitoxose	-	
52		Strophanthidin	C ₃₁ H ₄₄ O ₁₀	9	577.301	599.283	Acetyl-Digitoxose	-	
53		Strophanthidin	C ₃₁ H ₄₄ O ₁₁	8.08	593.296	615.277	Acetyl-Deoxyhexose	-	
54		Strophanthidin	C ₃₁ H ₄₄ O ₁₁	8.41	593.296	615.277	Acetyl-Deoxyhexose	-	
55	Erychroside	Strophanthidin	C ₃₄ H ₅₀ O ₁₃	6.8	667.333	689.315	Digitoxose	Xylose	535.291
56		Strophanthidin	C ₃₄ H ₅₀ O ₁₄	6.45	683.328	705.311	Deoxyhexose	Xylose	551.271
57		Strophanthidin	C ₃₅ H ₅₂ O ₁₃	7.14	681.348	703.33	Digitoxose	Deoxyhexose	535.291
58	Erysimoside*	Strophanthidin	C ₃₅ H ₅₂ O ₁₄	6.48	697.344	719.325	Digitoxose	Glucose	535.287

59		Strophanthidin	C ₃₅ H ₅₂ O ₁₄	6.72	697.344	719.325	Deoxyhexose	Deoxyhexose	551.282
60	Cheirotoxin	Strophanthidin	C ₃₅ H ₅₂ O ₁₅	6.22	713.338	735.32	Gulomethylose	Glucose	551.286
61		Strophanthidin	C ₃₆ H ₅₂ O ₁₄	7.48	709.344	731.326	Acetyl-Digitoxose	Xylose	577.301
62		Strophanthidin	C ₃₆ H ₅₂ O ₁₄	7.89	709.344	731.326	Acetyl-Digitoxose	Xylose	577.301
63		Strophanthidin	C ₃₇ H ₅₄ O ₁₅	7.12	739.365	761.336	Acetyl-Digitoxose	Glucose	577.301
64		Strophanthidin	C ₃₇ H ₅₄ O ₁₅	7.15	739.365	761.336	Acetyl-Digitoxose	Glucose	577.301
65		Strophanthidin	C ₃₇ H ₅₂ O ₁₆	7.51	753.331	775.316	Digitoxose	C ₈ H ₁₂ O ₈	535.289
66		<i>Strophanthidin-formate</i>	C ₃₀ H ₄₂ O ₁₁	6.49	579.28	601.263	Digitoxose	-	
67		<i>Strophanthidin-formate</i>	C ₃₀ H ₄₂ O ₁₁	6.99	579.28	601.263	Digitoxose	-	
68		<i>Strophanthidin-formate</i>	C ₃₀ H ₄₂ O ₁₂	6.14	595.275	617.257	Deoxyhexose	-	
69		<i>Strophanthidin-formate</i>	C ₃₀ H ₄₂ O ₁₂	6.21	595.275	617.257	Deoxyhexose	-	
70		<i>Strophanthidin-formate</i>	C ₃₀ H ₄₂ O ₁₂	6.56	595.275	617.257	Deoxyhexose	-	
71		<i>Strophanthidin-formate</i>	C ₃₀ H ₄₂ O ₁₃	4.8	611.271	633.252	Glucose	-	
72		<i>Strophanthidin-formate</i>	C ₃₂ H ₄₄ O ₁₂	8.01	621.291	643.278	Acetyl-Digitoxose	-	
73		<i>Strophanthidin-formate</i>	C ₃₅ H ₅₀ O ₁₅	6.27	711.323	733.305	Digitoxose	Xylose	579.282
74		<i>Strophanthidin-formate</i>	C ₃₅ H ₅₀ O ₁₆	5.88	727.317	749.299	Deoxyhexose	Xylose	595.275
75		<i>Strophanthidin-formate</i>	C ₃₆ H ₅₂ O ₁₅	6.62	725.338	747.32	Digitoxose	Deoxyhexose	579.282
76		<i>Strophanthidin-formate</i>	C ₃₆ H ₅₂ O ₁₆	5.67	741.333	763.315	Digitoxose	Glucose	579.282
77		<i>Strophanthidin-formate</i>	C ₃₆ H ₅₂ O ₁₆	5.95	741.333	763.315	Digitoxose	Glucose	579.282
78		<i>Strophanthidin-formate</i>	C ₃₈ H ₅₂ O ₁₈	7.01	797.323	819.306	Digitoxose	C ₈ H ₁₂ O ₈	579.282
79		Bipindogenin	C ₂₉ H ₄₄ O ₁₀	4.59	553.302	575.282	Deoxyhexose	-	

80		Bipindogenin	C ₂₉ H ₄₄ O ₁₀	4.83	553.302	575.282	Deoxyhexose	-
81		Bipindogenin	C ₂₉ H ₄₄ O ₁₀	4.9	553.302	575.282	Deoxyhexose	-
82		Bipindogenin	C ₂₉ H ₄₄ O ₁₀	5.76	553.302	575.282	Deoxyhexose	-
83		Bipindogenin	C ₂₉ H ₄₄ O ₁₀	5.98	553.302	575.282	Deoxyhexose	-
84		Bipindogenin	C ₂₉ H ₄₄ O ₁₀	6.64	553.302	575.282	Deoxyhexose	-
85		Bipindogenin	C ₃₁ H ₄₆ O ₁₁	5.53	595.312	617.296	Acetyl-Deoxyhexose	-
86		Bipindogenin	C ₃₁ H ₄₆ O ₁₁	6.28	595.312	617.296	Acetyl-Deoxyhexose	-
87		Bipindogenin	C ₃₅ H ₅₄ O ₁₅	5.22	715.365	737.346	Deoxyhexose	Glucose 553.301
88	Nigescigenin digitoxoside	Nigescigenin	C ₂₉ H ₄₂ O ₁₀	5.05	551.288	573.27	Digitoxose	-
89		Nigescigenin	C ₂₉ H ₄₂ O ₁₁	4.73	567.278	589.259	Deoxyhexose	-
90		Nigescigenin	C ₂₉ H ₄₂ O ₁₁	6.8	567.278	589.259	Deoxyhexose	-
91		Nigescigenin	C ₂₉ H ₄₂ O ₁₁	6.99	567.278	589.259	Deoxyhexose	-
92		Nigescigenin	C ₃₁ H ₄₄ O ₁₂	7.78	609.292	631.271	Acetyl-Deoxyhexose	-
93		Nigescigenin	C ₃₁ H ₄₄ O ₁₂	8.23	609.292	631.271	Acetyl-Deoxyhexose	-
94		Nigescigenin	C ₃₅ H ₅₂ O ₁₃	4.69	683.329	705.311	Digitoxose	Xylose 551.284
95		Nigescigenin	C ₃₅ H ₅₂ O ₁₄	4.92	697.344	719.324	Digitoxose	C ₆ H ₁₂ O ₅ 551.281
96		Nigescigenin	C ₃₅ H ₅₂ O ₁₅	6.29	713.338	735.32	Digitoxose	Glucose 551.298
97	Glucocanescein	Nigescigenin	C ₃₅ H ₅₂ O ₁₆	6.1	729.333	751.316	Gulomethylose	Glucose 567.284