Polygenic routes lead to parallel altitudinal adaptation in *Heliosperma pusillum* (Caryophyllaceae)

Aglaia Szukala[1,2*], Jessica Lovegrove-Walsh[1], Hirzi Luqman[4], Simone Fior[4], Thomas Wolfe[5], Božo Frajman[3], Peter Schönswetter[3] and Ovidiu Paun[1*]

[1] *Department of Botany and Biodiversity Research, University of Vienna, Vienna, Austria.*

[2] *Vienna Graduate School of Population Genetics, Vienna, Austria.*

[3] *Department of Botany, University of Innsbruck, Innsbruck, Austria.*

[4] *Department of Environmental System Science, ETH Zürich, Zürich, Switzerland.*

[5] *Institute for Forest Entomology, Forest Pathology and Forest Protection, BOKU, Vienna, Austria.*

*authors for correspondence:

Aglaia Szukala, aglaia.szukala@univie.ac.at

Ovidiu Paun, ovidiu.paun@univie.ac.at

Abstract

Understanding how organisms adapt to the environment is a major goal of modern biology. Parallel evolution - the independent evolution of similar phenotypes in different populations - provides a powerful framework to explore this question. Here, we quantified the degree of gene expression and functional parallelism across replicated ecotype formation in *Heliosperma pusillum* (Caryophyllaceae) and gained insights into the architecture of adaptive traits. Population structure analyses and demographic modelling confirm the previously formulated hypothesis of parallel polytopic divergence of montane and alpine ecotypes. We detect a large proportion of differentially expressed genes (DEGs) underlying adaptation of each replicate ecotype pair, with a strikingly low amount of shared DEGs across pairs. Functional enrichment of DEGs reveals that the traits affected by divergent gene expression are the same across ecotype pairs, in strong contrast to the non-shared genetic basis. The remarkable redundancy of differential gene expression indicates that diverged adaptive traits are highly polygenic. We conclude that polygenic traits appear key to opening multiple routes for adaptation, widening the adaptive potential of organisms.

Keywords: Polygenic adaptation, Parallel divergence, RNA-seq, Demography, Ecotypes;

Introduction

Independent instances of adaptation with similar phenotypic outcomes (i.e., parallel evolution) are powerful avenues for exploring the mechanisms and timescale of adaptation and divergence (Arendt and Reznick 2008; Turner et al. 2010; Agrawal 2017; Buckley et al. 2019; Knotek et al. 2020). Evolutionary replicates offer insight into the constraints on evolution and help disentangle the nonrandom or more "predictable" actions of natural selection from confounding stochastic effects such as drift and demography (Lee and Coop 2019). In particular, parallel formation of conspecific ecotypes (Nosil et al. 2009; Nosil et al. 2017), are pivotal to enhancing our understanding of the processes leading to adaptation in response to a changing environment.

A number of studies have shown that parallel evolution can be driven by either standing genetic variation, possibly shared across lineages through pre- or post-divergence gene flow (Colosimo et al. 2005; Jones et al. 2012; Soria-Carrasco et al. 2014; Van Belleghem et al. 2018; Alves et al. 2019; Thompson et al. 2019; Louis et al. 2020), or, more rarely, by recurrent *de novo* mutations with large phenotypic effects (Hoekstra et al. 2006; Chan et al. 2010; Zhen et al. 2012; Projecto-Garcia et al. 2013; Tan et al. 2020). These sources of adaptive variation produce phenotypic similarities via the same genetic locus, regardless if it was acquired independently or present in the ancestral gene pool (Stern 2013).

On the other hand, there is compelling evidence of phenotypic convergence resulting from non-parallel signatures of adaptation (Elmer et al. 2014; Yeaman et al. 2016; Rellstab et al. 2020), even among closely related populations (Wilkens and Strecker 2003; Steiner et al. 2009; Fischer et al. 2021), and replicated laboratory evolution (Cooper et al. 2003; Barghi et al. 2019). A typical example is the convergent evolution of a lighter coat pigmentation in beach mouse populations of the Gulf of Mexico and the Atlantic Coasts driven by different mutations (Steiner et al. 2009).

Such cases suggest that evolutionary replicates can follow diverse routes and relatively few molecular constraints exist in the evolution of adaptive traits (Arendt and Reznick 2008; Losos 2011). The degree of convergence during adaptation to similar selective pressures across taxa reveals that genomic signatures of adaptation are often redundant (Wilkens and Strecker 2003; Mandic et al. 2018; Fischer et al. 2021). Parallel evolution can involve highly heterogeneous routes depending on variation in gene flow, strength of selection, effective population size, demographic history, and extent of habitat differentiation, leading to different degrees of parallelism. This complex range of processes including non-parallel to parallel trajectories have also been described using the more comprehensive term *continuum of (non)parallel evolution* (Stuart et al. 2017; Bolnick et al. 2018).

Recently, a quantitative genetics view of the process of adaptation has gained attention among evolutionary biologists (Barghi et al. 2020), complementing existing models on adaptation via selective sweeps. Accordingly, selection can act on different combinations of loci, each of small effect, leading to shifts in the trait mean through changes in multiple loci within the same molecular pathway (Hermisson and Pennings 2017; Höllinger et al. 2019). Thus, a key feature of polygenic adaptation is that different combinations of adaptive alleles can contribute to the selected phenotype (Barghi et al. 2020). This heterogeneity among loci, termed genetic redundancy (Nowak et al. 1997; Barghi et al. 2019), can lead to non-parallel genomic changes in populations evolving under the same selective pressure. Footprints of selection acting on polygenic traits have been detected in a wide range of study systems, such as in fish (Therkildsen et al. 2019) and in cacao plants (Hämälä et al. 2020), which can also foster convergent adaptive responses and phenotypes during independent divergence events (Lim et al. 2019; Rougeux et al. 2019; Hämälä et al. 2020).

A current major challenge is predicting adaptive responses of populations and species to environmental change. Despite several advances, it is still unclear which adaptive signatures

are expected to be consistent across evolutionary replicates, especially when selection acts on complex traits. An important aspect to investigate is the architecture of adaptive traits (simple/monogenic, oligogenic or polygenic). A polygenic architecture may facilitate alternative pathways leading to the same phenotypic innovation, enhancing the probability of parallel evolution (Boyle et al. 2017), and, as a consequence, the adaptive potential of populations.

To date, only a handful of cases of parallel evolution have been extensively studied in plants (Roda et al. 2013; Yeaman et al. 2016; Trucchi et al. 2017; Konečná et al. 2019; James et al. 2020; Rellstab et al. 2020; Tan et al. 2020). Altitudinal ecotypes of *Heliosperma pusillum* s.l. (Waldst. and Kit.) Rchb. (Caryophyllaceae) offer a system to study this process. In the Alps, this species includes an alpine ecotype (1,400–2,300 m above sea level) widely distributed across the mountain ranges of southern and central Europe, and a montane ecotype (500–1,300 m) endemic to the south-eastern Alps. The latter was previously described from scattered localities as *H. veselskyi* Janka, but the two ecotypes are highly interfertile (Bertel, Hülber, et al. 2016) and isolation-by-distance analyses confirmed their conspecificity (Trucchi et al. 2017). While the alpine ecotype has a relatively continuous distribution in moist screes above the timberline, the montane ecotype forms small populations (typically < 100 individuals) below overhanging rocks.

Previous work (Bertel, Buchner, et al. 2016; Bertel et al. 2018) reported substantial abiotic differences between the habitats preferred by the two ecotypes. For example, differences in average temperature (montane: warm vs. alpine: cold), temperature amplitude, the degree of humidity (montane: dry vs. alpine: humid) and light availability (montane: shade vs. alpine: full sunlight) were found between the two altitudinal sites. Moreover, metagenomics (Trucchi et al. 2017) showed evidence of distinct microbial communities in the respective phyllospheres. The two ecotypes also differ significantly in their physiological response to light

108 and humidity conditions in a common garden (Bertel, Buchner, et al. 2016). Finally, the
109 montane ecotype is covered by a dense glandular indumentum, which is absent in the alpine
110 populations (Frajman and Oxelman 2007; Bertel et al. 2017). Whether this phenotypic
111 difference in trichome density confers differential adaptation in their respective environment
112 is unclear.

113 Both ecotypes show higher fitness at their native sites in reciprocal transplantation
114 experiments (Bertel et al. 2018), confirming an adaptive component to their divergence.
115 Common garden experiments across multiple generations further rejected the hypothesis of a
116 solely plastic response shaping the phenotypic divergence observed (Bertel et al. 2017). Most
117 importantly, population structure analyses based on genome-wide SNPs derived from
118 restriction site-associated DNA (RAD-seq) markers (Trucchi et al. 2017) supported a scenario
119 of five parallel divergence events across the six investigated ecotype pairs. Hereafter, we use
120 the term "ecotype pairs" to indicate single instances of divergence between alpine and montane
121 ecotypes across their range of co-occurrence.

122 The combination of ecological, morphological, and demographic features outlined
123 above renders *H. pusillum* a well-suited system to investigate the mechanisms driving local
124 recurrent altitudinal adaptation in the Alps. Here, we quantify the magnitude of gene expression
125 and functional parallelism across ecotype pairs, by means of RNA-seq analyses of wild plants
126 grown in a common garden. We also investigate the independent evolution of ecotype pairs
127 more in depth than previously. More specifically, this study asks (1) how shared are gene
128 expression differences between ecotypes among evolutionary replicates or, in other words, is
129 the adaptation to elevation driven by expression changes in specific genes or in different genes
130 affecting similar traits, (2) how shared is the adaptive functional variation encoded by
131 differentially expressed genes (DEGs) among evolutionary replicates, and (3) do we find
132 consistent signatures of selection on coding sequence variation across evolutionary replicates?
133

134 **Results**

135 Our main aim was to test the repeatability of the molecular patterns and functions that
136 distinguish the alpine from the montane ecotype in different ecotype pairs. To achieve this goal,
137 we performed differential expression (DE) analyses of four ecotype pairs (4 pairs × 2 ecotypes
138 × 3 replicates, i.e., 24 individuals in total; Fig. 1a and b, Table S1) grown in a common garden
139 to remove any differential environmental effects that might confound gene expression in the
140 samples. To identify genetic variants under selection we extended the sampling by including
141 41 additional transcriptomes of individuals from a transplantation experiment (Szukala A et al.
142 unpublished data; Table S1). Two alpine individuals of pair 3 (A3b and A3c, Table S1) were
143 found highly introgressed with genes from the alpine population of pair 4 (Fig. S1a), and have
144 been discarded from subsequent genetic analyses, retaining a total of 63 individuals for further
145 analyses based on SNPs. This dataset was also used to test the hypothesis of parallel ecotype
146 divergence in *H. pusillum* suggested by Trucchi et al (2017).

147 **Reference genome assembly and annotation.** We first assembled *de novo* a draft
148 genome for an individual of the alpine ecotype of *H. pusillum* (generated via artificial selfing
149 over three generations) using 192.3 Gb (ca 148×) Illumina paired-end reads and 14.9 Gb (ca
150 11.5×) PacBio single-molecule long reads. A hybrid assembly recovered a total length of 1.21
151 Gb of scaffolds corresponding to 93% of the estimated genome size (1C = 1.3 pg, (Temsch et
152 al. 2010)). The draft genome was split into 75,567 scaffolds with an N50 size of 41,610 bp.
153 RepeatModeler v.1.0.11 (http://www.repeatmasker.org/RepeatModeler/) identified 1,021
154 repeat families making up roughly 71% of the recovered genome. This high proportion of
155 repetitive elements aligns well with observations in other plant genomes.

156 Structural annotations (Stanke et al. 2006) identified 25,661 protein-coding genes with
157 an average length of 4,570 bp (Fig. S2a and b). All protein-coding genes were found on 8,632
158 scaffolds that belong to the longest tail of the contig length distribution (Fig. S2d).

159 Nevertheless, we also observe in our assembly comparatively long contigs that do not contain
160 any gene models (Fig. S2d). Of this set of genes, 17,009 could be functionally annotated (Götz
161 et al. 2008; Haas et al. 2013). We evaluated the completeness of the genome assembly by
162 searching our gene models against the BUSCO v.3 embryophyta dataset (Simão et al. 2015).
163 A total of 83.2% of the set of single-copy conserved BUSCO genes were found within our
164 annotated genes (Fig. S2c).

165 **Genetic diversity and structure.** To explore population diversity and structure we
166 filtered a dataset of 7,107 putatively neutral SNPs at unlinked four-fold degenerate (FFD) sites
167 from 63 individuals representing the four ecotype pairs (Fig. 1b, Table S1). Within-population
168 allelic diversity (average pairwise nucleotide diversity, π, and Watterson's theta, θ_w), Tajima's
169 D, as well as between-population differentiation (F_{ST}), are reported in Table S2. Average π
170 showed similar values across alpine and montane populations, ranging from $\pi_{A5} = 0.140 \pm 0.12$
171 to $\pi_{A1} = 0.173 \pm 0.12$ in the alpine ecotype, and from $\pi_{M3} = 0.143 \pm 0.12$ to $\pi_{M5} = 0.171 \pm 0.13$
172 in the montane. Watterson's theta ranged from $\theta_{w\text{-}A5} = 0.130 \pm 0.11$ to $\theta_{w\text{-}A4} = 0.136 \pm 0.12$ and
173 from $\theta_{w\text{-}M3} = 0.116 \pm 0.08$ to $\theta_{w\text{-}M5} = 0.157 \pm 0.11$ in the alpine and montane ecotype,
174 respectively. We did not observe a clear alpine versus montane distinction of within-population
175 allelic diversity. Global Tajima's D estimates were always positive, but close to 0 (Table S2,
176 Fig. S3), suggesting that these populations are within neutral-equilibrium expectations, and that
177 both alpine and montane populations were not affected by major changes in population size in
178 the recent past.

179 Averaged pairwise F_{ST} tended to be slightly higher between montane than between
180 alpine populations (weighted F_{ST} = 0.28-0.56 in alpine, and weighted F_{ST} = 0.39-0.52 in
181 montane; Table S2). Between-ecotype F_{ST} was lower than F_{ST} between pairs, except in the case
182 of pair 4 (weighted $F_{ST} = 0.48$), consistent with overall high expression differentiation between
183 ecotypes in this pair as described below.

184 We further investigated the population structure with principal component analyses
185 (PCA) and an admixture plot, both based on genotype likelihoods computed in ANGSD
186 v.0.931 (Korneliussen et al. 2014). In the PCA (Fig. 1c) the analyzed populations cluster by
187 geography, in line with previous results (Trucchi et al. 2017). The first component (15.2% of
188 explained variance, Fig. 1c) shows a clear east-west separation of the ecotype pairs. The second
189 component (12.4% of explained variance, Fig. 1c) places ecotype pair 5 closer to pair 1 and
190 most distant from pair 3 showing a north-south separation.

191 We performed two rounds of population structure inference using NgsAdmix v.32
192 (Skotte et al. 2013), to test the effects of uneven sample size on the inferred clusters. We
193 compared the results inferred using the set of 63 accessions to those inferred when randomly
194 subsampling all populations to three individuals (i.e., the minimum number of individuals per
195 population in our dataset). With uneven sampling, we observed that the individuals from
196 populations with reduced sampling size (i.e., ecotype pair 4) tended to be assigned to
197 populations of higher sampling density (Fig. S1b), an otherwise known problem affecting
198 population structure analyses (Puechmaille 2016; Meirmans 2019). Consistent with the
199 clustering observed in the PCA, pair 5 was first separated from the rest pairs (K = 2, Fig. 1d).
200 The best three Ks, as evaluated using the Evanno method (Evanno et al. 2005), were 2, 3 and
201 7, in this order, confirming an enhanced separation of pair 5 from the rest, while the two
202 ecotypes in this pair are the least diverged (K = 7, Fig. 1d), consistent with a lower degree of
203 expression differentiation in this pair (Fig. 2).

204 **Demographic model selection, parallelism and gene flow.** We tested two contrasting
205 topologies for each combination of two ecotype pairs (Fig. 3): one model assuming a single
206 origin (1-origin) of each ecotype, and one assuming independent between-ecotype divergence
207 across geographic localities (2-origins). Additionally, for each topology two scenarios were
208 evaluated: one in absence of migration between populations (strict isolation, SI) and one with

209 continuous migration between demes (isolation with migration, IM). In line with the results
210 from the population structure analyses our expectation was to find higher migration rates
211 between ecotypes within each ecotype pair (solid lines in Fig. 3). To evaluate which model
212 better explains our data we used a composite-likelihood maximization method implemented in
213 fastSimcoal2 v.2.6.0.3 (Excoffier et al. 2013). Delta Akaike information criteria (ΔAIC) for
214 each model tested are summarized in Table S3a and c.

215 Our SI simulations showed that the 2-origins topology is always preferred. When pair
216 3 and 4 were analysed, the difference between the performance of the two models was small.
217 In all cases, models allowing gene flow achieved a higher likelihood than SI models (Table
218 S3a). The 2-origins scenario again achieved a better likelihood in 5 out of 6 ecotype pair
219 comparisons. The 1-origin model was preferred when pairs 3-4 were simulated. For each
220 parameter we took as a final estimate the 95% confidence intervals CI of the ten best model
221 estimates. The CI of the times of divergence and effective population size (N_e) from the best
222 model estimates are reported in Table S3b. We computed migration rate estimates for each
223 model including both directions of migration for all combinations of ecotype populations from
224 two pairs (Table S3d). We found migration rates to be generally higher between ecotypes in
225 each pair, and low between different ecotype pairs (upper limit of the CI always below 0.002)
226 across all comparisons and scenarios tested (Table S3d).

227 **Patterns of differential gene expression between ecotypes.** We analysed gene
228 expression in a common garden to identify genes with divergent expression between ecotypes,
229 as these are hypothesized to underlie phenotypic differentiation and adaptation to different
230 altitudinal niches. After trimming genes with low mean expression counts per million (cpm)
231 across samples we retained a dataset of 16,389 genes on which we performed DE analyses.

232 A major proportion of DEGs were found to be unique to each pair (colored area of the
233 bars in Fig. 2a and b). This pattern was particularly enhanced in pair 5, in which ca 85% of

234 DEGs were not shared with other pairs, while ca 65%, 70% and 80% of DEGs were unique to
235 pairs 3, 1 and 4, respectively (Fig. 2, Fig. S4). Although the overlap of DEGs was significantly
236 higher than chance expectations ($p < 0.01$) in several comparisons, our analyses recovered an
237 overall low number of shared DEGs. In contrast to expectations, we found across all ecotype
238 pairs that only two and zero genes were consistently over- and under-expressed in the montane
239 compared to the alpine ecotype, respectively.

240 The number of DEGs varied across ecotype pairs. DEGs were almost four times higher
241 in pair 4 (highest degree of expression differentiation) compared to pair 5 (lowest degree of
242 expression differentiation), while the difference in DEGs was less pronounced between pairs 1
243 and 3. This result is consistent with the PCA of normalized read counts (Fig. S5a) and the
244 multidimensional scaling plot of gene expression (Fig. S5b). Overall, the relative degree of
245 expression differentiation between ecotypes at different geographic localities is consistent with
246 their degree of genetic differentiation (F_{ST}, Table S2). The second component of the PCA of
247 gene expression (13.8% of variance explained, Fig. S5a), as well as the second dimension of
248 logFC of the multidimensional scaling analysis (Fig. S5b), tends to separate the two ecotypes.
249 Interestingly, gene expression appears more uniform across the montane accessions compared
250 to the alpine ones, even if the overall expression divergence between different populations was
251 not significantly different between ecotypes (Wilcoxon signed rank test $p = 0.56$; Fig. S6 and
252 Table S4).

253 **Parallel multilocus gene expression variation.** We performed a conditioned (partial)
254 redundancy analysis (cRDA) of gene expression to elucidate if a different analytical framework
255 would have more power to detect common genes with opposite expression patterns between
256 ecotypes across all evolutionary replicates. Redundancy analysis is thought to be a good
257 approach to detect changes between conditions (in our case, ecotypes), even when such
258 differences are subtle and possibly masked by other factors (Forester et al. 2018).

259 We found that 1.8% of total expression variation was explained by divergence between
260 montane and alpine ecotypes across all ecotype pairs (Fig. 4), consistent with low overlap of
261 DEGs across evolutionary replicates in DE analyses. Also consistent with the low number of
262 shared DEGs, the ANOVA test of the full model was not significant ($F = 1.39$, $p = 0.18$),
263 confirming that most expression differences between ecotypes in our dataset do not follow
264 consistent routes across ecotype pairs. We further searched for cRDA outliers to identify genes
265 with consistent, albeit subtle, changes in expression across ecotypes. The transcript score was
266 transformed into a z-score with a distribution ranging from -3.55 to 3.43 (Fig. S7). We
267 identified 115 ($p < 0.01$) and 739 ($p < 0.05$) outlier genes with consistently opposite expression
268 direction between ecotypes across all ecotype pairs.

269 **Ecological and biological significance of DEGs.** In stark contrast to the low overlap
270 at the level of individual genes affected by DE, we observed evidence of convergence in the
271 enriched biological functions across DEG lists of each ecotype pair. Gene ontologies (GO)
272 terms enriched (adjusted $p < 0.05$) in genes that tended to be underexpressed in the montane
273 ecotype across ecotype pairs included photosynthesis, oxidation-reduction processes -
274 potentially related to acclimation responses related to photosynthesis, protein phosphorylation
275 and responses to biotic stress (Fig. 5, Table S5). By contrast, GO terms enriched in genes
276 overexpressed in the montane ecotype comprise transmembrane transport activity and
277 regulation of water content, probably involved in response to water deficit and drought (Fig. 5,
278 Table S5). The degree of over- and underexpression of convergent GO terms across pairs
279 varied, depending on the specific function of the genes affecting the same molecular pathway.
280 We also analyzed enriched biological processes in cRDA gene outliers, since these genes
281 possibly underlie biologically and ecologically relevant adaptive traits. Consistent with the DE
282 results, cRDA outlier genes were significantly enriched for photosynthesis-related functions
283 and response to water deprivation.

284 In the GO enrichment analysis of the DEG lists we did not find significantly enriched
285 GO terms related to trichome development, despite the obvious morphological difference
286 between the two ecotypes, given that multicellular glandular trichomes are present in montane
287 plants but not in alpine. We observed that some genes known to be involved in trichome
288 formation and found to be expressed in our transcriptomes were significantly differentially
289 expressed in some of the ecotype pairs but not in others, or showed consistent changes in
290 expression between ecotypes even if not significant after FDR correction (examples shown in
291 Fig. 6). As opposed to the DE analysis, in the cRDA analysis we found a significant enrichment
292 for the term root hair elongation (GO:0048767), a GO term that shares functionality with the
293 multicellular trichome development pathway (Benítez et al. 2007). We subsequently identified
294 gene candidates underlying this trait, as they show subtle gene expression differences between
295 ecotypes that were consistent across all four comparisons analyzed. IBR3, a Indole-3-butyric
296 acid response gene, known to promote hair elongation (Strader et al. 2010; Velasquez et al.
297 2016) was always overexpressed in the montane ecotype (Fig. 6). This same gene was also
298 significantly differentially expressed in three out of four ecotype pairs in previous DEG
299 analyses before correction of p-values for multiple testing (Fig. 6). Another candidate, the PID
300 (PIDox) gene, a positive regulator of auxin efflux known to suppress root hair growth when
301 overexpressed (Lee and Cho 2006), was found to be underexpressed in the montane ecotype in
302 pairs 1, 4 and 5. Overall, our analyses found low overlap of expression changes in genes
303 controlling trichome formation across ecotype pairs of *H. pusillum*.

304 **(Non-)Shared adaptive outlier loci.** To identify possible candidate genes under
305 divergent selection in independent divergence events, we searched for coding genomic regions
306 with pronounced allelic divergence between ecotypes in pairs 1 and 3. We excluded ecotype
307 pairs 4 and 5 from this analysis because of the low number of individuals available from each
308 population. To detect SNPs with genotype divergence between ecotypes we used the method

309 implemented in pcadapt (Luu et al. 2017) which is based on principal component analysis. We
310 then searched for genes containing outlier SNPs that are shared by ecotype pairs 1 and 3.
311 After variant calling and filtering we retained a dataset with 116,075 and 86,900
312 variable biallelic sites in pair 1 and 3, respectively, identified across 8,051 genes. Principal
313 component analysis of each pair shows that genetic variation well separates the two ecotypes
314 (PC1 = 5.8% and 14.3% in pairs 1 and 3, respectively, Fig. S8). In both ecotype pairs, the alpine
315 ecotype shows a higher genotypic variance (Fig. S8), which is consistent with higher variance
316 of gene expression in this ecotype compared to the montane (Fig. S5). In pair 1 we found 445
317 outlier SNPs distributed across a total of 141 genes. In pair 3 we found 444 outlier SNPs
318 distributed across a total of 123 genes. Eighty-seven SNPs and 19 genes containing outlier
319 SNPs were shared by both pairs, a number significantly higher than expected by chance ($p <$
320 0.0001 for both genes and SNPs). Our results also show that the majority of outlier genes is not
321 shared by different ecotype pairs, consistent with a mostly independent divergence history of
322 each ecotype pair and very low overlap of DEGs. Functional annotations of the 19 shared genes
323 containing outlier SNPs are reported in Table S6. Among those candidate genes, we found
324 genes involved in response to drought and salt stress (APUM5, AHA4, PYL8), oxidation-
325 reduction processes (POA2, NADPH-cytochrome P450 reductase, CCR1), light signal
326 transduction (PIA2 and a putative LOV domain containing protein) and flavonoid biosynthesis
327 (HIDH). Genes containing outliers were not differentially expressed.
328

329 **Discussion**

330 Parallel evolution has long been recognised as a powerful process to study adaptation,
331 overcoming intrinsic limitations of studies on natural populations that often miss replication
332 (Elmer and Meyer 2011). In this work, we aimed to investigate the genetic basis of adaptation

333 to different elevations in the plant *Heliosperma pusillum*. We asked in particular to what extent
334 different ecotype pairs show signatures of parallel evolution in this system.

335 Our genetic structure analyses and coalescence-based demographic modelling were in
336 line with a scenario of parallel, polytopic ecotype divergence, as suggested previously by a
337 marked dissimilarity of the genomic landscape of differentiation between ecotype pairs
338 revealed by RAD-seq data (Trucchi et al. 2017). In our demographic investigations, parallel
339 divergence always obtained greater support under a strict isolation model. Still, models
340 including low amounts of gene flow were shown to be more likely. Additionally, in one
341 comparison (i.e., including ecotype pairs 3 and 4) the single origin IM scenario aligned more
342 closely with the data than the two origins IM. This result is consistent with greater co-ancestry
343 observed for these two pairs with respect to other comparisons (Fig. 1c and d). Nevertheless,
344 the estimates of migration rates between different ecotype pairs were overall extremely low
345 (i.e., always lower than 1.2e-03), indicating that each ecotype pair diverged in isolation from
346 other pairs, even when it is not straightforward to distinguish between the different models (i.e.
347 1-origin vs 2-origins).

348 Our results from selection scans showed that only few diverged genes, likely under
349 selection during adaptation to different elevations, were shared between the two ecotype pairs
350 analyzed (i.e., pair 1 and 3), while ca 85% of putatively adaptive loci were unique to each pair.
351 This high degree of unique outliers, consistent with a previous investigation (Trucchi et al.
352 2017), supports a scenario of mainly independent evolutionary history of different ecotype
353 pairs. However, we cannot exclude the possibility that a few shared loci, likely from standing
354 genetic variation, might have played a role in shaping the ecotype divergence of different
355 evolutionary replicates in our system.

356 Global Tajima's D estimates were close to 0, suggesting that the recent past of all these
357 populations was not affected by major bottlenecks or population expansions. Consistently,

358 within-population diversity was similar across montane and alpine ecotypes, likely reflecting
359 ancestral variation before altitudinal divergence. Due to the low number of individuals
360 available for ecotype pairs 4 (three individuals per ecotype) and 5 (four individuals per
361 ecotype), these estimates should be considered with caution. However, previous work using an
362 RNA-seq-derived dataset of synonymous variants similar to ours (Fraïsse et al. 2018) showed
363 that model selection based on the joint site frequency spectrum is robust to the numbers of
364 individuals and loci. Nevertheless, future analyses should aim for enlarged sampling sizes.

365 We further asked how consistent across divergence events are the molecular processes
366 underlying ecotype formation. We screened the expression profiles of four ecotype pairs grown
367 in a common garden to shed light on the genetic architecture of the adaptive traits involved in
368 parallel adaptation to divergent elevations as well as to warmer/dry vs. colder/humid
369 conditions. Our analyses showed that gene expression changes between ecotypes is genetically
370 determined and not a plastic response due to environmental differences. We found strikingly
371 few DEGs shared across all four ecotype pairs, with most DEGs unique to one ecotype pair,
372 suggesting that convergent phenotypes do not consistently rely on changes in expression of
373 specific genes. This pattern was most pronounced in ecotype pair 5, which we also showed to
374 bear a lower degree of shared ancestry with the other pairs in genetic structure analyses (Fig.
375 1c-d). Given that ecotype pair 5 is the most eastern in terms of geographic distribution, it can
376 be hypothesized that this pair represents a more distinct lineage, as break zones in the
377 distribution of genetic diversity and distribution of biota have been identified to the West of
378 this area of the Alps (Thiel-Egenter et al. 2011). Our sampling was not appropriate to further
379 test hypotheses of biogeographic nature. Even so, our results suggest that parallel evolution is
380 analyzed at different levels of coancestry in our dataset. This implies that parallel signatures of
381 ecotype evolution can decrease significantly, even within a relatively small geographic range.
382 This view is in line with previous findings of unexpectedly heterogeneous differentiation

383 between freshwater and marine sticklebacks across the globe, including more distant lineages
384 (Fang et al. 2020).

385 Despite the low parallelism in gene activity, we identified across the ecotype pairs a
386 high reproducibility of the biological processes related to ecological divergence (i.e., different
387 water and light availability, and biotic stress) at the two elevations. Functional enrichment of
388 responses to biotic stress are consistent with the biotic divergence between the two habitat
389 types, featuring distinct potentially pathogenic microbiomes (Trucchi et al. 2017). The
390 dichotomy of convergence in enriched GO terms, but a low amount of shared DEGs, indicates
391 that different redundant genes likely concur to shape similar phenotypic differentiation, as
392 expected under polygenic adaptation (Barghi et al. 2020).

393 The presence (montane ecotype) or absence (alpine ecotype) of multicellular glandular
394 hairs on the plants represents a striking morphological difference in our system. Trichome
395 formation has been studied extensively in Brassicaceae, especially in *Arabidopsis*, where this
396 trait is controlled by a relatively simple regulatory pathway shared across the family (Hülskamp
397 et al. 1994; Hülskamp 2004; Hilscher et al. 2009; Pesch and Hülskamp 2009; Tominaga-Wada
398 et al. 2011; Chopra et al. 2019). Still, a certain degree of genetic redundancy has been shown
399 to underlie trichome formation in *Arabidopsis* (Khosla et al. 2014). Studies on other plant
400 lineages, such as cotton (Machado et al. 2009), snapdragons (Tan et al. 2020), *Artemisia* (Shi
401 et al. 2018) and tomato (Chang et al. 2018), highlighted that the genetic basis of multicellular
402 glandular trichomes formation does not always involve the same loci as in *Arabidopsis.*
403 Trichome formation outside of the Brassicaceae family likely involves convergent changes in
404 different genetic components (Serna and Martin 2006; Tan et al. 2020), and has been reported
405 to be initiated even as an epigenetic response to herbivory in *Mimulus guttatus* (Scoville et al.
406 2011).

407 We expected to find evidence of specific genes controlling trichome development in
408 our transcriptome dataset. However, we did not observe a consistent change in regulation of
409 particular genes underlying trichome initiation and elongation across ecotype pairs. This can
410 be caused by sampling only one ontogenetic stage during leaf development, which might not
411 be the most informative for a specific trait. Nonetheless, it is interesting to see that some genes
412 known to underlie hair initiation, elongation and malformation in other plant species were
413 differentially expressed in some ecotype pairs, but not in all of them.

414 Analyses of replicated evolution in laboratory experiments on bacteria (Cooper et al.
415 2003; Fong et al. 2005), yeast (Nguyen Ba et al. 2019) and *Drosophila* (Barghi et al. 2019)
416 have provided insights about adaptation, showing that redundant trajectories can lead to the
417 same phenotypic optimum, when selection acts on polygenic traits. In line with other studies
418 on diverse organisms including whitefish (Rougeux et al. 2019), hummingbirds (Lim et al.
419 2019), snails (Ravinet et al. 2016) and frogs (Sun et al. 2018), our results suggest that
420 convergent phenotypes can be achieved via changes in different genes affecting the same
421 molecular pathway and, ultimately, adaptive trait, and that this polygenic basis might facilitate
422 repeated adaptation to different elevations via alternative routes.

423 In conclusion, this study adds evidence to recent findings showing that polygenic traits
424 and genetic redundancy open multiple threads for adaptation, providing the substrate for
425 reproducible outcomes in convergent divergence events. Future studies using transcriptomics
426 as well as genomic approaches should focus on genotype-by-environment interactions, e.g., in
427 reciprocal transplantation experiments, to further deepen our understanding of the process of
428 adaptation in *H. pusillum*.

429

430 **Materials and methods**

431 **Sampling, library preparation and sequencing.** We performed DE analyses on 24
432 plants grown in common garden settings at the Botanical Garden of the University of
433 Innsbruck, Austria. Wild seeds were collected from four alpine/montane ecotype pairs in the
434 south-eastern Alps (Fig. 1b, Table S1). The numbering of localities is consistent with that used
435 in Bertel et al. (2018) and the acronyms corresponding to Trucchi et al. (2017) are added in
436 Table S1. All seeds were germinated on the same day and the seedlings were grown in uniform
437 conditions. One week before RNA fixation, the plants were brought to a climate chamber
438 (Percival PGC6L set to 16 h 25 °C three lamps/8 h 15 °C no lamps). Then, 400 mg of green
439 fresh stalk-leaf material, sampled at a similar developmental stage for all individuals, was fixed
440 in RNAlater in the same morning and kept at -80 °C until extraction. Total RNA was extracted
441 from ca 90 mg leaves using the mirVana miRNA Isolation Kit (Ambion) following the
442 manufacturer's instructions. Residual DNA has been digested with the RNase-Free DNase Set
443 (Qiagen); the abundant ribosomal RNA was depleted by using the Ribo-Zero rRNA Removal
444 Kit (Illumina). RNA was then quantified with a NanoDrop2000 spectrophotometer (Thermo
445 Scientific), and quality assessed using a 2100 Bioanalyzer (Agilent). Strand-specific libraries
446 were prepared with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New
447 England Biolabs). Two individual RNA-seq libraries were pooled and sequenced per Illumina
448 HiSeq 2500 lane with single-end reads (100 bp) at the NGS Facility at the Vienna BioCenter
449 Core Facilities (VBCF; https://www.viennabiocenter.org/). Two samples (A1a and A4b) were
450 sequenced with paired-end reads (150 bp) with the initial aim of assembling reference
451 transcriptomes.

452 The population genetic analyses (i.e., SNP-based) were performed on a larger dataset
453 including in total 63 individuals from the four population pairs (Supporting information Table
454 S1). For this we added RNA-seq samples from ecotype pairs 1 and 3 (Fig. 1b), which were

455 sampled within a different experiment. The procedure used to prepare the RNA-seq libraries
456 was the same as described above, except that 21 individual samples have been pooled and
457 sequenced together with single-end reads (100 bp) on Illumina NovaSeq S1 at the Vienna
458 BioCenter Core Facilities.

459 **Genomic DNA extraction and sequencing.** A reference genome was assembled using
460 short and long read technologies from an alpine individual that descended from population 1,
461 from a selfed line over three generations. DNA was extracted from leaves using a CTAB
462 protocol adapted from Cota-Sánchez et al. (2006). DNA for long reads was extracted from
463 etiolated tissue after keeping the plant for one week under no light conditions. Illumina libraries
464 were prepared with IlluminaTruSeq DNA PCR-free kits (Illumina) and sequenced as 150 bp
465 paired-end reads on Illumina HiSeq X Ten by Macrogen Inc. (Korea). PacBio library
466 preparation and sequencing of four SMRT cells on a Sequel I instrument was done at the
467 sequencing facility of the Vienna BioCenter Core Facilities.

468 **Reference genome assembly and annotation.** We performed a hybrid *de novo*
469 genome assembly using ca 120x paired-end Illumina and ca 15x PacBio reads with MaSuRCA
470 v.3.2.5 (Zimin et al. 2013). The assembled genome was structurally annotated *ab initio* using
471 Augustus (Stanke et al. 2006) and GeneMark-ET (Lomsadze et al. 2014), as implemented in
472 BRAKER1 v.2.1.0 (Hoff et al. 2016) with the options *--softmasking=1 --filterOutShort*.
473 Mapped RNA-seq data from three different samples was used to improve *de novo* gene finding.

474 A transcriptome was assembled using Trinity v.2.4.0 (Haas et al. 2013) to be used in
475 MAKER-P annotation as expressed sequence tag (EST). We used as additional evidence the
476 transcriptome of the closely related *Silene vulgaris* (Sloan et al. 2011). The annotation was
477 further improved using MAKER-P v.2.31.10 (Campbell et al. 2014) supplying gene models
478 identified using BRAKER1 and after generating a custom repeat library for masking using
479 RepeatModeler v.1.0.11 (http://www.repeatmasker.org/RepeatModeler/) as implemented in

480 MAKER-P. Gene models identified by both BRAKER1 and MAKER-P were functionally
481 annotated using Blast2GO (Götz et al. 2008; Haas et al. 2013). BUSCO v.2.0 (Simão et al.
482 2015) was used for quality assessment of the assembled genome and annotated gene models.

483 **Genetic diversity and structure.** RNA-seq data was demultiplexed using
484 BamIndexDecoder v.1.03 (http://wtsi-npg.github.io/illumina2bam/#BamIndexDecoder) and
485 raw sequencing reads were cleaned to remove adaptors and quality filtered using trimmomatic
486 v.0.36 (Bolger et al. 2014). Individual reads were aligned to the reference genome using STAR
487 v.2.6.0c (Dobin et al. 2013). Mapped files were sorted according to the mapping position and
488 duplicates were marked and removed using Picard v.2.9.2
489 (https://broadinstitute.github.io/picard/). Subsequently, we used a pipeline implemented in
490 ANGSD v.0.931 (Korneliussen et al. 2014) to estimate genotype likelihoods. The latter might
491 be more reliable than genotype calling for low coverage segments (Korneliussen et al. 2014),
492 in particular when handling data with strongly varying sequencing depth among regions and
493 individuals such as RNA-seq. Briefly, after mapping and removing duplicates, individual bam
494 files were processed using GATK v.3.7.0 function IndelRealigner to locally improve read
495 alignments around indels. Subsequently, ANGSD was run to compute posterior probabilities
496 for the three possible genotypes at each variant locus (considering only bi-allelic SNPs), taking
497 into account the observed allelic state in each read, the sequencing depth and the Phred-scaled
498 quality scores. ANGSD was run with the options *-GL 2 -doMajorMinor 1 -doMaf 1 -SNP_pval*
499 *2e-6 -minMapQ 20 -minQ 20 -minInd 12 -minMaf 0.045 -doGlf 2*. A significant portion of
500 RNA-seq data includes protein coding regions expected to be under selection. To investigate
501 genetic structure and demography the SNPs dataset was further filtered to keep FFD sites using
502 the Bioconductor package VariantAnnotation in R (Obenchain et al. 2014).

503 A covariance matrix computed from the genotype likelihoods of FFD SNPs at unlinked
504 positions (i.e., one per 10 Kb windows) was used for principal component analysis using

505 PCAngsd (Meisner and Albrechtsen 2018). To test for admixture, we run NgsAdmix (Skotte
506 et al. 2013) on genotype likelihoods at FFD unlinked sites. The number of clusters tested for
507 the admixture analysis ranged from K = 1 to K = 9. The seed for initializing the EM algorithm
508 was set to values ranging from 10 to 50 to test for convergence. Finally, the K best explaining
509 the variance observed in the data was evaluated using the Evanno method (Evanno et al. 2005)
510 in CLUMPAK (http://clumpak.tau.ac.il/bestK.html). Result plotting was performed using R.

511 For each population we estimated the average global Watterson's theta (θ_w) and average
512 pairwise nucleotide diversity (π), whereas to test for departures from mutation/drift equilibrium
513 we computed Tajima's D (Tajima 1989). Global average estimates of these statistics were
514 computed using ANGSD and custom bash scripts, implementing a sliding window approach
515 with windows of 50 Kb and a step of size 10 Kb.

516 We estimated between-population differentiation as F_{ST} for all pairs of populations at
517 high and low elevation respectively, as well as for pairs of ecotypes across localities. F_{ST}
518 statistics were carried out in ANGSD using the folded joint site frequency spectra (jSFS) for
519 all population pairs as summary statistics. Given that no suitable outgroup sequence was
520 available, the ancestral state was unknown. As a consequence, we observed a deviation from
521 the expected SFS for some populations (i.e. a high frequency of sites with fixed derived alleles)
522 when polarizing toward the major allele throughout the alpine populations. Therefore, we
523 produced site allele frequency likelihoods using ANGSD settings *-dosaf 1 -GL 2 -minQ 20 -P*
524 *8 -skipTriallelic 1 -doMajorMinor 1 - anc reference.genome.fasta,* limiting the analysis to the
525 set of FFD sites using the *-sites* option. Finally, we used the *-fold* option to fold the spectra
526 when using realSFS (for further analyses in ANGSD) or using a custom R script to fold the
527 spectra into fastsimcoal2 format (for coalescent simulations in fastsimcoal2).

528 **Testing alternative demographic scenarios.** We evaluated which demographic
529 scenario (1-origin vs 2-origins) explains our data using fastSimcoal2 v.2.6.0.3 (Excoffier et al.

530 2013). We tested four populations at a time, i.e. with two ecotype pairs in each simulation,
531 using for each analysis the jSFS for all six combinations of populations as summary statistics.
532 For all models we let the algorithm estimate the effective population size (N), the mutation rate
533 (μ) and the time of each split (T1, T2 and T3, Fig. 3). Although N, μ and the time of split
534 between ecotypes in each pair have been previously estimated by Trucchi et al (2017), we
535 started with broad search ranges for the parameters to not constrain a priori the model. The
536 final priors of the simulations were set for a mutation rate between 1e-8 and 1e-10, the effective
537 population size between 50 and 50,000 (alpine populations) and 50 and 5,000 (montane
538 populations), and for the time of each split between 1,000 and 100,000 generations ago. We
539 forced T1 to predate T2 and T3, and performed separate simulations setting T2 > T3 and T3 >
540 T2, respectively. For the models including gene flow, migration rate (m) between any pair of
541 demes was initially set to a range between 10e-10 and two.

542 The generation time in *H. pusillum* was reported to be 1 year (Flatscher et al. 2012;
543 Trucchi et al. 2017). While most populations in the montane zone flower during the first year
544 after germination, this is not the case in the alpine environment, where plants usually start to
545 flower in the second year after germination. Therefore, 1 year is most likely an underestimation
546 of the intergeneration interval, which is more realistically around 3 years. While this parameter
547 does not affect the overall results in terms of topology, it should be considered carefully in
548 terms of divergence times between ecotypes that were previously hypothesized to be post-
549 glacial (Flatscher et al. 2012; Trucchi et al. 2017).

550 FastSimcoal2 was run excluding monomorphic sites (*-0* option). We performed
551 200,000 simulations and ran up to 50 optimizations (ECM) cycles to estimate the parameters.
552 To find the global optimum of the best combination of parameter estimates, we performed 60
553 replicates of each simulation run. The MaxEstLhood is the maximum estimated likelihood
554 across all replicate runs, while the MaxObsLhood is the maximum possible value for the

555 likelihood if there was a perfect fit of the expected to the observed site frequency spectrum.
556 We report the difference between these two estimates (ΔL) for each model and ΔAIC scores
557 (i.e., the difference between the AIC for the best possible model and the tested model) to
558 compare models with different numbers of parameters. Finally, the parameter estimations of
559 the best run were used to simulate the expected jSFS and test the goodness of fit of the topology
560 plus parameter estimates to the observed data.

561 **Differential gene expression analysis.** Only unique read alignments were considered
562 to produce a table of counts using FeatureCounts v.1.6.3 (Liao et al. 2014) with the option *-t*
563 *gene* to count only reads mapping to genes. DE analyses were performed using the
564 Bioconductor package EdgeR v.3.24.3 (Robinson et al. 2010). The count matrix was filtered,
565 keeping only genes with mean counts per million (cpm) higher than 1. Data normalization to
566 account for library depth and RNA composition was performed using the weighted trimmed
567 mean of M-values (TMM) method. The estimateDisp() function of edgeR was used to estimate
568 the trended dispersion coefficients across all expressed tags by supplying a design matrix with
569 ecotype pair and ecotype information for each sample. We implemented a generalized linear
570 model (glm) to find gene expression differences between low and high elevation ecotypes by
571 taking into account the effects of the covariates ecotype and ecotype pair on gene expression.
572 A likelihood ratio test (lrt) was used to test for DE genes between ecotypes in each pair. The
573 level of significance was adjusted using Benjamini-Hochberg correction of p-values to account
574 for multiple testing (threshold of FDR < 0.05). The statistical significance of the overlaps
575 between lists of DEGs was tested using a hypergeometric test implemented in the Bioconductor
576 package SuperExactTest (Wang et al. 2015).

577 **Functional interpretation of DEG.** We performed separate GO terms enrichment
578 analysis for each ecotype pair, and gave special attention to functions that were shared among
579 lists of DEG. Fisher test statistics implemented in the Bioconductor package topGO v.2.34.0

580 (https://bioconductor.org/packages/release/bioc/html/topGO.html) were run with the algorithm
581 "weight01" to test for over-representation of specific functions conditioned on neighbouring
582 terms. Multiple testing correction of p-values (FDR correction) was applied and significance
583 was assessed below a threshold of 0.05. DEG were also explicitly searched for protein coding
584 genes and transcription factors underlying the formation of trichomes and visually checked
585 using R.

586 **Detection of multilocus gene expression variation.** To detect gene expression
587 changes underlying adaptive traits with a strongly polygenic basis we performed a cRDA of
588 the gene expression data using the R Package vegan v.2.5-6 (Oksanen et al. 2019). The cRDA
589 approach is well suited to identify groups of genes showing expression changes that covary
590 with the "ecotype" variable while controlling for population structure (Bourret et al. 2014;
591 Forester et al. 2018). As a table of response variables in the cRDA we used the matrix of read
592 counts after filtering using mean cpm higher than 1 as in the DE analysis. First, the cRDA
593 includes a multiple regression step of gene expression on the explanatory variable "ecotype".
594 In our case, the RDA was conditioned to remove the effects of the geographic ecotype pair
595 using the formula "~ ecotype + Condition(pair)". In the second step, a principal component
596 analysis (PCA) of the fitted values from the multiple regression is performed to produce
597 canonical axes, based on which an ordination in the space of the explanatory variable is
598 performed. The first axis of the cRDA therefore shows the variance explained by the
599 constrained variable "ecotype", while the second axis is the first component of the PCA nested
600 into the RDA, representing the main axis of unconstrained variance. The significance of the
601 cRDA was tested with ANOVA and 1,000 permutations. Each gene was assigned a cRDA
602 score that is a measure of the degree of association between the expression level of a gene and
603 the variable "ecotype". Outliers were defined as genes with scores above the significance

604 thresholds of ± 2 and, respectively, ± 2.6 standard deviations from the mean score of the
605 constrained axis, corresponding to p-value thresholds of 0.05 and 0.01, respectively.

606 **SNPs calling and detection of selection outliers.** In order to perform selection scans,
607 we performed variants calling following best standard practices implemented in GATK
608 v.4.1.8.1 (Van der Auwera and O'Connor 2020). Briefly, after sorting mapped files and
609 duplicates removal (see above), we used the split'N'trim function to split reads with Ns in the
610 CIGAR string and trim overhangs. We called variants using HaplotypeCaller with the option -
611 *ERC GVCF* and merged multiple samples in gvcf format using the GenomicsDBImport utility
612 with the *-L* option to operate in parallel on multiple genomic intervals. Finally, we used
613 GenotypeGVCFs to perform joint genotyping. The resulting vcf file was processed using the
614 vcfallelicprimitives modality implemented in vcflib v.1.0.2 (https://github.com/vcflib/vcflib)
615 with the options *--keep-info --keep-geno* to split indels into multiple SNPs. Filtering was
616 performed using vcftools v.0.1.16 (Danecek et al. 2011) with the options *--max-alleles 2 --min-*
617 *alleles 2 --minDP 4 --minGQ 20 --minQ 30 --remove-indels --max-missing 0.5*.

618 Selection outliers analyses were carried out on ecotype pairs 1 and 3, for which we had
619 a minimum of 10 individuals in each population analyzed. To detect outlier SNPs potentially
620 under divergent selection during ecotype adaptation to different elevations we followed the
621 method implemented in the R package pcadapt based on Principal Component Analysis (Luu
622 et al. 2017). After filtering the vcf file, we used vcftools to obtain two separate vcf files for pair
623 1 and 3 and retained loci with a minimum minor allele frequency 0.1. Plink (Purcell et al. 2007)
624 was used to convert vcf files into the plink binary bed format. We used the pcadapt() function
625 of pcadapt with the LD.clumping option to perform thinning using a window size of 200 SNPs
626 and a $r^2 = 0.1$ (default parameters). Outlier SNPs for ecotype divergence were defined as loci
627 showing a strong correlation with the first principal component, which in both pairs separates

the ecotypes (Fig. S8). P-values were corrected using the Bonferroni method and a threshold of adjusted $p < 0.05$ was chosen to assess significance.

Data accessibility. Scripts used to perform the analyses can be found under this gitHub repository https://github.com/aglaszuk.

Acknowledgements

This work was financially supported by the Austrian Science Fund (FWF): DK W1225-B20 and Y661-B16. We thank Nicholas Barton, Andrew Clark, Virginie Courtier-Orgogozo, Joachim Hermisson, Thibault Leroy, Magnus Nordborg, John Parsch, Christian Schlötterer, Emiliano Trucchi and Alex Widmer for insightful comments and feedback. We thank Juliane Baar, Marie Huber, Carles Ferré Ortega and Daniela Paun for their support during laboratory work and data acquisition, Marianne Magauer and Daniela Pirkebner for producing the selfed line, as well as Martina Imhiavan and Daniel Schlorhaufer from the Botanical Garden of the University of Innsbruck for the cultivation of the plants.

References

Agrawal AA. 2017. Toward a predictive framework for convergent evolution: integrating natural history, genetic mechanisms, and consequences for the diversity of life. *Am Nat.* 190:1–12.

Alves JM, Carneiro M, Cheng JY, Lemos de Matos A, Rahman MM, Loog L, Campos PF, Wales N, Eriksson A, Manica A, et al. 2019. Parallel adaptation of rabbit populations to myxoma virus. *Science* 363:1319–1326.

Arendt J, Reznick D. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol Evol.* 23:26–32.

Barghi N, Hermisson J, Schlötterer C. 2020. Polygenic adaptation: a unifying framework to understand positive selection. *Nat Rev Genet.* 21:769–781.

Barghi N, Tobler R, Nolte V, Jakšić AM, Mallard F, Otte KA, Dolezal M, Taus T, Kofler R, Schlötterer C. 2019. Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLoS Biol.* 17:e3000128.

Benítez M, Espinosa-Soto C, Padilla-Longoria P, Díaz J, Alvarez-Buylla ER. 2007. Equivalent genetic regulatory networks in different contexts recover contrasting spatial cell patterns that resemble those in *Arabidopsis* root and leaf epidermis: a dynamic model. *Int J Dev Biol*. 51:139-55.

Bertel C, Buchner O, Schönswetter P, Frajman B, Neuner G. 2016. Environmentally induced and (epi-)genetically based physiological trait differentiation between *Heliosperma pusillum* and its polytopically evolved ecologically divergent descendent, *H. veselskyi* (Caryophyllaceae: Sileneae). *Bot J Linn Soc.* 182:658–669.

Bertel C, Hülber K, Frajman B, Schönswetter P. 2016. No evidence of intrinsic reproductive isolation between two reciprocally non-monophyletic, ecologically differentiated mountain plants at an early stage of speciation. *Evol Ecol.* 30:1031–1042.

Bertel C, Schönswetter P, Frajman B, Holzinger A, Neuner G. 2017. Leaf anatomy of two reciprocally non-monophyletic mountain plants (*Heliosperma* spp.): does heritable adaptation to divergent growing sites accompany the onset of speciation? *Protoplasma* 254:1411–1420.

Bertel C, Rešetnik I, Frajman B, Erschbamer B, Hülber K, Schönswetter P. 2018. Natural selection drives parallel divergence in the mountain plant *Heliosperma pusillum* s.l. *Oikos* 127:1355–1367.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.

Bolnick DI, Barrett RDH, Oke KB, Rennison DJ, Stuart YE. 2018. (Non)Parallel Evolution. *Annu Rev Ecol Evol S.* 49:303–330.

Bourret V, Dionne M, Bernatchez L. 2014. Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: polygenic multilocus analysis surpasses genome scan. *Mol Ecol.* 23:4444–4457.

Boyle EA, Li YI, Pritchard JK. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169:1177–1186.

Buckley J, Widmer A, Mescher MC, De Moraes CM. 2019. Variation in growth and defence traits among plant populations at different elevations: Implications for adaptation to climate change. *J. Ecol.* 107:2478–2492.

Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome Annotation and Curation Using MAKER and MAKER-P. *Curr Protoc Bioinformatics* 48:4.11.1–39.

Chang J, Yu T, Yang Q, Li C, Xiong C, Gao S, Xie Q, Zheng F, Li H, Tian Z, et al. 2018. Hair, encoding a single C2H2 zinc-finger protein, regulates multicellular trichome formation in tomato. *Plant J.* 96:90–102.

Chan YF, Marks ME, Jones FC, Villarreal G Jr, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327:302–305.

Chopra D, Mapar M, Stephan L, Albani MC, Deneer A, Coupland G, Willing E-M, Schellmann S, Schneeberger K, Fleck C, et al. 2019. Genetic and molecular analysis of trichome development in *Arabis alpina*. *Proc Natl Acad Sci U S A.* 116:12078–12083.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G Jr, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307:1928–1933.

Cooper TF, Rozen DE, Lenski RE. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc Natl Acad Sci U S A.*100:1072–1077.

Cota-Sánchez JH, Remarchuk K, Ubayasena K. 2006. Ready-to-use DNA extracted with a CTAB method adapted for herbarium specimens and mucilaginous plant tissue. *Plant Mol Biol Rep.* 24:161–167.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.

Elmer KR, Fan S, Kusche H, Spreitzer ML, Kautt AF, Franchini P, Meyer A. 2014. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nat Commun.* 5:5168.

Elmer KR, Meyer A. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol Evol.* 26:298–306.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol.* 14:2611–2620.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet.* 9:e1003905.

Fang B, Kemppainen P, Momigliano P, Feng X, Merilä J. 2020. On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nat Ecol Evol.* 4:1105–1115.

Fischer EK, Song Y, Hughes KA, Zhou W, Hoke KL. 2021. Nonparallel transcriptional divergence during parallel adaptation. *Mol. Ecol.* 30: 1516–1530.

Flatscher R, Frajman B, Schönswetter P, Paun O. 2012. Environmental heterogeneity and phenotypic divergence: can heritable epigenetic variation aid speciation? *Genet. Res. Int.* 2012:698421.

Fong SS, Joyce AR, Palsson BØ. 2005. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res.* 15:1365–1372.

Forester BR, Lasky JR, Wagner HH, Urban DL. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Mol Ecol.* 27:2215–2233.

Fraïsse C, Roux C, Gagnaire P-A, Romiguier J, Faivre N, Welch JJ, Bierne N. 2018. The divergence history of European blue mussel species reconstructed from Approximate

Bayesian Computation: the effects of sequencing techniques and sampling strategies. *PeerJ*. 6:e5198.

Frajman B, Oxelman B. 2007. Reticulate phylogenetics and phytogeographical structure of *Heliosperma* (Sileneae, Caryophyllaceae) inferred from chloroplast and nuclear DNA sequences. *Mol Phylogenetics Evol.* 43:140–155.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36:3420–3435.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 8:1494–1512.

Hämälä T, Guiltinan MJ, Marden JH, Maximova SN, dePamphilis CW, Tiffin P. 2020. Gene expression modularity reveals footprints of polygenic adaptation in *Theobroma cacao*. *Mol Biol Evol.* 37:110–123.

Hermisson J, Pennings PS. 2017. Soft sweeps and beyond: Understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods Ecol Evol.* 8:700–716.

Hilscher J, Schlötterer C, Hauser M-T. 2009. A single amino acid replacement in ETC2 shapes trichome patterning in natural *Arabidopsis* populations. *Curr. Biol.* 19:1747–1751.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313:101–104.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32:767–769.

Höllinger I, Pennings PS, Hermisson J. 2019. Polygenic adaptation: From sweeps to subtle frequency shifts. *PLoS Genet.* 15:e1008035.

Hülskamp M. 2004. Plant trichomes: a model for cell differentiation. *Nat Rev Mol Cell Biol.* 5:471–480.

Hülskamp M, Misŕa S, Jürgens G. 1994. Genetic dissection of trichome cell development in *Arabidopsis*. *Cell* 76:555–566.

James ME, Arenas-Castro H, Groh JS, Engelstädter J, Ortiz-Barrientos D, unpublished data, https://www.biorxiv.org/content/10.1101/2020.02.05.936401v3, last accessed July 2, 2021.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:55–61.

Khosla A, Paper JM, Boehler AP, Bradley AM, Neumann TR, Schrick K. 2014. HD-Zip Proteins GL2 and HDG11 Have Redundant Functions in *Arabidopsis* Trichomes, and GL2 Activates a Positive Feedback Loop via MYB23. *Plant Cell* 26:2184–2200.

Knotek A, Konečná V, Wos G, Požárová D, Šrámková G, Bohutínská M, Zeisek V, Marhold K, Kolář F. 2020. Parallel Alpine Differentiation in *Arabidopsis arenosa*. *Front. Plant Sci.* 11:561526.

Konečná V, Nowak MD, Kolář F. 2019. Parallel colonization of subalpine habitats in the central European mountains by *Primula elatior*. *Sci. Rep.* 9:3294.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinform.* 15:356.

Lee KM, Coop G. 2019. Population genomics perspectives on convergent adaptation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374:20180236.

Lee SH, Cho H-T. 2006. PINOID Positively Regulates Auxin Efflux in *Arabidopsis* Root Hair Cells and Tobacco Cells. *Plant Cell* 18:1604–1616.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930.

Lim MCW, Witt CC, Graham CH, Dávalos LM. 2019. Parallel molecular evolution in pathways, genes, and sites in high-elevation hummingbirds revealed by comparative transcriptomics. *Genome Biol. Evol.* 11:1552–1572.

Lomsadze A, Burns PD, Borodovsky M. 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* 42:e119.

Losos JB. 2011. Convergence, adaptation, and constraint. *Evolution* 65:1827–1840.

Louis M, Galimberti M, Archer F, Berrow S, Brownlow A, Fallon R, Nykänen M, O'Brien J, Roberston KM, Rosel PE, et al., unpublished data, https://www.biorxiv.org/content/10.1101/2020.10.05.325159v12020, last accessed July 2, 2021.

Luu K, Bazin E, Blum MGB. 2017. pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Mol Ecol Resour.* 17:67–77.

Machado A, Wu Y, Yang Y, Llewellyn DJ, Dennis ES. 2009. The MYB transcription factor GhMYB25 regulates early fibre and trichome development. *Plant J.* 59:52–62.

Mandic M, Ramon ML, Gerstein AC, Gracey AY, Richards JG. 2018. Variable gene transcription underlies phenotypic convergence of hypoxia tolerance in sculpins. *BMC Evol. Biol.* 18:163.

Meirmans, P.G. 2019. Subsampling reveals that unbalanced sampling affects STRUCTURE results in a multi-species dataset. *Heredity* 122:276–287.

Meisner J, Albrechtsen A. 2018. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics* 210:719–731.

Nguyen Ba AN, Cvijović I, Rojas Echenique JI, Lawrence KR, Rego-Costa A, Liu X, Levy SF, Desai MM. 2019. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature* 575:494–499.

Nosil P, Feder JL, Flaxman SM, Gompert Z. 2017. Tipping points in the dynamics of speciation. *Nat Ecol Evol*. 1:0001.

Nosil P, Harmon LJ, Seehausen O. 2009. Ecological explanations for (incomplete) speciation. *Trends Ecol Evol.* 24:145–156.

Nowak MA, Boerlijst MC, Cooke J, Smith JM. 1997. Evolution of genetic redundancy. *Nature* 388:167–171.

Obenchain V, Lawrence M, Carey V, Gogarten S, Shannon P, Morgan M. 2014. VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics* 30:2076–2078.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, et al. 2019. vegan: Community Ecology Package. R package version 2.5-6. Retrieved from https ://CRAN.R-proje ct.org/packa ge=vegan

Pesch M, Hülskamp M. 2009. One, two, three…models for trichome patterning in *Arabidopsis*? *Curr Opin Plant Biol.* 12:587–592.

Projecto-Garcia J, Natarajan C, Moriyama H, Weber RE, Fago A, Cheviron ZA, Dudley R, McGuire JA, Witt CC, Storz JF. 2013. Repeated elevational transitions in hemoglobin function during the evolution of Andean hummingbirds. *Proc Natl Acad Sci U S A.* 110:20669–20674.

Puechmaille SJ. 2016. The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Mol Ecol Resour.* 16:608–627.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81:559–575.

Ravinet M, Westram A, Johannesson K, Butlin R, André C, Panova M. 2016. Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Mol Ecol.* 25:287–305.

Rellstab C, Zoller S, Sailer C, Tedder A, Gugerli F, Shimizu KK, Holderegger R, Widmer A, Fischer MC. 2020. Genomic signatures of convergent adaptation to Alpine environments in three Brassicaceae species. *Mol. Ecol.* 29:4250–4365.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.

Roda F, Ambrose L, Walter GM, Liu HL, Schaul A, Lowe A, Pelser PB, Prentis P, Rieseberg LH, Ortiz-Barrientos D. 2013. Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. *Mol. Ecol.* 22:2941–2952.

Rougeux C, Gagnaire P-A, Praebel K, Seehausen O, Bernatchez L. 2019. Polygenic selection drives the evolution of convergent transcriptomic landscapes across continents within a Nearctic sister species complex. *Mol. Ecol.* 28:4388–4403.

Scoville AG, Barnett LL, Bodbyl-Roels S, Kelly JK, Hileman LC. 2011. Differential regulation of a MYB transcription factor is correlated with transgenerational epigenetic inheritance of trichome density in *Mimulus guttatus*. *New Phytol.* 191:251–263.

Serna L, Martin C. 2006. Trichomes: different regulatory networks lead to convergent structures. *Trends Plant Sci.* 11:274–280.

Shi P, Fu X, Shen Q, Liu M, Pan Q, Tang Y, Jiang W, Lv Z, Yan T, Ma Y, et al. 2018. The roles of AaMIXTA1 in regulating the initiation of glandular trichomes and cuticle biosynthesis in *Artemisia annua*. *New Phytol.* 217:261–276.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.

Skotte L, Korneliussen TS, Albrechtsen A. 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics* 195:693–702.

Sloan DB, Keller SR, Berardi AE, Sanderson BJ, Karpovich JF, Taylor DR. (2011) *De novo* transcriptome assembly and polymorphism detection in the flowering plant *Silene vulgaris* (Caryophyllaceae). *Mol Ecol Resour.* 12:333–343.

Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, Buerkle CA, Feder JL, Bast J, Schwander T, et al. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. *Science* 344:738–742.

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

Steiner CC, Römpler H, Boettger LM, Schöneberg T, Hoekstra HE. 2009. The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. *Mol. Biol. Evol.* 26:35–45.

Stern DL. 2013. The genetic causes of convergent evolution. *Nat Rev Genet.* 14:751–764.

Strader LC, Culler AH, Cohen JD, Bartel B. 2010. Conversion of endogenous indole-3-butyric acid to indole-3-acetic acid drives cell expansion in *Arabidopsis* seedlings. *Plant Physiol.* 153:1577–1586.

Stuart YE, Veen T, Weber JN, Hanson D, Ravinet M, Lohman BK, Thompson CJ, Tasneem T, Doggett A, Izen R, et al. 2017. Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nat Ecol Evol* 1:158.

Sun Y-B, Fu T-T, Jin J-Q, Murphy RW, Hillis DM, Zhang Y-P, Che J. 2018. Species groups distributed across elevational gradients reveal convergent and continuous genetic adaptation to high elevations. *Proc Natl Acad Sci U S A.* 115:10634–10641.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

Tan Y, Barnbrook M, Wilson Y, Molnár A, Bukys A, Hudson A. 2020. Shared Mutations in a Novel Glutaredoxin Repressor of Multicellular Trichome Fate Underlie Parallel Evolution of *Antirrhinum* Species. *Curr Biol.* 30:1357–1366.

Temsch EM, Temsch W, Ehrendorfer-Schratt L, Greilhuber J. 2010. Heavy metal pollution, selection, and genome size: the species of the Žerjav study revisited with flow cytometry. *J Bot.* 2010:596542.

Therkildsen NO, Wilder AP, Conover DO, Munch SB, Baumann H, Palumbi SR. 2019. Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. *Science* 365:487–490.

Thiel-Egenter C, Alvarez N, Holderegger R, Tribsch A, Englisch T, Wohlgemuth T, Colli L, Gaudeul M, Gielly L, Jogan N, et al. 2010. Break zones in the distributions of alleles and species in alpine plants. *J Biogeogr.* 38:772–782.

Thompson KA, Osmond MM, Schluter D. 2019. Parallel genetic evolution and speciation from standing variation. *Evol Lett.* 3:129–141.

Tominaga-Wada R, Ishida T, Wada T. 2011. New Insights into the Mechanism of Development of *Arabidopsis* Root Hairs and Trichomes. *Int Rev Cell Mol. Biol.* 286:67–106.

Trucchi E, Frajman B, Haverkamp THA, Schönswetter P, Paun O. 2017. Genomic analyses suggest parallel ecological divergence in *Heliosperma pusillum* (Caryophyllaceae). *New Phytol.* 216:267–278.

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV. 2010. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.* 42:260–263.

Van Belleghem SM, Vangestel C, De Wolf K, De Corte Z, Möst M, Rastas P, De Meester L, Hendrickx F. 2018. Evolution at two time frames: Polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. *PLoS Genet.* 14:e1007796.

Van der Auwera GA, O'Connor BD. 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition). O'Reilly Media.

Velasquez SM, Barbez E, Kleine-Vehn J, Estevez JM. 2016. Auxin and Cellular Elongation. *Plant Physiol.* 170:1206–1215.

Wang M, Zhao Y, Zhang B. 2015. Efficient Test and Visualization of Multi-Set Intersections. *Sci Rep.* 5:16923.

Wilkens H, Strecker U. 2003. Convergent evolution of the cavefish *Astyanax* (Characidae, Teleostei): genetic evidence from reduced eye-size and pigmentation. *Biol J Linn Soc.* 80:545–554.

Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA, Smets P, Wang T, Gray LK, et al. 2016. Convergent local adaptation to climate in distantly related conifers. *Science* 353:1431–1433.

Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. 2012. Parallel Molecular Evolution in an Herbivore Community. *Science* 337:1634–1637.

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29:2669–2677.

Figures

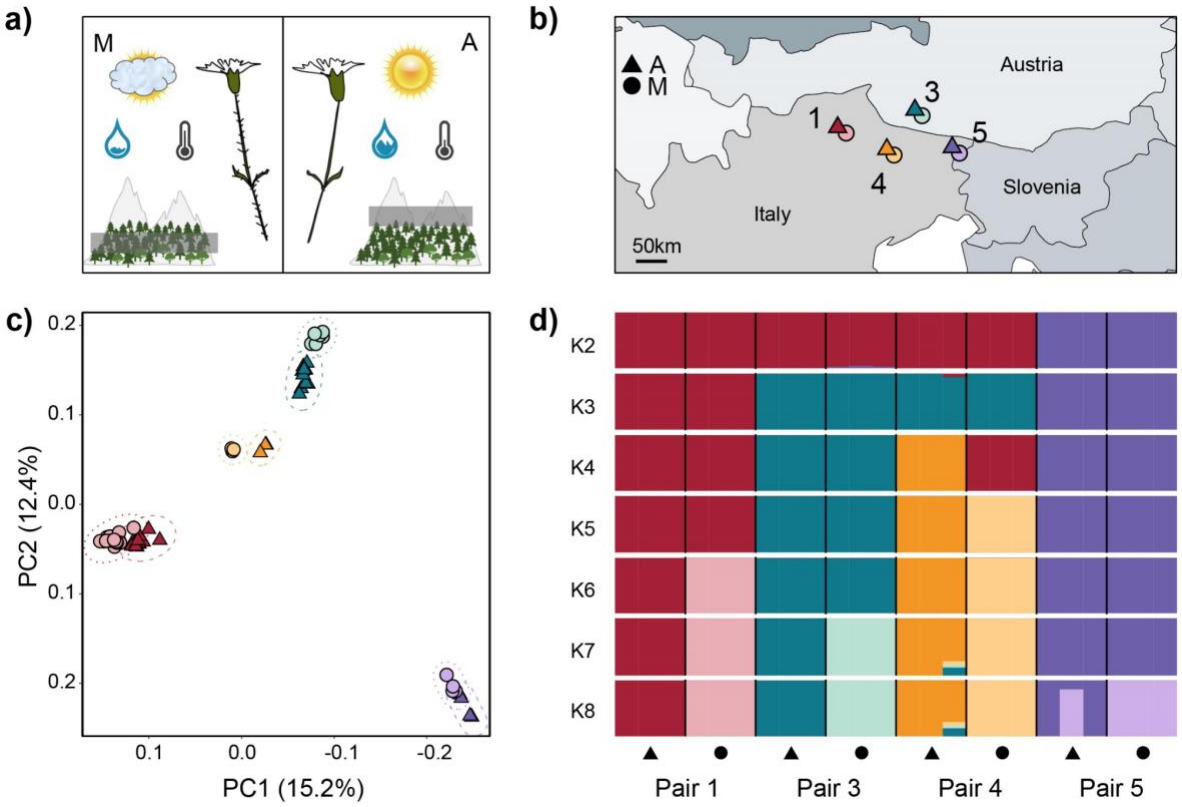

Figure 1. Study system, sampling setup, and genetic variation among four montane (M) - alpine (A) ecotype pairs of *Heliosperma pusillum*. Color coding of populations is consistent across panels. The numbering of the ecotype pairs is consistent with a previous publication on the same system (Bertel et al. 2018). **(a)** Graphic description of the main ecological and morphological differences between the ecotypes. **(b)** Geographic map showing the location of the analyzed populations in the southeastern Alps. **(c)** Clustering of individuals along the first two vectors of a principal component analysis. **(d)** Bar plot showing the assignment of individuals to the clusters identified by NgsAdmix for K = 2 through 8.

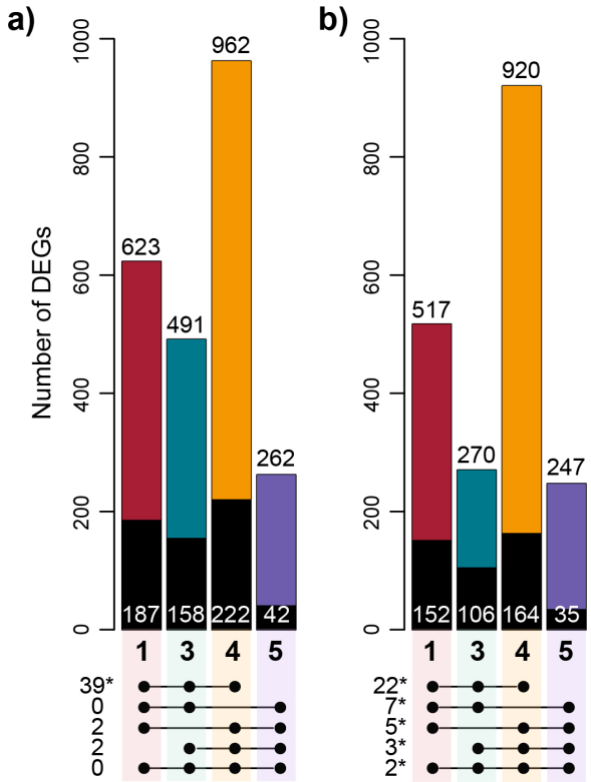

Figure 2. Differentially expressed genes (DEGs) at each ecotype pair show low overlap across different pairs. Colors represent the ecotype pair as in Fig. 1. Histograms show the number of DEGs (FDR < 0.05) underexpressed **(a)** and overexpressed **(b)** in the montane ecotype compared to the alpine in each pair. Colored and black areas of the bars show the amount of DEGs unique to each ecotype pair and, respectively, shared with at least one other pair. Numbers reported on top of the bars show the total amount of DEGs between ecotypes per pair and category. Numbers on the black areas show the amount of DEGs shared with at least one other pair. Linked dots below bars show the amount of shared DEGs between at least three pairs. Stars indicate that the overlap is significantly higher than chance expectations (hypergeometric test, $p < 0.01$).

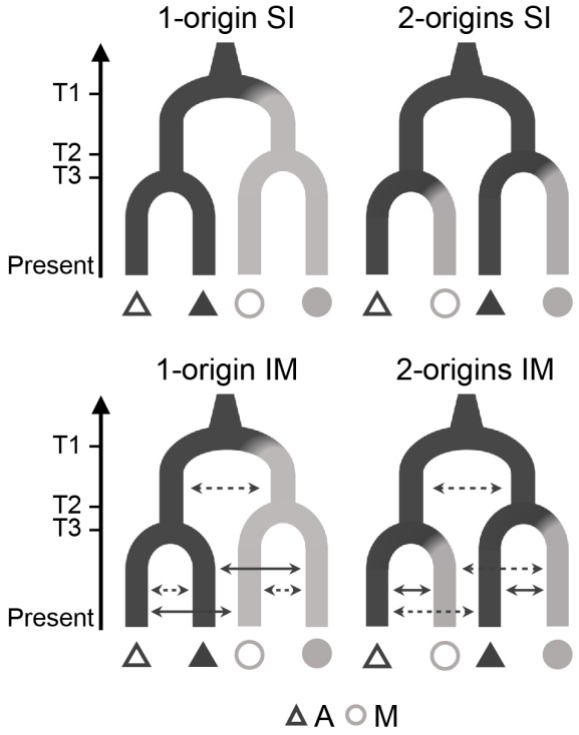

Figure 3. Alternative topologies tested using Fastsimcoal2 for all combinations of two ecotype pairs. Strict isolation (SI, upper panels) and isolation with migration (IM, lower panels) were modeled. Solid arrows in the IM models indicate higher migration rates expected between ecotypes at each locality according to population structure results. Divergence times T2 and T3 were allowed to vary (i.e., T2 > T3 but also T3 > T2 were modeled), whereas T1 was always the oldest event. Triangles and circles represent populations of the alpine (A) and the montane (M) ecotype, respectively. Filled and empty symbols represent different ecotype pairs.

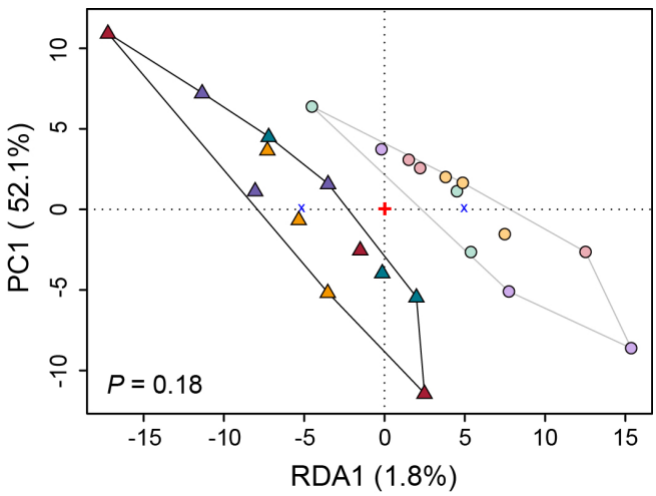

Figure 4. Expression divergence between accessions of the alpine and the montane ecotypes captured with conditioned redundancy analysis (cRDA). Colors represent the populations as in Fig. 1. Grey and black clusters correspond to montane and alpine ecotypes, respectively. The ANOVA test of the full model was not significant (p = 0.18), confirming that most expression differences between ecotypes in our dataset do not follow consistent routes across ecotype pairs.

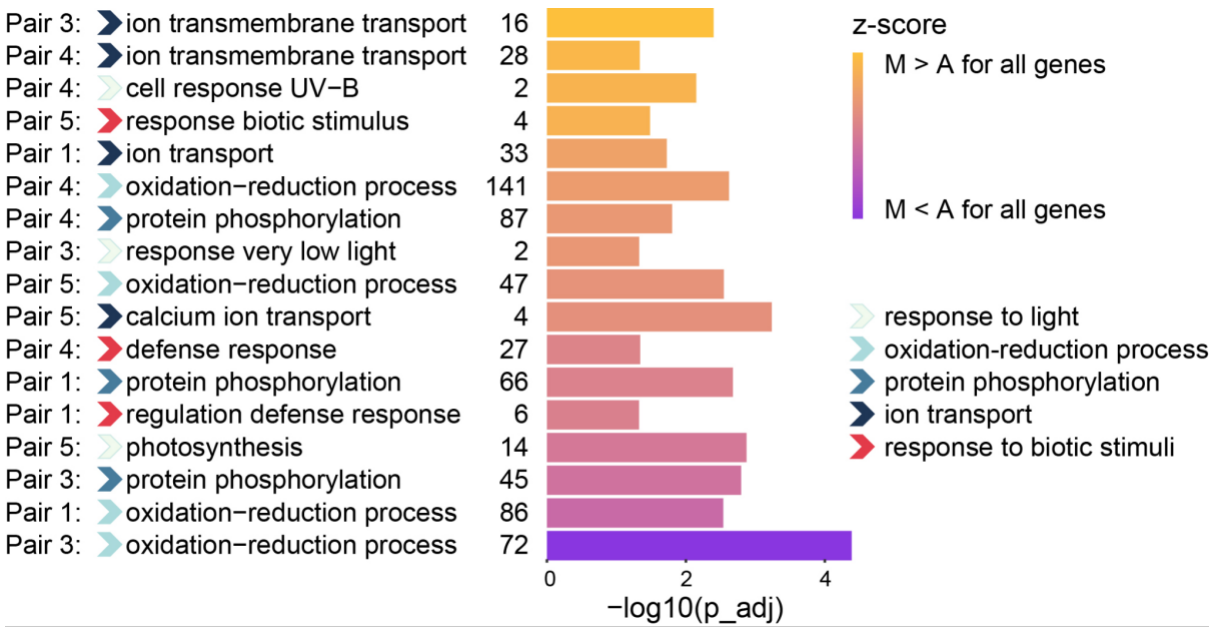

Figure 5. Functional enrichment of differentially expressed genes (DEGs) showing that across ecotype pairs similar biological processes are linked to adaptation to the different elevations. Height of the bars shows the significance of the enriched GO terms (Fisher's test). Numbers left of the bars show the number of DEGs underlying the corresponding GO term. Descriptions of GO terms were shortened for visual clarity. The ecotype pair in which a certain term is found to be enriched is specified on the left side of the plot. The z-score (color scale of the bars) was computed based on the log fold-change of gene expression, whereas positive and negative values show over- and underexpression in the montane ecotype respectively.

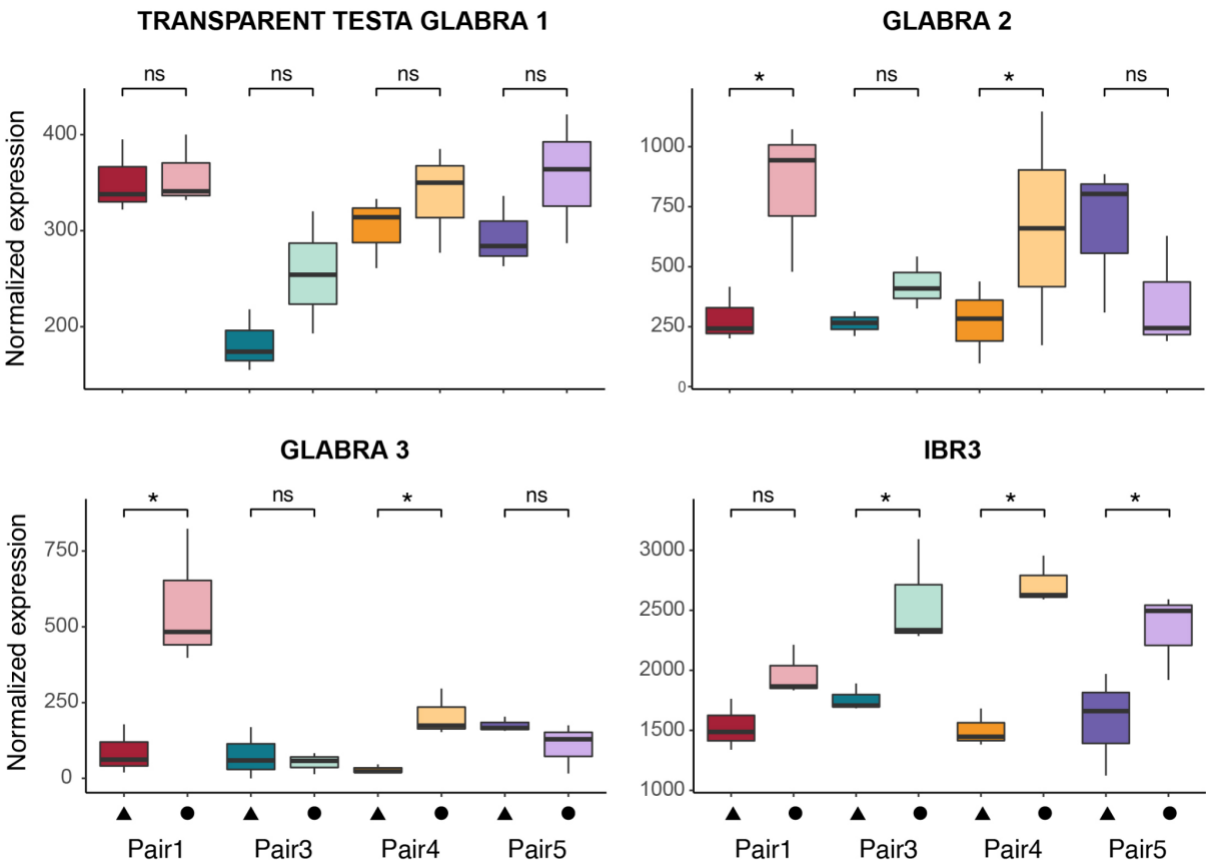

Figure 6. Examples of expression of genes known to be related to trichome formation and elongation in plants. Colors represent the populations as in Fig. 1. Stars indicate significant differential expression ($p < 0.05$) before false discovery rate (FDR) correction. Non-significant differences are marked with ns.