A framework to quantify phenotypic and genotypic parallel evolution

James, M. E.[1], Wilkinson, M. J.[1], North, H. L[1,2], Engelstädter, J.[1], and D. Ortiz-Barrientos[1].

[1]The University of Queensland, School of Biological Sciences, St. Lucia QLD 4072, Australia. [2]Current address: University of Cambridge, Department of Zoology, Downing St., Cambridge CB2 14 3EJ, UK.

Abstract

The independent and repeated adaptation of populations to similar environments often results in the evolution of similar forms. This phenomenon creates a strong correlation between phenotype and environment and is referred to as 'parallel evolution.' However, there is ongoing debate as to when we should call a system either phenotypically or genotypically 'parallel.' Here, we suggest a novel and simple framework to quantify parallel evolution at the genotypic and phenotypic levels. We apply this framework to coastal ecotypes of an Australian wildflower, *Senecio lautus*. Our findings show that *S. lautus* populations inhabiting similar environments have evolved strikingly similar phenotypes. These phenotypes have arisen via mutational changes affecting common biological functions but occurring in different genes. Our work not only provides a common framework to study the repeated evolution of forms in nature, but also highlights *S. lautus* as an excellent system to study parallel evolution.

Introduction

The independent and repeated adaptation of populations to similar environments often results in the evolution of similar phenotypes (Schluter, 2000) and is often seen in a wide variety of animal taxa (e.g., Johannesson *et al.*, 1993; Nosil *et al.*, 2002; Elmer *et al.*, 2010; Ravinet *et al.*, 2013; Soria-Carrasco *et al.*, 2014; Perreault-Payette *et al.*, 2017) and in some plants (e.g., Foster *et al.*, 2007; Trucchi *et al.*, 2017; Cai *et al.*, 2019; Konečná *et al.*, 2019). When independent populations evolve from similar initial conditions, this phenomenon is referred to as 'parallel evolution' (Schluter & Nagel, 1995; Bolnick *et al.*, 2018), The correlation that arises between phenotype and environment during parallel evolution provides strong evidence for the role of natural selection in creating new forms. This is because it is unlikely that similar phenotypes would have evolved multiple times purely by chance (Lenormand *et al.*, 2009 but see Losos, 2011). Systems of parallel evolution are unique as they provide natural replicates of the evolutionary process, enabling researches to study the genetics and ecology of how adaptation proceeds in nature (Lenormand *et al.*, 2016).

The evolution of similar phenotypes within the same environment can arise via independent and repeated selection on the same nucleotide site or gene (reviewed in Wood *et al.*, 2005; Christin *et al.*, 2010; Stern, 2013). Similar phenotypes can also arise by selection on entirely different genes, although often from the same functional pathway (e.g., Smith & Rausher, 2011; Kowalko *et al.*, 2013; Roda *et al.*, 2013b; Laporte *et al.*, 2015; Perreault-Payette *et al.*, 2017; Cassin-Sackett *et al.*, 2019). Here, different genetic routes are able to produce similar phenotypic outcomes across populations, even though selection does not act upon the same allele or gene. Despite progress, we remain largely ignorant as to what type of genetic architecture is the most common. There is also controversy as to which level of biological organisation (nucleotide site, gene, or biological function) should be viewed as genotypically 'parallel'. This makes it difficult to conceptualise parallel evolution across both the phenotype and genotype.

The debate surrounding the term *parallel evolution* remains sprightly but confusing (Haas & Simpson, 1946; Arendt & Reznick, 2008; Lenormand *et al.*, 2016; Stuart, 2019; Thompson *et al.*, 2019). Some researchers argue that *parallel* systems arise only when the same nucleotide site or gene is repeatedly involved in the evolution of similar phenotypes (Rosenblum *et al.*, 2014). However, others maintain that when similar phenotypes arise from the contribution of mutations in different genes, they should still be classed as cases of parallel evolution (Lim *et*

al., 2019). The term *parallel* has sometimes been disregarded, and *convergence* has replaced it (Arendt & Reznick, 2008). Nonetheless, it is generally accepted that *convergence* refers to the evolution of similar forms from distantly related taxa, and *parallel* is the evolution of similar forms from similar initial conditions in closely related taxa – a terminology that I am also adopting here (Speed & Arbuckle, 2017; Bolnick *et al.*, 2018 but see Lenormand *et al.*, 2016). To reduce the ensuing confusion when the term *parallel evolution* is used in isolation, some researchers have proposed we should separate *parallel evolution*, or *parallelism*, into two distinct components: the phenotype and the genotype (Elmer & Meyer, 2011).

Despite separating parallel evolution into phenotypic and genotypic components, researches often implement different approaches to measure parallelism within each study system. However, different interpretations of parallel phenotypic and genotypic evolution rely on different statistical approaches (Bolnick *et al.*, 2018). Consequently, it remains difficult to directly compare between studies, possibly obfuscating the processes creating variation in the origin and maintenance of adaptations and our understanding of parallelism. We propose a novel and simple framework to understand parallel evolution (Figure 1) where we show how measures of parallelism reveal how natural selection creates patterns at different scales of phenotypic and genotypic variation.

A framework to measure parallel evolution

To characterise the extent of phenotypic and genotypic parallelism within a system, it is necessary to rigorously demonstrate that populations adapting to similar environments (collectively referred to as an 'ecotype') have arisen multiple times independently. We refer the reader to our previous analyses of parallel evolution in *Senecio lautus* (James *et al.*, 2020) and to systems such as the marine snail, *Littorina saxatilis* (Quesada *et al.*, 2007; Johannesson *et al.*, 2010; Bierne *et al.*, 2013; Butlin *et al.*, 2014; Pérez-Pereira *et al.*, 2017), and threespine stickleback, *Gasterosteus aculeatus* (Colosimo *et al.*, 2005; Chan *et al.*, 2010; Dean *et al.*, 2019; Marques *et al.*, 2019) where one can find some of the strongest evidence for the independent origin of populations, and to the increasing number of potential cases of parallel evolution in plants (Foster *et al.*, 2007; Ostevik *et al.*, 2012; Trucchi *et al.*, 2017; Cai *et al.*, 2019; Konečná *et al.*, 2019).

81 *Phenotypic parallel evolution*

82 Researchers of parallelism traditionally ask to what extent ecotypes are phenotypically
83 distinct and which traits contribute to these differences (see Bolnick *et al.*, 2018) for a
84 detailed review of approaches). More recently, studies of parallelism (e.g., Elmer *et al.*, 2014;
85 Kusche *et al.*, 2015; Oke *et al.*, 2017; Stuart *et al.*, 2017; Jacobs *et al.*, 2018; Pilakouta *et al.*,
86 2019) have shifted to also quantify phenotypic similarities between each replicate pair using
87 geometric approaches such as phenotypic change vector analysis (Collyer & Adams, 2007;
88 Adams & Collyer, 2009; Collyer *et al.*, 2015). Unfortunately, geometric approaches do not
89 quantify overall differences between ecotypes, so replicate pairs can show little parallelism
90 yet contribute to strong parallel phenotypic differences between ecotypes. To maximise the
91 benefits of both methods we propose to quantify phenotypic and genotypic parallelism with
92 multiple approaches, which can be classified at broad- and narrow-sense scales (Figure 1).

93 In the broad-sense, phenotypic parallelism quantifies the net contribution of all replicate
94 populations within each ecotype to the overall phenotypic differences between ecotypes
95 (Figure 1). It can be broken into two components: 'parallel ecotype', and 'parallel trait'.
96 'Parallel ecotype' describes the multivariate differences between ecotypes, such that a highly
97 parallel system would show no overlap in the trait space occupied by each ecotype. 'Parallel
98 trait' describes which univariate traits are different between ecotypes such that they can
99 easily predict the habitat in which an individual lives, from trait attributes alone. In the
100 narrow-sense, phenotypic parallelism quantifies and compares similarities between each
101 replicate pair (which we have called 'parallel pair'). Narrow-sense parallelism gives insight
102 into which pairs are most similar to each other, and which pairs deviate from the common
103 trend within the system (Figure 1). These levels of phenotypic parallelism reveal both the net
104 contribution of replicate pairs to the general differences between ecotypes (broad-sense), as
105 well as the relative similarities in divergence between each replicate population (narrow-
106 sense).

107 *Genotypic parallel evolution*

108 Genotypic parallelism often measures the proportion of shared outlier loci between replicate
109 populations to ask whether the same alleles or genes are responsible for the evolution of
110 similar phenotypes (Wood *et al.*, 2005). In most cases, this will be difficult because causal
111 alleles might not be sampled (Catchen *et al.*, 2017; Lowry *et al.*, 2017; McKinney *et al.*,

112 2017), only their indirect linkage effects might be detected by outlier loci (Hoban *et al.*,
113 2016), and random fixations can accumulate between populations over time as genetic drift
114 increases dissimilarity between population pairs. If knowledge of causal alleles is available,
115 then one must demonstrate that a given site has been under repeated and independent
116 selection in each replicate pair (Lee & Coop, 2017, 2019). Typically, these analyses are
117 undertaken comparing individual replicate pairs (e.g., Roda *et al.*, 2013b; Lamichhaney *et al.*,
118 2017; Cassin-Sackett *et al.*, 2019), although some studies detect broad, or 'global' outliers by
119 comparing the aggregate of all populations within each ecotype (e.g., Jones *et al.*, 2012;
120 Kautt *et al.*, 2012). In the absence of repeated selection on the same allele or gene,
121 parallelism can be manifested at the functional level, such that different genes under selection
122 participating in the same predicted biological function contribute to the pattern of phenotypic
123 parallelism in a system. It is frequently asked whether predicted biological functions are
124 consistently enriched across independent populations (Smith & Rausher, 2011; Kowalko *et*
125 *al.*, 2013; Roda *et al.*, 2013b; Perreault-Payette *et al.*, 2017; Cassin-Sackett *et al.*, 2019).
126 These analyses can inform us about how genotypically parallel a system is at different scales
127 of divergence, so we can also view genotypic parallel evolution as broad-sense or narrow-
128 sense (Figure 1).

129 Broad-sense genotypic parallelism quantifies the net contribution of all replicate populations
130 within each ecotype to the overall genetic differences between ecotypes (Figure 1). We can
131 measure it at different levels of biological organisation: the nucleotide site, gene, and
132 biological function. For instance, a highly parallel nucleotide site would clearly differentiate
133 the two ecotypes. Such parallel alleles might have originated repeatedly and independently
134 gone to fixation in each population inhabiting similar habitats, or might have recurrently
135 increased in frequency from standing genetic variation. Broad-sense parallelism at the level
136 of the biological function asks which, if any, predicted functions are enriched with divergent
137 genes between the ecotypes. An enriched function might suggest it has been repeatedly
138 involved in adaptation across replicate populations.

139 On the other hand, narrow-sense genotypic parallelism quantifies which outlier nucleotide
140 sites and genes are the most similar across replicate pairs, and how similar the enriched
141 biological functions are for each pair (Figure 1). It asks which genetic differences distinguish
142 the ecotypes of each replicate pair and at what level of organisation. Patterns of genotypic
143 divergence can be complex and create a mismatch between genotypic and phenotypic

144 parallelism. For instance, all replicate pairs might have biological functions that diverged in
145 parallel, but in each case recruiting different genes. This will lead to lack of genotypic
146 parallelism at the nucleotide site and gene level both in the broad- and narrow-sense. Such a
147 system can still be highly parallel at the phenotypic level. We expect this to be common when
148 parallel evolution occurs on complex traits via polygenic adaptation (Gompel &
149 Prud'homme, 2009; Ralph & Coop, 2010; Rosenblum *et al.*, 2014; Yeaman, 2015;
150 MacPherson & Nuismer, 2017).

151 Our framework to measure parallel evolution synthesises traditional and new approaches to
152 quantify both phenotypic and genotypic parallelism. By categorising analyses into the broad-
153 and narrow-sense scales, researchers gain information about not only the overall patterns
154 within a system but also the specific similarities of replicate pairs. Our framework is useful
155 for systems with the repeated and independent origin of populations, where pairs of ecotypes
156 have adapted to contrasting environments. Nevertheless, our framework can still be used for
157 single origin systems containing multiple secondary contacts with gene flow between
158 ecotypes, or species. Here, researchers can still examine the dynamics of repeated population
159 divergence and parallel maintenance of species differences in the face of gene flow, but
160 cannot interpret populations as being from independent origins. Here, we examine the extent
161 of phenotypic and genotypic parallel evolution within an Australian wildflower species
162 complex, *Senecio lautus*. We use both broad- and narrow-sense measures and demonstrate
163 the importance of quantifying parallelism at these different scales of divergence.

164 **The *Senecio lautus* species complex**

165 The *Senecio lautus* species complex contains a variety of ecotypes adapted to contrasting
166 environments. The Dune and Headland ecotypes are of particular interest as they consist of
167 multiple parapatric Dune-Headland population pairs along the Australian coastline and are
168 often sister taxa in the phylogeny (Roda *et al.*, 2013a; Melo *et al.*, 2019; James *et al.*, 2020).
169 Despite the close geographic proximity between populations of a pair (i.e. ecotypes within
170 each locality), there is little to no gene flow between them, suggesting a large number of
171 independent and repeated origins (see James *et al.*, 2020). There is a strong association
172 between overall morphology and habitat in this coastal system: Dune plants are erect with
173 few branches, colonising the sandy dunes, whereas Headland individuals grow on rocky
174 headlands and are prostrate with many branches (Figure 2D). Populations maintain their
175 phenotypes when grown in common garden conditions, suggesting that phenotypic plasticity

176 within the system is weak (Walter *et al.*, 2016, 2018a; Wilkinson *et al.*, 2019). Previous work
177 with *S. lautus* in common garden conditions has identified a suite of divergent traits between
178 Dune and Headland populations, which include characteristics related to plant architecture
179 and leaf morphology (Walter *et al.*, 2018a). However, we lack a comprehensive
180 characterisation of how parallel the phenotypes and genotypes are within natural populations,
181 and how this affects divergence at the level of the ecotype and replicate pair.

182 To assess the extent of phenotypic and genotypic parallelism in *S. lautus*, we use 22 Dune
183 and Headland populations (including nine replicate Dune-Headland pairs, two allopatric
184 Dune populations, and two allopatric Headland populations). We first quantify how
185 phenotypically distinct the Dune and Headland ecotypes are, and which replicate pairs are the
186 most phenotypically similar. We then expand previous work (Roda *et al.*, 2013b) to ask
187 whether similar genetic mechanisms underly these repeated phenotypes (i.e. repeated
188 selection on the same nucleotide site, gene, or biological function). In addition, we ask
189 whether the variation in the extent of parallelism can be attributed to non-stochastic
190 processes, including levels of gene flow and within ecotype environmental variation.

191 **Methods**

192 **Phenotypic parallelism**

193 To quantify the extent of phenotypic parallelism within *S. lautus*, we measured a suite of
194 plant architecture and leaf traits from 20 Dune and Headland populations along the coast of
195 Australia ($n_{mean} = 30$, $n_{total} = 605$; Figure 2A, 2B; Table S1). These populations include eight
196 Dune-Headland pairs along a continuum from geographically parapatric to allopatric (four of
197 which are sister taxa, inferred from previous phylogenetic analyses) as well as four allopatric
198 populations that do not have a parapatric pair. We sampled mature (flowering) plants evenly
199 across the geographic range of each population, ensuring that each plant was more than one
200 metre apart. We measured six architecture traits (vegetative height, widest width, narrowest
201 width, main stem angle, main stem diameter, and secondary branch angle) and seven leaf
202 traits (area, perimeter, width, height, elongation, compactness, dissection and circularity;
203 Table S2). All architectural traits were measured in the field, and we sampled three secondary
204 branch leaves per plant for leaf morphometric analysis in ImageJ v1.51 (Schneider *et al.*,
205 2012). Leaves were scanned at 600 dpi on a CanoScan 9000F scanner and ImageJ was used
206 to automatically extract leaf shape characteristics. The same populations and traits were

measured in the glasshouse, and are discussed in (Walter, 2016) and (Wilkinson, 2019): overall, phenotypic variation and differences in the wild are highly correlated with those measured under controlled conditions.

In ~11% of sampled plants, we were unable to measure all six architectural traits (such as main stem diameter and angle). In these cases, we took the average of the population to impute the trait value for that individual. We ran the below analyses with and without these individuals and obtained consistent results. We report the analyses undertaken using the population means for the missing data. All phenotypic analyses were undertaken in R v3.4.2 (R Core Team, 2017). Traits were log transformed and standardised to have a mean of 0 and standard deviation of 1. We calculated pairwise correlations between all traits and removed four traits with high correlations across all populations (>0.8; Table S2). These correlated traits added minimal additional phenotypic information and are thus effectively redundant.

Relative contributions of broad- and narrow-sense phenotypic parallelism

We initially assessed the relative contributions of broad- and narrow-sense phenotypic parallelism within *S. lautus*. More specifically, we performed linear models at the multivariate level of the ecotype (MANOVA: *traits* = *ecotype* + *pair* + *ecotype x pair*), as well as the univariate level of each trait (ANOVA: *trait* = *ecotype* + *pair* + *ecotype x pair*) and extracted the partial effect sizes (partial η^2; Langerhans & DeWitt, 2004) for each term in the ANOVAs, (and Wilk's partial η^2 for the MANOVA) using the *etasq* function in the *heplots* package (Fox *et al.*, 2018) in R. As these models require population pairs, we excluded two Headland allopatric populations (H03 and H07) and two Dune allopatric populations (D09 and D35). The partial effect size of the *ecotype* term denotes how much of the phenotypic variation is explained by the overall differences between ecotypes, whereas the *pair* and *interaction* terms indicate how much variation is unique to replicate pairs. Therefore, when the effect size for the *ecotype* term is larger than the *pair* or *interaction*, this indicates that broad-sense phenotypic parallelism predominates over narrow-sense. After assessing these relative contributions (at the level of the ecotype and trait), we further explored the detailed patterns of broad- and narrow-sense phenotypic parallelisms within *S. lautus*.

236 **Broad-sense phenotypic parallelism**

237 **Parallel ecotype:** To ask whether ecotypes are phenotypically distinct within multivariate
238 space, we performed a one-way MANOVA (MANOVA: *traits* = *ecotype*) across the 20
239 Dune and Headland populations, where *traits* denotes the multivariate response variable of all
240 traits, and *ecotype* is a fixed effect of Dune or Headland. We also split traits into an
241 architectural and a leaf trait-set to ask whether phenotypic differences between ecotypes
242 depend on the trait category. To ask whether we can predict the ecotype each individual
243 belongs to, based on their phenotype, we performed K-means clustering with the Hartigan-
244 Wong algorithm (Hartigan & Wong, 1979). We used 25 random initial configurations and
245 retained the run with the smallest sums of squares of the individuals to their assigned cluster
246 centre, and then calculated the proportion of individuals assigned to their correct ecotype. We
247 also performed a linear discriminant analysis across all traits to ask which linear trait
248 combination best explains the phenotypic differences between Dune and Headland ecotypes.

249 **Parallel trait:** To ask which univariate traits are the most parallel between the Dune and
250 Headland ecotypes we undertook three approaches: vote counting, trait-by-trait linear
251 models, and R^2 analysis. As the vote counting and linear model approaches require
252 population pairs, we only included population pairs for these analyses (i.e. we again excluded
253 the two Headland allopatric populations (H03 and H07) and two Dune allopatric populations
254 (D09 and D35). We excluded two Headland allopatric populations (H03 and H07) and two
255 Dune allopatric populations (D09 and D35) for the following analyses. For vote counting, we
256 calculated the mean trait value for the Dune and Headland of each replicate pair and asked
257 whether there was a consistent increase or decrease in the trait value all replicate pairs (two-
258 sided dependent-samples sign-tests). However, this vote counting approach ignores trait
259 effect-size, and also has low statistical power when the sample size (number of replicate
260 pairs) is small. Therefore, we also used trait-by-trait linear models (*trait* = *ecotype* + *pair* +
261 *ecotype x pair*) to ask whether there was a significant main effect of ecotype for each trait. To
262 also quantify the extent of phenotypic parallelism between ecotypes, we extracted the
263 proportion of phenotypic variance explained by the ecotype term (R^2) for each trait from one-
264 way ANOVAs performed on the population means (see Langerhans, 2018). Following
265 Langerhans (2018), we considered a trait strongly parallel when $R^2 \geq 0.50$, moderately
266 parallel when $0.50 > R^2 \geq 0.33$, and weakly parallel when $R^2 < 0.33$. We then examined
267 which traits were consistently viewed as parallel across each of the three approaches.

268 **Narrow-sense phenotypic parallelism**

269 **Parallel pair:** To quantify the specific phenotypic similarities and differences between
270 replicate population pairs, we used Phenotypic Change Vector Analysis (PCVA; Collyer &
271 Adams, 2007; Adams & Collyer, 2009; Collyer *et al.*, 2015). Within multivariate phenotypic
272 space, PCVA quantifies both 1) the amount of divergence, and 2) the contribution of traits to
273 divergence between replicate pairs. The procedure is as follows: the phenotypic centroid
274 (multivariate mean) is calculated per population. For each population pair, their centroids are
275 connected with a vector. The length (L) of this vector quantifies how divergent the two
276 populations are – the longer the length, the more divergent. The difference in length (ΔL)
277 between vectors thus denotes the difference in the magnitude of divergence between two
278 replicate population pairs. The two pairs are considered parallel with regards to the
279 magnitude of their divergence if ΔL is not statistically different from zero (Bolnick *et al.*,
280 2018).

281 The contribution of traits to divergence is measured by the angle between vectors (θ). A large
282 angle between two pairs ($\theta >> 0°$) suggests the traits contributing to population divergence
283 are quite different between the pairs. The contribution of traits is considered parallel when the
284 angle is not statistically different from zero ($\theta \approx 0°$). Using R code modified from Collyer &
285 Adams (2007), we calculated ΔL and θ for all pairwise comparisons between replicate pairs
286 and performed permutations to test for statistical significance (see *Supplementary R code* for
287 details). To ensure this analysis was robust and not dominated by a single trait, we repeated
288 the calculations of ΔL and θ an additional nine times with removing a single trait each time.
289 We observed consistent results across all calculations, suggesting our results are not
290 dominated by a single trait.

291 **Genotypic parallelism**

292 To quantify the extent of genotypic parallelism within *Senecio lautus*, we collected leaf
293 samples from nine population pairs along the coast of Australia ($n_{mean} = 56$, $n_{total} = 1009$;
294 Figure 2A, 2B; Table S1). DNA was extracted using a modified CTAB protocol (Clarke,
295 2009). We created reduced representation libraries by using a two-enzyme Genotyping-by-
296 Sequencing (GBS) approach (modified from Poland *et al.*, 2012), with Pst1 and Msp1. We
297 created seven equimolar pools (192 individuals per pool), which were size-selected on the
298 BluePippin (2% DF Marker V1, 300-500bp; Sage Science). Pooled libraries were sent to

299 Beijing Genomics Institute (BGI) for sequencing on seven lanes of the HiSeq4000, with
300 100bp paired-end sequencing.

301 BGI removed forward barcodes and quality filtered the raw reads to remove reads containing
302 Illumina adaptors, low quality reads (> 50% of bases < Q10), and reads with > 10% Ns. We
303 trimmed reverse barcodes with *TagCleaner* standalone v0.12 (Schmieder *et al.*, 2010). Reads
304 were mapped to the *S. lautus* reference PacBio genome v1.0 (Wilkinson, 2019) with *BWA-*
305 *MEM* v0.7.15 (Li & Durbin, 2009; Li, 2013). *PicardTools* v2.7.0 (Broad Institute, 2019) was
306 used to clean aligned reads and to add read groups (PCR duplicates were not marked for
307 removal). We jointly called all variant and invariant sites for each population with *FreeBayes*
308 v1.1.0 (Garrison & Marth, 2012). Using *VCFtools* v0.1.15 (Danecek *et al.*, 2011), we
309 retained sites if they were present in > 50% of individuals, had a minimum quality score of
310 30, and a minimum minor allele count of 1. We then filtered for a minimum depth of 3 for a
311 genotype call. Individuals were removed if they contained > 40% missing data. We then
312 filtered for a maximum mean depth of 100, and a minimum mean depth of 10. We filtered for
313 missing data per population, removing sites if they contained > 50% of missing data within
314 each population. We refiltered for an overall missing data of 20%. Indels were removed with
315 *vcflib* (Garrison, 2016). We then filtered for population-specific Hardy Weinberg Equilibrium
316 using the *filter_hwe_by_pop.pl* script within *dDocent* (Puritz *et al.*, 2014). We filtered for an
317 overall minor allele count of 5, retaining 9,686 variable sites across all populations. See
318 James *et al.*, (2020) for further details on DNA extraction, library preparation and
319 bioinformatics.

320 **Broad-sense genotypic parallelism**

321 **Parallel nucleotide site:** We first assessed the relative contributions of broad- and narrow-
322 sense parallelism for each of the 9,686 sequenced SNPs across populations. More
323 specifically, we first used PLINK v1.9 (Purcell *et al.*, 2007) to normalise each SNP by
324 conducting a PCA and extracting the first eigenvector. For each SNP we then performed
325 linear models (ANOVA: *SNP* = *ecotype* + *pair* + *ecotype x pair*) and we extracted the partial
326 effect sizes (partial η^2) for each term in the model. As above, the partial effect size of the
327 *ecotype* term denotes how much of the variation for each SNP is explained by the overall
328 differences between ecotypes, whereas the *pair* and *interaction* terms indicate how much
329 variation is unique to replicate pairs. Therefore, when the effect size for the *ecotype* term is
330 larger than the *pair* or *interaction*, this indicates that broad-sense predominates over narrow-

331 sense. To plot this data as a frequency distribution, we calculated each SNP's distance from a
332 1:1 line, by subtracting the effect size for either the *pair* or *interaction* term from the *ecotype*
333 term in the model. Positive values indicate a larger contribution of broad- over narrow-sense
334 parallelism, and vice versa for negative values.

335 After assessing these relative contributions, we further explored the detailed patterns of
336 broad- parallelism at the level of the nucleotide site. More specifically we combined a
337 number of approaches to detect outliers between the Dune and Headland ecotypes. To ask
338 which nucleotide sites distinguish the ecotypes (overall Dune vs Headland), we undertook
339 two approaches: 1) detected outliers comparing all Dune populations vs all Headland
340 populations, and 2) detected outliers separately for each Dune-Headland pair, and asked
341 whether there were concordant allele frequency differences across all pairs.

342 Approach 1: To detect broad-sense outliers between the ecotypes, we undertook three outlier
343 detection methods: the top 1% from the distribution of F_{ST} values, the top 1% from the
344 distribution of cluster separation scores (CSS), and those SNPs identified by *BayeScan* (Foll
345 & Gaggiotti, 2008). To remove low frequency variants (Bhatia *et al.*, 2013), we filtered this
346 dataset for MAF 0.05, retaining 9,269 SNPs. F_{ST} per SNP (Weir & Cockerham, 1984) was
347 calculated within *VCFtools* (Danecek *et al.*, 2011), and we selected the top 1%
348 (corresponding to $F_{ST} > 0.353$, which also corresponds to an average change in allele
349 frequency, Δp, of 0.47, SD = 0.080; Figure S1A). We calculated the cluster separation score
350 (Jones *et al.*, 2012) per SNP using a custom R script (see *Supplementary R code* for details).
351 CSS is a genetic distance-based measure that quantifies the average divergence between
352 Dune and Headland clusters, after accounting for the variance within each ecotype. It ranges
353 between 0 and 1, with higher values indicating a more distinct separation between Dune and
354 Headland ecotypes. We selected the top 1% of CSS (corresponding to CSS > 0.523, an
355 average Δp of 0.45, SD = 0.087; Figure S1B). We also detected outliers in *BayeScan* v2.1
356 (Foll & Gaggiotti, 2008). *BayeScan* implements a reversible-jump MCMC algorithm to
357 estimate posterior odds comparing a model with and without selection for each SNP. We
358 used default parameters with a prior odd of ten, meaning the neutral model is ten times more
359 likely than the model with selection (as recommended for datasets of our size; Foll &
360 Gaggiotti, 2008). Results were robust to increasing the prior odds to 100. We categorised
361 SNPs as highly differentiated if they contained a posterior probability > 0.91, corresponding
362 to a Bayes Factor of > 10, and corresponding to an average Δp of 0.38, SD = 0.121. For this

363 approach 1, we classified SNPs as outliers if they were detected in at least two of the three
364 methods (Figure S1C).

365 Approach 2: To detect more subtle signals of broad-sense outliers between the ecotypes, we
366 asked whether there were concordant allele frequency changes across replicate pairs.
367 Specifically, we first detected outliers separately for each Dune-Headland pair (see methods
368 below). If a nucleotide site was highly differentiated in at least one pair, we compared allele
369 frequencies across all pairs for the site, and we asked whether the change in allele frequency
370 for each replicate pair was in the same direction across all pairs (tested using two-sided
371 dependent-samples sign-tests in R). Our overall best candidates for broad-sense SNP
372 parallelism are those that overlap between the two methods – i.e. they show high
373 differentiation between ecotypes (approach 1), with concordant allele frequency chances
374 across replicate pairs (approach 2).

375 To ask whether the candidate outliers from any of the approaches above fall within genic or
376 non-genic regions, we mapped a previously created *S. lautus* transcriptome (Liu, 2014) to the
377 reference PacBio genome v1.0 (Wilkinson *et al.*, 2019) with *minimap2* v2.17 (Li, 2018)
378 using default parameters. We considered each transcript a separate gene, which included all
379 isoforms. As the transcriptome excludes introns, we still considered SNPs mapped to the
380 reference genome that fall between two segments of the same transcript as a genic SNP. All
381 other SNPs were considered non-genic, which are expected to include variants in regulatory
382 and repetitive regions as well as in genic regions with unknown homologous genes in other
383 plants. We excluded SNPs that had > 1 gene mapping to it.

384 **Parallel gene:** We assessed the relative contributions of broad- and narrow-sense parallelism
385 for each gene across populations. We again normalised the data, retaining the first
386 eigenvector for each gene. For each gene we then performed linear models (ANOVA: *gene* =
387 *ecotype* + *pair* + *ecotype x pair*) and we extracted the partial effect sizes (partial η^2) for each
388 term in the model. We again plotted this as a frequency distribution (see above for details).
389 After assessing these relative contributions, we further explored the detailed patterns of
390 broad- parallelism at the level of the gene according to the outlier analysis above. More
391 specifically, we considered a gene an outlier if it harboured at least one outlier SNP
392 according to the two above approaches (i.e. showed high differentiation between ecotypes
393 (approach 1), with concordant allele frequency chances across replicate pairs (approach 2)).
394 We asked whether the broad-sense nucleotide-site outliers were located across multiple genes

395 or the same gene. To assign orthologue genes, we used *BLASTx* (Altschul *et al.*, 1990) to
396 obtain a RefSeq code (Pruitt, 2004) for each *S. lautus* gene. We searched the RefSeq protein
397 database for *Arabidopsis thaliana* proteins that match our target genes using an E-value
398 threshold of $< 10^{-6}$. We used the web-based version of *DAVID* v6.8 (Huang *et al.*, 2009b; a)
399 to obtain the predicted functional annotation of each *S. lautus* gene sequenced in this work.

400 **Enriched biological function:** To ask whether the outliers were enriched for any functional
401 categories, we used *DAVID* to conduct a gene-enrichment analysis using 'functional
402 annotation clustering'. Functional annotation clustering groups similar functional terms into
403 'clusters' to avoid redundant annotations. The *Arabidopsis thaliana* genome was used as a
404 genetic background.

405 **Narrow-sense genotypic parallelism**

406 **Parallel nucleotide site:** We asked which nucleotides sites are highly differentiated (i.e.
407 outliers) in each replicate pair, and whether the same nucleotide sites are outliers across
408 multiple pairs (i.e. across multiple localities). We first filtered each separate pair for MAF
409 0.05, retaining between 5,513 and 8,875 SNPs per pair (mean = 7,461 SNPs; SD = 1,038). As
410 above, we identified outliers for each replicate pair using a combination of F_{ST}, CSS, and
411 *BayeScan*. Instead of selecting the top 1% of F_{ST} and CSS values (which is highly dependent
412 on the number of sampled SNPs) we chose stringent cut-offs of F_{ST} and CSS > 0.95. Within
413 *BayeScan*, we considered SNPs with a posterior probability > 0.91 as highly differentiated,
414 which corresponds to a Bayes Factor > 10. We considered a SNP an outlier if it was detected
415 in at least two of the approaches. We used custom R scripts to calculate the overlap of
416 outliers across replicate pairs. To ask whether the outliers shared between pairs were greater
417 than expected by chance, we used a hypergeometric distribution function, *phyper*, within R to
418 evaluate the likelihood of sharing outliers between localities.

419 **Parallel gene:** As above, we considered a gene to be an outlier if it harboured at least one
420 differentiated SNP, and used *DAVID* to obtain the functional annotation of each gene. We
421 again used a hypergeometric distribution to ask whether shared outlier genes between
422 localities were greater than expected by chance.

423 **Parallel biological function:** For each replicate pair, we conducted a gene-enrichment
424 analysis using 'functional annotation clustering' in *DAVID* (as described above). For each
425 pair we selected clusters containing at least one significantly enriched functional term (P <

426 0.05). We then compared these enriched terms across pairs to ask whether any biological
427 functions were repeatedly enriched across pairs. We used a two-sided dependent-samples
428 sign-test to ask if the number of enriched pairs per predicted functional category differed
429 from chance. We compared the distributions of the proportions of shared outlier nucleotide
430 sites, outlier genes and enriched biological functions across pairs using a two-sided X^2-test
431 with continuity correction in R using the *prop.test* function.

432 **Variation in phenotypic parallelism**

433 We asked whether the variation in phenotypic parallelism within the system (i.e. differences
434 in divergence (ΔL) and the contribution of traits (θ) between replicate pairs) could be
435 explained by demographic factors. All analyses were undertaken in R. Specifically, we used
436 gene flow estimates from James *et al.*, (2020) to ask whether gene flow constrains
437 divergence, where we predict pairs with higher gene flow to be more phenotypically similar
438 (linear model: *phenotypic length (L) = gene flow*; Table S3). As most pairs experience
439 minimal gene flow and are effectively allopatric, we do not expect gene flow to be a major
440 constraining force in most population pairs. We also used divergence time estimates from
441 James *et al.*, (2020) to ask whether older pairs show more phenotypic divergence than
442 younger pairs as they have experienced more genetic drift over time (linear model:
443 *phenotypic length (L) = divergence time*; Table S3). We also reasoned that populations
444 adapting to more contrasting environments should have greater phenotypic differences (linear
445 model: *phenotypic length (L) = environmental distance*; Table S3). We used environmental
446 distances from previous work in *S. lautus* (see Roda *et al.*, 2013b). In addition, we asked
447 whether pairs that were more phenotypically similar (ΔL and θ) shared more outlier
448 nucleotide sites, genes, and biological functions using Mantel tests (Mantel 1967) with 999
449 permutations.

450 ## Results

451 **Relative contributions of broad- and narrow-sense phenotypic parallelism**

452 At the multivariate level of the ecotype (MANOVA: *traits = ecotype + pair + ecotype x*
453 *pair*), the partial effect size of the ecotype term (Wilks partial $\eta^2 = 0.86$) was larger than both
454 the pair (Wilks partial $\eta^2 = 0.23$; Figure S3) and the interaction term (Wilks partial $\eta^2 = 0.19$;
455 Figure 2C). At the univariate level of the trait, the *ecotype* effect size was larger than both the

456 pair (Figure S3) and interaction term (Figure 2C) for most traits (more data points above the
457 dotted line than below, see Table S4 for details). The larger effect sizes for the ecotype terms
458 suggest that the phenotypic variation within the system is mainly explained by differences
459 between ecotypes rather than replicate pairs: broad-sense phenotypic differences between
460 ecotypes predominates over narrow-sense in *S. lautus*.

461 **Phenotypic parallelism**

462 **Broad-sense phenotypic parallelism**

463 **Parallel ecotype:** We found striking visual differences between the mean Dune and
464 Headland phenotypes for both architecture and leaf characteristics based on linear
465 discriminant function analysis (Figure 2D, 3C). In multivariate space, Dune and Headland
466 ecotypes clearly clustered into two distinct groups (Figure 2E; Pillai's Trace = 0.73, $F_{1, 603}$ =
467 175.13, $P < 2.2 \times 10^{-16}$). This pattern held true when traits were separated into architecture
468 (Figure S2A; Pillai's Trace = 0.63, $F_{1, 603}$ = 202.42, $P < 2.2 \times 10^{-16}$) and leaf categories
469 (Figure S2B; Pillai's Trace = 0.61, $F_{1, 603}$ = 233.15, $P < 2.2 \times 10^{-16}$). Across all traits, K-
470 means clustering analysis correctly assigned 95% of Dune individuals into the correct cluster,
471 and 87% of Headland individuals, further suggesting the majority of individuals within an
472 ecotype are more phenotypically similar than between ecotypes. When traits were split into
473 architecture and leaf, these numbers were slightly reduced. For architecture traits alone, 91%
474 of Dunes and 82% of Headlands were assigned to the correct cluster. For leaf traits, 93% of
475 Dunes and 78% of Headlands were correctly assigned.

476 We performed a linear discriminant analysis (LDA) on all traits to ask which linear
477 combination of traits best explains the phenotypic differences between Dune and Headland
478 ecotypes. The LDA was strongly loaded by leaf area and secondary branch angle, followed
479 by leaf dissection, leaf circularity, and widest width of the plant (Figure 3C). All traits were
480 loaded in the same direction, except for widest width, leaf dissection, and leaf circularity. The
481 LDA suggests that divergence between ecotypes is multivariate and has occurred on the
482 majority of measured traits, and a single trait does not dominate the phenotypic differences
483 between ecotypes.

484 **Parallel trait:** We first used vote-counting to quantify whether the traits in the Dune and
485 Headland populations of each pair have evolved in the same direction. For all traits, at least
486 six of the eight pairs evolved in parallel (Figure 3A). Four of the nine traits had all pairs

evolving in the same direction and only these four traits were considered statistically parallel. Trait-by-trait linear models revealed a significant main effect of ecotype for each trait (Table S4), suggesting there are differences between Dune and Headland populations for all traits. The proportion of phenotypic variance explained by the ecotype term (R^2) for linear models with population means revealed six strongly parallel traits ($R^2 \geq 0.50$), one moderately parallel trait ($0.50 > R^2 \geq 0.33$), and two weakly parallel traits ($R^2 < 0.33$; Figure 3B). Four traits (vegetative height, secondary branch angle, leaf area and leaf elongation) were consistently viewed as parallel across each of the three approaches, two traits (main stem angle and main stem diameter) were parallel within two of the approaches, and three traits (widest width, leaf dissection and leaf circularity) were parallel within only one approach (see Table S5 for more details).

Overall, we observed strong broad-sense phenotypic parallelism within *S. lautus* at the level of both the ecotype and trait. To address whether each replicate pair is phenotypically similar, we next asked whether the phenotypic differences between Dune and Headlands depend on replicate pairs.

Narrow-sense phenotypic parallelism

Parallel pair: We used Phenotypic Change Vector Analysis (PCVA) to ask how parallel each replicate pair is in terms of amount of divergence, and the contribution of traits to divergence. Within multivariate phenotypic space, there were different levels of divergence (ΔL) between replicate pairs (Figure 4A, 4B). Considering all traits, the mean ΔL ($\pm$SE) between pairs was 1.7 ± 0.15. Out of the 28 pairwise comparisons, we only observed nine statistically parallel comparisons (i.e. $\Delta L \approx 0$; 32.1% of pairwise comparisons; Table S6). Therefore, most population pairs have different amounts of divergence between the Dune and Headland populations. When we split the traits into architecture and leaf categories, we recovered more phenotypic parallelism (Figure S4, S5). We observed ten statistically parallel comparisons for architecture traits (mean ΔL 1.0 ± 0.12; Table S7), and thirteen statistically parallel for leaf traits (mean ΔL 0.93 ± 0.14; Table S8).

The contribution of traits to divergence (θ) was quite variable across pairs (Figure 4A, 4C). Out of the 28 pairwise comparisons, only one angle was parallel, i.e. not significantly different from zero ($\theta \approx 0°$; 3.6% of pairwise comparisons; Table S9), indicating that a different contribution of traits is involved in the divergence between most pairs. The mean

518 angle (±SE) between population pairs was 39.5 ± 2.1°, and all angles were acute, with a
519 largest of 62.8°. When traits were split into architecture and leaf categories, we again
520 recovered more phenotypic parallelism. We observed nine statistically parallel angles for
521 architecture traits (mean angle 29.8 ± 3.0°; Table S10) and four statistically parallel for leaf
522 traits (42.6 ± 3.4°; Table S11).

523 Overall, narrow-sense phenotypic parallelism (i.e. the similarities between replicate pairs)
524 was weak within the system. Most pairs had different magnitudes of divergence (ΔL), with
525 different contribution of traits to divergence (θ).

526 **Genotypic parallelism**

527 **Broad-sense genotypic parallelism**

528 **Parallel nucleotide site:** The relative contributions of broad- and narrow-sense parallelism
529 revealed that very few SNPs explained more variance between ecotypes than between pairs
530 (Figure 5A, 5C, S6A, S6C). More specifically, only 6.3% of SNPs contained a partial effect
531 size of the ecotype term that was larger than the interaction term (i.e. those above the dashed
532 line in Figure 5A which are > 0 in Figure 5C). This dramatically decreases to 0.68% when we
533 only consider SNPs which are > 0.1. This indicates that parallel evolution at the level of the
534 nucleotide site within the system is largely predominated by differences between replicate
535 pairs, suggesting that adaptation in each pair is occurring via different SNPs.

536 Using the outlier approach to detect highly differentiated SNPs between Dune vs Headland,
537 we detected 93 parallel nucleotide sites (1% of sequenced SNPs) which were considered
538 broad-sense outliers comparing all Dune vs all Headland populations (approach 1). Fifty-four
539 of these SNPs fall within genic regions, whereas thirty-nine are in non-genic regions.
540 Although this may imply an overrepresentation of outlier SNPs in genic regions, we must
541 note our GBS data is biased towards genic regions (due to the enzymes preferentially cutting
542 within gene-rich regions). When considering outliers detected within each replicate pair that
543 contained concordant allele frequency differences across pairs (approach 2), we detected 15
544 parallel nucleotide sites (0.16% of SNPs; Figure S7). These 15 SNPs had concordant allele
545 frequency differences in either all nine (S-statistic = 9, P = 0.004), or eight (S-statistic = 8, P
546 = 0.04) replicate pairs. Nine of these SNPs fall within genic regions, whereas six do not. Five
547 SNPs were detected as outliers in both approaches (three genic, two non-genic; Figure 6A;

548 red dots in Figure 5A). These five nucleotide sites are considered the best candidates for
549 broad-sense parallelism as they show both high differentiation between ecotypes, with
550 concordant allele frequency changes in each replicate pair. The average difference in allele
551 frequency between Dune and Headlands for the three genic SNPs was 0.55 (SD = 0.096),
552 whereas the average for the two non-genic SNPs was 0.57 (SD = 0.107). This suggests that
553 possibly both genic and regulatory changes contribute to parallel evolution in the system.

554 **Parallel gene:** The relative contributions of broad- and narrow-sense parallelism at the level
555 of the gene revealed that very few genes explained more variance between ecotypes than
556 between pairs (Figure 5B, 5D, S6B, S6D). More specifically, only 6.3% of SNPs contained a
557 partial effect size of the ecotype term that was larger than the interaction term (i.e. those
558 above the dashed line in Figure 5A which are > 0 in Figure 5C). Although this number
559 dramatically decreases to 1.4% when we only consider genes which are > 0.1. This indicates
560 that differences between replicate pairs largely predominates parallel evolution at the level
561 gene within the system, suggesting that adaptation in each pair is likely occurring at different
562 genes.

563 Of the five candidate outlier SNPs identified above using the outlier approach (i.e. that
564 showed high differentiation between ecotypes (approach 1), with concordant allele frequency
565 chances across replicate pairs (approach 2)), the three genic SNPs fall within three separate
566 genes, two of which have homologs within *Arabidopsis* (Table S12; red dots in Figure 5B).
567 These two genes are a galactose oxidase/kelch repeat superfamily protein (AT5G04420;
568 Figure 6A first panel) and a basic salivary proline-rich-like protein (AT5G14540; Figure 6A
569 last panel). The proteins are both located in the cytosol and are both widely expressed within
570 the plant in structures including the leaves, stem, reproductive organs, roots, and seeds
571 (Klepikova *et al.*, 2016). Looking at approach 1 and 2 separately, the 54 outlier genic SNPs
572 detected in approach 1 fall in 49 separate genes, of which 44 have homologs within
573 *Arabidopsis*. The majority of these genes have functions in the nucleus, cytoplasm, plasma
574 membrane, mitochondria and chloroplast, and are involved in processes including ion
575 transport, transcription, response to heat, response to water deprivation, DNA repair and
576 embryo development (see Table S12 for details). For approach 2, the nine outlier genic SNPs
577 from approach 2 fall in nine separate genes, of which seven have homologs within
578 *Arabidopsis*. These genes have functions in the nucleus, cytoplasm, plasma membrane and

579 mitochondria, and are involved in processes including ion transport, aminoacylation, embryo
580 development and DNA repair (see Table S12 for details).

581 **Parallel biological function:** Because we only sampled two candidate broad-sense outlier
582 genes that had predicted functions, we did not perform an enrichment analysis on them.
583 However, we performed enrichment analysis on 1) the 54 genes from approach 1 (of which
584 48 contained *Arabidopsis* homologues), and 2) the nine genes from approach 2 (of which
585 seven contained *Arabidopsis* homologues). For approach one, the biological category of
586 helicases was significantly enriched (P = 0.0344; P-value is the EASE score, a modified
587 Fisher Exact P-value). The functions of the genes within this helicase category are involved
588 in DNA replication, recombination and repair, and often associated with tolerance to drought
589 and salinity in plants (Owttrim, 2006). There were no significantly enriched functions with
590 approach 2 (P > 0.05).

591 **Narrow-sense genotypic parallelism**

592 **Parallel nucleotide site:** We observed shared outliers in pairwise comparisons between
593 localities: on average, 157 outlier SNPs (SD = 74.5) were shared between pairs (Table S13).
594 For each pairwise comparison, the shared SNPs were greater than expected by chance,
595 suggesting some degree of narrow-sense parallelism at the level of the nucleotide site. We
596 detected six nucleotide sites that were outliers in seven pairs (Figure 6B), however, there
597 were no outlier SNPs common across all nine pairs, suggesting that even though there are
598 common outliers between individual pairs, this does not translate to common outliers across
599 the system as a whole. This is because within narrow-sense genotypic parallelism, a parallel
600 nucleotide site across all pairs has to be separately detected as an outlier in each replicate
601 pair. Conversely, a broad-sense outlier may reflect strong ecotypic differentiation even if
602 some nucleotide sites that are not statistical outliers in every replicate pair.

603 **Parallel gene:** Pairwise comparisons between localities revealed an average of 124 outlier
604 genes shared between pairs (SD = 54.6). The shared outlier genes between all comparisons
605 were greater than expected by chance, except for one comparison (D14-H15 vs D32-H12;
606 Table S14). We detected 39 genes that were outliers in all nine or eight pairs (Figure 6B), of
607 which 36 contained homologs in *Arabidopsis* These genes have functions in the nucleus,
608 cytoplasm, plasma membrane, chloroplast, vacuole and endoplasmic reticulum, and are
609 involved in processes including ion transport, transcription, seed development, response to

610 auxin, response to heat, response to salt stress, embryo development and cell growth (see
611 Table S12 for details).

612 **Parallel biological function:** For each replicate pair we conducted a gene-enrichment
613 analysis using outlier genes to ask whether any biological functions were enriched and
614 whether the same functions were repeatedly enriched across replicate pairs. Across all pairs
615 there were 17 enriched functions (Figure 6C; Table S15). No function was repeatedly
616 enriched in all replicate pairs, although two functions (chloroplast and nucleotide-
617 binding/ATP-binding) were enriched across eight replicate pairs, this being greater than
618 expected by chance (S-statistic = 8, P = 0.04). Within the chloroplast category, the majority
619 of outlier genes across pairs have functions in the chloroplast, cytosol, mitochondria, nucleus,
620 vacuole and the cell wall, and are involved in processes including oxidation reduction,
621 response to light, translation, proteolysis, protein phosphorylation and protein folding (see
622 Table S16 for details).

623 Within the nucleotide-binding/ATP-binding category, the majority of outlier genes across
624 pairs have functions in the chloroplast, nucleus, plasma membrane, cytosol, mitochondria,
625 cytoplasm, and are involved in processes including protein phosphorylation, protein folding,
626 transcription, aminoacylation, ion transport, and response to stress (see Table S17 for details).
627 Across the nine replicate pairs, the distribution of the proportion of shared outlier nucleotide
628 sites was significantly different to the outlier genes ($X^2 = 279.65$, df = 8, $P < 2.2x10^{-16}$), and
629 biological functions ($X^2 = 361.95$, df = 8, $P < 2.2x10^{-16}$). The distributions of the genes and
630 biological functions were not significantly different across the nine replicate pairs ($X^2=14.52$,
631 df=8, P = 0.069).

632 **Variation in phenotypic parallelism**

633 Gene flow did not constrain phenotypic divergence, as there was no relationship between
634 levels of gene flow and the lengths of phenotypic vectors (L) between ecotypes within a
635 locality for the Dune to Headland gene flow ($F_{1,6} = 0.007$, P = 0.934, $R^2 = 0.001$), Headland
636 to Dune gene flow ($F_{1,6} = 1.17$, P = 0.321, $R^2 = 0.163$), and absolute gene flow ($F_{1,6} = 0.061$,
637 P = 0.463, $R^2 = 0.092$). There was also no relationship between divergence time between
638 ecotypes and L within a locality ($F_{1,6} = 0.321$, P = 0.592, $R^2 = 0.051$). Environmental distance
639 did not relate to how phenotypically divergent (θ) a population pair was ($F_{1,3} = 0.046$, P =
640 0.843, $R^2 = 0.015$), although we treat this data with caution due to low sample size of the

641 environmental data. Population pairs that were more phenotypically similar (i.e. smaller ΔL)
642 did not share more outlier SNPs, genes or biological functions (Mantel test SNPs: r = -0.215,
643 P = 0.894; Mantel test genes: r = -0.164, P = 0.835, Mantel test biological functions: r = -
644 0.179, P = 0.837). Population pairs with similar contribution of traits to divergence (i.e.
645 smaller θ) also did not share more outlier SNPs, genes, or biological functions (Mantel test
646 SNPs: r = -0.493, P = 0.989; Mantel test genes: r = -0.347, P = 0.865; Mantel test biological
647 functions: r = -0.337, P = 0.93).

648 **Discussion**

649 We have developed an integrative framework to measure parallel evolution at the phenotypic
650 and genotypic levels. Our approach aims to quantity the overall variation between ecotypes
651 (broad-sense parallelism) as well as the variation between replicate pairs (narrow-sense
652 parallelism). Applying our framework to *Senecio lautus*, we have demonstrated that the
653 independent adaptation of populations to dune and headland environments has resulted in the
654 repeated evolution of similar phenotypes. In each replicate population, these ecotypes have
655 mainly diverged via different mutational changes in different genes, but often belonging to
656 the same predicted biological function, further adding credence to the independent evolution
657 of these ecotypes along the Australian coast. These results are consistent with the
658 demographic history of the system detailed in James *et al.*, (2020), where evolution of each
659 population is highly independent due to strong isolation by distance, and augmented by the
660 strong local adaptation measured in multiple reciprocal transplant experiments in the system
661 (Melo *et al.*, 2014; Richards *et al.*, 2016; Richards & Ortiz-Barrientos, 2016; Walter *et al.*,
662 2016, 2018b; a; Wilkinson *et al.*, 2019). Below we discuss the framework we proposed to
663 understand parallel evolution in light of *S. lautus* and consider how the genetics and ecology
664 of coastal ecotypes impact the likelihood of phenotypic and genotypic parallelism. We finish
665 by discussing when our framework should be used and address potential caveats, and how it
666 could contribute to future studies of parallel evolution in nature.

667 **Phenotypic parallelism**

668 In *S. lautus*, the phenotypic differences between ecotypes was stronger than the variation
669 arising from replicate pairs in both multivariate and univariate space. For instance, the partial
670 effect size of the ecotype was 86% for multivariate phenotypic differences, whereas the
671 partial effect size of replicate pairs was only 19% (Figure 2C). This was also true for most

traits when considered alone, where ecotypic differences explained most variation in vegetative height, the angle of secondary branches, and the size of leaves (Figure 2C, Table S4). These traits largely capture the overall growth habits of prostrate and erect in the system. Partitioning variance between ecotypes and replicate pairs is a simple statistical representation of broad- versus narrow-sense parallelism and in this case, it reveals broad-sense phenotypic parallelism is very strong in the *S. lautus* system. To see this in greater detail, we now discuss results from our other analyses of broad-sense parallelism in *S. lautus* and compare them to other systems where parallel evolution seems strong.

Broad-sense phenotypic parallelism (i.e. the net contribution of all independent replicate populations within each ecotype to the overall phenotypic differences between ecotypes) was strong in the coastal Dune and Headland ecotypes. At the level of the ecotype we observed striking phenotypic differences between ecotypes in multivariate trait space, with little overlap between ecotypes (Figure 1C). This strong broad-sense phenotypic parallelism is seen in other empirical parallel evolution systems such as lake-stream stickleback on Haida Gwaii in Canada (Deagle *et al.*, 2012), and dwarf-normal lake whitefish (Laporte *et al.*, 2015), where there is a clear phenotypic separation of ecotypes in the first two dimensions of multivariate space. In contrast, some systems of parallel evolution have lower broad-sense phenotypic parallelism, where there is overlap between ecotypes in phenotypic space. For instance, benthivorous-planktivorous Arctic charr (Jacobs *et al.*, 2018), benthic-limnetic cichlid fishes (Elmer *et al.*, 2014) and lake-stream stickleback on Vancouver Island in Canada (Stuart *et al.*, 2017) do not have each ecotype occupying a separate region in the first two dimensions of multivariate space, thus likely having lower broad-sense phenotypic parallelism (note that some overlap in multivariate variance can still lead to significant broad-sense parallelism).

Broad-sense phenotypic parallelism at the level of the trait was high in *S. lautus*. This was true for all measured traits, except for leaf dissections, that seem to vary more between replicate pairs than between ecotypes. This pattern is rather different to other systems where the contribution of traits to ecotypic differences has been calculated. For instance in stickleback ecotypes, Stuart *et al.*, (2017) found that the variation in most traits is explained mainly by differences between replicate pairs rather than between ecotypes, suggesting little parallelism at the level of individual traits. This discrepancy might disappear if more traits are measured in *S. lautus*, but the phenotypic dimensionality of the system does not seem to be

704 very high: of the 16 traits we measured, we discarded seven highly correlated traits. In other
705 studies where more traits have been measured in *S. lautus* populations (Walter, 2016; Walter
706 *et al.*, 2018a), strong genetic correlations exist amongst a variety of vegetative traits
707 suggesting strong interdependence between morphological modules such as leaf and plant
708 architecture.

709 In contrast, narrow-sense phenotypic parallelism (i.e. the similarities between replicate pairs,
710 quantified with PCVA) in *S. lautus* was weak, with most pairs having different magnitudes of
711 divergence (32.1% of pairwise comparisons were parallel, i.e. $\Delta L \approx 0$) and different
712 contribution of traits to divergence (3.6% of pairwise comparisons were parallel, i.e. $\theta \approx 0°$).
713 Although the PCVA approach is relatively new within studies of parallel evolution, a recent
714 study in 16 lake-stream stickleback pairs discovered less parallelism than *S. lautus*, with
715 27.5% of pairwise comparisons having parallel magnitudes of divergence, and only 0.83% of
716 pairwise comparisons with parallel contribution of traits to divergence (Stuart *et al.*, 2017).
717 As more systems adopt the PCVA approach we will be able to directly compare the amount
718 of parallelism between replicate populations across different study systems, although the
719 creation of null distributions for the process remains unexplored.

720 The differences between the strong broad-sense and weak narrow-sense scales of parallelism
721 could be due to environmental heterogeneity within each habitat type (Bolnick *et al.*, 2018;
722 Thompson *et al.*, 2019). This is seen within stickleback, where environmental differences
723 explain variation in phenotypes between replicate pairs (Stuart *et al.*, 2017). Within *S. lautus*,
724 the sand dune habitats are highly homogenous, whereas the rocky headland habitats are more
725 variable across localities (Roda *et al.*, 2013b) and are harder to transplant in to other
726 headlands (Walter *et al.*, 2018a). Differences in environmental variance within each ecotype
727 can modulate variation in phenotypic adaptation, and partially explain why narrow-sense
728 phenotypic parallelism is low in *S. lautus*. In our phenotypic vector analyses this trend can be
729 observed: Dune populations cluster but Headland populations are scattered in vector space
730 (Figure 4A).

731 One interpretation for these results comes when considering the nature of fitness landscapes
732 in *S. lautus*. The fitness landscape across sand dunes might have a common narrow fitness
733 peak whereas the fitness landscape across headlands might be broad and rugged. As a
734 consequence, every Headland population might be evolving in a slightly unique way, but
735 within the realm of closely related fitness peaks where selection coefficients covary more

736 strongly within headland environments than relative to other habitats like the sand dunes
737 (Lobkovsky & Koonin, 2012). Also consistent with this expectation, previous work in *S.*
738 *lautus* have shown that multivariate phenotypic space is more restricted in homogenous sand
739 dunes than in the variable rocky headlands (Roda *et al.*, 2013b; Walter *et al.*, 2018a). More
740 telling, within previous reciprocal transplants, Dune populations survive equally well across
741 other sand dunes, but Headlands do not survive well across other rocky headlands (Walter *et*
742 *al.*, 2016, 2018b), suggesting that Headland populations have adapted to different local
743 fitness peaks and that fitness trade-offs might depend on slight differences in fitness
744 landscapes within an ecotype. Overall, it seems that differences in the structure of fitness
745 landscapes within habitats could produce little narrow-sense phenotypic parallelism, yet
746 strong broad-sense phenotypic differences between ecotypes.

747 A second interpretation for the decoupling between broad- and narrow-sense phenotypic
748 parallelism comes from the flexibility and functional redundancy of complex genetic
749 architectures (Storz, 2016; Yeaman *et al.*, 2018). Within each ecotype, the fitness optimum
750 may encompass multiple phenotypic solutions with different genetic architectures, a so called
751 'degenerate phenotype-fitness map' (Salazar-Ciudad & Marín-Riera, 2013). Phenotypic
752 variation within each ecotype could therefore arise if the phenotypes of populations are
753 functionally equivalent and when mutations are limiting, with the same mutations not being
754 readily available during each independent colonisation of a new habitat (Gompel &
755 Prud'homme, 2009; Lenormand *et al.*, 2009, 2016; Stern & Orgogozo, 2009; Chevin *et al.*,
756 2010; Rosenblum *et al.*, 2014; Dittmar *et al.*, 2016; Stoltzfus & McCandlish, 2017; Blount *et*
757 *al.*, 2018). Mutational constraints and functional equivalence (both genotypic and
758 phenotypic), therefore, could result in slight phenotypic variations in each population of a
759 given ecotype, with each of these phenotypic solutions at (or close to) similar fitness
760 optimum for each ecotype.

761 Finally, the nature of parallelism in a system can depend on the nature of selection driving
762 adaptation and on whether selection drives the evolution of one or multiple traits. For
763 instance, ecotypic description of systems, such as sticklebacks and cichlids, is often binary
764 and relies on correlating a large environmental category to a general perception of the
765 phenotypic differences between populations. In *S. lautus*, Dune and Headland ecotypes are
766 qualitatively different in their growth habit, where Dune individuals grows erect and tall, and
767 Headland individuals always grow prostrate forming matts over the ground. In *S. lautus*,

768 however, many traits seem to contribute simultaneously to the qualitative descriptions of
769 prostrate and erect (Walter *et al.*, 2018a), suggesting that parallel evolution relies on
770 coordinated trait evolution for size and for shape. In sticklebacks, on the other hand, only a
771 few traits seem to be highly parallel across replicate pairs (Stuart *et al.*, 2017). One simple
772 explanation for these discrepancies is whether parallel evolution relies mostly on the origin of
773 complex or simple adaptations, and on whether few genes of large effect (e.g., EDA gene in
774 sticklebacks, (Colosimo *et al.*, 2005) or many of small effect (e.g., genetics of gravitropism in
775 *S. lautus*; Wilkinson *et al.*, 2019) drive the genetic architecture of adaptation. Our genotypic
776 results might help shed some light on these different explanations for patterns of phenotypic
777 parallelism in *S. lautus*.

778 **Genotypic parallelism**

779 Within *S. lautus*, the relative contributions of broad- and narrow-sense genotypic parallelism
780 revealed that very few SNPs and genes explained more variance between ecotypes than
781 between pairs. This suggests there is little broad-sense genotypic parallelism within the
782 system, where divergence is explained by differences between replicate pairs at the level of
783 the SNP and gene. This is also evident when detecting outliers between the ecotypes (i.e. the
784 genetic differences between all Dunes vs all Headlands), where very few nucleotide sites and
785 genes distinguished the ecotypes. Narrow-sense genotypic parallelism within the system (i.e.
786 the genetic similarities between Dune-Headland replicate pairs) was strongest at the level of
787 the biological function, although there was some sharing of common outlier nucleotide sites
788 and genes across replicate population pairs. These results suggest that adaptation in *S. lautus*
789 could be rather flexible at lower levels of organisation (the nucleotide site and gene), and
790 only parallel at the level of the biological function. Parallelism at the biological function level
791 might be common within nature (e.g., Smith & Rausher, 2011; Kowalko *et al.*, 2013; Roda *et*
792 *al.*, 2013b; Laporte *et al.*, 2015; Perreault-Payette *et al.*, 2017; Cassin-Sackett *et al.*, 2019)
793 given that there are fewer biological functions than there are genes or nucleotide sites
794 (Tenaillon *et al.*, 2012; Tiffin & Ross-Ibarra, 2014). In this regard, our study is rather
795 conservative as we have only defined narrow-sense parallelism in function when the function
796 was separately enriched across either all nine, or eight population pairs. Even in those few
797 enriched functions, the set of sequenced genes that were divergent in each population pair
798 were often different (Table S16, S17), further suggesting that adaptation in *S. lautus* might be

799 quite robust and possibly is underpinned by polygenic adaptation and functional redundancy
800 of alleles at different loci (Tiffin & Ross-Ibarra, 2014).

801 A recent study in this system has demonstrated that hormone signalling, specifically the
802 auxin signalling pathway is divergent between Dune and Headland populations, indicating
803 that integrated and modular phenotypes might be key to the patterns of parallel evolution we
804 have observed at the genotypic level in this system (Roda *et al.*, 2013b; Wilkinson *et al.*,
805 2019). Consistent with this observation, we have sequenced divergent genes involved in the
806 auxin pathway, although these genes are not consistently differentiated across all population
807 pairs, giving further evidence that genotypic adaptation in *S. lautus* is most parallel at the
808 level of the biological function (Table S18). Future studies on the molecular basis of
809 adaptation should focus on the concomitant contribution of many genes to phenotypic
810 variation and to their shared cellular and physiological roles, as it is likely that variation in
811 regulatory networks might underlie a large fraction of the adaptive space in organisms (Boyle
812 *et al.*, 2017; VanWallendael *et al.*, 2019).

813 The demographic history of species can also alter the patterns of genotypic parallelism
814 (Conte *et al.*, 2012; Ord & Summers, 2015). The most trivial reason is the contribution of
815 drift to diversity and divergence, particularly when populations are rather small. In such
816 scenarios, not surprisingly, many outlier loci will randomly accumulate in each population
817 pair, thus reducing levels of both broad- and narrow-sense genotypic parallelism. This is one
818 argument for grouping outliers through their cellular and physiological function, as drift will
819 be less likely to create functional similarities across population pairs. Future studies of
820 parallelism should not only identify outlier loci that are unlikely to have drifted to fixation,
821 but also ask if they have functional relationships amongst them. Although, more challenging,
822 identifying loci with concordant patterns of divergence, even if they are not outliers per se,
823 can also reveal sets of genes that contribute to consistent patterns of divergence across
824 populations and that are thus likely to be responsible for polygenic signatures of adaptation
825 (Berg & Coop, 2014; Tiffin & Ross-Ibarra, 2014).

826 **Broad- and narrow-sense parallel evolution framework**

827 Within the literature, the debate surrounding the term 'parallel evolution' manifests because
828 parallelism can arise at different scales of phenotypic and genotypic divergence, as well as
829 different genetic levels of biological organisation. Here, our framework proposes to

830 categorise parallel evolution into the broad- and narrow-sense scales (at both the phenotypic
831 and genotypic levels) to help quantify both the overall patterns between ecotypes, as well as
832 the similarities of each replicate pair. We suggest that researchers need to be explicit when
833 referring to either phenotypic or genotypic patterns of parallelism, whether these patterns
834 exist across the system as a whole, or only between particular replicate pairs, and at what
835 level of biological organisation the genotypic parallelism occurs (i.e. nucleotide site, gene or
836 biological function).

837 A 'perfectly' parallel system at the level of the phenotype would contain distant phenotypic
838 differences between ecotypes, (i.e. strong broad-sense phenotypic parallelism), with replicate
839 pairs that are effectively identical in their phenotypic differences (i.e. strong narrow-sense
840 phenotypic parallelism). Such a system would imply that evolution is extremely repeatable
841 and predicable at the level of the phenotype (Blount *et al.*, 2018), where every instance of
842 repeated evolution has resulted in the same phenotypic adaptations. However, this may be
843 highly unlikely within natural systems due to a range of factors such as environmental
844 heterogeneity within each habitat, the interplay between the genotype, phenotype and fitness
845 landscapes, genetic constraint, and stochastic forces such as genetic drift (Lenormand *et al.*,
846 2009, 2016; Rosenblum *et al.*, 2014; Fraïsse & Welch, 2019).

847 A 'perfectly' parallel system at the level of the genotype would show the same nucleotide
848 sites involved in adaptation across replicate pairs, with evidence that the mutation(s) have
849 undergone repeated and independent selection in each population. Again, this may be highly
850 unlikely within natural systems, especially in systems with phenotypic redundancy (with
851 many phenotypic solutions to the same problem), when adaptive traits are polygenic
852 (Yeaman, 2015), and is highly dependent on the availability of adaptive mutations (Chevin *et*
853 *al.*, 2010; Ralph & Coop, 2010; Stoltzfus & McCandlish, 2017; Bailey *et al.*, 2018). Within
854 genotypic parallelism, the nested levels of biological organisation imply that parallelism at
855 the lowest level (i.e. the nucleotide site) will translate to parallelism at the gene and
856 biological function levels. We therefore must keep in mind that when parallelism occurs at a
857 given level (i.e. the biological function), it will not necessarily occur at levels below (i.e. the
858 gene and nucleotide site).

859 *The effects of sampling on parallelism*

860 The ability to detect genotypic parallelism is impacted by sampling. Obvious pitfalls exist for
861 methods that do not sample the entire genome. For instance, reduced representation libraries
862 sparsely sample genomic regions and may be biased towards genic or non-genic regions
863 depending on the choice of enzymes. Here, we by-pass some of these problems as our
864 enzyme of choice cuts disproportionally gene rich regions of the genome. Overall, we
865 mapped ~60-70% of RAD tags to the transcriptome of *S. lautus,* suggesting that traditional
866 problems with RADs do not affect us (Tiffin & Ross-Ibarra, 2014; Lowry *et al.*, 2017).
867 Nonetheless, in optimal cases like our study, sequencing of reduced representation libraries,
868 or even whole genomes, might still fail to sample the loci involved in adaptation due to lack
869 of sequencing of important regions, but also because of the effects of background selection,
870 linkage, and how demography can affect the distribution of F_{ST} throughout the genome (see
871 Hoban *et al.*, 2016 for a review).

872 Additionally, common F_{ST} -based approaches, such as those used within this study can fail to
873 detect other structural aspects of the genome that can be involved in adaptation
874 (Wellenreuther *et al.*, 2019), including copy number variation (Schrider *et al.*, 2016; Nelson
875 *et al.*, 2019), recombination (Reeve *et al.*, 2016; Ortiz-Barrientos & James, 2017; Samuk *et*
876 *al.*, 2019), inversions (Kirkpatrick & Barton, 2006; Lowry & Willis, 2010; Faria *et al.*, 2019),
877 and transposons (González & Petrov, 2009; Schrader & Schmitz, 2019). Furthermore, the
878 alleles contributing to adaptation are also difficult to detect if adaptation has proceeded by
879 many alleles of small effect, which may be quite common in highly polygenic traits
880 (Yeaman, 2015). In such systems, divergent loci are difficult to detect as they contain low
881 F_{ST}, which further makes detecting genetic parallelism highly unlikely in these systems.
882 Finally, traditional approaches that sample SNPs across the genome will fail to detect
883 variation in gene expression levels, which may also be involved in adaptation during parallel
884 evolution (e.g., Rivas *et al.*, 2018; Verta & Jones, 2019).

885 The quantification of phenotypes may also experience sampling bias. We may be
886 unknowingly biased towards measuring traits that are divergent between ecotypes, rather than
887 the overall phenotype of an organism. We might also overlook parallelism in traits we did not
888 or could not measure, for instance extended adaptive phenotypes such flowering times within
889 plants (Blackman *et al.*, 2011; Christie & Strauss, 2019). Researchers may also disregard
890 traits that are difficult to measure (such as the root networks of plants), or the ability of an

organism to physiologically respond to a stimulus (such as tropisms in plants). We must also be aware that the likelihood of detecting parallelism is highly dependent on which traits are measured as well as the number of traits (Stayton, 2008), suggesting that theories of multi-trait evolution could help us develop better null hypothesis for parallel evolution while accounting for correlations between traits, including those traits that are highly pleiotropic (Yeaman, 2015; De Lisle & Bolnick, 2020).

The geometry and statistics of parallel evolution

When quantifying narrow-sense phenotypic parallelism with vectors, the interpretation of phenotypic change in lengths (ΔL) and angles (θ) between replicate pairs should be treated with caution. This is due to how phenotypic similarities manifest both within each replicate pair and between ecotypes. Consider two replicate pairs where the populations within each pair are more phenotypically similar than the populations belonging to the same ecotype (i.e. they had different common ancestors; Bolnick *et al.*, 2018). However, due to their phenotypic configuration in multivariate space, we may still measure the same change in length ($\Delta L \approx 0$) and angle ($\theta \approx 0°$) between pairs. This is apparent in the work by Elmer *et al.*, (2014), where they showed strong ecological parallelism in two replicate limnetic-benthic cichlid replicate pairs, despite the phenotypic differences being larger between pairs than between ecotypes within a pair (see Figure 3 from Elmer *et al.*, 2014). This highlights the necessity for systems of parallel evolution to have strong broad-sense phenotypic differences between all populations within each ecotype, or at a biological level that can be measured across all populations, such as the ecological axis used in Elmer *et al.*, (2014). This can help enrich the interpretation of parallel evolution even in some of the most famous cases of parallel evolution (e.g., Stuart *et al.*, 2017). In *S. lautus*, we have found a striking case of strong phenotypic differences Dune and Headland ecotypes despite variable levels of narrow-sense parallelism, suggesting that the origin of Headland populations is likely from a common ancestor. Further investigation of other biological axes of organisation will further complement our current view of this system, such as the contributions of physiology and network evolution to adaptation and divergence.

Comparing levels of parallel evolution across systems

To advance the field of parallel evolution, it is necessary to undertake statistical approaches that allow us to compare levels of parallelism across systems on a common scale. It would

922 also be useful to have key criteria to define the amount of parallelism in a system. Rather
923 than classifying evolution as "parallel" or "non-parallel", we should strive to place each
924 system on a continuum of parallel evolution, as suggested by the recent work by Bolnick *et*
925 *al.*, (2018) which discusses the "(non)-parallel" continuum. Not only does our framework
926 pave the way for utilising some common approaches, but it also attempts to clarify the
927 continuum of parallel evolution into broad- and narrow-sense scales at both the genotypic and
928 phenotypic levels. This is important as our work has shown that parallel evolution can
929 manifest in different ways at each of these scales.

930 One potential way forward to compare between systems is to use variance partitioning
931 methods (such as partial effect sizes; partial η^2) to gain an overall understanding of the
932 relative contributions of broad- and narrow-sense parallelism. This can be undertaken at the
933 multivariate and univariate levels of the phenotype as well as the level of the SNP and gene
934 for the genotype. We can then compare the distribution of effect sizes at these different levels
935 across study systems. Systems that have more broad-sense parallelism will show a larger
936 proportion of partial effect sizes (either at the phenotypic or genotypic level) where the
937 variance is explained by differences between the ecotypes rather than replicate pairs.
938 Researchers may also benefit from approaches such as the K-means clustering analysis,
939 where we can directly compare the percentage of individuals correctly assigned to each
940 ecotype (also see the work on "exchangeability" by Hendry *et al.*, 2013). For instance, for a
941 system with high broad-sense phenotypic parallelism at the level of the ecotype, we would be
942 able to easily predict which ecotype all individuals belong to based on their phenotypic traits.

943 Recent work by De Lisle & Bolnick (2020) advocate using theory from quantitative genetics
944 to quantify the direction of evolutionary change in multivariate space across population pairs.
945 It has advantages over the PCVA approach as rather than interpreting individual vectors and
946 angles, it uses the full matrix of correlations between populations and asks which dimensions
947 of multivariate space explain a significant amount of variance. One can then extract the
948 loadings from these dimensions to quantity the combination of traits the contribute to the
949 divergence between populations. Parallelism can be assessed by asking whether the loadings
950 are in the same direction across replicate populations. The work by De Lisle & Bolnick
951 (2020) also paves the way for generating biologically meaningful null hypotheses, which is
952 lacking in the field of parallel evolution (Bolnick *et al.*, 2018). Future work on null
953 hypotheses should also strive to model the likelihood of phenotypic or genotypic parallelism

954 while taking into account variance in factors such as gene flow, environmental heterogeneity
955 and the genetic architecture of adaptive traits (e.g., Thompson *et al.*, 2019).

956 To further advance the field of parallel evolution, researchers must also strive to specifically
957 identify the causal genes for adaptation by directly linking them to adaptive phenotypes and
958 further demonstrating that these phenotypes confer a fitness advantage to native populations.
959 In systems where adaptation is governed by selection on the same nucleotide sites (e.g.,
960 (Colosimo *et al.*, 2005), researchers must demonstrate whether the selected loci arose via
961 standing genetic variation or new mutations (Lee & Coop, 2017, 2019). The field of parallel
962 evolution also requires a more thorough understanding of how the demographic history of
963 populations impacts the likelihood of genotypic and phenotypic parallelism. Furthermore,
964 understanding the contributions of phenotypic plasticity within each system (by comparing
965 natural populations to those reared in common garden conditions; Oke *et al.*, 2016) will help
966 shed light on whether phenotypic parallelism is enhanced or dampened by non-genetic
967 changes.

968 We have outlined a framework to quantify both phenotypic and genotypic parallel evolution
969 within empirical systems. Applying this framework to *S. lautus*, we have demonstrated strong
970 phenotypic differences between the Dune and Headland ecotypes, although most replicate
971 pairs are somewhat phenotypically unique. Genotypic parallelism across pairs is mainly
972 driven by different mutations in different genes, but belonging to similar biological functions.
973 Our work provides progress towards a common framework to measure and compare parallel
974 evolution across natural systems.

975 Acknowledgements

976 We are grateful to A. Nguyen Vu for assisting with sample collection. We thank Ortiz-
977 Barrientos laboratory members for insightful comments on previous versions of this
978 manuscript. S. Chenoweth and M. Blows provided very useful feedback on M.E. James' PhD
979 dissertation. This work was funded by The University of Queensland and the Australian
980 Research Council grants to DO, and by an Australian Postgraduate Award and a Graduate
981 School International Travel Award to MEJ.

Author contributions

MEJ and DO conceived the project. MEJ, MW, HN and JE undertook sample collection and the field phenotypic measurements. MEJ extracted DNA, prepared libraries and performed bioinformatics. MEJ undertook all data analysis, and JE assisted with R code. MEJ and DO wrote the paper with input from all authors. DO is the mentor and supervisor for the research program.

Conflicts of interest

We do not have any conflicts of interest.

Data archival

Data will be uploaded to Dryad upon acceptance of the manuscript.

References

Adams, D.C. & Collyer, M.L. 2009. A general framework for the analysis of phenotypic trajectories in evolutionary studies. *Evolution* **63**: 1143–1154.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

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

Bailey, S.F., Guo, Q. & Bataillon, T. 2018. Identifying drivers of parallel evolution: A regression model approach. *Genome Biology and Evolution* **10**: 2801–2812.

Berg, J.J. & Coop, G. 2014. A population genetic signal of polygenic adaptation. *PLoS Genetics* **10**: e1004412.

Bhatia, G., Patterson, N., Sankararaman, S. & Price, A.L. 2013. Estimating and interpreting FST: The impact of rare variants. *Genome Research* **23**: 1514–1521.

Bierne, N., Gagnaire, P.A. & David, P. 2013. The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Current Zoology* **59**: 72–86.

Blackman, B.K., Michaels, S.D. & Rieseberg, L.H. 2011. Connecting the sun to flowering in sunflower adaptation. *Molecular Ecology* **20**: 3503–3512.

Blount, Z.D., Lenski, R.E. & Losos, J.B. 2018. Contingency and determinism in evolution: Replaying life's tape. *Science* **362**: eaam5979.

Bolnick, D.I., Barrett, R.D.H., Oke, K.B., Rennison, D.J. & Stuart, Y.E. 2018. (Non)Parallel evolution. *Annual Review of Ecology, Evolution, and Systematics* **49**: 303–330.

Boyle, E.A., Li, Y.I. & Pritchard, J.K. 2017. An expanded view of complex traits: From polygenic to omnigenic. *Cell* **169**: 1177–1186.

Broad Institute. 2019. *PicardTools*. Broad Institute, GitHub Repository. http://broadinstitute.github.io/picard/.

Butlin, R.K., Saura, M., Charrier, G., Jackson, B., André, C., Caballero, A., *et al.* 2014. Parallel evolution of local adaptation and reproductive isolation in the face of gene flow. *Evolution* **68**: 935–949.

Cai, Z., Zhou, L., Ren, N.N., Xu, X., Liu, R., Huang, L., *et al.* 2019. Parallel speciation of wild rice associated with habitat shifts. *Molecular Biology and Evolution* **36**: 875–889.

Cassin-Sackett, L., Callicrate, T.E. & Fleischer, R.C. 2019. Parallel evolution of gene classes, but not genes: Evidence from Hawai'ian honeycreeper populations exposed to avian malaria. *Molecular Ecology* **28**: 568–583.

Catchen, J.M., Hohenlohe, P.A., Bernatchez, L., Funk, W.C., Andrews, K.R. & Allendorf, F.W. 2017. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources* **17**: 362–365.

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Shapiro, M.D., Brady, S.D., *et al.* 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**: 302–305.

Chevin, L.M., Martin, G. & Lenormand, T. 2010. Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution: heterogeneous mutation effects across loci. *Evolution* **64**: 3213–3231.

Christie, K. & Strauss, S.Y. 2019. Reproductive isolation and the maintenance of species boundaries in two serpentine endemic Jewelflowers. *Evolution* **73**: 1375–1391.

Christin, P.A., Weinreich, D.M. & Besnard, G. 2010. Causes and evolutionary significance of genetic convergence. *Trends in Genetics* **26**: 400–405.

Clarke, J.D. 2009. Cetyltrimethyl Ammonium Bromide (CTAB) DNA miniprep for plant DNA isolation. *Cold Spring Harbor Protocols* **2009**: Pdb.prot5177.

Collyer, M.L. & Adams, D.C. 2007. Analysis of two-state multivariate phenotypic change in ecological studies. *Ecology* **88**: 683–692.

Collyer, M.L., Sekora, D.J. & Adams, D.C. 2015. A method for analysis of phenotypic change for phenotypes described by high-dimensional data. *Heredity* **115**: 357–365.

Colosimo, P.F., Hosemann, K.E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., *et al.* 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* **307**: 1928–1933.

Conte, G.L., Arnegard, M.E., Peichel, C.L. & Schluter, D. 2012. The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B: Biological Sciences* **279**: 5039–5047.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

De Lisle, S.P. & Bolnick, D.I. 2020. A multivariate view of parallel evolution. BioRxiv [Preprint]. https://biorxiv.org/lookup/doi/10.1101/2020.01.26.920439.

Deagle, B.E., Jones, F.C., Chan, Y.F., Absher, D.M., Kingsley, D.M. & Reimchen, T.E. 2012. Population genomics of parallel phenotypic evolution in stickleback across stream–lake ecological transitions. *Proceedings of the Royal Society B: Biological Sciences* **279**: 1277–1286.

Dean, L.L., Magalhaes, I.S., Foote, A., D'Agostino, D., McGowan, S. & MacColl, A.D.C. 2019. Admixture between ancient lineages, selection, and the formation of sympatric stickleback species-pairs. *Molecular Biology and Evolution* **36**: 2481–2497.

Dittmar, E.L., Oakley, C.G., Conner, J.K., Gould, B.A. & Schemske, D.W. 2016. Factors influencing the effect size distribution of adaptive substitutions. *Proceedings of the Royal Society B: Biological Sciences* **283**.

Elmer, K.R., Fan, S., Kusche, H., Luise Spreitzer, M., Kautt, A.F., Franchini, P., *et al.* 2014. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nature Communications* **5**: 5168.

Elmer, K.R., Kusche, H., Lehtonen, T.K. & Meyer, A. 2010. Local variation and parallel evolution: morphological and genetic diversity across a species complex of neotropical crater lake cichlid fishes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 1763–1782.

Elmer, K.R. & Meyer, A. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in Ecology & Evolution* **26**: 298–306.

Faria, R., Johannesson, K., Butlin, R.K. & Westram, A.M. 2019. Evolving inversions. *Trends in Ecology & Evolution* **34**: 239–248.

Foll, M. & Gaggiotti, O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* **180**: 977–993.

Foster, S.A., McKinnon, G.E., Steane, D.A., Potts, B.M. & Vaillancourt, R.E. 2007. Parallel evolution of dwarf ecotypes in the forest tree *Eucalyptus globulus*. *New Phytologist* **175**: 370–380.

Fox, J., Friendly, M. & Monette, G. 2018. *heplots: Visualizing Tests in Multivariate Linear Models*. R package version 1.3-5. https://CRAN.R-project.org/package=heplots.

Fraïsse, C. & Welch, J.J. 2019. The distribution of epistasis on simple fitness landscapes. *Biology Letters* **15**: 20180881.

Garrison, E. 2016. *Vcflib, a simple C++ library for parsing and manipulating VCF files.* https://github.com/vcflib/vcflib.

Garrison, E. & Marth, G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv. arXiv:1207.3907v2 [q-bio.GN].

Gompel, N. & Prud'homme, B. 2009. The causes of repeated genetic evolution. *Developmental Biology* **332**: 36–47.

González, J. & Petrov, D.A. 2009. The adaptive role of transposable elements in the Drosophila genome. *Gene* **448**: 124–133.

Haas, O. & Simpson, G.G. 1946. Analysis of some phylogenetic terms, with attempts at redefinition. *Proceedings of the American Philosophical Society* **90**: 319–349.

Hartigan, J.A. & Wong, M.A. 1979. A K-Means clustering algorithm. *Applied Statistics* **28**: 100.

Hendry, A.P., Kaeuffer, R., Crispo, E., Peichel, C.L. & Bolnick, D.I. 2013. Evolutionary inferences from the analysis of exchangeability. *Evolution* **67**: 3429–3441.

Hoban, S., Kelley, J.L., Lotterhos, K.E., Antolin, M.F., Bradburd, G., Lowry, D.B., *et al.* 2016. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist* **188**: 379–397.

Huang, D.W., Sherman, B.T. & Lempicki, R.A. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**: 1–13.

Huang, D.W., Sherman, B.T. & Lempicki, R.A. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**: 44–57.

Jacobs, A., Carruthers, M., Yurchenko, A., Gordeeva, N., Alekseyev, S., Hooker, O., *et al.* 2018. Convergence in form and function overcomes non-parallel evolutionary histories in a Holarctic fish. BioRxiv [Preprint]. https://doi.org/10.1101/265272.

James, M.E., Arenas-Castro, H., Groh, J.S., Engelstädter, J. & Ortiz-Barrientos, D. 2020. Highly replicated evolution of parapatric ecotypes. BioRxiv [Preprint].

Johannesson, K., Johannesson, B. & Rolán-Alvarez, E. 1993. Morphological differentiation and genetic cohesiveness over a microenvironmental gradient in the marine snail *Littorina saxatilis*. *Evolution* **47**: 1770–1787.

Johannesson, K., Panova, M., Kemppainen, P., André, C., Rolán-Alvarez, E. & Butlin, R.K. 2010. Repeated evolution of reproductive isolation in a marine snail: unveiling mechanisms of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 1735–1747.

Jones, F.C., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., *et al.* 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**: 55–61.

Kautt, A.F., Elmer, K.R. & Meyer, A. 2012. Genomic signatures of divergent selection and speciation patterns in a 'natural experiment', the young parallel radiations of Nicaraguan crater lake cichlid fishes. *Molecular Ecology* **21**: 4770–4786.

Kirkpatrick, M. & Barton, N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* **173**: 419–434.

Klepikova, A.V., Kasianov, A.S., Gerasimov, E.S., Logacheva, M.D. & Penin, A.A. 2016. A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *The Plant Journal* **88**: 1058–1070.

Konečná, V., Nowak, M.D. & Kolář, F. 2019. Parallel colonization of subalpine habitats in the central European mountains by *Primula elatior*. *Scientific Reports* **9**: 3294.

Kowalko, J.E., Rohner, N., Linden, T.A., Rompani, S.B., Warren, W.C., Borowsky, R., *et al.* 2013. Convergence in feeding posture occurs through different genetic loci in independently evolved cave populations of *Astyanax mexicanus*. *Proceedings of the National Academy of Sciences* **110**: 16933–16938.

Kusche, H., Elmer, K.R. & Meyer, A. 2015. Sympatric ecological divergence associated with a color polymorphism. *BMC Biology* **13**: 82.

Lamichhaney, S., Fuentes-Pardo, A.P., Rafati, N., Ryman, N., McCracken, G.R., Bourne, C., *et al.* 2017. Parallel adaptive evolution of geographically distant herring populations on both sides of the North Atlantic Ocean. *Proceedings of the National Academy of Sciences* **114**: E3452–E3461.

Langerhans, R.B. 2018. Predictability and parallelism of multitrait adaptation. *Journal of Heredity* **109**: 59–70.

Langerhans, R.B. & DeWitt, T.J. 2004. Shared and unique features of evolutionary diversification. *The American Naturalist* **164**: 335–349.

Laporte, M., Rogers, S.M., Dion-Côté, A.M., Normandeau, E., Gagnaire, P.A., Dalziel, A.C., *et al.* 2015. RAD-QTL mapping reveals both genome-level parallelism and different genetic architecture underlying the evolution of body shape in lake whitefish (*Coregonus clupeaformis*) species pairs. *Genes|Genomes|Genetics* **5**: 1481–1491.

Lee, K.M. & Coop, G. 2017. Distinguishing among modes of convergent adaptation using population genomic data. *Genetics* **207**: 1591–1619.

Lee, K.M. & Coop, G. 2019. Population genomics perspectives on convergent adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**: 20180236.

Lenormand, T., Chevin, L.M. & Bataillon, T. 2016. Parallel evolution: what does it (not) tell us and why is it (still) interesting? In: *Chance in Evolution*. Chicago, USA: University Chicago Press.

Lenormand, T., Roze, D. & Rousset, F. 2009. Stochasticity in evolution. *Trends in Ecology & Evolution* **24**: 157–165.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN].

Li, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.

Li, H. & Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Lim, M.C.W., Witt, C.C., Graham, C.H. & Dávalos, L.M. 2019. Parallel molecular evolution in pathways, genes, and sites in high-elevation hummingbirds revealed by comparative transcriptomics. *Genome Biology and Evolution* **11**: 1573–1585.

Liu, H. 2014. Developing genomic resources for an emerging ecological model species *Senecio lautus*. PhD Thesis. The University of Queensland.

Lobkovsky, A.E. & Koonin, E.V. 2012. Replaying the tape of life: quantification of the predictability of evolution. *Frontiers in Genetics* **3**.

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

Lowry, D.B., Hoban, S., Kelley, J.L., Lotterhos, K.E., Reed, L.K., Antolin, M.F., *et al.* 2017. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources* **17**: 142–152.

Lowry, D.B. & Willis, J.H. 2010. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology* **8**: e1000500.

MacPherson, A. & Nuismer, S.L. 2017. The probability of parallel genetic evolution from standing genetic variation. *Journal of Evolutionary Biology* **30**: 326–337.

Marques, D.A., Lucek, K., Sousa, V.C., Excoffier, L. & Seehausen, O. 2019. Admixture between old lineages facilitated contemporary ecological speciation in Lake Constance stickleback. *Nature Communications* **10**: 4240.

McKinney, G.J., Larson, W.A., Seeb, L.W. & Seeb, J.E. 2017. RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry *et al*. (2016). *Molecular Ecology Resources* **17**: 356–361.

Melo, M.C., Grealy, A., Brittain, B., Walter, G.M. & Ortiz-Barrientos, D. 2014. Strong extrinsic reproductive isolation between parapatric populations of an Australian groundsel. *New Phytologist* **203**: 323–334.

Melo, M.C., James, M.E., Roda, F., Bernal-Franco, D., Wilkinson, M.J., Liu, H., *et al.* 2019. Evidence for mutation-order speciation in an Australian wildflower. BioRxiv [Preprint]. doi: https://doi.org/10.1101/692673.

Nelson, T.C., Monnahan, P.J., McIntosh, M.K., Anderson, K., MacArthur-Waltz, E., Finseth, F.R., *et al.* 2019. Extreme copy number variation at a TRNA ligase gene affecting phenology and fitness in yellow monkeyflowers. *Molecular Ecology* **28**: 1460–1475.

Nosil, P., Crespi, B.J. & Sandoval, C.P. 2002. Host-plant adaptation drives the parallel evolution of reproductive isolation. *Nature* **417**: 440–443.

Oke, K.B., Bukhari, M., Kaeuffer, R., Rolshausen, G., Räsänen, K., Bolnick, D.I., *et al.* 2016. Does plasticity enhance or dampen phenotypic parallelism? A test with three lake-stream stickleback pairs. *Journal of Evolutionary Biology* **29**: 126–143.

Oke, K.B., Rolshausen, G., LeBlond, C. & Hendry, A.P. 2017. How parallel is parallel evolution? A comparative analysis in fishes. *The American Naturalist* **190**: 1–16.

Ord, T.J. & Summers, T.C. 2015. Repeated evolution and the impact of evolutionary history on adaptation. *BMC Evolutionary Biology* **15**: 137.

Ortiz-Barrientos, D. & James, M.E. 2017. Evolution of recombination rates and the genomic landscape of speciation. *Journal of Evolutionary Biology* **30**: 1519–1521.

Ostevik, K.L., Moyers, B.T., Owens, G.L. & Rieseberg, L.H. 2012. Parallel ecological speciation in plants? *International Journal of Ecology* **2012**: 1–17.

Owttrim, G.W. 2006. RNA helicases and abiotic stress. *Nucleic Acids Research* **34**: 3220–3230.

Pérez-Pereira, N., Quesada, H. & Caballero, A. 2017. Can parallel ecological speciation be detected with phylogenetic analyses? *Molecular Phylogenetics and Evolution* **116**: 149–156.

Perreault-Payette, A., Muir, A.M., Goetz, F., Perrier, C., Normandeau, E., Sirois, P., *et al.* 2017. Investigating the extent of parallelism in morphological and genomic divergence among lake trout ecotypes in Lake Superior. *Molecular Ecology* **26**: 1477–1497.

Pilakouta, N., Humble, J.L., Hill, I.D.C., Kristjánsson, B.K., Skúlason, S., Killen, S.S., *et al.* 2019. Testing the predictability of morphological evolution in naturally warmed stickleback populations. BioRxiv [Preprint]. doi: https://dx.doi.org/10.1101/609933.

Poland, J.A., Brown, P.J., Sorrells, M.E. & Jannink, J.L. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme Genotyping-by-Sequencing approach. *PLoS ONE* **7**: e32253.

Pruitt, K.D. 2004. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* **33**: D501–D504.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., *et al.* 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**: 559–575.

Puritz, J.B., Hollenbeck, C.M. & Gold, J.R. 2014. *dDocent*: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* **2**: e431.

Quesada, H., Posada, D., Caballero, A., Morán, P. & Rolán-Alvarez, E. 2007. Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. *Evolution* **61**: 1600–1612.

R Core Team. 2017. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Ralph, P. & Coop, G. 2010. Parallel adaptation: One or many waves of advance of an advantageous allele? *Genetics* **186**: 647–668.

Ravinet, M., Prodöhl, P.A. & Harrod, C. 2013. Parallel and nonparallel ecological, morphological and genetic divergence in lake-stream stickleback from a single catchment. *Journal of Evolutionary Biology* **26**: 186–204.

Reeve, J., Ortiz-Barrientos, D. & Engelstädter, J. 2016. The evolution of recombination rates in finite populations during ecological speciation. *Proceedings of the Royal Society B: Biological Sciences* **283**: 20161243.

Richards, T.J. & Ortiz-Barrientos, D. 2016. Immigrant inviability produces a strong barrier to gene flow between parapatric ecotypes of *Senecio lautus*. *Evolution* **70**: 1239–1248.

Richards, T.J., Walter, G.M., McGuigan, K. & Ortiz-Barrientos, D. 2016. Divergent natural selection drives the evolution of reproductive isolation in an Australian wildflower. *Evolution* **70**: 1993–2003.

Rivas, M.J., Saura, M., Pérez-Figueroa, A., Panova, M., Johansson, T., André, C., *et al.* 2018. Population genomics of parallel evolution in gene expression and gene sequence during ecological adaptation. *Scientific Reports* **8**: 16147.

Roda, F., Ambrose, L., Walter, G.M., Liu, H.L., Schaul, A., Lowe, A., *et al.* 2013a. Genomic evidence for the parallel evolution of coastal forms in the *Senecio lautus* complex. *Mol Ecol* **22**: 2941–2952.

Roda, F., Liu, H., Wilkinson, M.J., Walter, G.M., James, M.E., Bernal, D.M., *et al.* 2013b. Convergence and divergence during the adaptation to similar environments by an Australian groundsel. *Evolution* **67**: 2515–2529.

Rosenblum, E.B., Parent, C.E. & Brandt, E.E. 2014. The molecular basis of phenotypic convergence. *Annual Review of Ecology, Evolution, and Systematics* **45**: 203–226.

Salazar-Ciudad, I. & Marín-Riera, M. 2013. Adaptive dynamics under development-based genotype–phenotype maps. *Nature* **497**: 361–364.

Samuk, K., Manzano-Winkler, B., Ritz, K.R. & Noor, M.A.F. 2019. Natural selection shapes variation in genome-wide recombination rate in *Drosophila pseudoobscura*. BioRxiv [Preprint]. https://dx.doi.org/10.1101/787382.

Schluter, D. 2000. *The ecology of adaptive radiation*. Oxford, UK: Oxford University Press.

Schluter, D. & Nagel, L.M. 1995. Parallel speciation by natural selection. *The American Naturalist* **146**: 292–301.

Schmieder, R., Lim, Y.W., Rohwer, F. & Edwards, R. 2010. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. *BMC Bioinformatics* **11**: 341.

Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**: 671–675.

Schrader, L. & Schmitz, J. 2019. The impact of transposable elements in adaptive evolution. *Molecular Ecology* **28**: 1537–1549.

Schrider, D.R., Hahn, M.W. & Begun, D.J. 2016. Parallel evolution of copy-number variation across continents in *Drosophila melanogaster*. *Molecular Biology and Evolution* **33**: 1308–1316.

Smith, S.D. & Rausher, M.D. 2011. Gene loss and parallel evolution contribute to species difference in flower color. *Molecular Biology and Evolution* **28**: 2799–2810.

Soria-Carrasco, V., Gompert, Z., Comeault, A.A., Farkas, T.E., Parchman, T.L., Johnston, J.S., *et al.* 2014. Stick insect genomes reveal natural selection's role in parallel speciation. *Science* **344**: 738–742.

Speed, M.P. & Arbuckle, K. 2017. Quantification provides a conceptual basis for convergent evolution. *Biological Reviews* **92**: 815–829.

Stayton, C.T. 2008. Is convergence surprising? An examination of the frequency of convergence in simulated datasets. *Journal of Theoretical Biology* **252**: 1–14.

Stern, D.L. 2013. The genetic causes of convergent evolution. *Nature Reviews Genetics* **14**: 751–764.

Stern, D.L. & Orgogozo, V. 2009. Is genetic evolution predictable? *Science* **323**: 746–751.

Stoltzfus, A. & McCandlish, D.M. 2017. Mutational biases influence parallel adaptation. *Molecular Biology and Evolution* **34**: 2163–2172.

Storz, J.F. 2016. Causes of molecular convergence and parallelism in protein evolution. *Nature Reviews Genetics* **17**: 239–250.

Stuart, Y.E. 2019. Divergent uses of "parallel evolution" during the history of *The American Naturalist*. *The American Naturalist* **193**: 11–19.

Stuart, Y.E., Veen, T., Weber, J.N., Hanson, D., Ravinet, M., Lohman, B.K., *et al.* 2017. Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nature Ecology & Evolution* **1**: 0158.

Tenaillon, O., Rodriguez-Verdugo, A., Gaut, R.L., McDonald, P., Bennett, A.F., Long, A.D., *et al.* 2012. The molecular diversity of adaptive convergence. *Science* **335**: 457–461.

Thompson, K.A., Osmond, M.M. & Schluter, D. 2019. Parallel genetic evolution and speciation from standing variation. *Evolution Letters* **3**: 129–141.

Tiffin, P. & Ross-Ibarra, J. 2014. Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution* **29**: 673–680.

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

VanWallendael, A., Soltani, A., Emery, N.C., Peixoto, M.M., Olsen, J. & Lowry, D.B. 2019. A molecular view of plant local adaptation: Incorporating stress-response networks. *Annual Review of Plant Biology* **70**: 559–583.

Verta, J.P. & Jones, F.C. 2019. Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *eLife* **8**: e43785.

Walter, G.M. 2016. The genetic and ecological basis of diversification. PhD Thesis. The University of Queensland.

Walter, G.M., Aguirre, J.D., Blows, M.W. & Ortiz-Barrientos, D. 2018a. Evolution of genetic variance during adaptive radiation. *The American Naturalist* **191**: E108–E128.

Walter, G.M., Wilkinson, M.J., Aguirre, J.D., Blows, M.W. & Ortiz-Barrientos, D. 2018b. Environmentally induced development costs underlie fitness tradeoffs. *Ecology* **99**: 1391–1401.

Walter, G.M., Wilkinson, M.J., James, M.E., Richards, T.J., Aguirre, J.D. & Ortiz-Barrientos, D. 2016. Diversification across a heterogeneous landscape. *Evolution* **70**: 1979–1992.

Weir, B.S. & Cockerham, C.C. 1984. Estimating F-Statistics for the analysis of population structure. *Evolution* **38**: 1358.

Wellenreuther, M., Mérot, C., Berdan, E. & Bernatchez, L. 2019. Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. *Molecular Ecology* **28**: 1203–1209.

Wilkinson, M.J. 2019. The genetic basis of adaptive evolution and divergence in an Australian wildflower. PhD Thesis. The University of Queensland.

Wilkinson, M.J., Roda, F., Walter, G.M., James, M.E., Nipper, R., Walsh, J., *et al.* 2019. Divergence in hormone signalling links local adaptation and hybrid failure. BioRxiv [Preprint]. https://doi.org/10.1101/845354.

Wood, T.E., Burke, J.M. & Rieseberg, L.H. 2005. Parallel genotypic adaptation: when evolution repeats itself. *Genetica* **123**: 157–170.

Yeaman, S. 2015. Local adaptation by alleles of small effect. *The American Naturalist* **186**: S74–S89.

Yeaman, S., Gerstein, A.C., Hodgins, K.A. & Whitlock, M.C. 2018. Quantifying how constraints limit the diversity of viable routes to adaptation. *PLOS Genetics* **14**: e1007717.

Figures

Figure 1. A framework to measure phenotypic and genotypic parallel evolution

Parallel evolution can be broken down into two components: the phenotype and the genotype. These two components can be measured in the broad-sense (the overall differences between ecotypes), and the narrow-sense (the specific similarities between replicate pairs). Broad-sense phenotypic parallelism can be measured at the level of the ecotype and trait, whereas narrow-sense phenotypic parallelism is measured at the level of the pair. Genotypic parallelism can be measured at different biological levels of organisation in both the broad- and narrow-sense scales: the nucleotide site, gene, and biological function.

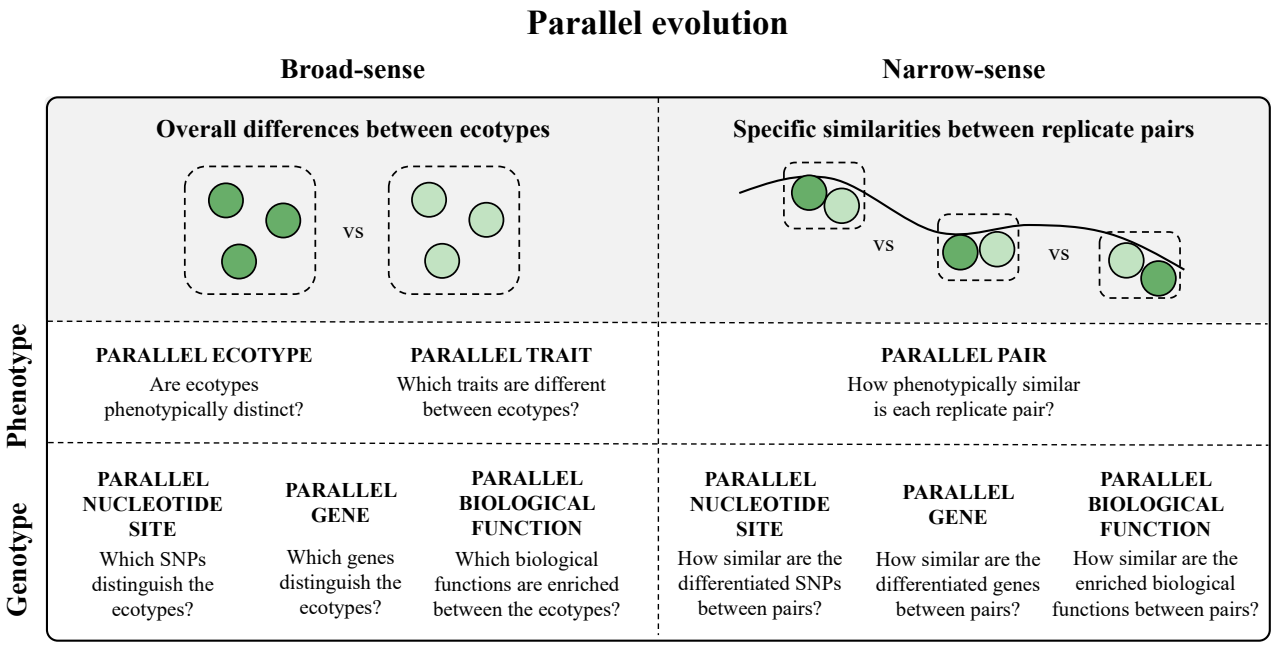


Figure 2. Broad-sense phenotypic parallelism in *S. lautus*: parallel ecotype

(A) Sampling locations of the 22 Dune (orange) and Headland (green) *Senecio lautus* populations along the coast of Australia. **(B)** Maximum likelihood phylogeny of Dune and Headland populations implemented in IQ-TREE. Numbers on each node represent the SH-alRT support (%), followed by the ultrafast bootstrap support (%). Modified within permission from James *et al.*, (2020). The intermediate population (H12A) is not included within this study. **(C)** Partial effect sizes (partial η^2) for the ecotype and the interaction (ecotype * pair) for the trait-by-trait linear models, each dot representing a single trait. The red dot represents Wilk's partial effect size for all traits combined within the MANOVA. Dashed line is a 1:1 ratio, where points above the line represent a larger contribution of broad- to narrow-sense divergence. See Table S4 for details. **(D)** Schematic diagram of Dune and Headland ecotypes based on mean trait values from linear discriminant analysis shown in Figure 3C. **(E)** Principle component analysis of Dune and Headland phenotypes (five architecture and four leaf traits) across 20 populations. Ecotypes are delimited by 70% probability ellipses.

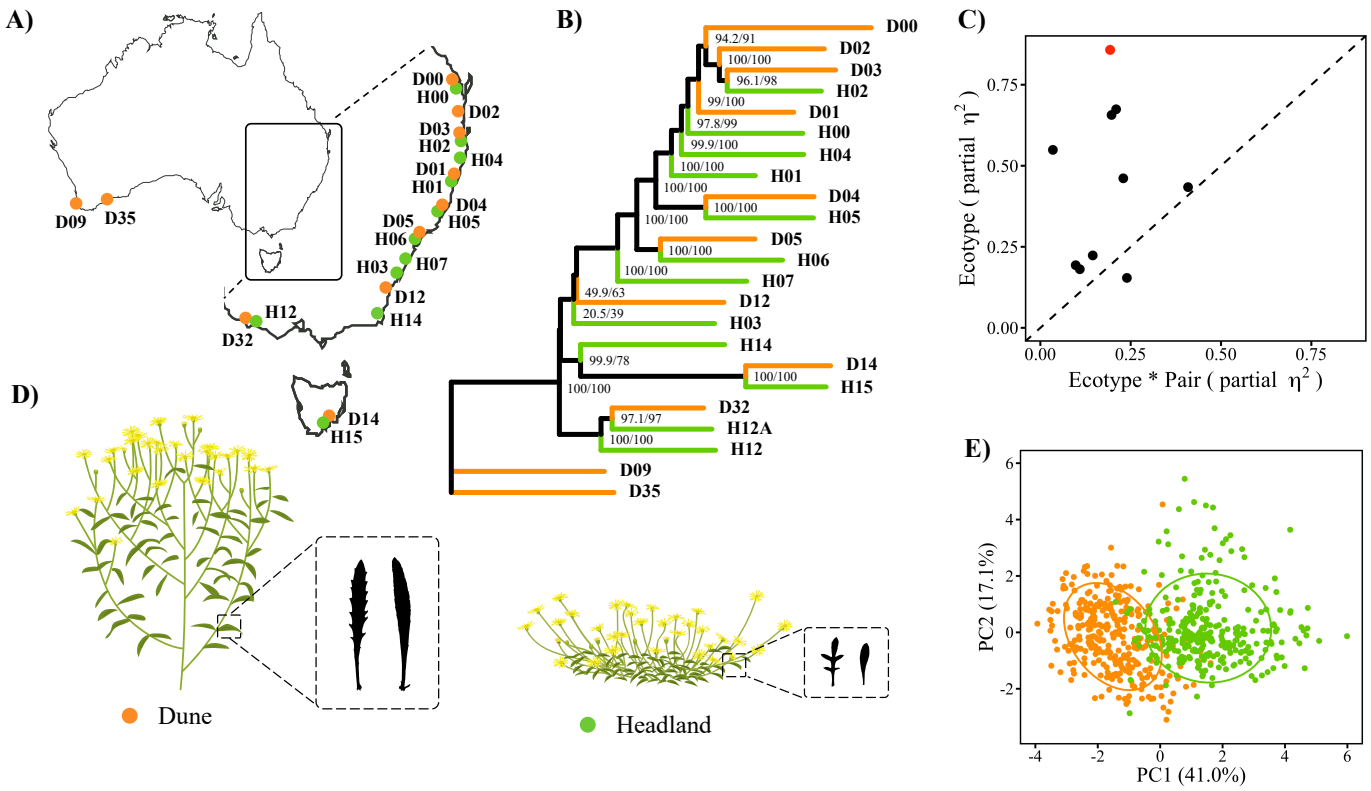
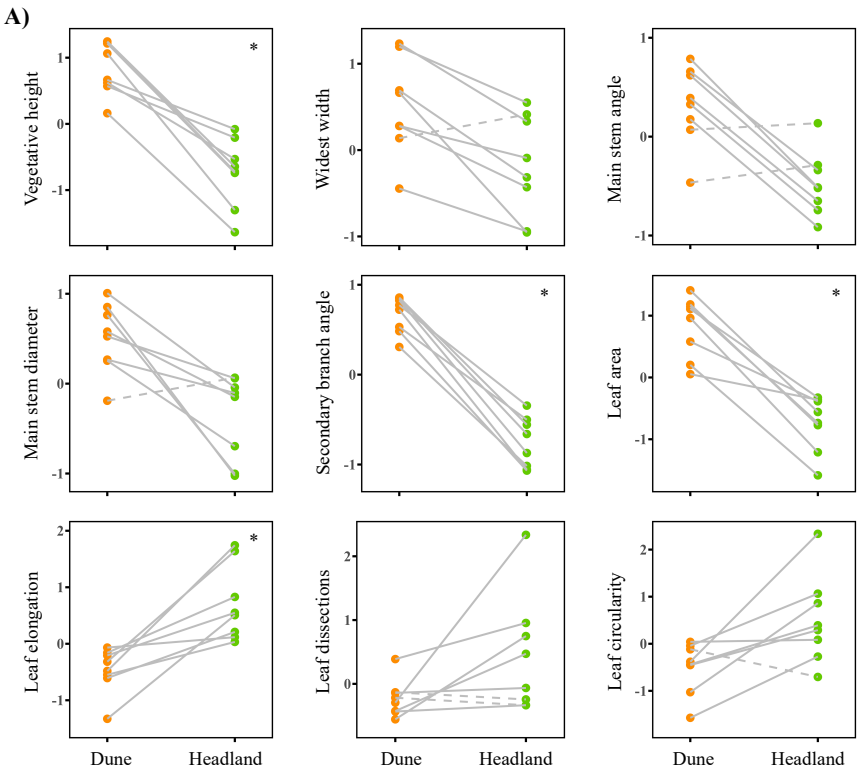
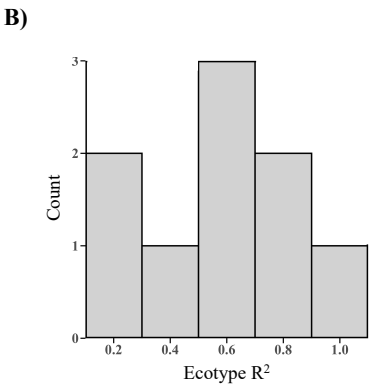


Figure 3. Broad-sense phenotypic parallelism in *S. lautus*: parallel trait

(A) Vote-counting for five architecture and four leaf traits across eight replicate pairs. Dots represent the mean trait value for each population (N = 30). Lines connect the Dune (orange) populations to their Headland (green) pair. Dashed lines represent pairs whose Dune-Headland trait value is in the opposite direction from the majority of pairs. Asterisks denote significance (* P<0.01). **(B)** Frequency distribution of the ecotype R^2 values (proportion of phenotypic variance explained by the ecotype term) for each trait from one-way ANOVAs performed on the population means. **(C)** Linear discriminant analysis of all traits between Dune and Headland ecotypes.





C)

Trait Category	Trait	LDA
Architecture	Vegetative height	**-0.26**
	Widest width	**0.41**
	Main stem angle	-0.15
	Main stem diameter	**-0.36**
	Secondary branch angle	**-0.61**
Leaf	Area	**-0.78**
	Elongation	-0.01
	Dissection	**0.51**
	Circularity	**0.45**

Figure 4. Narrow-sense phenotypic parallelism in *S. lautus*: parallel pair

(A) PC1 and PC2 for five architecture and four leaf traits across eight replicate Dune-Headland pairs. Each dot represents the population centroid (multivariate phenotypic mean), ± SE. The Dune (orange) and Headland (green) populations of a replicate pair are connected with a line. **(B)** Frequency distribution of the 28 pairwise phenotypic divergences (ΔL) between Dune-Headland replicate pairs (Supplementary Table S6). **(C)** Frequency distribution of the 28 pairwise contribution of traits (θ) between Dune-Headland replicate pairs (Supplementary Table S9).

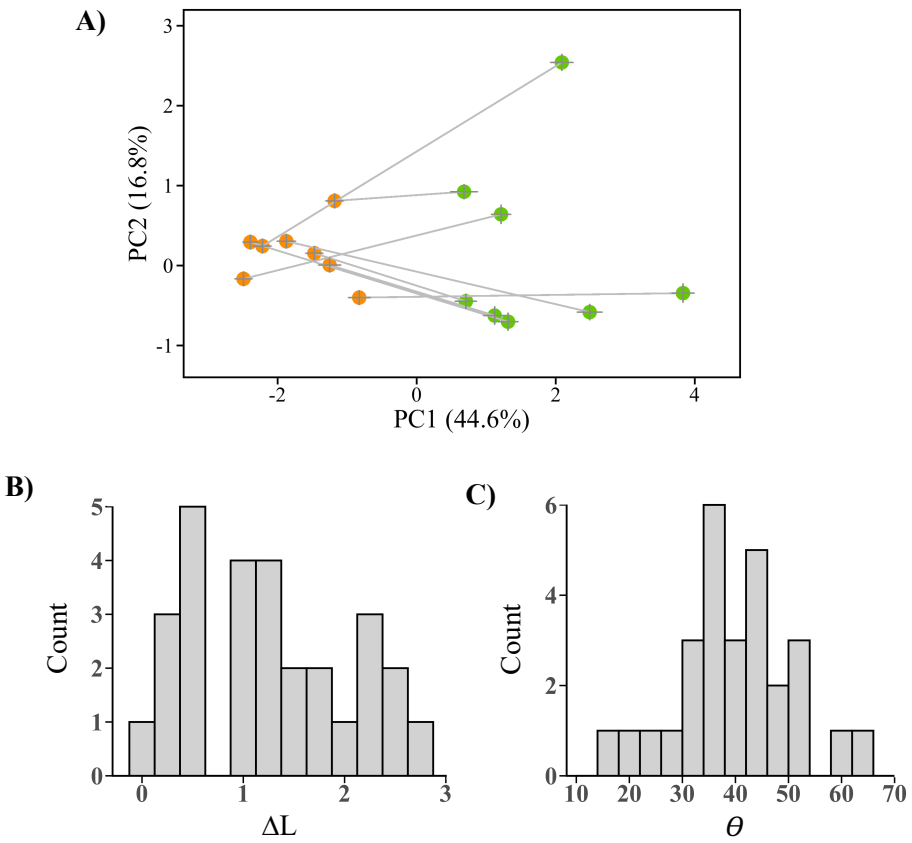


Figure 5. Relative contributions of broad- and narrow-sense genotypic parallelism in *S. lautus*

Partial effect sizes (partial η^2) for the ecotype and the interaction (ecotype * pair) for the linear models for nucleotide sites **(A)** genes **(B)**. Each dot represents either a single SNP **(A)** or gene **(B)**. Dashed line is a 1:1 ratio, where points above the line represent a larger contribution of broad- to narrow-sense parallelism. The red dotes denotes broad-sense outlier SNPs **(A)** and genes **(B)**. Partial effect size frequency distributions for the nucleotide sites **(C)** and genes **(D)**. Values above zero represent a larger contribution of broad- to narrow-sense parallelism.

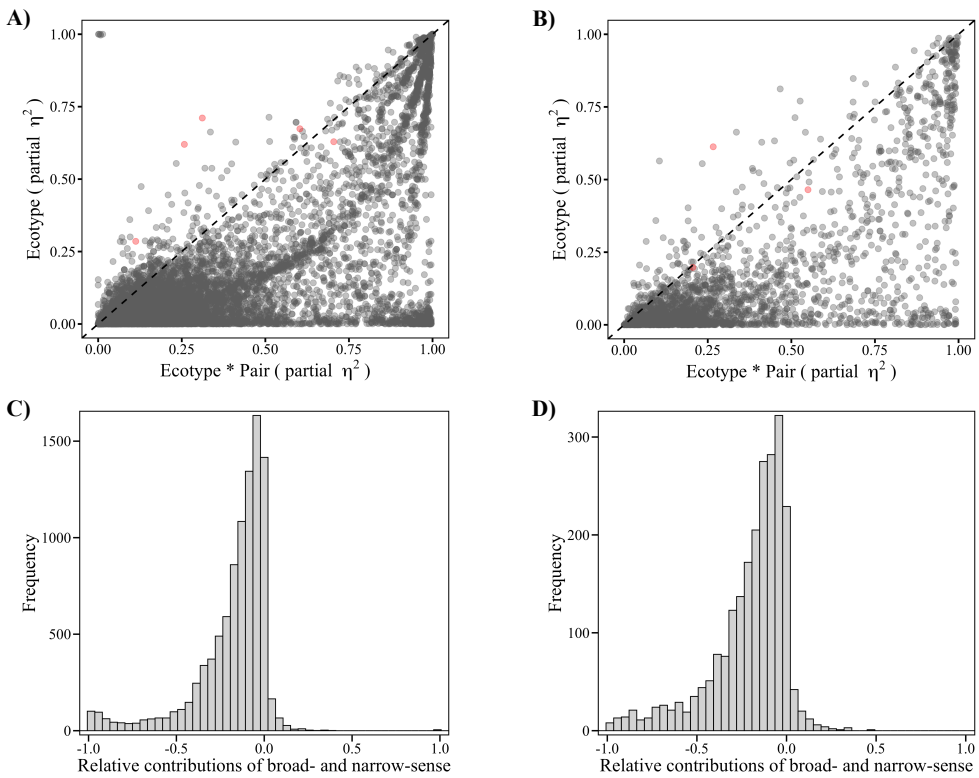
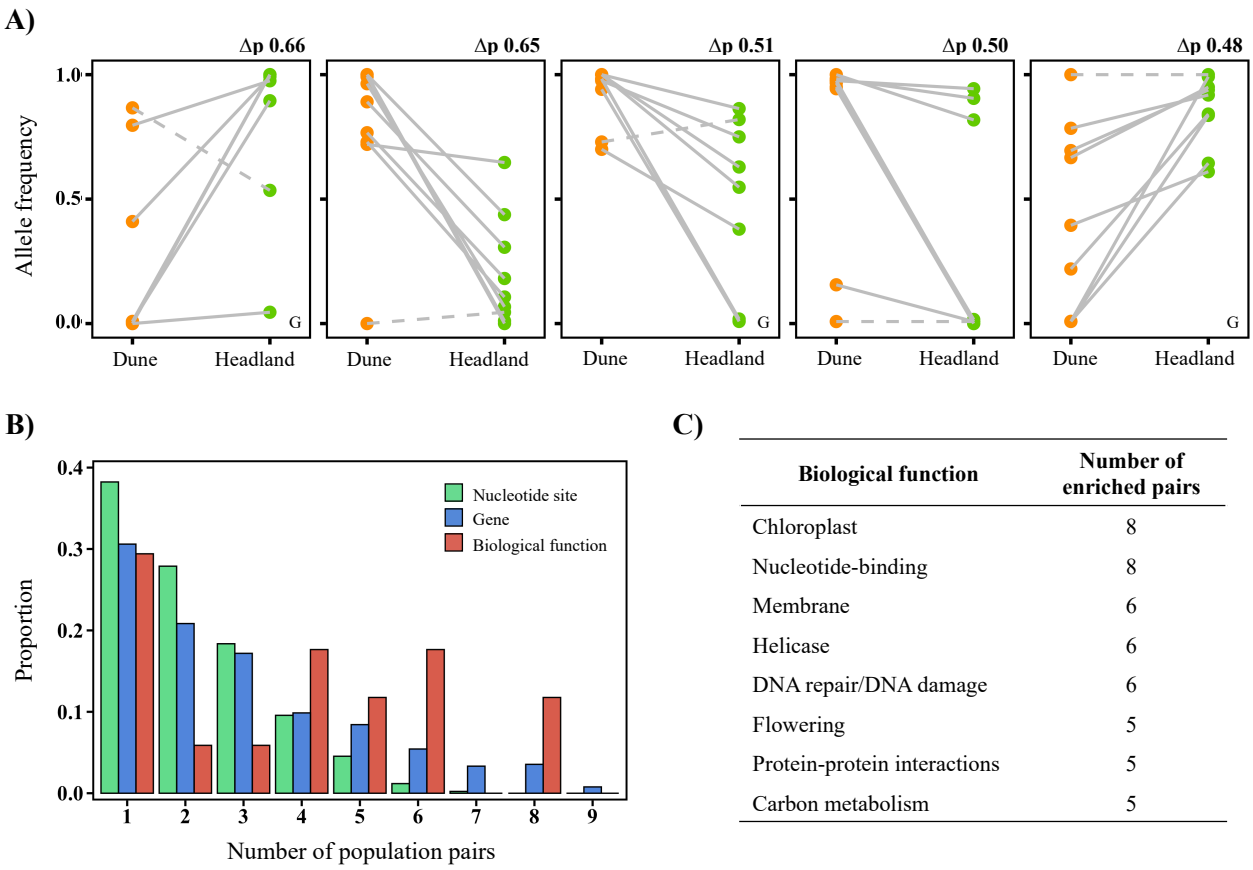


Figure 6. Broad- and narrow-sense genotypic parallelism in *S. lautus*

(A) Broad-sense parallel nucleotide sites showing high differentiation between the Dune-Headland ecotypes as well as concordant allele frequency changes across replicate pairs. Dots represent the allele frequency value (of the reference allele) for each population. Lines connect the Dune (orange) populations to their Headland (green) pair. Dashed lines represent pairs whose Dune-Headland change in allele frequency is in the opposite direction from the majority of pairs. Δ*p* denotes the overall change in allele frequency between the ecotypes. *G* denotes SNPs that occur within genic regions. **(B)** Proportion of outlier nucleotide sites, outlier genes, and enriched biological functions shared across the nine replicate pairs. **(C)** Enriched biological functions shared across five or more replicate population pairs.



C)

Biological function	Number of enriched pairs
Chloroplast	8
Nucleotide-binding	8
Membrane	6
Helicase	6
DNA repair/DNA damage	6
Flowering	5
Protein-protein interactions	5
Carbon metabolism	5

Supplementary tables and figures

Table S1. *Senecio lautus* sampling locations in this study

Sampling locations of the 22 *Senecio lautus* Dune and Headland populations. Coordinates represent the mid-point of each population. N genetics and N phenotype represent the sample sizes used for the genotypic and phenotypic analysis respectively. Note: N genetics corresponds to the final number of individuals after removing those with low coverage. Pairs in bold are sister-taxa within the phylogeny.

Clade	Population code	Location	Ecotype	Pair	Coordinates	N genetics	N phenotype
Eastern	D00	QLD: Stradbroke Island	Dune	D00-H00	S27° 31.153' E153° 30.189'	62	30
Eastern	H00	QLD: Stradbroke Island	Headland	D00-H00	S27° 26.140' E153° 32.749'	63	30
Eastern	D02	QLD: Southport	Dune	D02-H04	S27° 56.846' E153° 25.736'	62	30
Eastern	H04	NSW: Byron Bay	Headland	D02-H04	S28° 38.060' E153° 38.268'	62	30
Eastern	D03	NSW: Cabarita	Dune	**D03-H02**	S28° 19.794' E153° 34.264'	61	30
Eastern	H02	NSW: Cabarita	Headland	**D03-H02**	S28° 21.013' E153° 34.676'	61	30
Eastern	D01	QLD: Lennox Head	Dune	D01-H01	S28° 46.858' E153° 35.655'	60	30
Eastern	H01	QLD: Lennox Head	Headland	D01-H01	S28° 48.813' E153° 36.313'	58	30
Eastern	D04	NSW: Coffs Harbour	Dune	**D04-H05**	S30° 18.946' E153° 08.142'	62	30
Eastern	H05	NSW: Coffs Harbour	Headland	**D04-H05**	S30° 18.741' E153° 08.676'	62	30
Eastern	D05	NSW: South West Rocks	Dune	**D05-H06**	S30° 53.027' E153° 04.037'	62	30
Eastern	H06	NSW: South West Rocks	Headland	**D05-H06**	S30° 52.710' E153° 04.549'	62	30
Southern	H07	NSW: Port Macquarie	Headland	-	S31° 28.526' E152° 56.219	-	30
Southern	H03	NSW: Kiama	Headland	-	S34° 40.301' E150° 51.704'	-	29
Southern	D12	NSW: Bermagui	Dune	D12-H14	S36° 28.346' E150° 03.581'	62	31
Southern	H14	NSW: Green Cape	Headland	D12-H14	S37° 15.748' E150° 02.991'	62	30
Southern	D32	VIC: Cape Bridgewater	Dune	**D32-H12**	S38° 19.631' E141° 23.772'	62	30
Southern	H12	VIC: Cape Bridgewater	Headland	**D32-H12**	S38° 22.728' E141° 22.018'	63	30
Southern	D14	TAS: Port Arthur	Dune	**D14-H15**	S43° 10.550' E147° 51.267'	12	-
Southern	H15	TAS: Port Arthur	Headland	**D14-H15**	S43° 11.240' E147° 50.672'	11	-
Western	D35	WA: Isthmus Hill	Dune	-	S35° 05.885' E117° 59.182'	-	34
Western	D09	WA: Leeuwin-Naturaliste National Park	Dune	-	S33° 46.239' E114° 59.541'	-	31

Table S2. *Senecio lautus* phenotypic traits

Traits with asterisks were removed from the analysis due to high correlations (>0.8).

Trait category	Trait	Description
Architecture	Vegetative height	Base of plant to highest vegetative leaf (not including flowers)
	Widest width	Widest with of plant between vegetative leaves
	Narrowest width*	Narrowest width of plant between vegetative leaves
	Main stem angle	Angle of main stem measured from base of soil
	Main stem diameter	Diameter of main stem measured 1cm above soil
	Secondary branch angle	Average angle of secondary branches
Leaf	Area	Leaf area
	Perimeter*	Leaf perimeter
	Width*	Leaf width
	Height*	Leaf height
	Elongation	Length to width ratio
	Compactness*	Squared perimeter to area ratio
	Dissection	Perimeter to length ratio
	Circularity	4π (area/perimeter^2)

Table S3. Demographic and environmental data for *S. lautus* population pairs

Rates of gene flow (2Nm) and divergence times were calculated in *fastsimcoal2* (see James *et al.*, (2020) for details). 2Nm (D -> H): gene flow since secondary contact from the Dune to the Headland. 2Nm (H -> D): gene flow since secondary contact from the Headland to the Dune. See (Roda *et al.*, 2013b) for details of the environmental distances.

Pair	2Nm (D -> H)	2Nm (H -> D)	2Nm (ABS)	Divergence time (years)	Environmental distance
D00-H00	0.2176	0.2830	0.065424	71945	3.23
D02-H04	0.2697	0.1720	0.097695	82539	NA
D03-H02	0.1590	0.4722	0.313209	44190	10.26
D01-H01	0.6241	0.3671	0.256984	71918	13.68
D04-H05	**1.3942**	**1.5024**	0.108212	30707	6.31
D05-H06	0.4049	0.4325	0.027621	67927	NA
D12-H14	0.2188	0.1787	0.040083	47929	5.14
D14-H15	0.3952	0.5187	0.123542	110018	NA
D32-H12	**5.5694**	**5.2694**	0.300057	38706	NA

Table S4. Trait-by-trait linear models and partial effect sizes

F-values and P-values for the ecotype (Ecotype $F_{1,465}$; Ecotype P-value) and interaction (Ecotype x Pair $F_{7,465}$; Ecotype x Pair P-value) for trait-by-trait linear models (*trait = ecotype + pair + ecotype x pair*). Partial effect sizes (partial η^2) for the Ecotype, Pair and Interaction (Ecotype x Pair) for each trait.

Trait Category	Trait	Ecotype $F_{1,465}$	Ecotype P-value	Ecotype x Pair $F_{7,465}$	Ecotype x Pair P-value	Ecotype partial η^2	Pair partial η^2	Ecotype x Pair partial η^2
Architecture	Vegetative height	962.65	$< 2.2 \times 10^{-16}$ ***	17.63	$< 2.2 \times 10^{-16}$ ***	0.67	0.26	0.21
	Widest width	102.62	$< 2.2 \times 10^{-16}$ ***	8.14	2.3×10^{-9} ***	0.18	0.30	0.11
	Main stem angle	111.35	$< 2.2 \times 10^{-16}$ ***	7.20	3.4×10^{-8} ***	0.19	0.06	0.10
	Main stem diameter	133.75	$< 2.2 \times 10^{-16}$ ***	11.26	3.3×10^{-13} ***	0.22	0.07	0.14
	Secondary branch angle	566.05	$< 2.2 \times 10^{-16}$ ***	2.39	0.02092 *	0.55	0.08	0.03
Leaf	Area	889.52	$< 2.2 \times 10^{-16}$ ***	16.30	$< 2.2 \times 10^{-16}$ ***	0.66	0.26	0.20
	Elongation	398.11	$< 2.2 \times 10^{-16}$ ***	19.79	$< 2.2 \times 10^{-16}$ ***	0.46	0.26	0.23
	Dissection	84.64	$< 2.2 \times 10^{-16}$ ***	20.92	$< 2.2 \times 10^{-16}$ ***	0.15	0.25	0.24
	Circularity	356.68	$< 2.2 \times 10^{-16}$ ***	45.92	$< 2.2 \times 10^{-16}$ ***	0.43	0.46	0.41

Table S5. Summary of parallel phenotypic traits in *S. lautus*

Summary of the three approaches to measure phenotypic parallelism per trait (vote-counting, trait-by-trait linear models, and R^2). 'Yes' denotes a trait was either statistically significant using the vote-counting approach (Vote-counting), had a significant ecotype term in the trait-by-trait linear models (Linear models), or an $R^2 \geq 0.50$ for one-way ANOVAs performed on the population means (Ecotype R^2). Traits in bold are consistently viewed as parallel across each of the three approaches.

Trait Category	Trait	Vote-counting	Linear models	Ecotype R^2
Architecture	**Vegetative height**	Yes	Yes	Yes
	Widest width	No	Yes	No
	Main stem angle	No	Yes	Yes
	Main stem diameter	No	Yes	Yes
	Secondary branch angle	Yes	Yes	Yes
Leaf	**Area**	Yes	Yes	Yes
	Elongation	Yes	Yes	Yes
	Dissection	No	Yes	No
	Circularity	No	Yes	No

Table S6. Pairwise change in lengths (ΔL) for all traits in coastal ecotypes of *S. lautus*

Values below the diagonal represent the change in lengths, above the diagonal are the P-values. Green cells donate parallel pairs (P>0.01).

	D00-H00	**D01-H01**	**D02-H04**	**D03-H02**	**D04-H05**	**D05-H06**	**D12-H14**	**D32-H12**
D00-H00	-	0.3239	0.0001	0.0001	0.0001	0.4949	0.0001	0.0001
D01-H01	0.27	-	0.0001	0.0001	0.0001	0.0763	0.0001	0.0001
D02-H04	2.45	2.72	-	0.0548	0.0002	0.0001	0.0001	0.5683
D03-H02	1.86	2.13	0.59	-	0.0435	0.0001	0.0367	0.1627
D04-H05	1.27	1.54	1.18	0.59	-	0.0003	0.8839	0.0003
D05-H06	0.20	0.47	2.26	1.66	1.08	-	0.0007	0.0001
D12-H14	1.24	1.51	1.22	0.62	0.04	1.04	-	0.0005
D32-H12	2.29	2.56	0.16	0.43	1.02	2.10	1.05	-

Table S7. Pairwise change in lengths (ΔL) for architecture traits in coastal ecotypes of *S. lautus*

Values below the diagonal represent the change in lengths, above the diagonal are the P-values. Green cells donate parallel pairs (P>0.01).

	D00-H00	**D01-H01**	**D02-H04**	**D03-H02**	**D04-H05**	**D05-H06**	**D12-H14**	**D32-H12**
D00-H00	-	0.1366	0.0001	0.0001	0.0001	0.6428	0.0016	0.0008
D01-H01	0.44	-	0.0001	0.0001	0.0001	0.0394	0.0002	0.0001
D02-H04	1.55	1.99	-	0.0926	0.4009	0.0001	0.0477	0.1153
D03-H02	2.08	2.52	0.53	-	0.0056	0.0001	0.0001	0.0017
D04-H05	1.32	1.76	0.23	0.77	-	0.0001	0.1633	0.3648
D05-H06	0.14	0.58	1.41	1.94	1.18	-	0.0070	0.0032
D12-H14	0.94	1.38	0.61	1.14	0.37	0.80	-	0.6626
D32-H12	1.07	1.52	0.48	1.01	0.24	0.94	0.13	-

Table S8. Pairwise change in lengths (ΔL) for leaf traits in coastal ecotypes of *S. lautus*

Values below the diagonal represent the change in lengths, above the diagonal are the P-values. Green cells donate parallel pairs (P>0.01).

	D00-H00	**D01-H01**	**D02-H04**	**D03-H02**	**D04-H05**	**D05-H06**	**D12-H14**	**D32-H12**
D00-H00	-	0.9650	0.0001	0.0740	0.0239	0.4917	0.0006	0.0001
D01-H01	0.01	-	0.0001	0.1268	0.0509	0.5771	0.0046	0.0001
D02-H04	1.90	1.89	-	0.0001	0.0001	0.0001	0.0015	0.4492
D03-H02	0.34	0.34	1.56	-	0.6506	0.3263	0.0352	0.0001
D04-H05	0.43	0.42	1.47	0.09	-	0.1646	0.0674	0.0001
D05-H06	0.14	0.13	1.77	0.21	0.30	-	0.0052	0.0001
D12-H14	0.81	0.80	1.10	0.46	0.37	0.67	-	0.0001
D32-H12	2.08	2.07	0.17	1.73	1.64	1.94	1.27	-

Table S9. Pairwise angles (θ) for all traits in coastal ecotypes of *S. lautus*

Values below the diagonal represent the angles (in degrees), above the diagonal are the P-values. Green cells donate parallel pairs (P>0.01).

	D00-H00	**D01-H01**	**D02-H04**	**D03-H02**	**D04-H05**	**D05-H06**	**D12-H14**	**D32-H12**
D00-H00	-	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
D01-H01	43.38	-	0.0001	0.0001	0.0001	0.0001	0.0018	0.0001
D02-H04	43.73	49.83	-	0.0001	0.0001	0.0001	0.0001	0.0001
D03-H02	31.00	50.52	40.33	-	0.0105	0.0001	0.0001	0.0001
D04-H05	26.57	42.70	42.93	14.51	-	0.0001	0.0001	0.0001
D05-H06	61.69	49.81	62.81	39.29	39.30	-	0.0001	0.0001
D12-H14	50.10	35.99	34.81	34.08	35.69	35.34	-	0.0010
D32-H12	50.27	34.82	45.36	33.59	33.74	24.43	20.14	-

Table S10. Pairwise angles (θ) for architecture traits in coastal ecotypes of *S. lautus*

Values below the diagonal represent the angles (in degrees), above the diagonal are the P-values. Green cells donate parallel pairs (P>0.01).

	D00-H00	**D01-H01**	**D02-H04**	**D03-H02**	**D04-H05**	**D05-H06**	**D12-H14**	**D32-H12**
D00-H00	-	0.0021	0.0025	0.0049	0.0022	0.0003	0.0013	0.0005
D01-H01	49.70	-	0.0001	0.0001	0.0001	0.0005	0.0002	0.0001
D02-H04	26.78	62.07	-	0.2492	0.0270	0.0002	0.0002	0.0161
D03-H02	23.10	60.56	9.27	-	0.0395	0.0001	0.0009	0.0043
D04-H05	27.06	50.46	15.63	13.48	-	0.0198	0.1135	0.5276
D05-H06	43.11	50.21	30.17	30.37	18.95	-	0.0051	0.0835
D12-H14	31.30	44.17	26.61	22.66	13.08	23.40	-	0.1199
D32-H12	33.60	50.15	18.37	19.47	8.49	17.08	14.04	-

Table S11. Pairwise angles (θ) for leaf traits in coastal ecotypes of *S. lautus*

Values below the diagonal represent the angles (in degrees), above the diagonal are the P-values. Green cells donate parallel pairs (P>0.01).

	D00-H00	**D01-H01**	**D02-H04**	**D03-H02**	**D04-H05**	**D05-H06**	**D12-H14**	**D32-H12**
D00-H00	-	0.0014	0.0001	0.0001	0.0012	0.0001	0.0001	0.0001
D01-H01	37.73	-	0.0001	0.1043	0.0172	0.0001	0.0531	0.0009
D02-H04	54.25	40.54	-	0.0001	0.0001	0.0001	0.0001	0.0001
D03-H02	31.56	19.33	58.04	-	0.0907	0.0001	0.0002	0.0001
D04-H05	21.88	26.54	60.06	11.42	-	0.0001	0.0001	0.0001
D05-H06	75.54	48.76	82.96	44.26	54.12	-	0.0001	0.0001
D12-H14	63.73	26.53	40.49	42.27	51.98	43.90	-	0.0060
D32-H12	58.83	24.12	56.47	29.60	40.89	26.79	20.31	-

Table S12. Broad- and narrow-sense outlier genes between replicate population pairs from coastal ecotypes of *S. lautus*

Functional annotation of the outlier genes within *DAVID*. Gene ontology is the *DAVID GOTERM_BP_DIRECT* (denoted below with *BP*), where absent, the *GOTERM_CC_DIRECT* was used (*CC*), and where absent, the *GOTERM_MF_DIRECT* (*MF*) was used. *BS – A1* denotes genes detected as broad-sense outliers using approach 1 (see methods). *BS – A2* denotes genes detected as broad-sense outliers using approach 2. *NS* denotes genes detected in the narrow-sense that are common to either all nine (9) or eight (8) replicate pairs.

Gene	Gene ontology	BS - A1	BS - A2	NS
Galactose oxidase/kelch repeat superfamily protein(AT5G04420)	CC = GO:0005634~nucleus,GO:0005737~cytoplasm,GO:0005829~cytosol,	Y	Y	N
basic salivary proline-rich-like protein (DUF1421)(AT5G14540)	CC = GO:0005634~nucleus,GO:0005829~cytosol,	Y	Y	N
Metallo-hydrolase/oxidoreductase superfamily protein(AT1G25375)	CC = GO:0009536~plastid,GO:0016021~integral component of membrane,	Y	N	9
basic helix-loop-helix (bHLH) DNA-binding superfamily protein(AT1G29950)	BP = GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,	Y	N	9
nodulin MtN21 /EamA-like transporter family protein(UMAMIT14)	BP = GO:0044746~amino acid transmembrane export,GO:0048316~seed development,GO:0098712~L-glutamate import across plasma membrane,	Y	N	8
ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein(AT3G63380)	BP = GO:0070588~calcium ion transmembrane transport,	Y	N	N
Argonaute family protein(AGO6)	BP = GO:0006306~DNA methylation,GO:0006342~chromatin silencing,GO:0006351~transcription, DNA-templated,GO:0006417~regulation of translation,GO:0030422~production of siRNA involved in RNA interference,GO:0051607~defense response to virus,	Y	N	N
Chaperone protein htpG family protein(CR88)	BP = GO:0006457~protein folding,GO:0006950~response to stress,GO:0009408~response to heat,GO:0009414~response to water deprivation,GO:0009651~response to salt stress,GO:0009704~de-etiolation,GO:0009793~embryo development ending in seed dormancy,GO:0010157~response to chlorate,GO:0045037~protein import into chloroplast stroma,	Y	N	N

Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein(AT4G31350)	CC = GO:0005739~mitochondrion,GO:0016021~integral component of membrane,	Y	N	N
EMS-MUTAGENIZED BRI1 SUPPRESSOR 1(EBS1)	BP = GO:0006486~protein glycosylation,GO:0009626~plant-type hypersensitive response,GO:0009751~response to salicylic acid,GO:0010204~defense response signaling pathway, resistance gene-independent,GO:0016051~carbohydrate biosynthetic process,GO:0018279~protein N-linked glycosylation via asparagine,GO:0030968~endoplasmic reticulum unfolded protein response,GO:0046283~anthocyanin-containing compound metabolic process,GO:0051084~'de novo' posttranslational protein folding,GO:0071712~ER-associated misfolded protein catabolic process,	Y	N	N
FMN-linked oxidoreductases superfamily protein(AT3G63510)	BP = GO:0006808~regulation of nitrogen utilization,GO:0055114~oxidation-reduction process,	Y	N	N
GDSL-like Lipase/Acylhydrolase superfamily protein(AT1G33811)	BP = GO:0016042~lipid catabolic process,	Y	N	N
HSI2-like 1(HSL1)	BP = GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0010030~positive regulation of seed germination,GO:2000034~regulation of seed maturation,	Y	N	N
Haloacid dehalogenase-like hydrolase (HAD) superfamily protein(AT5G44730)	BP = GO:0008152~metabolic process,	Y	N	N
Ku80 family protein(KU80)	BP = GO:0000723~telomere maintenance,GO:0006281~DNA repair,GO:0006302~double-strand break repair,GO:0006303~double-strand break repair via nonhomologous end joining,GO:0006310~DNA recombination,GO:0009408~response to heat,GO:0015074~DNA integration,GO:0032508~DNA duplex unwinding,GO:0071480~cellular response to gamma radiation,GO:0071481~cellular response to X-ray,	Y	N	N
Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family(AT3G54200)	CC = GO:0005634~nucleus,GO:0005886~plasma membrane,GO:0009506~plasmodesma,GO:0016021~integral component of membrane,GO:0046658~anchored component of plasma membrane,	Y	N	N
Leucine-rich receptor-like protein kinase family protein(AT5G56040)	BP = GO:0006468~protein phosphorylation,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,GO:0010082~regulation of root meristem growth,	Y	N	N
NADP-malic enzyme 3(NADP-ME3)	BP = GO:0006090~pyruvate metabolic process,GO:0006108~malate metabolic process,GO:0051260~protein homooligomerization,GO:0055114~oxidation-reduction process,	Y	N	N
Plant invertase/pectin methylesterase inhibitor superfamily(AT5G27870)	BP = GO:0042545~cell wall modification,GO:0045490~pectin catabolic process,	Y	N	N
RECQ helicase SIM(RECQSIM)	BP = GO:0000724~double-strand break repair via homologous recombination,GO:0006281~DNA repair,GO:0032508~DNA duplex unwinding,	Y	N	N
RING/FYVE/PHD zinc finger superfamily protein(AT5G16680)	BP = GO:0006355~regulation of transcription, DNA-templated,	Y	N	N

RNA helicase, ATP-dependent, SK12/DOB1 protein(MTR4)	BP = GO:0000460~maturation of 5.8S rRNA,GO:0006364~rRNA processing,GO:0006401~RNA catabolic process,GO:0031125~rRNA 3'-end processing,	Y	N	N
RNA-binding (RRM/RBD/RNP motifs) family protein(AT2G44710)	MF = GO:0000166~nucleotide binding,GO:0003676~nucleic acid binding,GO:0003723~RNA binding,	Y	N	N
Ribosomal protein S7e family protein(AT3G02560)	BP = GO:0006364~rRNA processing,GO:0006412~translation,GO:0042274~ribosomal small subunit biogenesis,	Y	N	N
TRF-like 10(TRFL10)	BP = GO:0009751~response to salicylic acid,GO:0031627~telomeric loop formation,	Y	N	N
alpha-crystallin domain 32.1(ACD32.1)	BP = GO:0009408~response to heat,	Y	N	N
casein kinase 1-like protein 2(CKL2)	BP = GO:0006897~endocytosis,GO:0008360~regulation of cell shape,GO:0016055~Wnt signaling pathway,GO:0018105~peptidyl-serine phosphorylation,	Y	N	N
cation-chloride co-transporter 1(CCC1)	BP = GO:0006812~cation transport,GO:0006814~sodium ion transport,	Y	N	N
chloroplastic acetylcoenzyme A carboxylase 1(CAC1)	BP = GO:0006633~fatty acid biosynthetic process,	Y	N	N
electron transport SCO1/SenC family protein(HCC1)	BP = GO:0008535~respiratory chain complex IV assembly,GO:0009793~embryo development ending in seed dormancy,GO:0033617~mitochondrial respiratory chain complex IV assembly,GO:0045454~cell redox homeostasis,	Y	N	N
embryo defective 2737(emb2737)	BP = GO:0009793~embryo development ending in seed dormancy,	Y	N	N
endoplasmatic reticulum retrieval protein 1B(RER1B)	BP = GO:0006890~retrograde vesicle-mediated transport, Golgi to ER,	Y	N	N
ethylene response factor 110(ERF110)	BP = GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0009873~ethylene-activated signaling pathway,GO:0045893~positive regulation of transcription, DNA-templated,GO:0048510~regulation of timing of transition from vegetative to reproductive phase,	Y	N	N
gamma subunit of Mt ATP synthase(ATP3)	BP = GO:0015986~ATP synthesis coupled proton transport,GO:0015992~proton transport,	Y	N	N
hypothetical protein(AT2G44850)	CC = GO:0005576~extracellular region,	Y	N	N
inter alpha-trypsin inhibitor, heavy chain-like protein(AT1G72500)	CC = GO:0005576~extracellular region,GO:0005886~plasma membrane,	Y	N	N

lysine-ketoglutarate reductase/saccharopine dehydrogenase bifunctional enzyme(AT4G33150)	BP = GO:0019477~L-lysine catabolic process,GO:0033512~L-lysine catabolic process to acetyl-CoA via saccharopine,GO:0080167~response to karrikin,	Y	N	N
modifier of snc1(MOS1)	BP = GO:0016569~covalent chromatin modification,GO:0040029~regulation of gene expression, epigenetic,	Y	N	N
pol-like 5(PLL5)	BP = GO:0006470~protein dephosphorylation,GO:0048366~leaf development,	Y	N	N
proline transporter 1(PROT1)	BP = GO:0006865~amino acid transport,GO:0015824~proline transport,	Y	N	N
sucrose transporter 4(SUT4)	BP = GO:0005985~sucrose metabolic process,GO:0015770~sucrose transport,	Y	N	N
symplekin(AT1G27595)	BP = GO:0071333~cellular response to glucose stimulus,	Y	N	N
tetratricopeptide repeat (TPR)-containing protein(TSK)	BP = GO:0000086~G2/M transition of mitotic cell cycle,GO:0009933~meristem structural organization,GO:0009934~regulation of meristem structural organization,GO:0016569~covalent chromatin modification,GO:0051301~cell division,	Y	N	N
zeaxanthin epoxidase (ZEP) (ABA1)(ABA1)	BP = GO:0006970~response to osmotic stress,GO:0009408~response to heat,GO:0009414~response to water deprivation,GO:0009688~abscisic acid biosynthetic process,GO:0010114~response to red light,GO:0010182~sugar mediated signaling pathway,GO:0016123~xanthophyll biosynthetic process,GO:0044550~secondary metabolite biosynthetic process,GO:0055114~oxidation-reduction process,	Y	N	N
AAA-type ATPase family protein(EMB3144)	BP = GO:0006508~proteolysis,GO:0009793~embryo development ending in seed dormancy,	N	Y	N
cyclic nucleotide gated channel 1(CNGC1)	BP = GO:0006813~potassium ion transport,GO:0006816~calcium ion transport,GO:0042391~regulation of membrane potential,	N	Y	N
glutamine-tRNA ligase, putative / glutaminyl-tRNA synthetase, putative / GlnRS(OVA9)	BP = GO:0006412~translation,GO:0006424~glutamyl-tRNA aminoacylation,GO:0006425~glutaminyl-tRNA aminoacylation,GO:0048481~plant ovule development,	N	Y	N
hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase(HCT)	BP = GO:0009809~lignin biosynthetic process,GO:0009963~positive regulation of flavonoid biosynthetic process,GO:0010252~auxin homeostasis,	N	Y	N
zinc ion binding protein(AT5G65740)	BP = GO:0006281~DNA repair,	N	Y	N
ARF GAP-like zinc finger-containing protein ZIGA4(ZIGA4)	BP = GO:0046777~protein autophosphorylation,	N	N	9

BRI1-associated receptor kinase(BAK1)	BP = GO:0002229~defense response to oomycetes,GO:0006468~protein phosphorylation,GO:0008219~cell death,GO:0009742~brassinosteroid mediated signaling pathway,GO:0016049~cell growth,GO:0042742~defense response to bacterium,GO:0050832~defense response to fungus,	N	N	9
MOS4-associated complex 3B(MAC3B)	BP = GO:0000349~generation of catalytic spliceosome for first transesterification step,GO:0006281~DNA repair,GO:0006397~mRNA processing,GO:0008380~RNA splicing,GO:0016567~protein ubiquitination,GO:0042742~defense response to bacterium,GO:0045087~innate immune response,GO:0070534~protein K63-linked ubiquitination,	N	N	9
AGAMOUS-like 42(AGL42)	BP = GO:0000165~MAPK cascade,GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0009838~abscission,GO:0009909~regulation of flower development,GO:0010150~leaf senescence,GO:0030154~cell differentiation,GO:0045944~positive regulation of transcription from RNA polymerase II promoter,GO:0080187~floral organ senescence,	N	N	8
Auxin-responsive GH3 family protein(GH3.1)	BP = GO:0009733~response to auxin,	N	N	8
CW7(CW7)	CC = GO:0005634~nucleus,	N	N	8
Early-responsive to dehydration stress protein (ERD4)(AT1G69450)	BP = GO:0006811~ion transport,	N	N	8
Fes1A(Fes1A)	BP = GO:0009408~response to heat,GO:0009651~response to salt stress,	N	N	8
Glycosyl transferase family 4 protein(AT3G57220)	BP = GO:0000271~polysaccharide biosynthetic process,GO:0006047~UDP-N-acetylglucosamine metabolic process,GO:0006487~protein N-linked glycosylation,GO:0006488~dolichol-linked oligosaccharide biosynthetic process,GO:0019348~dolichol metabolic process,	N	N	8
Homeodomain-like superfamily protein(AT5G52660)	BP = GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0009651~response to salt stress,GO:0009723~response to ethylene,GO:0009737~response to abscisic acid,GO:0009739~response to gibberellin,GO:0009751~response to salicylic acid,GO:0009753~response to jasmonic acid,GO:0042752~regulation of circadian rhythm,GO:0046686~response to cadmium ion,	N	N	8
LMBR1-like membrane protein(AT5G65290)	CC = GO:0009507~chloroplast,GO:0016021~integral component of membrane,	N	N	8
Leucine-rich repeat transmembrane protein kinase(GSO1)	BP = GO:0006468~protein phosphorylation,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,GO:0009793~embryo development ending in seed dormancy,GO:0090558~plant epidermis development,	N	N	8
Major facilitator superfamily protein(AT3G16180)	BP = GO:0006857~oligopeptide transport,GO:0080055~low-affinity nitrate transport,	N	N	8
Malectin/receptor-like protein kinase family protein(AT4G39110)	BP = GO:0006468~protein phosphorylation,	N	N	8

Myosin heavy chain-related protein(ZYP1a)	BP = GO:0007129~synapsis,GO:0007131~reciprocal meiotic recombination,GO:0051301~cell division,	N	N	8
NAD(P)-linked oxidoreductase superfamily protein(AT1G59960)	BP = GO:0009651~response to salt stress,GO:0055114~oxidation-reduction process,	N	N	8
P-loop containing nucleoside triphosphate hydrolases superfamily protein(AT4G30100)	CC = GO:0005634~nucleus,GO:0009506~plasmodesma,	N	N	8
Protein phosphatase 2C family protein(AT1G47380)	BP = GO:0006470~protein dephosphorylation,	N	N	8
RNA-binding CRS1 / YhbY (CRM) domain-containing protein(CAF2)	BP = GO:0000373~Group II intron splicing,GO:0006397~mRNA processing,	N	N	8
SAC3/GANP/Nin1/mts3/eIF-3 p25 family(SAC3C)	CC = GO:0005634~nucleus,GO:0005739~mitochondrion,	N	N	8
TGACG motif-binding factor 4(TGA4)	BP = GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0009409~response to cold,GO:0042742~defense response to bacterium,	N	N	8
Tetratricopeptide repeat (TPR)-like superfamily protein(AT1G02370)	CC = GO:0005739~mitochondrion,	N	N	8
UBQ, helicase-c and DEAD-like helicase domain-containing protein(AT5G08110)	BP = GO:0010501~RNA secondary structure unwinding,	N	N	8
actin-related protein 5(ARP5)	BP = GO:0006281~DNA repair,GO:0006355~regulation of transcription, DNA-templated,GO:0007275~multicellular organism development,GO:0010082~regulation of root meristem growth,GO:0030029~actin filament-based process,GO:0080036~regulation of cytokinin-activated signaling pathway,	N	N	8
adenylate kinase 1(ADK1)	BP = GO:0046686~response to cadmium ion,	N	N	8
beta-glucosidase 47(BGLU47)	BP = GO:0005975~carbohydrate metabolic process,GO:1901657~glycosyl compound metabolic process,	N	N	8
cell elongation protein / DWARF1 / DIMINUTO (DIM)(DWF1)	BP = GO:0006694~steroid biosynthetic process,GO:0009416~response to light stimulus,GO:0009808~lignin metabolic process,GO:0009826~unidimensional cell growth,GO:0009834~plant-type secondary cell wall biogenesis,GO:0016132~brassinosteroid biosynthetic process,GO:0055114~oxidation-reduction process,	N	N	8
glyoxal oxidase-related protein(AT1G75620)	CC = GO:0005576~extracellular region,	N	N	8

heat-inducible transcription repressor (DUF639)(AT3G18350)	CC = GO:0005739~mitochondrion,GO:0016021~integral component of membrane,	N	N	8
microtubule-associated proteins 70-2(MAP70-2)	BP = GO:0007010~cytoskeleton organization,	N	N	8
mitochondrially targeted single-stranded DNA binding protein(MTSSB)	BP = GO:0006260~DNA replication,	N	N	8
pentatricopeptide (PPR) repeat-containing protein(AT1G62914)	CC = GO:0005739~mitochondrion,	N	N	8
zinc finger CCCH domain-containing protein 44(AT3G51120)	CC = GO:0005634~nucleus,	N	N	8

Table S13. Shared outlier nucleotide sites between replicate population pairs from coastal ecotypes of *S. lautus*

Values below the diagonal represent the number of shared outlier nucleotide sites. Values above the diagonal are the P-values of the probability that the common nucleotide sites are shared by chance, calculated from the hypergeometric distribution.

	D00-H00	**D01-H01**	**D02-H04**	**D03-H02**	**D04-H05**	**D05-H06**	**D12-H14**	**D14-H15**	**D32-H12**
D00-H00	-	3.75E-58	7.72E-54	3.98E-37	5.31E-51	8.48E-42	8.13E-50	2.74E-07	1.68E-32
D01-H01	169	-	3.59E-58	9.08E-52	1.14E-62	1.52E-48	1.39E-51	4.07E-05	6.49E-33
D02-H04	189	161	-	1.58E-49	3.52E-32	1.10E-53	1.20E-37	1.80E-03	3.40E-44
D03-H02	178	164	187	-	8.27E-46	1.36E-46	3.76E-44	2.63E-05	6.47E-36
D04-H05	195	174	155	192	-	5.31E-51	3.60E-44	1.07E-05	4.69E-34
D05-H06	165	144	173	176	178	-	2.27E-58	1.15E-11	7.98E-49
D12-H14	226	185	193	222	216	218	-	6.91E-10	7.40E-68
D14-H15	31	21	21	28	28	35	41	-	1.13E-03
D32-H12	207	164	215	220	210	214	311	30	-

Table S14. Shared outlier genes between replicate population pairs from coastal ecotypes of *S. lautus*

Values below the diagonal represent the number of shared outlier genes. Values above the diagonal are the P-values of the probability that the common genes are shared by chance, calculated from the hypergeometric distribution.

	D00-H00	**D01-H01**	**D02-H04**	**D03-H02**	**D04-H05**	**D05-H06**	**D12-H14**	**D14-H15**	**D32-H12**
D00-H00	-	2.10E-40	2.76E-53	1.54E-34	6.83E-33	2.47E-39	9.75E-24	2.05E-05	1.01E-19
D01-H01	123	-	1.13E-52	1.18E-44	5.03E-45	2.07E-35	2.88E-29	7.56E-04	1.79E-22
D02-H04	156	128	-	8.25E-47	2.06E-34	1.26E-42	5.36E-31	3.48E-05	1.15E-34
D03-H02	137	122	142	-	1.11E-34	1.78E-35	2.78E-22	7.95E-05	1.16E-24
D04-H05	141	127	133	136	-	6.85E-34	1.52E-20	2.30E-03	7.76E-20
D05-H06	146	114	140	134	138	-	1.32E-32	1.25E-08	5.61E-24
D12-H14	171	140	168	158	164	179	-	1.50E-03	2.34E-17
D14-H15	32	22	29	29	27	37	40	-	0.10091
D32-H12	168	133	177	166	167	170	231	35	-

Table S15. Shared biological functions between replicate population pairs from coastal ecotypes of *S. lautus*

Summary of the 17 enriched biological functions across the 9 replicate pairs inferred in *DAVID*. P-values represent the EASE score, a modified Fisher Exact P-value. 'NA' denotes the function was not enriched within the pair.

Biological function	D00-H00	D02-H04	D03-H02	D01-H01	D04-H05	D05-H06	D12-H14	D14-H15	D32-H12
Chloroplast	7.25E-04	4.91E-05	1.08E-04	6.63E-03	8.56E-03	3.02E-02	3.57E-10	NA	4.79E-05
Nucleotide-binding	1.12E-05	3.05E-03	7.00E-07	5.32E-04	7.74E-04	2.33E-02	7.87E-03	NA	3.77E-07
Membrane	8.50E-04	NA	NA	3.55E-05	1.22E-02	2.25E-02	1.67E-03	NA	9.48E-03
Helicase	2.38E-03	NA	7.20E-03	NA	2.69E-02	1.62E-02	8.08E-04	NA	3.90E-02
DNA repair/DNA damage	NA	3.81E-02	4.32E-02	7.01E-04	4.65E-02	NA	1.42E-03	NA	3.31E-02
Flowering	NA	NA	3.36E-04	5.95E-03	NA	3.60E-02	NA	2.79E-02	1.24E-02
Protein-protein interactions	NA	NA	NA	8.00E-03	1.90E-02	2.96E-02	NA	3.38E-02	1.32E-03
Carbon metabolism	2.37E-02	NA	NA	2.35E-02	4.40E-02	NA	NA	5.93E-03	2.14E-02
Zinc finger	2.47E-02	2.25E-02	NA	NA	NA	1.73E-02	2.28E-03	NA	NA
Nucleus	NA	2.59E-02	1.34E-02	NA	NA	NA	4.36E-03	NA	4.44E-03
ATPase dependent activity	NA	NA	4.08E-03	NA	NA	4.02E-03	4.20E-04	3.55E-02	NA
Catalytic activity	NA	NA	1.34E-02	NA	NA	6.62E-03	5.54E-04	NA	NA
Stress response	NA	NA	NA	NA	NA	3.18E-02	9.14E-03	NA	NA
Auxin pathway	NA	NA	NA	NA	NA	1.16E-02	NA	NA	NA
Amino acid transport	NA	NA	NA	NA	NA	NA	NA	NA	4.29E-03
Glycoprotein	1.63E-02	NA	NA	NA	NA	NA	NA	NA	NA
Endoplasmic reticulum	NA	NA	NA	NA	NA	3.80E-02	NA	NA	NA

Table S16. Chloroplast outlier genes

Functional annotation of the outlier genes within the enriched biological function of 'chloroplast'. Gene ontology is the *DAVID GOTERM_BP_DIRECT* (denoted below with *BP*), and where absent, the *GOTERM_CC_DIRECT* was used (*CC*). *Number of replicate* pairs denotes how many replicate pairs the gene was detected as an outlier.

Gene	Gene ontology	Number replicate pairs
mitochondrially targeted single-stranded DNA binding protein(MTSSB)	BP = GO:0006260~DNA replication,	7
pentatricopeptide (PPR) repeat-containing protein(AT1G62914)	CC = GO:0005739~mitochondrion,	7
RNA-binding CRS1 / YhbY (CRM) domain-containing protein(CAF2)	BP = GO:0000373~Group II intron splicing,GO:0006397~mRNA processing,	7
Tetratricopeptide repeat (TPR)-like superfamily protein(AT1G02370)	CC = GO:0005739~mitochondrion,	7
chloroplastic acetylcoenzyme A carboxylase 1(CAC1)	BP = GO:0006633~fatty acid biosynthetic process,	6
FtsH extracellular protease family(FtsHi1)	BP = GO:0006508~proteolysis,GO:0009658~chloroplast organization,GO:0009793~embryo development ending in seed dormancy,GO:0010020~chloroplast fission,GO:0043572~plastid fission,	6
Ribosomal L28 family(PRPL28)	BP = GO:0006412~translation,	6
Ribosomal protein L1p/L10e family(EMB3126)	BP = GO:0000470~maturation of LSU-rRNA,GO:0006412~translation,	6
translocase inner membrane subunit 44-2(TIM44-2)	BP = GO:0006626~protein targeting to mitochondrion,GO:0030150~protein import into mitochondrial matrix,	6
Tubulin/FtsZ family protein(FTSZ2-1)	BP = GO:0010020~chloroplast fission,	6
zeaxanthin epoxidase (ZEP) (ABA1)(ABA1)	BP = GO:0006970~response to osmotic stress,GO:0009408~response to heat,GO:0009414~response to water deprivation,GO:0009688~abscisic acid biosynthetic process,GO:0010114~response to red light,GO:0010182~sugar mediated signaling pathway,GO:0016123~xanthophyll biosynthetic process,GO:0044550~secondary metabolite biosynthetic process,GO:0055114~oxidation-reduction process,	6

ATP-dependent Clp protease(AT1G33360)	BP = GO:0006457~protein folding,GO:0006508~proteolysis,GO:0030163~protein catabolic process,	5
electron transport SCO1/SenC family protein(HCC1)	BP = GO:0008535~respiratory chain complex IV assembly,GO:0009793~embryo development ending in seed dormancy,GO:0033617~mitochondrial respiratory chain complex IV assembly,GO:0045454~cell redox homeostasis,	5
gamma subunit of Mt ATP synthase(ATP3)	BP = GO:0015986~ATP synthesis coupled proton transport,GO:0015992~proton transport,	5
lon protease 4(LON4)	BP = GO:0001666~response to hypoxia,GO:0006515~misfolded or incompletely synthesized protein catabolic process,GO:0007005~mitochondrion organization,GO:0034599~cellular response to oxidative stress,GO:0051131~chaperone-mediated protein complex assembly,GO:0051260~protein homooligomerization,GO:0070407~oxidation-dependent protein catabolic process,GO:0090296~regulation of mitochondrial DNA replication,	5
NAD(P)-binding Rossmann-fold superfamily protein(Tic62)	BP = GO:0015031~protein transport,	5
Transketolase family protein(AT2G34590)	BP = GO:0006096~glycolytic process,GO:0008152~metabolic process,GO:0048868~pollen tube development,GO:0055114~oxidation-reduction process,	5
vacuolar proton ATPase A3(VHA-A3)	BP = GO:0007035~vacuolar acidification,GO:0015986~ATP synthesis coupled proton transport,GO:0015991~ATP hydrolysis coupled proton transport,GO:0031669~cellular response to nutrient levels,GO:0032119~sequestering of zinc ion,GO:0043181~vacuolar sequestering,GO:0070072~vacuolar proton-transporting V-type ATPase complex assembly,GO:0071472~cellular response to salt stress,	5
ACT domain repeat 6(ACR6)	BP = GO:0008152~metabolic process,	4
allene oxide synthase(AOS)	BP = GO:0006633~fatty acid biosynthetic process,GO:0006952~defense response,GO:0009611~response to wounding,GO:0009620~response to fungus,GO:0009695~jasmonic acid biosynthetic process,GO:0009753~response to jasmonic acid,GO:0016125~sterol metabolic process,GO:0019373~epoxygenase P450 pathway,GO:0031407~oxylipin metabolic process,GO:0031408~oxylipin biosynthetic process,GO:0050832~defense response to fungus,GO:0055114~oxidation-reduction process,	4
alpha/beta-Hydrolases superfamily protein(AT4G36530)	CC = GO:0005737~cytoplasm,GO:0009507~chloroplast,GO:0009941~chloroplast envelope,	4
Chaperone protein htpG family protein(CR88)	BP = GO:0006457~protein folding,GO:0006950~response to stress,GO:0009408~response to heat,GO:0009414~response to water deprivation,GO:0009651~response to salt stress,GO:0009704~de-etiolation,GO:0009793~embryo development ending in seed dormancy,GO:0010157~response to chlorate,GO:0045037~protein import into chloroplast stroma,	4
folate transporter 1(FOLT1)	BP = GO:0006412~translation,GO:0006810~transport,GO:0015884~folic acid transport,	4
GDSL-like Lipase/Acylhydrolase superfamily protein(AT1G29670)	BP = GO:0016042~lipid catabolic process,	4
glucose 6-phosphate/phosphate translocator 1(GPT1)	BP = GO:0007033~vacuole organization,GO:0008643~carbohydrate transport,GO:0009553~embryo sac development,GO:0009624~response to nematode,GO:0009793~embryo development ending in seed dormancy,GO:0010152~pollen maturation,GO:0015713~phosphoglycerate transport,GO:0015714~phosphoenolpyruvate transport,GO:0015760~glucose-6-phosphate transport,GO:0034389~lipid particle organization,GO:0035436~triose phosphate transmembrane transport,	4

Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase family protein(DMS11)	BP = GO:0006281~DNA repair,GO:0006282~regulation of DNA repair,GO:0006952~defense response,GO:0016569~covalent chromatin modification,GO:0031935~regulation of chromatin silencing,GO:0080188~RNA-directed DNA methylation,GO:1902290~positive regulation of defense response to oomycetes,	4
ILITYHIA(ILA)	BP = GO:0006417~regulation of translation,GO:0009682~induced systemic resistance,GO:0033554~cellular response to stress,GO:0033674~positive regulation of kinase activity,GO:0042742~defense response to bacterium,GO:0045087~innate immune response,	4
LMBR1-like membrane protein(AT5G65290)	CC = GO:0009507~chloroplast,GO:0016021~integral component of membrane,	4
Lysyl-tRNA synthetase, class II(OVA5)	BP = GO:0006418~tRNA aminoacylation for protein translation,GO:0006430~lysyl-tRNA aminoacylation,GO:0048481~plant ovule development,	4
Malectin/receptor-like protein kinase family protein(AT4G39110)	BP = GO:0006468~protein phosphorylation,	4
MOS4-associated complex 3B(MAC3B)	BP = GO:0000349~generation of catalytic spliceosome for first transesterification step,GO:0006281~DNA repair,GO:0006397~mRNA processing,GO:0008380~RNA splicing,GO:0016567~protein ubiquitination,GO:0042742~defense response to bacterium,GO:0045087~innate immune response,GO:0070534~protein K63-linked ubiquitination,	4
Myosin heavy chain-related protein(AT5G52280)	CC = GO:0009507~chloroplast,	4
NAD(P)-linked oxidoreductase superfamily protein(AT1G59960)	BP = GO:0009651~response to salt stress,GO:0055114~oxidation-reduction process,	4
NADH-dependent glutamate synthase 1(GLT1)	BP = GO:0006537~glutamate biosynthetic process,GO:0008652~cellular amino acid biosynthetic process,GO:0009735~response to cytokinin,GO:0019676~ammonia assimilation cycle,GO:0042128~nitrate assimilation,GO:0046686~response to cadmium ion,GO:0048589~developmental growth,GO:0055114~oxidation-reduction process,GO:0097054~L-glutamate biosynthetic process,	4
P-loop containing nucleoside triphosphate hydrolases superfamily protein(RIN1)	BP = GO:0000492~box C/D snoRNP assembly,GO:0006281~DNA repair,GO:0006338~chromatin remodeling,GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0006357~regulation of transcription from RNA polymerase II promoter,GO:0009908~flower development,GO:0016573~histone acetylation,GO:0030154~cell differentiation,GO:0048507~meristem development,GO:2000072~regulation of defense response to fungus, incompatible interaction,	4
Plant heme oxygenase (decyclizing) family protein(TED4)	BP = GO:0006788~heme oxidation,GO:0006979~response to oxidative stress,GO:0009585~red, far-red light phototransduction,GO:0009813~flavonoid biosynthetic process,GO:0010019~chloroplast-nucleus signaling pathway,GO:0010024~phytochromobilin biosynthetic process,GO:0010075~regulation of meristem growth,GO:0010119~regulation of stomatal movement,GO:0015979~photosynthesis,GO:0016117~carotenoid biosynthetic process,GO:0042167~heme catabolic process,GO:0055072~iron ion homeostasis,GO:0071494~cellular response to UV-C,	4
thylakoid ATP/ADP carrier(TAAC)	BP = GO:0006412~translation,GO:0006810~transport,GO:0010117~photoprotection,GO:0010206~photosystem II repair,GO:0055085~transmembrane transport,	4
alpha-glucan phosphorylase 2(PHS2)	BP = GO:0005975~carbohydrate metabolic process,GO:0005980~glycogen catabolic process,GO:0009414~response to water deprivation,GO:0046686~response to cadmium ion,	3
calcium dependent protein kinase 1(CPK1)	BP = GO:0006468~protein phosphorylation,GO:0009738~abscisic acid-activated signaling pathway,GO:0018105~peptidyl-serine phosphorylation,GO:0035556~intracellular signal transduction,GO:0046777~protein autophosphorylation,	3

chlorophyll A/B binding protein 1(CAB1)	BP = GO:0009416~response to light stimulus,GO:0009765~photosynthesis, light harvesting,GO:0009768~photosynthesis, light harvesting in photosystem I,GO:0015979~photosynthesis,GO:0018298~protein-chromophore linkage,	3
chloroplast heat shock protein 70-1(cpHsc70-1)	BP = GO:0006457~protein folding,GO:0009408~response to heat,GO:0009409~response to cold,GO:0016032~viral process,GO:0045036~protein targeting to chloroplast,GO:0046686~response to cadmium ion,	3
FKBP-like peptidyl-prolyl cis-trans isomerase family protein(AT5G13410)	BP = GO:0000413~protein peptidyl-prolyl isomerization,GO:0018208~peptidyl-proline modification,GO:0061077~chaperone-mediated protein folding,	3
GTP binding Elongation factor Tu family protein(AT5G60390)	BP = GO:0006414~translational elongation,GO:0046686~response to cadmium ion,	3
integral membrane hemolysin-III-like protein(AT2G45250)	BP = GO:0006351~transcription, DNA-templated,GO:0007049~cell cycle,	3
Major facilitator superfamily protein(AT2G32040)	BP = GO:0006810~transport,GO:0051958~methotrexate transport,	3
NAD(P)-binding Rossmann-fold superfamily protein(AT1G10310)	BP = GO:0006760~folic acid-containing compound metabolic process,GO:0055114~oxidation-reduction process,	3
NADP-malic enzyme 3(NADP-ME3)	BP = GO:0006090~pyruvate metabolic process,GO:0006108~malate metabolic process,GO:0051260~protein homooligomerization,GO:0055114~oxidation-reduction process,	3
Nucleotide-diphospho-sugar transferase family protein(AT1G14590)	CC = GO:0009507~chloroplast,GO:0009543~chloroplast thylakoid lumen,	3
PPR containing-like protein(AT1G04590)	CC = GO:0009507~chloroplast,	3
pyruvate kinase family protein(AT3G49160)	BP = GO:0006096~glycolytic process,	3
Ribosomal protein S7e family protein(AT3G02560)	BP = GO:0006364~rRNA processing,GO:0006412~translation,GO:0042274~ribosomal small subunit biogenesis,	3
SLH domain protein(AT3G25680)	CC = GO:0009507~chloroplast,GO:0009536~plastid,GO:0009941~chloroplast envelope,GO:0016021~integral component of membrane,	3
Tetratricopeptide repeat (TPR)-like superfamily protein(AT3G24000)	CC = GO:0005739~mitochondrion,	3
thylakoid lumenal protein (Mog1/PsbP/DUF1795-like photosystem II reaction center PsbP family protein)(AT1G77090)	BP = GO:0015979~photosynthesis,	3
TLC ATP/ADP transporter(AT5G05310)	CC = GO:0005886~plasma membrane,GO:0009507~chloroplast,GO:0016021~integral component of membrane,	3

UbiA prenyltransferase family protein(G4)	BP = GO:0015995~chlorophyll biosynthetic process,	3
UDP-galactose transporter 6(UTR6)	BP = GO:0008643~carbohydrate transport,GO:0015739~sialic acid transport,GO:0015780~nucleotide-sugar transport,	3
UDP-glucosyl transferase 85A5(UGT85A5)	BP = GO:0008152~metabolic process,GO:0009813~flavonoid biosynthetic process,GO:0052696~flavonoid glucuronidation,	3
calmodulin-binding protein(PSL4)	BP = GO:0006491~N-glycan processing,GO:0042742~defense response to bacterium,	2
Cupredoxin superfamily protein(DRT112)	BP = GO:0009411~response to UV,GO:0017148~negative regulation of translation,GO:0046688~response to copper ion,GO:0055070~copper ion homeostasis,GO:0055114~oxidation-reduction process,	2
cytochrome P450, family 77, subfamily A, polypeptide 5 pseudogene(CYP77A5P)	BP = GO:0008152~metabolic process,	2
D-3-phosphoglycerate dehydrogenase(PGDH)	BP = GO:0006564~L-serine biosynthetic process,GO:0055114~oxidation-reduction process,	2
dehydroascorbate reductase 1(DHAR3)	BP = GO:0006749~glutathione metabolic process,GO:0009735~response to cytokinin,GO:0010731~protein glutathionylation,	2
embryo defective 2737(emb2737)	BP = GO:0009793~embryo development ending in seed dormancy,	2
fructose-bisphosphate aldolase 1(FBA1)	BP = GO:0006096~glycolytic process,GO:0006098~pentose-phosphate shunt,GO:0046686~response to cadmium ion,	2
FTSH protease 11(FTSH11)	BP = GO:0006508~proteolysis,GO:0009408~response to heat,GO:0009644~response to high light intensity,GO:0010304~PSII associated light-harvesting complex II catabolic process,	2
global transcription factor group E8(GTE8)	BP = GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,	2
Glycosyl hydrolase family 38 protein(AT3G26720)	BP = GO:0006013~mannose metabolic process,GO:0006517~protein deglycosylation,	2
Glycosyl transferase, family 35(PHS1)	BP = GO:0005975~carbohydrate metabolic process,GO:0005980~glycogen catabolic process,GO:0007129~synapsis,GO:0009231~riboflavin biosynthetic process,GO:0009266~response to temperature stimulus,GO:0009414~response to water deprivation,GO:0009738~abscisic acid-activated signaling pathway,GO:0055114~oxidation-reduction process,GO:1901135~carbohydrate derivative metabolic process,	2
high cyclic electron flow 1(HCEF1)	BP = GO:0005985~sucrose metabolic process,GO:0005986~sucrose biosynthetic process,GO:0006000~fructose metabolic process,GO:0006001~fructose catabolic process,GO:0006002~fructose 6-phosphate metabolic process,GO:0006094~gluconeogenesis,GO:0009409~response to cold,GO:0009735~response to cytokinin,GO:0009773~photosynthetic electron transport in photosystem I,GO:0015979~photosynthesis,GO:0019253~reductive pentose-phosphate cycle,GO:0030388~fructose 1,6-bisphosphate metabolic process,	2

HR-like lesion-inducing protein-like protein(AT1G04340)	CC = GO:0009507~chloroplast,GO:0016021~integral component of membrane,	2
isopropylmalate dehydrogenase 2(IMD2)	BP = GO:0009082~branched-chain amino acid biosynthetic process,GO:0009098~leucine biosynthetic process,GO:0055114~oxidation-reduction process,	2
ketol-acid reductoisomerase(AT3G58610)	BP = GO:0008652~cellular amino acid biosynthetic process,GO:0009082~branched-chain amino acid biosynthetic process,GO:0009097~isoleucine biosynthetic process,GO:0009099~valine biosynthetic process,GO:0046686~response to cadmium ion,GO:0055114~oxidation-reduction process,	2
Kinesin motor family protein(AT4G39050)	BP = GO:0007018~microtubule-based movement,GO:0031347~regulation of defense response,	2
L-O-methylthreonine resistant 1(OMR1)	BP = GO:0006566~threonine metabolic process,GO:0008652~cellular amino acid biosynthetic process,GO:0009097~isoleucine biosynthetic process,	2
Leucine-rich repeat protein kinase family protein(AT3G47090)	BP = GO:0006468~protein phosphorylation,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,	2
Major facilitator superfamily protein(AT1G64890)	BP = GO:0006810~transport,	2
Mitochondrial substrate carrier family protein(AT5G19760)	BP = GO:0006412~translation,GO:0006810~transport,GO:0006835~dicarboxylic acid transport,GO:0006839~mitochondrial transport,GO:0035674~tricarboxylic acid transmembrane transport,GO:0055085~transmembrane transport,	2
NAD(P)-linked oxidoreductase superfamily protein(AT1G60690)	BP = GO:0055114~oxidation-reduction process,	2
NAD(P)H dehydrogenase (quinone)s(LTO1)	BP = GO:0010207~photosystem II assembly,GO:0055114~oxidation-reduction process,	2
P-type ATP-ase 1(PAA1)	BP = GO:0006511~ubiquitin-dependent protein catabolic process,GO:0009767~photosynthetic electron transport chain,GO:0030001~metal ion transport,GO:0035434~copper ion transmembrane transport,GO:0055070~copper ion homeostasis,	2
phosphoglycerate/bisphosphoglycerate mutase family protein(AT5G22620)	BP = GO:0008152~metabolic process,	2
photosystem II subunit P-1(PSBP-1)	BP = GO:0015979~photosynthesis,GO:0019684~photosynthesis, light reaction,GO:0042742~defense response to bacterium,	2
Rhodanese/Cell cycle control phosphatase superfamily protein(AT4G27700)	BP = GO:0007568~aging,	2
ribosomal protein L3 plastid(RPL3P)	BP = GO:0006412~translation,	2
RNA-binding (RRM/RBD/RNP motifs) family protein(AT1G67950)	CC = GO:0009507~chloroplast,	2

Sodium Bile acid symporter family(BASS2)	BP = GO:0006814~sodium ion transport,GO:0006849~plasma membrane pyruvate transport,GO:1901475~pyruvate transmembrane transport,	2
succinate dehydrogenase subunit 4(SDH4)	BP = GO:0006099~tricarboxylic acid cycle,	2
Tetratricopeptide repeat (TPR)-like superfamily protein(AT1G78915)	CC = GO:0009507~chloroplast,GO:0009535~chloroplast thylakoid membrane,GO:0016021~integral component of membrane,	2
thioredoxin M-type 4(TRX-M4)	BP = GO:0000103~sulfate assimilation,GO:0006109~regulation of carbohydrate metabolic process,GO:0006457~protein folding,GO:0006662~glycerol ether metabolic process,GO:0006979~response to oxidative stress,GO:0009735~response to cytokinin,GO:0034599~cellular response to oxidative stress,GO:0043085~positive regulation of catalytic activity,GO:0043086~negative regulation of catalytic activity,GO:0045454~cell redox homeostasis,GO:0055114~oxidation-reduction process,	2
trafficking protein particle complex subunit-like protein(AT2G47960)	CC = GO:0009507~chloroplast,	2
Transducin/WD40 repeat-like superfamily protein(AT3G06880)	BP = GO:0007165~signal transduction,	2
transmembrane protein(AT4G27390)	CC = GO:0009507~chloroplast,GO:0016021~integral component of membrane,	2
Transmembrane proteins 14C(AT3G57280)	BP = GO:0010208~pollen wall assembly,GO:0015908~fatty acid transport,GO:0055088~lipid homeostasis,GO:0071668~plant-type cell wall assembly,GO:1902001~fatty acid transmembrane transport,	2
valyl-tRNA synthetase / valine-tRNA ligase (VALRS)(TWN2)	BP = GO:0006418~tRNA aminoacylation for protein translation,GO:0006438~valyl-tRNA aminoacylation,GO:0009793~embryo development ending in seed dormancy,	2
AAA-type ATPase family protein(AT4G02480)	CC = GO:0005634~nucleus,GO:0009507~chloroplast,	1
AAA-type ATPase family protein(EMB3144)	BP = GO:0006508~proteolysis,GO:0009793~embryo development ending in seed dormancy,	1
ABC-2 and Plant PDR ABC-type transporter family protein(PEN3)	BP = GO:0006855~drug transmembrane transport,GO:0009627~systemic acquired resistance,GO:0009737~response to abscisic acid,GO:0009817~defense response to fungus, incompatible interaction,GO:0015691~cadmium ion transport,GO:0031348~negative regulation of defense response,GO:0042344~indole glucosinolate catabolic process,GO:0042742~defense response to bacterium,GO:0052544~defense response by callose deposition in cell wall,GO:0071366~cellular response to indolebutyric acid stimulus,	1
acetyl-CoA synthetase(ACS)	BP = GO:0006083~acetate metabolic process,GO:0006631~fatty acid metabolic process,GO:0019427~acetyl-CoA biosynthetic process from acetate,	1
alpha-N-acetylglucosaminidase family / NAGLU family(CYL1)	BP = GO:0008152~metabolic process,	1
alpha/beta-Hydrolases superfamily protein(AT5G17670)	CC = GO:0009507~chloroplast,GO:0009941~chloroplast envelope,	1

anthranilate phosphoribosyltransferase(AT1G70570)	BP = GO:0000162~tryptophan biosynthetic process,GO:0006177~GMP biosynthetic process,GO:0008152~metabolic process,	1
ARM repeat superfamily protein(AT3G06530)	BP = GO:0000462~maturation of SSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA),GO:0045943~positive regulation of transcription from RNA polymerase I promoter,	1
ARM repeat superfamily protein(TOR1)	BP = GO:0007275~multicellular organism development,GO:0009826~unidimensional cell growth,GO:0010031~circumnutation,	1
aspartate aminotransferase 3(ASP3)	BP = GO:0006103~2-oxoglutarate metabolic process,GO:0006520~cellular amino acid metabolic process,GO:0006531~aspartate metabolic process,GO:0006536~glutamate metabolic process,GO:0009058~biosynthetic process,GO:0010150~leaf senescence,	1
ATP citrate lyase (ACL) family protein(AT2G20420)	BP = GO:0006099~tricarboxylic acid cycle,GO:0046686~response to cadmium ion,	1
calcium-dependent protein kinase 6(CDPK6)	BP = GO:0006468~protein phosphorylation,GO:0009651~response to salt stress,GO:0009738~abscisic acid-activated signaling pathway,GO:0010119~regulation of stomatal movement,GO:0010359~regulation of anion channel activity,GO:0018105~peptidyl-serine phosphorylation,GO:0035556~intracellular signal transduction,GO:0046777~protein autophosphorylation,	1
Chaperone DnaJ-domain superfamily protein(ARC6)	BP = GO:0006457~protein folding,GO:0009658~chloroplast organization,GO:0010020~chloroplast fission,	1
chloride channel E(CLC-E)	BP = GO:0006821~chloride transport,	1
chloride channel F(CLC-F)	BP = GO:0006821~chloride transport,GO:1902476~chloride transmembrane transport,	1
chloroplast RNA binding protein(CRB)	BP = GO:0000272~polysaccharide catabolic process,GO:0005996~monosaccharide metabolic process,GO:0006364~rRNA processing,GO:0007623~circadian rhythm,GO:0009409~response to cold,GO:0009611~response to wounding,GO:0009658~chloroplast organization,GO:0010468~regulation of gene expression,GO:0032544~plastid translation,GO:0042631~cellular response to water deprivation,GO:0042742~defense response to bacterium,GO:0045727~positive regulation of translation,GO:0045893~positive regulation of transcription, DNA-templated,	1
Class II aaRS and biotin synthetases superfamily protein(OVA6)	BP = GO:0006418~tRNA aminoacylation for protein translation,GO:0006433~prolyl-tRNA aminoacylation,GO:0009553~embryo sac development,GO:0010109~regulation of photosynthesis,GO:0048316~seed development,GO:0048481~plant ovule development,	1
CRM family member 3A(CFM3A)	BP = GO:0000373~Group II intron splicing,GO:0006397~mRNA processing,GO:0048316~seed development,	1
CRT (chloroquine-resistance transporter)-like transporter 2(CLT2)	BP = GO:0002229~defense response to oomycetes,GO:0034635~glutathione transport,GO:0046686~response to cadmium ion,	1
DEA(D/H)-box RNA helicase family protein(AT2G01130)	BP = GO:0006396~RNA processing,	1
Dihydrodipicolinate reductase, bacterial/plant(AT3G59890)	BP = GO:0009089~lysine biosynthetic process via diaminopimelate,GO:0019877~diaminopimelate biosynthetic process,GO:0055114~oxidation-reduction process,	1

Dihydrolipoamide succinyltransferase(AT5G55070)	BP = GO:0006099~tricarboxylic acid cycle,GO:0006979~response to oxidative stress,GO:0033512~L-lysine catabolic process to acetyl-CoA via saccharopine,	1
DNAJ heat shock N-terminal domain-containing protein(AT2G42750)	BP = GO:0006457~protein folding,	1
DnaJ/Hsp40 cysteine-rich domain superfamily protein(LQY1)	BP = GO:0010206~photosystem II repair,	1
elongation defective 1 protein / ELD1 protein(KOB1)	BP = GO:0007049~cell cycle,GO:0009416~response to light stimulus,GO:0009663~plasmodesma organization,GO:0009737~response to abscisic acid,GO:0009738~abscisic acid-activated signaling pathway,GO:0009749~response to glucose,GO:0009826~unidimensional cell growth,GO:0009831~plant-type cell wall modification involved in multidimensional cell growth,GO:0010078~maintenance of root meristem identity,GO:0010215~cellulose microfibril organization,GO:0030154~cell differentiation,GO:0030244~cellulose biosynthetic process,GO:0040008~regulation of growth,GO:0051301~cell division,GO:0060560~developmental growth involved in morphogenesis,GO:0071482~cellular response to light stimulus,GO:2001009~regulation of plant-type cell wall cellulose biosynthetic process,	1
elongation factor Ts (DUF810)(AT2G25800)	CC = GO:0009507~chloroplast,	1
EMS-MUTAGENIZED BRI1 SUPPRESSOR 1(EBS1)	BP = GO:0006486~protein glycosylation,GO:0009626~plant-type hypersensitive response,GO:0009751~response to salicylic acid,GO:0010204~defense response signaling pathway, resistance gene-independent,GO:0016051~carbohydrate biosynthetic process,GO:0018279~protein N-linked glycosylation via asparagine,GO:0030968~endoplasmic reticulum unfolded protein response,GO:0046283~anthocyanin-containing compound metabolic process,GO:0051084~'de novo' posttranslational protein folding,GO:0071712~ER-associated misfolded protein catabolic process,	1
F-box/kelch-repeat protein(AT4G00893)	CC = GO:0009507~chloroplast,	1
FAD-binding Berberine family protein(AT1G30700)	BP = GO:0055114~oxidation-reduction process,	1
FAD-binding Berberine family protein(AT4G20840)	BP = GO:0055114~oxidation-reduction process,	1
FAD/NAD(P)-binding oxidoreductase family protein(AT3G04650)	BP = GO:0055114~oxidation-reduction process,	1
FMN-linked oxidoreductases superfamily protein(AT3G63510)	BP = GO:0006808~regulation of nitrogen utilization,GO:0055114~oxidation-reduction process,	1
GLU-ADT subunit B(GATB)	BP = GO:0006412~translation,GO:0006424~glutamyl-tRNA aminoacylation,GO:0032543~mitochondrial translation,GO:0070681~glutaminyl-tRNAGln biosynthesis via transamidation,	1
glutamate receptor 3.5(GLR3.5)	BP = GO:0006816~calcium ion transport,GO:0006874~cellular calcium ion homeostasis,GO:0009416~response to light stimulus,GO:0009845~seed germination,GO:0019722~calcium-mediated signaling,GO:0070584~mitochondrion morphogenesis,GO:0071230~cellular response to amino acid stimulus,	1
glyceraldehyde-3-phosphate dehydrogenase B subunit(GAPB)	BP = GO:0006006~glucose metabolic process,GO:0006096~glycolytic process,GO:0009409~response to cold,GO:0009416~response to light stimulus,GO:0009744~response to sucrose,GO:0019253~reductive pentose-phosphate cycle,GO:0046686~response to cadmium ion,GO:0055114~oxidation-reduction process,	1

Glycogen/starch synthases, ADP-glucose type(SS1)	BP = GO:0010021~amylopectin biosynthetic process,GO:0019252~starch biosynthetic process,	1
GRAM domain family protein(VAD1)	BP = GO:0006915~apoptotic process,GO:0009626~plant-type hypersensitive response,GO:0009723~response to ethylene,GO:0009751~response to salicylic acid,GO:0009873~ethylene-activated signaling pathway,GO:0042742~defense response to bacterium,GO:0043069~negative regulation of programmed cell death,	1
Guanylate-binding family protein(AT5G46070)	BP = GO:0006955~immune response,	1
Heat shock protein DnaJ with tetratricopeptide repeat-containing protein(TPR16)	BP = GO:0006457~protein folding,	1
Hydroxyproline-rich glycoprotein family protein(CHUP1)	BP = GO:0009902~chloroplast relocation,	1
Insulinase (Peptidase family M16) protein(AT1G51980)	BP = GO:0006508~proteolysis,GO:0006627~protein processing involved in protein targeting to mitochondrion,GO:0009651~response to salt stress,GO:0009735~response to cytokinin,GO:0055114~oxidation-reduction process,	1
Isocitrate/isopropylmalate dehydrogenase family protein(AT5G14590)	BP = GO:0006099~tricarboxylic acid cycle,GO:0006102~isocitrate metabolic process,	1
LETM1-like protein(AT5G06220)	BP = GO:0006875~cellular metal ion homeostasis,	1
Leucine-rich repeat protein kinase family protein(AT3G28450)	BP = GO:0006468~protein phosphorylation,GO:0006952~defense response,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,GO:1900150~regulation of defense response to fungus,GO:1900425~negative regulation of defense response to bacterium,	1
light-harvesting chlorophyll-protein complex I subunit A4(LHCA4)	BP = GO:0009409~response to cold,GO:0009644~response to high light intensity,GO:0009645~response to low light intensity stimulus,GO:0009765~photosynthesis, light harvesting,GO:0009768~photosynthesis, light harvesting in photosystem I,GO:0015979~photosynthesis,GO:0018298~protein-chromophore linkage,GO:0080167~response to karrikin,	1
Mitochondrial substrate carrier family protein(BOU)	BP = GO:0006412~translation,GO:0006810~transport,GO:0006839~mitochondrial transport,GO:0015822~ornithine transport,	1
NADH-ubiquinone dehydrogenase(EMB1467)	BP = GO:0006979~response to oxidative stress,GO:0009793~embryo development ending in seed dormancy,GO:0009853~photorespiration,GO:0042773~ATP synthesis coupled electron transport,GO:0045333~cellular respiration,	1
non-intrinsic ABC protein 11(ABCI13)	BP = GO:0006869~lipid transport,	1
OPC-8:0 CoA ligase1(OPCL1)	BP = GO:0008152~metabolic process,GO:0009611~response to wounding,GO:0009695~jasmonic acid biosynthetic process,GO:0009698~phenylpropanoid metabolic process,	1
Peptidase M50 family protein(EGY1)	BP = GO:0006508~proteolysis,GO:0007275~multicellular organism development,GO:0009416~response to light stimulus,GO:0009630~gravitropism,GO:0009658~chloroplast organization,GO:0009723~response to ethylene,GO:0009959~negative gravitropism,GO:0010027~thylakoid membrane organization,GO:0010207~photosystem II assembly,GO:0043157~response to cation stress,GO:0048564~photosystem I assembly,GO:0060359~response to ammonium ion,	1

phosphatidylglycerolphosphate synthase 1(PGP1)	BP = GO:0006655~phosphatidylglycerol biosynthetic process,GO:0006661~phosphatidylinositol biosynthetic process,GO:0008654~phospholipid biosynthetic process,GO:0010027~thylakoid membrane organization,	1
phosphoglucan, water dikinase(PWD)	BP = GO:0005975~carbohydrate metabolic process,GO:0016310~phosphorylation,	1
photosystem I subunit l(PSAL)	BP = GO:0015979~photosynthesis,GO:0019684~photosynthesis, light reaction,	1
Polynucleotidyl transferase, ribonuclease H-like superfamily protein(AT3G25430)	BP = GO:0006397~mRNA processing,GO:0009451~RNA modification,	1
protein EMBRYO DEFECTIVE 2247(EMB2247)	BP = GO:0006418~tRNA aminoacylation for protein translation,GO:0006438~valyl-tRNA aminoacylation,GO:0009793~embryo development ending in seed dormancy,	1
pyrimidine d(PYRD)	BP = GO:0006207~'de novo' pyrimidine nucleobase biosynthetic process,GO:0009220~pyrimidine ribonucleotide biosynthetic process,GO:0044205~'de novo' UMP biosynthetic process,GO:0055114~oxidation-reduction process,	1
Rieske (2Fe-2S) domain-containing protein(AT1G71500)	BP = GO:0010207~photosystem II assembly,GO:0042128~nitrate assimilation,GO:0042549~photosystem II stabilization,GO:0055114~oxidation-reduction process,	1
RNA-metabolising metallo-beta-lactamase family protein(emb2746)	BP = GO:0009658~chloroplast organization,GO:0009793~embryo development ending in seed dormancy,GO:0009942~longitudinal axis specification,GO:0060918~auxin transport,	1
Rubisco methyltransferase family protein(AT5G14260)	BP = GO:0018026~peptidyl-lysine monomethylation,	1
semialdehyde dehydrogenase family protein(AT1G14810)	BP = GO:0006520~cellular amino acid metabolic process,GO:0009086~methionine biosynthetic process,GO:0009088~threonine biosynthetic process,GO:0009089~lysine biosynthetic process via diaminopimelate,GO:0009097~isoleucine biosynthetic process,GO:0055114~oxidation-reduction process,	1
Serine/Threonine-kinase(PRIN2)	BP = GO:0009642~response to light intensity,GO:0010468~regulation of gene expression,	1
shikimate kinase 1(SK1)	BP = GO:0008652~cellular amino acid biosynthetic process,GO:0009073~aromatic amino acid family biosynthetic process,GO:0009423~chorismate biosynthetic process,GO:0016310~phosphorylation,GO:0019632~shikimate metabolic process,	1
Thioredoxin superfamily protein(AT1G21350)	BP = GO:0045454~cell redox homeostasis,	1
Thioredoxin superfamily protein(AT2G37240)	BP = GO:0055114~oxidation-reduction process,	1
Transducin/WD40 repeat-like superfamily protein(MSI1)	BP = GO:0006260~DNA replication,GO:0006349~regulation of gene expression by genetic imprinting,GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0007275~multicellular organism development,GO:0008283~cell proliferation,GO:0009555~pollen development,GO:0009793~embryo development ending in seed dormancy,GO:0009908~flower development,GO:0009909~regulation of flower development,GO:0010026~trichome differentiation,GO:0010214~seed coat development,GO:0016569~covalent chromatin modification,GO:0031507~heterochromatin assembly,GO:0045787~positive regulation of cell cycle,GO:0048316~seed development,GO:0048366~leaf development,GO:0070828~heterochromatin organization,GO:2000653~regulation of genetic imprinting,	1

transmembrane/coiled-coil protein (DUF726)(AT4G36210)	CC = GO:0009507~chloroplast,GO:0016021~integral component of membrane,	1
Trypsin family protein(AT5G45030)	CC = GO:0009507~chloroplast,	1
UDP-Glycosyltransferase superfamily protein(AT4G14090)	BP = GO:0008152~metabolic process,GO:0009698~phenylpropanoid metabolic process,GO:0009718~anthocyanin-containing compound biosynthetic process,GO:0009813~flavonoid biosynthetic process,GO:0052696~flavonoid glucuronidation,	1

Table S17. Nucleotide-binding/ATP-binding outlier genes

Functional annotation of the outlier genes within the enriched biological function of 'nucleotide-binding/ATP-binding'. Gene ontology is the *DAVID GOTERM_BP_DIRECT* (denoted below with *BP*), and where absent, the *GOTERM_CC_DIRECT* was used (*CC*). *Number of replicate* pairs denotes how many replicate pairs the gene was detected as an outlier.

Gene	Gene ontology	Number replicate pairs
BRI1-associated receptor kinase(BAK1)	BP = GO:0002229~defense response to oomycetes,GO:0006468~protein phosphorylation,GO:0008219~cell death,GO:0009742~brassinosteroid mediated signaling pathway,GO:0016049~cell growth,GO:0042742~defense response to bacterium,GO:0050832~defense response to fungus,	8
Leucine-rich repeat transmembrane protein kinase(GSO1)	BP = GO:0006468~protein phosphorylation,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,GO:0009793~embryo development ending in seed dormancy,GO:0090558~plant epidermis development,	8
Malectin/receptor-like protein kinase family protein(AT4G39110)	BP = GO:0006468~protein phosphorylation,	8
UBQ, helicase-c and DEAD-like helicase domain-containing protein(AT5G08110)	BP = GO:0010501~RNA secondary structure unwinding,	8
adenylate kinase 1(ADK1)	BP = GO:0046686~response to cadmium ion,	7
casein kinase 1-like protein 2(CKL2)	BP = GO:0006897~endocytosis,GO:0008360~regulation of cell shape,GO:0016055~Wnt signaling pathway,GO:0018105~peptidyl-serine phosphorylation,	7
RNA helicase, ATP-dependent, SK12/DOB1 protein(MTR4)	BP = GO:0000460~maturation of 5.8S rRNA,GO:0006364~rRNA processing,GO:0006401~RNA catabolic process,GO:0031125~rRNA 3'-end processing,	7
ATP-dependent Clp protease(AT1G33360)	BP = GO:0006457~protein folding,GO:0006508~proteolysis,GO:0030163~protein catabolic process,	6
FtsH extracellular protease family(FtsHi1)	BP = GO:0006508~proteolysis,GO:0009658~chloroplast organization,GO:0009793~embryo development ending in seed dormancy,GO:0010020~chloroplast fission,GO:0043572~plastid fission,	6
GTP binding Elongation factor Tu family protein(AT5G60390)	BP = GO:0006414~translational elongation,GO:0046686~response to cadmium ion,	6
Protein kinase superfamily protein(AT5G56460)	BP = GO:0006468~protein phosphorylation,	6

RECQ helicase SIM(RECQSIM)	BP = GO:0000724~double-strand break repair via homologous recombination,GO:0006281~DNA repair,GO:0032508~DNA duplex unwinding,	6
ROOT GROWTH DEFECTIVE 3(RGD3)	BP = GO:1902185~positive regulation of shoot apical meristem development,	6
Tubulin/FtsZ family protein(FTSZ2-1)	BP = GO:0010020~chloroplast fission,	6
ATP binding cassette subfamily B19(ABCB19)	BP = GO:0008361~regulation of cell size,GO:0009637~response to blue light,GO:0009639~response to red or far red light,GO:0009640~photomorphogenesis,GO:0009733~response to auxin,GO:0009734~auxin-activated signaling pathway,GO:0009926~auxin polar transport,GO:0009958~positive gravitropism,GO:0010218~response to far red light,GO:0010540~basipetal auxin transport,GO:0010541~acropetal auxin transport,GO:0043481~anthocyanin accumulation in tissues in response to UV light,GO:0048364~root development,GO:0048443~stamen development,GO:0048527~lateral root development,GO:0055085~transmembrane transport,GO:0060918~auxin transport,	5
ATP-binding casette family G25(ABCG25)	BP = GO:0009737~response to abscisic acid,GO:0080168~abscisic acid transport,	5
DNA ligase IV(ATLIG4)	BP = GO:0000723~telomere maintenance,GO:0006266~DNA ligation,GO:0006273~lagging strand elongation,GO:0006302~double-strand break repair,GO:0006303~double-strand break repair via nonhomologous end joining,GO:0006310~DNA recombination,GO:0006974~cellular response to DNA damage stimulus,GO:0010165~response to X-ray,GO:0015074~DNA integration,GO:0051103~DNA ligation involved in DNA repair,GO:0071897~DNA biosynthetic process,	5
Dynamin related protein 5A(DRP5A)	BP = GO:0000911~cytokinesis by cell plate formation,GO:0048285~organelle fission,	5
Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase family protein(DMS11)	BP = GO:0006281~DNA repair,GO:0006282~regulation of DNA repair,GO:0006952~defense response,GO:0016569~covalent chromatin modification,GO:0031935~regulation of chromatin silencing,GO:0080188~RNA-directed DNA methylation,GO:1902290~positive regulation of defense response to oomycetes,	5
Leucine-rich repeat receptor-like protein kinase family protein(AT3G24240)	BP = GO:0006468~protein phosphorylation,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,GO:0010082~regulation of root meristem growth,GO:0010449~root meristem growth,GO:0016032~viral process,	5
lon protease 4(LON4)	BP = GO:0001666~response to hypoxia,GO:0006515~misfolded or incompletely synthesized protein catabolic process,GO:0007005~mitochondrion organization,GO:0034599~cellular response to oxidative stress,GO:0051131~chaperone-mediated protein complex assembly,GO:0051260~protein homooligomerization,GO:0070407~oxidation-dependent protein catabolic process,GO:0090296~regulation of mitochondrial DNA replication,	5
Lysyl-tRNA synthetase, class II(OVA5)	BP = GO:0006418~tRNA aminoacylation for protein translation,GO:0006430~lysyl-tRNA aminoacylation,GO:0048481~plant ovule development,	5
MIRO-related GTP-ase 1(MIRO1)	BP = GO:0007005~mitochondrion organization,GO:0007264~small GTPase mediated signal transduction,GO:0009793~embryo development ending in seed dormancy,GO:0009860~pollen tube growth,	5
P-glycoprotein 20(ABCB20)	BP = GO:0010315~auxin efflux,GO:0010540~basipetal auxin transport,GO:0010541~acropetal auxin transport,GO:0055085~transmembrane transport,	5
P-loop containing nucleoside triphosphate hydrolases superfamily protein(RIN1)	BP = GO:0000492~box C/D snoRNP assembly,GO:0006281~DNA repair,GO:0006338~chromatin remodeling,GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0006357~regulation of transcription from RNA polymerase II promoter,GO:0009908~flower	5

	development,GO:0016573~histone acetylation,GO:0030154~cell differentiation,GO:0048507~meristem development,GO:2000072~regulation of defense response to fungus, incompatible interaction,	
Protein kinase superfamily protein(AT1G09600)	BP = GO:0006468~protein phosphorylation,	5
Protein kinase superfamily protein(AT1G26970)	BP = GO:0006468~protein phosphorylation,	5
Protein kinase superfamily protein(AT5G55560)	BP = GO:0006468~protein phosphorylation,GO:0035556~intracellular signal transduction,	5
SNF1 kinase homolog 10(KIN10)	BP = GO:0003006~developmental process involved in reproduction,GO:0005975~carbohydrate metabolic process,GO:0006468~protein phosphorylation,GO:0006633~fatty acid biosynthetic process,GO:0009594~detection of nutrient,GO:0009738~abscisic acid-activated signaling pathway,GO:0010050~vegetative phase change,GO:0010150~leaf senescence,GO:0010182~sugar mediated signaling pathway,GO:0035556~intracellular signal transduction,GO:0042128~nitrate assimilation,GO:0080022~primary root development,	5
SOS3-interacting protein 1(SIP1)	BP = GO:0006468~protein phosphorylation,GO:0007165~signal transduction,	5
calcium dependent protein kinase 1(CPK1)	BP = GO:0006468~protein phosphorylation,GO:0009738~abscisic acid-activated signaling pathway,GO:0018105~peptidyl-serine phosphorylation,GO:0035556~intracellular signal transduction,GO:0046777~protein autophosphorylation,	4
Chaperone protein htpG family protein(CR88)	BP = GO:0006457~protein folding,GO:0006950~response to stress,GO:0009408~response to heat,GO:0009414~response to water deprivation,GO:0009651~response to salt stress,GO:0009704~de-etiolation,GO:0009793~embryo development ending in seed dormancy,GO:0010157~response to chlorate,GO:0045037~protein import into chloroplast stroma,	4
diacylglycerol kinase1(DGK1)	BP = GO:0006952~defense response,GO:0007205~protein kinase C-activating G-protein coupled receptor signaling pathway,GO:0016310~phosphorylation,GO:0035556~intracellular signal transduction,	4
glutamine-tRNA ligase, putative / glutaminyl-tRNA synthetase, putative / GlnRS(OVA9)	BP = GO:0006412~translation,GO:0006424~glutamyl-tRNA aminoacylation,GO:0006425~glutaminyl-tRNA aminoacylation,GO:0048481~plant ovule development,	4
heavy metal atpase 5(HMA5)	BP = GO:0010273~detoxification of copper ion,GO:0030001~metal ion transport,GO:0046688~response to copper ion,	4
HSI2-like 1(HSL1)	BP = GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0010030~positive regulation of seed germination,GO:2000034~regulation of seed maturation,	4
Ku80 family protein(KU80)	BP = GO:0000723~telomere maintenance,GO:0006281~DNA repair,GO:0006302~double-strand break repair,GO:0006303~double-strand break repair via nonhomologous end joining,GO:0006310~DNA recombination,GO:0009408~response to heat,GO:0015074~DNA integration,GO:0032508~DNA duplex unwinding,GO:0071480~cellular response to gamma radiation,GO:0071481~cellular response to X-ray,	4
Leucine-rich repeat protein kinase family protein(AT3G47090)	BP = GO:0006468~protein phosphorylation,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,	4
Leucine-rich repeat protein kinase family protein(PRK3)	BP = GO:0006468~protein phosphorylation,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,GO:0080092~regulation of pollen tube growth,	4

Leucine-rich repeat transmembrane protein kinase(AT1G07650)	BP = GO:0006468~protein phosphorylation,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,GO:0016310~phosphorylation,	4
P-glycoprotein 21(ABCB21)	BP = GO:0010315~auxin efflux,GO:0010540~basipetal auxin transport,GO:0055085~transmembrane transport,GO:0060919~auxin influx,	4
STRUBBELIG-receptor family 3(SRF3)	BP = GO:0006468~protein phosphorylation,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,GO:0007623~circadian rhythm,	4
TLC ATP/ADP transporter(AT5G05310)	CC = GO:0005886~plasma membrane,GO:0009507~chloroplast,GO:0016021~integral component of membrane,	4
ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein(AT3G63380)	BP = GO:0070588~calcium ion transmembrane transport,	3
chloroplast heat shock protein 70-1(cpHsc70-1)	BP = GO:0006457~protein folding,GO:0009408~response to heat,GO:0009409~response to cold,GO:0016032~viral process,GO:0045036~protein targeting to chloroplast,GO:0046686~response to cadmium ion,	3
cyclic nucleotide gated channel 1(CNGC1)	BP = GO:0006813~potassium ion transport,GO:0006816~calcium ion transport,GO:0042391~regulation of membrane potential,	3
DEAD/DEAH box RNA helicase family protein(UAP56b)	BP = GO:0000398~mRNA splicing, via spliceosome,GO:0006406~mRNA export from nucleus,GO:0006974~cellular response to DNA damage stimulus,GO:0010468~regulation of gene expression,GO:0010501~RNA secondary structure unwinding,GO:0046686~response to cadmium ion,	3
dicer-like 4(DCL4)	BP = GO:0006353~DNA-templated transcription, termination,GO:0006396~RNA processing,GO:0009616~virus induced gene silencing,GO:0009944~polarity specification of adaxial/abaxial axis,GO:0010050~vegetative phase change,GO:0010216~maintenance of DNA methylation,GO:0010267~production of ta-siRNAs involved in RNA interference,GO:0010492~maintenance of shoot apical meristem identity,GO:0010599~production of lsiRNA involved in RNA interference,GO:0035196~production of miRNAs involved in gene silencing by miRNA,GO:0048608~reproductive structure development,GO:0051214~RNA virus induced gene silencing,GO:0051607~defense response to virus,	3
FAD-binding Berberine family protein(AT1G30700)	BP = GO:0055114~oxidation-reduction process,	3
Guanylate-binding family protein(AT5G46070)	BP = GO:0006955~immune response,	3
Leucine-rich receptor-like protein kinase family protein(AT5G56040)	BP = GO:0006468~protein phosphorylation,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,GO:0010082~regulation of root meristem growth,	3
multidrug resistance-associated protein 5(ABCC5)	BP = GO:0009651~response to salt stress,GO:0030007~cellular potassium ion homeostasis,GO:0055085~transmembrane transport,GO:1901527~abscisic acid-activated signaling pathway involved in stomatal movement,GO:1990069~stomatal opening,	3
multidrug resistance-associated protein 6(ABCC8)	CC = GO:0000325~plant-type vacuole,GO:0005774~vacuolar membrane,GO:0005886~plasma membrane,GO:0016021~integral component of membrane,	3
NIMA-related serine/threonine kinase 1(NEK1)	BP = GO:0006468~protein phosphorylation,	3

non-intrinsic ABC protein 12(NAP12)	CC = GO:0005576~extracellular region,GO:0005886~plasma membrane,GO:0009506~plasmodesma,GO:0016021~integral component of membrane,	3
P-loop containing nucleoside triphosphate hydrolases superfamily protein(SDE3)	BP = GO:0009616~virus induced gene silencing,GO:0035194~posttranscriptional gene silencing by RNA,	3
proline-rich extensin-like receptor kinase 10(PERK10)	BP = GO:0006468~protein phosphorylation,	3
pyruvate kinase family protein(AT3G49160)	BP = GO:0006096~glycolytic process,	3
SNF2 domain-containing protein / helicase domain-containing protein / HNH endonuclease domain-containing protein(AT5G07810)	CC = GO:0005634~nucleus,	3
STRUBBELIG-receptor family 7(SRF7)	BP = GO:0006468~protein phosphorylation,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,	3
TCP-1/cpn60 chaperonin family protein(AT5G16070)	BP = GO:0006457~protein folding,	3
ABC-2 and Plant PDR ABC-type transporter family protein(PEN3)	BP = GO:0006855~drug transmembrane transport,GO:0009627~systemic acquired resistance,GO:0009737~response to abscisic acid,GO:0009817~defense response to fungus, incompatible interaction,GO:0015691~cadmium ion transport,GO:0031348~negative regulation of defense response,GO:0042344~indole glucosinolate catabolic process,GO:0042742~defense response to bacterium,GO:0052544~defense response by callose deposition in cell wall,GO:0071366~cellular response to indolebutyric acid stimulus,	2
ABC-2 type transporter family protein(ABCG22)	BP = GO:0009414~response to water deprivation,GO:0010148~transpiration,	2
ARF GAP-like zinc finger-containing protein ZIGA4(ZIGA4)	BP = GO:0046777~protein autophosphorylation,	2
ATP-dependent DNA helicase 2 subunit Ku70-like protein(KU70)	BP = GO:0000723~telomere maintenance,GO:0006281~DNA repair,GO:0006303~double-strand break repair via nonhomologous end joining,GO:0006310~DNA recombination,GO:0009408~response to heat,GO:0032508~DNA duplex unwinding,GO:0071480~cellular response to gamma radiation,GO:0071481~cellular response to X-ray,	2
calcium-dependent protein kinase 6(CDPK6)	BP = GO:0006468~protein phosphorylation,GO:0009651~response to salt stress,GO:0009738~abscisic acid-activated signaling pathway,GO:0010119~regulation of stomatal movement,GO:0010359~regulation of anion channel activity,GO:0018105~peptidyl-serine phosphorylation,GO:0035556~intracellular signal transduction,GO:0046777~protein autophosphorylation,	2
F-box/RNI-like superfamily protein(AT1G13570)	CC = GO:0005634~nucleus,	2
FTSH protease 11(FTSH11)	BP = GO:0006508~proteolysis,GO:0009408~response to heat,GO:0009644~response to high light intensity,GO:0010304~PSII associated light-harvesting complex II catabolic process,	2
growth-regulating factor 1(GRF1)	BP = GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0009624~response to nematode,GO:0019760~glucosinolate metabolic process,GO:0048364~root development,GO:0048366~leaf development,GO:0061062~regulation of nematode larval development,	2

Inositol-pentakisphosphate 2-kinase family protein(AT1G22100)	BP = GO:0032958~inositol phosphate biosynthetic process,GO:0052746~inositol phosphorylation,	2
Kinesin motor family protein(AT4G39050)	BP = GO:0007018~microtubule-based movement,GO:0031347~regulation of defense response,	2
Leucine-rich repeat receptor-like protein kinase family protein(AT4G26540)	BP = GO:0006468~protein phosphorylation,GO:0010082~regulation of root meristem growth,	2
MAP kinase 19(MPK19)	BP = GO:0007165~signal transduction,	2
Minichromosome maintenance (MCM2/3/5) family protein(MCM3)	BP = GO:0006268~DNA unwinding involved in DNA replication,GO:0006270~DNA replication initiation,GO:0007049~cell cycle,	2
NIMA-related serine/threonine kinase 1(NEK1)	BP = GO:0006468~protein phosphorylation,	2
nuclear poly(a) polymerase(nPAP)	BP = GO:0006378~mRNA polyadenylation,GO:0006397~mRNA processing,GO:0031123~RNA 3'-end processing,GO:0043631~RNA polyadenylation,	2
P-type ATP-ase 1(PAA1)	BP = GO:0006511~ubiquitin-dependent protein catabolic process,GO:0009767~photosynthetic electron transport chain,GO:0030001~metal ion transport,GO:0035434~copper ion transmembrane transport,GO:0055070~copper ion homeostasis,	2
pleiotropic drug resistance 12(ABCG40)	BP = GO:0002229~defense response to oomycetes,GO:0006855~drug transmembrane transport,GO:0009723~response to ethylene,GO:0009751~response to salicylic acid,GO:0009753~response to jasmonic acid,GO:0010193~response to ozone,GO:0015692~lead ion transport,GO:0046865~terpenoid transport,GO:0080168~abscisic acid transport,	2
protein kinase family protein / protein phosphatase 2C (PP2C) family protein(AT2G40860)	BP = GO:0006468~protein phosphorylation,	2
Protein kinase superfamily protein(AT1G18160)	BP = GO:0006468~protein phosphorylation,	2
Protein kinase superfamily protein(CRLK1)	BP = GO:0006468~protein phosphorylation,GO:0009409~response to cold,GO:0009631~cold acclimation,	2
receptor-like protein kinase 2(RPK2)	BP = GO:0006468~protein phosphorylation,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,GO:0009409~response to cold,GO:0009414~response to water deprivation,GO:0009808~lignin metabolic process,GO:0009846~pollen germination,GO:0009942~longitudinal axis specification,GO:0009945~radial axis specification,GO:0010073~meristem maintenance,GO:0010152~pollen maturation,GO:0048508~embryonic meristem development,GO:0048653~anther development,GO:0051260~protein homooligomerization,	2
RHO-related protein from plants 9(ROP9)	BP = GO:0007264~small GTPase mediated signal transduction,GO:0009738~abscisic acid-activated signaling pathway,	2
shikimate kinase 1(SK1)	BP = GO:0008652~cellular amino acid biosynthetic process,GO:0009073~aromatic amino acid family biosynthetic process,GO:0009423~chorismate biosynthetic process,GO:0016310~phosphorylation,GO:0019632~shikimate metabolic process,	2

sucrose transporter 4(SUT4)	BP = GO:0005985~sucrose metabolic process,GO:0015770~sucrose transport,	2
sulfate transporter 1;3(SULTR1;3)	CC = GO:0005887~integral component of plasma membrane,	2
ubiquitin-activating enzyme 1(UBA1)	BP = GO:0006464~cellular protein modification process,GO:0006511~ubiquitin-dependent protein catabolic process,GO:0016567~protein ubiquitination,GO:0046686~response to cadmium ion,GO:0051707~response to other organism,	2
ubiquitin-conjugating enzyme 25(UBC25)	BP = GO:0016567~protein ubiquitination,	2
valyl-tRNA synthetase / valine-tRNA ligase (VALRS)(TWN2)	BP = GO:0006418~tRNA aminoacylation for protein translation,GO:0006438~valyl-tRNA aminoacylation,GO:0009793~embryo development ending in seed dormancy,	2
with no lysine (K) kinase 8(WNK8)	BP = GO:0006468~protein phosphorylation,GO:0009909~regulation of flower development,GO:0009911~positive regulation of flower development,GO:0010228~vegetative to reproductive phase transition of meristem,GO:0035556~intracellular signal transduction,GO:0046777~protein autophosphorylation,GO:0048573~photoperiodism, flowering,	2
3-ketoacyl-CoA synthase 5(KCS5)	BP = GO:0000038~very long-chain fatty acid metabolic process,GO:0006633~fatty acid biosynthetic process,GO:0009409~response to cold,GO:0009416~response to light stimulus,GO:0042335~cuticle development,GO:0048868~pollen tube development,	1
AAA-type ATPase family protein(AT4G02480)	CC = GO:0005634~nucleus,GO:0009507~chloroplast,	1
AAA-type ATPase family protein(EMB3144)	BP = GO:0006508~proteolysis,GO:0009793~embryo development ending in seed dormancy,	1
acetoacetyl-CoA thiolase 2(ACAT2)	BP = GO:0006635~fatty acid beta-oxidation,GO:0008152~metabolic process,GO:0008299~isoprenoid biosynthetic process,GO:0009793~embryo development ending in seed dormancy,GO:0009846~pollen germination,GO:0009860~pollen tube growth,GO:0016125~sterol metabolic process,	1
acetyl-CoA synthetase(ACS)	BP = GO:0006083~acetate metabolic process,GO:0006631~fatty acid metabolic process,GO:0019427~acetyl-CoA biosynthetic process from acetate,	1
acyl-CoA binding protein 2(ACBP2)	BP = GO:0001666~response to hypoxia,GO:0006869~lipid transport,GO:0010288~response to lead ion,	1
Alg9-like mannosyltransferase family(EBS3)	BP = GO:0006486~protein glycosylation,GO:0006490~oligosaccharide-lipid intermediate biosynthetic process,GO:0016310~phosphorylation,GO:0030433~ER-associated ubiquitin-dependent protein catabolic process,	1
alpha-glucan phosphorylase 2(PHS2)	BP = GO:0005975~carbohydrate metabolic process,GO:0005980~glycogen catabolic process,GO:0009414~response to water deprivation,GO:0046686~response to cadmium ion,	1
arabinose kinase(ARA1)	BP = GO:0019566~arabinose metabolic process,	1
aspartate aminotransferase 3(ASP3)	BP = GO:0006103~2-oxoglutarate metabolic process,GO:0006520~cellular amino acid metabolic process,GO:0006531~aspartate metabolic process,GO:0006536~glutamate metabolic process,GO:0009058~biosynthetic process,GO:0010150~leaf senescence,	1

ATP citrate lyase (ACL) family protein(AT2G20420)	BP = GO:0006099~tricarboxylic acid cycle,GO:0046686~response to cadmium ion,	1
cation calcium exchanger 4(CCX4)	BP = GO:0006812~cation transport,GO:0006813~potassium ion transport,GO:0006814~sodium ion transport,GO:0055085~transmembrane transport,	1
Cation efflux family protein(AT1G51610)	BP = GO:0006812~cation transport,	1
cellulose synthase family protein(IRX1)	BP = GO:0006970~response to osmotic stress,GO:0009414~response to water deprivation,GO:0009834~plant-type secondary cell wall biogenesis,GO:0009863~salicylic acid mediated signaling pathway,GO:0009867~jasmonic acid mediated signaling pathway,GO:0009873~ethylene-activated signaling pathway,GO:0010116~positive regulation of abscisic acid biosynthetic process,GO:0030244~cellulose biosynthetic process,GO:0042742~defense response to bacterium,GO:0050832~defense response to fungus,GO:0052386~cell wall thickening,GO:0071555~cell wall organization,	1
chloride channel E(CLC-E)	BP = GO:0006821~chloride transport,	1
chlorophyll A/B binding protein 1(CAB1)	BP = GO:0009416~response to light stimulus,GO:0009765~photosynthesis, light harvesting,GO:0009768~photosynthesis, light harvesting in photosystem I,GO:0015979~photosynthesis,GO:0018298~protein-chromophore linkage,	1
chromomethylase 2(CMT2)	BP = GO:0006306~DNA methylation,GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0010426~DNA methylation on cytosine within a CHH sequence,GO:0016569~covalent chromatin modification,GO:0034605~cellular response to heat,	1
Class II aaRS and biotin synthetases superfamily protein(OVA6)	BP = GO:0006418~tRNA aminoacylation for protein translation,GO:0006433~prolyl-tRNA aminoacylation,GO:0009553~embryo sac development,GO:0010109~regulation of photosynthesis,GO:0048316~seed development,GO:0048481~plant ovule development,	1
Class II aminoacyl-tRNA and biotin synthetases superfamily protein(AT4G31180)	BP = GO:0006422~aspartyl-tRNA aminoacylation,GO:0046686~response to cadmium ion,GO:0050832~defense response to fungus,	1
Core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein(AT4G31350)	CC = GO:0005739~mitochondrion,GO:0016021~integral component of membrane,	1
cysteine-rich RLK (RECEPTOR-like protein kinase) 2(CRK2)	BP = GO:0006952~defense response,GO:0010193~response to ozone,GO:0046777~protein autophosphorylation,	1
DEA(D/H)-box RNA helicase family protein(AT2G01130)	BP = GO:0006396~RNA processing,	1
DNA repair/recombination protein(RAD54)	BP = GO:0000724~double-strand break repair via homologous recombination,GO:0006281~DNA repair,GO:0007067~mitotic nuclear division,GO:0010332~response to gamma radiation,GO:0016032~viral process,GO:0045003~double-strand break repair via synthesis-dependent strand annealing,GO:0051301~cell division,GO:0051701~interaction with host,GO:0071248~cellular response to metal ion,GO:0071480~cellular response to gamma radiation,	1
endoplasmatic reticulum retrieval protein 1B(RER1B)	BP = GO:0006890~retrograde vesicle-mediated transport, Golgi to ER,	1
FAD-binding Berberine family protein(AT4G20840)	BP = GO:0055114~oxidation-reduction process,	1

flavodoxin family protein / radical SAM domain-containing protein(AT1G75200)	BP = GO:0008033~tRNA processing,GO:0055114~oxidation-reduction process,	1
fructose-bisphosphate aldolase 1(FBA1)	BP = GO:0006096~glycolytic process,GO:0006098~pentose-phosphate shunt,GO:0046686~response to cadmium ion,	1
gamma-tubulin(TUBG1)	BP = GO:0000911~cytokinesis by cell plate formation,GO:0000914~phragmoplast assembly,GO:0007020~microtubule nucleation,GO:0009553~embryo sac development,GO:0009624~response to nematode,GO:0010103~stomatal complex morphogenesis,GO:0031122~cytoplasmic microtubule organization,GO:0048366~leaf development,GO:0048768~root hair cell tip growth,GO:0051641~cellular localization,	1
GLU-ADT subunit B(GATB)	BP = GO:0006412~translation,GO:0006424~glutamyl-tRNA aminoacylation,GO:0032543~mitochondrial translation,GO:0070681~glutaminyl-tRNAGln biosynthesis via transamidation,	1
glucan synthase-like 12(GSL12)	BP = GO:0006075~(1->3)-beta-D-glucan biosynthetic process,GO:0008360~regulation of cell shape,GO:0071555~cell wall organization,	1
Glycogen/starch synthases, ADP-glucose type(SS1)	BP = GO:0010021~amylopectin biosynthetic process,GO:0019252~starch biosynthetic process,	1
Glycosyl transferase family 4 protein(AT3G57220)	BP = GO:0000271~polysaccharide biosynthetic process,GO:0006047~UDP-N-acetylglucosamine metabolic process,GO:0006487~protein N-linked glycosylation,GO:0006488~dolichol-linked oligosaccharide biosynthetic process,GO:0019348~dolichol metabolic process,	1
golgin Putative 1(GC1)	BP = GO:0000301~retrograde transport, vesicle recycling within Golgi,GO:0007030~Golgi organization,GO:0031640~killing of cells of other organism,GO:0050832~defense response to fungus,GO:0055114~oxidation-reduction process,	1
heat shock cognate protein 70-1(HSC70-1)	BP = GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0006457~protein folding,GO:0009408~response to heat,GO:0009409~response to cold,GO:0009615~response to virus,GO:0010187~negative regulation of seed germination,GO:0042742~defense response to bacterium,GO:0046686~response to cadmium ion,GO:0050832~defense response to fungus,GO:0090332~stomatal closure,GO:0098542~defense response to other organism,	1
histone acetyltransferase of the CBP family 1(HAC1)	BP = GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0006357~regulation of transcription from RNA polymerase II promoter,GO:0006473~protein acetylation,GO:0009294~DNA mediated transformation,GO:0009908~flower development,GO:0016573~histone acetylation,GO:0048573~photoperiodism, flowering,	1
HXXXD-type acyl-transferase family protein(AT5G23970)	CC = GO:0005737~cytoplasm,	1
HXXXD-type acyl-transferase family protein(PMAT2)	BP = GO:0009636~response to toxic substance,	1
hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase(HCT)	BP = GO:0009809~lignin biosynthetic process,GO:0009963~positive regulation of flavonoid biosynthetic process,GO:0010252~auxin homeostasis,	1
inositol transporter 1(INT1)	BP = GO:0015798~myo-inositol transport,GO:0055085~transmembrane transport,	1
Insulinase (Peptidase family M16) protein(AT1G51980)	BP = GO:0006508~proteolysis,GO:0006627~protein processing involved in protein targeting to mitochondrion,GO:0009651~response to salt stress,GO:0009735~response to cytokinin,GO:0055114~oxidation-reduction process,	1

kinase associated protein phosphatase(KAPP)	BP = GO:0006470~protein dephosphorylation,GO:0007165~signal transduction,GO:0009934~regulation of meristem structural organization,GO:0016310~phosphorylation,	1
kinase family with leucine-rich repeat domain-containing protein(AT5G25930)	BP = GO:0006468~protein phosphorylation,	1
L-fucokinase/GDP-L-fucose pyrophosphorylase(FKGP)	BP = GO:0005975~carbohydrate metabolic process,GO:0016310~phosphorylation,GO:0042352~GDP-L-fucose salvage,	1
Leucine-rich repeat protein kinase family protein(AT3G28450)	BP = GO:0006468~protein phosphorylation,GO:0006952~defense response,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,GO:1900150~regulation of defense response to fungus,GO:1900425~negative regulation of defense response to bacterium,	1
light-harvesting chlorophyll-protein complex I subunit A4(LHCA4)	BP = GO:0009409~response to cold,GO:0009644~response to high light intensity,GO:0009645~response to low light intensity stimulus,GO:0009765~photosynthesis, light harvesting,GO:0009768~photosynthesis, light harvesting in photosystem I,GO:0015979~photosynthesis,GO:0018298~protein-chromophore linkage,GO:0080167~response to karrikin,	1
lon protease 2(LON2)	BP = GO:0006515~misfolded or incompletely synthesized protein catabolic process,GO:0016485~protein processing,GO:0016558~protein import into peroxisome matrix,GO:0016560~protein import into peroxisome matrix, docking,GO:0040007~growth,GO:0048527~lateral root development,	1
Mitochondrial substrate carrier family protein(BOU)	BP = GO:0006412~translation,GO:0006810~transport,GO:0006839~mitochondrial transport,GO:0015822~ornithine transport,	1
Myosin heavy chain-related protein(AT5G52280)	CC = GO:0009507~chloroplast,	1
non-intrinsic ABC protein 11(ABCI13)	BP = GO:0006869~lipid transport,	1
OPC-8:0 CoA ligase1(OPCL1)	BP = GO:0008152~metabolic process,GO:0009611~response to wounding,GO:0009695~jasmonic acid biosynthetic process,GO:0009698~phenylpropanoid metabolic process,	1
oxoprolinase 1(OXP1)	BP = GO:0006751~glutathione catabolic process,	1
P-loop containing nucleoside triphosphate hydrolases superfamily protein(AT3G58570)	BP = GO:0010501~RNA secondary structure unwinding,	1
Peroxidase superfamily protein(AT3G28200)	BP = GO:0006979~response to oxidative stress,GO:0042744~hydrogen peroxide catabolic process,GO:0055114~oxidation-reduction process,	1
phosphoglucan, water dikinase(PWD)	BP = GO:0005975~carbohydrate metabolic process,GO:0016310~phosphorylation,	1
phytochrome E(PHYE)	BP = GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0009584~detection of visible light,GO:0009585~red, far-red light phototransduction,GO:0017006~protein-tetrapyrrole linkage,GO:0018298~protein-chromophore linkage,	1
phytosulfokin receptor 1(PSKR1)	BP = GO:0006468~protein phosphorylation,GO:0007169~transmembrane receptor protein tyrosine kinase signaling pathway,GO:0009611~response to wounding,GO:0031347~regulation of defense response,GO:0045087~innate immune response,	1

PLANT CADMIUM RESISTANCE 2(PCR2)	BP = GO:0006979~response to oxidative stress,	1
Plant invertase/pectin methylesterase inhibitor superfamily(AT5G27870)	BP = GO:0042545~cell wall modification,GO:0045490~pectin catabolic process,	1
poly(A) polymerase 1(PAPS1)	BP = GO:0006378~mRNA polyadenylation,GO:0008285~negative regulation of cell proliferation,GO:0009908~flower development,GO:0031123~RNA 3'-end processing,GO:0043631~RNA polyadenylation,GO:0045824~negative regulation of innate immune response,GO:0048366~leaf development,GO:0048451~petal formation,	1
poly(A) polymerase 3(PAPS3)	BP = GO:0006351~transcription, DNA-templated,GO:0006378~mRNA polyadenylation,	1
protein EMBRYO DEFECTIVE 2247(EMB2247)	BP = GO:0006418~tRNA aminoacylation for protein translation,GO:0006438~valyl-tRNA aminoacylation,GO:0009793~embryo development ending in seed dormancy,	1
Protein kinase superfamily protein(AT1G79640)	BP = GO:0006468~protein phosphorylation,	1
Protein kinase superfamily protein(AT3G07070)	BP = GO:0006468~protein phosphorylation,	1
Protein kinase superfamily protein(AT5G11850)	BP = GO:0006468~protein phosphorylation,	1
Protein kinase superfamily protein(AT5G14720)	CC = GO:0005634~nucleus,GO:0005737~cytoplasm,GO:0005886~plasma membrane,	1
Protein kinase superfamily protein(emb1187)	BP = GO:0006646~phosphatidylethanolamine biosynthetic process,GO:0008654~phospholipid biosynthetic process,GO:0009793~embryo development ending in seed dormancy,GO:0016310~phosphorylation,	1
Protein kinase superfamily protein(MHK)	BP = GO:0006468~protein phosphorylation,	1
Protein kinase superfamily protein(PERK5)	BP = GO:0046777~protein autophosphorylation,	1
Protein kinase superfamily protein(PID)	BP = GO:0006468~protein phosphorylation,GO:0009733~response to auxin,GO:0009734~auxin-activated signaling pathway,GO:0009926~auxin polar transport,GO:0009958~positive gravitropism,GO:0035556~intracellular signal transduction,GO:0048766~root hair initiation,GO:0048767~root hair elongation,GO:0048825~cotyledon development,GO:0048827~phyllome development,GO:0080167~response to karrikin,	1
Ras-related small GTP-binding family protein(ARA7)	BP = GO:0006897~endocytosis,GO:0007264~small GTPase mediated signal transduction,	1
RHOMBOID-like protein 15(RBL15)	BP = GO:0006508~proteolysis,	1
RING/U-box superfamily protein(AT5G43420)	BP = GO:0016567~protein ubiquitination,	1

RNA polymerase II degradation factor-like protein (DUF1296)(AT1G29370)	CC = GO:0005634~nucleus,	1
S-adenosyl-L-methionine-dependent methyltransferases superfamily protein(AT1G29470)	BP = GO:0032259~methylation,	1
S-adenosyl-L-methionine-dependent methyltransferases superfamily protein(AT2G27590)	BP = GO:0032259~methylation,	1
S-adenosyl-L-methionine-dependent methyltransferases superfamily protein(AT3G60910)	BP = GO:0032259~methylation,	1
S-locus lectin protein kinase family protein(AT4G27290)	BP = GO:0006468~protein phosphorylation,GO:0006952~defense response,GO:0048544~recognition of pollen,	1
SPA1-related 3(SPA3)	BP = GO:0009585~red, far-red light phototransduction,	1
sucrose nonfermenting-like protein(SNF4)	BP = GO:0000266~mitochondrial fission,GO:0005975~carbohydrate metabolic process,GO:0009859~pollen hydration,GO:0016559~peroxisome fission,GO:0042149~cellular response to glucose starvation,GO:0045859~regulation of protein kinase activity,GO:0046777~protein autophosphorylation,GO:2000377~regulation of reactive oxygen species metabolic process,	1
TCP-1/cpn60 chaperonin family protein(AT3G11830)	BP = GO:0006457~protein folding,GO:0046686~response to cadmium ion,	1
tetratricopeptide repeat (TPR)-containing protein(TSK)	BP = GO:0000086~G2/M transition of mitotic cell cycle,GO:0009933~meristem structural organization,GO:0009934~regulation of meristem structural organization,GO:0016569~covalent chromatin modification,GO:0051301~cell division,	1
Transmembrane amino acid transporter family protein(AT1G47670)	BP = GO:0003333~amino acid transmembrane transport,GO:0006865~amino acid transport,	1
tryptophan aminotransferase related 2(TAR2)	BP = GO:0009684~indoleacetic acid biosynthetic process,GO:0009723~response to ethylene,GO:0009793~embryo development ending in seed dormancy,GO:0009851~auxin biosynthetic process,GO:0009908~flower development,GO:0009958~positive gravitropism,GO:0010078~maintenance of root meristem identity,GO:0010087~phloem or xylem histogenesis,GO:0010588~cotyledon vascular tissue pattern formation,GO:0042742~defense response to bacterium,GO:0043562~cellular response to nitrogen levels,GO:0048367~shoot system development,GO:0048467~gynoecium development,GO:0048527~lateral root development,GO:0048825~cotyledon development,GO:0080022~primary root development,	1
UbiA prenyltransferase family protein(G4)	BP = GO:0015995~chlorophyll biosynthetic process,	1
ubiquitin conjugating enzyme 4(UBC4)	BP = GO:0006511~ubiquitin-dependent protein catabolic process,GO:0009408~response to heat,GO:0016567~protein ubiquitination,	1
UDP-glucosyl transferase 85A5(UGT85A5)	BP = GO:0008152~metabolic process,GO:0009813~flavonoid biosynthetic process,GO:0052696~flavonoid glucuronidation,	1
Xanthine/uracil permease family protein(AT2G34190)	BP = GO:0055085~transmembrane transport,	1

Table S18. Auxin outlier genes

Functional annotation of the outlier genes associated with the auxin pathway. Gene ontology is the *DAVID GOTERM_BP_DIRECT*. *Number of replicate* pairs denotes how many replicate pairs the gene was detected as an outlier.

Gene	Gene ontology	Number replicate pairs
Auxin-responsive GH3 family protein(GH3.1)	GO:0009733~response to auxin,	8
zinc induced facilitator-like 1(ZIFL1)	GO:0009414~response to water deprivation,GO:0009630~gravitropism,GO:0010540~basipetal auxin transport,GO:0048364~root development,GO:0055085~transmembrane transport,GO:0080167~response to karrikin,GO:0090333~regulation of stomatal closure,	6
SWAP (Suppressor-of-White-APricot)/surp domain-containing protein(TGH)	GO:0006396~RNA processing,GO:0006397~mRNA processing,GO:0009734~auxin-activated signaling pathway,GO:0009926~auxin polar transport,GO:0010087~phloem or xylem histogenesis,GO:0030422~production of siRNA involved in RNA interference,GO:0035196~production of miRNAs involved in gene silencing by miRNA,GO:0040008~regulation of growth,	6
ATP binding cassette subfamily B19(ABCB19)	GO:0008361~regulation of cell size,GO:0009637~response to blue light,GO:0009639~response to red or far red light,GO:0009640~photomorphogenesis,GO:0009733~response to auxin,GO:0009734~auxin-activated signaling pathway,GO:0009926~auxin polar transport,GO:0009958~positive gravitropism,GO:0010218~response to far red light,GO:0010540~basipetal auxin transport,GO:0010541~acropetal auxin transport,GO:0043481~anthocyanin accumulation in tissues in response to UV light,GO:0048364~root development,GO:0048443~stamen development,GO:0048527~lateral root development,GO:0055085~transmembrane transport,GO:0060918~auxin transport,	6
Transcriptional factor B3 family protein / auxin-responsive factor AUX/IAA-like protein(NPH4)	GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0007275~multicellular organism development,GO:0009630~gravitropism,GO:0009638~phototropism,GO:0009723~response to ethylene,GO:0009733~response to auxin,GO:0009734~auxin-activated signaling pathway,GO:0009785~blue light signaling pathway,GO:0010311~lateral root formation,GO:0040008~regulation of growth,GO:0045893~positive regulation of transcription, DNA-templated,GO:0048366~leaf development,GO:0048527~lateral root development,	5
P-glycoprotein 20(ABCB20)	GO:0010315~auxin efflux,GO:0010540~basipetal auxin transport,GO:0010541~acropetal auxin transport,GO:0055085~transmembrane transport,	5

hypothetical protein (duplicated DUF1399)(AT4G37900)	GO:0006979~response to oxidative stress,GO:0009650~UV protection,GO:0009734~auxin-activated signaling pathway,GO:0010928~regulation of auxin mediated signaling pathway,GO:0071470~cellular response to osmotic stress,GO:2000028~regulation of photoperiodism, flowering,	5
tryptophan aminotransferase related 2(TAR2)	GO:0009684~indoleacetic acid biosynthetic process,GO:0009723~response to ethylene,GO:0009793~embryo development ending in seed dormancy,GO:0009851~auxin biosynthetic process,GO:0009908~flower development,GO:0009958~positive gravitropism,GO:0010078~maintenance of root meristem identity,GO:0010087~phloem or xylem histogenesis,GO:0010588~cotyledon vascular tissue pattern formation,GO:0042742~defense response to bacterium,GO:0043562~cellular response to nitrogen levels,GO:0048367~shoot system development,GO:0048467~gynoecium development,GO:0048527~lateral root development,GO:0048825~cotyledon development,GO:0080022~primary root development,	4
sec7 domain-containing protein(GN)	GO:0000911~cytokinesis by cell plate formation,GO:0001736~establishment of planar polarity,GO:0006897~endocytosis,GO:0007155~cell adhesion,GO:0009793~embryo development ending in seed dormancy,GO:0009826~unidimensional cell growth,GO:0009880~embryonic pattern specification,GO:0009942~longitudinal axis specification,GO:0010087~phloem or xylem histogenesis,GO:0010274~hydrotropism,GO:0010311~lateral root formation,GO:0010540~basipetal auxin transport,GO:0015031~protein transport,GO:0016192~vesicle-mediated transport,GO:0032012~regulation of ARF protein signal transduction,GO:0032509~endosome transport via multivesicular body sorting pathway,GO:0048209~regulation of vesicle targeting, to, from or within Golgi,GO:0048765~root hair cell differentiation,GO:0050790~regulation of catalytic activity,GO:0071555~cell wall organization,	4
P-glycoprotein 21(ABCB21)	GO:0010315~auxin efflux,GO:0010540~basipetal auxin transport,GO:0055085~transmembrane transport,GO:0060919~auxin influx,	4
hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase(HCT)	GO:0009809~lignin biosynthetic process,GO:0009963~positive regulation of flavonoid biosynthetic process,GO:0010252~auxin homeostasis,	4
FKBP-type peptidyl-prolyl cis-trans isomerase family protein(PAS1)	GO:0000413~protein peptidyl-prolyl isomerization,GO:0009734~auxin-activated signaling pathway,GO:0009735~response to cytokinin,GO:0009736~cytokinin-activated signaling pathway,GO:0009793~embryo development ending in seed dormancy,GO:0009826~unidimensional cell growth,GO:0009880~embryonic pattern specification,GO:0018208~peptidyl-proline modification,GO:0030010~establishment of cell polarity,GO:0030154~cell differentiation,GO:0042761~very long-chain fatty acid biosynthetic process,GO:0048364~root development,GO:0048513~animal organ development,GO:0048527~lateral root development,GO:0061077~chaperone-mediated protein folding,	3
auxin response factor 6(ARF6)	GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0009733~response to auxin,GO:0009734~auxin-activated signaling pathway,GO:0009908~flower development,	3

Walls Are Thin 1(WAT1)	GO:0000162~tryptophan biosynthetic process,GO:0006810~transport,GO:0006952~defense response,GO:0008652~cellular amino acid biosynthetic process,GO:0009734~auxin-activated signaling pathway,GO:0009826~unidimensional cell growth,GO:0009834~plant-type secondary cell wall biogenesis,GO:0009851~auxin biosynthetic process,GO:0010315~auxin efflux,GO:0071555~cell wall organization,GO:0090355~positive regulation of auxin metabolic process,GO:0090358~positive regulation of tryptophan metabolic process,	2
translation initiation factor 3 subunit H1(TIF3H1)	GO:0001731~formation of translation preinitiation complex,GO:0006413~translational initiation,GO:0009733~response to auxin,GO:0009737~response to abscisic acid,GO:0009738~abscisic acid-activated signaling pathway,GO:0009744~response to sucrose,GO:0009749~response to glucose,GO:0034286~response to maltose,GO:0045948~positive regulation of translational initiation,	2
structural constituent of nuclear pore(EMB2766)	GO:0006405~RNA export from nucleus,GO:0006606~protein import into nucleus,GO:0010930~negative regulation of auxin mediated signaling pathway,GO:0051028~mRNA transport,GO:0071426~ribonucleoprotein complex export from nucleus,	2
cullin 1(CUL1)	GO:0006511~ubiquitin-dependent protein catabolic process,GO:0007049~cell cycle,GO:0009733~response to auxin,GO:0009734~auxin-activated signaling pathway,GO:0009753~response to jasmonic acid,GO:0009793~embryo development ending in seed dormancy,GO:0009867~jasmonic acid mediated signaling pathway,GO:0009873~ethylene-activated signaling pathway,GO:0010087~phloem or xylem histogenesis,GO:0010265~SCF complex assembly,GO:0042752~regulation of circadian rhythm,GO:0042787~protein ubiquitination involved in ubiquitin-dependent protein catabolic process,GO:0048366~leaf development,	2
Chalcone and stilbene synthase family protein(TT4)	GO:0006810~transport,GO:0006979~response to oxidative stress,GO:0009058~biosynthetic process,GO:0009611~response to wounding,GO:0009629~response to gravity,GO:0009715~chalcone biosynthetic process,GO:0009733~response to auxin,GO:0009753~response to jasmonic acid,GO:0009813~flavonoid biosynthetic process,GO:0009926~auxin polar transport,GO:0010224~response to UV-B,GO:0031540~regulation of anthocyanin biosynthetic process,	2
auxin canalization protein (DUF828)(AT3G22810)	GO:0007165~signal transduction,	2
Transmembrane amino acid transporter family protein(AUX1)	GO:0001736~establishment of planar polarity,GO:0003333~amino acid transmembrane transport,GO:0006865~amino acid transport,GO:0009624~response to nematode,GO:0009723~response to ethylene,GO:0009733~response to auxin,GO:0009734~auxin-activated signaling pathway,GO:0009926~auxin polar transport,GO:0009958~positive gravitropism,GO:0010311~lateral root formation,GO:0048765~root hair cell differentiation,GO:0048829~root cap development,	1
RNA-metabolising metallo-beta-lactamase family protein(emb2746)	GO:0009658~chloroplast organization,GO:0009793~embryo development ending in seed dormancy,GO:0009942~longitudinal axis specification,GO:0060918~auxin transport,	1

Protein kinase superfamily protein(PID)	GO:0006468~protein phosphorylation,GO:0009733~response to auxin,GO:0009734~auxin-activated signaling pathway,GO:0009926~auxin polar transport,GO:0009958~positive gravitropism,GO:0035556~intracellular signal transduction,GO:0048766~root hair initiation,GO:0048767~root hair elongation,GO:0048825~cotyledon development,GO:0048827~phyllome development,GO:0080167~response to karrikin,	1
Flavin-binding monooxygenase family protein(YUC2)	GO:0009851~auxin biosynthetic process,GO:0055114~oxidation-reduction process,	1
auxin response factor 8(ARF8)	GO:0006351~transcription, DNA-templated,GO:0006355~regulation of transcription, DNA-templated,GO:0009733~response to auxin,GO:0009734~auxin-activated signaling pathway,GO:0009908~flower development,	1
Auxin efflux carrier family protein(AT2G17500)	GO:0009733~response to auxin,GO:0009734~auxin-activated signaling pathway,GO:0009926~auxin polar transport,GO:0010252~auxin homeostasis,GO:0010311~lateral root formation,GO:0040009~regulation of growth rate,GO:0055085~transmembrane transport,	1

Figure S1. Patterns of Dune-Headland FST, CSS and Δp in *S. lautus*

(A) Relationship between the change in allele frequency (Δp) and F_{ST} comparing all Dune vs all Headland individuals, each datapoint representing a SNP. Red denotes the top 1% F_{ST} values. **(B)** Relationship of Δp to the cluster separation score (CSS). Red denotes the top 1% CSS values. **(C)** Relationship of F_{ST} and CSS. Red denotes the SNPs considered as outliers, being detected as outliers in at least two of the following approaches: top 1% F_{ST}, top 1% CSS, and *BayeScan* posterior probability > 0.91.

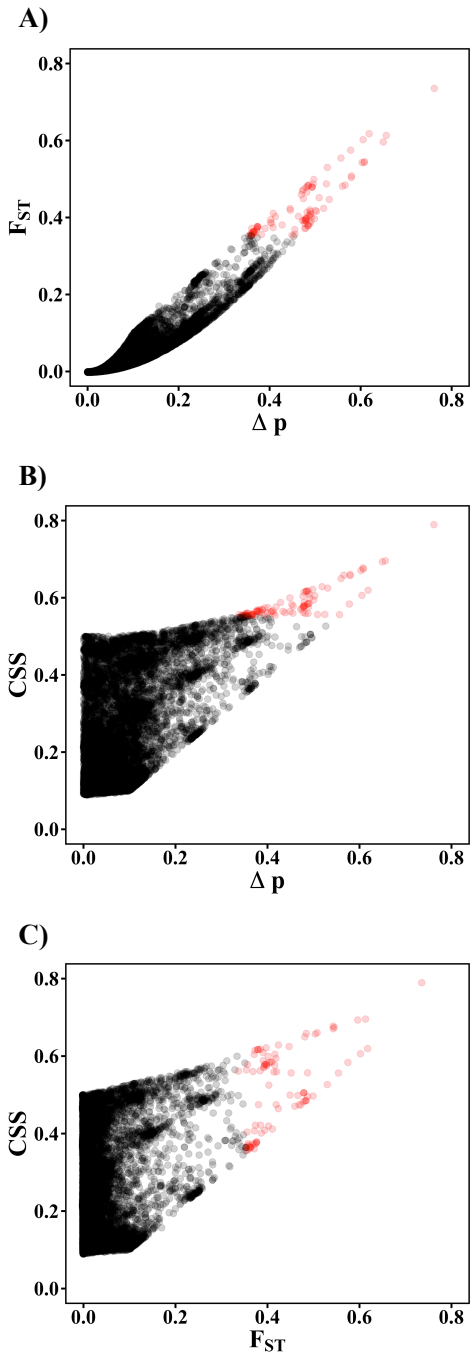


Figure S2. Principle components analysis of architecture and leaf traits in coastal ecotypes of *S. lautus*

Principle component analysis of Dune (orange) and Headland (green) phenotypes for **(A)** five architecture, and **(B)** four leaf traits across 20 populations. Ecotypes are delimited by 70% probability ellipses.

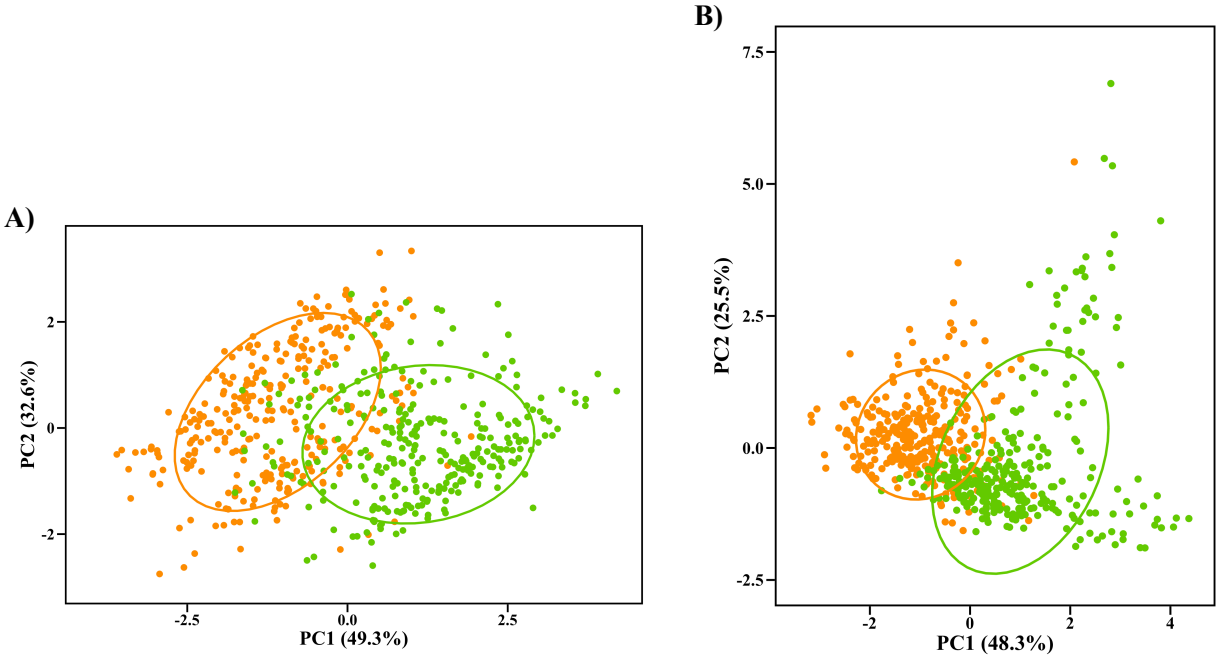


Figure S3. Trait-by-trait effect sizes in coastal ecotypes of *S. lautus*

Partial effect sizes (partial η^2) for the ecotype and pair for the trait-by-trait linear models, each dot representing a single trait. The red represents the partial effect size for all traits combined within the MANOVA. Dashed line is a 1:1 ratio, where points above the line represent a larger contribution of broad- to narrow-sense divergence. See Table S4 for details.

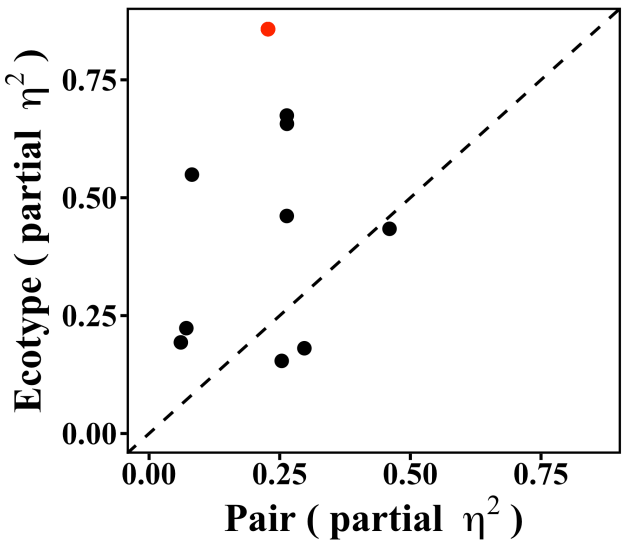


Figure S4. Phenotypic Change Vector Analysis for architecture traits in costal ecotypes from *S. lautus*

(A) PC1 and PC2 for five architecture traits across eight replicate Dune-Headland pairs. Each dot represents the population centroid (multivariate phenotypic mean), ± SE. The Dune (orange) and Headland (green) populations of a replicate pair are connected with a line. **(B)** Frequency distribution of the 28 pairwise phenotypic divergences (ΔL) between Dune-Headland replicate pairs (Supplementary Table S7). **(C)** Frequency distribution of the 28 pairwise contribution of traits (θ) between Dune-Headland replicate pairs (Supplementary Table S10).

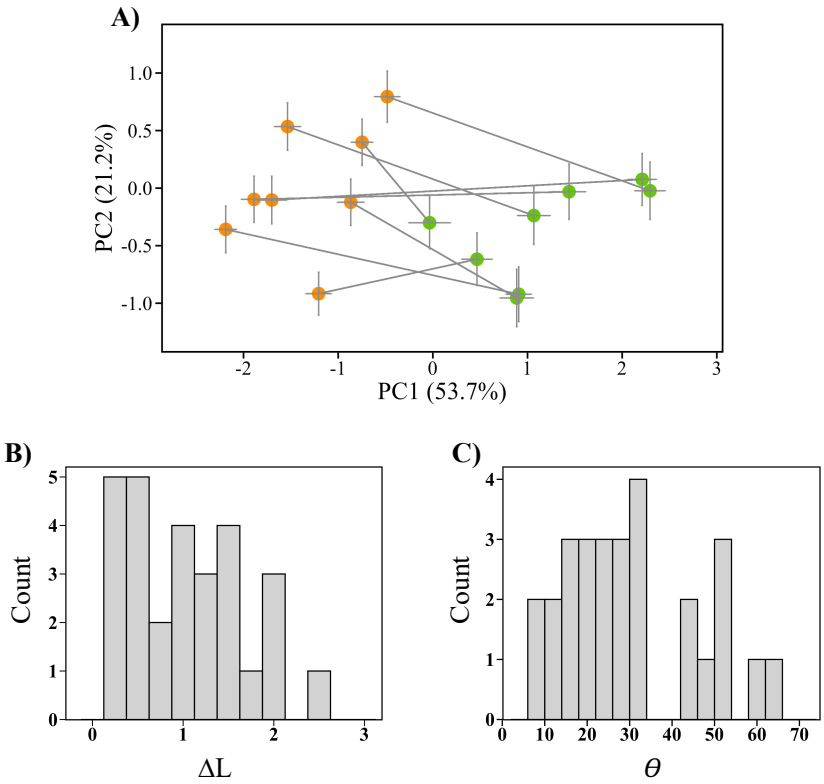


Figure S5. Phenotypic Change Vector Analysis for leaf traits in costal ecotypes from *S. lautus*

(A) PC1 and PC2 for four leaf traits across eight replicate Dune-Headland pairs. Each dot represents the population centroid (multivariate phenotypic mean), ± SE. The Dune (orange) and Headland (green) populations of a replicate pair are connected with a line. **(B)** Frequency distribution of the 28 pairwise phenotypic divergences (ΔL) between Dune-Headland replicate pairs (Supplementary Table S8). **(C)** Frequency distribution of the 28 pairwise contribution of traits (θ) between Dune-Headland replicate pairs (Supplementary Table S11).

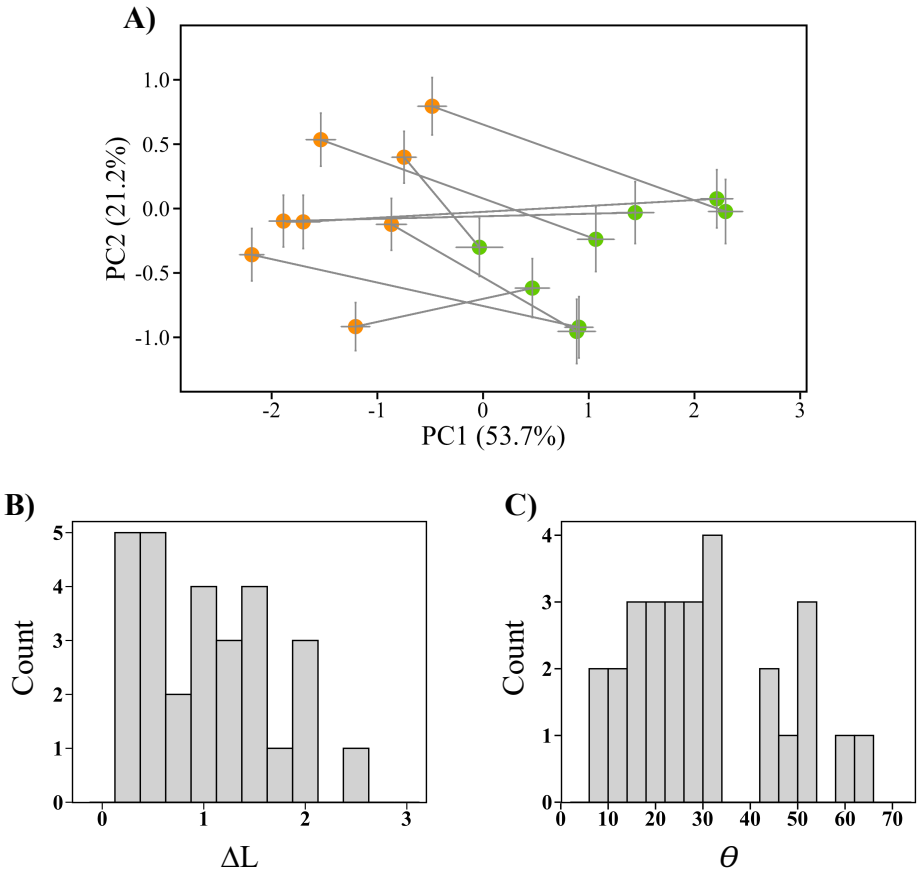


Figure S6. Relative contributions of broad- and narrow-sense genotypic parallelism

Partial effect sizes (partial η^2) for the ecotype and the pair for the linear models for nucleotide sites **(A)** genes **(B)**. Each dot represents either a single SNP **(A)** or gene **(B)**. Dashed line is a 1:1 ratio, where points above the line represent a larger contribution of broad- to narrow-sense parallelism. The red dotes denotes broad-sense outlier SNPs **(A)** and genes **(B)**. Partial effect size frequency distributions for the nucleotide sites **(C)** and genes **(D)**. Values above zero represent a larger contribution of broad- to narrow-sense parallelism.

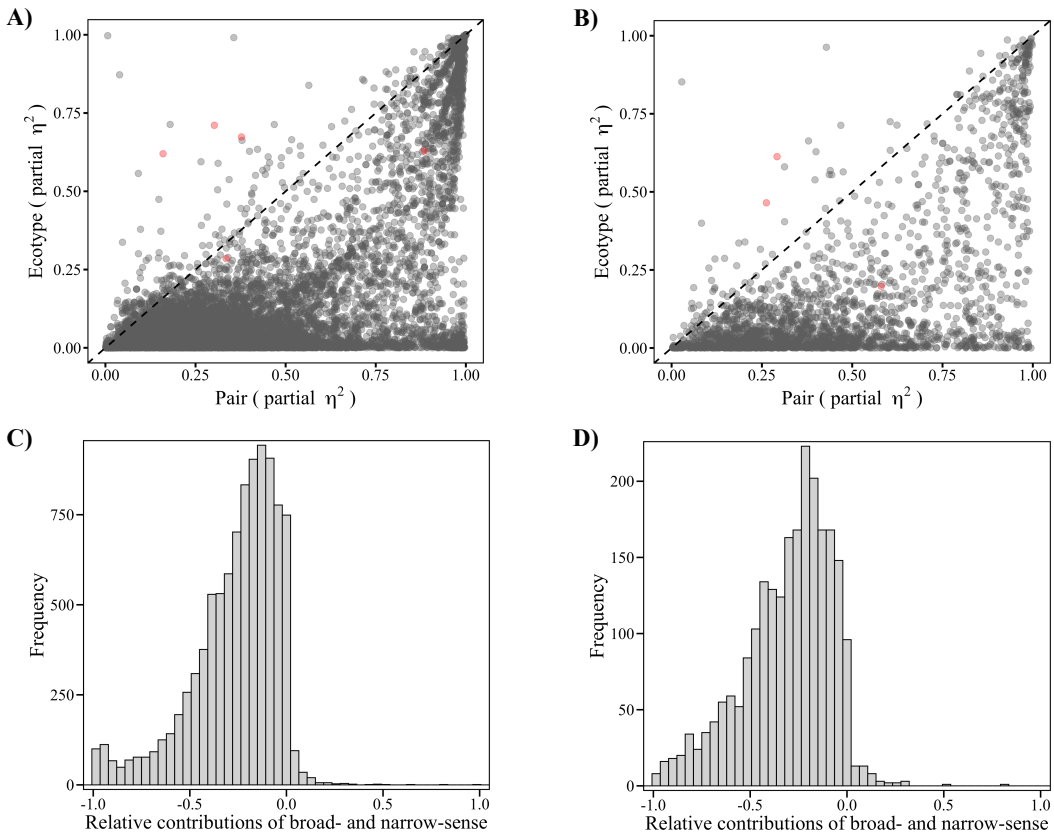
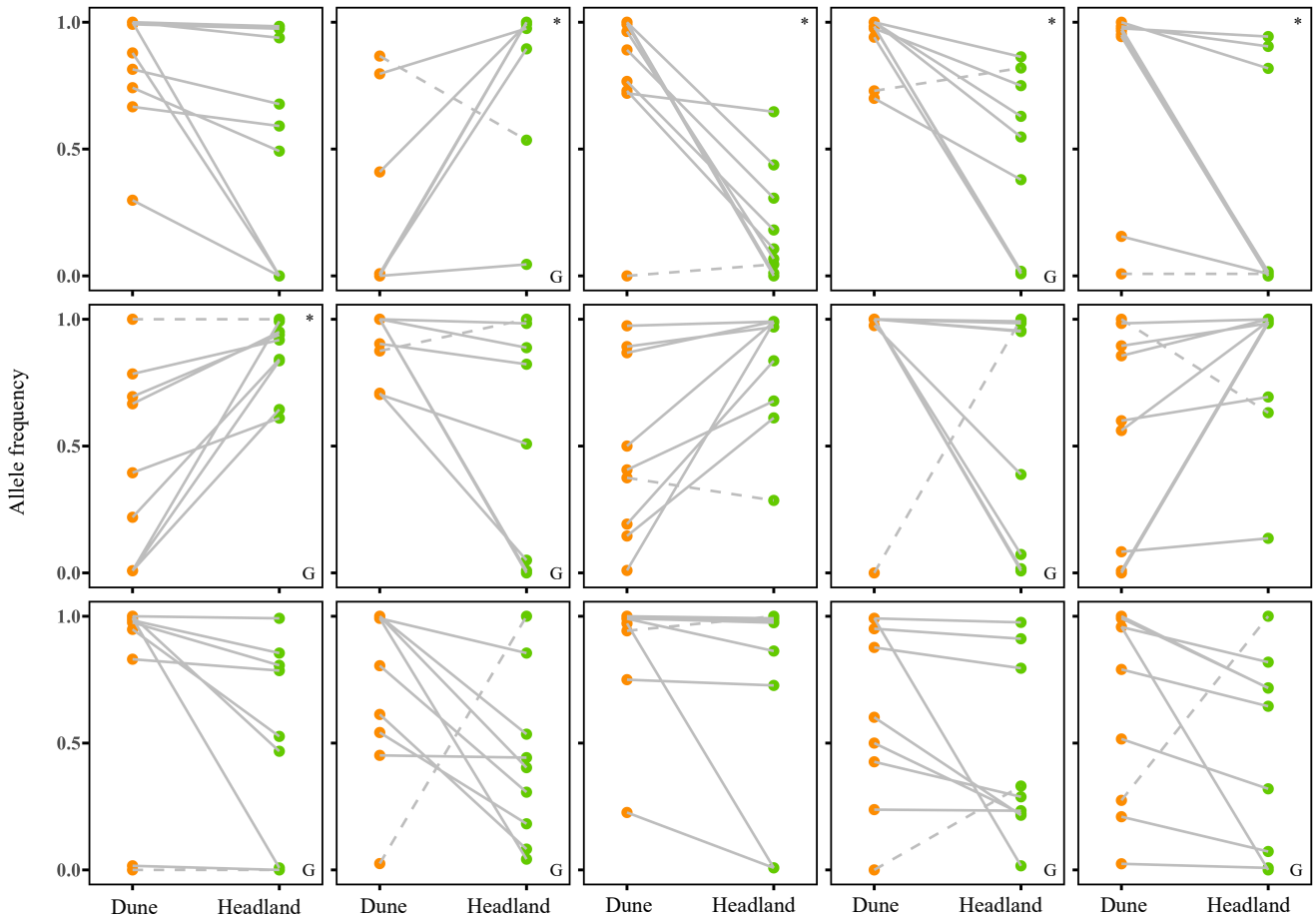


Figure S7. Broad-sense parallel nucleotide sites with concordant allele frequency differences

Broad-sense parallel nucleotide sites with concordant allele frequency differences (approach 2) in either all nine (S-statistic = 9, P = 0.004), or eight (S-statistic = 8, P = 0.04) replicate pairs. Dots represent the allele frequency value (of the reference allele) for each population. Lines connect the Dune (orange) populations to their Headland (green) pair. Dashed lines represent pairs whose Dune-Headland change in allele frequency is in the opposite direction from the majority of pairs. *G* denotes SNPs that occur within genic regions. Asterisks denote SNPs that were also detected as outliers within approach 1 (see *Methods* for details).



Supplementary R code

R code for Phenotypic Change Vector Analysis

```
# The following function estimates vector lengths, contrast, angles, and
# associated p values for differences between traits in multivariate space
# between two pairs of populations. Code adapted from Collyer & Adams (2007)

CollyerAdamsPCVA <- function(y.mat, x.mat, nPermutations = 9999) {
# First step: finalise the two x matrices (full and reduced)
x.mat.full<-x.mat
x.mat.red<-x.mat.full[,-4] # This removes the coding for the interaction effect

# Second step, estimate parameters for full and reduced models
b.mat.full<-solve(  (t(x.mat.full)%*%x.mat.full)  ) %*% (t(x.mat.full)%*%y.mat)
b.mat.red<-solve((t(x.mat.red)%*%x.mat.red))%*%(t(x.mat.red)%*%y.mat)

# Third step, Estimate LS means

# In this case, but not necessarily always, a = species and b = locality
# Vectors will be calculated for 2 species, a1 (P. jordani) and a2 (P. teyahalee)
# In our example the following dummy variables were used:
# P. jordani = 1, P. teyahalee = -1, sympatry = 1, allopatry = -1
# Thus, LS means could be calculated as follows:

a1b1<-cbind(1,1,1,1)
a1b2<-cbind(1,1,-1,-1)
a2b1<-cbind(1,-1,1,-1)
a2b2<-cbind(1,-1,-1,1)

x.ls.full<-rbind(a1b1,a1b2,a2b1,a2b2)
x.ls.red<-x.ls.full[,-4]

obs.ls.full<-x.ls.full%*%b.mat.full  # Observed ls means (full model)
obs.ls.red<-x.ls.red%*%b.mat.red     # Observed ls means (reduced model)

# Fourth Step, vector and statistics calculations

obs.a1.vect<-obs.ls.full[1,]-obs.ls.full[2,] # These are the phenotypic change vectors
obs.a2.vect<-obs.ls.full[3,]-obs.ls.full[4,]

obs.d.a1<-c(sqrt(t(obs.a1.vect)%*%obs.a1.vect)) # These are lengths of vectors
obs.d.a2<-c(sqrt(t(obs.a2.vect)%*%obs.a2.vect))

obs.contrast<-abs(obs.d.a1-obs.d.a2)
obs.angle<-c(acos(t((obs.a1.vect)/obs.d.a1)%*%((obs.a2.vect)/obs.d.a2)))
obs.angle<-obs.angle*180/pi  # This step is only necessary to convert radians to degrees

# Fifth Step, set-up permutation procedure

y.hat<-x.mat.red%*%b.mat.red     # Predicted values from reduced model
y.res<-y.mat-y.hat               # Resdiuals of reduced mode (these are the permuted units)

# PERMUTATION PROCEDURE

# Need to set-up distributions to be generated

dist.d1<-NULL
dist.d1<-rbind(dist.d1,obs.d.a1) # Observed value is first random value
dist.d2<-NULL
dist.d2<-rbind(dist.d2,obs.d.a2) # Observed value is first random value
dist.contrast<-NULL
dist.contrast<-rbind(dist.contrast,obs.contrast) # Observed value is first random value
dist.angle<-NULL
dist.angle<-rbind(dist.angle,obs.angle) # Observed value is first random value

# In addition to saving random values, it is wise to save the outcome of comparisons
# of observed and random values.
# This can be done with logical statements (below).
# Separate distributions are created for these comparisons.
# The 'p' indicates that these distributions will be used
```

```
# to calculate empirical probabilities.

pdist.contrast<-NULL
pdist.contrast<-rbind(pdist.contrast,1)
pdist.angle<-NULL
pdist.angle<-rbind(pdist.angle,1)

for(i in 1:nPermutations){
		y.res.rand <- y.res[sample(1:nrow(y.res)), ]

		# Create random values
		y.rand<-y.hat+y.res.rand

		# Estimate parameters
		b.mat.rand<-solve((t(x.mat.full)%*%x.mat.full))%*%(t(x.mat.full)%*%y.rand)

		# Calculate LS means
		rand.ls.full<-x.ls.full%*%b.mat.rand

		# Repeat fourth step for random data!

		rand.a1.vect<-rand.ls.full[1,]-rand.ls.full[2,] # These are the phenotypic change vectors
		rand.a2.vect<-rand.ls.full[3,]-rand.ls.full[4,]

		rand.d.a1<-c(sqrt(t(rand.a1.vect)%*%rand.a1.vect)) # These are lengths of vectors
		rand.d.a2<-c(sqrt(t(rand.a2.vect)%*%rand.a2.vect))

		rand.contrast<-abs(rand.d.a1-rand.d.a2)
		rand.angle<-c(acos(t((rand.a1.vect)/rand.d.a1)%*%((rand.a2.vect)/rand.d.a2)))
		rand.angle<-rand.angle*180/pi  # This step is only necessary to convert radians to degrees

		# Append distributions

		dist.d1<-rbind(dist.d1,rand.d.a1)
		dist.d2<-rbind(dist.d2,rand.d.a2)
		dist.contrast<-rbind(dist.contrast,rand.contrast)
		dist.angle<-rbind(dist.angle,rand.angle)

		aa<-ifelse(rand.contrast>=obs.contrast,1,0)
		bb<-ifelse(rand.angle>=obs.angle,1,0)

		pdist.contrast<-rbind(pdist.contrast,aa)
		pdist.angle<-rbind(pdist.angle,bb)
}

# Empirical probabilities are calculated as follows

 p.contrast<-sum(pdist.contrast)/(nPermutations+1)
 p.angle<-sum(pdist.angle)/(nPermutations+1)

 return(c(length.v1 = obs.d.a1,
	length.v2 = obs.d.a2,
	contrast = obs.contrast,
	p.contrast = p.contrast,
	angle = obs.angle,
	p.angle = p.angle))
}


# The following R code generates the required files to be used with the CollyerAdamsPCVA function

# 'master.file' is the entire phenotyping file, first column the pair, second column the ecotype,third column the population ID, the other columns the phenotype data
master.file <- read.delim ("path-to-file/allTraitsPairs", header = TRUE)

# extract the unique elements of the first column (the pair)
pairs <- unique(master.file$Pair)

# create all possible pairwise comparisons
pairwise <-combn(pairs, 2)

# create empty data frame for the results
results <- data.frame(pair1 = pairwise[1,],
		pair2 = pairwise[2,],
		length.v1 = NA,
		length.v2 = NA,
		contrast = NA,
		p.contrast = NA,
		angle = NA,
```

```
                p.angle = NA)

for(i in 1:ncol(pairwise)) {
        # subset master file to generate proto-y.mat
        y.mat <- master.file[ master.file$Pair %in% pairwise[, i], ]

        # generate x.mat
        x.mat <- matrix(NA, nrow = nrow(y.mat), ncol = 4)  # empty matrix
        x.mat[,1] <- 1  # intercept
        x.mat[,3] <- 2 * (y.mat$Ecotype == "Dune") - 1  # 1 for dunes, -1 for headlands
        x.mat[,2] <- 2 * (y.mat$Pair == pairwise[1,i]) - 1  # 1 for pair 1, -1 for pair 2
        x.mat[,4] <- x.mat[,2] * x.mat[,3]  # interaction

        # modify y.mat to fit requirements
        y.mat <- as.matrix(y.mat[, -(1:3)])

        # run CollyerAdamsPCVA function on the two matrices
        results[i, 3:ncol(results)] <- CollyerAdamsPCVA(y.mat, x.mat)
}

write.table(results,file="path-to-file/PCVA_results.txt",row.names=FALSE,col.names=TRUE,quote=FALSE)
```

R code for cluster separation score (CSS)

```
#Define the path
path <- "/Users/maddiejames/Desktop/ESC/MDS/"
#Get a list of the input files (all files within the path)
allFiles <- list.files(path = path, pattern = "\\.mds$")
#Create a dataframe to store the data
CSSsnp <- data.frame(allFiles)
names <- c("CSS")
CSSsnp[ , names] <- NA
#Open each file and calculate CSS
for(k in 1:length(allFiles)) {
  MDSall<-read.delim(paste0(path,allFiles[k]), header=T, sep="")
  #Create the ecotype column
  ecotype<-substr(MDSall$IID, start = 1, stop = 1)
  MDSall<-cbind(Ecotype = ecotype, MDSall)
  #Replace the D with Dune and H with Headland
  MDSall$Ecotype <- gsub('D', 'Dune', MDSall$Ecotype)
  MDSall$Ecotype <- gsub('H', 'Headland', MDSall$Ecotype)
  #Subet the data into Dunes and Headlands
  MDSdune <- MDSall[MDSall$Ecotype=='Dune', ]
  MDSheadland <- MDSall[MDSall$Ecotype=='Headland', ]

  #Calculate the number of Dune individuals
  m <- nrow(MDSdune)
  #Calculate the number of Headland individuals
  n <-  nrow(MDSheadland)

  #Calculate the between group distance
  sumBetween <- 0
  for(i in 1:m) {
    for(j in 1:n) {
      sumBetween <- sumBetween + abs(MDSdune$C1[i] - MDSheadland$C1[j])
    }
  }

  #Calculate the within Dune distance
  sumWithinDune <- 0
  for(i in 1:(m-1)) {
    sumWithinDune <- sumWithinDune + abs(MDSdune$C1[i] - MDSdune$C1[i+1])
  }

  #Calculate the within Headland distance
  sumWithinHeadland <- 0
  for(j in 1:(n-1)) {
    sumWithinHeadland <- sumWithinHeadland + abs(MDSheadland$C1[j] - MDSheadland$C1[j+1])
  }

  #Calculate the CSS
  CSSsnp$CSS[k] <- sumBetween/(m*n) - (m+n)*(sumWithinDune/(m^2 * (m-1)) + sumWithinHeadland/(n^2 * (n-1)))
}

  #Print the results
write.table(CSSsnp, file="/Users/maddiejames/Desktop/ESC/CSSsnp_ESC.txt", row.names = FALSE, col.names = TRUE, quote = FALSE)
```