Evidence for mutation-order speciation in Australian wildflower

Maria C. Melo¶, Maddie E. James¶, Federico Roda, Diana Bernal-Franco, Melanie J. Wilkinson, Huanle Liu, Greg M. Walter, and Daniel Ortiz-Barrientos*

¶These authors contributed equally to this work.

The University of Queensland, School of Biological Sciences, St. Lucia, Queensland 4072, Australia.

*Corresponding author:

Daniel Ortiz-Barrientos

The University of Queensland

School of Biological Sciences

St. Lucia, Queensland 4072, Australia

E-mail: d.ortizbarrientos@uq.edu.au

Phone: 617-3365-1767

Fax: 617-3365-1655

Keywords: Parallel Speciation, Natural Selection, Reproductive Isolation, Gene Flow, Clines, *Senecio*

In a number of animal species, divergent natural selection has repeatedly and independently driven the evolution of reproductive isolation between populations adapted to contrasting, but not to similar environments[1]. This process is known as parallel ecological speciation, and examples in plants are enigmatically rare[2]. Here, we perform a comprehensive test of the ecological speciation hypothesis in an Australian wildflower where parapatric populations found in coastal sand dunes (Dune ecotype) and headlands (Headland ecotype) have repeatedly and independently diverged in growth habit. Consistent with a role for divergent natural selection driving the evolution of reproductive isolation, we found that Dune populations with erect growth habit were easy to transplant across sand dunes, were largely interfertile despite half-a-million years of divergence, and were reproductively isolated from equally divergent Headland populations with prostrate growth habit. However, we unexpectedly discovered that both extrinsic and intrinsic reproductive isolation has evolved between prostrate Headland populations, suggesting that populations evolving convergent phenotypes can also rapidly become new species. Mutation-order speciation[2], where the random accumulation of adaptive alleles create genetic incompatibilities between populations inhabiting similar habitats, provides a compelling explanation for these complex patterns of reproductive isolation. Our results suggest that natural selection can drive speciation effectively, but environmental and genetic complexity might make parallel ecological speciation uncommon in plants despite strong morphological convergence.

Convergent and divergent phenotypes can evolve when populations adapt repeatedly to contrasting environments. For instance, stickleback fishes adapted to freshwater lakes from the northern hemisphere lack armour and are similar in size when compared to their armoured and smaller relatives from the ocean[3]. This pattern of convergence and divergence is also seen in cave fishes[4], stick insects[5], marine snails[6], eucalypts[7], and dandelions[8,9], and is thought to be driven by the repeated and independent action of divergent natural selection between multiple population pairs occupying contrasting environments.

Senecio lautus is a species complex from Australia and the pacific islands, generally known as the common groundsel, that has repeatedly evolved two coastal ecotypes adapted to contrasting environments[10]. The Dune (D) ecotype occupies the sheltered areas behind the sand dunes of the beach, whereas the Headland (H) ecotype lives in exposed and salty rocky outcrops and cliff tops that interrupt the beach. These ecotypes have contrasting growth habits, where Dunes grow erect and Headlands grow prostrate and form mats on top of the ground[10,11] (Fig. 1A). Multiple transplant experiments have shown that Dune and Headland populations are exposed to strong divergent natural selection[12-14], which creates strong ecologically-dependent reproductive isolation between them[12,14,15]. However, it remains unknown whether intrinsic reproductive isolation has evolved as a by-product of adaptation to contrasting or similar environments. Here we test the fundamental predictions of the parallel speciation hypothesis for the evolution of both extrinsic and intrinsic reproductive isolation in *S. lautus* by using a range of glasshouse, field, and laboratory experiments.

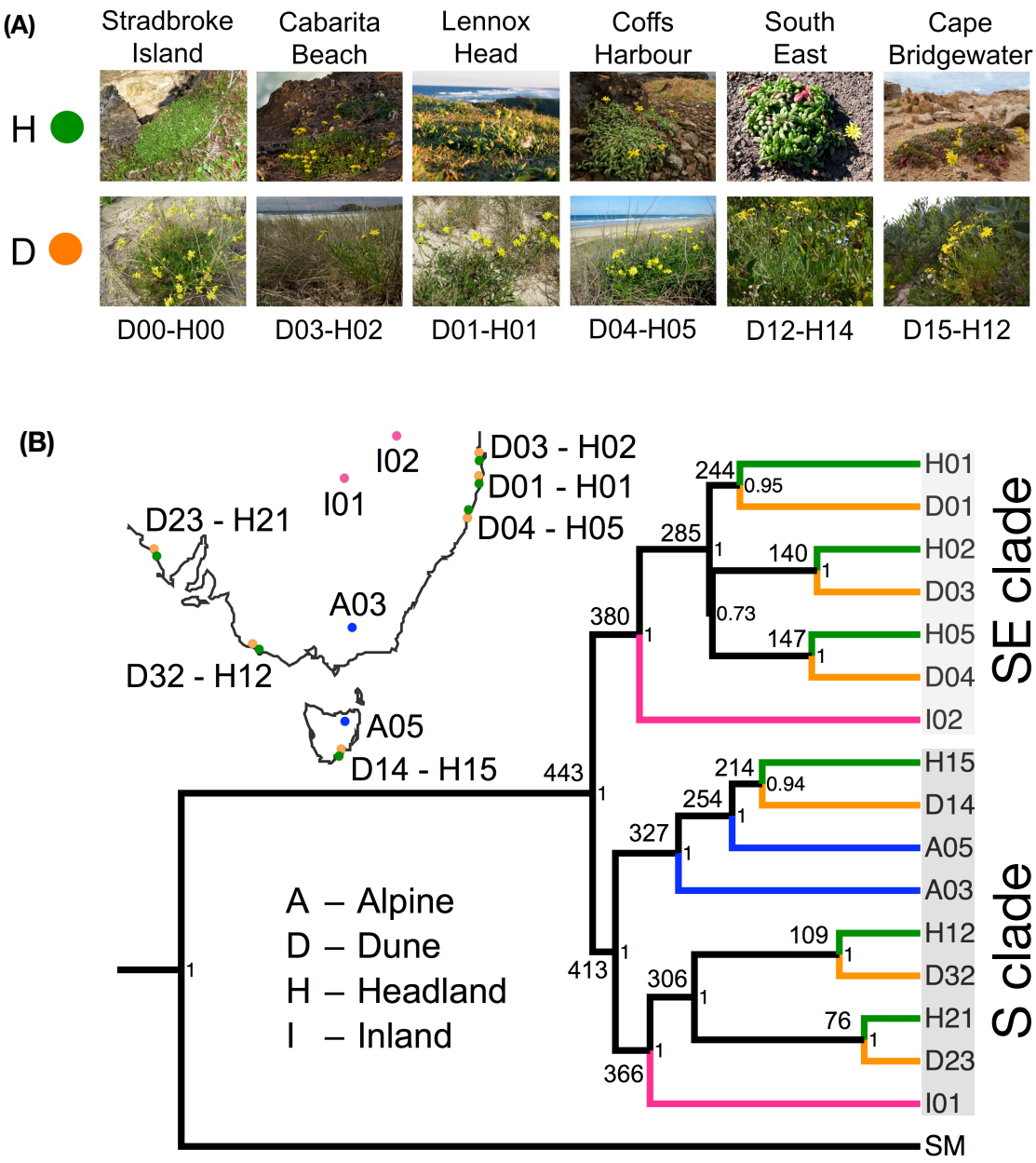

Fig. 1. A) Morphological differences between the two ecotypes: Dunes grow erect and have big leaves, while Headlands grow prostrate and have smaller succulent leaves. Modified from[10], with permission from the authors. **B)** Geographic distribution of 16 populations in the study and phylogeny based on 13 neutral markers using Bayesian inference. SE and S respectively refer to the South-Eastern and Southern clades observed in the phylogeny. Numbers above nodes are estimates of divergence time obtained in IMa2, whereas numbers below nodes are credible posterior probabilities obtained in BEAST*.

We first confirmed if divergent natural selection drove the independent and repeated evolution of these two coastal ecotypes (Fig. 1A). As we observed before[10], Dune and Headland populations that live next to each other (parapatric DH pairs) are phylogenetically more related to each other than to any other Dune or Headland population, respectively (Fig. 1B, Table S1). Using 13 neutral nuclear loci common to sixteen populations of *S. lautus* we show that DH pairs are sister taxa (see Table S5-6 for genetic details for each locus). DH pairs sorted into two main clades (South-Eastern, and Southern), which we also detected in a population genetics analysis of the same loci using the STRUCTURE[35] algorithm (Supp. Fig. 1). These results suggest that coastal ecotypes in *S. lautus* have evolved at least twice independently.

Next, we asked if the independent evolution of Dune and Headland populations led to the evolution of reproductive isolation between among occupying contrasting environments (DxH crosses or transplants), but not among populations found in similar environments (DxD and HxH crosses or transplants). We performed crosses in common garden conditions both within and between the South-Eastern (SE) and Southern (S) clades, thus correcting for divergence time, and reanalysed previous transplant experiments in the field across multiple Dune and Headland habitats[13]. In each cross, we measured the ability to produce hybrid seed as well as ability to germinate relative to the average of the parents. In transplant experiments in the field we measured the viability component of fitness by tracking individual mortality since germination.

Consistent with parallel speciation driven by divergent natural selection, we found that populations that inhabit similar environments were on average reproductively and ecologically compatible, while populations inhabiting different environments were reproductively and ecologically isolated. We asked if reproductive and ecological compatibility was similar between Dune populations or between Headland populations. Surprisingly, we found that only Dune populations were easily crossed or transplanted into each other's habitat, whereas Headland populations were difficult to cross and transplant between headland sites (Fig. 3 and 4, Supplementary Table S3 and S4). Altogether, it appears that costal parapatric pairs of *S. lautus* have evolved by divergent natural selection, but Headland populations have done it uniquely every time, thus creating further reproductive isolation among them.

Contrary to expectations from the parallel speciation hypothesis, our results indicate that the replicated evolution of the same morphology does not guarantee reproductive compatibility between populations. A straightforward explanation is adaptive polygenic divergence between populations inhabiting similar environments. Here, populations evolving similar values of a complex trait accumulate adaptive differences in random order[1,2] some of which incidentally cause intrinsic reproductive isolation. This speciation mode is referred to as mutation-order speciation, and it has been used as an alternative view to

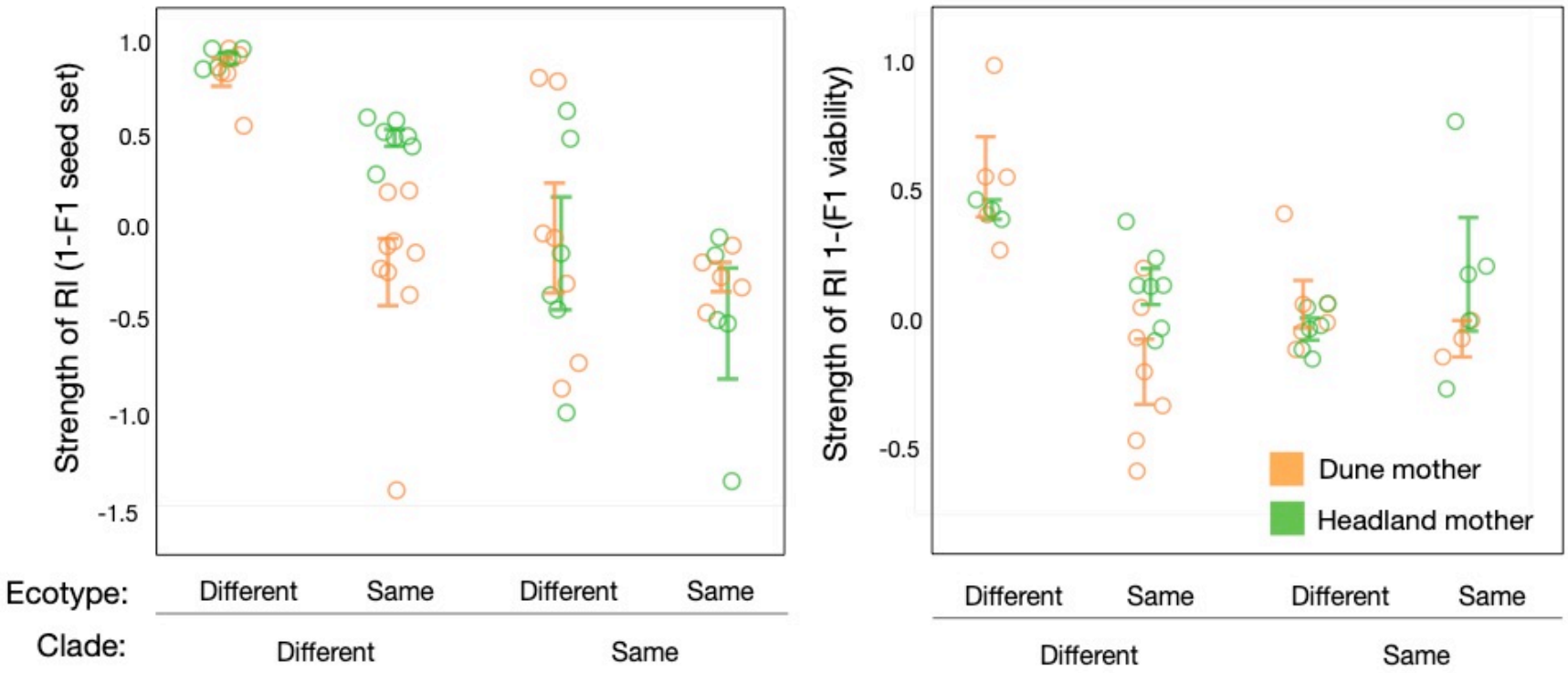

Fig. 3. F1 hybrid seed set and F1 hybrid inviability in crosses between coastal populations in *Senecio lautus*. Each panel represents a combination of crosses within ecotype (DxD, with Dune mother; HxH, with Headland mother) or between ecotypes (DxH, with Dune mother; HxD, with Headland mother) from the same or from a different clade. Positive values of Strength of RI imply that hybrids perform worse than parents, and negative values that hybrids perform better than parents.

speciation driven by divergent natural selection. Our findings suggest that the two processes might have more in common than previously anticipated.

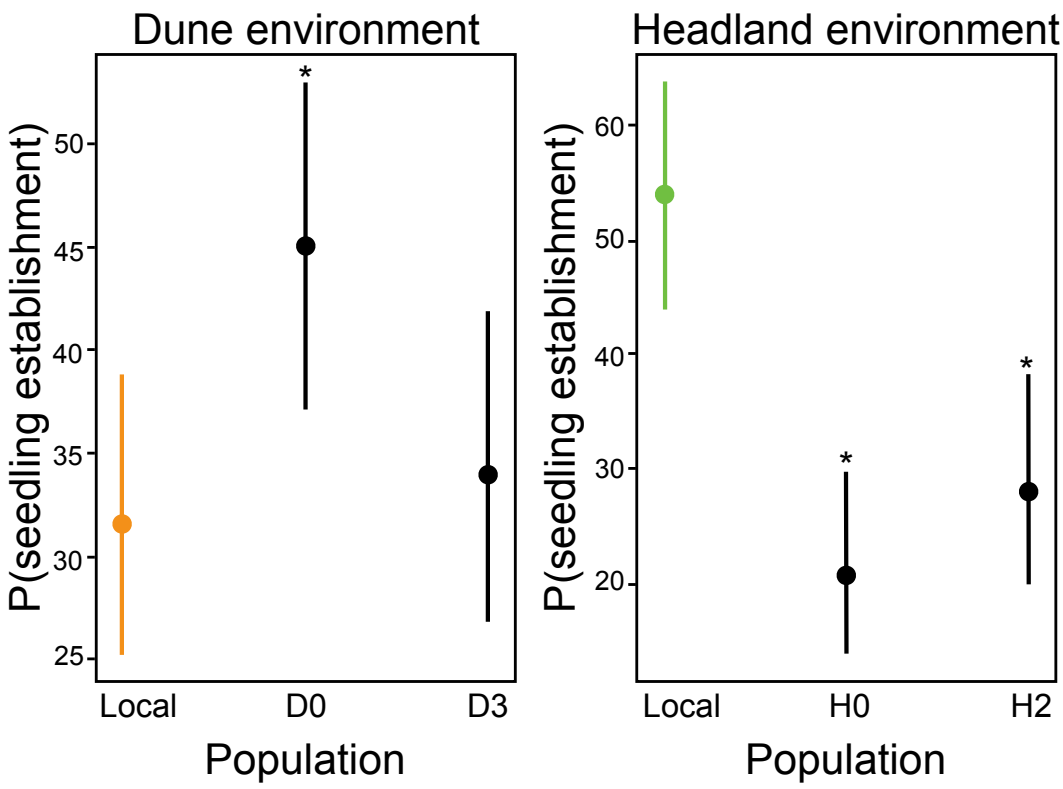

Figure 4. Field performance of three replicate populations from each ecotype in their native habitat at Lennox Head (NSW, pair D01-H01) in the form of Estimated S (survival). Probability of reaching seedling establishment was similar for all Dune populations (or there was hybrid vigour), but was much higher for the local \than the other two Headland populations (asterisks denote significant difference to local population).

Previous results in *S. lautus* have suggested that DH pairs are divergent in similar signalling pathways even if the differentiated genes are not common across populations pairs[16]. One of these pathways controls auxin movement, which is known to affect variation in growth habit in *Arabidopsis* and other model plant systems. Re-use of a genetic pathway might be an easy solution for creating

phenotypic replicates and confer local adaptation, yet it might incidentally lead to the evolution of intrinsic reproductive isolation as hormone signalling diverges between them. This might explain why parallel ecological speciation might be rare in some organisms, such as plants where many developmental and integrated phenotypes rely on morphogens.

METHODS

Molecular analyses using nuclear genes

Study populations and sample preparation

Leaf samples were collected from six parapatric *Senecio lautus* Dune and Headland populations (*S. pinnatifiolius* var. *pinnatifolius*, *S. pinnatifolius* var. *maritimus*) as well as two Inland and two Alpine populations (*S. pinnatifolius* var *serratus* and *S. pinnatifolius* var. *alpinus*) (Table S1 and map in Fig. 1, and supplementary materials[10] for more details). *Senecio madagascariensis*, a closely related species to *S. lautus* from Africa was included as an outgroup[10]. We extracted DNA with a modified CTAB protocol and samples were standardised to 30ng/µL[10].

Library construction and sequencing

We undertook targeted re-sequencing in the Fluidigm Access Array system[17], which enables the simultaneous amplification of 48 primers across 48 individuals, whilst barcoding individuals for next generation sequencing. Individuals were barcoded using the 4-primer PCR process with 454 barcodes[17]. Individuals were randomised and pooled in equimolar quantities into two pools, and sent for sequencing at the Beijing Genomics Institute. Pools were cleaned with an Agencourt AMPure XP purification kit, and then sequenced using the emPCR (Lib-A) kit for bi-directional sequencing on two lanes of the Roche GS FLX Titanium platform.

SNP calling

After trimming barcodes with TagCleaner[18], reads that had more than 40% low quality bases (<Q20), had Ns higher than 2%, or were shorter than 50bp were removed. PRGmatic [19] was used for read alignment. The PRGmatic pipeline aligns reads into loci and constructs haplotypes for each individual using dependent programs including CAP3[20], BWA[21], SAMtools[22], and VarScan[23]. We used default parameters and an overlap of 100bp for read alignment, as suggested for 454 data[19]. A locus was only considered for further analyses if haplotypes were assigned in six or more individuals per population. BLAT[24] was used to map loci from PRGmatic to the expected amplicons. 26 loci were selected and aligned in MUSCLE[25] and subsequently used in phylogenetic ananylses and neutrality tests. See Tables S5-S7 for genes and primer sequences.

DNA sequence polymorphism analyses

Deviations from neutrality[26] for each gene within each population were tested using HKA[27], Tajima's D[28] and Fu and Li's tests[29] using DNAsp. Loci selected for phylogenetic and population genetics analyses did not reject neutrality via the HKA test, or via the other two tests if HKA was impossible to perform because a locus did not amplify in the outgroup species. As a consequence, 26 loci were used in estimating gene flow between different pairwise comparisons between populations, and 13 neutral loci across populations were used for the phylogeny (Table S5-S7).

Phylogenetic analysis and population differentiation for neutral markers

A Bayesian phylogenetic analysis was performed using *Beast for species tree estimation[30]. Analysis was performed with a chain 300,000,000 long using a strict molecular clock. ITS was used as the reference locus, with a mutation rate for herbaceous plants of $4.13*10^{-9}$ subs/site/year [31], from which the rate of the other genes were estimated. According to Bayesian Information Criterion values found using jModeltest[32], the best substitution model for 17 out of the 26 genes was the HKY model[33]. We used a *Yules* species process for species tree estimation (this assumes that lineages split at a constant rate).

We explored population structure across the species complex by inferring the number of genetic clusters (K) using STRUCTURE v2.3.4[34]. This was undertaken at two levels: *i)* including all populations, and *ii)* by pairs (six parapatric and two allopatric). STRUCTURE was run using an admixture model and the correlated allele frequency model[35]. Parameters used for *i)* K=1-16, 20 iterations per K, burnin 100,000, MCMC 100,000; and for *ii)* K=1-9, 20 iterations per K, burnin 100,000, MCMC 100,000, as suggested by *Gilbert et al.* (2012) . To choose the most likely K value, both methods by Pritchard *et al*. (2000) and Evanno *et al*. (2005) were examined. Because both methods tend to overestimate K, and high K values did not add any additional clustering information compared to smaller Ks, the smallest K that captured the major structure in the data was chosen, ensuring that all summary statistics converged.

Estimates of reproductive isolation

Seeds from 30 individuals from four population pairs, two from the eastern coast (D01-H01 and D04-H05) and two from the southern coast (D32-H12 and D23-H21), were collected and stored in dry conditions at 4°C in the Ortiz-Barrientos Laboratory at The University of Queensland, Australia. Seeds were scarified (1mm trimmed at the micropyle side) and germinated on moist filter paper in petri dishes. Seeds were kept in dark and controlled conditions for three days to induce root elongation, and subsequently placed under the light for 7 days to induce vegetative growth. Seedlings were then transplanted into 0.25L pots filled with standard potting mix and transferred to a glasshouse with constant temperature (25°C) and 12h:12h light:dark cycle. After two months, flowering individuals were crossed by rubbing flower heads over a period of three to five days to create seed stocks for each population.

F1 hybrid seed set

15 families for each population were germinated and grown under glasshouse conditions (described above). Intra- and inter-population crosses were performed twice a day by gently rubbing flower heads (capitula): each flower head was crossed at least three times to saturate the number of fertilised florets. Crosses were blind with respect to research assistants but not with respect to MCM who supervised the experiments. We calculated seed set by estimating the proportion of fertilised (filled seeds)[14] seeds in flower heads. We divided the number of fertilised seeds in an inter-population cross by the average number of seeds produced in parental intra-population crosses; we subtracted this fraction

from one to calculate postmating prezygotic reproductive isolation in the system[17].

F1 viability at early stages of the life cycle

We tested for the viability of F1 hybrids between multiple inter-population crosses (see above) by counting the proportion of seeds that germinated in relation to the parental germination proportions. A total of 260 families equally distributed across all parental and hybrid cross types were germinated, placing five seeds of each family into moist filter paper in petri dishes (one family per petri dish). We placed petri dishes into trays (20 per tray) while randomising the position of all families. Germination conditions are the same as described above, except for tray position on the shelves, which in this case was switched daily. We did not scarify seeds to investigate the intrinsic ability of embryos to germinate. Data collection was blind, where cells in a tray were assigned a number that was cross-referenced with the original genotype and family ID at the end of the experiment. We calculated F1 inviability by estimating the proportion of germinated seeds in a cross, divided by the average number of seeds that germinated of the two parental populations.

Strength of reproductive isolation

Estimates of reproductive isolation were done for two reproductive isolating barriers in the following way: *Hybrid seed set (I)* in the glasshouse or whether the proportion of fertilised seeds in a flower head from an inter-population cross differed from an intra-population cross as $RI_{seed\ set}=1-(Pf_{inter}/Pf_{intra})$ were Pf stands for proportion fertilised. *Hybrid viability (I)*, or whether hybrid seedlings

germinated equally well as their parents in the glasshouse, as H_{hf}=1-(v_{F1}/$v_{parents}$), where v_{F1} is the average germination of F1 hybrids, and $v_{parents}$ is the average germination of the two parents.

Estimates of local and non-local field performance

To identify whether local adaptation was stronger among populations for the Headland versus the Dune ecotype, we reanalysed field transplant experiments that used seeds from three populations of each ecotype (n=150-180 seeds/population/transplant environment). Methods of the transplants into the two natural environments at Lennox Head (NSW), including the local population are described in detail in Walter et al. 2016, but briefly, we fixed each seed to a toothpick using non-drip supa glue. The toothpicks were then placed in a fully randomised grid under natural conditions in both the sand dunes and rocky headland at Lennox Head. We used six blocks to sample each environment. Shadecloth (50%) was suspended 15cm above the grids and seeds watered for three weeks to replicate ideal germination conditions. We recorded seedling emergence as a binary trait, we then tracked seedling survival and recorded whether plants reached 10 leaves (as a measure of seedling establishment).

To understand whether local populations performed better than non-local populations (from the same habitat), we first implemented a survival analysis within the R package 'coxme'[39] where survival was censored at day 320, and we included population as a fixed effect, and environmental block as a random effect. Next, to test whether local populations had a higher probability of reaching seedling establishment, we implemented a generalised linear mixed effects

model within the R package 'lme4'[40]. To do so, we included seedling establishment as a binary response variable, with population as a fixed effect, and environmental block as a random effect. In both analyses we wanted to test whether the two non-local populations performed more poorly than the local population. Thus, for each analysis we designated the local population as the intercept, testing each non-local population relative to the intercept.

References

1 Schluter, D. Evidence for Ecological Speciation and Its Alternative. *Science* **323**, 737-741, doi:10.1126/science.1160006 (2009).

2 Ostevik, K. L., Moyers, B. T., Owens, G. L. & Rieseberg, L. H. Parallel ecological speciation in plants? *International Journal of Ecology* **2012** (2012).

3 Boughman, J. W., Rundle, H. D. & Schluter, D. Parallel evolution of sexual isolation in sticklebacks. *Evolution* **59**, 361-373, doi:10.1111/j.0014-3820.2005.tb00995.x (2005).

4 Strecker, U., Hausdorf, B. & Wilkens, H. Parallel speciation in *Astyanax* cave fish (Teleostei) in Northern Mexico. *Molecular Phylogenetics and Evolution* **62**, 62-70 (2012).

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

6 Johannesson, K. *et al.* 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, doi:10.1098/rstb.2009.0256 (2010).

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

8 Vijverberg, K., Kuperus, P., Breeuwer, J. & Bachmann, K. Incipient adaptive radiation of New Zealand and Australian Microseris (Asteraceae): an amplified fragment length polymorphism (AFLP) study. *Journal of Evolutionary Biology* **13**, 997-1008 (2000).

9 Vijverberg, K., Mes, T. H. & Bachmann, K. Chloroplast DNA evidence for the evolution of *Microseris* (Asteraceae) in Australia and New Zealand after long-distance dispersal from western North America. *American Journal of Botany* **86**, 1448-1463 (1999).

10 Roda, F. *et al.* Genomic evidence for the parallel evolution of coastal forms in the Senecio lautus complex. *Molecular ecology* **22**, 2941-2952, doi:10.1111/mec.12311 (2013).

11 Thompson, I. Taxonomic studies of Australian *Senecio* (Asteraceae): 5. The *S. pinnatifolius/S. lautus* complex. *Muelleria* **21**, 23-76 (2005).

12 Richards, T. J., Walter, G. M., McGuigan, K. & Ortiz-Barrientos, D. Divergent natural selection drives the evolution of reproductive isolation in an Australian wildflower. *Evolution; international journal of organic evolution* **70**, 1993-2003, doi:10.1111/evo.12994 (2016).

13 Walter, G. M., Wilkinson, M. J., Aguirre, J. D., Blows, M. W. & Ortiz-Barrientos, D. Environmentally induced development costs underlie fitness tradeoffs. *Ecology* **99**, 1391-1401, doi:doi:10.1002/ecy.2234 (2018).

14 Melo, M. C., Grealy, A., Brittain, B., Walter, G. M. & Ortiz-Barrientos, D. Strong extrinsic reproductive isolation between parapatric populations of an Australian groundsell. *New Phytologist* **doi: 10.1111/nph.12779** (2014).

15 Richards, T. J. & Ortiz-Barrientos, D. Immigrant inviability produces a strong barrier to gene flow between parapatric ecotypes of Senecio lautus. *Evolution; international journal of organic evolution* **70**, 1239-1248, doi:10.1111/evo.12936 (2016).

16 Roda, F. *et al.* Convergence and divergence during the adaptation to similar environments by an Australian groundsel. *Evolution; international journal of organic evolution* **67**, 2515-2529, doi:10.1111/evo.12136 (2013).
17 Moonsamy, P. V. *et al.* 201-P Use of the Fluidigm® access Array™ system provides simplified amplicon library preparation in next generation sequencing for high throughput HLA genotyping. *Human Immunology* **72**, S142 (2011).
18 Schmieder, R., Lim, Y., Rohwer, F. & Edwards, R. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. *BMC Bioinformatics* **11**, 341 (2010).
19 Hird, S. M., Brumfield, R. T. & Carstens, B. C. PRGmatic: an efficient pipeline for collating genome‐enriched second‐generation sequencing data using a ‘provisional‐reference genome’. *Molecular Ecology Resources* **11**, 743-748 (2011).
20 Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. *Genome Research* **9**, 868-877 (1999).
21 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
22 Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
23 Koboldt, D. C. *et al.* VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**, 2283-2285, doi:10.1093/bioinformatics/btp373 (2009).
24 Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Research* **12**, 656-664 (2002).
25 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797 (2004).
26 Kimura, M. *The neutral theory of molecular evolution*. (Cambridge University Press, 1985).
27 Hudson, R. R., Kreitman, M. & Aguade, M. A Test of Neutral Molecular Evolution Based on Nucleotide Data. *Genetics* **116**, 153-159 (1987).
28 Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595 (1989).
29 Fu, Y.-X. & Li, W.-H. Statistical tests of neutrality of mutations. *Genetics* **133**, 693-709 (1993).
30 Heled, J. & Drummond, A. J. Bayesian inference of species trees from multilocus data. *Molecular biology and evolution* **27**, 570-580 (2010).
31 Kay, K., Whittall, J. & Hodges, S. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. *BMC Evolutionary Biology* **6**, 36 (2006).
32 Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818, doi:10.1093/bioinformatics/14.9.817 (1998).
33 Hasegawa, M., Kishino, H. & Yano, T.-a. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of molecular evolution* **22**, 160-174 (1985).

34 Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959 (2000).
35 Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587 (2003).
36 Gilbert, K. J. *et al.* Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program structure. *Molecular Ecology* **21**, 4925-4930, doi:10.1111/j.1365-294X.2012.05754.x (2012).
37 Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**, 2611-2620, doi:10.1111/j.1365-294X.2005.02553.x (2005).
38 Ramsey, J., Bradshaw Jr, H. & Schemske, D. Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* **57**, 1520-1534 (2003).
39 Huebner, M. & Therneau, T. Mixed Effects Cox Models For Gene Set Analysis In Lung Cancer. *Genet Epidemiol* **36**, 157-157 (2012).
40 Bates, D., Machler, M., Bolker, B. M. & Walker, S. C. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw* **67**, 1-48 (2015).