

# Current Biology

## Adaptive Evolution Is Common in Rapid Evolutionary Radiations

### Highlights

- Adaptive evolution is common across rapid evolutionary radiations
- The frequency of adaptive evolution correlates negatively with population size
- Deleterious substitutions are more common in evolutionary radiations
- Small population sizes may play a central role in rapid evolutionary radiations

### Authors

Bruno Nevado, Edgar L.Y. Wong,  
Owen G. Osborne, Dmitry A. Filatov

### Correspondence

[bruno.nevado@plants.ox.ac.uk](mailto:bruno.nevado@plants.ox.ac.uk)

### In Brief

The processes driving evolutionary radiations remain unclear. Nevado et al. show that adaptive evolution is more frequent in evolutionary radiations compared to slowly diversifying lineages and that the frequency of adaptive evolution correlates negatively with population size—supporting a role for small population size in rapid diversification.



# Adaptive Evolution Is Common in Rapid Evolutionary Radiations

Bruno Nevado,<sup>1,3,\*</sup> Edgar L.Y. Wong,<sup>1</sup> Owen G. Osborne,<sup>1,2</sup> and Dmitry A. Filatov<sup>1</sup>

<sup>1</sup>Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

<sup>2</sup>Present address: Molecular Ecology and Fisheries Genetics Laboratory, School of Natural Sciences, Bangor University, Bangor LL57 2UW, UK

<sup>3</sup>Lead Contact

\*Correspondence: [bruno.nevado@plants.ox.ac.uk](mailto:bruno.nevado@plants.ox.ac.uk)

<https://doi.org/10.1016/j.cub.2019.07.059>

## SUMMARY

One of the most long-standing and important mysteries in evolutionary biology is why biological diversity is so unevenly distributed across space and taxonomic lineages. Nowhere is this disparity more evident than in the multitude of rapid evolutionary radiations found on oceanic islands and mountain ranges across the globe [1–5]. The evolutionary processes driving these rapid diversification events remain unclear [6–8]. Recent genome-wide studies suggest that natural selection may be frequent during rapid evolutionary radiations, as inferred from work in cichlid fish [9], white-eye birds [10], new world lupins [11], and wild tomatoes [12]. However, whether frequent adaptive evolution is a general feature of rapid evolutionary radiations remains untested. Here we show that adaptive evolution is significantly more frequent in rapid evolutionary radiations compared to background levels in more slowly diversifying lineages. This result is consistent across a wide range of angiosperm lineages analyzed: 12 evolutionary radiations, which together comprise 1,377 described species, originating from some of the most biologically diverse systems on Earth. In addition, we find a significant negative correlation between population size and frequency of adaptive evolution in rapid evolutionary radiations. A possible explanation for this pattern is that more frequent adaptive evolution is at least partly driven by positive selection for advantageous mutations that compensate for the fixation of slightly deleterious mutations in smaller populations.

## RESULTS AND DISCUSSION

### Positive Selection Is More Frequent in Rapid Evolutionary Radiations

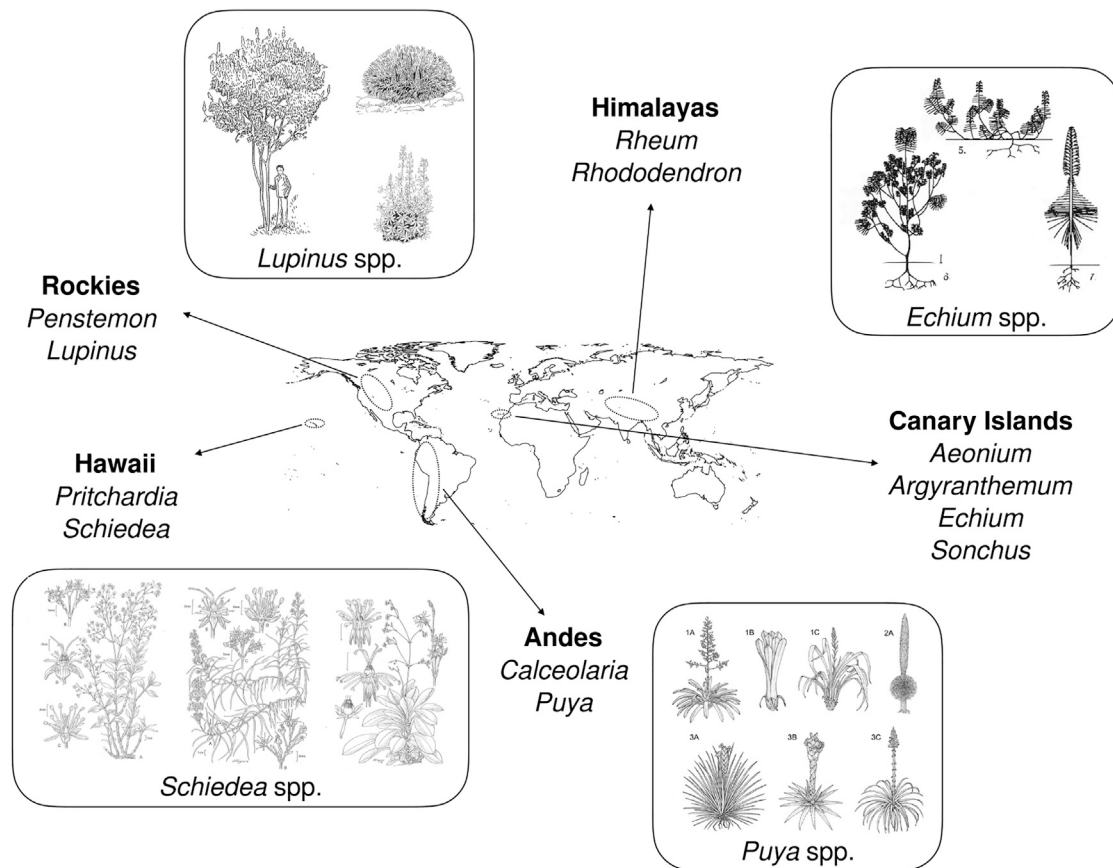
A central unanswered question in our understanding of rapid evolutionary radiations concerns the extent to which rapid diversification is driven by natural selection as opposed to extrinsic factors such as geographic isolation [7, 8]. In this context,

evolutionary radiations have often been classified as either adaptive or non-adaptive depending on the perceived role of natural selection in diversification [7, 8]. However, it is likely that both adaptive and non-adaptive processes play a role in every evolutionary radiation [13]. Furthermore, evolutionary radiations underpinned by variation in physiological or behavioral traits can more easily be perceived as non-adaptive, compared to those involving more conspicuous morphological traits, causing a bias in our understanding of the extent and distribution of adaptive radiations in nature. An alternative way to understand the role of natural selection in rapid diversification is to quantify the extent to which natural selection drives evolution in these systems and, in particular, ask whether adaptive evolution is more common in lineages that diversify rapidly compared to those that do not.

To test whether adaptive evolution is a common feature of evolutionary radiations, we estimate the genome-wide frequency of adaptive evolution occurring across 12 rapid evolutionary radiations of angiosperms (Figure 1; Table S1) and compare these estimates to background values obtained in previous studies [11, 17]. We focus on evolutionary radiations from some of the most species-rich biodiversity hotspots in the world (the Andes, the Rockies, the Himalayas, Hawaii, and the Canary Islands), including both oceanic island chains and large mountain ranges—geographic regions that have long fascinated evolutionary biologists due to the unparalleled abundance of rapidly diversifying lineages of many different taxonomic groups [1–3]. For each evolutionary radiation analyzed, we obtained fresh tissue for RNA extraction from multiple species from different Botanic Gardens throughout the UK (Table S1), with the exception of *Lupinus* and *Schiedea*, for which plants were grown from seed in greenhouses in Oxford. For lineages for which more than 8 species were available and produced high-quality RNA extracts, we randomly selected 8 species for RNA sequencing with the Illumina platform (average of 54 million reads per sample). For background lineages, the same number of species per lineage ( $n = 6–8$ ) was obtained from GenBank during our previous work [11].

To estimate the frequency of positive selection acting on coding sequences during diversification of each lineage analyzed, we applied methods based on the estimation of the nonsynonymous-to-synonymous substitution-rate ratio along phylogenies [18, 19] ( $\omega = dN/dS$ ) to genome-wide datasets obtained with RNA sequencing. We assembled transcriptomes *de novo* for each species and identified single-copy orthogroups [20]





**Figure 1. Geographic Origin of the Twelve Plant Evolutionary Radiations Analyzed in This Study**

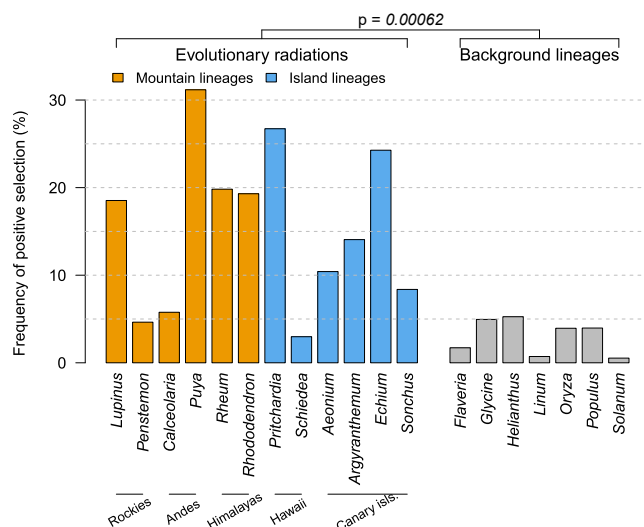
Insets showcase the phenotypic diversity of some of the genera studied. Line drawings of *Lupinus* kindly provided by Rosemary Wise; *Schiedea* adapted from [14]; *Puya* adapted from [15]; and *Echium* adapted from [16]. See also Table S1.

(i.e., groups of orthologous genes present as single copy in all species) within each lineage. Without prejudice to the role of gene duplications during rapid evolutionary radiations [9], in this work, we focus on single-copy genes to avoid the confounding effect of gene duplications on phylogenetic estimation and analysis of selection. On average, we analyzed 7,459 single-copy orthogroups containing at least 6 congeneric species within each lineage and tested 66,982 genes for adaptive evolution across 12 plant evolutionary radiations.

We found that the genome-wide frequency of positive selection is variable across evolutionary radiations, with the percentage of genes evolving under positive selection within each lineage ranging from 2.9% in *Schiedea* to 31% in *Puya* (Figure 2; Table S2). The percentage of genes evolving under positive selection was significantly higher (Welch's two-sample *t* test,  $p = 0.00062$ ) in the evolutionary radiations studied here compared to the more slowly diversifying background lineages analyzed previously [11] (Figure 2), and non-synonymous mutations evolving under positive selection (population-scaled selective coefficient,  $S > 1$ ) were significantly more frequent in evolutionary radiations compared to background lineages (Welch's two-sample *t* test,  $p = 0.0087$ ; Figure S1). There were no significant differences in the frequency of positive selection (mean = 14.4% versus 16.5%; Welch's two-sample *t* test,

$p = 0.71$ ) or the pairwise dN/dS (mean = 0.27 versus 0.23; Welch's two-sample *t* test,  $p = 0.32$ ) between evolutionary radiations from islands and those from mountain ranges.

Our analysis of adaptive evolution assumes a known phylogenetic relationship between species within each lineage, across which rates of synonymous and non-synonymous divergence are estimated. If the underlying phylogeny is incorrect, false positive results can occur [21, 22]. We find that phylogenetic error is an unlikely source for the pattern we observe for three reasons. First, phylogenetic reconstruction using maximum-likelihood and coalescent-based approaches returned identical and well-supported topologies in most cases, with only six species (in four plant lineages) having conflicting phylogenetic positions in the two approaches (see STAR Methods). Furthermore, the percentage of genes evolving under positive selection in evolutionary radiations compared to more slowly diversifying background lineages remains significantly different (Welch's two-sample *t* test,  $p = 0.0028$ ) after exclusion of these four lineages. Second, we excluded from our analyses genes that showed significant phylogenetic conflict with the overall species tree. Reassuringly, this approach resulted in exclusion of 68.5% of genes within *Aeonium* (and 0.2% to 26.8% in other lineages, Table S2), which reflects the polyploid events within this genus [23] and shows that our approach is able to detect and exclude



**Figure 2. Positive Selection Is More Frequent in Rapid Evolutionary Radiations**

Bar plot shows the percentage of genes evolving under positive selection detected within each plant lineage analyzed (evolutionary radiations vs background lineages, Welch's two-sample t-test,  $p = 0.00062$ ). See also Figures S1 and S2 and Table S2.

genes with conflicting phylogenetic histories. Third, across all lineages analyzed in this study, the percentage of genes evolving under positive selection correlated positively (two-sided Spearman's rank correlation test,  $\rho = 0.66$ ,  $p = 0.0028$ ) with the average pairwise dN/dS (Figure S2A), which does not depend on phylogenetic information and thus would not be affected by incomplete lineage sorting or hybridization.

A potential caveat of this result is that more rapid speciation should lead to shorter internal branches, which can cause increased variance in estimates of dN/dS and potentially lead to false inferences of positive selection. If this were the case, we would expect more frequent inference of positive selection on lineages that exhibit more frequent phylogenetic conflict and lower synonymous divergence (due to faster speciation and/or younger age). While the percentage of genes under positive selection correlated negatively with mean synonymous divergence across all lineages studied (two-sided Spearman's rank correlation test,  $\rho = -0.46$ ,  $p = 0.0477$ , Figure S2B), there was no significant correlation between these two variables within evolutionary radiations (two-sided Spearman's rank correlation test,  $\rho = -0.48$ ,  $p = 0.11$ ) nor background lineages (two-sided Spearman's rank correlation test,  $\rho = -0.32$ ,  $p = 0.49$ ) separately nor between the percentage of genes evolving under positive selection and the amount of phylogenetic incongruence across all lineages (two-sided Spearman's rank correlation test,  $\rho = 0.41$ ,  $p = 0.0834$ , Figure S2C). A linear regression model taking as independent variables lineage type (evolutionary radiation versus background lineage), mean synonymous divergence, and amount of phylogenetic conflict further confirmed that only the classification as an evolutionary radiation or a background lineage was a significant predictor of the percentage of genes evolving under positive selection in each lineage ( $p < 0.05$ ; overall model fit,  $R^2 = 0.324$ ). Thus, our finding that adaptive evolution is more frequent in evolutionary radiations

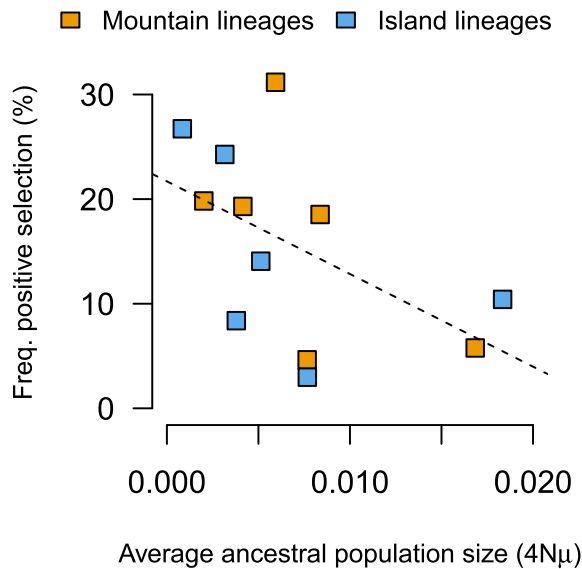
does not seem to be driven by biases inherent in our analysis but rather represents a true pattern observed across the plant lineages analyzed.

Our analysis is based on sampling of a large array of species comprising very different life history traits, generation times, fecundity, sizes or habits. The amount of phylogenetic conflict observed varied markedly among lineages (Table S2), suggesting that the most frequent speciation modes (e.g., founder effect versus parapatric speciation [24]) may differ among the evolutionary radiations analyzed. Thus, our finding of consistently more frequent adaptive evolution in evolutionary radiations compared to more slowly diversifying background lineages is generalizable to other plant lineages and perhaps to rapid evolutionary radiations in general. The unequivocal result of more frequent adaptive evolution, across such a wide range of angiosperm species, points to a general rule of evolutionary radiations: the pervasive role of adaptive evolution.

### What Determines the Rate of Adaptive Evolution during Rapid Diversification?

More frequent adaptive evolution during rapid diversification may be caused by adaptation to a diverse range of new ecological niches [1]. If this were the case, we would expect a positive correlation between the niche volume—defined as a hyper-volume in multi-dimensional ecological space [25]—occupied by an evolutionary radiation and the rate of diversification and frequency of adaptive evolution observed. This scenario also implies a positive correlation between population size and frequency of adaptive evolution, as population size determines the rate of emergence of advantageous mutations driving adaptation to novel niches [26], even though other sources of adaptive genetic variation—including from standing genetic variation or hybridization—may also be involved [27–29]. To test this hypothesis, we estimated niche volume, diversification rates, and ancestral population sizes for the 12 evolutionary radiations studied (Table S2). Here, we focus on the analyses within evolutionary radiations to understand why the frequency of adaptive evolution varies among the rapidly diversifying lineages. Furthermore, ecological modeling for widely distributed background lineages would be less reliable due to lack of clear limits on their geographic distributions. We note, however, that diversification rates are higher in evolutionary radiations compared to background lineages (by definition) and that population sizes of lineages in oceanic islands are typically smaller than their mainland counterparts [e.g., 30, 31], and the same is likely true of the plant lineages inhabiting young mountain ranges considered here. Our analyses revealed a positive but non-significant relationship between diversification rate and niche volume (two-sided Spearman's rank correlation test,  $\rho = 0.65$ ,  $p = 0.066$ ; Figure S3), and a negative non-significant relationship between niche volume and frequency of positive selection (two-sided Spearman's rank correlation test,  $\rho = -0.24$ ,  $p = 0.44$ ); while population size (averaged over all ancestral nodes of each phylogeny) and frequency of positive selection showed a significant negative correlation (two-sided Spearman correlation test,  $\rho = -0.58$ ,  $p = 0.048$ ; Figure 3).

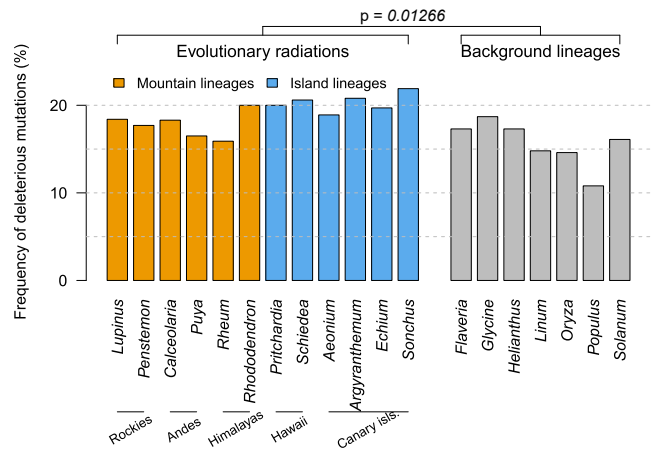
Our results thus do not support the hypothesis that more frequent adaptive evolution in evolutionary radiations is driven



**Figure 3. Adaptive Evolution Is More Common in Smaller Populations**

Across the twelve rapid evolutionary radiations studied, the percentage of genes evolving under positive selection shows a significant negative correlation with the average ancestral population size in each lineage (two-sided Spearman's rank correlation test,  $\rho = -0.58$ ,  $p = 0.048$ ). See also Figure S3 and Table S2.

primarily by adaptation to novel ecological niches. Instead, they suggest a central role for population size, with smaller population sizes driving more frequent adaptive evolution. While unexpected under simpler population genetics models [26], this result can be explained in light of the accumulation of slightly deleterious mutations in smaller populations, causing proteins to be further away from their fitness optimum and promoting more frequent compensatory mutations in the same or different genes [32, 33]. Indeed, using a homology-based functional prediction approach [34], we found that non-synonymous substitutions more frequently have deleterious effects in evolutionary radiations compared to background lineages (mean = 19.1% versus 15.7%; Welch's two-sample t test,  $p = 0.0127$ ; Figure 4) and that average ancestral population size and frequency of deleterious substitutions within evolutionary radiations show a negative but non-significant relationship (two-sided Spearman's rank correlation test,  $\rho = -0.20$ ,  $p = 0.53$ ). This has important implications for our understanding of the evolution of reproductive isolation during rapid evolutionary radiations. In fact, population sizes of species forming evolutionary radiations on oceanic islands and mountain ranges are typically small [30, 31], and speciation in these systems is often accompanied by recurrent bottlenecks followed by range expansions, which can further promote fixation of slightly deleterious mutations [35, 36]. Deleterious mutations and compensatory adaptive mutations can thus be quickly and independently fixed in different populations, causing reproductive isolation, because hybrid individuals would frequently carry deleterious mutations without the corresponding compensatory mutation.



**Figure 4. Non-Synonymous Substitutions More Frequently Have Deleterious Effects in Rapid Evolutionary Radiations**

Bar plot shows the percentage of non-synonymous substitutions that are putatively deleterious (Provean score  $< -2.5$ ) within each plant lineage analyzed (evolutionary radiations vs background lineages, Welch's two-sample t-test,  $p = 0.01266$ ).

## Conclusions

In this work, we investigated the evolutionary processes driving evolutionary radiations in plant lineages from two habitats where rapid species radiations are most common: oceanic islands (Hawaii, Canary Islands) and mountain ranges (Andes, Rockies, Himalayas). We demonstrate that, genome-wide, adaptive evolution is significantly more frequent in evolutionary radiations, compared to more slowly diversifying plant lineages. This generalizes previous results [9–12, 37] and suggests a central role for adaptive evolution during rapid diversification. In addition, we find a significant negative effect of population size on adaptive evolution in evolutionary radiations. This suggests that widespread adaptive evolution may be partly driven by demographic effects that cause fixation of deleterious mutations and associated compensatory adaptive mutations.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Taxonomic sampling
  - RNA extraction and sequencing
  - Pre-processing and transcriptome assembly
  - Orthology inference
  - Phylogenetic inference
  - Frequency of positive selection genome-wide
  - Distribution of selection coefficients
  - Functional impact of non-synonymous substitutions
  - Diversification rates
  - Ancestral population sizes
  - Niche volumes



- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cub.2019.07.059>.

## ACKNOWLEDGMENTS

This work was supported by grants from the Natural Environment Research Council (grant NE/K004352/1 to D.A.F.), the University of Oxford (JFF grant 161/044 to D.A.F.), and Edinburgh Genomics (Edinburgh Genomics poster prize at the UK Plant Evolution Symposium, 2014, to O.G.O.). We thank the staff at the Wellcome Trust Centre (Oxford) and Edinburgh Genomics (Edinburgh) for sequencing and computational support and acknowledge the use of the University of Oxford Advanced Research Computing facility in carrying out this work (<https://doi.org/10.5281/zenodo.22558>). We thank the following persons and institutions for providing plant material for this study: Simon Hiscock, Tom Price, and Ben Jones, Oxford Univ. BG; Catherine Kidner and Pete Brownless, RBG Edinburgh; Beverley Glover, Samuel Brockington, and Pete Atkinson, Cambridge Univ. BG; Sue Medway and Nick Bailey, Chelsea Physic Garden; Alex Papadopoulos, RBG Kew; Chris Kidd, Ventnor BG; Andrew Smith and Juan Beltran, Oxford Univ.; Stephen Weller and Ann Sakai, Univ. California-Irvine; Colin Hughes and Guy Atchison, Univ. Zurich; and the United States Department of Agriculture. We thank Rosemary Wise for line drawings of *Lupinus* used in Figure 1. We thank Varvara Fazalova for comments and discussions on multiple versions of this manuscript and an anonymous reviewer for suggesting the inclusion of the analysis of functional impact of mutations with PROVEAN.

## AUTHOR CONTRIBUTIONS

B.N. and D.A.F. conceived and designed the study. B.N., E.L.Y.W., and O.G.O. collected the data. B.N. analyzed the data and wrote the manuscript. All authors contributed to editing the manuscript and agree with its content.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 3, 2019

Revised: July 8, 2019

Accepted: July 19, 2019

Published: September 5, 2019

## REFERENCES

- Schluter, D. (2000). *The Ecology of Adaptive Radiation* (Oxford University Press).
- Stroud, J.T., and Losos, J.B. (2016). Ecological opportunity and adaptive radiation. *Annu. Rev. Ecol. Evol. Syst.* 47, 507–532.
- Simpson, G.G. (1953). *The Major Features of Evolution* (Columbia University Press).
- Linder, H.P. (2008). Plant species radiations: where, when, why? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3097–3105.
- Hughes, C.E., and Atchison, G.W. (2015). The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. *New Phytol.* 207, 275–282.
- Simões, M., Breitkreuz, L., Alvarado, M., Baca, S., Cooper, J.C., Heins, L., Herzog, K., and Lieberman, B.S. (2016). The evolving theory of evolutionary radiations. *Trends Ecol. Evol.* 31, 27–34.
- Rundell, R.J., and Price, T.D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends Ecol. Evol.* 24, 394–399.
- Givnish, T.J. (2015). Adaptive radiation versus ‘radiation’ and ‘explosive diversification’: why conceptual distinctions are fundamental to understanding evolution. *New Phytol.* 207, 297–303.
- Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E., et al. (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513, 375–381.
- Cornetti, L., Valente, L.M., Dunning, L.T., Quan, X., Black, R.A., Hébert, O., and Savolainen, V. (2015). The genome of the “Great Speciator” provides insights into bird diversification. *Genome Biol. Evol.* 7, 2680–2691.
- Nevado, B., Atchison, G.W., Hughes, C.E., and Filatov, D.A. (2016). Widespread adaptive evolution during repeated evolutionary radiations in New World lupins. *Nat. Commun.* 7, 12384.
- Pease, J.B., Haak, D.C., Hahn, M.W., and Moyle, L.C. (2016). Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol.* 14, e1002379.
- Donoghue, M.J., and Sanderson, M.J. (2015). Confluence, synnovation, and depauperons in plant diversification. *New Phytol.* 207, 260–274.
- Wagner, W.L., Weller, S.G., and Sakai, A. (2005). Monograph of *Schiedea* (Caryophyllaceae-Alsinoideae). *Syst. Bot. Monogr.* 72.
- Hornung-Leoni, C.T., and Sosa, V. (2008). Morphological phylogenetics of *Puya* subgenus *Puya* (Bromeliaceae). *Bot. J. Linn. Soc.* 156, 93–110.
- Böhle, U.-R., Hilger, H.H., and Martin, W.F. (1996). Island colonization and evolution of the insular woody habit in *Echium* L. (Boraginaceae). *Proc. Natl. Acad. Sci. USA* 93, 11740–11745.
- Gossmann, T.I., Song, B.-H., Windsor, A.J., Mitchell-Olds, T., Dixon, C.J., Kapralov, M.V., Filatov, D.A., and Eyre-Walker, A. (2010). Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Mol. Biol. Evol.* 27, 1822–1832.
- Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A.-M.K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449.
- Nielsen, R., and Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929–936.
- Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 157.
- Mendes, F.K., and Hahn, M.W. (2016). Gene tree discordance causes apparent substitution rate variation. *Syst. Biol.* 65, 711–721.
- Pie, M.R. (2006). The influence of phylogenetic uncertainty on the detection of positive Darwinian selection. *Mol. Biol. Evol.* 23, 2274–2278.
- Mort, M.E., Soltis, D.E., Soltis, P.S., Francisco-Ortega, J., and Santos-Guerra, A. (2001). Phylogenetic relationships and evolution of Crassulaceae inferred from matK sequence data. *Am. J. Bot.* 88, 76–91.
- Coyne, J.A., and Orr, H.A. (2004). *Speciation* (Sinauer).
- Hutchinson, G.E. (1957). Population studies - animal ecology and demography - concluding remarks. *Cold Spring Harb. Sym.* 22, 415–427.
- Lanfear, R., Kokko, H., and Eyre-Walker, A. (2014). Population size and the rate of evolution. *Trends Ecol. Evol.* 29, 33–41.
- Anderson, E. (1949). *Introgressive Hybridization* (J. Wiley).
- Arnold, M.L. (2015). *Divergence with Genetic Exchange* (Oxford University Press).
- Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends Ecol. Evol.* 19, 198–207.
- Frankham, R. (1997). Do island populations have less genetic variation than mainland populations? *Heredity* 78, 311–327.
- James, J.E., Lanfear, R., and Eyre-Walker, A. (2016). Molecular evolutionary consequences of island colonization. *Genome Biol. Evol.* 8, 1876–1888.
- Cherry, J.L. (1998). Should we expect substitution rate to depend on population size? *Genetics* 150, 911–919.

33. Hartl, D.L., and Taubes, C.H. (1996). Compensatory nearly neutral mutations: selection without adaptation. *J. Theor. Biol.* **182**, 303–309.
34. Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., and Chan, A.P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* **7**, e46688.
35. Charlesworth, J., and Eyre-Walker, A. (2007). The other side of the nearly neutral theory, evidence of slightly advantageous back-mutations. *Proc. Natl. Acad. Sci. USA* **104**, 16992–16997.
36. Peischl, S., Dupanloup, I., Kirkpatrick, M., and Excoffier, L. (2013). On the accumulation of deleterious mutations during range expansions. *Mol. Ecol.* **22**, 5972–5982.
37. Naumenko, S.A., Logacheva, M.D., Popova, N.V., Klepikova, A.V., Penin, A.A., Bazykin, G.A., Etingova, A.E., Mugue, N.S., Kondrashov, A.S., and Yampolsky, L.Y. (2017). Transcriptome-based phylogeny of endemic Lake Baikal amphipod species flock: fast speciation accompanied by frequent episodes of positive selection. *Mol. Ecol.* **26**, 536–553.
38. Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
39. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512.
40. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.
41. Löytynoja, A., and Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl. Acad. Sci. USA* **102**, 10557–10562.
42. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.
43. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19** (Suppl 6), 153.
44. Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704.
45. Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**, 1246–1247.
46. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.
47. Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556.
48. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
49. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659.
50. Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Curr. Zool.* **61**, 854–865.
51. Gehrig, H.H., Winter, K., Cushman, J., Borland, A., and Taybi, T. (2000). An improved RNA isolation method for succulent plant species rich in polyphenols and polysaccharides. *Plant Mol. Biol. Report.* **18**, 369–376.
52. Kosiol, C., Holmes, I., and Goldman, N. (2007). An empirical codon model for protein sequence evolution. *Mol. Biol. Evol.* **24**, 1464–1479.
53. Yang, Y., and Smith, S.A. (2014). Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Mol. Biol. Evol.* **31**, 3081–3092.
54. Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web servers. *Syst. Biol.* **57**, 758–771.
55. Shimodaira, H., and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**, 1114.
56. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300.
57. Nielsen, R., and Yang, Z. (2003). Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. *Mol. Biol. Evol.* **20**, 1231–1239.
58. Magallón, S., and Sanderson, M.J. (2001). Absolute diversification rates in angiosperm clades. *Evolution* **55**, 1762–1780.
59. Rannala, B., and Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**, 1645–1656.
60. Zhang, C., Zhang, D.-X., Zhu, T., and Yang, Z. (2011). Evaluation of a bayesian coalescent method of species delimitation. *Syst. Biol.* **60**, 747–761.
61. Burgess, R., and Yang, Z. (2008). Estimation of hominoid ancestral population sizes under bayesian coalescent models incorporating mutation rate variation and sequencing errors. *Mol. Biol. Evol.* **25**, 1979–1994.
62. Hijmans, R., van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J.A., Lamigueiro, O.P., Bevan, A., Racine, E.B., Shortridge, A., et al. (2017). Raster: geographic data analysis and modeling. 2.6 Edition.
63. Blonder, B., Lamanna, C., Violle, C., and Enquist, B.J. (2014). The *n*-dimensional hypervolume. *Glob. Ecol. Biogeogr.* **23**, 595–609.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Biological Samples</b>		
Fresh tissues from multiple plant species were obtained from Botanic Gardens as detailed in <a href="#">Table S1</a> .	Multiple	Multiple
Seeds from multiple plant species were obtained from multiple sources as detailed in <a href="#">Table S1</a> .	Multiple	Multiple
<b>Deposited Data</b>		
Newly sequenced data was deposited in GenBank's Short Read Archive, BioProject ID PRJNA491458	This study	PRJNA491458
Sequence data re-used in this study and available from previous work was obtained from GenBank as detailed in <a href="#">Table S1</a> .	GenBank SRA	Multiple
<b>Software and Algorithms</b>		
fastqc	<a href="#">[38]</a>	V0.11
trim_galore	<a href="http://www.bioinformatics.babraham.ac.uk">www.bioinformatics.babraham.ac.uk</a>	V0.4
trinity	<a href="#">[39]</a>	V2.5
trimmomatic	<a href="#">[40]</a>	N/A
Trandecoder	<a href="#">[39]</a>	V3.0
orthofinder	<a href="#">[20]</a>	V1.1
prank	<a href="#">[41]</a>	v140110
raxml	<a href="#">[42]</a>	V8.0
Astral	<a href="#">[43]</a>	V5.6
phym1	<a href="#">[44]</a>	V3.1
Consel	<a href="#">[45]</a>	V1.2
paml	<a href="#">[46, 47]</a>	V4.8
provean	<a href="#">[34]</a>	V1.1
psiblast	<a href="#">[48]</a>	V2.2
cd-hit	<a href="#">[49]</a>	V4.8
bpp	<a href="#">[50]</a>	V4.0
Statistical analysis performed in R	<a href="http://www.R-project.org">www.R-project.org</a>	V3.4.3
Custom software developed for this project available from <a href="https://github.com/brunonevado/trimming">https://github.com/brunonevado/trimming</a>	This study	N/A

### LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Bruno Nevado ([bruno.nevado@plants.ox.ac.uk](mailto:bruno.nevado@plants.ox.ac.uk)). This study did not generate new unique reagents.

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

For analysis of species from evolutionary radiations, fresh plant material was collected from different Botanic Gardens throughout the UK (detailed in [Table S1](#)) from the following species: *Aeonium aureum*, *Aeonium decorum*, *Aeonium gomerense*, *Aeonium holochrysum*, *Aeonium lindleyi*, *Aeonium simsii*, *Aeonium urbicum*, *Aeonium virgineum*, *Argyranthemum foeniculaceum*, *Argyranthemum frutescens*, *Argyranthemum haovarytheum*, *Argyranthemum maderense*, *Argyranthemum sventenii*, *Argyranthemum tenerifae*, *Calceolaria cavanillesii*, *Calceolaria corymbosa* ssp. *Montana*, *Calceolaria fothergillii*, *Calceolaria integrifolia*, *Calceolaria pinifolia*, *Calceolaria thyrsiflora*, *Calceolaria uniflora*, *Echium acanthocarpum*, *Echium candicans*, *Echium giganteum*, *Echium hierrense*, *Echium leucophaeum*, *Echium simplex*, *Echium webbii*, *Echium wildpretii*, *Penstemon barbatus* subsp. *Barbatus*, *Penstemon cardwellii*, *Penstemon fruticosus* subsp. *scouleri*, *Penstemon hartwegii*, *Penstemon kunthii*, *Penstemon pinifolius*, *Penstemon procerus*,



*Penstemon virens*, *Pritchardia affinis*, *Pritchardia elliptica*, *Pritchardia glabrata*, *Pritchardia hillebrandii*, *Pritchardia lowreyana*, *Pritchardia schattaueri*, *Puya alpestris* ssp *alpestris*, *Puya berteroniana*, *Puya chilensis*, *Puya coerulea*, *Puya ferruginea*, *Puya mirabilis*, *Puya raimondii*, *Puya spathacea*, *Rheum acuminatum*, *Rheum alexandrae*, *Rheum forestii*, *Rheum franzenbachii*, *Rheum officinale*, *Rheum palmatum*, *Rhododendron barbatum*, *Rhododendron cerasinum*, *Rhododendron decorum*, *Rhododendron falconeri*, *Rhododendron lowndesii*, *Rhododendron macabeum*, *Rhododendron williamsianum*, *Sonchus acaulis*, *Sonchus arboreus*, *Sonchus bommulleri*, *Sonchus canariensis*, *Sonchus hierrensis*, *Sonchus pinnatus*. Additionally, fresh plant material was obtained from plants grown from seed in the greenhouse in the Department of Plant Sciences (Oxford) for the following species: *Lupinus albicaulis*, *Lupinus albifrons*, *Lupinus arboreus*, *Lupinus formosus*, *Lupinus latifolius*, *Lupinus nevadensis*, *Lupinus perennis*, *Lupinus rivularis*, *Schiedea globosa*, *Schiedea hookeri*, *Schiedea kaalae*, *Schiedea kealiae*, *Schiedea ligustrina*, *Schiedea nuttallii*, *Schiedea pentandra*, *Schiedea trinervis*.

For analysis of background lineages, Illumina sequence data was downloaded from GenBank SRA database for the following species: *Flaveria anomala*, *Flaveria bidentis*, *Flaveria brownii*, *Flaveria chlorifolia*, *Flaveria pubescens*, *Flaveria ramosissima*, *Flaveria robusta*, *Flaveria trinervia*, *Glycine canescens*, *Glycine clandestina*, *Glycine max*, *Glycine soja*, *Glycine syndetika*, *Glycine tomentella*, *Helianthus annuus*, *Helianthus argophyllus*, *Helianthus bolanderi*, *Helianthus debilis*, *Helianthus exilis*, *Helianthus petiolaris*, *Helianthus praecox*, *Linum bienne*, *Linum grandiflorum*, *Linum hirsutum*, *Linum lewisii*, *Linum macraei*, *Linum perenne*, *Linum tenuifolium*, *Linum usitatissimum*, *Oryza barthii*, *Oryza glumipatula*, *Oryza granulata*, *Oryza meyeriana*, *Oryza nivara*, *Oryza officinalis*, *Oryza rufipogon*, *Oryza sativa*, *Populus euphratica*, *Populus pruinosa*, *Populus tomentosa*, *Populus tremula*, *Populus tremuloides*, *Populus trichocarpa*, *Solanum berthaultii*, *Solanum dulcamara*, *Solanum habrochaites*, *Solanum incanum*, *Solanum melongena*, *Solanum pennellii*, *Solanum phureja*, *Solanum torvum*.

## METHOD DETAILS

### Taxonomic sampling

We collected fresh leaves from 6 to 8 species from each focal plant lineage, most of which were obtained from live plants sampled in Botanic Gardens throughout the UK (Table S1). Material collected was immediately frozen in dry ice or liquid nitrogen, taken to the lab and kept at  $-80^{\circ}\text{C}$  until extraction.

### RNA extraction and sequencing

We extracted total RNA from each sample using the QIAGEN RNeasy Plant Mini kit, following the manufacturer's instructions and applying the optional DNase digestions step. For some species this extraction protocol performed poorly, thus we used a modified CTAB extraction protocol followed by column cleaning from RNeasy Plant Mini Kit (QIAGEN) as follows: 1) 600ul CTAB buffer (with 1%  $\beta$ -mercaptoethanol) was added to the samples and mixed; 2) 600ul of chloroform-isoamyl alcohol (SEVAG) was added, mixed and centrifuged at maximum speed for five minutes; 3) the supernatant was transferred to a new tube and the steps 2-3 were repeated; 4) the supernatant was transferred to a new tube and equal amount of isopropanol was added, mixed and centrifuged at maximum speed for 15 min; 5) the supernatant was removed and the pellet was washed with 75% ethanol; 6) excess ethanol was removed and 75ul water was added to dissolve the pellet for 15 min at  $65^{\circ}\text{C}$ ; 7) 340ul RTL buffer (from the RNeasy Plant Mini Kit) and 200ul 100% ethanol were added; 8) manufacturer's protocol from step 5 onward was followed with additional DNase treatment. For samples of *Schiedea* and *Puya*, we used the Trizol extraction protocol [51] followed by DNase treatment. For all samples, RNA integrity was assessed with the Agilent 2100 Bioanalyzer. Paired-end Illumina libraries were prepared by the WTCHG Oxford genomics or the Edinburgh Genomics facilities. Samples were multiplexed and sequenced on the Illumina HiSeq platform with an average of 8 samples per lane.

### Pre-processing and transcriptome assembly

We quality-checked short read sequence data using FASTQC v0.11 [38] and used TRIM\_GALORE v 0.4 (available from [www.bioinformatics.babraham.ac.uk](http://www.bioinformatics.babraham.ac.uk)) to remove Illumina adaptors. We assembled *de novo* transcriptomes for each species using TRINITY v 2.5 [39] with default settings and included an initial trimming step to remove low quality bases with TRIMMOMATIC [40]. We identified putative coding sequences within the longest isoform of each gene using TRANDECODER v 3.0 [39] with default settings.

### Orthology inference

To identify orthologous genes within each evolutionary radiation we used ORTHOFINDER v 1.1 [20], retaining for subsequent analyses only single-copy orthologous genes sequenced in at least 6 species. For background lineages we included in our analyses only the single-copy orthologous genes identified in previous work [11]. We initially aligned the coding sequences of each orthologous gene from multiple species using PRANK v 140110 [41] with the codon substitution matrix method [52]. Manual inspection revealed that these alignments often contained alignment errors or included paralogous regions. To deal with this, we applied a series of alignment cleaning and trimming steps using stringent criteria. First, we excluded positions with more than 50% missing data and sequences shorter than 10 bp, and obtained a phylogenetic tree for each gene using RAXML v 8.0 [42] with the GTRCAT model. We calculated the average tip length for each species and the average internal edge length (both standardized by total tree length) across all orthologous gene trees obtained for each plant lineage. We then recursively trimmed tips and internal edges that were too long – for each species we trimmed tips that had a relative length above 99% of the distribution seen across all trees; and for each resulting tree we trimmed edges with relative length above 99% seen across all genes, in each case keeping the resulting subtree with most

sequences. Visual inspection revealed that paralogous errors were removed with these steps, however alignment errors remained in some instances, particularly at the end of sequences or in short sequences between gaps. To address the former issue, we trimmed the beginning and end of sequences that showed high divergence – we calculated average divergence to all other sequences in the orthogroup and, moving in windows of 9 bp, trimmed the beginning and end of sequences when average divergence on the window was above 20x the average across the sequence. To address the later issue, we trimmed stretches of sequences with less than 10 contiguous nucleotides and flanked by more than 8 gaps on each side. The resulting trimmed alignments were then realigned using PRANK with the codon model. Genes with fewer than four species remaining after these cleaning and trimming steps were excluded from subsequent analysis.

The cleaning steps described above were based on published pipelines for inference of orthologous genes using RNaseq data of non-model organism [53], but used a faster orthology inference method (with ORTHOFINDER) and clade-specific thresholds for identifying paralogs and poorly aligned regions. This analysis was implemented in R and Perl scripts (available from <https://github.com/brunonevado/trimming>). Manual inspection of original and final alignments revealed that paralogy and alignment errors were thoroughly removed, although at the cost of removing some regions that did not show any obvious errors – for instance, most sequences that contained variable positions at the end or start of sequences were trimmed. We see no reason to expect this to bias the results of subsequent analyses, whereas failing to remove alignment errors is more likely to lead to spurious results.

### Phylogenetic inference

To obtain a phylogenetic tree for each plant lineage, we used a Maximum-Likelihood supermatrix approach – we concatenated all orthologous genes, removed positions with more than 50% missing data, and used RAXML to obtain a phylogenetic tree using the GTRCAT nucleotide substitution model, with node support estimated with 100 rapid bootstraps [54]. To assess the impact of incomplete lineage sorting on phylogenetic inference, for each lineage we used an alternative coalescent-based approach implemented in ASTRAL V 5.6 [43]. To estimate a species tree with this approach, we first estimated gene trees for each orthologous gene with RAXML (with the same settings as above), and then used these gene trees to estimate a supertree in ASTRAL. Maximum-Likelihood and coalescent-based approaches returned identical topologies for most lineages, with only six species showing conflicting phylogenetic positions in the two approaches (Figure S4).

The resulting unrooted Maximum-Likelihood phylogenetic tree (hereafter the species tree) was used for subsequent analysis of selection, but for estimation of ancestral population sizes a rooted tree was required. For each evolutionary radiation, we downloaded proteomes from related species (Table S3) from uniprot (<https://www.uniprot.org/443/>) and from the 1000 plants project (<https://sites.google.com/a/ualberta.ca/onekp/>). Single copy orthologous genes between ingroup and outgroup were obtained with ORTHOFINDER and their protein sequences aligned with PRANK. Resulting protein alignments were concatenated and a phylogenetic tree obtained with RAXML using the PROTGAMMAAUTO option, constraining the phylogenetic relationships within the ingroup to the unrooted phylogeny obtained previously. The resulting phylogeny was used to infer the most basal split within the species tree, such that the ingroup phylogeny could be rooted for subsequent analysis.

To infer the extent of phylogenetic conflict among genes within each plant lineage, we performed the Shimodaira-Hasegawa test (SH-test) [55]. For each gene, we compared the fit of an unconstrained gene tree to the species tree obtained with all genes. We used PHYLML V 3.1 [44] with the GTR+ $\Gamma$  model to optimize the likelihood of the unconstrained gene tree and the species trees, and CONSEL V 1.2 [45] to perform the SH-test. Genes that significantly preferred the gene tree over the species tree (FDR-corrected  $p < 0.05$ ) were excluded from subsequent analyses of selection.

### Frequency of positive selection genome-wide

To estimate the frequency of positive selection acting on coding sequences during diversification in each plant lineage we used methods based on the estimation of the nonsynonymous to synonymous substitution rate ratio ( $\omega = dN/dS$ ) along phylogenies [18, 19]. We used the sites models implemented in the program CODEML from the package PAML V 4.8 [46, 47], and fitted two alternative models: model M7, which allows only for sites evolving under neutrality and purifying selection; and model M8 which additionally allows a class of sites evolving under positive selection. We compared the fit of the two models with a Likelihood Ratio Test (LRT) with 2 degrees of freedom, and corrected resulting significance values with the False Discovery Rate (FDR) method [56].

We performed this analysis of selection for each orthologous gene of each plant lineage, after excluding codon positions with more than 50% missing data, and sequences with more than 50% missing codons. We also excluded genes with fewer than six species or 99 codons, genes without any variation, genes with too high synonymous substitution rates ( $dS > 2$ ) that could indicate remaining alignment errors, and genes that showed significant phylogenetic conflict (SH-test FDR correct  $p < 0.05$ ). For all tests, equilibrium codon frequencies were estimated from the average nucleotide frequencies at the three codon positions (codonFreq = 2), and positions with missing data were included (cleandata = 0).

In addition to the phylogeny-based analysis of the genome-wide frequency of positive selection, we estimated  $\omega$  between each pair of species within each plant lineage. To do so we concatenated the coding sequences of all genes tested above, and used the Maximum-Likelihood pairwise estimation method in CODEML (runmode = -2). The pairwise approach does not allow distinguishing positive selection from relaxation of purifying selection, and assumes a single  $\omega$  value for all sites, but provides an estimate of  $\omega$  that is not dependent on the assumed phylogeny.

### Distribution of selection coefficients

To estimate the distribution of selective coefficients ( $S$ ) of non-synonymous mutations in each plant lineage, we used the inverse mapping from the distribution of  $\omega$  among sites [57]. For each plant lineage, the distribution of  $\omega$  was calculated by concatenating the coding sequences of all genes used in the analysis of the genome-wide frequency of positive selection (above) and running model M5 in CODEML (model = 0; NSsites = 5). This model assumes that  $\omega$  is gamma-distributed among codon sites with parameters  $\alpha$  and  $\beta$  [18]. The distribution of selective coefficients can then be calculated from Equation 4 of Nielsen and Yang [57]:

$$f(S) = \frac{(\beta e^S S / h(S))^\alpha e^{-\beta e^S S / h(S)} (h(S) - S)}{S h(S) \Gamma(\alpha)}, \quad -\infty < S < \infty,$$

where  $h(S) = (e^S - 1)$ . This inverse mapping from  $\omega$  to  $S$  makes some simplifying assumptions: population sizes are constant among lineages; all non-synonymous mutations on the same site have the same selective coefficient; there is no interference in the fixation process of multiple mutations; and there are never more than 2 alleles segregating on the same site [57].

### Functional impact of non-synonymous substitutions

To infer the functional impact of non-synonymous substitutions we used the algorithm implemented in PROVEAN v 1.1 [34] using PSIBLAST v 2.2 [48] and CD-HIT v 4.8 [49]. For each plant lineage we randomly selected 250 genes from the set of genes tested for selection for this analysis (except for *Populus*, for which all 126 genes available were used, Table S2). For each gene, we generated a consensus sequence, identified non-synonymous substitutions and predicted their effect using PROVEAN by blasting against the NCBI non-redundant protein database. We classified substitutions as putatively deleterious if their Provean score was below  $-2.5$ , and calculated the percentage of substitutions that were putatively deleterious for each lineage.

### Diversification rates

To test whether the frequency of adaptive evolution correlates with diversification rate, for each evolutionary radiation analyzed we estimated diversification rate ( $r$ ) from the number of species ( $n$ ) and expected crown age ( $t$ ) as

$$r = \frac{\log(n) - \log(2)}{t}$$

(Equation 4 in Magallon and Sanderson [58]). Number of species and crown age of each lineages were obtained from previous studies (Table S2).

### Ancestral population sizes

To estimate ancestral population sizes along the phylogenies of each evolutionary radiation, we used BPP v 4.0 [50]. This program applies the multispecies coalescent model in a Bayesian framework to estimate ancestral population sizes and times of species split, while accounting for incomplete lineage sorting and low information content of each individual gene. We used the analysis A00, which estimates the parameters of the multispecies coalescent (population sizes and times of divergence) along a fixed topology [59].

The analysis under the multispecies coalescent assumes free recombination between loci, no recombination within loci, neutral evolution of all loci, and no gene flow between species [59]. To accommodate the assumptions of neutral evolution and no recombination within loci we excluded genes evolving under positive selection (as estimated previously), and used only the first 300 nucleotides of each gene. We cannot rule out gene flow between species, but this should result in predictable biases (overestimation of population sizes, underestimation of divergence times) and simulations suggest a low amount of migration should not result in strong biases [60]. Preliminary runs showed that analysis of all genes within each plant lineage was not computationally feasible. We thus performed three runs for each analysis, in each case randomly sampling 1,000 loci from each dataset.

For each analysis performed with BPP we used weak priors on both population sizes and divergence times using an inverse gamma distribution with mean 0.002 (population sizes) and 0.001 (divergence times). We allowed for variable mutation rate between loci – modeled with the random-rates model of [61] using a Dirichlet distribution (prior  $\alpha = 2$ ) – and included positions with ambiguity data (cleandata = 0). Each run consisted of 10,000 burnin steps, followed by 100,000 samples taken every 10<sup>th</sup> generation. For each analysis, estimates of ancestral population sizes were compared and summarized across runs using tracer (available from <http://tree.bio.ed.ac.uk/software/tracer/>).

Within each plant lineage the estimates of ancestral population sizes were somewhat variable between independent runs, which is expected given that each run used a random subset of 1,000 loci. Parameters with estimates that were particularly variable across runs (i.e., with effective sample size < 100 when summarized across all runs), were excluded from subsequent analysis, and average estimates for other parameters were obtained across runs.

### Niche volumes

To estimate the niche volume occupied by each evolutionary radiation, we used occurrence data from the Global Biodiversity Information Facility (GBIF, <https://www.gbif.org/>), and 34 environmental variables including elevation data, soil quality, solar irradiation and bioclimatic variables (Table S4).

For the occurrence data, for each evolutionary radiation we included all congeneric occurrences and restricted the geographic origin of records to the Andean Mountains, the Rockies, the Himalayas, the Canary Islands or the Hawaii archipelago, using map-drawn polygons. In addition, we excluded records without exact geographic coordinates (ignoring records flagged with `COORDINATE_ROUNDED`, or with coordinates with fewer than 3 decimal places) and, for lineages from mountainous regions, records with elevation below 2000 m (or without elevation information). The number of records obtained for each lineage was highly variable, which can confound the analysis. We thus randomly selected the same number of records ( $n = 70$ ) from each lineage for subsequent analysis.

All environmental variables were resampled to the same resolution with the R function `RESAMPLE` from the `RASTER` package [62], and cell grids without occurrence data were excluded from subsequent analysis. To reduce dimensionality, we performed a Principal Component Analysis (PCA) with the R function `PRINCOMP` from the `STATS` package, using the correlation matrix to standardize the variables. We then quantified niche volume for each plant evolutionary radiation from the first 3 principal component axes within environmental space using the R function `HYPERVOLUME_GAUSSIAN` from the `HYPERVOLUME` package [63].

## QUANTIFICATION AND STATISTICAL ANALYSIS

All statistics were calculated in R v 3.4.3. *t* tests estimated variance separately for each group, used Welch's modification to the degrees of freedom and were two-sided. Tests for association between variables used Spearman's *rho* and were two-sided.

## DATA AND CODE AVAILABILITY

Raw sequence data generated for this study is available from GenBank's Short Read Archive, BioProject ID PRJNA491458. Custom software developed for this project is available from [https://github.com/brunonevado/trimming\\_](https://github.com/brunonevado/trimming_).