Widespread adaptive evolution during repeated evolutionary radiations in New World lupins

Bruno Nevada¹, Guy W. Atchison², Colin E. Hughes² & Dmitry A. Filatov¹

The evolutionary processes that drive rapid species diversification are poorly understood. In particular, it is unclear whether Darwinian adaptation or non-adaptive processes are the primary drivers of explosive species diversifications. Here we show that repeated rapid radiations within New World lupins (Lupinus, Leguminosae) were underpinned by a major increase in the frequency of adaptation acting on coding and regulatory changes genome-wide. This contrasts with far less frequent adaptation in genomes of slowly diversifying lupins and all other plant genera analysed. Furthermore, widespread shifts in optimal gene expression coincided with shifts to high rates of diversification and evolution of perenniality, a putative key adaptation trait thought to have triggered the evolutionary radiations in New World lupins. Our results reconcile long-standing debate about the relative importance of protein-coding and regulatory evolution, and represent the first unambiguous evidence for the rapid onset of lineage- and genome-wide accelerated Darwinian evolution during rapid species diversification.
Rapid radiations of species are fundamental and ubiquitous components of evolutionary diversification across the Tree of Life in diverse organismal groups and geographical and ecological settings\(^1\)–\(^3\), and gave rise to many major groups of organisms such as flowering plants\(^1\). Recent studies are revealing widespread and repeated disparities in rates of species and trait diversification among lineages and through time, documenting an ever-growing number of rapidly evolving and often exceptionally species-rich clades, and shedding light on the extrinsic circumstances and intrinsic traits potentially triggering radiations\(^4\)–\(^7\). However, the mechanisms and processes, especially the genomic basis underlying why and how rapid species diversification happens, remain poorly understood.

Rapid diversification is thought to result from both adaptive (for example, ecological speciation) and non-adaptive (for example, geographic isolation) processes\(^8\)–\(^28\). When adaptive processes dominate, these lineages are called ‘adaptive radiations’, while the term ‘non-adaptive radiation’ has been put forth for rapidly diversifying lineages dominated by non-adaptive processes\(^8\)–\(^11\). While most radiations probably involve both adaptive and non-adaptive processes, very few studies have investigated the relative frequency of neutral and adaptive processes in rapidly versus slowly radiating groups\(^1\)–\(^2\),\(^14\) and the extent to which adaptive changes have wide-ranging effects throughout the genome remains unclear\(^15\). Instead, most work has focused on rates of species diversification and morphological evolution, and their possible correlations with traits, geological events and ecological shifts\(^7\),\(^16\). However, rates of diversification tell us little about the underlying evolutionary processes and the role of adaptation\(^8\),\(^11\), while the rate of morphological evolution is not informative about natural selection acting on non-morphological (for example, physiological or behavioural) traits, which may underpin adaptive radiation but remain undetected\(^17\). Genome-wide analysis of natural selection at the level of gene sequences and their expression levels, as pursued here, represents a much more direct approach to quantify the role of adaptation in rapid species diversifications.

High rates of DNA evolution, accelerated coding-sequence evolution, accumulation of gene duplications during speciation and sharing of genetic variation between species through ancestral polymorphisms or natural hybridization have been suggested as key genomic features underpinning evolutionary radiations\(^1\),\(^2\),\(^14\),\(^18\),\(^19\). However, the scarcity of well-sampled comparative studies of genome-wide selective pressures in rapidly and slowly radiating groups of species means that how rapid diversification unfolds at the genetic level remains unclear. In particular, it is unclear whether rapid speciation and ecological and morphological evolution are driven by changes in many genes throughout the genome, or in just a few key regulatory genes. Furthermore, whether changes in protein-coding sequences or expression levels of genes play the predominant role in adaptation is actively debated and remains unresolved\(^20\),\(^21\).

The genus *Lupinus* (Leguminosae) provides an ideal model system for understanding the genetic basis of rapid species diversification. It is a species-rich clade of \(~280\) species\(^22\),\(^23\), with a hypothesized ancestral area in the Old World, and two dispersals to the New World within the last \(10–12\) Myr (refs 23,24). The majority of species diversity occurs in the Americas, with only \(13\) species in Europe and North Africa. Within the New World, *Lupinus* spp. exhibit very diverse growth forms (ephemeral annuals, prostrate perennial herbs, acaulescent rosettes, stem rosettes, woody shrubs and small trees) and thrive in a very wide array of ecosystems and climates (coastal dunes, chaparral, sagebrush steppe and grasslands, but especially open mountain forests, meadows and disturbed slopes, extending to sub-alpine/alpine elevations) across an exceptionally wide altitudinal range (from sea level to \(4,900\) m, close to the upper global elevation limit for plant growth), and spanning \(~100^\circ\) of latitude from \(~30^\circ\)S in the Andes to \(~70^\circ\)N in Alaska. On the basis of extensive taxon sampling and phylogenetic analyses, a set of parallel and nested radiations with significantly higher net diversification rates compared with background levels in the genus (net diversification rate \(r = 0.10–0.48\) lineages per million years) were identified within the New World: the western North American perennial lupins with \(58\) species, an estimated age of \(4.6\) Ma and \(r = 0.46–1.76\); the eastern South American clade (\(35\) spp, \(6.5\) Ma and \(r = 0.36–1.33\)), the Mexican and Central American species (\(46\) spp, \(2.7\) Ma and \(r = 1.19–6.15\)) and the Andean clade (\(\sim 81\) spp, \(2.7\) Ma and \(r = 1.56–5.21\))\(^22\),\(^24\). Net diversification rates in these clades are amongst the highest documented for plants\(^23\), exhibit an up to 12-fold increase over background rates observed in *Lupinus*\(^22\), and are apparently still accelerating, suggesting that these radiations may be in the early explosive phase of radiation\(^25\). The rapid radiations within New World lupins are strongly associated with the derived evolution of perennial life history and dispersal into montane ecosystems\(^22\),\(^24\). It is thought that the shift to perennial habit allowed colonization of higher altitude regions, and enabled the accelerated diversification of growth forms via ecological release in response to ecological opportunities in physiographically heterogeneous montane environments\(^23\),\(^25\). The Andean lupins are particularly notable for being among the most rapidly diversifying plant clades documented to date, with an exceptionally high rate of net species diversification\(^22\), in line with other high-altitude Andean grassland plant radiations which are predominantly young and fast\(^26\), and comparable to rates of diversification documented for cichlid fish radiations in African lakes\(^25\),\(^27\). In contrast to these rapid evolutionary radiations, the annual, lowland species of *Lupinus* exhibit significantly lower species diversification rates, including the Old World species, the eastern North American clade and the western North American annual species\(^22\),\(^24\),\(^28\).

Here we investigate the role of natural selection during rapid evolutionary radiations in New World lupins. Taking advantage of the striking among-lineage variation in species diversification rates within this genus, we show that rapid species diversification is underpinned by rapid onset of lineage- and genome-wide accelerated adaptive evolution affecting both coding sequences and expression levels of genes. Our results support a preponderant role for adaptive evolution in evolutionary radiations, and reconcile a long-standing debate on the relative contribution of coding and regulatory changes in evolution.

**Results**

**Transcriptome sequencing.** To analyse selective pressures at a significant fraction of genes across the genome we obtained transcriptome sequence (RNA-seq) data from 55 New World lupin species (67 individuals totalling 2.6 billion Illumina HiSeq reads, Supplementary Data 1). We included multiple representative species from *Lupinus* lineages with contrasting diversification rates: the slowly diversifying North American annual lineages (NAA, 11 species); and the rapidly diversifying North American Perennial lineages (NAP, 11 species) and the Andean clade (26 species, including nine as yet undescribed species). We also included species representing the other New World lineages: the eastern South American, eastern North American and Mexican lupins. The notable advantage of the RNA-seq approach is the ability to track adaptation at both sequence and gene expression levels, to address long-standing debate about the relative importance of structural and regulatory evolution in adaptation\(^20\),\(^21\). We identified and analysed coding sequences and normalized expression values for 6,013 orthologous genes present...
in at least 30 species (Supplementary Fig. 1) including both the rapidly and slowly radiating lineages of lupins.

**Phylogeny and diversification rates.** Phylogenetic reconstruction using maximum-likelihood (ML) on the concatenated data set and coalescent-based species tree methods returned very similar results (Supplementary Fig. 2 and Supplementary Note 1). Henceforth, we discuss only the results obtained with the coalescent-based approach, but results obtained with the ML phylogeny were similar throughout and are reported in Supplementary Files.

The phylogeny obtained (Fig. 1) is largely congruent with previous studies\(^{22,23}\), but exhibits robustly supported species relationships even within the very young Andean clade that previously lacked resolution (Supplementary Fig. 2). Phylogenetic reconstruction using only gene expression values resulted in a very similar topology (Supplementary Fig. 3). Estimated diversification rates confirm the previously reported increase in diversification rate in NAP, Mexican and Andean lineages\(^{22,23}\) (Supplementary Fig. 4).

**Selection on coding sequences.** To estimate the frequency of natural selection on coding-sequence evolution, we implemented dN/dS phylogenetic tests for each gene. As the power of these tests depends on the number of species used in the analysis\(^{29}\), we

---

**Figure 1 | Coalescent-based phylogeny of New World lupins based on 6,013 loci.** The black arrow marks the shift from annual lowland species to mostly perennial montane lupins, which coincided with a significant increase in net species diversification rate. All nodes have > 90% bootstrap support except those marked with black dots. Inset photographs illustrate the diversity of growth forms across the genus, which includes ephemeral annuals, prostrate perennial herbs, acaulescent rosettes, stem rosettes, woody shrubs and small trees (photos by C.E.H.). Inset map shows the geographic distribution of New World lupins, encompassing a variety of ecosystems and climates (coastal dunes, chaparral, sagebrush steppe and grasslands, but especially open mountain forests, meadows and disturbed slopes, extending to sub-alpine/alpine elevations) across an exceptionally wide altitudinal range (from sea level to 4,900 m). Inferred distribution ranges of the main Lupinus lineages are coloured to match branches on the phylogeny (adapted from ref. 22). The main subgroups studied are highlighted: the NAA, NAP and the Andean clade. Tree scale is in expected substitutions per site.
analysed separate subgroups of 10 randomly selected species of slow (NAA) or rapid (NAP and Andean) lineages, and the entire New World and Andean clades.

The percentage of genes evolving under positive selection ranged from 5 to 32% in different Lupinus lineages (Fig. 2a and Supplementary Fig. 5). We found a clear difference between slowly and rapidly diversifying lineages, with only 5.8% of positively selected genes in NAA, while 13.1–16.8% showed evidence of positive selection in the rapidly diversifying lineages (Supplementary Table 1). This result is robust regardless of the sub-sample of Andean species analysed (ANDs1 and ANDs2, Fig. 2a,b). For genes showing evidence of positive selection, the percentage of sites affected ($P_{sel}$) and the dN/dS estimates for these sites ($\omega_1$) varied widely within and between groups analysed, and were very high in some cases ($P_{sel}$ > 20% or $\omega_1$ > 100; Supplementary Fig. 6). Estimates of these parameters for data sets with low species divergence, as in our case, can be relatively imprecise, hence we focus on the likelihood ratio tests (LRT) that are robust even if accurate parameter estimates are not possible. Our results are consistent with the higher power for larger sample size (Fig. 2b), and are conservative given that the older age of the NAA subgroup should increase the power29, yet we see least positive selection in NAA. Similar results were obtained with the branch-site modelling framework (Supplementary Fig. 7 and Supplementary Note 2): the rapidly diversifying Andean clade showed higher rates of positive selection compared with NAA, while NAP exhibited intermediate values. Taken together, our analyses reveal a strong signal of accelerated protein-coding evolution associated with rapidly diversifying lineages, with up to three times as many genes evolving under positive selection compared with slowly diversifying lupins. Gene Ontology (GO) analysis revealed no over-represented terms in the set of genes under selection within each group, suggesting natural selection affected many different genes and gene functions.

The percentage of genes under selection, 40% when all 26 Andean species are considered, is strikingly high (Fig. 2b), particularly given the young age of this clade. Although no data sets of comparable size exist for other plant or animal genera to set this result in a wider context, population genetic studies indicate that over 50% of non-synonymous mutations are fixed by positive selection in most species analysed30–32. Conversely, the percentage of non-synonymous mutations fixed by positive selection is close to zero in humans33 and several plant species studied34,35. However, both sample size and methodology of analysis vary widely across studies, making direct comparisons difficult. To address this we collected and analysed previously published transcriptome data for seven other plant genera, each represented by six to eight species (Supplementary Data 2). To account for sample size differences we re-analysed random subsets of eight species from rapidly and slowly diversifying Lupinus lineages. We find that the percentage of genes evolving under positive selection in the slowly diversifying NAA lupins is similar to other plant genera, while the high frequency of positive selection within rapidly diversifying NAP and Andean lupins remains unmatched (Fig. 2c and Supplementary Table 2).

Selection on gene expression levels. To understand the evolution of gene expression levels during rapid diversification, we calculated a rescaled relative branch length index for each branch (RRBL; Fig. 3), which reflects overall changes in gene expression

**Figure 2 | Phylogenetic analysis of selection on coding sequences. (a)** Percentage of genes evolving under positive selection in the different Lupinus lineages (New World—all species; NAA; NAP; ANDs1 and ANDs2 are random subsets of 10 Andean species). (b) The effect of the number of species used on the power to detect selection. Because of missing data, different genes were sampled in a variable number of species. The graph shows the percentage of genes inferred to contain sites evolving under positive selection, for each group and according to the number of species sampled. Results are only shown for points where 100 or more genes were tested. (c) Percentage of genes evolving under positive selection across seven plant genera with transcriptomes available for sets of species, compared with the different Lupinus lineages studied. Each group was tested using between six and eight species. For a and c values depicted are averages across all genes sampled in at least six species, and numbers on top of bars denote number of genes tested for selection with each group.
values relative to sequence-based divergence along each branch. We obtained the expected distribution of RRBL values under a strictly neutral model of gene expression evolution by simulating gene expression changes along the observed phylogeny with Brownian motion (BM) processes. We found that the Andean clade is significantly enriched for positive RRBL values (two-sided binomial test, \( P = 2.294 \times 10^{-6} \)), and mean RRBL values were highest in the Andean lineage, lowest in the NAA and intermediate in the NAP (inset Fig. 3 and Supplementary Fig. 8). As we used the observed tree topology and branch lengths in the simulations and did not detect any clustering of high RRBL values, our results are unlikely to represent an artefact of the shorter branch lengths in the Andean clade, but instead indicate more frequent directional changes in gene expression values in this clade, possibly driven by positive selection. Results of RRBL analyses using subsets of species for which the same plant tissues (see Supplementary Data 1) were collected, returned identical results (Supplementary Fig. 8). Thus, the observation of higher expression divergence in the Andean clade is robust to the slight differences in tissue sampling between species.

To detect whether shifts in gene expression values occurred gradually and randomly, or in punctuated bursts during diversification of New World lupins, we performed Ornstein–Uhlenbeck (OU) tests on gene-expression values. These tests revealed that a model including at least one shift in optimal expression level was preferred for most genes (5,108; 93% of genes tested), and that these shifts were strongly clustered: 671 (13%) occurred close to the base of the Andean clade, subsuming the fastest diversifying lupin lineage\(^{22,23}\), while a further 608 (12%) occurred near the branch leading to the NAP (Fig. 3), closely matching the shift to mostly perennial life cycle and montane habitats. These two nodes appear as outliers in the distribution of gene expression shifts along the phylogeny (Supplementary Fig. 9). Similar to the analysis of RRBL values, clustering of gene expression shifts in these nodes is robust to differences in tissue sampling between species (Supplementary Fig. 9). Gene Ontology analysis revealed several over-represented terms in genes showing shifts in gene expression in outlier nodes, including cellular components (GO0009579 thylakoid, GO0009536 plastid), molecular functions (GO0039824 catalytic activity, GO0004518 nuclease activity and GO0016788 hydrolase activity) and biological processes (GO0015979 photosynthesis) (Supplementary Fig. 9).

**Discussion**

Our results represent the first systematic genome-wide analysis of natural selection at the molecular level across closely related congeneric lineages with sharply contrasting slow and fast species diversification rates. With dense sampling of species across lineages, we demonstrate that explosive species diversification in lupins is underpinned by widespread adaptive evolution at the protein sequence and gene-expression levels, indicating the importance of both processes for rapid diversification. Accelerated protein evolution is not an inherent feature across all lupins, as the slowly diversifying annual species exhibit levels of selection comparable to other plant genera. Instead, the increase in frequency of positive selection is restricted to the rapidly diversifying *Lupinus* lineages, illustrating the involvement of widespread Darwinian selection during rapid species diversification.

Accelerated coding-sequence evolution during rapid diversification has been suggested in other evolutionary radiations of both animals and plants. In cichlids, comparison of four East African species to the Nile tilapia (*Oreochromis niloticus*) showed higher dN/dS values for species belonging to the rapidly diversifying clades, but failed to distinguish a scenario of relaxed purifying selection from increased positive selection\(^{12}\). Furthermore, this study included only one or two species from each of the East African Lakes (Tanganyika, Malawi and Victoria), and a single closely related representative of the ‘slow radiating’ *O. niloticus* lineage—a lineage encompassing over 60 species and including several rapidly diversifying sub-lineages, and thus not necessarily the ideal representative of a slowly diversifying clade\(^{17}\). Similarly, in birds higher than average dN/dS values and an increase in the number of genes under positive selection were reported in the silveryeye, *Zosterops lateralis*\(^{14}\), but this analysis used a single representative species of the large and rapidly diversifying white-eye bird radiation, and compared it to distantly related species (over 20 Ma divergence). In plants, a high percentage of genes under selection was detected for species in the Hawaiian *Schiedea* radiation compared with other plant genera\(^{13}\), but the number of genes analysed (36) was small. Furthermore, a transcriptome-wide analysis of the wild tomato clade detected several genes experiencing positive selection\(^{18}\), but the authors did not compare these values to related lineages with lower diversification rates and thus the link between positive selection and diversification rates remained unclear. Our results thus represent the first unambiguous example of a rapid shift between low and high rates of coding-sequence evolution concomitant with a large increase in species diversification rates, and affecting a large number of species within the same lineages. Similar analyses of densely sampled phylogenies of other evolutionary radiations will be necessary, to conclude whether accelerated adaptive evolution is a general property of rapid species diversification.
The genome-wide effect of Darwinian selection during rapid diversification is illustrated by our finding that over 40% of genes have undergone adaptive evolution of coding sequences during the recent radiation of Andean lupins. The Andean clade spans almost 5,000 m of elevation with a few species at low altitudes and cohorts of different species at mid-, high- and very high altitudes, including a suite of species adapted to the extreme daily freeze-thaw conditions found between 4,500 and 4,900 m, close to the upper limits for plant growth. The high percentage of genes under selection in this clade suggests that rapid diversification in Andean lupins was at least partly driven by ecological divergence, as species adapted to the very diverse habitats spanning the recently formed and physiographically heterogeneous Andean mountains, in line with recent work showing a large role for ecologically driven diversification in other Andean clades18.

We found close associations between clusters of shifts in gene expression and shifts in species diversification rates, habitats and life history traits, suggesting that rapid gene expression divergence may underpin rapid evolutionary diversification. Interestingly, the very young Andean radiation shows a higher rate of expression divergence than the older NAP species. Both lineages show increases in net species diversification rates compared with annual lupins (NAA), but phenotypic diversity and the rate of morphological diversification are both highest within the Andes, with spectacular variation in plant habit and size, showing a 100-fold variation in plant height and ranging from minute prostrate mat-forming plants less than 5 cm tall, to cauliflorescent cushion-like perennial herbs, to large stem rosettes, to woody shrubs and trees 5–6 m in height and 10 cm stem diameter. Variation in growth forms and plant size is matched by similar ≥100-fold diversification in inflorescence size, disposition and form. The association between gene expression divergence and rates of morphological diversification supports the idea that gene expression plays a dominant role in morphological evolution, a theory first proposed more than 30 years ago16 and recently supported by empirical studies in model organisms21,57,38, but hitherto untested in non-model organisms or evolutionary radiations.

The phylogeny obtained in this study supports the paraphyly of the North American Perennial lupins with respect to the Mexican and Andean lineages. This suggests that North American perennial lupins represent a first foray into montane environments, conferring pre-adaptation to montane conditions in lineages that subsequently dispersed to the Mexican and Andean mountain ranges where they diversified rapidly and extensively22, a pattern seen in other plant groups pre-adapted to cold environments in mountain ranges such as the Andes39. In this regard, the New World lupin radiation is similar to that of the large adaptive radiation of cichlid fish in the East African Great Lakes, where the older Lake Tanganyika harbours a paraphyletic assemblage from which a lineage, the Haplochromini, emerged to colonize and radiate extensively in Lake Malawi and Lake Victoria40,41. The extremely high diversification rates in the latter two lakes are thus derived from lineages pre-adapted to lacustrine environments that originally evolved in Lake Tanganyika42,43. Together with the steep multi-dimensional ecological gradients found in both lakes and mountains (for example, in depth or elevation, water or air temperature, substrate or soil type), the arrival of species pre-adapted to available habitats seems a crucial factor driving the very fast diversification rates seen in such disparate groups as cichlids and lupins23,27.

Studies of other evolutionary radiations suggest an important role for genomic mechanisms that provide evolutionary flexibility and allow clades to diversify rapidly when new ecological opportunities arise. In particular, the recruitment of genes from gene duplication events, and the sharing of genetic variation between species through ancestral polymorphisms or natural hybridization, have been proposed as important drivers of rapid species diversification14,18,19. Chromosome rearrangements and genome duplication may have played a significant role in evolution of lupins, as chromosome numbers vary considerably throughout the genus (2n = 30 to 2n = 52) (ref. 22). However, this source of evolutionary novelty may be less significant in the species studied in this paper because chromosome number is constant (2n = 48) across the whole western New World clade encompassing the North American Annuals, North American Perennials, Mexican and Andean species. Nevertheless, the role of gene and genome duplication in adaptation will be an interesting area for future research.

Methods

Data collection. We obtained Lupinus spp. samples from different sources listed in Supplementary Data 1. We included multiple representative species from Lupinus lineages with contrasting slow or rapid diversification rates. To maximize the sampling of genes, we collected and homogenized different plant tissues; stem and leaf tissue were available for all samples, while flower tissue was collected only for species that flowered in the greenhouse (Supplementary Data 1). We sampled young leaves with stem, and flowers from developing buds. For each accession, tissues were collected from a single plant, up to 100 mg of tissue was flash-frozen in liquid nitrogen, ground to powder and used for total RNA extraction with the RNAasy Plant extraction kit (Qiagen), following the manufacturer’s instructions and performing the optional DNase digestion step. The integrity of extracted RNA was assessed with the Agilent 2100 Bioanalyzer, and samples with low integrity values were re-extracted. For each sample, paired-end Illumina libraries were constructed, samples multiplexed and sequencing performed on the Illumina Hi-seq sequencing platform with paired-end reads of size 100–125 bp.

Transcriptome assembly. We quality checked the short sequence reads of each species using FASTQC v 0.11 (available from http://www.bioinformatics.babraham.ac.uk/projects/fastqc). We used CUTADAPT v 1.4 (ref. 44) to remove illumina adaptors, and TRIMMOMATIC v 0.32 (ref. 45) to trim reads. We trimmed leading and trailing bases with quality below 5, and cropped reads if the average quality per base dropped below 15 over 4 consecutive bases. We discarded reads smaller than 36 bp after trimming. To assemble a de novo transcriptome for each individual, we used TRINITY v r20140413p1 (ref. 46) with default settings except for the minimum assembled contig length (set to 300 bp). We identified putative coding sequences (CDSs) within each transcript with TRANSDCODER (available from https://transdecoder.github.io/) using default settings.

Orthology inference. Using a clustering and phylogeny-based orthology inference method47–49 we identified 6,013 orthologous genes longer than 300 bp present in at least 30 samples (Supplementary Methods). Each orthologous gene was annotated with GO terms in Blast2GO (ref. 48) using the NCBI Non-redundant database (E-value threshold of 10−5 and considering only hits with minimum 50% coverage). GO enrichment tests for over-represented terms were performed using Fisher’s exact tests at false discovery rate (FDR) corrected P < 0.05 (ref. 49). Coding sequences of these 6,013 orthologous genes and their relative gene expression levels were used in all subsequent analyses.

Phylogenetic inference. To recover the phylogenetic relationships within the New World lupins sampled in this study, we employed three alternative approaches (Supplementary Methods): ML phylogenetic inference50 on the concatenated data set of all orthologs (supermatrix approach); coalescent-based phylogenetic construction51 using individual gene trees (species tree approach); and neighbour-joining based on the gene expression values. For all phylogenetic analyses we used only orthologous genes available in at least 30 species and with 300 bp remaining after trimming positions with more than 30% missing data.

Phylogenetic conflict. To infer the extent of phylogenetic conflict among genes, we performed the Shimodaira-Hasegawa test (SH-test)52 with each gene to compare the fit of an unconstrained gene tree to either of the two species trees (supermatrix or species tree). We used PHYML v 3.1 (ref. 53) with the GTR + F model to optimize the likelihood of the unconstrained gene tree and the two species trees. We then used the individual site likelihoods reported by PHYML to perform the SH-test in CONSEL v 1.2 (ref. 54), and excluded genes that significantly preferred the gene tree over the species tree (FDR-corrected P < 0.05).

Diversification rates. To test whether previously detected net diversification shifts within New World lupins are supported in our data set, we performed a Bayesian analysis of macroevolutionary mixtures55. We converted the sequence-based phylogenies obtained previously into ultrametric trees using the
CHRONOPL, which implements the penalized likelihood method. We arbitrarily set the age of the root to 1 (we were only interested in relative divergence rates), and estimated the optimal value for the rate smoothing parameter by cross-validation. We used the resulting ultrametric tree in the program RAMiS (ref. 55) to obtain estimates of net diversification rates for each branch with the speciation and extinction model. We used previously published estimates for the number of species within each lineage to set clade-specific sampling probabilities as follows: NAA = 0.41 (11 species sampled, 27 estimated); NAP = 0.17 (11/63); Andean clade = 0.32 (26/81); Mexican clade = 0.04 (2/46); and eastern New World species = 0.1 (4/39). To test for convergence, we ran and compared the results of three independent MCMC simulations, 10 million generations each, every 5,000 generations. The package RAMiS models 2.0 (ref. 58) was used to test for convergence, plot the estimated net diversification rates per branch, and identify the most credible set of distinct rate shift configurations.

Selection on coding sequences. To estimate the frequency of positive selection acting on coding sequences, we used methods based on the estimation of the non-synonymous to synonymous substitution rate ratio (ω = DN/DS) along phylogenetic branches for each gene using ‘sites models’—which allow ω to vary among sites within each gene—and ‘branch-site models’, which allows ω to vary among sites of the gene and branches of the phylogeny. We performed these analyses using only genes that did not show significant phylogenetic conflict (SH-test).

We used the program CODEML from the package PAML v 4.8 (ref. 61) to fit to each gene the ‘sites models’ M7 and M8 (ref. 59). Both models split the sites in an alignment into classes: model M7 considers only two classes of sites, viz. neutrally evolving (ω = 1) or negatively selected sites (ω<1); while model M8 additionally includes a class of sites allowed to evolve under positive selection (ω>1). We compared the fit of these models with a likelihood ratio test (LRT), and corrected resulting P values with the FDR method. For genes significantly rejecting M7 in favour of M8, we performed a second test, comparing the fit of models M8 and M8a. The latter has the same parameters as M8 but ω is fixed to 1 (ref. 62). We classified genes as evolving under positive selection when model M8 exhibited a significantly better fit than model M8a (LRT with 1 degree of freedom, FDR corrected P<0.05). We performed these ‘sites-models’ tests using the entire data set of New World lupins and the entire Andean assemblage, as well as separately for subgroups of 10 randomly selected species representing the slow radiating North American Annuals (NAA) and the fast radiating North American Perennials (NAP) (ref. 59,60). We analysed each gene using ‘sites models’—which allow ω to vary among sites within each gene—and ‘branch-site models’, which allows ω to vary among sites of the gene and branches of the phylogeny. We performed these analyses using only genes that did not show significant phylogenetic conflict (SH-test).

To identify genes containing sites that experienced episodic positive selection on specific lineages we used the adaptive branch-site random effects likelihood method (aBS-REL) implemented in the package CODEML v 2.2 (ref. 59). aBS-REL fits a variable number of site classes—some of which are allowed to experience positive selection—to each branch of the phylogeny. Using a model-selection procedure the method infers the optimal number of classes for each branch, and tests whether any of these classes is evolving under positive selection (ω>1). We used the same set of filters as for the ‘sites-models’ described above, and performed these tests for each gene in the data set including all New World lupins. The branch-site tests returned a list of branches under selection in specific lineages. For each gene, we compared an OU model allowing for a single number of ω processes along phylogenies can be compared using a LRT to infer shifts in trait value using the OU modelling framework. In addition, models allowing for a different number of ω processes along phylogenies can be compared using a LRT to infer shifts in optimal trait value in specific branches. For each gene, we compared an OU model allowing for a single optimal expression level for all the branches of the phylogeny (OU1-model) to models that allow a shift in the optimal expression value in a specific node of the phylogeny (OU2-models). We fitted OU2-models to each possible node in the phylogeny, and compared the likelihood of the best-fitting OU2-model to the OU1-model, using a LRT with 1 degree of freedom. For genes significantly rejecting the OU1-model compared to the best OU2-model (FDR corrected P<0.05), we recorded the node on which the shift in optimal gene expression level occurred. For each gene, we then summarized how many genes changed gene expression levels on each node of the phylogeny. For species with more than one accessions available, for each gene we used the mean expression level of all conspecifics. This mean-approach was recently shown to have negligible effect on both power and FDR to detect shifts in gene expression using the OU modelling framework . To further reduce the noise introduced by having a single expression level measurement for most species, we only considered OU2 models that included at least three species in each of the lineages defined.

Data availability. RNAseq data that support the findings of this study have been deposited in the NCBI Sequence Read Archive (SRA) with the accession codes SAMN04869551 to SAMN04869616. Other RNAseq data referenced in this study are available in GenBank with the accession codes SAMEA1378181 (Lupinus polyphyllus); SRR1141027, SRR1141029, SRR1141037, SRR1141040, SRR1165196, SRR1165206, SRR1166363 and SRR1166735 (Flaveria spp.); SRR1171830, SRR1171850, SRR1171942, SRR1210106, SRR452306 and SRR452313 (Glycine spp.); SRR1041637, SRR1041531, SRR1665780, SRR286582, SRR286604, SRR286626 and SRR286631 (Populus spp.); SRR957566 (Lonicum spp.); SRR957565 (Lonicum spp.); SRR957722, SRR1171995, SRR1220645, SRR1592601 and SRR1598571 (Oryza spp.); ERR206331, SRR1046128, SRR1508229, SRR166793, SRR450223 and SRR9001769 (Populus spp.); ERR185922, SRR1087902, SRR1104128, SRR122109, SRR1291243, SRR221349, SRR765527 and SRR799447 (Salolium spp.).

References
1. Whitfield, J. B. & Lockhart, P. J. Deciphering ancient rapid radiations. Trends Ecol. Evol. 22, 258–265 (2007).
2. Losos, J. B. & Mahler, D. L. in Evolution since Darwin: The First 150 Years. (eds Bell, M. A., Futuyma, D. J., Eanes, W. F. & Levinton, J. S.) 381–420 (Sinauer Associates, 2010).
3. Hughes, C. E., Nyffeler, R. & Linder, H. P. Evolutionary plant radiations: where, when, why and how? New Phytol. 207, 249–253 (2015).
4. Vendetti, C., Meade, A. & Pagel, M. Multiple routes to mammalian diversity. Nature 479, 393–396 (2011).
5. Rabosky, D. L. et al. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat. Comm.**on.***4, 1598 (2013).

6. Wagner, C. E., Harmon, L. J. & Seehausen, O. Ecological opportunity and sexual selection together predict adaptive radiation. *Nature***487, 366–369 (2012).

7. Donoghue, M. J. & Sanderson, M. J. Confluence, synnovation, and depauperons in plant diversification. *New Phytol.* ***207,*** 260–274 (2015).

8. Rundell, R. J. & Price, T. D. Adaptive radiation, non-adaptive radiation, ecological speciation and non-ecological speciation. *Trends Ecol. Evol.* ***24,*** 394–399 (2009).

9. Gittenberger, E. What about non-adaptive radiation? *Biol. J. Linnean Soc.***43, 263–272 (1991).

10. Givnish, T. J. Adaptive radiation versus ‘radiation’ and ‘explosive diversification’: why conceptual distinctions are fundamental to understanding evolutionary theory. *New Phytol.* ***207,*** 297–303 (2015).

11. Givnish, T. J. in *Molecular Evolution and Adaptive Radiation* (eds Givnish, T. J. & Sytsma, K. J.) 1–54 (Cambridge University Press, 1997).

12. Brawand, D. et al. The genomic substrate for adaptive radiation in African cichlid fish. *Nature***513, 375–381 (2014).

13. Kapralov, M. V., Votintseva, A. A. & Filatov, D. A. Molecular adaptation during a rapid adaptive radiation. *Mol. Biol. Evol.* ***30,*** 1051–1059 (2013).

14. Cornetti, L. et al. The genome of the ‘great spectator’ provides insights into bird diversification. *Genome Biol. Evol.* ***7,*** 2680–2691 (2015).

15. Vendetti, C. & Pagel, M. Speciation as an active force in promoting genetic evolution. *Trends Ecol. Evol.* ***25,*** 14–20 (2010).

16. Sanderson, M. J. Reappraising adaptive radiations. *Am. J. Bot.* ***85,*** 1650–1655 (1998).

17. Berner, D. & Sale wurger, B. The genomics of organismal diversification illuminated by adaptive radiations. *Trends Genet.* ***31,*** 491–499 (2015).

18. Pease, J. B., Haak, D. C., Hahn, M. W. & Moyle, L. C. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol.* ***14,*** e1002379 (2016).

19. Lamichhaney, S. et al. Evolution of Darwin’s finches and their beaks revealed by genome sequencing. *Nature***518, 371–375 (2015).

20. Hoekstra, H. E. & Coyne, J. A. The locus of evolution: evo devo and the genetics of adaptation. *Evolution***61, 995–1016 (2007).

21. Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* ***8,*** 206–216 (2007).

22. Drummond, C. S., Eastwood, R. J., Miotto, S. T. S. & Hughes, C. E. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): testing for key innovation with incomplete taxon sampling. *Syst. Biol.* ***61,*** 443–460 (2012).

23. Hughes, C. & Eastwood, R. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proc. Natl Acad. Sci. USA***103,*** 10301–10306 (2006).

24. Drummond, C. S. Diversification of *Lupinus* (Leguminosae) in the western New World: derived evolution of perennial life history and colonization of montane habitats. *Mol. Phylogenet. Evol.* ***48,*** 408–421 (2008).

25. Hughes, C. E. & Atchison, G. W. The ubiquity of alpine plant radiations: from *Globularia* to *Saxifraga*. *Evolution***61, 1341–1351 (2007).

26. Madigan, M. S., Corl, A. J. & Richardson, J. E. Universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* ***21,*** 3674–3676 (2005).

27. Seehausen, O. Process and pattern in cichlid radiations - inferences for understanding unusually high rates of evolutionary diversification. *New Phytol.* ***207,*** 304–312 (2015).

28. Koenen, E. J. M. et al. Exploring the tempo of species diversification in legumes. *S. Afr. J. Bot.* ***89,*** 19–30 (2013).

29. Anisimova, M., Bia³awski, J. P. & Yang, Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* ***18,*** 1585–1592 (2001).

30. Eyre-Walker, A. The genomic rate of adaptive evolution. *Proc. Natl Acad. Sci. USA***105**(Suppl 1): 11549–11555 (2008).

31. Stepien, S. M. et al. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *PLoS ONE***9, e89543 (2014).

32. Kohlbaur, M. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* ***29,*** 644–652 (2011).

33. Yang, Y. & Smith, S. A. Orthology inference in non-model organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. *Mol. Biol. Evol.* ***31,*** 3081–3092 (2014).

34. Rohlfs, R. V., Harrigan, P. & Nielsen, R. Modeling gene expression evolution and evolutionary dynamics on phylogenetic trees. *PLoS ONE***9, e1002379 (2014).

35. Robinson, E., Claude, J. & Strimmer, K. H. APE: analysis of phylogenetics and evolution in R language. *Bioinformatics***20, 289–290 (2004).

36. Rabosky, D. L. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE***9, e95943 (2014).

37. Ramljak, T. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* ***52,*** 696–704 (2003).

38. Shimodaira, H. & Hasegawa, M. ConSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics***17, 1246–1247 (2001).

39. Rabosky, D. L. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE***9, e89543 (2014).

40. Koblmu¨ller, S., Sefc, K. & Sturmbauer, C. in *Molecular Evolution and Adaptive Radiation* (eds Rabosky, D. L. & Sytsma, K. J.) 1–54 (Cambridge University Press, 2009).

41. Rabosky, D. L. Phylogenetic comparative analysis: a modelling approach for adaptive evolution. *Am Nat***164, 683–695 (2004).

42. Hansen, T. F. Stabilizing selection and the comparative analysis of adaptation. *Evolution***51, 1341–1351 (1997).

43. Seitz, B., Sefc, K. M., Less, G. et al. A rapid adaptive radiation. *Proc. Natl Acad. Sci. USA***105,*** 6504–6510 (2008).

44. Rohlfs, R. V., Harrigan, P. & Nielsen, R. Modeling gene expression evolution with an extended Ornstein-Uhlenbeck process accounting for within-species variation. *Mol. Biol. Evol.* ***31,*** 201–211 (2014).
Acknowledgements
This work was supported by grants from the Natural Environment Research Council (Grant NE/K004352/1 to D.A.F.) and the Swiss National Science Foundation (Grant 31003A_135522 to C.E.H.). We thank the staff at Edinburgh Genomics (University of Edinburgh), the Functional Genomics Centre Zurich (University of Zurich) and the Wellcome Trust Centre (Oxford) for sequencing and computational support, and acknowledge the use of the University of Oxford Advanced Research Computing facility in carrying out this work (http://dx.doi.org/10.5281/zenodo.22558).
We thank the United States Department of Agriculture and the Desert Legume Program (University of Arizona) for seedlots, the national authorities in Argentina, Bolivia, Ecuador, Mexico and Peru for permission to collect plant material, Rayko Jonas and Markus Meierhofer (Zurich) and Tim Wroe (Oxford) for help with growing plants. We thank O. Osborne for assistance in the analysis of non-lupin genera, and M. Chester for guidance in the lab.

Author contributions
D.A.F. conceived the study. B.N., C.E.H. and D.A.F. designed the research and wrote the manuscript. B.N. and G.W.A. generated the data. B.N. analysed the data supervised by D.A.F. All authors discussed the results and commented on the manuscript.

Additional information
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interest.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Nevado, B. et al. Widespread adaptive evolution during repeated evolutionary radiations in New World lupins. Nat. Commun. 7:12384 doi: 10.1038/ncomms12384 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2016