

Is specialization an evolutionary dead end? Testing for differences in speciation, extinction and trait transition rates across diverse phylogenies of specialists and generalists

E. H. DAY, X. HUA & L. BROMHAM

Centre for Macroevolution & Macroecology, Division of Ecology, Evolution and Genetics, Research School of Biology, Australian National University, Canberra, ACT, Australia

Keywords:

diversification;
macroevolution;
null models;
phylogenetics;
specialization.

Abstract

Specialization has often been claimed to be an evolutionary dead end, with specialist lineages having a reduced capacity to persist or diversify. In a phylogenetic comparative framework, an evolutionary dead end may be detectable from the phylogenetic distribution of specialists, if specialists rarely give rise to large, diverse clades. Previous phylogenetic studies of the influence of specialization on macroevolutionary processes have demonstrated a range of patterns, including examples where specialists have both higher and lower diversification rates than generalists, as well as examples where the rates of evolutionary transitions from generalists to specialists are higher, lower or equal to transitions from specialists to generalists. Here, we wish to ask whether these varied answers are due to the differences in macroevolutionary processes in different clades, or partly due to differences in methodology. We analysed ten phylogenies containing multiple independent origins of specialization and quantified the phylogenetic distribution of specialists by applying a common set of metrics to all datasets. We compared the tip branch lengths of specialists to generalists, the size of specialist clades arising from each evolutionary origin of a specialized trait and whether specialists tend to be clustered or scattered on phylogenies. For each of these measures, we compared the observed values to expectations under null models of trait evolution and expected outcomes under alternative macroevolutionary scenarios. We found that specialization is sometimes an evolutionary dead end: in two of the ten case studies (pollinator-specific plants and host-specific flies), specialization is associated with a reduced rate of diversification or trait persistence. However, in the majority of studies, we could not distinguish the observed phylogenetic distribution of specialists from null models in which specialization has no effect on diversification or trait persistence.

Introduction

Specialization, where a lineage evolves to utilize a significantly narrower range of resources than related lineages, is frequently cited as a potential evolutionary dead end, limiting capacity for future evolutionary

change or diversification (e.g. Cope, 1896; Haldane, 1951; Kelley & Farrell, 1998; Nosil & Mooers, 2005; Vamossi *et al.*, 2014). Specialization may be advantageous if it results in efficient selection for adaptation to a narrow niche, reducing the cost of trade-offs by abandoning traits needed to utilize a wider range of resources (Futuyma & Moreno, 1988). However, these shorter term microevolutionary benefits may come at the cost of longer term macroevolutionary success (e.g. Agnarsson *et al.*, 2006; Anacker *et al.*, 2011; Forister *et al.*, 2012; Armbruster, 2014). Specialization on one resource may lead to the irreversible loss of traits

Correspondence: Lindell Bromham, Research School of Biology, Australian National University, Building 116, Daley Road, Canberra, ACT 0200, Australia.
Tel.: +61 2 6125 9545; fax: +61 2 6125 5573;
e-mail: lindell.bromham@anu.edu.au

required for utilizing a wider range of niches, leading specialist lineages down a 'blind alley' that limits future evolutionary potential (Haldane, 1951). Thus, narrowing of resource use may limit the capacity of specialist species to persist if the environment changes, leading to higher extinction rate. Efficient selection for specialization may leave specialists with a restricted distribution and insufficient variation for adaptation to new niches, leading to reduced speciation rate. Therefore, lineages of specialists might have higher rates of extinction and lower rates of speciation than generalists and may also have limited capacity to evolve to exploit more generalist resources.

These predicted patterns have been tested experimentally (e.g. Kassen, 2002; Buckling *et al.*, 2003) and in the field (e.g. Griffith & Sultan, 2012), modelled theoretically (e.g. Ravnigné *et al.*, 2009), examined using genomic data (e.g. McBride, 2007) and tested against the fossil record (e.g. Colles *et al.*, 2009). But increasingly, the influence of specialization on macroevolution has been tested using phylogenetic analyses, as the proposed long-term effects of specialization – raised extinction, reduced speciation and irreversibility – might produce distinct patterns on phylogenies. Phylogenetic tests of the influence of specialization on macroevolution have used a range of approaches, including a variety of methods for ancestral state reconstruction to compare transition rates from generalists to specialists and vice versa (e.g. Stephens & Wiens, 2003; Nosil & Mooers, 2005; Stireman, 2005), and maximum-likelihood modelling of speciation and extinction rates to test whether specialists have reduced diversification rates (e.g. Fernández-Mazuecos *et al.*, 2013; Hardy & Otto, 2014).

The results of phylogenetic tests of the influence of specialization on the tempo and mode of macroevolution have been mixed. Some studies have found that specialists have lower diversification rates (e.g. Fernández-Mazuecos *et al.*, 2013), and others that generalists have lower diversification rates (e.g. Hardy & Otto, 2014). While some studies find that transitions from generalists to specialists are more common than the reverse (e.g. Tripp & Manos, 2008), others find more frequent transitions from specialists to generalists (e.g. Stireman, 2005; Johnson *et al.*, 2009), or no consistent directionality of transitions (e.g. Janz *et al.*, 2001). The mixed pattern of results may be because there is no uniform effect of specialization on macroevolutionary patterns, so there is no general pattern across different phylogenies. Each study may therefore reveal interesting processes at work in that particular system, but have limited generalizability to macroevolution as a whole.

However, a confounding factor is that these phylogenetic studies also differ in their methods and approaches to testing the hypothesis that specialization is an evolutionary dead end. Concerns have been raised that methodological artefacts may obscure processes or generate misleading patterns in the data. Methods and

assumptions used for ancestral state reconstruction may have a strong influence on conclusions regarding relative rates of transitions between generalist and specialist states (Stephens & Wiens, 2003; Stireman, 2005). The BiSSE models commonly used for estimating speciation, extinction and trait transition rates (Maddison *et al.*, 2007; FitzJohn *et al.*, 2009) are often applied to datasets too small to give reliable results (Davis *et al.*, 2013), or where the assumptions of the model may be violated (e.g., if the transition to specialization is associated with speciation events: Goldberg & Igc, 2012). Furthermore, results of phylogenetic studies of specialization are sometimes presented without being tested against a null model, or in cases where the null model of equal diversification or transition rates of specialists and generalists cannot be rejected (Kelley & Farrell, 1998; Fernández-Mazuecos *et al.*, 2013). Evaluating the power to detect signatures of reduced diversification in specialists will be particularly important where specialists tend to be recently derived and dispersed on the phylogeny, which may limit ability to accurately estimate speciation or extinction rates (Kelley & Farrell, 1998).

In order to address some of these issues, we use phylogenetic analyses combined with macroevolutionary models to reanalyse ten data sets, to test the hypothesis that specialists have reduced macroevolutionary capacity due to increased extinction rate, decreased speciation rate and/or reduced capacity to transition to a generalist state. These data sets span a range of taxa (plants, vertebrates and invertebrates) and a range of forms of specialization (diet, host and pollinators). We used a number of phylogenetic measures to detect an impact of specialization on macroevolutionary processes: tip length (branch length connecting a terminal taxon with the node shared with its closest relative), number of tips (species) arising from each inferred origin of a specialist trait and a measure of phylogenetic scatter (how dispersed across the tips of the phylogeny specialists tend to be). For each of these metrics, we generate a null distribution for each data set, and we test the relative fit of a range of macroevolutionary models to explain the observed phylogenetic distribution to detect whether specialization may be considered an evolutionary dead end in each case.

Materials and methods

Case studies

We analysed ten published phylogenies that were previously used to examine macroevolutionary patterns of specialization in particular clades (Table 1). These case studies were selected following a literature search using the key words 'specialization', 'evolution', 'dead end' and 'diversification'. Because we require ultrametric phylogenies with branch lengths for our analysis, we only included a study provided that: (i) the phylogeny

Table 1 List of ten case studies, with details on the taxa and the definition of specialization used in each case study, the number of species included in each phylogeny, and the number of species on each phylogeny that are specialists.

	Taxon	Clade	Type of specialization	Definition of specialization	Phylogeny size	No. of specialists	Reference
1	Plants	<i>Linaria</i>	Pollinator	Plants with narrow corollas (assumes this corresponds to pollinator specificity)	29	7	Fernández-Mazuecos <i>et al.</i> (2013)
2	Butterflies	Heliconiini	Diet	Feed on plants with phylogenetic diversity < 0.1	33	16	Hardy & Otto (2014)
3	Butterflies	Nymphalini	Diet	Feed on plants within one family	25	15	Janz <i>et al.</i> (2001)
4	Lice	<i>Columbicola</i>	Host	Parasitize one host species	38	30	Johnson <i>et al.</i> (2009)
5	Bees	<i>Andrena</i>	Diet	Collect pollen from one plant tribe	35	20	Larkin <i>et al.</i> (2008)
6	Flatworms	<i>Cichlidogyrus/Scutogyrus</i>	Host	Parasitize one host species or genera	26	16	Mendlová & Šimková (2014)
7	Parrots	Psittiformes	Diet	Feed exclusively on nectar	78	13	Schweizer <i>et al.</i> (2014)
8	Flatworms	<i>Dactylogyrus</i>	Host	Parasitize one host species within a local area	51	41	Šimková <i>et al.</i> (2006)
9	Flies	Tachinidae	Host	Parasitize fewer than three host families	49	39	Stireman (2005)
10	Plants	<i>Ruellia</i>	Pollinator	Plants that are pollinated by hawkmoths	115	15	Tripp & Manos (2008)

was estimated using DNA sequences; (ii) the DNA sequences and relevant trait information (i.e. species assigned as specialists or generalists) were publically available; and (iii) the study inferred multiple independent origins of the specialist trait. We included only studies where specialists use a narrower range of resources than the generalists – so, for example, we did not include life-history specialists that have a reduced number of larval stages (Cieslak *et al.*, 2014) or different reproductive mode (Pyron & Burbrink, 2014). Because we require multiple independent instances of specialization to test for any general influence on diversification rates or trait persistence, we were unable to test groups in which there is only a single origin of specialization, which biases case study selection towards traits that may be gained repeatedly within a clade, rather than very uncommon cases of specialization that evolve only rarely.

In this study, we categorized specialists and generalists as binary states, with specialists as state ‘1’ and generalists as state ‘0’. We assumed the same definitions and categorizations of specialists and generalists as the original published study (details in Appendix S1). Although degree of specialization exists on a continuum for some traits (e.g. diet breadth), it is necessary to treat specialization as discrete trait for many phylogenetic tests, including trait-dependent diversification models such as BiSSE (Maddison *et al.*, 2007; FitzJohn *et al.*, 2009). However, we recognize that the patterns observed may be contingent on the way that specialists and generalists have been categorized.

Phylogenies

Our analyses required phylogenies with ultrametric branch lengths (where distances from the root of the tree to every tip are equal), so when this was unavail-

able from the original publication it was necessary to estimate ultrametric branch lengths for the given topology. For one case study (Hardy & Otto, 2014), an ultrametric phylogeny was available, so we used this phylogeny for our analyses. However, for the remaining nine studies, only sequence data or phylogenies that were not ultrametric were available, so we aligned available sequence data and estimated ultrametric branch lengths on a topology constrained to match the published phylogeny using BEAST v2.1.3 (Bouckaert *et al.*, 2014). For one case study (Johnson *et al.*, 2009), the phylogeny estimation process did not converge using a constrained topology, so for this phylogeny, we estimated the topology as well as branch lengths.

To estimate phylogenies, we obtained DNA sequences from GenBank and aligned them in GENEIOUS v6.1.8 (Kearse *et al.*, 2012) using the MUSCLE alignment plugin for noncoding genes and the MUSCLE translational alignment plugin for protein-coding genes (Edgar, 2004). We manually checked and corrected the alignments and removed ambiguous sections. We then estimated phylogenies in BEAST v2.1.3 (Bouckaert *et al.*, 2014) using a general time-reversible (GTR) model of nucleotide evolution with separate partitions for each gene and codon position, a relaxed log-normal clock and birth–death priors. We ran a single MCMC chain for > 30 000 000 generations, sampling trees every 3000 generations. If the estimation process did not converge using these settings, we reduced the complexity of the model, for example using a Hasegawa, Kishino and Yano (HKY) model of nucleotide evolution and yule priors (see Appendix S1). We evaluated adequate posterior sample size and stationarity through visual inspection of trace plots in TRACER v1.5 (Rambaut & Drummond, 2012), when the effective sample size for all parameters was > 200. We selected the maximum clade credibility (MCC) tree from the posterior distribution of phylogenies using TREEANNOTATOR v2.0.3

(Bouckaert *et al.*, 2014) with the first 5% of trees discarded as burnin.

Quantifying phylogenetic distributions

To test whether specialists have different macroevolutionary dynamics from generalists, we need a way of comparing the rate of key macroevolutionary processes between specialist and generalist lineages. Specifically, we wish to ask whether specialization affects the inferred speciation and extinction rates of a lineage or influences the trait transition rate. As trait-associated changes in speciation, extinction and transition rates can leave their mark on phylogenies, we ought to be able to use phylogenetic patterns to test for these effects of traits on macroevolutionary processes (Maddison *et al.*, 2007; Bromham *et al.*, 2016; Hua & Bromham, 2016). To quantify the phylogenetic distribution of specialists, we used three different phylogenetic metrics to express different qualities of a phylogenetic distribution of traits that may reflect underlying macroevolutionary processes (Bromham *et al.*, 2016), using the R package 'PHYLOMETRICS' (Hua & Bromham, 2016) implemented in R v1.3.1 (R Development Core Team, 2015).

One common way to assess the effect of a trait on macroevolutionary dynamics is to use a model-fitting approach to find the model that gives the most likely explanation of the observed data, then ask whether the best-fit model infers different parameter values for the specialist and generalist lineages. One limitation of these model-fitting approaches is that the best-fitting model may not be sufficiently better supported by the data than alternative models, making it difficult to reject the null hypothesis that specialization has no influence on diversification or persistence. Furthermore, in some cases, even the best-fitting model may be a poor description of the data, so in addition to testing against a null hypothesis, we need to be able to test the adequacy of fit of the best model, in order to be able to reject even the best-fit model as an adequate description of the data.

Tip length

If specialization leads to lower speciation rates, this could result in long tip lengths because specialists are less likely to diversify. Here, 'tip' refers to the branch connecting terminal taxa to the node shared with their closest relative. A lineage that has reduced speciation rate may persist through evolutionary time but not give rise to many daughter lineages, so the distance between any extant species with the specialized trait and its nearest relative may be increased. We used the tip age rank sum (TARS) metric to test whether specialists have longer tip lengths than generalists, with significance determined by the nonparametric Wilcoxon rank-sum test (Bromham *et al.*, 2016).

Number of tips per origin

If specialization causes high rates of extinction or frequent loss of the specialized trait, then specialized lineages are unlikely to persist for long enough to give rise to large clades of descendants. Therefore, we might expect specialists to be scattered across the tips of the phylogeny, in shallow clades with recent origins of the specialist trait (i.e. a 'tippy' distribution: Kelley & Farrell, 1998; Stireman, 2005). For any given number of specialist species, we would expect them to arise from a larger number of independent origins than for a trait that does not influence rates of extinction or trait loss. Therefore, we would expect the number of estimated origins of specialization to be high relative to the number of specialist species. We used number of tips per origin (NoTO) to infer the minimum number of origins required to explain the observed distribution of specialists across the tips of the phylogeny. To generate a null model of the expected distribution across the phylogeny for a trait that has no influence on extinction rate or trait loss, we used a stochastic model of trait evolution along the same topology, fixing the number of specialists to the number in the observed data. We used this procedure to generate 1000 simulated data sets from which to calculate the significance of any difference between the observed data and the null distribution (Bromham *et al.*, 2016).

Phylogenetic scattering

If specialization causes either high rates of extinction or raised rate of loss of the specialist trait, then there will be relatively few large clades of taxa sharing the specialized trait, and most observed specialist lineages will arise from relatively recent origins of the specialist trait. This will result in specialist taxa having few close relatives that share the specialist trait, which will lead to a relatively high degree of phylogenetic scatter. We assessed the degree of phylogenetic scatter of specialists using the sum of sister clade differences (SSCD), which calculates the sum of absolute difference between pairs of daughter species descending from each node, with statistical significance determined by comparison with stochastically simulated traits similarly to NoTO (number of tips per origin). We also estimated phylogenetic scatter using the Fritz & Purvis D statistic (FPD) (Fritz & Purvis, 2010); however, the results were similar to SSCD, so we reported the results of FPD in Table S1.

Macroevolutionary models

Each of the phylogenetic tests described above incorporates a null model test, where we ask whether the observed value of each metric is significantly different from the distribution of values we would expect for any given data set if specialization evolves under a stochastic process. However, we also need to know whether the observed distribution of the metric is dis-

tinguishable from a null model where specialization has no effect on lineage diversification or persistence (i.e. when speciation, extinction and trait transition rates are equal between specialists and generalists). Furthermore, we need to test whether the observed data are consistent with a model where specialization is an evolutionary dead end (it increases the extinction rate, lowers the speciation rate and/or influences trait transition rate) and whether the observed distribution of specialists could have arisen under any alternative macroevolutionary processes. Alternative macroevolutionary models can be generated and tested against the observed phylogenetic distribution of specialists, using the parameter space (speciation, extinction and trait transition rates) of the observed phylogeny to generate expected values of these metrics under different macroevolutionary scenarios.

To develop alternative models of trait evolution for each data set, we estimated maximum-likelihood values for parameters in the binary-state speciation and extinction (BiSSE) model (Maddison *et al.*, 2007; FitzJohn *et al.*, 2009) using the 'DIVERSITREE' package in R (FitzJohn, 2012). These parameters are: speciation rate for specialists (λ_1) and generalists (λ_0), extinction rate for specialists (μ_1) and generalists (μ_0), and rates of gain of the specialist trait (q_{01}) and loss of the specialist trait (q_{10}). Next, to generate distributions of traits under a range of parameter values, we estimated eight alternative parameter sets by constraining one, two or three of the parameters to be equal for specialists and generalists. As irreversibility has been cited as an important influence in the macroevolutionary patterns of specialization, we also generated a parameter set by constraining the rate of loss of the specialized trait to be zero ($q_{10} = 0$). In total, nine different parameter sets were estimated for each data set, representing alternative models of trait evolution (Tables S2 and S3). For example, constraining all parameters to be equal between specialists and generalists ($\lambda_1 = \lambda_0$ and $\mu_1 = \mu_0$ and $q_{01} = q_{10}$) creates a null model where specialization and generalization have equal effect on diversification rates and trait persistence.

To test whether we would expect each of these models to produce the observed phylogenetic distribution of specialists, we used 'backward simulation' to generate phylogenies with the same number of specialists and generalists as the observed phylogeny, taking into account the sampling fraction (completeness) of the phylogeny (Hua & Bromham, 2016). Using this method, we simulated 100 trees constrained under each parameter set for each case study. Under some parameter sets, in particular when rate of extinction or trait loss was high, lineages did not coalesce into one ancestral state due to an overall negative diversification rate. In these cases, extinction rate was reduced to increase the overall diversification rate to allow for tree simulation (Table S3). We then computed the values of TARS

(tip length), NoTO (number of tips per origin), SSCD (phylogenetic scatter) and FPD (Fritz & Purvis D statistic) for each of the 100 simulated trees for each parameter set. We rejected a given macroevolutionary model as an adequate explanation for the observed data if the values of any one of these metrics on the observed phylogeny were significantly different from the distribution of values from the simulated trees (if $P \leq 0.01$ or $P \geq 0.99$ following a Bonferroni correction, where P represents the proportion of simulated metric values that are more extreme than the observed metric values). We set the significance level at 0.01 because four metrics were used to test each of the simulated macroevolutionary models against the observed data.

In addition, we used the Cramér test (Baringhaus & Franz, 2004), a nonparametric multivariate t -test, to assess whether each model can adequately explain the phylogenetic distribution of traits as described by the four metrics. We performed the test using the R package 'CRAMER' (Franz, 2014) and scaled each metric before the test. For this test, we rejected a given macroevolutionary model as a plausible explanation for the observed data if $P < 0.05$.

Results

Two of the ten case studies showed phylogenetic patterns of specialization that were consistent with specialization being an evolutionary dead end. Firstly, hawkmoth-pollinated plants in the genus *Ruellia* (Tripp & Manos, 2008) have fewer species per origin (pNoTO = 0.013) and are more phylogenetically scattered (pSSCD = 0.002) than expected if the trait had evolved under a stochastic process (Fig. 1, Table 2). We also rejected a null model where specialization and generalization have equal influence on diversification for the *Ruellia* dataset using both the univariate ($P < 0.01$) and multivariate analyses ($P = 0.025$). Several macroevolutionary models could not be rejected, including models where specialists have higher extinction rates, higher rates of trait loss, or lower speciation rates than generalists (Fig. 2). Secondly, species of parasitoid flies in the family Tachinidae (Stireman, 2005) that are host-specific (parasitize fewer than three host families or fewer than eight host species) show evidence of reduced speciation rate, as specialists have significantly longer tip lengths than generalists (pTARS = 0.997, Table 2). We rejected a null model where specialization and generalization have equal influence on diversification rates for the Tachinid flies data set using the univariate analysis ($P < 0.01$) although this is not statistically significant for the non-parametric multivariate analysis ($P = 0.067$). Models that we could not reject are those in which specialists have lower speciation rates than generalists, where the specialist trait is highly labile (with high rates of gain and loss) or where specialization is an irreversible trait

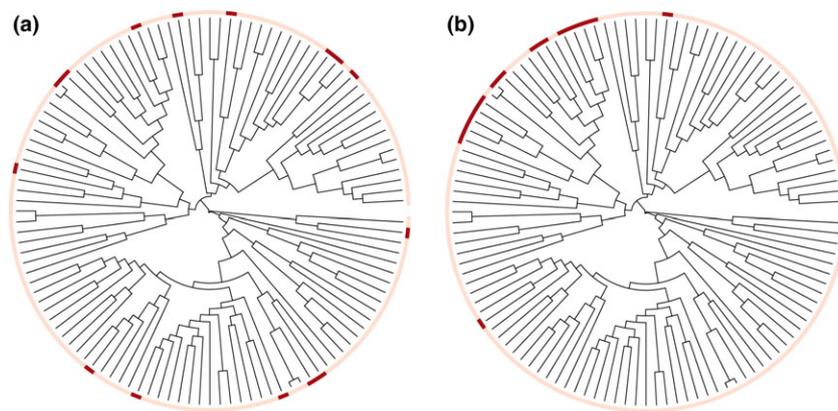


Fig. 1 (a) The observed distribution of hawkmoth-pollinated species (dark red) on a phylogeny of the genus *Ruellia* compared to (b) a randomly selected example of the expectation under Brownian motion evolution generated by simulation of Brownian motion traits on the tree.

Table 2 The *P*-values for the three metrics (pTARS, pNoTO and pSSCD) for ten case studies (see Table 1).

Case study	Taxa	pTARS	pNoTO	pSSCD
1 Fernández-Mazuecos <i>et al.</i> (2013)	<i>Linaria</i>	0.651	0.224	0.078
2 Hardy & Otto (2014)	Heliconiini	0.399	0.226	0.032*
3 Janz <i>et al.</i> (2001)	Nymphalini	0.850	0.796	0.624
4 Johnson <i>et al.</i> (2009)	<i>Columbicola</i>	0.228	1.00*	0.215
5 Larkin <i>et al.</i> (2008)	<i>Andrena</i>	0.297	0.084	0.612
6 Mendlová & Šimková (2014)	<i>Cichlidogyrus/Scutogyrus</i>	0.384	0.041*	0.071
7 Schweizer <i>et al.</i> (2014)	Psittiformes	0.300	0.994*	0.998*
8 Šimková <i>et al.</i> (2006)	<i>Dactylogyrus</i>	0.041*	1.00*	0.528
9 Stireman (2005)	Tachinidae	0.997*	1.00*	0.541
10 Tripp & Manos (2008)	<i>Ruellia</i>	0.631	0.013*	0.002*

An asterisk (*) indicates statistical significance for a metric ($P < 0.05$ or $P > 0.95$). pTARS < 0.05 indicates that specialists have significantly shorter tip lengths than generalists, while pTARS > 0.95 indicates that specialists have significantly longer tip lengths than generalists, as determined by a Wilcoxon rank-sum test. pNoTO < 0.05 indicates that each inferred origin of specialization gives rise to fewer species than expected under a stochastic process, while pNoTO > 0.95 indicates that each origin gives rise to more species than expected. pSSCD < 0.05 indicates that specialists are more scattered across the phylogeny than expected under a stochastic process, while pSSCD > 0.95 indicates that they are more clustered together than expected.

(Fig. 3). Therefore, our results provide support for specialization being an evolutionary dead end in these two lineages, with specialists having higher extinction rates, lower speciation rates, higher rates of trait loss or being irreversible.

The remaining eight case studies did not provide clear evidence for an influence of specialization on lineage diversification or trait persistence. In seven studies (Janz *et al.*, 2001; Šimková *et al.*, 2006; Larkin *et al.*, 2008; Johnson *et al.*, 2009; Fernández-Mazuecos *et al.*, 2013; Hardy & Otto, 2014; Mendlová & Šimková, 2014), we could not reject the null models tested. For

these studies, we cannot reject the hypothesis that specialization is an evolutionary dead end, but also we cannot reject the null hypotheses that specialists are no more prone to higher extinction rates, lower speciation rates or higher trait loss than generalists. In contrast, in nectarivorous parrots (Schweizer *et al.*, 2014), the phylogenetic distribution of specialists was inconsistent with all models we tested, which may indicate that this case study fails to meet model assumptions, or there are too few origins of specialization (see Discussion).

Discussion

The influence of specialization on macroevolutionary potential has been debated for over a century. In recent decades, there has been a growing number of studies using phylogenetic analyses to test the hypothesis that specialization is an 'evolutionary dead end' because it decreases speciation rate, increases extinction rate or reduces capacity for future evolutionary change. Such studies, which encompass a range of taxa and types of specialization, have produced a wide range of conclusions. In this study, we wished to ask whether the variation in outcomes of phylogenetic studies of the macroevolutionary consequences of specialization was in part due to differences in analytical techniques employed. Therefore, we selected ten representative case studies and applied a consistent methodology to quantify the phylogenetic patterns of specialists and generalists, in an attempt to uncover any consistent differences in speciation, extinction and trait transition rates between specialist and generalist lineages. We did find evidence that specialization is an evolutionary dead end in some groups; however, our results are inconclusive for other groups as we could not reject the null models tested.

Our methods have a number of advantages, including: the use of phylogenetic metrics that allow detection and comparison of evolutionary patterns and processes; the use of null models of trait evolution to test for significant departures from expected distributions if specialization

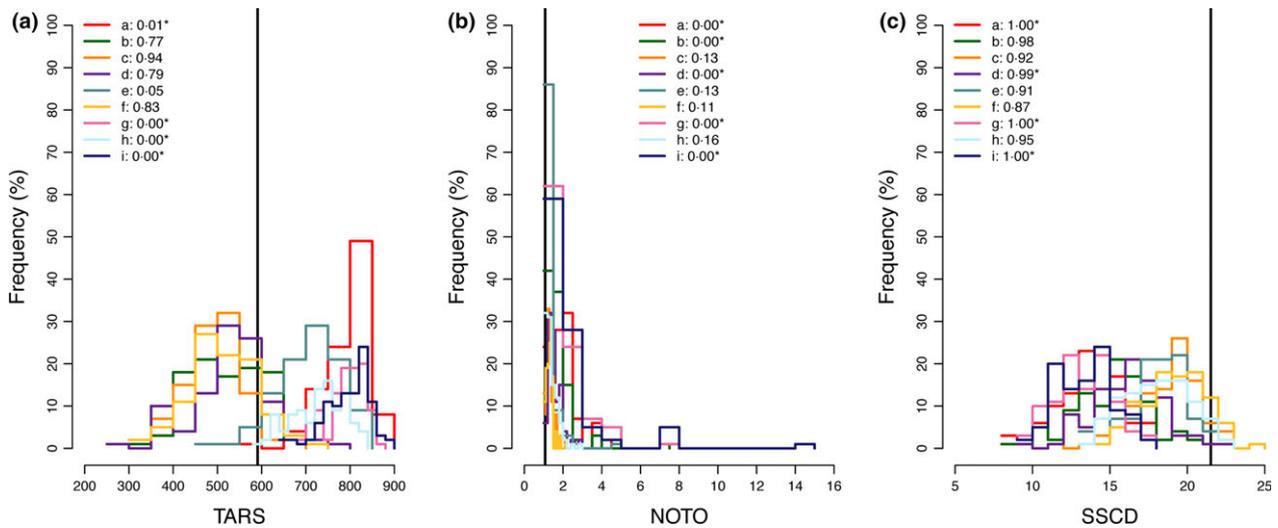


Fig. 2 The observed value of (a) TARS (tip length), (b) NoTO (number of tips per origin) and (c) SSCD (phylogenetic scatter) for hawkmoth-pollinated *Ruellia* (black vertical lines) compared to the distribution of values generated under nine alternative macroevolutionary models (coloured histograms indicated in the figure legend; for model parameters, see Table S3). The *P*-values in the figure legend correspond to the proportion of simulated phylogenies that have metric values less than or equal to the observed metric values. Following a Bonferroni correction, we reject a model as inconsistent with the observed data if $P \leq 0.01$ or $P \geq 0.99$ for any one of the three metrics (indicated by an asterisk).

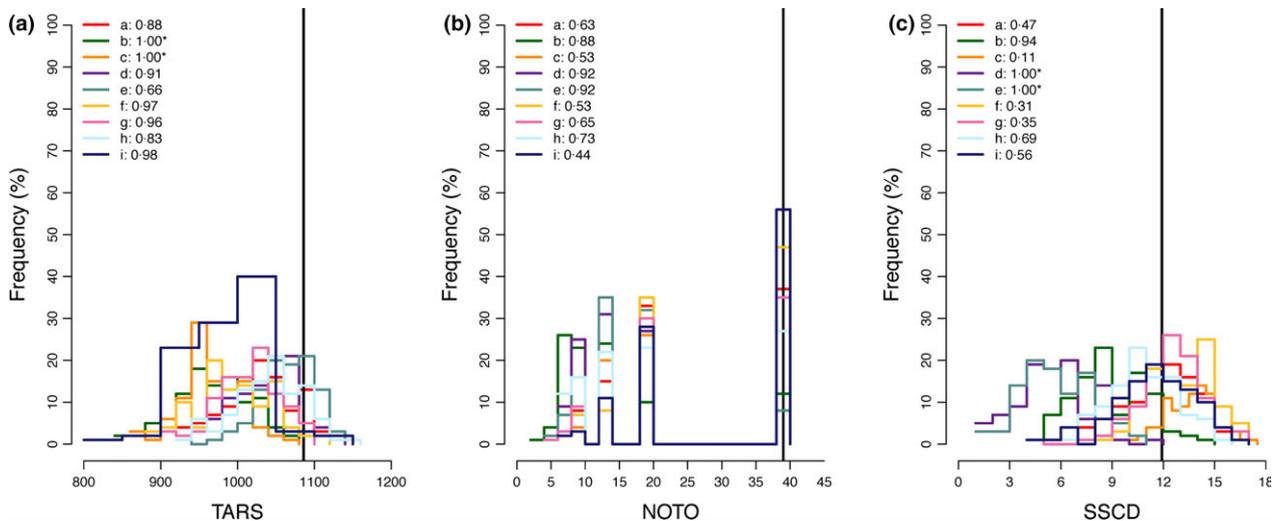


Fig. 3 The observed value of (a) TARS (tip length), (b) NoTO (number of tips per origin) and (c) SSCD (phylogenetic scatter) for host-specific Tachinid flies (black vertical lines) compared to the distribution of values generated under nine alternative macroevolutionary models (coloured histograms indicated in the figure legend; for model parameters, see Table S3). The *P*-values in the figure legend correspond to the proportion of simulated phylogenies that have metric values less than or equal to the observed metric values. Following a Bonferroni correction, we reject a model as inconsistent with the observed data if $P \leq 0.01$ or $P \geq 0.99$ for any one of the three metrics (indicated by an asterisk).

has no effect on lineage persistence or diversification; and tests of model adequacy that allow us to ask whether the observed pattern of specialists could have been produced under models of macroevolution where specialization is an evolutionary dead end. Our analyses overcome some identified weaknesses of previous studies. For

example, we do not directly rely on the reconstruction of ancestral character states, which has been cited as a potential weakness in studies of directionality of character state transitions (Stireman, 2005).

Our results are consistent with specialization being an evolutionary dead end in at least some cases. For two

data sets, we found that specialists had reduced capacity for diversification. The phylogenetic distribution of hawkmoth pollination in *Ruellia* (Tripp & Manos, 2008) is consistent with specialists having a reduced speciation rate, higher extinction rate or increased rate of trait loss. Narrow host specificity in Tachinid flies (Stireman, 2005) is consistent with a reduced speciation rate or an influence of specialization on trait transition rates (although the observed pattern could be predicted by two alternative models – high trait lability or irreversibility of the specialist trait). Both of these specialist traits have evolved repeatedly, potentially indicating that they are advantageous or neutral in the short-term but that specialist lineages suffer longer term disadvantages due to raised rates of extinction or trait loss.

In seven of the case studies, we could not reject models where specialization reduces the capacity for persistence or diversification, but neither could we reject the null model of no difference in macroevolutionary capacity between specialists and generalists. It may be that specialization has no influence on macroevolutionary dynamics in these clades. However, it is also possible that we had insufficient power to detect an underlying influence of specialization on speciation, extinction or trait transition rates. Statistical power of our tests to detect macroevolutionary mechanisms will depend on several factors. For example, the power of NoTO (number of tips per origin) to detect a model where specialists have low speciation rates increases with higher tip ratio bias (when the proportion of species in one binary state is large), but power to detect models of high extinction or lability using SSCD (phylogenetic scatter) increases with lower tip ratio bias (when the proportion of species in each binary state approaches 50 : 50) (Bromham *et al.*, 2016; Hua & Bromham, 2016). Incomplete phylogenetic sampling can also decrease statistical power to reject models, and it is possible that BiSSE parameter estimates used for developing alternative models are inappropriate considering the small size of most phylogenies (Davis *et al.*, 2013).

The metric-based approach to detecting macroevolutionary patterns has a number of advantages over common model-fitting approaches. Under a model-fitting approach, it is difficult to generate an appropriate null model against which to test observed patterns. Some methods circumvent this problem by approximating a null distribution using a chi-squared distribution, as is performed in likelihood ratio tests, but this can give misleading results in some data sets, leading to potentially high error rates (Boettiger *et al.*, 2012). Furthermore, model-fitting approaches are limited by the goodness of fit of the models being tested (Pennell *et al.*, 2015). If these approaches are not combined with a test of model adequacy, the best-fitting model may be accepted even if it is a very poor description of the underlying evolutionary process (Rabosky & Goldberg, 2015; Hua & Bromham, 2016). The metric-based approach illustrated here has a number of advantages

over the model-fitting approach. Backwards simulation makes the construction of a null model tractable, and we can directly estimate the statistical power of each summary statistic under many alternative models. We can also combine many different summary statistics to test deviation in various properties of the trait distribution on phylogenies. These features of the metric-based approach make it much easier to test the power of any test to detect a given macroevolutionary process (Bromham *et al.*, 2016). Power testing of model-based approaches like BiSSE for any given data set or model is more challenging, and BiSSE is known to have low power for data sets where the phylogeny has fewer than 300 tips and where < 10% of the tips have one character state (e.g. are specialists: Davis *et al.*, 2013). So while model-fitting approaches are a very useful way of exploring macroevolutionary dynamics, the usefulness of methods such as BiSSE is limited in many cases. None of the case studies explored here are large enough to be reliably investigated using BiSSE (Table 1), and yet we detected significant macroevolutionary patterns in two of these cases using the metric-based approaches.

It is possible that the failure to detect a consistent macroevolutionary pattern associated with specialization is due to the very heterogeneous traits and species to which the term ‘specialist’ is applied (Holmes, 1977; Ferry-Graham *et al.*, 2002). A broad range of traits have been used to test whether specialization is an evolutionary dead end, including traits as disparate as flowers designed to attract moths (Tripp & Manos, 2008), reduction in number of larval instars in cave beetles (Cieslak *et al.*, 2014), docile behaviour in spider colonies (Pruitt, 2013), viviparity (live birth) in lizards (Pincheira-Donoso *et al.*, 2013), host switching in aphids (Moran, 1988) and psammophily (living in sand) in springtails (D’Haese, 2000). It may be that the mechanisms underlying the macroevolutionary patterns vary between different kinds of specialization. Indeed, it might be surprising if such different ways of being considered specialists were to show common macroevolutionary patterns. Even if the underlying evolutionary mechanisms were the same, there is often no simple way of delineating the specialists from the generalists, and different definitions or thresholds can change the results of macroevolutionary analyses (see Results). The challenge is to find definitions of specialization that are sufficiently general to allow testing across the widest range of lineages, yet specific enough to allow targeting of shared macroevolutionary processes.

In one case study, we rejected all tested models as inconsistent with the observed phylogenetic distribution (Schweizer *et al.*, 2014). It is possible that this study does not meet the assumptions of the trait-dependent diversification models used (Goldberg & Iqic, 2012; Magnuson-Ford & Otto, 2012; Rabosky & Goldberg, 2015). Furthermore, there are only four inferred origins

of nectarivory in parrots in this data set (Schweizer *et al.*, 2014) so a general association between nectarivory and diversification may be difficult to ascertain due to lack of independent origins.

From our results we can draw four conclusions:

- 1 *Specialization can be an evolutionary dead end:* In two case studies, we can reject null models in favour of ‘dead end’ models where specialists have reduced speciation, increased extinction rates, high trait lability or irreversibility of specialization. In all but one of the remaining studies, we fail to reject models in which specialists have different speciation, extinction and trait transition rates from generalists; therefore, we cannot tell in these cases whether specialization is a dead end or not. A test on a wider range of case studies, especially those with larger data sets with a greater number of independent origins of the specialist trait, would be needed to determine how generalizable the hypothesis that specialization is an evolutionary dead end is.
- 2 *Detection of ‘dead end’ patterns is partly dependent on definition of specialization:* In their classic review of the macroevolutionary consequences of specialization, Futuyma & Moreno (1988) stated that ‘specialization must lie in the eye of the beholder’, highlighting the context dependent and variable ways that ecological specialization is defined (Berenbaum, 1996; Devictor *et al.*, 2010; Barrett & Heil, 2012). We have tested a range of types of specialization – host specificity, diet breadth and pollinator range – but we have used the definition of specialization provided in the original publication. In some cases, degree of specialization is a continuous or nonbinary trait, so the cut-off when defining specialists and generalists can be somewhat arbitrary. As the particular definition of specialization changes, or the way that species are categorized into binary groups of ‘specialists’ and ‘generalists’ varies between studies, the proportion and phylogenetic distribution of specialists and generalists will also vary. Therefore, the definition of specialization adopted may influence the detection of macroevolutionary patterns. In several case studies, the original publication provided multiple trait states for the trait of interest, such as different pollinator types (Tripp & Manos, 2008) and different definitions of specialization at local and global scales (Šimková *et al.*, 2006; Mendlová & Šimková, 2014). In these cases, we tested alternative schemes for dividing traits into binary categories of specialist and generalist (see Appendix S1). In each of these cases, only one of the alternative definitions of specialization resulted in a phylogenetic distribution of specialists that was significantly different from a null distribution. It may be that a similar result would be achieved in other data sets if specialization was redefined. Therefore, we can conclude that definitions can make a difference to whether significant macroevolutionary patterns are detected or not.
- 3 *Observed patterns must be tested against appropriate null models:* Detecting cases of evolutionary dead ends from phylogenies requires not only a way of quantifying the distribution of specialists but also generating null distributions against which to test observed patterns. This is particularly important given that relevant phylogenetic data sets are often relatively small, and the number of origins of specializations relatively few, reducing the power to determine when a ‘tippy’ distribution of specialists represents the trace of an interesting macroevolutionary phenomenon and when it is simply the result of the random distribution of a small number of arbitrarily defined species. Phylogenetic null models are complicated by the need to compare the observed distribution to an expected phylogenetic distribution conditioned on the same size and topology with an equivalent number of specialists and generalist species, to ask whether the observed number and distribution of specialists could have been produced under a model where specialization has no effect on persistence or diversification. We have shown how the use of null models can highlight cases where an apparent macroevolutionary pattern does not have significant statistical support, because the same pattern could have been produced by processes other than those of interest. Our use of phylogenetic metrics quantifying the tip length of specialist taxa, the relative size and depth of clades of specialists and the degree of phylogenetic scatter provides a tractable way of producing null models, because we can generate null distributions of these metrics by evolving a stochastic trait along the same topology. We can then compare the observed metric values to the null distribution to detect significant departures. These metrics are all available in the R package ‘PHYLOMETRICS’ (Hua & Bromham, 2016), which also allows users to specify their own metrics and develop alternative null model tests.
- 4 *Tests of model inadequacy allow poorly fitting models to be rejected:* Classical model-fitting approaches allow researchers to compare the fit of different models to the observed patterns in their data. But the best-fitting model may be a relatively poor description of the data. Rather than being forced to accept the best-fitting model, we need a way of asking whether any of the models provide a reasonable description of the data (Boettiger *et al.*, 2012). To do this, we need to ask what the chances are that the observed pattern could be produced under a given macroevolutionary model. Standard ‘forward’ simulation of phylogenetic data, where a model of speciation, extinction and trait transition is used to ‘grow’ a tree until the desired number of tips is reached, is inefficient for this purpose, because relatively few of the simulated trees will match the observed data in the precise number of specialist and generalist tips. Instead, we employ ‘backward simulation’ that fixes the number of specialists and generalists and uses a macroevolu-

tionary birth–death model to construct possible histories of speciation, extinction and trait transition that could have given rise to these observed species. If none of the models used to simulate the data lead to trees that have similar metric scores to the observed data, then we can conclude that none of the models provide an adequate description of the data. Therefore, it is possible to reject all models, as occurred for one case study (Schweizer *et al.*, 2014).

Conclusion

By applying consistent phylogenetic tests to ten diverse case studies, we find evidence that specialization can significantly reduce the macroevolutionary potential of lineages. Reduced speciation, increased extinction and high trait loss can lead to ‘evolutionary dead ends’ where specialists evolve often but have reduced capacity to persist or diversify compared to related generalist species. However, in the majority of case studies, we could not reject either the null model or the dead end model, indicating generally low power in these data sets to discriminate macroevolutionary hypotheses for the evolution of specialization. In one case study, all models tested were rejected, illustrating the need for tests of model inadequacy when testing macroevolutionary hypotheses from phylogenetic patterns. The methods we present, based on phylogenetic metrics describing tip length, phylogenetic scatter and clade size, provide tractable ways to measure patterns of trait distribution on phylogenies and generate null distributions under a range of macroevolutionary models. This provides a convenient way of assessing statistical power, which may be improved by larger data sets with greater number of independent origins of specialization. Our results also suggest that investigating the effect of definitions of specialization on the observed phylogenetic patterns and inferred macroevolutionary mechanisms is an important line of future enquiry.

Conflict of interest

The authors declared that they have no conflict of interest.

References

- Agnarsson, I., Avilés, L., Coddington, J.A. & Maddison, W.P. 2006. Sociality in theridiid spiders: repeated origins of an evolutionary dead end. *Evolution* **60**: 2342–2351.
- Anacker, B.L., Whittall, J.B., Goldberg, E.E. & Harrison, S.P. 2011. Origins and consequences of serpentine endemism in the California flora. *Evolution* **65**: 365–376.
- Armbruster, W.S. 2014. Floral specialization and angiosperm diversity: phenotypic divergence, fitness trade-offs and realized pollination accuracy. *AoB Plants* **6**: plu003.
- Baringhaus, L. & Franz, C. 2004. On a new multivariate two-sample test. *J. Multivar. Anal.* **88**: 190–206.
- Barrett, L.G. & Heil, M. 2012. Unifying concepts and mechanisms in the specificity of plant–enemy interactions. *Trends Plant Sci.* **17**: 282–292.
- Berenbaum, M.R. 1996. Introduction to the symposium: on the evolution of specialization. *Am. Nat.* **148**: S78–S83.
- Boettiger, C., Coop, G. & Ralph, P. 2012. Is your phylogeny informative? Measuring the power of comparative methods. *Evolution* **66**: 2240–2251.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D. *et al.* 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**: e1003537.
- Bromham, L., Hua, X. & Cardillo, M. 2016. Detecting macroevolutionary self-destruction from phylogenies. *Syst. Biol.* **65**: 109–127.
- Buckling, A., Wills, M.A. & Colegrave, N. 2003. Adaptation limits diversification of experimental bacterial populations. *Science* **302**: 2107–2109.
- Cieslak, A., Fresneda, J. & Ribera, I. 2014. Life-history specialization was not an evolutionary dead-end in Pyrenean cave beetles. *Proc. R. Soc. Lond. B: Biol. Sci.* **281**: 20132978.
- Colles, A., Liow, L.H. & Prinzing, A. 2009. Are specialists at risk under environmental change? Neoeological, paleoecological and phylogenetic approaches. *Ecol. Lett.* **12**: 849–863.
- Cope, E.D. 1896. *The Primary Factors of Organic Evolution*. The Open Court Publishing Company, Chicago.
- Davis, M.P., Midford, P.E. & Maddison, W.P. 2013. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC Evol. Biol.* **13**: 38.
- Devictor, V., Clavel, J., Julliard, R., Lavergne, S., Mouillot, D., Thuiller, W. *et al.* 2010. Defining and measuring ecological specialization. *J. Appl. Ecol.* **47**: 15–25.
- D’Haese, C. 2000. Is Psammophily an evolutionary dead end? A phylogenetic test in the genus *Willemia* (Collembola: Hypogastruridae). *Cladistics* **16**: 255–273.
- Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792–1797.
- Fernández-Mazuecos, M., Blanco-Pastor, J.L., Gómez, J.M. & Vargas, P. 2013. Corolla morphology influences diversification rates in bifid toadflaxes (*Linaria* sect. *Versicolores*). *Ann. Bot.* **112**: 1705–1722.
- Ferry-Graham, L.A., Bolnick, D.I. & Wainwright, P.C. 2002. Using functional morphology to examine the ecology and evolution of specialization. *Integr. Comp. Biol.* **42**: 265–277.
- FitzJohn, R.G. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol. Evol.* **3**: 1084–1092.
- FitzJohn, R.G., Maddison, W.P. & Otto, S.P. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* **58**: 595–611.
- Forister, M.L., Dyer, L.A., Singer, M.S., Stireman, J.O. & Lill, J.T. 2012. Revisiting the evolution of ecological specialization, with emphasis on insect–plant interactions. *Ecology* **93**: 981–991.
- Franz, C. 2014. cramer: Multivariate nonparametric Cramer-Test for the two-sample-problem. R package version 0.9-1. <http://CRAN.R-project.org/package=cramer>
- Fritz, S.A. & Purvis, A. 2010. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conserv. Biol.* **24**: 1042–1051.
- Futuyma, D.J. & Moreno, G. 1988. The evolution of ecological specialization. *Annu. Rev. Ecol. Syst.* **19**: 207–233.

- Goldberg, E.E. & Igic, B. 2012. Tempo and mode in plant breeding system evolution. *Evolution* **66**: 3701–3709.
- Griffith, T. & Sultan, S.E. 2012. Field-based insights to the evolution of specialization: plasticity and fitness across habitats in a specialist/generalist species pair. *Ecol. Evol.* **2**: 778–791.
- Haldane, J.B.S. 1951. *Everything has a History*. Allen & Unwin, London.
- Hardy, N.B. & Otto, S.P. 2014. Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. *Proc. R. Soc. Lond. B: Biol. Sci.* **281**: 20132960.
- Holmes, E.B. 1977. Is specialization a dead end? *Am. Nat.* **111**: 1021–1026.
- Hua, X. & Bromham, L. 2016. Phylometrics: an R package for detecting macroevolutionary patterns, phylogenetic metrics and backward tree simulation under a binary state birth-death model. *Methods Ecol. Evol.* doi:10.1111/2041-210X.12531.
- Janz, N., Nyblom, K. & Nylin, S. 2001. Evolutionary dynamics of host-plant specialization: a case study of the tribe Nymphalini. *Evolution* **55**: 783–796.
- Johnson, K.P., Malenke, J.R. & Clayton, D.H. 2009. Competition promotes the evolution of host generalists in obligate parasites. *Proc. R. Soc. Lond. B: Biol. Sci.* **276**: 3921–3926.
- Kassen, R. 2002. The experimental evolution of specialists, generalists, and the maintenance of diversity. *J. Evol. Biol.* **15**: 173–190.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S. *et al.* 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.
- Kelley, S.T. & Farrell, B.D. 1998. Is specialization a dead end? The phylogeny of host use in dendroctonus bark beetles (*Scolytidae*). *Evolution* **52**: 1731–1743.
- Larkin, L.L., Neff, J.L. & Simpson, B.B. 2008. The evolution of a pollen diet: host choice and diet breadth of *Andrena* bees (Hymenoptera: Andrenidae). *Apidologie* **39**: 133–145.
- Maddison, W.P., Midford, P.E. & Otto, S.P. 2007. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* **56**: 701–710.
- Magnuson-Ford, K. & Otto, S.P. 2012. Linking the investigations of character evolution and species diversification. *Am. Nat.* **180**: 225–245.
- McBride, C.S. 2007. Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proc. Natl. Acad. Sci. USA* **104**: 4996–5001.
- Mendlová, M. & Šimková, A. 2014. Evolution of host specificity in monogeneans parasitizing African cichlid fish. *Parasit. Vectors* **7**: 69.
- Moran, N.A. 1988. The evolution of host-plant alternation in aphids: evidence for specialization as a dead end. *Am. Nat.* **132**: 681–706.
- Nosil, P. & Mooers, A.Ø. 2005. Testing hypotheses about ecological specialization using phylogenetic trees. *Evolution* **59**: 2256–2263.
- Pennell, M.W., FitzJohn, R.G., Cornwell, W.K. & Harmon, L.J. 2015. Model adequacy and the macroevolution of angiosperm functional traits. *Am. Nat.* **186**: E33–E50.
- Pincheira-Donoso, D., Tregenza, T., Witt, M.J. & Hodgson, D.J. 2013. The evolution of viviparity opens opportunities for lizard radiation but drives it into a climatic cul-de-sac. *Glob. Ecol. Biogeogr.* **22**: 857–867.
- Pruitt, J.N. 2013. A real-time eco-evolutionary dead-end strategy is mediated by the traits of lineage progenitors and interactions with colony invaders. *Ecol. Lett.* **16**: 879–886.
- Pyron, R.A. & Burbrink, F.T. 2014. Early origin of viviparity and multiple reversions to oviparity in squamate reptiles. *Ecol. Lett.* **17**: 13–21.
- R Development Core Team 2015. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria.
- Rabosky, D.L. & Goldberg, E.E. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Syst. Biol.* **64**: 340–355.
- Rambaut, A. & Drummond, A.J. 2012. Tracer v1.5. from <http://beast.bio.ed.ac.uk/Tracer>.
- Ravigné, V., Dieckmann, U. & Olivieri, I. 2009. Live where you thrive: joint evolution of habitat choice and local adaptation facilitates specialization and promotes diversity. *Am. Nat.* **174**: E141–E169.
- Schweizer, M., Guntert, M., Seehausen, O., Leuenberger, C. & Hertwig, S.T. 2014. Parallel adaptations to nectarivory in parrots, key innovations and the diversification of the Loricinae. *Ecol. Evol.* **4**: 2867–2883.
- Šimková, A., Verneau, O., Gelnar, M. & Morand, S. 2006. Specificity and specialization of congeneric monogeneans parasitizing cyprinid fish. *Evolution* **60**: 1023–1037.
- Stephens, P.R. & Wiens, J.J. 2003. Ecological diversification and phylogeny of emydid turtles. *Biol. J. Linn. Soc.* **79**: 577–610.
- Stireman, J.O. 2005. The evolution of generalization? Parasitoid flies and the perils of inferring host range evolution from phylogenies. *J. Evol. Biol.* **18**: 325–336.
- Tripp, E.A. & Manos, P.S. 2008. Is floral specialization an evolutionary dead-end? Pollination system transitions in *Ruellia* (Acanthaceae). *Evolution* **62**: 1712–1737.
- Vamosi, J.C., Armbruster, W.S. & Renner, S.S. 2014. Evolutionary ecology of specialization: insights from phylogenetic analysis. *Proc. R. Soc. Lond. B: Biol. Sci.* **281**: 20142004.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1 Supplementary data collection and methods.

Table S1 Results for FPD metric.

Table S2 List of alternative macroevolutionary models.

Table S3 Model parameter estimates for each case study.

Figure S1 Observed FPD value compared to FPD under simulated models for hawkmoth-pollinated *Ruellia* plants.

Figure S2 Observed FPD value compared to FPD under simulated models for hosts specific Tachinid flies.

Data deposited at Dryad: doi: 10.5061/dryad.s6217

Received 30 November 2015; revised 29 February 2016; accepted 12 March 2016