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

Contents

Supplementary data collection and methods

In this supplementary text, we outline our data collection and tree estimation processes for the ten case studies used in this paper (Table 1). The studies are discussed in alphabetical order by the first author's surname. All of our estimated phylogenies are available on DataDryad. Tables and figures listed in square brackets correspond to tables and figures in the original publication.

Fernández-Mazuecos et al. (2013)

We collected accession numbers for one nuclear (ITS) and three plastid DNA regions (*rpl32*-*trn*L^{UAG}, *trn*K-*mat*K and *trn*S-*trn*G) from the supplementary material of Fernández-Mazuecos *et al.* (2013) [Data Table S1]. We collected data for 29 species of herbaceous plants in the genus *Linaria*, specifically the Bifid Toadflaxes from *Linaria* sect. Versicolores, and one outgroup, *Antirrhinum graniticum.* Initially, we estimated phylogenetic trees separately for ITS and plastid DNA regions, because separate topologies (species relationships) were reported in the original publication depending on the type of DNA used to estimate the phylogeny [Figure 3A and 3B]. We used BEAST v2.1.3 (Bouckaert *et al.*, 2014) to estimate ultrametric branch lengths on the constrained topologies using a GTR + Γ model of nucleotide evolution, separate partitions for each gene, a relaxed log-normal clock, and birth-death priors. During the tree estimation process, the phylogeny based on ITS DNA had high effective sample size for all parameters, however the phylogeny based on plastid DNA failed to converge given the constrained topology. Therefore, we only used the phylogeny based on ITS DNA for all further analyses for this case study.

In this case study, specialization is defined in terms of corolla morphology, with specialists having narrow-tubed flowers and generalists having wide-tubed flowers.

We obtained the corolla morphology for each of the 29 species by directly recording the trait data from Fernández-Mazuecos *et al.* (2013) [Table 1 – see under "morphological type"; Table S3]. As in the original publication, we assigned species with morphological type III as specialists and species with morphological type I or II as generalists due to morphometric similarity between these two types.

Hardy & Otto (2014)

Since time-calibrated ultrametric trees estimated in BEAST for this case study were available from DataDryad (http://doi.org/10.5061/dryad.bv689), we used the published Maximum Clade Credibility (MCC) tree for our analyses. The original authors estimated the trees in BEAST v.1.7.5 (Drummond & Rambaut, 2007) using a HKY model of nucleotide substitution, birth-death prior and an uncorrelated relaxed log-normal clock with the topology fixed on a Maximum Likelihood estimate of the tree previously performed in RAxML v7.0.4 (Stamatakis, 2006). The MCC tree we obtained from the DataDryad package had 63 species of butterflies in the tribe Heliconiini (subfamily: Heliconiinae), however we trimmed the phylogeny to contain only the 33 species for which trait data were available.

We used the same definition of specialization as the original authors, where specialists are diet-specific and feed on few hosts and generalists feed on many. The study's authors assigned specialists and generalists based on a quantitative measure of host breadth (phylogenetic diversity: PD), with specialists identified as species that feed on hosts with a PD < 0.1. We obtained the trait data for each species from the same DataDryad package as above.

Janz et al. (2001)

The phylogeny was not made available through the original publication, so we collected the Genbank accession numbers for one mitochondrial gene (*nd1*) and one nuclear gene (*wingless*) from Nylin *et al.* (2001) for 25 species in the butterfly tribe Nymphalini and one outgroup, *Argynnis paphia.* This list of 25 species matched the list of species in Janz *et al.* (2001) [Figure 2], although there have since been several name changes for species so we adjusted species names based on Genbank at the time of writing this paper. To obtain the topology with which to constrain our tree estimation process, we directly read the topology off the phylogeny in Janz *et al.* (2001) [Figure 2]. With a constrained topology, we then used BEAST v2.1.3 (Bouckaert *et al.*, 2014) to estimate ultrametric branch lengths. The trees did not converge using our default parameter settings, so we simplified the model by using a HKY + Γ model of nucleotide evolution with separate partitions for each gene, a relaxed log-normal clock, and a coalescent prior.

We also collected trait data from Janz *et al.* (2001) [Figure 2]. In this case study, specialization is defined as monophagy, whereby specialist species feed only on plants within one host family. Therefore any species that fed on one host family was considered a specialist and species that fed on more than one host family were considered generalists.

Johnson et al. (2009)

No phylogeny was available through the original publication so we collected Genbank accession numbers for one nuclear gene (EF-1α) and two mitochondrial genes (12S and COI) from Johnson *et al.* (2007) [Appendix A] for species in the genus *Columbicola.* Although there were 46 species listed in the phylogeny of

Johnson *et al.* (2009), we were unable to locate DNA sequences for several species from the publication or on Genbank, therefore we had to exclude these species from our analysis: *Columbicola rodmani*, *Columbicola harbisoni*, *Columbicola wombeyi*, *Columbicola koopae*, *Columbicola eowilsoni*, *Columbicola smithae*, *Columbicola waiteae*, and *Columbicola masoni*. In the phylogeny of Johnson *et al.* (2009), there were often multiple samples of a single species represented so to prevent biasing from infraspecific taxa, we chose one representative from each species using a random number generator. After these modifications, our species list consisted of 38 species within the genus *Columbicola* and one outgroup, *Oxyliperurus chiniri*.

The phylogeny in Johnson *et al.* (2009) was a strict consensus parsimony tree, with several polytomies that we resolved using a random number generator to generate a bifurcating topology. However, when we attempted to estimate ultrametric branch lengths on a constrained topology, the trees failed to converge. When we allowed the topology to vary in BEAST v2.1.3 (Bouckaert *et al.*, 2014), and estimated trees using a GTR + Γ model of nucleotide evolution with separate partitions for each gene and codon position, a relaxed log-normal clock, and a birth-death prior, the estimation process converged with high effective sample size for all parameters. Therefore, for this case study, we estimated the branch lengths and topology from the data such that the topology differs from the phylogeny in Johnson et al. (2009).

In this case study, specialization is defined as parasitizing a single host species, while generalists are defined as parasitizing more than one host species. We collected trait data for each species on our phylogeny by directly recording the data from Johnson et al. (2009) [Figure 1].

Larkin et al. (2008)

An ultrametric phylogeny was not available for this case study, so we collected Genbank accession numbers for two mitochondrial genes (COI and COII including the intervening tRNA-leucine) and one nuclear gene (EF-1α) from Larkin *et al.* (2006) [Table 1] for 35 species of bees in the genus *Andrena* and one outgroup, *Ancylandrena larreae.* One species in the phylogeny in Larkin et al. (2008), *Andrena hirticincta*, did not have accession numbers listed in Larkin *et al.* (2006), however we were able to obtain accession numbers for this species directly from Genbank for EF-1α (accession number: EF601180). The 35 species for which we collected sequences were a subset of the species listed in the phylogeny of Larkin *et al.* (2006), as we only used the species with available trait data in Larkin *et al.* (2008) [Figure 2]. To obtain the topology with which to constrain our tree estimation process, we directly used the topology from in Larkin *et al.* (2008) [Figure 2]. With a constrained topology, we then used BEAST v2.1.3 (Bouckaert *et al.*, 2014) to estimate ultrametric branch lengths. The trees did not converge using our default parameter settings, so we simplified the model by using a HKY + Γ model of nucleotide evolution, separate partitions for each gene, a relaxed log-normal clock, and a yule prior.

We obtained trait data for each species from Larkin et al. (2008) [Figure 2]. In this case study, specialization is defined as diet-breadth. Larkin et al. (2008) defined specialist bees as having more than 90% of pollen gains found on a bee species deriving from a single plant tribe, while generalists collected pollen from multiple tribes.

Mendlová et al. (2014)

We collected Genbank accession numbers for two partial ribosomal RNA genes (SSU and LSU) and the entire ITS1 nuclear gene region from Mendlová *et al.* (2012) for 25 species in the genus *Cichlidogyrus* and 3 species in the genus *Scutogyrus*. For two of the species listed in the phylogeny of Mendlová and Šimková (2014), *Cichlidogyrus dracolemma* and *Cichlidogyrus nageus*, we could not locate accession numbers listed in Mendlová *et al.* (2012) or independently on Genbank, so we excluded these two species from our analysis. We used the topology from Mendlová and Šimková (2014) [Figure 2] to constrain our phylogeny. We then used BEAST v2.1.3 (Bouckaert *et al.*, 2014) to estimate ultrametric branch lengths on the given topology, using a GTR + Γ model of nucleotide evolution, separate partitions for each gene, a relaxed log-normal clock, and birth-death priors.

In this case study, specialization is defined as host specificity and we collected trait data for each species from the supplementary information of Mendlová and Šimková (2014). The study's authors considered two separate definitions of specialization, by considering host-parasite records from Senegal only (local specificity), and considering all host-parasite records for African cichlids (global specificity). At both the local and global scale, they categorized species according to their strictness of specificity on a scale from 1 to 4: (1) Strict specialist (parasitizes only one host species); (2) Intermediate specialist (parasitizes two or more congeneric host species); (3) Intermediate generalist (parasitizing noncongeneric species of the tribe Tilapiini); (4) Generalist (parasitizing noncongeneric species of at least two different tribes. For our analyses, we needed to convert trait data to binary groups of "specialists" and "generalists", so did this in several ways to take into account the original authors' varied definitions of specialization. We therefore defined specialization in the following ways:

1. Strict and intermediate specialists at the global level
2. Strict specialists at the global level
3. Strict and intermediate specialists at the local level
4. Strict specialists at the local level

Two species, *Chichlidogyrus sclerosus* and *Scutogyrus minus*, did not have local host specificity data so these species were excluded when specialists were defined at the local level (variations 3 & 4), but included when defined at the global level (variations 1 & 2). We found that the phylogenetic distribution of specialists was only significantly different from a stochastic process when strict and intermediate specialists were defined at the global level (variation 1), so we pursued this definition of specialization for further analysis on this case study (i.e. comparing to alternative macroevolutionary models). For the other definitions of specialization (variations 2-4), none of the metric values were significantly different than expected by chance (results not shown).

Schweizer et al. (2014)

We collected Genbank accession numbers for three partial nuclear genes (c-mos, RAG-1 and Zenk) and one mitochondrial gene (ND2) from the supplementary material of Schweizer *et al.* (2014) for 78 parrot species within the order Psittaciformes and one outgroup, *Falco tinnunculus.* We used the topology given in Schweizer *et al.* (2014) [Figure 4] and randomly resolved polytomies to create a bifurcating topology using a random number generator. We then used BEAST v2.1.3 (Bouckaert *et al.*, 2014) to estimate ultrametric branch lengths on the given topology. The trees did not converge using our default parameter settings, so we modified the model by using a HKY + Γ model of nucleotide evolution with separate partitions for each gene, a relaxed log-normal clock, and a coalescent prior.

In this case study, specialization is defined as diet specificity, with specialists feeding exclusively on nectar. Generalists were defined as any species not feeding exclusively on nectar. We collected trait data for each species from Schweizer *et al.* (2014) [Figure 4].

Šimková et al. (2006)

We collected Genbank accession numbers for partial 18S rDNA and the entire nuclear ITS1 gene region for 51 species in the genus *Dactylogyrus* and two outgroups, *Thaparocleidus vistulensis* and *Clediodiscus pricei*, from (Šimková *et al.*, 2004) [Table 1]. We used the topology from Šimková *et al.* (2006) [Figure 3]. We then used BEAST v2.1.3 (Bouckaert *et al.*, 2014) to estimate ultrametric branch lengths on the given topology, using a GTR + Γ model of nucleotide evolution with separate partitions for each gene, a relaxed log-normal clock, and a birth-death prior.

We collected trait data for each species from Šimková *et al.* (2006) [Appendix 2].In this case study, specialization is defined as host specificity. The study's authors considered two separate definitions of specialization, by considering experimental host-parasite records within a local area in the Morava River basin in the Czech Republic (local specificity) and considering all host-parasite records for *Dactylogyrus* species (global specificity). Local host specificity was described in binary with specialists parasitizing one host species within the investigated area and generalists parasitizing more than one host species. Global host specificity was described on a scale from 1-5: (1) Specialist on a single host species; (2) Specialist on species of one genus; (3) Generalist living on related species of the same clade; (4) Generalist living on species of the same subfamily; (5) Generalist living on unrelated species. For our analyses, we needed to convert global specificity into binary groups of "specialists" and "generalists", but we wished to account for the different definitions of specialization given by the original authors. Therefore, we performed three separate analyses where we defined specialists as following:

1. Specialists were defined at the local level
2. Specialists were defined at the global level in categories 1 and 2
3. Specialists were defined at the global level in category 1 only

We found that the phylogenetic distribution of specialists was only significantly different from a stochastic process when specialists were defined at the local level (variation 1), so we pursued this definition of specialization for further analysis on this case study (i.e. comparing to alternative macroevolutionary models). For the other definitions of specialization (variations 2 and 3), none of the metric values were significantly different than expected by chance (results not shown).

Stireman (2005)

We collected Genbank accession numbers for EF-1α and 28S rRNA genes from Stireman (2002) [Table 1] for 49 species of Tachinid flies in the subfamily Exoristinae and one outgroup, *Metoposarcophaga sp.* In the list of species given in Stireman (2002) [Table 1], there were often multiple samples of one species. To prevent biasing from infraspecific taxa, we chose the sample with the most sequence data (i.e. where accession numbers for both genes were available instead of just one gene) to represent the species. If multiple samples had equal genetic information available, we used a random number generator to select a sample to represent the species. To obtain the topology with which to constrain our tree estimation process, we directly read the topology off the phylogeny in Stireman (2005) [Figure 3]. We then used BEAST v2.1.3 (Bouckaert *et al.*, 2014) to estimate ultrametric branch lengths on the given topology, using a GTR + Γ model of nucleotide evolution with separate partitions for each gene and codon position, a relaxed log-normal clock, and a birth-death prior.

In this case study, specialization is defined as host specificity, with specialists defined as having three or fewer host families and generalists as using hosts in five or more families. We collected the trait data for each species by directly recording from Stireman (2005) [Figure 3].

Tripp & Manos (2008)

While no ultrametric tree was available for this case study, we were able to obtain an alignment of the nuclear ribosomal ITS + 5.8S region and the chloroplast *trnG-trnR* region for 115 species in the genus *Ruellia* and one outgroup, *Sanchezia speciosa*, from TreeBASE under study ID S1995. We did not make any alterations to this list of species or the alignment. We used the topology in Tripp and Manos (2008) [Figure 2]. We then used BEAST v2.1.3 (Bouckaert *et al.*, 2014) to estimate ultrametric branch lengths on the given topology, using a GTR + Γ model of nucleotide evolution with separate partitions for each gene and codon position, a relaxed log-normal clock, and a birth-death prior.

We obtained the trait data for each species from Tripp and Manos (2008) [Figure 2]. In this case study, trait corresponded to corolla colour, which was assigned to one of four states: purple, red, white or yellow/green. Generalists were defined as species with purple corollas (52 species) because they are bee and insect pollinated, and accessible to a wide range of pollinators. The remaining corolla colours were assigned as different "types" of specialists with red (33 species), white (15 species) and yellow/green (15 species) corollas corresponding to hummingbird, hawkmoth and bat pollinated flowers respectively.

We wished to account for these different definitions of specialization, so we performed four separate tests, altering our definition of specialization in the following ways:

1. All red, white and yellow/green corolla species are specialists
2. Only red corolla species are specialists (hummingbird-pollinated plants)
3. Only white corolla species are specialists (hawkmoth-pollinated plants)
4. Only yellow/green corolla species are specialists (bat-pollinated plants)

In each case above, we assigned species with purple corollas as generalists and pruned any species that were neither purple nor the specialist being considered were from our phylogeny prior to analysis. This reduced the sample size from 115 species to 85, 67 and 67 species in cases 2-4 respectively. We adjusted the sampling fraction accordingly as species were pruned from the phylogeny. During our analyses, we found that the phylogenetic distribution of specialists was only significantly different from a stochastic process specialists were defined as hawkmoth-pollinated plants (variation 3), so we pursued this definition of specialization for further analysis on this case study (i.e. comparing to alternative macroevolutionary models). For the other definitions of specialization (variations 1, 2 and 4), none of the metric values were significantly different than expected by chance (results not shown).

References

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 Comp. Biol.* **10:** e1003537.

Drummond, A.J. & Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7:** 214.

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.

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.

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.

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., Reed, D.L., Hammond Parker, S.L., Kim, D. & Clayton, D.H. 2007. Phylogenetic analysis of nuclear and mitochondrial genes supports species groups for *Columbicola* (Insecta: Phthiraptera). *Mol. Phylogen. Evol.* **45:** 506-518.

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.

Larkin, L.L., Neff, J.L. & Simpson, B.B. 2006. Phylogeny of the Callandrena subgenus of Andrena (Hymenoptera: Andrenidae) based on mitochondrial and nuclear DNA data: polyphyly and convergent evolution. *Mol. Phylogen. Evol.* **38:** 330-343.

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.

Mendlová, M., Desdevises, Y., Civanova, K., Pariselle, A. & Šimková, A. 2012. Monogeneans of West African cichlid fish: evolution and cophylogenetic interactions. *PLoS One* **7:** e37268.

Mendlová, M. & Šimková, A. 2014. Evolution of host specificity in monogeneans parasitizing African cichlid fish. *Parasit. Vectors* **7:** 69.

Nylin, S., Nymblom, K., Ronquist, F., Janz, N., Belicek, J. & Källersjö, M. 2001. Phylogeny of Polygonia, Nymphalis and related butterflies (Lepidoptera: Nymphalidae): a total-evidence analysis. *Zool. J. Linn. Soc.* **132:** 441-468.

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 Loriinae. *Ecol. Evol.* **4:** 2867-2883.

Šimková, A., Morand, S., Jobet, E., Gelnar, M. & Verneau, O. 2004. Molecular phylogeny of congeneric monogenean parasites (*Dactylogyrus*): a case of intrahost speciation. *Evolution* **58:** 1001.

Šimková, A., Verneau, O., Gelnar, M. & Morand, S. 2006. Specificity and specialization of congeneric monogeneans parasitizing cyprinid fish. *Evolution* **60:** 1023-1037.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688-2690.

Stireman, J.O. 2002. Phylogenetic relationships of tachinid flies in subfamily Exoristinae (Tachinidae:Diptera) based on 28S rDNA and elongation factor-1α. *Syst. Entomol.* **27:** 409-435.

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.

Table S1: The results for the Fritz & Purvis D statistic for each of the ten case studies (Fritz & Purvis, 2010). FPD scales the difference between the observed SSCD value and expected SSCD value under Brownian motion by the difference between SSCD under a random tip-shuffle method and SSCD under Brownian motion (Fritz & Purvis, 2010). An absolute value of FPD = 1 means that the observed trait is randomly distributed across tips and FPD = 0 means the trait evolves as expected under Brownian motion. If a specialist trait is scattered across the phylogeny, we would expect FPD to be significantly positive, i.e. greater than 0.5.

	Case study	**FPD**
1	Fernández-Mazuecos et al. (2013)	0.905
2	Hardy & Otto (2014)	0.861
3	Janz et al. (2001)	-0.244
4	Johnson et al. (2009)	0.361
5	Larkin et al. (2008)	-0.152
6	Mendlová et al. (2014)	0.887
7	Schweizer et al. (2014)	-1.564
8	Šimková et al. (2006)	-0.141
9	Stireman (2005)	-0.022
10	Tripp & Manos (2008)	1.099

Table S2: List of models used to estimate alternative parameter sets, generated Binary State Speciation and Extinction (BiSSE) model (Maddison *et al.*, 2007; FitzJohn *et al.*, 2009).

Model	**Constraints**	
a)	All six parameters are free to vary	
b)	Speciation rates, extinction rates and transition rates are equal between specialists and generalists (the null expectation)	$\lambda_1 = \lambda_0$; $\mu_1 = \mu_0$; $q_{01} = q_{10}$
c)	Speciation rates and extinction rates are equal between specialists and generalists, but transition rates are free to vary	$\lambda_1 = \lambda_0$; $\mu_1 = \mu_0$
d)	Speciation rates and transition rates are equal between specialists and generalists, but extinction rates are free to vary	$\lambda_1 = \lambda_0$; $q_{01} = q_{10}$
e)	Extinction rates and transition rates are equal between specialists and generalists, but speciation rates are free to vary	$\mu_1 = \mu_0$; $q_{01} = q_{10}$
f)	Speciation rates are equal between specialists and generalists, but extinction rates and transition rates are free to vary	$\lambda_1 = \lambda_0$
g)	Extinction rates are equal between specialists and generalists, but speciation rates and transition rates are free to vary	$\mu_1 = \mu_0$
h)	Transition rates are equal between specialists and generalists, but speciation rates and extinction rates are free to vary	$q_{01} = q_{10}$
i)	Rate of trait loss is equal to zero, but all other parameters are free to vary (specialization is irreversible)	$q_{10} = 0$

Table S3: Maximum likelihood parameter estimates for ten published datasets. The six parameters are: speciation rate for generalists and specialists (λ_0, λ_1), extinction rate for generalists and specialists (μ_0, μ_1) and specialist trait gain and trait loss (q_{01}, q_{10}). Parameters were estimated using the Binary State Speciation and Extinction (BiSSE) model (Maddison *et al.*, 2007; FitzJohn *et al.*, 2009) under a range of constraints reflecting different possible impacts of specialization on speciation, extinction rates and trait transition rates (Table S1). Under some parameter sets, the simulated lineages did not coalesce into one ancestral lineage during the backwards simulation due to the negative overall diversification rate, so we adjusted the extinction rate to allow for tree simulation (adjusted value is given in parenthesis following the initial estimated value).

Table S3.1: Parameter estimates for Fernández-Mazuecos et al. (2013)

Constraints	λ_0	λ_1	μ_0	μ_1	q_{01}	q_{10}
a) No constraints	239	46.7	4.76E-06	34.4	60.0	2.06E-08
b) $\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$, $q_{01} = q_{10}$	208	208	2.65E-09	2.65E-09	63.8	63.8
c) $\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$	208	208	1.11E-05	1.11E-05	54.2	7.09E-06
d) $\lambda_0 = \lambda_1$, $q_{01} = q_{10}$	229	229	1.21E-11	235	94.2	94.2
e) $\mu_0 = \mu_1$, $q_{01} = q_{10}$	238	43.9	1.10E-09	1.10E-09	62.0	62.0
f) $\lambda_0 = \lambda_1$	234	234	6.20E-05	277 (233)	83.0	3.24E-05
g) $\mu_0 = \mu_1$	237	43.3	1.43E-07	1.43E-07	55.3	5.76E-08
h) $q_{01} = q_{10}$	238	4.39	4.35E-07	4.35E-06	62.0	62.0
i) $q_{10} = 0$	239	46.7	1.59E-10	34.4	60.0	0.00

Table S3.2: Parameter estimates for Hardy & Otto (2014)

Constraints	λ_0	λ_1	μ_0	μ_1	q_{01}	q_{10}
a) No constraints	3.29E-02	5.39E-02	1.24E-09	3.38E-08	4.93E-02	6.40E-02
b) $\lambda_0 = \lambda_1, \mu_0 = \mu_1, q_{01} = q_{10}$	4.34E-02	4.34E-02	1.82E-07	1.82E-07	7.10E-02	7.10E-02
c) $\lambda_0 = \lambda_1, \mu_0 = \mu_1$	4.33E-02	4.33E-02	6.33E-07	6.33E-07	6.40E-02	7.25E-02
d) $\lambda_0 = \lambda_1, q_{01} = q_{10}$	4.33E-02	4.33E-02	4.54E-06	6.68E-07	7.08E-02	7.08E-02
e) $\mu_0 = \mu_1, q_{01} = q_{10}$	3.52E-02	5.05E-02	6.42E-08	6.42E-08	7.43E-02	7.43E-02
f) $\lambda_0 = \lambda_1$	4.33E-02	4.33E-02	8.35E-07	2.36E-09	6.40E-02	7.24E-02
g) $\mu_0 = \mu_1$	3.29E-02	5.39E-02	8.98E-08	8.98E-08	4.92E-02	6.40E-02
h) $q_{01} = q_{10}$	3.54E-02	5.03E-02	2.56E-06	2.28E-06	7.40E-02	7.40E-02
i) $q_{10} = 0$	5.76E-02	2.56E-02	4.92E-10	1.80E-02	2.59E-02	0.00

Table S3.3: Parameter estimates for Janz et al. (2001)

Constraints	λ_0	λ_1	μ_0	μ_1	q_{01}	q_{10}
a) No constraints	234	272	5.95E-07	1.29E-06	25.4	57.4
b) $\lambda_0 = \lambda_1, \mu_0 = \mu_1, q_{01} = q_{10}$	264	264	4.85E-09	4.85E-09	56.6	56.6
c) $\lambda_0 = \lambda_1, \mu_0 = \mu_1$	264	264	6.55E-10	6.55E-10	27.2	53.7
d) $\lambda_0 = \lambda_1, q_{01} = q_{10}$	264	264	1.22E-08	2.58E-10	56.6	56.6
e) $\mu_0 = \mu_1, q_{01} = q_{10}$	242	271	9.13E-10	9.13E-10	61.1	61.1
f) $\lambda_0 = \lambda_1$	264	264	3.08E-10	3.35E-06	27.2	53.7
g) $\mu_0 = \mu_1$	234	273	8.55E-10	8.55E-10	25.4	57.4
h) $q_{01} = q_{10}$	242	271	4.35E-05	3.14E-06	61.1	61.1
i) $q_{10} = 0$	330	153	1.91E-09	1.05E-07	136	0.00

Table S3.4: Parameter estimates for Johnson et al. (2009)

Constraints	λ_0	λ_1	μ_0	μ_1	q_{01}	q_{10}
a) No constraints	4.21	8.51	6.28E-08	1.19E-09	3.00	1.69
b) $\lambda_0 = \lambda_1, \mu_0 = \mu_1, q_{01} = q_{10}$	7.91	7.91	4.63E-10	4.63E-10	1.23	1.23
c) $\lambda_0 = \lambda_1, \mu_0 = \mu_1$	7.91	7.91	2.45E-10	2.48E-10	30.9	8.48
d) $\lambda_0 = \lambda_1, q_{01} = q_{10}$	8.14	8.14	2.69	7.53-09	1.47	1.47
e) $\mu_0 = \mu_1, q_{01} = q_{10}$	4.14	8.46	1.79E-08	1.79E-08	1.45	1.45
f) $\lambda_0 = \lambda_1$	8.10	8.10	2.18	2.23-09	2.24	1.52
g) $\mu_0 = \mu_1$	4.21	8.51	2.36-09	2.36E-09	3.00	1.68
h) $q_{01} = q_{10}$	4.14	8.46	1.94E-05	5.30E-06	1.45	1.45
i) $q_{10} = 0$	9.20	6.52	1.76E-06	9.20E-09	5.38	0.00

Table S3.5: Parameter estimates for Larkin et al. (2008)

Constraints	λ_0	λ_1	μ_0	μ_1	q_{01}	q_{10}
a) No constraints	137	237	8.31E-06	1.30E-07	2.80E-07	64.4
b) $\lambda_0 = \lambda_1, \mu_0 = \mu_1, q_{01} = q_{10}$	212	212	1.27E-08	1.27E-08	60.6	60.6
c) $\lambda_0 = \lambda_1, \mu_0 = \mu_1$	212	212	4.38E-06	4.38E-06	1.21E-05	56.2
d) $\lambda_0 = \lambda_1, q_{01} = q_{10}$	212	212	9.93E-09	2.65E-09	60.6	60.6
e) $\mu_0 = \mu_1, q_{01} = q_{10}$	148	236	1.32E-12	1.32E-12	75.7	74.6
f) $\lambda_0 = \lambda_1$	212	212	3.94E-09	3.57E-06	2.96E-05	56.2
g) $\mu_0 = \mu_1$	137	237	2.19E-06	2.19E-06	5.42E-09	64.4
h) $q_{01} = q_{10}$	148	236	2.03E-05	5.57E-07	74.7	74.7
i) $q_{10} = 0$	249	144	6.87E-09	1.12E-05	94.8	0.00

Table S3.6: Parameter estimates for Mendlová et al. (2014)

Constraints	λ_0	λ_1	μ_0	μ_1	q_{01}	q_{10}
a) No constraints	1.44E-06	162	3.79E-05	105	52.6	55.2
b) $\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$, $q_{01} = q_{10}$	92.7	92.7	53.5	53.5	288	288
c) $\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$	92.7	92.7	53.5	53.5	434	284
d) $\lambda_0 = \lambda_1$, $q_{01} = q_{10}$	97.9	97.9	149	1.61E-08	202	202
e) $\mu_0 = \mu_1$, $q_{01} = q_{10}$	2.59E-09	114	32.6	32.6	93.6	93.6
f) $\lambda_0 = \lambda_1$	93.0	93.0	137	3.20E-08	340	276
g) $\mu_0 = \mu_1$	7.37E-10	114	28.1	28.1	34.9 (53.0)	72.2
h) $q_{01} = q_{10}$	9.62E-07	166	7.55E-06	111	55.5	55.5
i) $q_{10} = 0$	118	82.0	5.29E-10	156 (81.0)	103	0.00

Table S3.7: Parameter estimates for Schweizer et al. (2014)

Constraints	λ_0	λ_1	μ_0	μ_1	q_{01}	q_{10}
a) No constraints	260	296	2.15E-10	1.11E-05	7.30	1.11E-06
b) $\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$, $q_{01} = q_{10}$	264	264	4.87E-10	4.87E-10	7.18	7.18
c) $\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$	264	264	6.80E-07	6.80E-07	7.74	1.01E-04
d) $\lambda_0 = \lambda_1$, $q_{01} = q_{10}$	264	264	3.73E-07	8.48E-08	7.18	7.18
e) $\mu_0 = \mu_1$, $q_{01} = q_{10}$	260	299	1.15E-09	1.15E-09	6.78	6.78
f) $\lambda_0 = \lambda_1$	264	264	1.02E-06	4.10E-05	7.74	1.33E-06
g) $\mu_0 = \mu_1$	260	296	9.19E-08	9.19E-08	7.30	2.11E-07
h) $q_{01} = q_{10}$	260	299	2.00E-06	1.61E-05	6.78	6.78
i) $q_{10} = 0$	260	296	3.75E-07	1.29E-06	7.30	0.00

Table S3.8: Parameter estimates for Šimková et al. (2006)

Constraints	λ_0	λ_1	μ_0	μ_1	q_{01}	q_{10}
a) No constraints	77.8	170	7.24E-06	4.70E-09	7.57E-06	22.9
b) $\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$, $q_{01} = q_{10}$	158	158	5.45-09	5.45E-09	22.2	22.2
c) $\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$	158	158	4.03E-06	4.03E-06	9.61E-05	21.6
d) $\lambda_0 = \lambda_1$, $q_{01} = q_{10}$	158	158	8.74E-07	4.37E-09	22.2	22.2
e) $\mu_0 = \mu_1$, $q_{01} = q_{10}$	82.1	170	1.34E-10	1.64E-10	23.4	23.4
f) $\lambda_0 = \lambda_1$	160	160	158	6.29E-10	2.37E-06	22.4
g) $\mu_0 = \mu_1$	77.8	170	8.30E-06	8.30E-06	1.49E-06	22.9
h) $q_{01} = q_{10}$	82.1	170	6.58E-09	3.22E-09	23.4	23.4
i) $q_{10} = 0$	195	130	9.50E-08	2.51E-11	122	0.00

Table S3.9: Parameter estimates for Stireman (2005)

Constraints	λ_0	λ_1	μ_0	μ_1	q_{01}	q_{10}
a) No constraints	1432	528	815	4.83E-07	106	26.6
b) $\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$, $q_{01} = q_{10}$	797	797	298	298	22.5	22.5
c) $\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$	797	797	298	298	191	39.2
d) $\lambda_0 = \lambda_1$, $q_{01} = q_{10}$	733	733	8.78E-06	238	9.67	9.67
e) $\mu_0 = \mu_1$, $q_{01} = q_{10}$	822	533	28.4	28.4	10.3	10.3
f) $\lambda_0 = \lambda_1$	737	737	6.58E-08	261	174	15.3
g) $\mu_0 = \mu_1$	789	536	24.9	24.9	165	16.8
h) $q_{01} = q_{10}$	2710	548	2260	7.71E-08	42.2	42.2
i) $q_{10} = 0$	118	82.0	5.29E-10	156 (82.0)	103	0.00

Table S3.10: Parameter estimates for Tripp & Manos (2008)

Constraints	λ_0	λ_1	μ_0	μ_1	q_{01}	q_{10}
a) No constraints	218	18.2	1.56E-08	1.24E-05	40.7	2.30E-06
b) $\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$, $q_{01} = q_{10}$	197	197	2.32E-09	2.32E-09	34.5	34.5
c) $\lambda_0 = \lambda_1$, $\mu_0 = \mu_1$	197	197	1.91E-10	1.91E-10	62.8	206
d) $\lambda_0 = \lambda_1$, $q_{01} = q_{10}$	208	208	2.50E-10	109	46.2	46.2
e) $\mu_0 = \mu_1$, $q_{01} = q_{10}$	220	285	3.16E-11	3.16E-11	48.4	48.4
f) $\lambda_0 = \lambda_1$	197	197	4.18E-08	5.78E-07	62.8	206
g) $\mu_0 = \mu_1$	218	18.2	1.35E-07	1.35E-07	40.7	1.55E-08
h) $q_{01} = q_{10}$	220	28.5	9.44E-07	7.13E-09	48.4	48.4
i) $q_{10} = 0$	218	18.2	9.10E-09	1.66E-06	40.7	0.00

Figure S1: The observed value of FPD for hawkmoth-pollinated *Ruellia* (black vertical line) compared to the distribution of values generated under nine alternative macroevolutionary models (coloured histograms indicated in the figure legend; for model parameters, see Table S2). 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 ≤ 0.01 or p ≥ 0.99 (indicated by an asterisk).

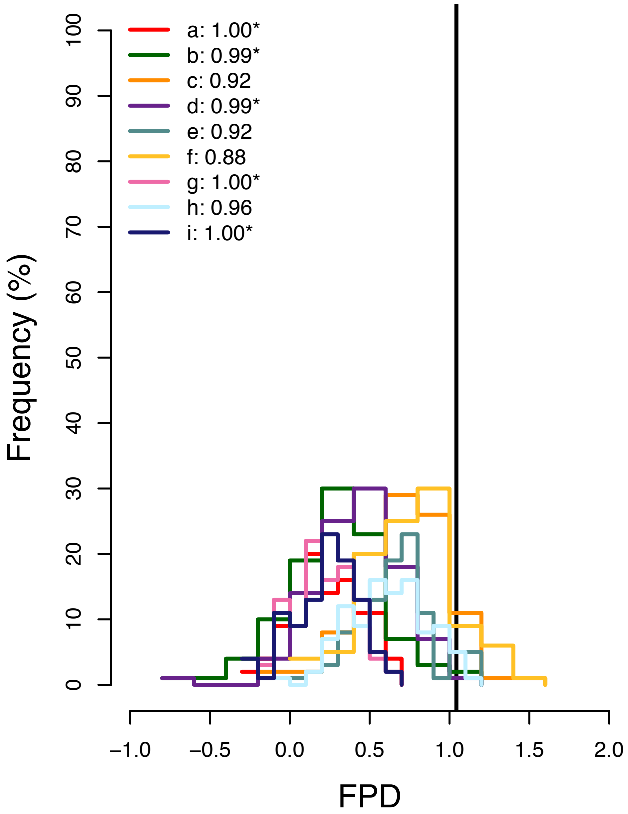

Figure S2: The observed value of FPD for host-specific Tachinid flies (black vertical line) compared to the distribution of values generated under nine alternative macroevolutionary models (coloured histograms indicated in the figure legend; for model parameters, see Table S2). The P-values in the figure 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 $p \geq 0.99$ (indicated by an asterisk).