Case Study 1: Bladderworts

[For the sake of this exercise, we will focus on the 1C genome size data only]

1.1.   Read **Bladderworts 1**: Given the hypothesis and findings, what should the implied DAG look like? (Hint: consider how the PGLS incorporates the effect of the phylogeny vs. that of the explanatory variables!)

1.2.   Read **Bladderworts 2**: Considering the results, do you find the support for the hypothesis convincing?
  a.   Considering the reconstruction of the COX mutations on the bladderworts phylogeny: If the DAG you drew above was representing the true causal relations, what would you expect the genome sizes to look like? Do the real genome sizes match that expectation?
  b.   If some of the data defies your expectations, can you consider those 'outliers' in a stochastic sense (i.e., to be expected from real observational data), or do they imply meaningful divergence from the hypothesis? Consider...
      i.   ...their position on the tree.
      ii.   ...their trait value in relation to the inferred effect size of the COX mutation.
      iii.   ...how they compare to values simulated under the same model.

[return to full class and present/discuss your answers]

1.3.   Read **Bladderworts 3**:
  a.   What might the inferred OU regimes for the genome sizes suggest about how a 'true' cause behind the observed genome shrinkage might be distributed (assuming it resembles a discrete trait)?
  b.   Can we reconcile this finding with our COX hypothesis? What points would we need to address?

[return to full class and present/discuss your answers]

1.4.   Drawing from both the paper's original findings and your hypothesized ones, what new alternative DAGs should we consider?

1.5.   What would we need to be able to determine which of these new DAGs is true? What new hypotheses would they generate?

[return to full class and present/discuss your answers]

Bladderworts 1

Zedek, F. *et al*. (2024) The smallest angiosperm genomes may be the price for effective traps of bladderworts. *Annals of Botany* **134**, 1131–1138.

Background Species of the carnivorous family Lentibulariaceae exhibit the smallest genomes in flowering plants. We explored the hypothesis that their minute genomes result from the unique mitochondrial cytochrome c oxidase (COX) mutation. The mutation may boost mitochondrial efficiency, which is especially useful for suction-bladder traps of Utricularia, but also increase DNA-damaging reactive oxygen species, leading to genome shrinkage through deletion-biased DNA repair. We aimed to explore the impact of this mutation on genome size, providing insights into genetic mutation roles in plant genome evolution under environmental pressures.

Methods We compiled and measured genome and mean chromosome sizes for 127 and 67 species, respectively, representing all three genera (Genlisea, Pinguicula and Utricularia) of Lentibulariaceae. We also isolated and analysed COX sequences to detect the mutation. Through phylogenetic regressions* and Ornstein–Uhlenbeck models of trait evolution, we assessed the impact of the COX mutation on the genome and chromosome sizes across the family.

Results Our findings reveal significant correlations between the COX mutation and smaller genome and chromosome sizes. Specifically, species carrying the ancestral COX sequence exhibited larger genomes and chromosomes than those with the novel mutation. This evidence supports the notion that the COX mutation contributes to genome downsizing, with statistical analyses confirming a directional evolution towards smaller genomes in species harbouring these mutations.

Conclusions Our study confirms that the COX mutation in Lentibulariaceae is associated with genome downsizing, probably driven by increased reactive oxygen species production and subsequent DNA damage requiring deletion-biased repair mechanisms. While boosting mitochondrial energy output, this genetic mutation compromises genome integrity and may potentially affect recombination rates, illustrating a complex trade-off between evolutionary advantages and disadvantages. Our results highlight the intricate processes by which genetic mutations and environmental pressures shape genome size evolution in carnivorous plants.

* To test whether species with or without the COX mutation differ in genome size [...], we employed phylogenetic regression as implemented in the R package Phylolm v.2.6.2 (Ho and Ane, 2014). We set the log10transformed 1C genome size [...] as a response variable and the state of the COX mutation (LS, CC and CS) as a categorical predictor. To select the best phylogenetic model for the regression's error term, we used the weighted Akaike information criterion (AICw; Akaike, 1978; Wagenmakers and Farrel, 2004).

** The phylogenetically corrected linear regression analysis revealed significant associations within Lentibulariaceae species. Species harbouring the ancestral LS mutation exhibited notably larger genomes ($P \ll 0.001$) [...] compared to those with either CC or CS mutations. [...] At the same time, no significant difference in genome size existed between these mutation categories [CC, CS] ($P = 0.144$). Furthermore, the differences in the state of the COX sequence explained 16.52% [...] of the variance in the genome [...] size [...]. Notably, the best model for error terms in the phylogenetic regression was the Ornstein–Uhlenbeck model (OUfixedRoot [...]), suggesting a directionality in the evolution of the genome or mean chromosome size of Lentibulariaceae.

Bladderworts 2

Zedek, F. *et al.* (2024) The smallest angiosperm genomes may be the price for effective traps of bladderworts. *Annals of Botany* **134**, 1131–1138.

We first replicated the analyses of Zedek et al. (2024), focusing on the phylogenetic regression of genome size and COX mutations, using the R package phylolm. The regression estimates different expected genome sizes for each COX-group (Figure 1a), while assuming that deviations from that expectation are distributed according to an OU-process on the phylogeny (residuals see Figure 1d).

We can interpret the distance between the expected genome sizes of the novel COX mutations (CC, CS) from the ancestral COX sequence (LS) as the effect sizes of those mutations. If the evolution of the CC and CS mutations were indeed leading to genome shrinkage, we would expect to see species' genome sizes distributed accordingly, i.e., clustering together based on their COX genotype (with some expected variation).
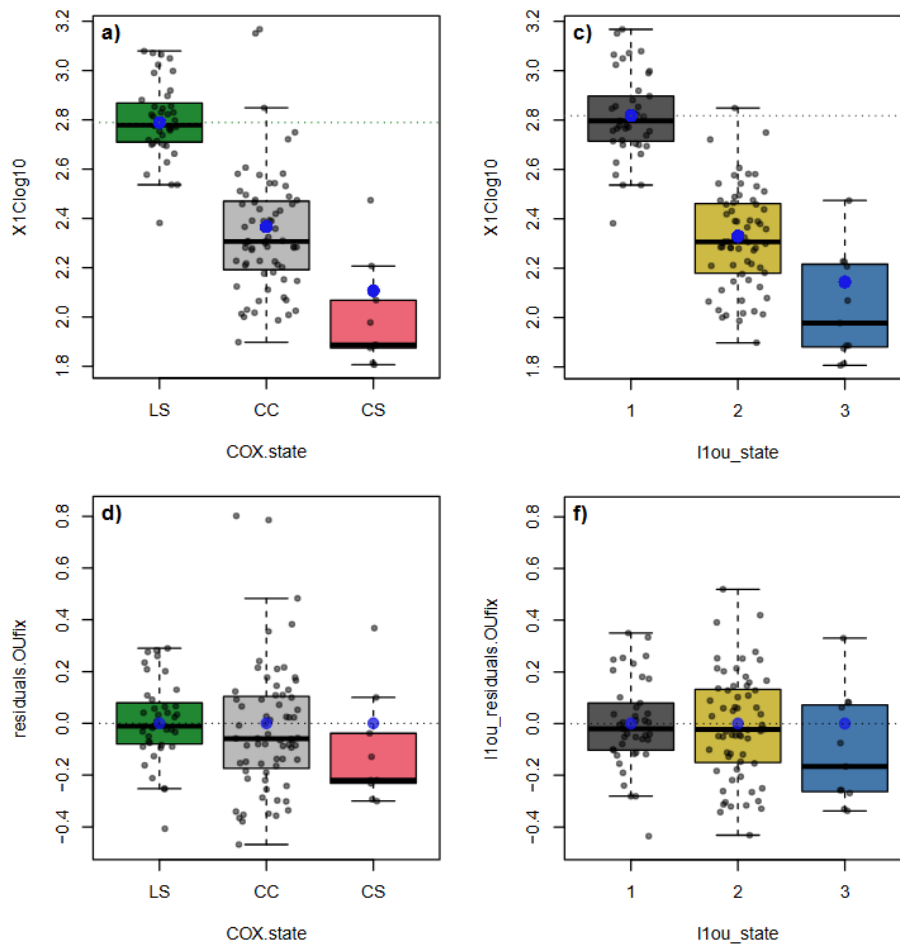

Figure 1: *Distribution of log10 genome sizes (a, c) and phylogenetic linear model residuals (d, f) between species carrying different COX mutations (a, d), or which are part of different OU-regimes (c, f). Gray points are genome sizes with added jitter, blue points are inferred expected values per group under the phylogenetic linear model.*

Figure 2: *Stochastic mapping of the ancestral states of COX mutations (left) and OU-regimes inferred through l1ou (right), along with the genome sizes for each species (center), coloured by the species' COX mutation.*
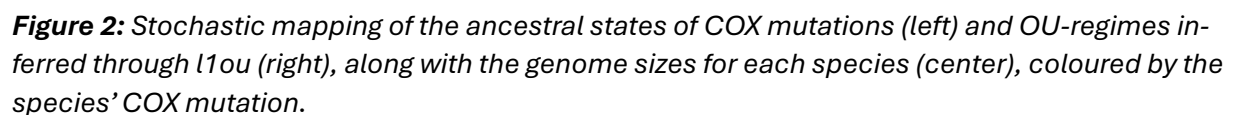
Figure 2 (left side) shows bladderworts' genome sizes as bars at the tips of the phylogeny with the branch colours showing reconstruction of the COX mutations.

Additionally, we performed a simplistic model adequacy test: we simulated 1000 new genome size datasets under the same model and the previously inferred parameters (i.e., PGLS, same tree, stochastic character map of COX genes, OU parameters, etc.). We then calculated the residuals of those simulated datasets and compared them with those from the original analysis (Figure 3 left). As a more specific summary statistic for our purposes, we compared the largest absolute residuals of each simulated dataset with the largest residual of the original analysis (Figure 3 right).
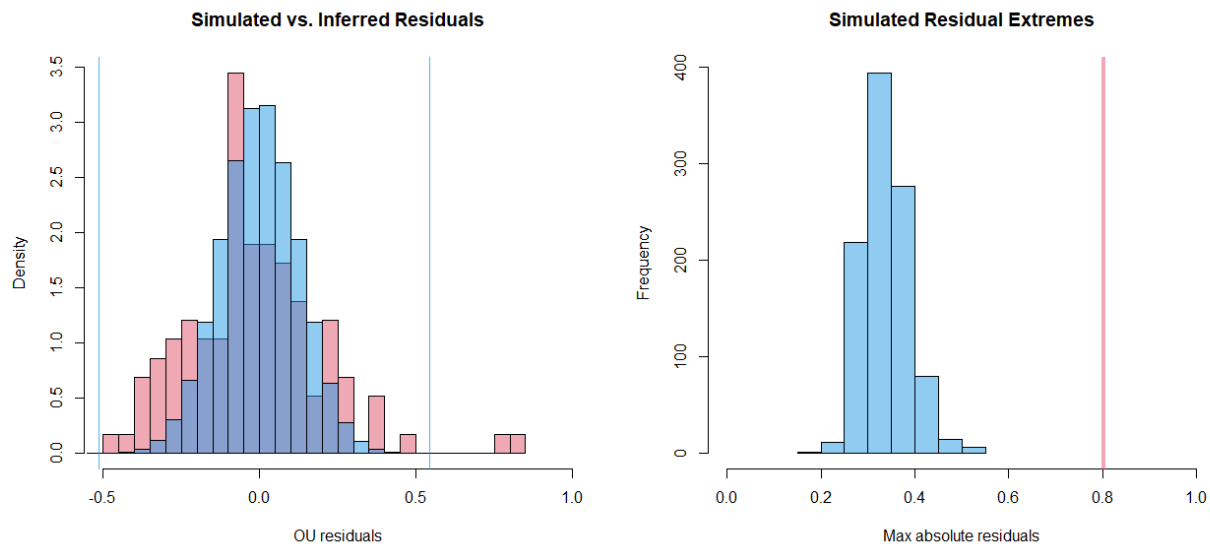

Figure 3: *Left: Distribution of inferred residuals from PGLS on genome size (red) and simulated ones (blue); displayed as densities to account for difference in number of samples, blue lines help to see the min/max range for the simulated residuals. Right: Distribution of largest (absolute) residual values from each of 1000 simulations, the red vertical line indicates the value of the largest residual from the empirical dataset.*

Bladderworts 3

Zedek, F. *et al.* (2024) The smallest angiosperm genomes may be the price for effective traps of bladderworts. *Annals of Botany* **134**, 1131–1138.

As an additional analysis, we quantify the observed heterogeneity in the genome sizes. We used the R package l1ou to estimate the optimal number and distribution of shifts needed to explain genome size differences under an Ornstein-Uhlenbeck process. The result is a shift each along the branches leading to *Utricularia*, and a clade within the *Genlisea* sections *Genlisea* and *Recurvatae* (Figure 2, right side). Strikingly, these clades essentially represent the same groupings as the COX mutations, except that four species of *Genlisea* sect. *Recurvatae* are now not grouped in the same regime as *Utricularia* anymore, but are instead split between *Genlisea* sect. *Genlisea* and the ancestral category.

We reanalyse genome size with the phylogenetic regression, using these newly inferred groups as explanatory variables instead, the results of which are shown in Figure 1c & 1f).

Case Study 2: Honeyeaters
[For exercise, we will focus on nectarivory and the rate of morphological evolution (esp. body-size, quantified by PC1) in the "background meliphagoids" only]

   2.1.   Read **Honeyeaters 1**: Given the hypothesis and findings (and the methods used), what should the implied DAG look like?

   2.2.   Read **Honeyeaters 2**: Considering the results, do you find the support for the hypothesis convincing?

   a.   Considering the reconstruction of nectarivory on the honeyeater phylogeny: If the DAG you drew above was representing the true causal relations, what would you expect the rates of body size evolution to look like? Do the inferred empirical rates match that expectation?

   b.   If the pattern in the data defies your expectations, can you attribute them to expected stochasticity, i.e., can you reconcile them with the result of the state-dependent rates analysis? Consider...

   i.   ... how clades who are nectarivorous/non-nectarivorous and who show different rates of body size evolution match up.

   ii.   ... the number and size of clades that show either pattern of trait-rate-combination.

[return to full class and present/discuss your answers]

   2.3.   Read **Honeyeaters 3**: What insights do we gain from the additional analysis?

   a.   What did we conceptually do when amalgamating traits and rate regimes? How can this help us assess the plausibility of nectarivory as a cause of higher rates of body size evolution? Does it change your verdict?

   b.   How can we reconcile this result with the one of the original analysis? Consider differences in evolutionary rates and clade sizes.

[return to full class and present/discuss your answers]

   2.4.   Based on the findings presented in the paper and your own considerations above, what new alternative DAGs should we consider?

   2.5.   What would we need to be able to determine which of these new DAGs is true? What new hypotheses would they generate?

[return to full class and present/discuss your answers]

Honeyeaters 1

The accumulation of exceptional ecological diversity within a lineage is a key feature of adaptive radiation resulting from diversification associated with the subdivision of previously underutilized resources. The invasion of unoccupied niche space is predicted to be a key determinant of adaptive diversification, and this process may be particularly important if the diversity of competing lineages within the area, in which the radiation unfolds, is already high. Here, we test whether the evolution of nectarivory resulted in significantly higher rates of morphological evolution*, more extensive morphological disparity, and a heightened buildup of sympatric species diversity in a large adaptive radiation of passerine birds (the honeyeaters, about 190 species) that have diversified extensively throughout continental and insular settings. We find that a large increase in rates of body size evolution and general expansion in morphological space followed an ancestral shift to nectarivory**, enabling the build-up of large numbers of co-occurring species that vary greatly in size, compared to related and co-distributed non-nectarivorous clades. These results strongly support the idea that evolutionary shifts into novel areas of niche space play a key role in promoting adaptive radiation in the presence of likely competing lineages.

*   Ran a multivariate BM model where the rate of body size evolution is allowed to differ between nectarivorous and non-nectarivorous lineages. Nectarivory was included via 1000 stochastic mappings of the binary trait using an ER model. Rate of morphological evolution was incorporated as the first principal component of a phylogenetic PCA of 4 selected traits (with PC1 largely aligning with body size).

** The model with trait-dependent BM rates was supported and inferred higher rates of body size evolution in nectarivorous lineages.
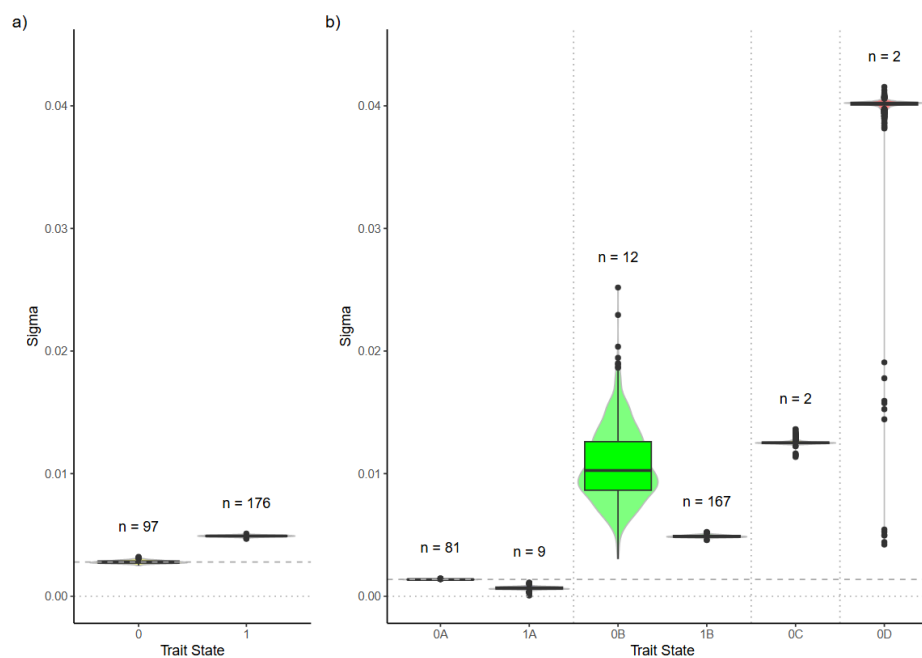

Figure 4: State dependent rates of Brownian motion for Meliphagidae body size (pPC1); a) rates for non-nectarivorous (0) and nectarivorous (1) lineages; b) rates for amalgamated states of nectarivory (0, 1) and BAMM rate regimes (A-D). Both analyses are based on 1000 stochastic mappings each, numbers above each box/violin indicate the number of tips in the respective state.

Honeyeaters 2

Marki, P. Z., Kennedy, J. D., Cooney, C. R., Rahbek, C. & Fjeldså, J. (2019) Adaptive radiation and the evolution of nectarivory in a large songbird clade. *Evolution* **73**, 1226–1240.

From Methods:

To assess the evolutionary origins of nectarivory among the Meliphagides, we reconstructed ancestral diets using stochastic character mapping (Bollback 2006) implemented in the R package phytools (Revell 2012; R Core Team 2016) [...] and estimated 1000 stochastic character maps using the ER model. [To quantify morphological diversity...], we performed a phylogenetic principal component (pPC) analysis upon the covariance matrix of the seven log-transformed variables (Revell 2009). [...] Combined, PC axes 1–4 explained 95% of the overall variance in the phylogenetic [...] PCA, and we thus focused our subsequent analyses and interpretations on these variables.

The invasion of novel niche space has been predicted to result in a decoupling of rates of ecomorphological evolution between the invading and noninvading clades (Rabosky 2017). To test this hypothesis, we compared the relative fit of different models of trait evolution using the R package mvMORPH (Clavel et al. 2015). Specifically, we compared a Brownian motion (BM) model with a single rate of trait evolution for all lineages (BM1) to a BM model with separate rates of trait evolution for nectarivorous and non-nectarivorous lineages (BMM). We fit these two models to each of the 1000 stochastic character maps. Univariate analyses were run for each of the first four pPC axes (pPC1–4). [...] Model support was assessed using Akaike information criterions corrected for small sample sizes (AICc) and Akaike weights. [...]

In addition, we also assessed finer scale lineage variation in the tempo and mode of meliphagoid morphological evolution using a variable rates model as implemented in BayesTraits version 2 [... which] uses reversible-jump Markov chain Monte Carlo algorithms (rjMCMC) and two scaling mechanisms to identify rate changes along single branches and for whole clades across the phylogeny (Venditti et al. 2011). In addition to BayesTraits, we also investigated another widely used framework for inferring variable rates of trait evolution across a phylogeny (BAMM version 2.5.0; Rabosky 2014; Rabosky et al. 2014a). The BAMM method attempts to identify the location and number of distinct macroevolutionary rate regimes on the phylogeny. The number of distinct regimes is modeled following a Poisson distribution, with rjMCMC used to sample different regimes that best explain the distribution of trait values at the tips of the tree. [...] We analyzed each of the four pPC axes calculated for the Meliphagides using the MCC tree as input.

From Results:

The ancestral estimation of the presence of nectar in the diet of the Meliphagides is strikingly characterized by a shift from a non-nectarivorous diet to one that can incorporate nectar in the common ancestor of honeyeaters (Fig. 5A). Nectarivory has also evolved independently among the pardalotes (family Pardalotidae) and among a handful of species of Australasian warblers (family Acanthizidae) that are members of the background meliphagoids. Among honeyeaters, loss of nectarivory has occurred independently on a number of more terminal branches, such as in the largely frugivorous genera *Melipotes* and *Macgregoria*, as well as in more insectivorous genera such as *Epthianura* and *Timeliopsis*.

A pPCA of the seven log-transformed morphological traits comparing honeyeaters against background meliphagoids showed that the first axis (pPC1) strongly reflected overall size, explaining

65.3% of the total variance in the morphological measurements. The next three axes (pPC2–4) were related to variation in Kipp's distance (pPC2), bill depth and width (pPC3), and bill length (pPC4), together explaining 29.8% of the variance. Visual comparisons of species scores on pPC axes 1–4 highlight the great morphological disparity and distinctiveness of the honeyeaters. First, the variance of body sizes (pPC1) exhibited by honeyeaters is much greater than that of the background meliphagoids. [...]

Comparisons of different models of trait evolution using mvMORPH provided strong support for a decoupling of trait diversification dynamics among nectarivorous and non-nectarivorous lineages. Models with separate rates of trait evolution (BMM) for nectarivorous and non-nectarivorous lineages represented the best fitting model for pPC1, pPC2, pPC4 [...], whereas a single-rate BM (BM1) model was the best fit for pPC3. For pPC1 [and] pPC4 [...], nectarivorous lineages were found to have a higher rate of evolution than non-nectarivorous lineages (Fig. 4a). For pPC2, nectarivorous lineages were found to have a lower rate of evolution than non-nectarivorous lineages. [...]

Model comparisons (using delta AIC) suggest that BayesTraits provides a branch-length transformation that results in the most Brownian-like distribution of the tip data among the range of models tested for all pPC axes (pPC1–4). Consequently, we focus our interpretation and discussion on the BayesTraits results (although those generated by BAMM were largely congruent, Fig. 5b).

The BayesTraits analyses [...] among the individual pPC axes provided strong support for a clade-wide shift to higher rates of trait evolution near the base of the honeyeater clade on pPC1 (PP = 0.90) and for three species of *Gerygone* among the background meliphagoids (PP = 0.83). No rate shifts in the univariate analysis of pPC2-4 were strongly supported (all PP < 0.7). [...]

Honeyeaters 3

We replicated the analyses of Marki et al. (2019, though we used BAMM to find shifts in rates and modes of the evolution of pPC1. The inferred best shift configuration includes a shift to higher rates at the base of Meliphagidae, as well as shifts to increased rates in two smaller clades (Figure 5b).

To revisit the plausibility of a causal effect of nectarivory, we can once again condition on the observed heterogeneity. We have already quantified the variation in the rate of body size evolution using BAMM, and can extract the rate-regime identity for each lineage from there. To condition the effect of nectarivory on the observed rate shifts, we simply amalgamate the trait states (0, 1) with the rate regimes (A-D), to crate the corresponding pseudotrait (while discarding the non-existing combinations). The subsequent analysis requires stochastic maps as input. So instead of inferring new mappings from the amalgamated pseudotrait states (and thereby distorting the true locations of the trait and rate shifts), we simply amalgamated the states on the existing stochastic maps of nectarivory with the BAMM edge states. The result were 1000 stochastic maps of nectarivory, conditional on the BAMM rate regimes (Figure 5c). Now we can infer the effect of nectarivory on rates of body size evolution, conditional on the inferred rate shifts (Figure 4b).
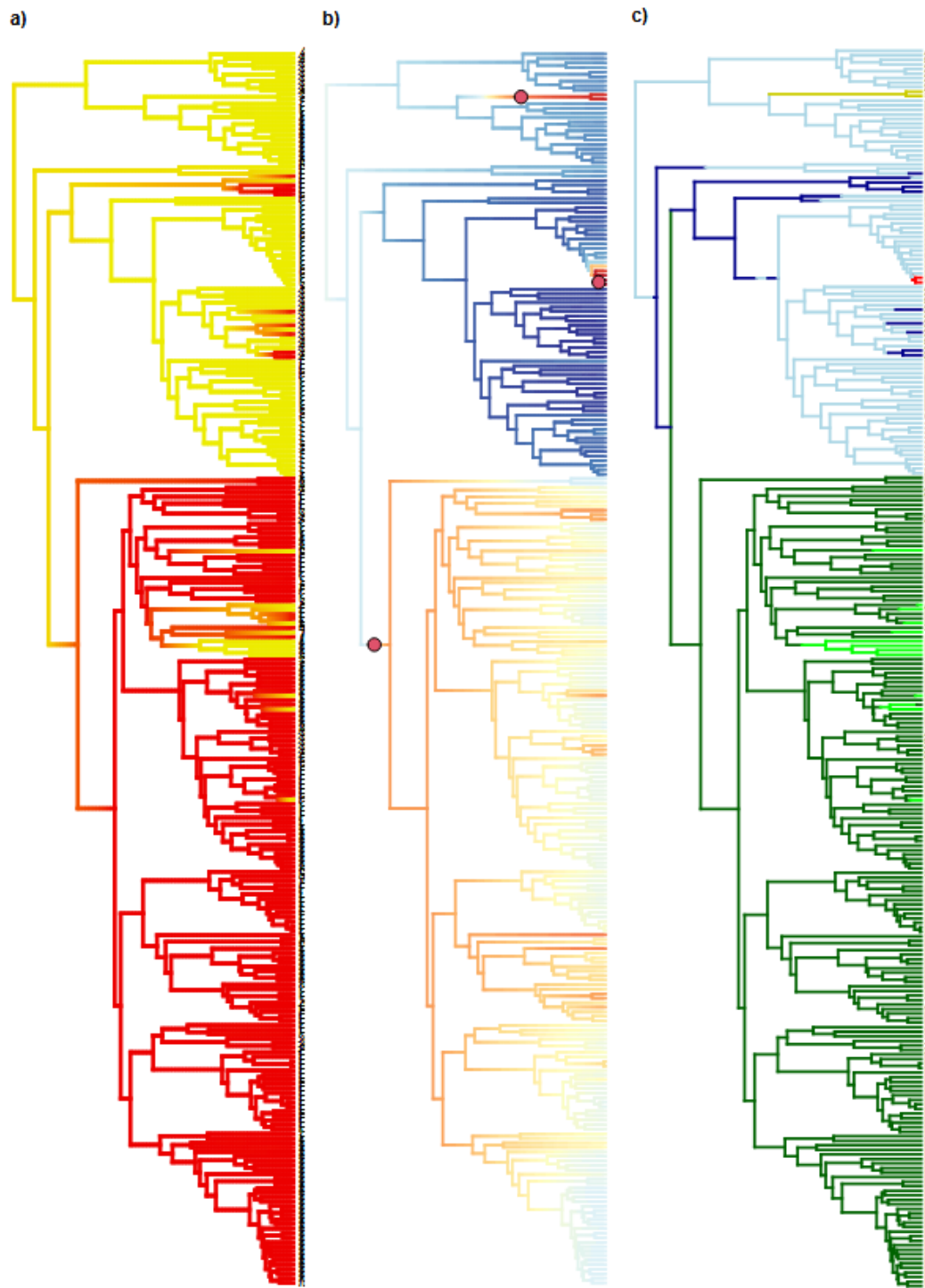
Figure 5: *Traits and rate regimes in Meliphagidae. a) Density plot of 1000 stochastic maps of the ancestral states of nectarivory (yellow: non-nectarivorous, red: nectarivorous) under the ER model; b) mean rates of body size evolution (pPC1, warmer colours indicating higher rates, Jenk's colour scheme) inferred in BAMM, with red dots indicating the rate shifts under the best regime; c) amalgamation of the nectarivory stochastic maps with the BAMM rate regimes (one example mapping; the BAMM regimes coloured in blue, green, red, and yellow, with the presence of nectarivory represented in the respectively darker colour).*