



Citation for published version:

Oyston, J, Wilkinson, M, Ruta, M & Wills, M 2022, 'Molecular phylogenies map to biogeography better than morphological ones', *Communications Biology*, vol. 5, 521. <https://doi.org/10.1038/s42003-022-03482-x>

DOI:

[10.1038/s42003-022-03482-x](https://doi.org/10.1038/s42003-022-03482-x)

Publication date:

2022

Document Version

Peer reviewed version

[Link to publication](#)

Publisher Rights

CC BY

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Molecular phylogenies map to biogeography better than morphological ones

Jack W. Oyston^{1*}(jwo22@bath.ac.uk), Mark Wilkinson², Marcello Ruta³
& Matthew A. Wills^{1*}(bssmaw@bath.ac.uk)

¹ Milner Centre for Evolution, Department of Biology & Biochemistry, University of Bath, Bath,
UK

² Vertebrates Division, Department of Life Sciences, Natural History Museum, Cromwell Road,
London, UK

³ School of Life Sciences, Joseph Banks Laboratories, College of Science, University of
Lincoln, Lincoln, UK

Abstract

Phylogenetic relationships are inferred principally from two classes of data: morphological and molecular. Currently, most phylogenies of extant taxa are inferred from molecules and when morphological and molecular trees conflict the latter are often preferred. Although supported by simulations, the superiority of molecular trees has rarely been assessed empirically. Here we test phylogenetic accuracy using two independent data sources: biogeographic distributions and fossil first occurrences. For 48 pairs of morphological and molecular trees we show that, on average, molecular trees provide a better fit to biogeographic data than their morphological counterparts and that biogeographic congruence increases over research time. We find no significant differences in stratigraphic congruence between morphological and molecular trees. These results have implications for understanding the distribution of homoplasy in morphological data sets, the utility of morphology as a test of molecular hypotheses and the implications of analysing fossil groups for which molecular data are unavailable.

Introduction

Phylogenies are essential in many areas of biology ¹, being widely utilised in evolutionary biology ^{2,3}, ecology ⁴, conservation ⁵, parasitology ⁶ and medicine ⁷. But what is the best way to produce an accurate phylogeny? Prior to the advent of molecular sequencing, morphology was the sole source of character data for phylogenetic inference in extant taxa ⁸. Since the 1990s ⁹, however, the balance has shifted dramatically in favour of phylogenomic data ¹⁰.

Studies of homoplasy and convergence demonstrate that morphological similarity can sometimes be a poor guide to evolutionary relationships ¹¹. While some argue that molecules should invariably have primacy in phylogenetic inference ¹², morphological and molecular data are often reciprocally illuminating, as shown in large-scale phylogenies of arthropods ¹³, reptiles and birds ¹⁴. This balanced approach, acknowledging that both types of data have strengths, is now common in systematics ^{15,16}. While phylogenetic hypotheses derived from morphology are often supported by molecular data ¹⁷, molecules have also overturned many long-standing morphological hypotheses ¹⁸. For example, phylogenomic analyses of placental mammals ¹⁹ have drastically altered the sequence of deep branching events traditionally supported by morphology ²⁰. Newly resulting mammal clades (e.g. Afrotheria, Atlantogenata, Boreoeutheria, Laurasiatheria) ²¹ are more congruent with their current geographic distributions, and have been named accordingly. Equally, molecular trees often conflict with each other, most notably when they are inferred using different sets of genes.

In the absence of known phylogenies, there can be no definitive assessment of the accuracy of branching patterns ^{22,23}. However, it is useful to evaluate conflicting trees using additional and independent criteria. Here we utilise two independent sources of data, namely biogeographic distributions and first stratigraphic occurrences. Before the cladistic revolution, biogeography was sometimes used to infer the relationships of extant taxa in combination with morphological data ^{24,25}. Although congruence with stratigraphy can be used as an ancillary criterion to choose between equally optimal trees for groups with a good fossil record, neither biogeographic ²⁶ nor stratigraphic data ²⁷⁻²⁹ are routinely used to infer phylogeny today.

Since Wallace and Darwin, observations on the geographic distributions of species have underpinned the development of evolutionary theory³⁰. Numerous studies have demonstrated non-random geographic patterns on evolutionary trees^{31,32}, and phylogenies are routinely used to test biogeographic hypotheses³³. Here, we employ biogeographic congruence as an ancillary test of competing phylogenetic hypotheses using a sample of 48 matched pairs of morphological and molecular trees of animals and plants at multiple taxonomic levels. By using randomisation tests to compare the fit of the same biogeographic regions on paired morphological and molecular trees of the same taxa, our approach controls for differences in tree size and balance to the extent that these influence our indices of fit. We demonstrate that molecular phylogenies fit biogeographic data significantly better than their morphological counterparts. This difference in biogeographic congruence is not simply explained by differences in tree shape, tree resolution or when the trees were first published, although more recently published trees do tend to perform better. Ancillary tests using biogeographic congruence are shown to perform at least as well as existing tests based on stratigraphic congruence. We therefore propose that tests of biogeographic congruence, in combination with other tests, represent a useful way of evaluating competing evolutionary trees.

Results

Testing biogeographic congruence

The process of summarising biogeographic data and assessing their fit onto trees is shown in Fig. 1 and described in detail in the Methods. Biogeographic occurrence data for extant taxa were compiled from the IUCN Red List of Threatened Species, Version 2019-2³⁴, the Global Biodiversity Information Facility (GBIF)³⁵ and The Reptile Database³⁶. These distributions were used to define regions of shared taxa that summarised their present-day distributions, combining adjacent regions that contained identical taxon sets (see Supplementary Methods). Regional distributions were encoded in a matrix in the form of presence/absence scores for each taxon in each region. The fit of these biogeographic characters to both morphological and molecular trees was assessed using the ensemble consistency index (CI) and retention

index (RI). However, our preferred index is a modified version of the homoplasy excess ratio³⁷, the biogeographic HER (bHER), derived from 10,000 random reassignments of biogeographic distribution data across terminals.

Phylogenies tend to be significantly congruent with biogeography

The overall congruence of phylogenies with biogeographic data was good: 54% of morphological and 65% of molecular trees had a significantly better fit than randomly permuted data at a p-value < 0.05 (and 69% of groups had one or both trees with a p-value < 0.05). Therefore, while biogeographic congruence for a minority of clades did not differ significantly from that expected by chance (e.g., Supplementary Figure 1), most groups showed significant patterns that could be used to discriminate between trees. Biogeography and phylogeny are often thought to be correlated for major clades at large geographic scales (e.g., the distribution of placental mammal orders on continents¹⁹; Fig. 2a), and we find compelling evidence for similar patterns at other taxonomic levels and geographic scales (Fig. 2b, Supplementary Figures 2, 3 & 4). Most biogeographic region matrices also had significantly non-random structure according to tree-independent permutation tail probability tests of pairwise character compatibility³⁸ (MCPTP tests: see Supplementary Methods). Our findings therefore support the use of biogeographic distribution data as an ancillary criterion for choosing between otherwise equally optimal trees, similar to the widespread practice adopted for stratigraphic congruence³⁹.

Molecular trees are more congruent with biogeography than morphological trees

Overall, biogeographic congruence was higher for our sample of molecular trees than for their morphological counterparts (Supplementary Figure 5: means of 0.322 vs. 0.305, medians of 0.277 vs. 0.276 for CI; means of 0.263 vs 0.228, medians of 0.211 vs. 0.183 for RI; Supplementary Figure 6: means of 0.188 vs 0.121, medians of 0.153 vs. 0.108 for bHER). These differences were significant for all measures of biogeographic congruence according to Wilcoxon paired signed-rank tests (Table 1: CI; $W = 685$, $Z = 2.22$, $rc = 0.384$, $p\text{-value} =$

0.027, RI; $W = 695$, $Z = 2.33$, $rc = 0.404$, $p\text{-value} = 0.0199$, bHER; $W = 888$, $Z = 3.08$, $rc = 0.51$, $p\text{-value} = 0.002$) across the 48 pairs of trees, with molecular trees having greater congruence on average, according to each index (Fig. 3). Two-tailed sign tests also demonstrated that molecular trees had greater biogeographic congruence more often than their morphological counterparts (Fig. 4, Supplementary Table 1). Our samples of molecular and morphological trees did not differ significantly in their balance (how symmetrical or pectinate they were), the degree to which CI & RI differed from randomly permuted data or any stratigraphic congruence measure tested. The bHER is our preferred index, since it controls for tree size, balance and the number of biogeographic regions. Considering only groups with significantly structured ($p\text{-value} < 0.05$) region matrices (Supplementary Table 2), we recovered a similar result for bHER ($W = 305$, $Z = 2.32$, $rc = 0.502$, $p\text{-value} = 0.019$, $n = 28$).

In order to further ensure that the observed differences in congruence were not the result of conflating factors (Supplementary Table 3), we also modelled CI, RI and bHER as a function of tree type (morphological or molecular), clade root node age, tree balance (using Colless's index⁴⁰), the number of geographic regions recognised, tree size (the number of terminal taxa), the ratio of characters to taxa (characters in the datasets used to generate the trees / the number of terminals), publication year and tree resolution expressed as the proportion of resolved nodes (number of internal nodes / (number of terminals – 2)). Multivariate linear regression models (Supplementary Table 4) supported publication year, number of biogeographic regions and the proportion of resolved nodes together as the best predictors of bHER, while CI was best predicted by the combination of data type (whether the tree was morphological or molecular), the age of the root node, the number of biogeographic regions, the number of terminal taxa and the ratio of phylogenetic characters to taxa. In contrast, the number of region characters, along with the root node age and the proportion of resolved nodes were the best predictors of the RI. Despite this, residuals from weighted robust regression models and from minimum adequate models (MAMs) selected by the Akaike

information criterion (AIC) showed a similar pattern to uncorrected values (Table 2), with CI and bHER demonstrating significantly greater biogeographic congruence for molecular trees (CI: $W = 994$, $Z = 4.16$, $rc = 0.69$, $p\text{-value} = 1.111 \times 10^{-5}$; bHER: $W = 827$, $Z = 2.45$, $rc = 0.406$, $p\text{-value} = 0.013$). Morphological trees contained more polytomies (Supplementary Table 5) and significantly fewer resolved nodes (Table 1), but there was still a significant difference between molecular and morphological bHER when groups with polytomous morphological trees were omitted ($n = 16$, $W = 179$, $Z = 2.12$, $rc = 0.603$, $p\text{-value} = 0.01459$).

Significant differences in bHER were also recovered comparing only groups with the same number of leaves in polytomies ($n = 16$, $W = 115$, $Z = 2.43$, $rc = 0.691$, $p\text{-value} = 0.01309$), only groups where 75% or more of the nodes in both trees were resolved ($n = 38$, $W = 537$, $Z = 2.41$, $rc = 0.449$, $p\text{-value} = 0.01485$) and groups which differed in their proportion of resolved nodes by 5% or less ($n = 16$, $W = 144$, $Z = 1.97$, $rc = 0.516$, $p\text{-value} = 0.04937$). Additionally, CI values showed no evidence of any correlation with the number of polytomies, number of branches in the polytomies or the proportion of resolved nodes (Supplementary Figure 7). While bHER showed evidence of significant but weak negative correlations with the number of branches in polytomies (Supplementary Figure 8b) and the proportion of resolved nodes (Supplementary Figure 8c), molecular trees still showed significantly greater congruence when comparing residual bHER values in each case (number of branches in polytomies: $W = 789$, $Z = 1.6$, $rc = 0.265$, $p\text{-value} = 0.03895$; proportion of resolved nodes: $W = 838$, $Z = 2.56$, $rc = 0.425$, $p\text{-value} = 0.009612$).

Whilst taxonomic sampling and clade age are, by definition, the same for each pair of morphological and molecular trees in our compilation, clade age itself might be expected to influence biogeographic fit. Both RI and bHER were weakly positively correlated with the log of clade root node age (Supplementary Figure 9: RI; $R^2 = 0.04437$, $p\text{-value} = 0.0394$; bHER; $R^2 = 0.05894$, $p\text{-value} = 0.01716$), indicating that phylogenies with earlier divergence times are more congruent with biogeography. In both cases residual values from linear regressions of fit metrics against log root node age still showed a significant difference between molecular

and morphological trees (RI: $W = 695$, $Z = 2.33$, $rc = 0.404$, $p\text{-value} = 0.0199$; bHER: $W = 888$, $Z = 3.08$, $rc = 0.51$, $p\text{-value} = 0.001684$). In addition, differences in fit metrics between morphological and molecular trees showed no evidence of any correlation with log root node age (Supplementary Figure 10). Any putative correlation between clade age and biogeographic fit is therefore insufficient to explain the differences between morphological and molecular trees observed here.

Morphological and molecular trees have similar stratigraphic congruence

Of our 48 pairs of morphological and molecular trees, 23 had at least 50% of terminals with a fossil record, and these were assessed for stratigraphic congruence (Supplementary Table 6). Our preferred index is the modified gap excess ratio (GER*)²⁷, since it is relatively insensitive to differences in tree shape (balance), tree size, and the distribution of first occurrence dates (although the latter two variables are constant for each of our pairs). Morphological and molecular trees (Supplementary Figure 11) had similar GER* values overall (0.774 and 0.780 respective means; 0.826 and 0.838 respective medians), and Wilcoxon signed-rank tests (Table 1) revealed no significant difference between the distributions of GER* values ($W = 90$, $Z = 0.196$, $rc = 0.0526$, $p\text{-value} = 0.8617$). We note that the highest stratigraphic congruence occurred more frequently in morphological ($n = 10$) than molecular trees ($n = 8$) (Supplementary Figure 12), but this difference was not significant (Supplementary Table 7: sign test; $n = 23$, $p\text{-value} = 0.21$). We observed similar results for the gap excess ratio (Supplementary Figure 13a: GER; $W = 91$, $Z = -0.523$, $rc = -0.133$, $p\text{-value} = 0.6142$), stratigraphic consistency index (Supplementary Figure 14a: SCI; $W = 140.5$, $Z = 1.33$, $rc = 0.338$, $p\text{-value} = 0.1913$) and modified Manhattan stratigraphic measure (Supplementary Figure 14b: MSM*; $W = 92$, $Z = -0.121$, $rc = -0.0316$, $p\text{-value} = 0.9198$). Although the power of statistical tests was likely impacted by reduced sample size, tests of biogeographic congruence using Wilcoxon signed-rank tests (Supplementary Table 8) and sign tests (Supplementary Table 9) showed significant differences for bHER when carried out on only those clades included in the stratigraphic analyses.

More recently published trees tend to be more biogeographically congruent

The history of systematic research is characterised by greater volumes of data being analysed with increasingly sophisticated methods and models⁴¹. All other factors being equal, we might therefore expect phylogenetic accuracy to increase over research time²¹. Across all 96 morphological and molecular trees, we observed significant positive correlation between publication year and bHER ($r_s = 0.257$, p-value = 0.012) and negative correlation between publication year and p-values from our biogeographic CI and RI ($r_s = -0.284$, p-value = 0.005). Hence, more recent trees tended to have higher biogeographic congruence (Supplementary Figure 15, Supplementary Table 10). A similar pattern was found for the bHER of the morphological trees considered alone ($r_s = 0.292$, p-value = 0.044), but was not significant for the molecular trees alone (bHER; $r_s = 0.184$, p-value = 0.210; CI & RI p-values; $r_s = -0.274$, p-value = 0.060). A significant minority (22 from 48) of our tree pairs had different publication dates, but we found no significant difference in the median publication years of the morphological and molecular partitions (Wilcoxon signed-rank $W = 59$, $Z = 0.947$, $rc = 0.297$, p-value = 0.362). An overall improvement in phylogenetic accuracy with research time may be driven partially by analysing increasing volumes of data, both in terms of number of taxa and numbers of characters. However, this trend cannot explain adequately the observed differences in biogeographic fit between pairs of morphological and molecular trees, as publication year was found to be a poor predictor of biogeographic congruence metrics in most cases (Supplementary Table 4) and residuals from linear regressions of congruence metrics against publication year were still significantly higher for molecular trees in each case (Wilcoxon signed-rank test: CI; $W = 769$, $Z = 2.5$, $rc = 0.423$, p-value = 0.01274, RI; $W = 760$, $Z = 2.4$, $rc = 0.406$, p-value = 0.01673, bHER; $W = 867$, $Z = 2.86$, $rc = 0.474$, p-value = 0.003649).

Discussion

The observation that biogeographic congruence is significantly greater than expected by chance alone for most of our clades (69% had one or both trees with CI & RI p-value < 0.005)

supports the use of biogeographic data as an ancillary test of phylogenetic accuracy. Moreover, median biogeographic congruence for our 48 molecular trees was significantly higher than for their morphological counterparts and biogeographic congruence was not a function of tree size and balance. Indeed, if our results are representative, biogeographic distribution may be a better ancillary test than the established criterion of stratigraphic congruence. Stratigraphic congruence might also be contingent on the method used for tree inference. For example, morphological trees constructed using maximum parsimony often show greater stratigraphic congruence than their Bayesian equivalents⁴², despite the increasing use of Bayesian methods with morphological data^{43,44}, although see^{45,46}. In this study, our ability to distinguish between morphological and molecular trees was likely limited by a small sample size ($n = 23$).

Molecular data offer several advantages over morphology. Firstly, molecular characters can be acquired in vastly greater numbers and more readily than morphological ones, and often with less taxonomic expertise⁴⁷. Secondly, published sequence data can be readily searched, repurposed and reanalysed alongside novel sequences. Despite efforts to systematically archive morphological character matrices and character descriptions⁴⁸, there is as yet no way to automatically produce iteratively larger morphological matrices in a manner analogous to that possible for molecular data⁴⁹. Both factors mean that it is often far easier to compile large molecular data sets than it is to compile equivalent volumes of morphological data. Thirdly, morphological systematists must make judgements concerning the homology of their characters and the way in which they are coded⁵⁰. Morphological variation is unlikely to be atomised in precisely the same manner by different systematists⁵¹, whereas it has been argued that *a priori* rules mitigate against subjectivity and promote repeatability in molecular systematics. Fourthly, a well-developed body of theory and empirical data facilitate sophisticated models of molecular evolution⁵², while mathematical models for morphological evolution are still in their infancy^{53,54}.

Of course, molecular phylogenetics is not without its own problems, including issues of homology (orthology detection, alignment, saturation and homoplasy), the dangers of model misspecification and systematic bias. Moreover, paralogy, incomplete lineage sorting and horizontal gene transfer mean that even accurate gene trees may be incongruent with species trees. However, all other things being equal, where molecular and morphological data yield conflicting trees, our results suggest that molecular trees are likely to be more accurate. Phylogenetic signals across multiple gene alignments are typically much stronger, and lead to higher bootstrap branch support and posterior probabilities than signals from morphology⁵⁵. Most morphological characters are binary and may be more prone to saturation than nucleotides and amino acids (assuming roughly equal rates of molecular and morphological character evolution). Many morphological characters are formulated to capture variation in different parts of the taxon sample. In so doing, however, they often incorporate assumptions about the way in which evolutionary transitions occurred. This is particularly true of characters whose states are logically contingent upon the states of others. For example, one character might code the presence or absence of a limb, while other characters might code for the morphology of bones within that limb. Where limbs are absent, these bone characters are often coded with “not applicable” scorings. Many morphological matrices therefore contain blocks of characters that are strongly conditionally dependent. However, morphological character matrices are, in theory, ‘infinitely extensible’ as newly discovered aspects of variation are accommodated in successive iterations by adding more characters and states. This approach to the accretion of morphological datasets might make characters less likely to show saturation through reversions to the same coded states but may make convergent gains more likely. This is particularly true if the initial hypotheses of transitions are incorrect. Convergence in morphological character states is common⁵⁶, even in characters that pass some of the conventional tests of homology⁵⁷ and have been hypothesised in the literature as homologous characters for decades⁵⁸.

While it is true that morphological trees tend to be less resolved, comparisons restricted to fully resolved trees have demonstrated that real incongruence in their primary phylogenetic signals⁵⁹ must account for the differing fits of morphological and molecular trees to biogeography. What we are unable to investigate further without access to the original data and comparative branch support metrics⁶⁰ is whether this incongruence is primarily due to lack of information or misleading information in morphological data. If, for example, incongruent relationships in morphological trees are less well supported by indices such as bootstrap⁶¹ or Bremer support⁶² than relationships which are congruent with biogeography, it would suggest that the biogeographic incongruence of morphological trees is partly attributable to a lack of strong signal in the morphological data.

Despite molecular trees typically showing greater biogeographic congruence, we found several cases where morphological trees have better fit than their molecular counterparts, such as dogs (Canidae), squirrels (Sciuridae), bats (Chiroptera), kangaroos (Macropodidae), conifers as a whole (Pinales) and pines (Pinaceae). However, in these cases, congruence values (and specifically bHER) only marginally favoured the morphological trees. Members of some these clades, such as conifers and bats, can disperse or travel over long distances and so may have large geographic ranges that limit the number of region characters and hence impact the power of our tests. Some morphological datasets may also contain characters that have evolved in response to particular environmental conditions (e.g., the pine dataset was based on cone morphology). This may increase congruence with biogeography when the regions within the clade's range broadly correspond with these environmental zones. Some clades (e.g., Canidae) were present in many more distinct biogeographic regions than the number of taxa in the dataset. As each region is defined by a unique grouping of taxa, a high number of regions relative to the number of taxa implies that the same taxa occur in different combinations in order to specify each distinct region. A 'mosaic pattern' of this type is likely to occur when at least some of the constituent taxa have fragmented rather than continuous distributions. This might, in turn, be indicative of frequent and rapid dispersal over long

distances. Such patterns are common in many clades, particularly large mammals^{63,64} which typically have wide-ranging distributions. Alternatively, or in addition, mosaic patterns might result from the rapid fragmentation of an original range. Since this occurs on much shallower timescales than the deeper divergences of the major branches in the phylogeny⁶⁵, the original biogeographic signal can be obscured.

Other problems that can impact accuracy, including long branch attraction and incomplete lineage sorting, are not unique to morphological data. While simulations suggest that likelihood and Bayesian analyses are more resilient to some of these issues⁶⁶, such methods are increasingly being applied to morphological data. For some clades, particularly mammals, it might be possible to estimate the likelihood of biogeographic character saturation. However, this would require independent data on the rate of biogeographic transitions (from either direct observations or population genetics), along with time calibrated phylogenies with scaled branch lengths. For most of the clades in this study such data do not exist and would require extensive effort to collect. More importantly, there is no reason why any such putative saturation effects should detrimentally impact biogeographic congruence for morphological trees more or less than their molecular counterparts. Therefore, while either morphological or molecular trees may show better congruence in a particular case, biogeographic congruence still provides a valuable ancillary test of phylogenetic accuracy.

The biogeographic distribution of extant species arises by two main processes: vicariance and dispersal⁶⁷. Vicariance is the division of an ancestral area of sympatry by a physical barrier to create allopatric populations that may ultimately speciate, while dispersal is the migration or diffusion of individuals from some centre of endemism⁶⁸. The relative importances of these two processes remain controversial and probably depend upon environment and time scale. Vicariance is often invoked as a result of the formation of land barriers such as mountains or oceans while dispersal is associated with repeated migrations away from a reservoir⁶⁹ or centre of endemism⁷⁰, as well as with biotic interchanges⁷¹. Species distribution patterns are unlikely to be purely vicariant or dispersive⁷² and may be shaped by additional factors such

as range expansions ⁷³, migrations ⁷⁴ and extinctions ⁷⁵. Regardless of which process dominated, we expect the geographic regions assessed here (which are analogous to the areas that would form the basis of area cladograms ⁷⁶) to show some level of congruence with phylogeny and to yield non-random distributions. While we concede that all our indices would be likely to yield higher values for a purely vicariant than a purely dispersive pattern, there is no reason why morphological or molecular trees should be preferentially more congruent with either pattern. It is possible that selection pressures that cause similar adaptations to evolve in similar environments might result in a bias in favour of morphological trees where 'convergent' geographical transitions have occurred. However similar phenomena may also occur in molecular datasets. For example, there is increasing evidence that horizontal gene transfers have happened numerous times in green plants ⁷⁷ and other eukaryotes ⁷⁸. Some of these genes are associated with traits that likely conferred a selective advantage in particular environments, such as vascular tissues in land plants, pathogen resistance and the C4 photosynthesis pathway in grasses, and herbivory in insects. Under certain circumstances, therefore, selection for traits expressed by horizontally transferred genes could also result in mitochondrial trees reflecting biogeography more closely than the true phylogeny. Determining the potential impact of these phenomena, as well as the roles of dispersal and vicariance in the specific biogeographic patterns seen here would require much more detailed analyses. It would necessitate combining independent population or observational data on biogeographic transitions with time-calibrated phylogenies at the species or population level. Such data and trees are lacking for most clades, and morphological phylogenies at this resolution are almost unheard of. While such work would be invaluable, it is vastly beyond the scope of this study and would prohibitively reduce our sample size of case studies.

Despite the superiority of molecular trees, the reciprocal illumination of morphological and molecular data and the simultaneous "total evidence" analysis of multiple data types remain instrumental in resolving the deep relationships of many otherwise recalcitrant clades including arthropods ¹⁷, echinoderms ⁷⁹, angiosperms ⁸⁰ and embryophytes ⁸¹. Even the major

revisions to the mammalian phylogeny supported by molecular analyses have prompted subsequent re-evaluation of morphological data. The latter have subsequently yielded results in broad agreement with phylogenomic trees. Biogeographic congruence of both morphological and molecular trees was found to improve over research time (publication date), indicating that the quality of morphological as well as molecular trees has improved. This is likely to have resulted not only from advances in methodology, but also a trend for increasing phylogenetic dataset size, regardless of the type of data being analysed. We also note the reciprocal illumination of published molecular and morphological phylogenies through research time, although the nature of this influence on subjective aspects of taxon choice, optimality criteria and character coding is difficult to assess. Molecular phylogenies often impact on new comparative morphological analyses (particularly by prompting the re-evaluation of hypotheses of homology) but morphological trees can also influence our understanding of molecular evolution and phylogeny. For example, several earlier multigene and genome-wide phylogenies of major arthropod groups yielded a clade comprising myriapods and chelicerates^{82,83}, a group so strikingly at odds with comparative morphological analyses that it was named “Paradoxapoda”⁸⁴. Such findings prompted a re-evaluation of analytical models for sequence data as well as the adequacy of taxon sampling for deep and ancient divergences⁸⁵.

More generally, we believe that the continued importance of morphological data in phylogenetic analyses is assured. Not only is phylogenetics built on a legacy of morphological research but approximately 98% of species are extinct, and morphology remains the only source of data for exclusively fossil taxa⁸⁶. Moreover, fossils often realise combinations of character states that are unknown from the extant biota⁸⁷, sample otherwise extinct or sparsely populated branches of the tree, and preserve the order in which character states have evolved, thereby enabling a better appreciation of evolutionary transitions (e.g., fish-tetrapod transition⁸⁸ or theropod-bird transition⁸⁹). A better understanding of morphological evolution and fossilisation biases Sansom and Wills⁹⁰, as well as broader character sampling

⁹¹ will be key to obtaining more accurate molecular tree calibrations. Despite the development of increasingly sophisticated clock models ⁹², there is often a paucity of good fossil calibration dates ⁹³. We hope that our study will stimulate further ancillary biogeographic and stratigraphic tests of phylogenies inferred from a variety of morphological, molecular and combined data sets using different methodologies.

Methods

Dataset Compilation

We initially obtained 106 animal and plant phylogenetic trees from 61 papers published between 1981 and 2015. These were reduced to 48 pairs of morphological and molecular trees for the same clades (Supplementary Table 11), derived from the same paper whenever possible. Phylogenies were taken from the main text of the paper where possible, with supplementary material only being used if trees were not present in the main paper. In cases where multiple morphological or molecular phylogenies were given, we used those preferred by the authors. If the authors expressed no preference, we selected trees which had the most taxa, most characters or were most resolved, in that order. Trees with the greatest possible overlap in taxon sets were selected, subsequently pruning unique leaves to yield identical taxon sets (46% of trees had different sources, 24% of trees had one or more taxa pruned, and these had a mean of 63% of leaves pruned). Most clades (73%) were terrestrial and freshwater vertebrates with strong patterns of endemism, but insect (13%) and plant (15%) clades were also included. Only 10% of clades contained any marine taxa, partly a function of the difficulties of accurately ascertaining and coding regions in these environments.

Coding Biogeographic Distributions

To assess biogeographic congruence, region characters summarising the distributions of taxa were defined from biogeographic occurrence data which could then be mapped onto phylogenies (Supplementary Figure 16). Biogeographic data were obtained primarily from The IUCN Red List of Threatened Species, Version 2019-2 ³⁴ and checked using data from the

Global Biodiversity Information Facility ³⁵ where available. The Reptile Database ³⁶ was used for the reptile clades in the study, which were frequently poorly represented in the IUCN and GBIF databases. Biogeographic data from these sources was then checked against any available data from the original publications. Biogeographic data were collected in two forms: taxon presences defined at the highest resolution of areas available (e.g., 'California', 'U.S.A.' or 'North America') and point occurrences. Point occurrences were synthesised into a list of presences for areas at the highest resolution of the online database. Our approach to coding was inclusive insofar as taxa known from multiple regions were recorded as present in all of these regions. For each clade, lists were combined to create a biogeographic character matrix of presence/absence characters for each recognised region (column). Taxa were scored "1" if present in and "0" if absent from the smallest discrete regions listed. If these regions were at different scales for different taxa, the larger region was broken up into its constituent sub regions to match the finest scale represented, with taxa coded as present in the larger region also coded as present in all the constituent sub-regions. A matrix of characters, rather than a single multistate character, allowed for taxa that were observed from more than one region. Regions were then checked to ensure that none of them overlapped or were duplicates of the same geographic area. This yielded a full list of the least inclusive regions in which the members of the clade were found. As the areas being combined were often defined geopolitically or at the limited spatial resolution of our data, the regions derived from them were only biogeographically meaningful if they contained unique information about how taxa are grouped in space. Therefore, to avoid over-splitting of regions, we combined pairs of closest geographically neighbouring regions with identical taxon presence/absences into a single larger region and continued this process until all regions had unique taxon presence/absences. As it was not uncommon for biogeographic region matrices to contain more regions than taxa after this process (as a difference in presence for one taxon was sufficient to define a distinct region) we merged regions with single unique taxa (autapomorphic region characters) into their geographically closest neighbours.

To test whether the resulting biogeographic region matrices could potentially inform phylogenetic inferences, we assessed their non-random structure using matrix compatibility permutation tail probability (MCPTP) tests³⁸ (Supplementary Methods). Two characters are incompatible if it is not possible to map them onto the same evolutionary tree without homoplasy. The test statistic is therefore the number of compatibilities (viz incompatibilities) between all pairs of characters in a matrix. Applying this test to the biogeographic character matrices is a means of assessing their congruent hierarchical signal (and thus the biogeographic information that they represent), in precisely the same manner as a parsimony PTP. Fewer incompatibilities indicate a more highly structured character matrix which is more likely to be phylogenetically informative. Significant non-random structure in the biogeographic data might be considered as a necessary pre-requisite for using those same data as an ancillary test of the accuracy of trees inferred from different data types. If differences in biogeographic congruence are truly indicative of the relative accuracy of morphological and molecular trees, then such differences should also be evident when considering only those biogeographic matrices with significantly non-random (potentially phylogenetic) signal.

Testing Biogeographic Congruence

We assessed the fit of the biogeographic matrices onto both morphological and molecular trees using the ensemble consistency index (CI), ensemble retention index (RI) and biogeographic HER (bHER) (Supplementary Table 12). We note that the CI is biased by tree size, and by tree shape and balance with certain types of characters⁹⁴ (e.g., irreversible and ordered). We therefore also measured congruence using a modification of the homoplasy excess ratio (HER) of Archie³⁷. Our biogeographic HER (bHER) was calculated by comparing the additional step length over and above the minimum necessary (the observed length for our data (L) minus the minimum possible given the number and nature of characters (MINL)) with the mean additional step length from lengths for biogeographically randomly permuted data (MEANNS) (randomly reassigning rows in the data matrix to the taxa 10,000 times, while holding tree topology constant). The bHER (or, more precisely, our modified MEANNS)

therefore differed from the HER in its original form by permuting rows of the matrix across taxa (rather than the entries within each column separately) and by calculating the length of the original and permuted biogeographic matrices on the morphological or molecular tree (rather than inferring a tree from these data). By permuting rows of codes across taxa (rather than each column of data across taxa independently), we ensured that there were no unrealised or unlikely combinations of regional distribution patterns. Specifically, $bHER = 1 - (L - MINL) / (MEANNS - MINL)$ (see Supplementary Methods for full details). A similar procedure was also used to produce a distribution of tree length values from randomly permuted biogeographic data, against which the original tree length could be compared to yield approximate p-values (the probability that a length as short or shorter could be observed for biogeographic data distributed at random on the tree). This is equivalent to a randomisation test for both CI and RI and will yield the same p-values for both metrics by definition. All analyses therefore accounted for the expected congruence if rows of region characters were randomly distributed across taxa. This was factored into how bHER was calculated, whilst for CI and RI it was controlled with an ancillary randomisation test. More specifically, this null expectation is factored into calculating MEANNS and therefore the scaling of the index. This ensured that, unlike CI and RI, bHER was already standardised relative to the expected fit of the region characters onto the tree of interest.

As most metrics were not normally distributed (Supplementary Table 13), non-parametric statistical tests were used in most cases. Correlations between biogeographic fit metrics and other variables of interest were assessed to determine whether confounding variables might affect our results. Breusch-Pagan tests indicated that the residuals from regressions between metrics of interest did not show significant heteroskedasticity in most but not all cases (Supplementary Table 14). Given that data might be non-normal, and relationships may be non-linear, Spearman-rank correlation was preferred, with Pearson's correlations also being calculated on the data after the identification and removal of outliers. Five groups contained molecular datasets far larger than all others (more than 9,000 characters) and were classed

as outliers. Each metric was tested against the number of phylogenetic characters in the source dataset (size: Supplementary Figure 17, Supplementary Table 15), the year in which the phylogeny was published (publication year: Supplementary Figure 15, Supplementary Table 12), the number of terminal taxa (taxa: Supplementary Figure 18, Supplementary Table 16), the ratio of region characters to terminal taxa (region characters/taxa: Supplementary Figure 19, Supplementary Table 17) and the ratio of phylogenetic characters to terminal taxa (S/T: Supplementary Table 18). The bHER, CI, RI and the p-values from CI & RI randomisation tests for morphological and molecular tree samples were compared using two-tailed paired Wilcoxon signed-rank tests using 'wilcox.test' in R. In each case, the functions 'wilcoxonZ' and 'wilcoxonPairedRC' from the package 'rcompanion' were used to calculate Z-scores and effect sizes as given by the matched-pairs rank biserial correlation coefficient. In addition, two-tailed sign tests were used to test whether selecting the most biogeographically congruent tree in each pair resulted in significantly more molecular or morphological trees being chosen than expected by chance.

Testing Stratigraphic Congruence

Data on the fossil record of each of the 48 clades in this study were collated from the Fossilworks portal of the Palaeobiology database ⁹⁵ (PBDB) and Benton 1993 ⁹⁶, as well as data within the source papers (Supplementary Methods). 23 Clades had published fossil data for at least 50% of their leaves, and so were judged suitable for tests of stratigraphic congruence. First and last occurrences for all taxa were assigned at the stage-level after O'Connor *et al.* 2016 ³⁹, using the International stratigraphic chart ⁹⁷, the Geologic Timescale 2004 ⁹⁸ and the GeoWhen database ⁹⁹. Low preservation potential and scarcity often ensure that first fossil occurrences lag behind true times of origin, while scarcity prior to the actual point of extinction mean that lineages are lost from the record prematurely (the 'Signor-Lipps effect'). Where stratigraphy was unresolved at the stage level, taxa were therefore assigned to the first stage in the time interval given for their first occurrence and the last interval of the time period for their last occurrence. Stratigraphic congruence was assessed using several

previously published and commonly utilised metrics, namely the stratigraphic consistency index (SCI), modified Manhattan stratigraphic measure (MSM*), the gap excess ratio and its modification (GER and GER*). The stratigraphic congruence of morphological and molecular trees was assessed using paired Wilcoxon signed-rank tests as well as sign tests, in a similar manner to that detailed for the biogeographic congruence tests.

Data Availability

The data that support the findings of this study are available on the website figshare (<https://figshare.com/>) with the identifier <https://doi.org/10.6084/m9.figshare.c.5946358>, in addition to being available from the authors upon request.

Code Availability

All custom scripts and programs used to calculate bHER, randomly permute region matrices and carry out MCPTP tests are available from the authors upon request.

Acknowledgements

We thank Tim Astrop for useful discussions and suggestions related to plotting the data as well as Tamás Székely, Polly Russell and Catherine Klein for useful discussions. JWO & MAW's work was funded by the John Templeton Foundation grants 61408 and 43915. MAW's work was funded by BBSRC grants BB/K015702/1 and BB/K006754/1, as well as BBSRC studentship 1923592.

Author Information

Affiliations

Milner Centre for Evolution, University of Bath, Bath, UK

Jack W. Oyston & Matthew A. Wills

Vertebrates Division, Life Sciences, Natural History Museum, London, UK

Mark Wilkinson

School of Life Sciences, College of Science, University of Lincoln

Marcello Ruta

Contributions

J.W.O. and M.A.W. conceived the study, devised tests of biogeographic congruence, developed the methods and theory, and wrote the paper. M.A.W. devised and scripted the bHER and other permutation tests. J.W.O. compiled the data, undertook all primary analyses and devised/drafted all figures. M.W. carried out the compatibility tests, analysed the data, and performed the simulations. M.W. and M.R. analysed data and contributed text to the introduction and discussion.

Corresponding authors

Correspondence to Jack W. Oyston or Matthew A. Wills.

Ethics declarations

Competing interests

The authors declare no competing interests.

Table 1: Biogeographic and stratigraphic congruence of morphological and molecular phylogenies.

Metric	Median (morphological trees)	Median (molecular trees)	Wilcoxon signed-rank test statistic (W)	Z-score	Effect size (rc)	P-value
Year	2003	2003.5	59	0.947	0.297	0.362
Size	91	2221.5	1174	6.01	0.997	2.132x10 ⁻¹⁴ *
S/T	4.852941	79.94878	1173	6	0.995	3.553x10 ⁻¹⁴ *
Res	0.9168956	0.9928571	534	2.75	0.519	0.006035*
C	0.2666665	0.2548875	508.5	-0.815	-0.135	0.4178
CI	0.276377	0.27705	685	2.22	0.384	0.027*
RI	0.183279	0.2110125	695	2.33	0.404	0.0199*
CI & RI p-value	0.025897	0.013849	373	-1.63	-0.279	0.104
bHER	0.1078203	0.1533195	888	3.08	0.51	0.002*
SCI	0.529	0.55	140.5	1.33	0.338	0.1913
MSM*	0.169	0.196	92	-0.121	-0.0316	0.9198
GER	0.571	0.588	91	-0.523	-0.133	0.6142
GER*	0.826	0.838	90	0.196	0.0526	0.8617

Results of paired Wilcoxon signed-rank tests on the two data partitions (morphological & molecular) for the following metrics: phylogeny publication year (Year), number of phylogenetic characters underpinning the source trees (Size), number of phylogenetic characters divided by number of taxa (S/T), proportion of resolved nodes (Res), Colless's index of tree balance (C), consistency index (CI), retention index (RI), probability of CI & RI values falling within the null distribution (CI & RI p-value), biogeographic homoplasy excess ratio (bHER), stratigraphic consistency index (SCI), the modified Manhattan stratigraphic measure (MSM*), the gap excess ratio (GER) and the modified gap excess ratio (GER*). The sample size is 46 trees for SCI, MSM*, GER, GER* and 96 trees for all other metrics. Effect sizes were calculated using the matched-pairs rank biserial correlation coefficient. Statistically significant results at the 95% confidence interval are indicated with an asterisk. n = 23 biologically independent pairs of morphological and molecular phylogenies for all stratigraphic congruence metrics (SCI, MSM*, GER, GER*) and N = 48 biologically independent pairs of morphological and molecular phylogenies for all other metrics.

Table 2: Biogeographic congruence metrics modelled by potential confounding variables.

Model	Linear Regression			Wilcoxon signed-rank test			
	AIC	R ²	P-value	W	Z-score	Effect size (rc)	P-value
CI ~ Age + log(Regions) + log(Taxa) + S/T	-152.119	0.5543	4.1x10 ⁻¹⁶	994	4.16	0.69	1.111x10 ⁻⁵
CI ~ Age + C + log(Regions) + log(Taxa) + S/T + Year + Res	-146.239	0.5397	5.683x10 ⁻¹⁴	1040	4.64	0.769	5.818x10 ⁻⁷
bHER ~ log(Regions) + Year + Res	-63.981	0.1139	0.0027	827	2.45	0.406	0.01349
bHER ~ Age + C + log(Regions) + log(Taxa) + S/T + Year + Res	-57.635	0.0894	0.0313	793	2.1	0.349	0.03511
RI ~ Age + log(Regions) + Res	-55.5291	0.1336	0.0010	568	0.3	0.0509	0.768
RI ~ Age + C + log(Regions) + log(Taxa) + S/T + Year + Res	-48.3497	0.1019	0.0199	529	-0.605	-0.1	0.5518

Results of models predicting measures of consistency index (CI), retention index (RI) and biogeographic homoplasy excess ratio (bHER) of geographic region characters from the age of the clade root (Age), Colless's index of tree balance (C), log of the number of regions (log(Regions)), log of the number of terminal taxa (log(Taxa)), number of phylogenetic characters divided by number of taxa (S/T), phylogeny publication year (Year) and the proportion of resolved nodes (Res). Both the model with all explanatory variables and the model with minimal Akaike information criterion (AIC) are given for each congruence measure. Wilcoxon signed-rank tests are between model residuals for morphological and molecular trees from weighted robust linear regression models and effect sizes were calculated using the matched-pairs rank biserial correlation coefficient (rc). N = 96 morphological and molecular phylogenies (regression models) and N = 48 biologically independent pairs of morphological and molecular phylogenies (Wilcoxon signed-rank tests).

Fig. 1: Testing the biogeographic congruence of phylogenetic trees.

(a) Defining biogeographic regions and coding taxon presences and absences. **1.** Occurrence data on the distribution of extant species is used to produce a list of biogeographic regions for the clade and to summarise ranges for the taxa in the published phylogenies. **2.** This distributional information is converted into a matrix of binary characters representing taxa in biogeographic regions, where 0 indicates the taxon is absent and 1 indicates the taxon is present. **3.** Characters in the occurrence matrix are mapped onto the morphological and molecular phylogeny selected for each clade, allowing standard measures of character fit (CI, RI) to be calculated for each tree. (b) Presence and absence codings in each matrix are randomly reassigned to taxa, keeping the presence and absence codings fixed for each row. Characters from the new randomly permuted matrix are mapped onto the original trees and both CI and RI are recalculated. The entire randomisation process is performed 10,000 times. (c) The 10,000 CI and RI values from matrices' biogeographic region reassignments form a null distribution of expected congruence values if taxa in the clade were randomly distributed in biogeographic regions. The observed CI and RI of region characters for a given tree is compared to the null distribution for that same tree to determine whether the observed biogeographic congruence value lies outside of the 95% confidence interval.

Fig. 2: Biogeographic congruence in morphological and molecular phylogenies.

Binary biogeographic region characters mapped onto paired morphological and molecular phylogenies. (a) Placental mammals (Eutheria) from O'Leary et al. 2013¹⁰⁰. (b) Caribbean boas (*Chilabothrus/Epicrates*), with the morphological tree taken from Kluge 1989¹⁰¹ and the molecular tree taken from Tolson 1987¹⁰². Regions for which the terminal taxon is coded present are represented as coloured pie slices. Consistency index (CI), retention index (RI) and biogeographic HER (bHER) values given are for the matrix of biogeographic region presences and absences, while CI & RI p-value is calculated using 10,000 randomly permuted region matrices.

Fig. 3: Differences in biogeographic congruence between morphological and molecular trees.

Boxplots of raw values and differences in values between morphological and molecular trees for the metrics of biogeographic congruence analysed in this study. **(a)** Consistency index (CI: $W = 685$, $Z = 2.22$, $rc = 0.384$, $p\text{-value} = 0.027$). **(b)** Retention index (RI: $W = 695$, $Z = 2.33$, $rc = 0.404$, $p\text{-value} = 0.0199$). **(c)** P-values for the CI & RI random permutations (CI & RI $p\text{-value}$: $W = 373$, $Z = -1.63$, $rc = -0.279$, $p\text{-value} = 0.104$). **(d)** Biogeographic HER (bHER: $W = 888$, $Z = 3.08$, $rc = 0.51$, $p\text{-value} = 0.002$). Boxes delimit the upper and lower quartiles of the data, while central bars are median values. Whiskers delimit plus or minus 1.5 times the inter-quartile range, from the first and third quartiles. Coloured lines connected pairs of values from the same clade, where red dashed lines indicate the morphological tree is most biogeographically congruent and green solid lines indicate the molecular tree is most biogeographically congruent. $N = 48$ biologically independent pairs of morphological and molecular phylogenies.

Fig. 4: The number of morphological and molecular trees most congruent with biogeography.

Comparison of the number of trees in each sample (morphological or molecular) with a greater biogeographic fit than its counterpart. **(a)** Consistency index (CI), grey bars show totals for the whole sample, coloured bars indicate totals in the subset significantly different from the expected null (CI & RI $p\text{-value} < 0.05$). **(b)** Retention index (RI), grey bars show totals for the whole sample, coloured bars indicate totals in the subset significantly different from the expected null ($p\text{-values}$ for the CI & RI from random permutations < 0.05). **(c)** P-values for the CI & RI random permutations (CI & RI $p\text{-value}$), where grey bars show totals for the whole sample, coloured bars are clades with values < 0.05 . **(d)** biogeographic HER (bHER), counts are for the whole dataset. Bars show the number of clades in each subset, with binomial confidence intervals calculated using the approach of Clopper and Pearson¹⁰³. $N = 48$ biologically independent pairs of morphological and molecular phylogenies.

References

- 1 Harvey, P. H. & Pagel, M. D. *The comparative method in evolutionary biology*. Vol. 239 (Oxford University Press, 1991).
- 2 Oyston, J. W., Hughes, M., Wagner, P. J., Gerber, S. & Wills, M. A. What limits the morphological disparity of clades? *Interface Focus* **5**, 0042, doi:10.1098/rsfs.2015.0042 (2015).
- 3 Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491**, 444-448 (2012).
- 4 Webb, C. O. Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *The American Naturalist* **156**, 145-155 (2000).
- 5 Purvis, A., Gittleman, J. L. & Brooks, T. *Phylogeny and conservation*. (Cambridge University Press, 2005).
- 6 Page, R. D. M. Parallel phylogenies: reconstructing the history of host-parasite assemblages. *Cladistics* **10**, 155-173 (1994).
- 7 Weaver, S. C. & Vasilakis, N. Molecular evolution of dengue viruses: contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. *Infection, Genetics and Evolution* **9**, 523-540 (2009).
- 8 Tassy, P. Trees before and after Darwin. *Journal of Zoological Systematics and Evolutionary Research* **49**, 89-101 (2011).
- 9 Heather, J. M. & Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **107**, 1-8 (2016).
- 10 Pyron, R. A. Post-molecular systematics and the future of phylogenetics. *Trends in Ecology & Evolution* **30**, 384-389 (2015).
- 11 Sansom, R. S. & Wills, M. A. Differences between hard and soft phylogenetic data. *Proceedings of the Royal Society B: Biological Sciences* **284**, 20172150, doi:10.1098/rspb.2017.2150 (2017).
- 12 Scotland, R. W., Olmstead, R. G. & Bennett, J. R. Phylogeny reconstruction: the role of morphology. *Systematic Biology* **52**, 539-548 (2003).
- 13 Regier, J. C. *et al.* Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. *Nature* **463**, 1079-1083 (2010).
- 14 Callender-Crowe, L. M. & Sansom, R. S. Osteological characters of birds and reptiles are more congruent with molecular phylogenies than soft characters are. *Zoological Journal of the Linnean Society* **194**, 1-13 (2022).
- 15 Wahlberg, N. *et al.* Synergistic effects of combining morphological and molecular data in resolving the phylogeny of butterflies and skippers. *Proceedings of the Royal Society B: Biological Sciences* **272**, 1577-1586 (2005).
- 16 He, L. *et al.* A molecular phylogeny of selligieoid ferns (Polypodiaceae): Implications for a natural delimitation despite homoplasy and rapid radiation. *Taxon* **67**, 237-249 (2018).

- 17 Fernández, R., Edgecombe, G. D. & Giribet, G. Phylogenomics illuminates the backbone of the Myriapoda Tree of Life and reconciles morphological and molecular phylogenies. *Scientific Reports* **8**, 1-7 (2018).
- 18 Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of eukaryotes. *Nature Reviews Microbiology* **15**, 711-723 (2017).
- 19 Asher, R. J., Bennett, N. & Lehmann, T. The new framework for understanding placental mammal evolution. *BioEssays* **31**, 853-864 (2009).
- 20 Shoshani, J. & McKenna, M. C. Higher taxonomic relationships among extant mammals based on morphology, with selected comparisons of results from molecular data. *Molecular Phylogenetics and Evolution* **9**, 572-584 (1998).
- 21 Beck, R. M. D. & Baillie, C. Improvements in the fossil record may largely resolve current conflicts between morphological and molecular estimates of mammal phylogeny. *Proceedings of the Royal Society B: Biological Sciences* **285**, 20181632, doi:10.1098/rspb.2018.1632 (2018).
- 22 Zou, Z. T. & Zhang, J. Z. Morphological and molecular convergences in mammalian phylogenetics. *Nature Communications* **7**, 1-9 (2016).
- 23 Hillis, D. M. Molecular versus morphological approaches to systematics. *Annual Review of Ecology and Systematics* **18**, 23-42 (1987).
- 24 Thompson, N. *Alfred Russell Wallace Contributions to the theory of Natural Selection, 1870, and Charles Darwin and Alfred Wallace, 'On the Tendency of Species to form Varieties' (Papers presented to the Linnean Society 30th June 1858)*. (Routledge, 2004).
- 25 Croizat, L. *Panbiogeography; or an introductory synthesis of zoogeography, phytogeography, and geology, with notes on evolution, systematics, ecology, anthropology, etc.*, Vol. 1, 2a & 2b (Published by the author, Caracas. , 1958).
- 26 Means, J. C. & Marek, P. E. Is geography an accurate predictor of evolutionary history in the millipede family Xystodesmidae? *PeerJ* **5**, e3854, doi:10.7717/peerj.3854 (2017).
- 27 Wills, M. A., Barrett, P. M. & Heathcote, J. F. The modified gap excess ratio (GER*) and the stratigraphic congruence of dinosaur phylogenies. *Systematic Biology* **57**, 891-904 (2008).
- 28 Fisher, D. C. Stratocladistics: integrating temporal data and character data in phylogenetic inference. *Annual Review of Ecology, Evolution and Systematics* **39**, 365-385 (2008).
- 29 Lazarus, D. B. & Prothero, D. R. The role of stratigraphic and morphologic data in phylogeny. *Journal of Paleontology* **58**, 163-172 (1984).
- 30 Camerini, J. R. Evolution, biogeography, and maps: an early history of Wallace's Line. *Isis* **84**, 700-727 (1993).
- 31 Upchurch, P., Hunn, C. A. & Norman, D. B. An analysis of dinosaurian biogeography: evidence for the existence of vicariance and dispersal patterns caused by geological events. *Proceedings of the Royal Society B: Biological Sciences* **269**, 613-621 (2002).
- 32 Ferreira, G. S., Bronzati, M., Langer, M. C. & Sterli, J. Phylogeny, biogeography and diversification patterns of side-necked turtles (Testudines: Pleurodira). *Royal Society Open Science* **5**, 171773, doi:10.1098/rsos.171773 (2018).
- 33 Ronquist, F. & Sanmartín, I. Phylogenetic methods in biogeography. *Annual Review of Ecology, Evolution, and Systematics* **42**, 441-464 (2011).

- 34 IUCN. *The IUCN Red List of Threatened Species. Version 2019-2.*, <<https://www.iucnredlist.org>> (2019).
- 35 GBIF.org. *GBIF Home Page*, <<https://www.gbif.org/>> (2019).
- 36 Uetz, P., Freed, P., Aguilar, R. & Hošek, J. *The reptile database.*, <<http://www.reptiledatabase.org>> (2019).
- 37 Archie, J. W. Homoplasy excess ratios: new indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. *Systematic Zoology* **38**, 253-269 (1989).
- 38 Wilkinson, M. On phylogenetic relationships within Dendrotriton (Amphibia: Caudata: Plethodontidae) is there sufficient evidence? *Herpetological Journal* **7**, 55-65 (1997).
- 39 O'Connor, A. & Wills, M. A. Measuring stratigraphic congruence across trees, higher taxa, and time. *Systematic Biology* **65**, 792-811 (2016).
- 40 Colless, D. H. Review of phylogenetics: the theory and practice of phylogenetic systematics. *Systematic Zoology* **31**, 100-104 (1982).
- 41 Lartillot, N. & Philippe, H. Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**, 1463-1472 (2008).
- 42 Sansom, R. S., Choate, P. G., Keating, J. N. & Randle, E. Parsimony, not Bayesian analysis, recovers more stratigraphically congruent phylogenetic trees. *Biology Letters* **14**, 20180263, doi:10.1098/rsbl.2018.0263 (2018).
- 43 Rosa, B. B., Melo, G. A. & Barbeitos, M. S. Homoplasy-based partitioning outperforms alternatives in Bayesian analysis of discrete morphological data. *Systematic Biology* **68**, 657-671 (2019).
- 44 Lucena, D. A. & Almeida, E. A. Morphology and Bayesian tip-dating recover deep Cretaceous-age divergences among major chrysidid lineages (Hymenoptera: Chrysididae). *Zoological Journal of the Linnean Society* **194**, 36-79 (2022).
- 45 O'Reilly, J. E. *et al.* Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biology Letters* **12**, 20160081, doi:10.1098/rsbl.2016.0081 (2016).
- 46 Smith, M. R. Bayesian and parsimony approaches reconstruct informative trees from simulated morphological datasets. *Biology Letters* **15**, 20180632, doi:10.1098/rsbl.2018.0632 (2019).
- 47 Wiens, J. The role of morphological data in phylogeny reconstruction. *Systematic Biology* **53**, 653-661 (2004).
- 48 O'Leary, M. A. & Kaufman, S. G. *MorphoBank 3.0: Web application for morphological phylogenetics and taxonomy.* , <<http://www.morphobank.org>> (2012).
- 49 de Queiroz, A. & Gatesy, J. The supermatrix approach to systematics. *Trends in Ecology & Evolution* **22**, 34-41 (2007).
- 50 Wilkinson, M. A comparison of two methods of character construction. *Cladistics* **11**, 297-308 (1995).

- 51 Brazeau, M. D. Problematic character coding methods in morphology and their effects. *Biological Journal of the Linnean Society* **104**, 489-498 (2011).
- 52 Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**, e88, doi:10.1371/journal.pbio.0040088 (2006).
- 53 O'Reilly, J. E., Puttick, M. N., Pisani, D. & Donoghue, P. C. Probabilistic methods surpass parsimony when assessing clade support in phylogenetic analyses of discrete morphological data. *Palaeontology* **61**, 105-118 (2018).
- 54 Keating, J. N., Sansom, R. S., Sutton, M. D., Knight, C. G. & Garwood, R. J. Morphological phylogenetics evaluated using novel evolutionary simulations. *Systematic Biology* **69**, 897-912 (2020).
- 55 Makarenkov, V. *et al.* Weighted bootstrapping: a correction method for assessing the robustness of phylogenetic trees. *BMC Evolutionary Biology* **10**, 1-16 (2010).
- 56 Stayton, C. T. The definition, recognition, and interpretation of convergent evolution, and two new measures for quantifying and assessing the significance of convergence. *Evolution* **69**, 2140-2153 (2015).
- 57 Sattler, R. Homology - a continuing challenge. *Systematic Botany* **9**, 382-394 (1984).
- 58 Jenner, R. A. & Schram, F. R. The grand game of metazoan phylogeny: rules and strategies. *Biological Reviews* **74**, 121-142 (1999).
- 59 Pisani, D. & Wilkinson, M. Matrix representation with parsimony, taxonomic congruence, and total evidence. *Systematic Biology* **51**, 151-155 (2002).
- 60 Arcila, D. *et al.* Testing the utility of alternative metrics of branch support to address the ancient evolutionary radiation of tunas, stromateoids, and allies (Teleostei: Pelagiaria). *Systematic Biology* **70**, 1123-1144 (2021).
- 61 Felsenstein, J. Phylogenies and the comparative method. *The American Naturalist* **125**, 1-15 (1985).
- 62 Bremer, K. Branch support and tree stability. *Cladistics* **10**, 295-304 (1994).
- 63 Johnson, W. E. *et al.* The late Miocene radiation of modern Felidae: a genetic assessment. *Science* **311**, 73-77 (2006).
- 64 Van der Made, J. Biogeography and climatic change as a context to human dispersal out of Africa and within Eurasia. *Quaternary Science Reviews* **30**, 1353-1367 (2011).
- 65 May, F., Rosenbaum, B., Schurr, F. M. & Chase, J. M. The geometry of habitat fragmentation: Effects of species distribution patterns on extinction risk due to habitat conversion. *Ecology and Evolution* **9**, 2775-2790 (2019).
- 66 Swofford, D. L. *et al.* Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Systematic Biology* **50**, 525-539 (2001).
- 67 Jaeger, J. J. & Martin, M. African marsupials - vicariance or dispersion? *Nature* **312**, 379-379 (1984).
- 68 Smith, B. T. *et al.* The drivers of tropical speciation. *Nature* **515**, 406-409 (2014).
- 69 Simkanin, C. *et al.* Exploring potential establishment of marine rafting species after transoceanic long-distance dispersal. *Global Ecology and Biogeography* **28**, 588-600 (2019).

- 70 Raxworthy, C. J., Forstner, M. R. J. & Nussbaum, R. A. Chameleon radiation by oceanic dispersal. *Nature* **415**, 784-787 (2002).
- 71 Stehli, F. G. & Webb, S. D. *The great American biotic interchange.*, Vol. 4 (Springer Science & Business Media, 2013).
- 72 Ronquist, F. Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Systematic Biology* **46**, 195-203, doi:10.2307/2413643 (1997).
- 73 Ricklefs, R. E. & Bermingham, E. The concept of the taxon cycle in biogeography. *Global Ecology and Biogeography* **11**, 353-361 (2002).
- 74 Ma, H. An analysis of the equilibrium of migration models for biogeography-based optimization. *Information Sciences* **180**, 3444-3464 (2010).
- 75 Yiming, L., Niemelä, J. & Dianmo, L. Nested distribution of amphibians in the Zhoushan archipelago, China: can selective extinction cause nested subsets of species? *Oecologia* **113**, 557-564 (1998).
- 76 Crisci, J. V., Katinas, L. & Posadas, P. *Historical Biogeography: An Introduction.* (Harvard University Press, 2003).
- 77 Chen, R. *et al.* Adaptive innovation of green plants by horizontal gene transfer. *Biotechnology Advances* **46**, 107671, doi:10.1016/j.biotechadv.2020.107671 (2021).
- 78 Schönknecht, G., Weber, A. P. & Lercher, M. J. Horizontal gene acquisitions by eukaryotes as drivers of adaptive evolution. *BioEssays* **36**, 9-20 (2014).
- 79 Smith, A. B. Echinoderm phylogeny: morphology and molecules approach accord. *Trends in Ecology & Evolution* **7**, 224-229 (1992).
- 80 Bateman, R. M., Hilton, J. & Rudall, P. J. Morphological and molecular phylogenetic context of the angiosperms: contrasting the 'top-down' and 'bottom-up' approaches used to infer the likely characteristics of the first flowers. *Journal of Experimental Botany* **57**, 3471-3503 (2006).
- 81 Morris, J. L. *et al.* The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences* **115**, E2274-E2283, doi:10.1073/pnas.1719588115 (2018).
- 82 Richter, S. The Tetraconata concept: hexapod-crustacean relationships and the phylogeny of Crustacea. *Organisms Diversity & Evolution* **2**, 217-237 (2002).
- 83 Dunn, C. W. *et al.* Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745-749 (2008).
- 84 Caravas, J. & Friedrich, M. Of mites and millipedes: recent progress in resolving the base of the arthropod tree. *BioEssays* **32**, 488-495 (2010).
- 85 Howard, R. J. *et al.* The Ediacaran origin of Ecdysozoa: integrating fossil and phylogenomic data. *Journal of the Geological Society*, doi:10.1144/jgs2021-107 (2022).
- 86 Newman, M. E. J. A model of mass extinction. *Journal of Theoretical Biology* **189**, 235-252 (1997).
- 87 Cobbett, A., Wilkinson, M. & Wills, M. A. Fossils impact as hard as living taxa in parsimony analyses of morphology. *Systematic Biology* **56**, 753-766 (2007).

- 88 Ruta, M., Krieger, J., Angielczyk, K. & Wills, M. A. The evolution of the tetrapod humerus: morphometrics, disparity, and evolutionary rates. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh* **109**, 351-369 (2018).
- 89 Puttick, M. N., Thomas, G. H. & Benton, M. J. High rates of evolution preceded the origins of birds. *Evolution* **68**, 1497-1510 (2014).
- 90 Sansom, R. S. & Wills, M. A. Fossilization causes organisms to appear erroneously primitive by distorting evolutionary trees. *Scientific Reports* **3**, 1-5 (2013).
- 91 Brinkworth, A., Sansom, R. & Wills, M. A. Phylogenetic incongruence and homoplasy in the appendages and bodies of arthropods: why broad character sampling is best. *Zoological Journal of the Linnean Society* **187**, 100-116 (2019).
- 92 Brown, J. W. & Smith, S. A. The past sure is tense: on interpreting phylogenetic divergence time estimates. *Systematic Biology* **67**, 340-353 (2018).
- 93 Barba-Montoya, J., Dos Reis, M. & Yang, Z. H. Comparison of different strategies for using fossil calibrations to generate the time prior in Bayesian molecular clock dating. *Molecular Phylogenetics and Evolution* **114**, 386-400 (2017).
- 94 Sanderson, M. J. & Donoghue, M. J. Patterns of variation in levels of homoplasy. *Evolution* **43**, 1781-1795 (1989).
- 95 Alroy, J. *Fossilworks: Gateway to the Paleobiology Database*, <<http://fossilworks.org>> (2019).
- 96 Benton, M. J. *The Fossil Record 2*. (Chapman & Hall, 1993).
- 97 Cohen, K. M., Harper, D. A. T. & Gibbard, P. L. *ICS International Chronostratigraphic Chart 2021/02*, <<http://www.stratigraphy.org/>> (2021).
- 98 Gradstein, F. & Ogg, J. Geologic time scale 2004—why, how, and where next! *Lethaia* **37**, 175-181 (2004).
- 99 Rohde, R. A. *The GeoWhen Database*, 2005).
- 100 O'Leary, M. A. *et al.* The placental mammal ancestor and the post-K-Pg radiation of placentals. *Science* **339**, 662-667 (2013).
- 101 Kluge, A. G. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Systematic Biology* **38**, 7-25 (1989).
- 102 Tolson, P. J. Phylogenetics of the boid snake genus *Epicrates* and Caribbean vicariance theory. *Occasional Papers of the Museum of Zoology, The University of Michigan* **715**, 1-68 (1987).
- 103 Clopper, C. J. & Pearson, E. S. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**, 404-413 (1934).