

SCIENTIFIC REPORTS



OPEN

A temporal banding approach for consistent taxonomic ranking above the species level

Ekaphan Kraichak^{1,2}, Ana Crespo³, Pradeep K. Divakar³, Steven D. Leavitt⁴ & H. Thorsten Lumbsch⁵

Comparable taxonomic ranks within clades can facilitate more consistent classifications and objective comparisons among taxa. Here we use a temporal approach to identify taxonomic ranks. This is an extension of the temporal banding approach including a Temporal Error Score that finds an objective cut-off for each taxonomic rank using information for the current classification. We illustrate this method using a data set of the lichenized fungal family Parmeliaceae. To assess its performance, we simulated the effect of taxon sampling and compared our method with the other temporal banding method. For our sampled phylogeny, 11 of the 12 included families remained intact and 55 genera were confirmed, whereas 32 genera were lumped and 15 genera were split. Taxon sampling impacted the method at the genus level, whereas yielded only insignificant changes at the family level. The other available temporal approach also gives a similar cutoff point to our method. Our approach to identify taxonomic ranks enables taxonomists to revise and propose classifications on an objective basis, changing ranks of clades only when inconsistent with most taxa in a phylogenetic tree. An R script to find the time point with the minimal temporal error is provided.

Conservative estimates of the number of species on this planet suggest that at least 8.7 million eukaryotes exist¹, with the fungal kingdom alone forecast to contain several million species². Overall, up to 100 million species are predicted to occupy the planet³. Regardless of the exact number, the sheer magnitude of species diversity requires a classification system that allows effective organization and communication of complex patterns of organismal diversity.

Beginning with Aristotle^{4,5}, organisms were classified according to their similarities. Linnaeus subsequently used a hierarchical classification and understood this pre-Darwinian “natural system” as a reflection of a divine plan of creation^{6,7}, as the idea of evolution was mostly alien to the scientists of the time^{4,5}. However, Linnaeus’ hierarchical classification system persisted and is still in use today. Attempts to develop non-hierarchical systems, such as a quinary system or a periodic system, were unsuccessful⁸.

In a phylogenetic context, taxonomic ranks in this hierarchical classification represent clades with a shared evolutionary history. Since members of a clade are derived from a common ancestor, they often share phenotypic traits. This helps explain why only seven years after Darwin’s seminal book, Haeckel could publish a tree of life, which reinterpreted the classification in a phylogenetic framework⁹. Taxonomic ranks in a hierarchical classification serve an important role for communication among biologists, and other disciplines, for comparative ecological, and conservation studies. However, it should be noted that these ranks are inherently arbitrary; hence there is no absolute definition of specific ranks, which partly explains disparities in ranking among classifications. Taxonomic circumscriptions of the same rank at varying phylogenetic scales lead to a number of potential biases when making comparisons among taxa. These discrepancies have been documented at higher-level taxonomic ranks, which have been shown to circumscribe vastly different phylogenetic scales in a number of cases^{8,10–14}. At the genus level, the current binomial system may lead to frequent name changes, thus making communication

¹Department of Botany, Kasetsart University, 50 Ngamwongwan Road, Ladyao, Chatuchak, Bangkok, 10900, Thailand. ²Center for Advanced Studies in Tropical Natural Resources, NRU-KU, Kasetsart University, Chatuchak, Bangkok, 10900, Thailand. ³Departamento de Biología Vegetal II, Facultad de Farmacia, Universidad Complutense de Madrid, Madrid, 28040, Spain. ⁴Department of Biology and M.L. Bean Life Science Museum, Brigham Young University, 4102 Life Science Building, Provo, UT, 84602, USA. ⁵Science & Education, Field Museum of Natural History, 1400 S Lake Shore Drive, Chicago, IL, 60605, USA. Correspondence and requests for materials should be addressed to E.K. (email: ekaphan.k@ku.th)

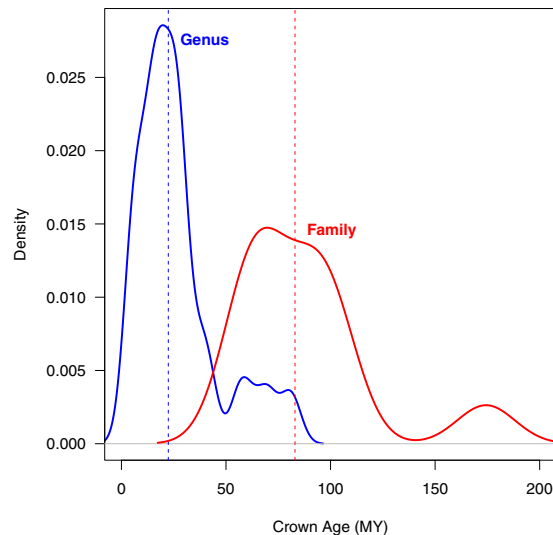


Figure 1. Distribution of crown ages of currently accepted genera (blue) and families (red) in the studied data set (Parmeliaceae, Ascomycota). The median for both ranks indicated by a dashed line.

difficult and potentially adding confusion^{15–19}. As a consequence, comparisons using higher-level ranks across different organismal groups are potentially flawed.

To make taxonomic ranks at supraspecific level more consistent and increase meaningful comparability among higher taxonomic ranks, the use of a standardized time of divergence has been suggested as a universal yardstick for the assignment of ranks^{8, 11–13, 20–22}. This means that groups of organisms are given the same rank if they originated in the standardized geological time period.

In this study we use a method that employs time-calibrated chronograms to identify upper and lower thresholds for taxonomic ranks based on the temporal banding approach^{11, 13}. To this end, we propose an extension of Holt and Jönsson’s Temporal Error Score²⁰ to find an objective cut-off for each taxonomic rank, using the information from the current classification. To illustrate the utility of this temporal banding approach, we apply this modified approach to the most comprehensive phylogeny of the hyperdiverse lichen-forming fungal family Parmeliaceae. We briefly describe the temporal approach with Temporal Error Score²⁰ and its computational limitations. Finally, we illustrate how the Temporal Error Score can be used to find the time point with the minimal error to modify the current classification. Through simulation studies we assess the sensitivity of this method to taxon sampling. Further, we provide an R script to find the time point with the minimal temporal error with this proposed method.

Temporal inconsistency at the same taxonomic ranks

The chronogram used for the demonstration of this method was reconstructed from a multi-locus alignment, containing six markers (ITS, mtSSU, nuLSU, *RPB-1*, *Tsr1*, *Mcm7*) from 340 taxa representing 81 currently accepted genera^{23, 24}. Due to a number of recent molecular phylogenetic studies on Parmeliaceae, the majority of the taxa are reciprocally monophyletic in this phylogeny²⁵. Fifty-two additional taxa from ten related families²⁶ were also included to determine the relationships among the families and to study the temporal inconsistency among the taxa of the same taxonomic ranks, particularly at the genus and family levels.

We used the function `getMRCA` in the R-package “ape” in order to find the crown age of currently accepted taxonomic groups and found that while the majority of genera and families in this cladogram are about the same age around the median (22 MY for genus and 83 MY for family), a fair number of taxa are far older or younger than the majority of clades (Fig. 1). As has been previously shown in different animal groups^{13, 20}, the ranges of taxon ages overlapped between the genus and family ranks, making some of these taxonomic units not comparable even within the same family or order.

Temporal Error Scores

In an attempt to achieve temporally consistent classification, Holt and Jönsson proposed the Temporal Error Scores as a way to assess the amount of deviation from a single temporal cut-off²⁰. The method requires that the tree has been dated (chronogram), and that each tip is assigned to a particular taxonomic group at each rank. This approach first finds the “empirical error” by calculating the difference between the cut-off point and the crown age of each currently accepted taxonomic group and summing it up as an error score. Then, since the empirical error is readily comparable to other scores, the “standardized score” is also calculated by dividing the empirical error with the score computed from random expectations, which are generated from splitting the tree randomly into the same number of monophyletic groups.

While this metric is useful for quantifying the error from the current classification, two issues make it difficult to apply this method with other systems. First, the temporal cut-off points were arbitrarily chosen to generate the roughly same number of monophyletic groups as the current classification, which may not serve as a good criterion, because the current taxonomic groups are also somewhat arbitrary, constantly subjected to either

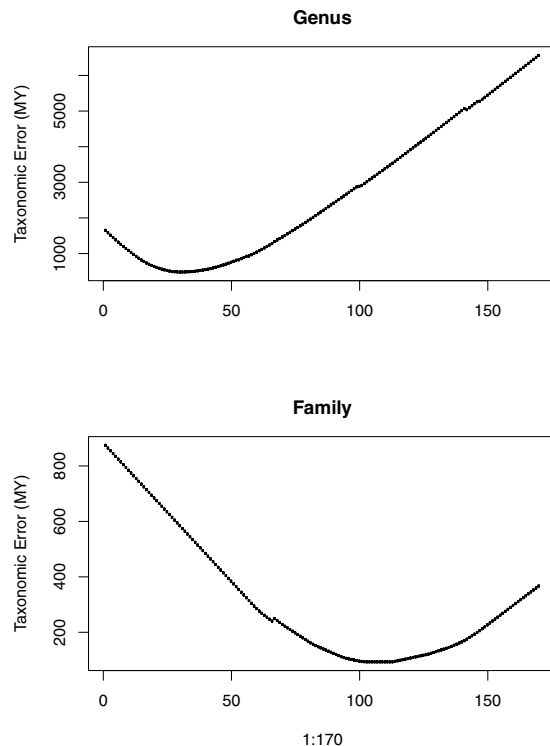


Figure 2. Distribution of empirical error score estimated from the tip (time = 0) to the root (the tree depth) for the genus level (A) and family level (B).

over-lumping or over-splitting. Second, the approach to generate random expectation can be computationally intensive and not readily reproducible. To address these issues, we extend the empirical error score to objectively select the cut-off point without the need to generate the random expectations.

Minimal Temporal Error Score

Using Holt & Jönsson's code for taxonomic errors²⁰, we applied a series of temporal thresholds from the tip (time = 0) to the root (the tree depth) and calculated an empirical error score at every 1 MY at both the genus and family levels. The temporal threshold with the minimum error score was then used as a cut-off point for the new classification with the temporal approach, in which we recognized all of the monophyletic groups that were more recent than the cut-off as individual taxa at that taxonomic rank. In cases where several time points produced the same error score, the average of those time points was used as a cut-off point. Because of inherently large confidence intervals at each node of the chronogram, we also calculated error margins around the cut-off point to allow some flexibility for reclassification, arbitrarily set to 5%. All of the procedures were performed in the statistical programming R. We also provided the R script to find the time point with minimal error, plot temporal banding on the chronogram, and reclassify based on the new time point on a data depository: <http://dx.doi.org/10.5061/dryad.p8n72>.

For our sampled phylogeny, the lowest empirical errors were found at 31 (29.45–32.55) and 107.5 (102.125–112.875) MY for the genus and family, respectively (Fig. 2). We then used the “cutree” function to reclassify the taxonomic groups, based on these time points. With the new cut-off point, 8 families remained intact with 2 families being split. For genera, 42 genera remain the same with 32 genera being lumped and 7 being split (Fig. 3; Supplementary materials). Details of the taxonomic changes are discussed in a companion paper in a mycological journal (manuscript in revision). Instead of using an arbitrary cut-off to maintain the number of taxonomic groups, here we provide a method to maintain the status of most taxonomic groups, while reclassifying the others, using the same age range within the same taxonomic rank.

Sensitivity to taxon sampling

Taxon sampling poses a challenge to virtually all molecular phylogenetic studies, as it is not always possible to acquire data for every single taxon in a lineage. Since the topology and branch lengths of chronograms have been shown to be sensitive to the amount of taxa being included^{27,28}, temporal methods are likely to be affected by taxon sampling, as they rely heavily on branch lengths. In order to determine the sensitivity of this method to the amount of included taxa, we simulated reduced taxon sampling in our sample tree by randomly removing 10% to 50% of the tips from the tree with the function “drop.tip.” Then, we used our proposed method to find the cut-off point with the minimum lowest empirical error for each of the 500 simulations.

The simulation results showed that the cut-off time point changed and got younger as more taxa were removed from the tree (Fig. 4). However, for the family level, the one standard deviation around the average from the

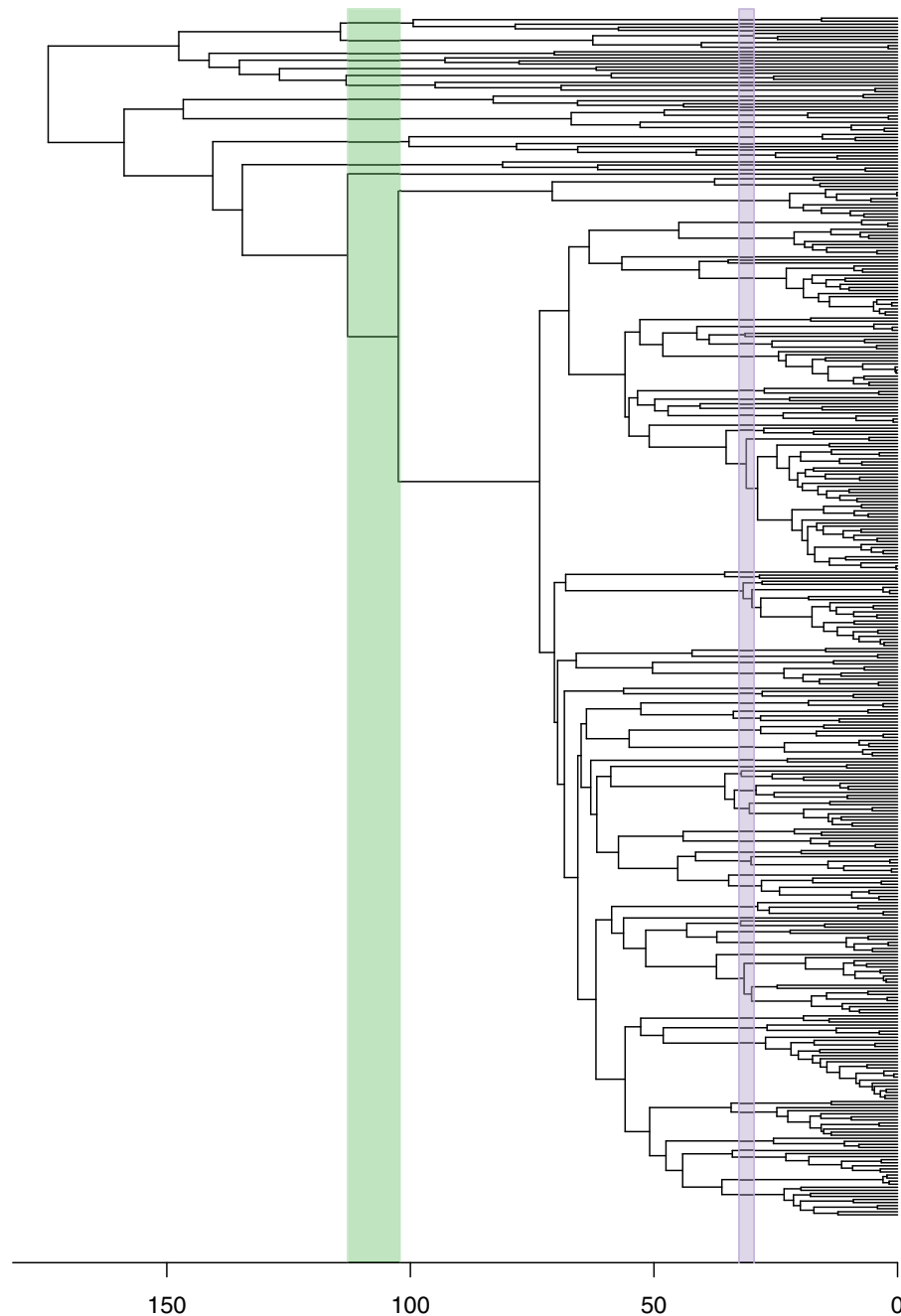


Figure 3. Time-calibrated phylogeny of Parmeliaceae and related families based on a multi-locus data set. Temporal bands for family rank (green) and genus level (pink) indicated.

simulations still fell within the 5% error margin of the selected time point in the current data. For the genus level, the cut-off points from the simulations were clearly younger than the most comprehensive dataset currently available, even when considering the standard deviation. These results suggest that taxon sampling is critical for the application of this method for temporal classification, but more so at the lower taxonomic ranks, e.g. genus, whereas the higher taxonomic ranks appear less impacted by lower sampling efforts.

Comparison with the Other Temporal Banding Method

Recently, Jönsson *et al.*²⁹ proposed a temporal banding approach that aims to minimize the disruption to the current taxonomy. They developed the metric called “Percent Consistency” which calculates the percentage of the taxa remaining intact after applying a certain cutoff time. This method then chooses the best cutoff time that yields the highest percent consistency. We ran our data through their public available R code and compared the resulting classification and the cutoff time at the genus and family level. The results showed that the cutoff time

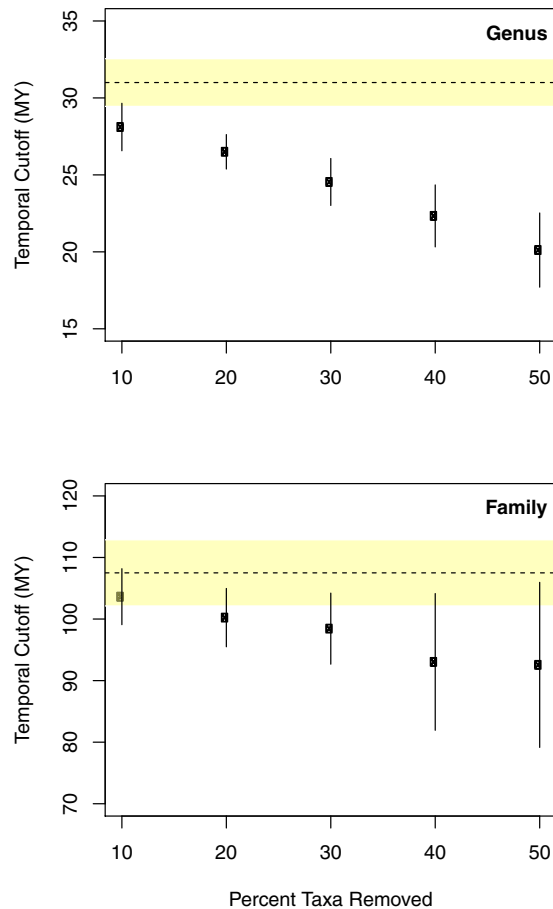


Figure 4. Results of simulations to evaluate sensitivity of temporal banding to taxon sampling. Temporal cut-off indicated as dots (with standard deviation) in relation to amount of removal of taxa from the data set. (A) Temporal cut-off at the genus level. (B) Temporal cut-off at the family level.

	Intact	Lumped	Split	Cutoff Time (MY)
Genus level				
Proposed method	55	32	15	31 (29.45–32.55)
Jönsson <i>et al.</i>	71	25	5	28.01
Family level				
Proposed method	11	0	1	107.5 (102.125–112.875)
Jönsson <i>et al.</i>	12	0	0	107.71

Table 1. Comparison of resulting new classification from the proposed temporal banding method and the method by Jönsson *et al.*²⁹.

for the genus level from the Jönsson *et al.* method (28 MY) was close to the lower end of the five-percent band from our method (29.45–32.55 MY), whereas the cutoff times for the family level were nearly identical (107.71 vs. 107.5 MY; Table 1). Our methods generally resulted in more changes in the assignment of genera and families than the Jönsson *et al.*'s method (Table 1). The discrepancy is due to our different metrics for choosing the optimal cutoff point.

The Jönsson *et al.* method²⁹ was specifically developed to maintain the current taxonomy, whereas our method relies on the calculation of temporal error scores and aimed more toward having comparable and temporally consistent taxonomic ranks. For many widely studied groups, such as mammals and birds, a temporal banding approach that minimize disruption to current taxonomy might be preferred, because any changes can affect many subsequent uses of the taxonomy. However, in other poorly studied groups, such as bryophytes and fungi, the existing assignment of supraspecific ranks are somewhat arbitrary and subjected to constant changes. In these groups, maintaining the current taxonomy is not a priority. Our method offers an alternative metric to find the temporal cutoff that does not solely focus on maintaining the current supraspecific taxonomic ranks.

Caveats

The effectiveness of the proposed temporal classification depends on the quality of chronogram reconstruction, which in turns relies on the amount and type of data, alignment methods and taxon sampling. As the minimum empirical error is derived for a particular tree, the absolute time point should not be used across different groups of organisms. For example, the 31-MY cut-off for the genus level cannot readily be applied to any other organismal groups, because it was calculated from the crown ages of the focused groups only.

Monotypic taxa are common across various supraspecific taxonomic ranks and pose a challenge as how to accurately determine the crown age of the taxa with only one tip. The calculation of the temporal error score that we implement here follows the algorithm by Holt and Jönsson²⁰, who explicitly state that this method of calculation does not use the crown age of the group and therefore is able to include monotypic taxa in the analysis. However, similar to their work, we did not include monotypic taxa for the crown age distribution analysis.

With different groups of organisms having different evolutionary histories and timelines, trying to find one universal cut-off for each taxonomic rank might not be productive. However, for groups of related organisms, the application of this method on a credible chronogram should allow to objectively find the temporal cut-off for classification of the same rank, while preserving taxonomic status of the majority of the taxa. The method provides an additional tool for erecting new taxa at supraspecific levels in an objective framework, adding to ongoing and growing discussion about temporal banding approaches to taxonomy.

References

- Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B. & Worm, B. How Many Species Are There on Earth and in the Ocean? *PLoS Biol* **9**, e1001127, doi:10.1371/journal.pbio.1001127 (2011).
- Blackwell, M. THE FUNGI: 1, 2, 3... 5.1 MILLION SPECIES? *American Journal of Botany* **98**, 426–438, doi:10.3732/ajb.1000298 (2011).
- May, R. M. Tropical arthropod species, more or less? *Science* **329**, 41–42, doi:10.1126/science.1191058 (2010).
- Mayr, E. *Systematics and the origin of species*. (Columbia University Press, 1942).
- Mayr, E. *Animal species and evolution*. (Harvard University Press, 1963).
- Linnaeus, C. *Species plantarum, exhibentes plantas rite cognitatas ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas*. Vol. I (II) (\Holmiae, 1753).
- Linnaeus, C. *Systema Naturae. Edit. XII*. (Stockholm, 1767).
- Hennig, W. *Phylogenetic systematics*. (University of Illinois Press, 1966).
- Haeckel, E. *Generelle Morphologie der Organismen*. (Georg Reimer, 1866).
- Johns, G. C. & Avise, J. C. A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. *Molecular Biology and Evolution* **15**, 1481–1490, doi:10.1093/oxfordjournals.molbev.a025875 (1998).
- Avise, J. C. & Johns, G. C. Proposal for a standardized temporal scheme of biological classification for extant species. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 7358–7363, doi:10.1073/pnas.96.13.7358 (1999).
- Avise, J. C. & Mitchell, D. Time to standardize taxonomies. *Systematic Biology* **56**, 130–133, doi:10.1080/10635150601145365 (2007).
- Avise, J. C. & Liu, J. X. On the temporal inconsistencies of Linnean taxonomic ranks. *Biological Journal of the Linnean Society* **102**, 707–714, doi:10.1111/j.1095-8312.2011.01624.x (2011).
- Castresana, J. Cytochrome b phylogeny and the taxonomy of great apes and mammals. *Molecular Biology and Evolution* **18**, 465–471, doi:10.1093/oxfordjournals.molbev.a003825 (2001).
- Bisby, F. A. & Hawksworth, D. L. What must be done to save systematics? *Regnum Vegetabile* **123**, 323–336 (1991).
- Michener, C. D. The possible use of unimonomial nomenclature to increase the stability of names in biology. *Systematic Zoology* **13**, 182–190, doi:10.2307/2411777 (1964).
- Cain, A. J. The post-Linnaean development of taxonomy. *Proceedings of the Linnean Society, London* **170**, 234–244, doi:10.1111/j.1095-8312.1959.tb00857.x (1959).
- Lumbsch, H. T. How objective are genera in euascomycetes? *Perspectives in Plant Ecology Evolution and Systematics* **5**, 91–101, doi:10.1078/1433-8319-00025 (2002).
- Nimis, P. L. A critical appraisal of modern generic concepts in lichenology. *Lichenologist* **30**, 427–438, doi:10.1017/S0024282992000422 (1998).
- Holt, B. G. & Jönsson, K. A. Reconciling Hierarchical Taxonomy with Molecular Phylogenies. *Systematic Biology* **63**, 1010–1017, doi:10.1093/sysbio/syu061 (2014).
- Zhao, R.-L. *et al.* Towards standardizing taxonomic ranks using divergence times – a case study for reconstruction of the *Agaricus* taxonomic system. *Fungal Diversity* **78**, 239–292, doi:10.1007/s13225-016-0357-x (2016).
- Talavera, G., Lukhtanov, V. A., Pierce, N. E. & Vila, R. Establishing criteria for higher-level classification using molecular data: the systematics of *Polymommatus* blue butterflies (Lepidoptera, Lycaenidae). *Cladistics* **29**, 166–192, doi:10.1111/j.1096-0031.2012.00421.x (2013).
- Divakar, P. K. *et al.* Evolution of complex symbiotic relationships in a morphologically derived family of lichen-forming fungi. *New Phytologist* **208**, 1217–1226, doi:10.1111/nph.13553 (2015).
- Singh, G. *et al.* Coalescent-Based Species Delimitation Approach Uncovers High Cryptic Diversity in the Cosmopolitan Lichen-Forming Fungal Genus *Protopermatelia* (Lecanorales, Ascomycota). *Plos One* **10**(5), e0124625, doi:10.1371/journal.pone.0124625 (2015).
- Crespo, A. *et al.* Phylogenetic generic classification of parmelioid lichens (Parmeliaceae, Ascomycota) based on molecular, morphological and chemical evidence. *Taxon* **59**, 1735–1753 (2010).
- Miadlikowska, J. *et al.* Multigene phylogenetic synthesis for 1307 fungi representing 1139 infrageneric taxa, 312 genera and 66 families of the class Lecanoromycetes (Ascomycota). *Molecular Phylogenetics and Evolution* **79**, 132–168, doi:10.1016/j.ympev.2014.04.003 (2014).
- Poux, C., Madsen, O., Glos, J., de Jong, W. W. & Vences, M. Molecular phylogeny and divergence times of Malagasy tenrecs: influence of data partitioning and taxon sampling on dating analyses. *BMC Evolutionary Biology* **8**, 102, doi:10.1186/1471-2148-8-102 (2008).
- Townsend, J. P. & Lopez-Giraldez, F. Optimal Selection of Gene and Ingroup Taxon Sampling for Resolving Phylogenetic Relationships. *Systematic Biology* **59**, 446–457, doi:10.1093/sysbio/syq025 (2010).
- Jönsson, K. A. *et al.* A supermatrix phylogeny of corvid passerine birds (Aves: Corvides). *Molecular Phylogenetics and Evolution* **94**, 87–94. Available at: doi:10.1016/j.ympev.2015.08.020 (2016).

Acknowledgements

The project was financially supported by the Spanish Ministerio de Ciencia e Innovación (projects CGL2013-42498-P), and the Negaunee Foundation (“The greatest radiation in the fungal kingdom”). E.K. was supported by a visiting scholarship from The Field Museum and funding from the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission.

Author Contributions

E.K., H.T.L. designed the research and wrote the paper. P.K.D., A.C., S.D.L. assembled molecular data and conducted phylogenetic analyses. E.K. produced the code and associated figure.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-02477-7](https://doi.org/10.1038/s41598-017-02477-7)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017