Supporting Information File 2

Supplementary Materials and Methods

The grass was greener: Repeated evolution of specialized morphologies and habitat shifts in ghost spiders following grassland expansion in South America

F.S. Ceccarelli, N. Mongiardino Koch, E.M. Soto, M.L. Barone, M.A. Arnedo, M.J. Ramírez

DNA extraction, Polymerase Chain Reactions and DNA sequencing

For the newly-generated sequences for this study, DNA was extracted from leg muscle tissue using the Qiagen DNeasy Blood and Tissue Kit, digesting the tissue at 56º C over-night with Proteinase K and following the manufacturer's protocol. For the Polymerase Chain Reactions (PCR), a mix was prepared including 1.5µl x10 PCR Buffer (Thermo Scientific), 10 µmoles $MgCl_2$, 0.25 µmoles of each dNTP, 0.4 µmoles of each primer, 0.1 µl Taq Polymerase (Thermo Scientific), 0.5 µl BSA, 1-2 µl genomic DNA and ddH_2O to bring the final volume to 15 µl. The primers used for amplification can be found in Table S2 of SI File 1. The thermal cycling profiles comprised an initial denaturing step at 95º C for 3 minutes, followed by 15 cycles of 30 seconds at 95º C, 30 seconds at the annealing temperature (51º C for nuclear and 45º C for mitochondrial gene fragments) and 45 minutes at 72º C; an additional 20 cycles were run with the annealing temperature lowered by 3º C and a final extension step of 10 minutes at 72º C was executed. PCR products were purified using ExosAP (Thermo Scientific) following the manufacturer's indications and sent for sequencing to Macrogen Inc., Korea. Sequences were edited based on the chromatograms in Sequencher v. 4.1.4 (GeneCodes corp.), where each protein-coding gene fragment was checked for stop-codons (indicating possible pseudogenes).

Phylogenetic analyses and node age estimations

Newly-edited sequences were combined with those from previous studies in an alignment per marker. The online version of MAFFT v. 7 (Katoh and Standley 2013) with the "Auto" strategy and a gap opening penalty of 1.53 was used to align the Histone 3-a (henceforth H3a) gene fragment, which was straightforward. The Cytochrome Oxidase C subunit 1 (henceforth COI) gene fragment was aligned based on its translation to protein in TranslatorX (Abascal et al. 2010) and for the 16S rDNA (henceforth 16S) and 28S rDNA (henceforth 28S) gene fragments the online version of T-coffee was used (Notredame et al. 2000; Di Tommaso et al.

2011) using the default settings. The Gblocks server (Castresana 2000) was used to remove poorly aligned positions for the 16S matrix, setting the less stringent options "allow for gap positions" and "less strict flanking positions" in the final alignment.

The chi-squared metric provided in the program TreePuzzle (Schmidt et al. 2002) was used to evaluate the molecular marker's utility for phylogenetic reconstruction, taking into consideration nucleotide compositional homogeneity and its importance for topological inference. Additionally, data partitioning strategies and nucleotide substitution models were selected using PartitionFinder v. 1.1.1. (Lanfear 2012) (see Table S3 of SI File 1 for details).

Separate phylogenetic trees were obtained for each molecular marker, to evaluate gene tree discordances, after which the matrices for the gene fragments were concatenated. Phylogenetic trees were obtained by Bayesian Inference (BI) and Maximum Likelihood (ML), the latter to corroborate results from the BI analyses. The ML tree was obtained by running the rapid hill climbing algorithm and setting the GTRGAMMA substitution model on a partitioned dataset in RAxML v. 8 (Stamatakis 2014). Nodal support was evaluated based on 1000 bootstrap replicates. For the BI phylogram, MrBayes v. 3.2.3 (Ronquist et al. 2012) was used, applying a Markov Chain Monte Carlo (MCMC) simulation of 20 million generations, sampling a tree every 2,000th generation for four chains in two independent runs. The data was partitioned and separate nucleotide substitution models were set as priors for the partitions based on the scheme and models selected by PartitionFinder, and parameters were unlinked across partitions. Consensus trees were obtained for each analysis using the "sumt" command and discarding the first 25% as burn-in, after checking the stationarity and correct mixing of the MCMC runs.

The program BEAST v. 1.8.2 (Drummond et al. 2012) was used to build a chronogram of the amaurobioidine taxa in this study by BI, estimating divergence times with fossil calibrations. In the absence of known fossils belonging to the focal subfamily, information from two fossil anyphaenines (outgroup taxa in this study) was used to apply

node age priors to the analysis. A fossil specimen postulated to belong to the genus *Anyphaena*, found in Baltic amber (Petrunkevitch 1946) and estimated to be between 33.9 and 37.2 Myr old, was used to set a uniform prior with a minimum age of 33.9 Myr on the constrained mrca of anyphaenines in this study. Since the assignment of this immature specimen to *Anyphaena* is not certain (Penney 2000), and given the fact that the sampling of extant species belonging to this genus is not extensive in this study, a more conservative calibration point was chosen. Information from a second fossil specimen belonging to the genus *Anyphaenoides* found in Dominican amber (Penney 2000) was used to set a uniform prior with a minimum age of 13.65 Myr to the mrca of *Anyphaenoides* and its sister group (given the fact that only one species of *Anyphaenoides* was available for our study). Further settings and parameters for the tree inference included partitioning the data by marker, unlinking the substitution and clock model priors for each partition and setting the most appropriate substitution model (as determined by PartitionFinder) to each partition. For the tree prior, a Birth-Death process was chosen, and an uncorrelated lognormal relaxed clock prior was set for each partition and the rates estimated based on a lognormal distribution around the mean (0), with an initial value and standard deviation of 1, providing a permissive range for the program to auto-optimize towards the true posterior values.

Two independent runs of 100 million generations each (sampling every 10,000th tree) were executed and the outcomes of the two runs were validated in Tracer v. 1.5 (Rambaut and Drummond 2007) to check that the effective sample sizes of the parameters were >200. The tree files were then combined using LogCombiner 1.8.2 (part of the BEAST package) and the maximum clade credibility (mcc) tree with mean node heights selected in TreeAnnotator 1.8.2, setting the burn-in at 10%. Nodal support was assessed based on posterior probability, and the plausibility of the node age estimates evaluated by comparing the ages to known geologic events (the appearance of the volcanic Juan Fernández Archipelago for endemic *Sanogasta* and *Philisca* species, see Soto and Ramírez 2012 and Soto et al. 2017) and to ages

estimated for particular clades in previous studies (Ceccarelli et al. 2016; Soto et al. 2017). Furthermore, to corroborate the node age estimates obtained by fossil-based calibration, another chronogram was built in BEAST using the same settings and parameters, but instead of constraining nodes with age priors, a prior was set for the COI substitution rate, based on rates from previous studies. More specifically, a normal prior with a mean of 0.0145 and standard deviation 0.0025 was set, to get 95% of the distribution to fall within the lowest and highest values obtained for spider COI substitution rates in previous studies (Arnedo and Gillespie 2006; Bidegaray-Batista and Arnedo 2011; Zhang and Li 2013; Ceccarelli et al. 2016; Soto et al. 2017). To visualize the topological uncertainties of the posterior BEAST trees, the program DensiTree v. 2.2.5 (Bouckaert and Heled 2014) was used. All phylogenetic analyses were executed through the CIPRES Science Gateway V. 3.1 (Miller et al. 2010).

References

Abascal, F., R. Zardoya, and M. J. Telford. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38:W7–W13.

Arnedo, M. A., and R. G. Gillespie. 2006. Species diversification patterns in the Polynesian jumping spider genus *Havaika* Prószyjski, 2001 (Araneae, Salticidae). Mol Phylogenet Evol. 41:472–495.

Bidegaray-Batista, L., and M. A. Arnedo. 2011. Gone with the plate: the opening of the Western Mediterranean basin drove the diversification of ground-dweller spiders. BMC Evol Biol. 11:317–331.

Bouckaert, R. R., and J. Heled. 2014. DensiTree 2: Seeing Trees Through the Forest. bioRxiv, http://dx.doi.org/10.1101/012401.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17:540–552.

Ceccarelli, F. S., B. D. Opell, C. R. Haddad, R. J. Raven, E. M. Soto, and M. J. Ramírez. 2016. Around the World in Eight Million Years: Historical Biogeography and Evolution of the Spray Zone Spider Amaurobioides (Araneae: Anyphaenidae). PloS one 11(10):e0163740.

Di Tommaso P., S. Moretti, I. Xenarios, M. Orobitg, A. Montanyola, J. M. Chang, J. F. Taly, and C. Notredame. 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. 39:W13-17.

Dietrich, C. H. 1999. The role of grasslands in the diversification of leafhoppers (Homoptera: Cicadellidae): a phylogenetic perspective. In Proceedings of the Fifteenth North American Prairie Conference (pp. 44-49). Bend, OR: Nat. Areas Assoc.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969–1973.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29:1695–1701.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. pp. 1–8 in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA.

Notredame, C., D. G. Higgins, and J. Heringa. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302:205–17.

Penney, D. 2000. Anyphaenidae in Miocene Dominican Republic amber (Arachnida, Araneae). J. Arachnol. 28:223–226.

Petrunkevitch, A. 1946. Fossil Spiders in the Collection of the American Museum of Natural History. Am. Mus. Novitates 1328:1–36.

Rambaut, A., and A. J. Drummond. 2007. Tracer v1.5. University of Edinburgh, Edinburgh, U.K. <http://beast.bio.ed.ac.uk/Tracer>.

Ronquist, F., M. Teslenko, F. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, , L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst. Biol. 61:539–542.

Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. Tree-Puzzle: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504.

Soto, E. M. and M. J. Ramírez. 2012. Revision and phylogenetic analysis of the spider genus Philisca Simon (Araneae: Anyphaenidae, Amaurobioidinae). Zootaxa 3443:1–65.

Soto, E. M., F. M. Labarque, F. S. Ceccarelli, M. A. Arnedo, J. Pizarro-Araya, and M. J. Ramírez. (2017). The life and adventures of an eight-legged castaway: Colonization and diversification of Philisca ghost spiders on Robinson Crusoe Island (Araneae, Anyphaenidae). Mol. Phylogenet. Evol. 107:132–141.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.

Zhang, Y, and S. Li. 2013. Ancient lineage, young troglobites: recent colonization of caves by Nesticella spiders. BMC Evol Biol. 13:183–192.