Electronic Supplementary Material

Taxon sampling

The phylogeny includes representatives of every higher taxonomic group to have a good representation of the higher diversity of brittle-stars. All families are represented in the phylogeny and have been systematically sampled, but deep-sea species are slightly underrepresented as species records have been gathered from specimens deposited in several scientific collections, and in which shallow-water samples are more abundant. Pie charts showing the distribution of bathomes across species for each family are given below (shelf-break = 200m, Figure S1).

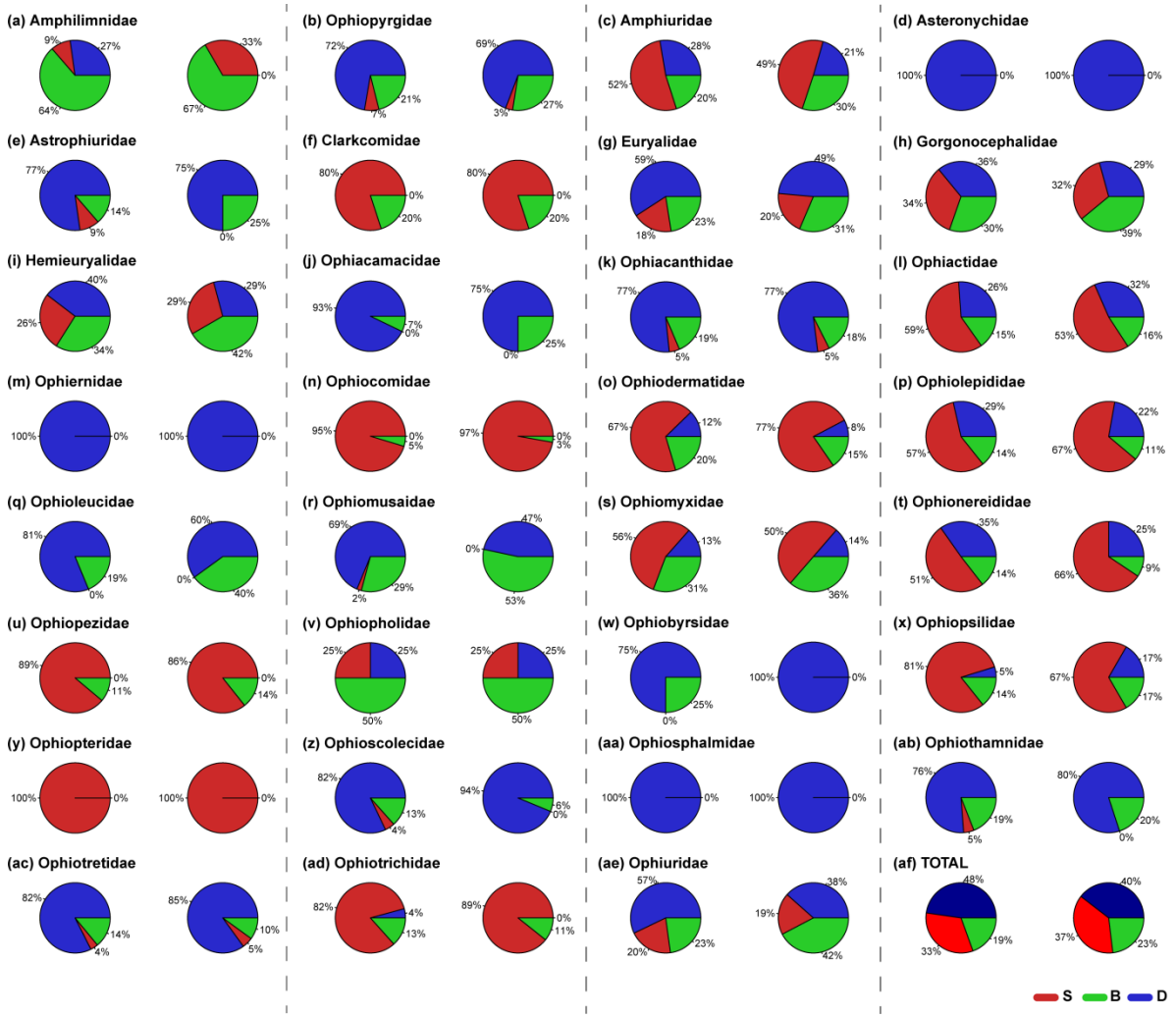

Figure S1. Percentage of species distributed in each bathome grouped by family for the global dataset (left) and for the phylogeny (right).

Phylogenetic structure of bathomes

As bathomes were coded as a discrete character (S, B and D), using maximum parsimony (MP) we expected to find few changes if bathomes were highly conserved through the phylogeny. On the contrary, if there was no conservatism, we would expect to find the trait distributed randomly through the phylogeny, which would be reflected in a higher number of bathome shifts. Niche conservatism is not just a marine phenomenon, for example land plants more frequently colonise similar biomes even after dispersing long distances [1].

We calculated the number of steps required for the parsimony reconstruction of bathomes on the phylogeny using Mesquite 3.01 [2] and coding depth as an ordered character (no direct transitions between S and D were allowed in any direction). We generated two groups of 10,000 datasets each by: (1) randomly choosing the depth distribution of 688 species from the global dataset for each new dataset (ALLrandom), and (2) reshuffling the tip states from the phylogeny for each new dataset (TREEreshuffle). We also calculated the number of parsimony steps for the observed data. Our null hypothesis of bathomes being randomly distributed for the phylogeny was rejected as the observed number of transitions was outside the 95% CI of both randomised and reshuffled datasets.

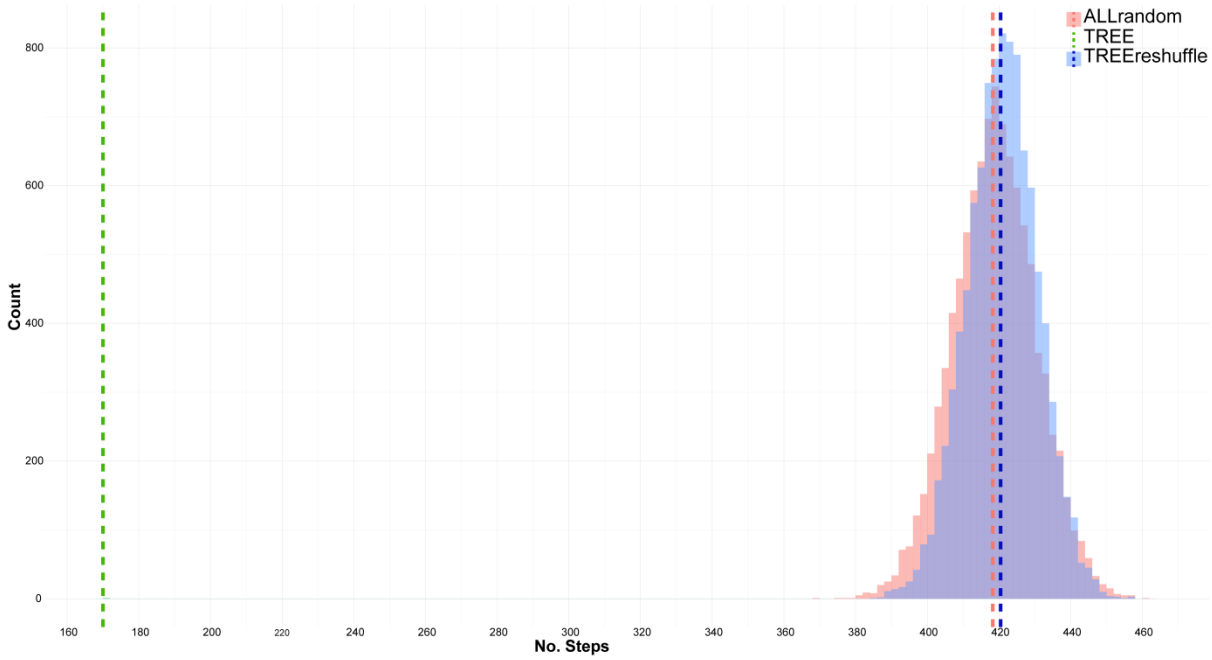

Figure S2. Number of parsimony steps for the observed data (green, n = 170) and for the 10,000 datasets randomly taken from the global dataset (pink, $mean_{ALL}$ = 418.3) and reshuffling the tip states (blue, $mean_{TREE}$ = 420.4). Mean values are shown in dotted lines.

Asymmetry in transition rates

The asymmetry in transition rates was explored using eight different models were depth was coded as a continuous character, either ordered or unordered. Analyses were run using the *fitDiscrete* function in the R [3] package 'Geiger'[4] and setting to 'TRUE' the Hessian element control in order to estimate 95% CI for each parameter (Table S1). For the ordered model ARD6, where shifts of bathome in different directions were estimated to be different and direct transitions between S and D were allowed, CI values were not estimated.

Model	lnL	No. rates	AIC	ΔAIC	Rates [95% CI]					
					SB	DB	BS	BD	SD	DS
ER	-512.41	1	1026.8	25.9	0.008 [0.007, 0.01]	0.008 [0.007, 0.01]	0.008 [0.007, 0.01]	0.008 [0.007, 0.01]	0 [0, 0]	0 [0, 0]
SYM	-506.471	2	1016.9	16	0.005 [0.004, 0.007]	0.012 [0.009, 0.017]	0.005 [0.004, 0.007]	0.012 [0.009, 0.017]	0 [0, 0]	0 [0, 0]
EC	-504.531	2	1013.1	12.2	0.007 [0.005, 0.009]	0.007 [0.005, 0.009]	0.012 [0.009, 0.015]	0.012 [0.009, 0.015]	0 [0, 0]	0 [0, 0]
E2C	-500.78	3	1007.6	6.7	0.004 [0.002, 0.007]	0.008 [0.007, 0.011]	0.011 [0.009, 0.014]	0.011 [0.009, 0.014]	0 [0, 0]	0 [0, 0]
EC2	-501.967	3	1009.9	9	0.007 [0.006, 0.009]	0.007 [0.006, 0.009]	0.007 [0.004, 0.013]	0.017 [0.012, 0.024]	0 [0, 0]	0 [0, 0]
ARD4	-496.437	4	1000.9	0	0.005 [0.003, 0.007]	0.012 [0.008, 0.017]	0.007 [0.004, 0.011]	0.022 [0.014, 0.034]	0 [0, 0]	0 [0, 0]
ARD6	-496.437	6	1004.9	4	0.005 NA	0.012 NA	0.007 NA	0.022 NA	4.41E-20 NA	2.31E-19 NA

Table S1. Fit of each of the eight models analysed under ML. Estimated rates for each direction of transition is shown, with 95% CI values.

Aditionally, we explored the effect of the cut-off value to delimit bathomes in the estimated transition rates. We generated four more datasets using 100, 150, 250 and 300 m as the shelf-break value, and analysed them using under the ARD model for comparison using *fitDiscrete*. The highest estimated rate of bathome shift in four datasets was B→D, same as with the 200m cut-off but there were differences in the other directions as shown in Figure S3.

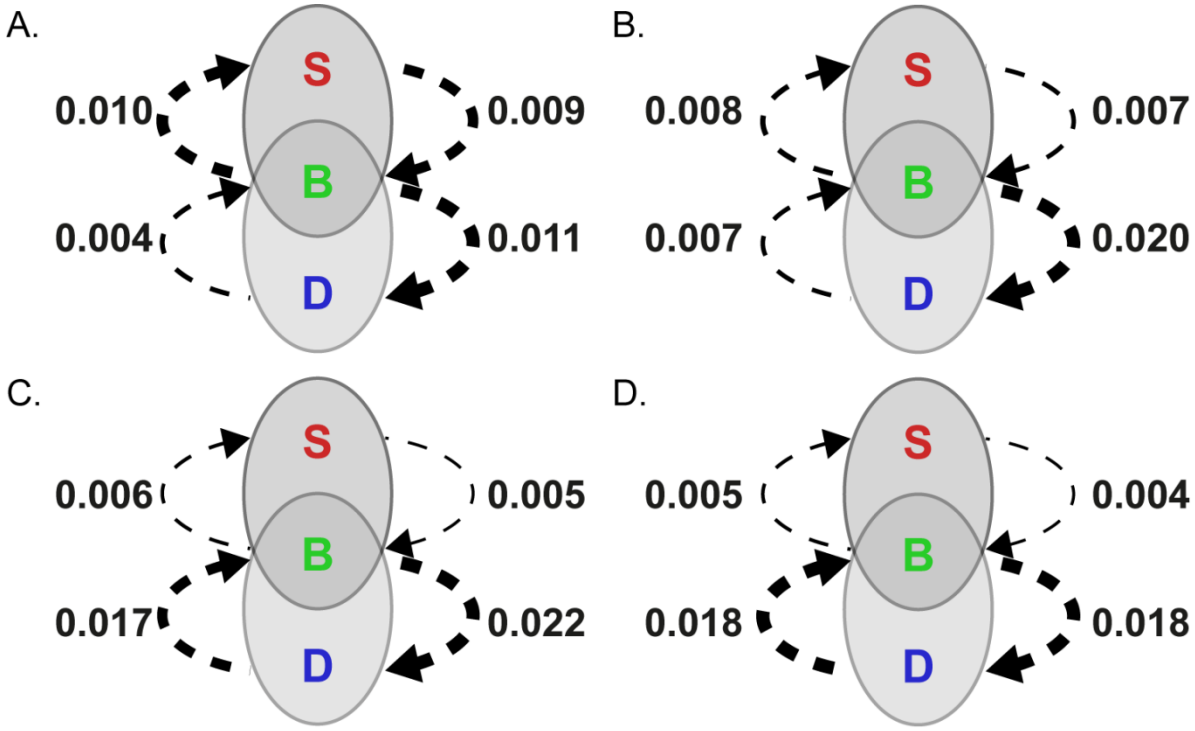

Figure S3. Estimated rates for bathome shifts for each pair, using different shelf-break values: (A.) 100 m, (B.) 150 m, (C.) 250 m and (D.) 300 m. Bathome shift is indicated with the arrows and the width is proportional to the shift rate value.

Incomplete taxon sampling

In order to broadly understand the effect of incomplete taxon in parameter estimation, we assessed the accuracy of the model when sample size decreases. A phylogeny with 2,120 extant taxa was simulated under a birth-death rate model ($\lambda = 1.0$ and $\mu = 0.5$, as reported for ophiuroids [5]), and an origination time of 270 Mya (function *sim.bd.taxa.age*, package 'TreeSim' [6]). The size of this phylogeny is the same as the total known species of ophiuroids considered for this study. Bathomes (S, B, D) for each tip from the large simulated phylogeny were simulated based on the Q matrix obtained for our dataset using *fitDiscrete* under the ARD (4 rates) model (function *sim.char* from R package 'Geiger' [4]). Smaller phylogenies were generated by randomly pruning 40 tips, until a phylogeny of only 40 taxa

was obtained. Each of them was analysed under the ARD4 model and the estimated parameters are presented in the plot below. Figure S4 shows that, even when tips are pruned randomly and affect the proportion of character states at the tips, the model estimated similar parameters until the sampling is less than 20% of the original data was considered.

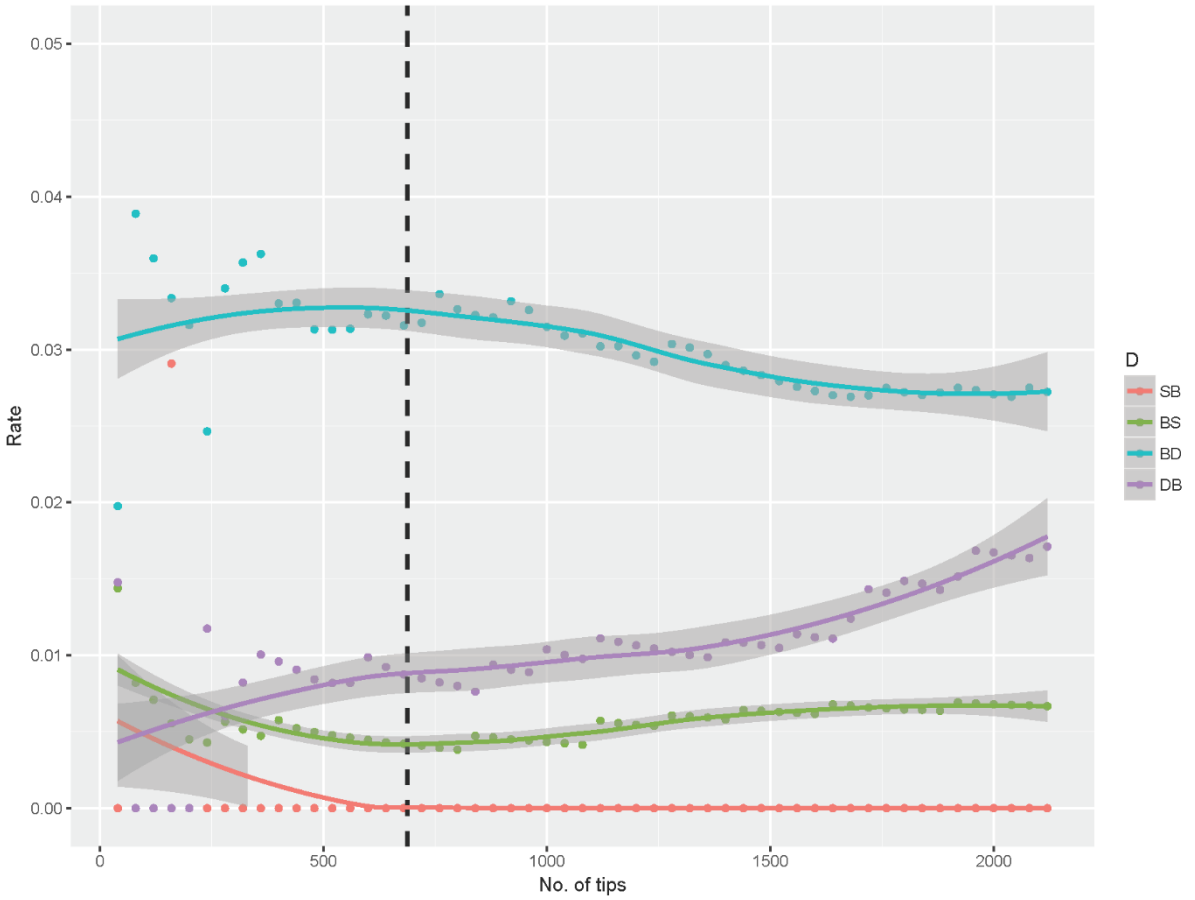

Figure S4. Plot shows that estimated parameters under ARD model are congruent with the observed parameters in the complete dataset, even when only 20% of the data is considered. Dashed line indicates 688 tips, which is the size of our phylogeny.

Heterotachy and ancestral reconstruction – Extended Methods

All analyses were performed on a single 688 taxa ultrametric tree with fixed topology and branch lengths. Heterotachy, the heterogeneity in rates of evolution of bathymetric ranges across different lineages, was investigated by implementing Bayesian inference models in

BEAST v 1.8.3 [7] using a ‘random local clock’ (RLC) [8]. Four different models, constraining shifts between bathomes and considering the depth stratum character as ordered (S←→B and B←→D) or unordered (also S←→D), were included using a ‘random local clock’ (RLC) [8] and a strict clock. The analyses were performed using a standard continuous-time Markov Chain (CTMC) with priors as in Ukuwela *et al* [9] and Lemey *et al* [10].

All eight models were run using the same parameters. Population size was assumed to remain constant through time. A MCMC wide exponential prior (mean=0.1) was used for the overall transition rate, and scale operators were used to allow rate heterogeneity in specific models. Two independent runs were performed for each model to ensure convergence. The length of each chain was 120 million steps, sampled every 12,000 and yielding a total of 10,000 samples. ESS (Effective Sample Size) was used to evaluate mixing behaviour of the chains. The MCMC, path sampling (PS) and stepping stone (SS) log files from the independent chains for each models were combined using LogCombiner [7], after removing burn-in (10%), and analysed in Tracer v1.6 [11]. Model selection was based on the log marginal likelihood for each model estimated by PS and SS method.

Estimation of ancestral bathomes was carried out simultaneously using a stochastic mapping based on a MCMC approach (markovJumpsTreeLikelihoood [12]), which maps the bathome shifts on each sampled tree and records the number of shifts for each pair (Table S2). The number of bathome shifts per pair were logged independently for each sampled tree. Only 9,000 trees from both runs (55% burn-in) were combined using LogCombiner [7] and used to estimate the state probability at nodes and the number of shift biomes from the stochastic mapping on the fixed tree in TreeAnnotator [7] (see Figure S5).

Model	Clock	Rates (transitions)						
		SB	SD	BS	BD	DS	DB	Overall
SYM	Strict	0.568 [3.61E-3, 1.406] *(52.1)*	0 n/a *(0)*	0.568 [3.61E-3, 1.41] *(48.7)*	1.250 [8.49E-3, 3.076] *(118.8719)*	0 n/a *(0)*	1.250 [8.49E-3, 3.076] *(128.2)*	0.012 [9.2863E-3, 0.0148] *(348.0135)*
	RLC	0.5691 [7.27E-3, 1.397] *(52.4)*	0 n/a *(0)*	0.569 [7.27E-3, 1.397] *(48.7)*	1.255 [0.024, 3.055] *(118.4)*	0 n/a *(0)*	1.255 [7.27E-3, 1.397] *(128.4)*	0.012 [9.2401E-3, 0.0148] *(347.861)*
ARD4	Strict	0.431 [0.062, 0.903] *(55.8)*	0 n/a *(0)*	0.613 [0.088, 1.33] *(46.7)*	1.781 [0.327, 3.631] *(142.0)*	0 n/a *(0)*	0.965 [0.143, 1.93] *(145.8)*	0.0149 [0.0109, 0.0195] *(390.1974)*
	RLC	0.439 [0.057, 0.924] *(56.1)*	0 n/a *(0)*	0.620 [0.086, 1.359] *(46.8)*	1.802 [0.287, 3.629] *(142.2)*	0 n/a *(0)*	0.975 [0.150, 1.986] *(146.0)*	0.0149 [0.0108, 0.0195] *(391.0889)*
ARD6	Strict	0.683 [0.174, 1.325] *(54.4)*	0.033 [4.59E-8, 0.108] *(1.7)*	0.951 [0.194, 1.900] *(44.6)*	2.777 [0.808, 5.163] *(136.0)*	0.033 [1.18E-6, 0.105] *(2.0)*	1.510 [0.482, 2.855] *(140.4)*	0.0144 [0.0106, 0.0187] *(379.2002)*
	RLC	0.684 [0.173, 1.334] *(54.7)*	0.034 [1.03E-6, 0.109] *(1.7)*	0.934 [0.206, 1.896] *(44.3)*	2.773 [0.751, 5.088] *(137.1)*	0.033 [2.38E-6, 0.106] *(2.1)*	1.507 [0.396, 2.759] *(140.9)*	0.0144 [0.0104, 0.0188] *(380.6135)*
ER	Strict	1 n/a *(45.9)*	1 n/a *(28.1)*	1 n/a *(23.7)*	1 n/a *(33.4)*	1 n/a *(30.8)*	1 n/a *(73.8)*	0.0076952 [6.229E-3, 9.2324E-3] *(235.5587)*
	RLC	1 n/a *(47.5)*	1 n/a *(30.154)*	1 n/a *(25.9)*	1 n/a *(36.3)*	1 n/a *(32.1)*	1 n/a *(74.2)*	0.0080421 [6.3773E-3, 9.6931E-3] *(246.1247)*

Table S2. Estimated transition rates, 95%HPD and mean number of inferred events for every model using a 'strict' and RLC evaluated under BI.

REFERENCES

1. Crisp M.D., Arroyo M.T., Cook L.G., Gandolfo M.A., Jordan G.J., McGlone M.S., Weston P.H., Westoby M., Wilf P., Linder H.P. 2009 Phylogenetic biome conservatism on a global scale. *Nature* **458**(7239), 754-756. (doi:10.1038/nature07764).
2. Maddison W.P., Maddison D.R. 2015 Mesquite: a modular system for evolutionary analysis. (3.04 ed.
3. Team R.C. 2016 R: A language and environment for statistical computing. R Foundation for Statistical Computing. (Vienna, Austria, R Foundation for Statistical Computing.
4. Harmon L.J., Weir J.T., Brock C.D., Glor R.E., Challenger W. 2008 GEIGER: investigating evolutionary radiations. *Bioinformatics* **24**(1), 129-131. (doi:10.1093/bioinformatics/btm538).
5. O'Hara T.D., Hugall A.F., Thuy B., Stohr S., Martynov A.V. 2017 Restructuring higher taxonomy using broad-scale phylogenomics: the living Ophiuroidea. *Molecular Phylogenetics and Evolution* **accepted**.
6. Stadler T. 2011 Simulating trees with a fixed number of extant species. *Systematic Biology* **60**.
7. Drummond A.J., Rambaut A. 2007 BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214. (doi:10.1186/1471-2148-7-214).
8. Drummond A.J., Suchard M.A. 2010 Bayesian random local clocks, or one rate to rule them all. *BMC Biol* **8**, 114. (doi:10.1186/1741-7007-8-114).
9. Ukuwela K.D.B., Lee M.S.Y., Rasmussen A.R., de Silva A., Mumpuni, Fry B.G., Ghezellou P., Rezaie-Atagholipour M., Sanders K.L. 2016 Evaluating the drivers of Indo-Pacific biodiversity: speciation and dispersal of sea snakes (Elapidae: Hydrophiinae). *Journal of Biogeography* **43**(2), 243-255. (doi:10.1111/jbi.12636).
10. Lemey P., Rambaut A., Drummond A.J., Suchard M.A. 2009 Bayesian phylogeography finds its roots. *PLoS Computational Biology* **5**(9), e1000520. (doi:10.1371/journal.pcbi.1000520).
11. Rambaut A., Suchard M.A., Xie D., Drummond A.J. 2014 Tracer. (1.6 ed.

12. King B., Lee M.S.Y. 2015 Ancestral State Reconstruction, Rate Heterogeneity, and the Evolution of Reptile Viviparity. *Systematic biology* **64**(3), 532-544. (doi:10.1093/sysbio/syv005).