

The computational time required for running the less sensitive to LBA site-heterogeneous models implemented in PhyloBayes_mpi on our phylogenomic datasets was extremely important. In particular, given the high combinatorial range of heterogeneity across sites, obtaining a full convergence of MCMC chains for our complete dataset (97 species x 401,632 positions) may be difficult, even impossible with our current computational resources. Given limited computational resources and the large datasets to be analysed, we chose a strategy with approximations akin to the Approximate Bayesian Computation approach (ABC, see Beaumont 2010, Sunnåker et al. 2013) through gene jackknifing.

Use of gene jackknifing to overcome limitations of the MCMC convergence of PhyloBayes_mpi with site-heterogeneous models

We introduced multiple large-scale perturbations to the dataset by applying jackknifing, both on genes (100 or 10 replicates) and species (4 different taxon samplings) starting every analysis from a random tree. It should be noted that our gene jackknifing strategy was particularly strict as each of the jackknife replicates was computed from a randomly chosen 25% sub-sample of the complete dataset. These large-scale perturbations are expected to be much more important than possible minor errors due to imperfect convergence of the MCMC chains, allowing to infer an accurate estimate of jackknife proportions (JPs) despite an approximation in the inference of the topology for individual replicates.

First, a careful observation of every individual jackknife replicate clearly showed that:

1) The lack of topological convergence almost exclusively affected the same few clades with short internal branch lengths and for which the contradicting topologies differ by one NNI change (e.g. *Mnemiopsis*+*Cestum* or the position of *Lucernariopsis*). These are the same set of clades that received low JP supports in our analyses (<75% in fig. 3) and it is therefore very likely that **possible lack of topological convergence did not lead to incorrect but highly supported relationships in our analyses.**

2) **The phylogenetic position of sponges is highly supported:** In our analysis with 100 jackknife replicates with 90 species (Fig. 3), 95% of the replicates yielded sponges as sister to all other metazoans, including 88% with maximal support (PP=1). Two replicates yielded a topology in which ctenophores and sponges are sister-groups and only three replicates placed ctenophores as sister to all other animals.

3) The chosen approximation in the convergence of the chains is likely to have no effect on our conclusions given that **we obtained a remarkable congruency between independent gene jackknifing replicates for 4 different taxon samplings** (97, 90, 80 and 70 species), with identical topologies obtained, except for the LBA of ctenophores and hexactinellids in the case of 70 species, and a statistically unsupported NNI change of the position of *Salpingoeca dolichothecata*.

Impact of possible convergence issues on the estimation of jackknife proportions

Second, we conducted several experiments to check that our approach provides accurate estimates of both the topology and the jackknife proportions.

1) We checked by eye that chains reach a plateau for all parameters, and for 10 randomly chosen chains we followed the evolution of the Effective Sampling Size (ESS) for more cycles (>15,000) than for all the others (6,000 cycles). For 7 chains, we obtained ESS > 50 for all parameters, while for the remaining 3 parameters (loglik and to a lesser extent statent and stataalpha) ESS remains below 50. Most importantly, in agreement with our long experience with convergence of chains in PhyloBayes, **the topology remains almost identical with much longer chains**. More precisely, we computed the consensus topologies for each chain by sampling trees between 3,000 and 6,000 cycles and between 3,000 and >15,000 cycles. For 7 chains, the topologies are identical, despite important differences in the SSE values, and for the three remaining chains, a single NNI change was observed (position of *Leuconia* (JP=91% in our Fig.3), of *Codosiga* (JP=45%) and of *Salpingoeca dolichothecata* (JP=35%)). This experiment suggests that **the convergence of the CAT chains towards the correct topology is rapid and does not require ESS > 50 for all parameters**.

2) For our 100 jackknife replicates, we separated the 50 chains with the worst SSE for the likelihood and the 50 ones with the best SSE, and computed independently the consensus. Obviously for 71 nodes with a JP=100, we observed no differences and for the remaining 16 nodes, the JP values are highly similar (see Table below), in the range of the fluctuation expected for a sample size of 50 (the fluctuations are, as expected, the highest when JP is close to 50% and are negligible when JP>90%). This demonstrates that **the topology of our figure 3 and its associate JP values are not affected by SSE values < 50**, because the topological convergence is already reached.

3) To further check that convergence in topology occurs well before the convergence of all the parameters, in particular the ones of the Dirichlet process, we analysed the chains before our burn-in of 3,000 cycles. More precisely, we computed consensus topologies by only using the sample from 1,000 to 2,000 cycles or from 2,000 to 3,000 cycles. Remarkably, for the 71 nodes with a JP of 100% in our fig. 3, we also obtained a JP of 100%, demonstrating that for highly supported nodes, convergence of the PhyloBayes chains is accurate and quite rapid. For the remaining 16 nodes, the JP values are strikingly similar to the ones of our figure 3, despite corresponding to independent samples: the greatest difference is 4% (91% versus 87% and 43% versus 39%, see Table below). This indicates that **chains that are not perfectly converged for several parameters** (in particular the ones related to the Dirichlet process) **are almost perfectly converged in terms of topology**.

Conclusions

In summary, the gene jackknife approach we proposed allowed an accurate estimate of the topology and of its statistical support, even if the convergence of some individual chains is not perfect. This is likely because: 1) the fine tuning of the Dirichlet process parameters has almost no impact on the topology, and 2) the possible lack of topological convergence of the chains only affects nodes that are poorly supported, so that this random error is drowned within the large variance generated by gene jackknifing.

Overall, these observations demonstrate that sponges being placed as sister to all other animals is the result of the robust congruence of an ambitious jackknife strategy that is not impacted neither by gene sampling, nor by species sampling, nor by a lack of chain convergence.

Table : JP values recovered for the 16 nodes that were not supported by 100% JP. JP values are showed as in Fig. 3 (3,000+), separating the 50 best and the 50 worst chains in term of loglik SSE values (3,000- best or worst SSE loglik), and prior to the burn-in used in our study (from 1,000 to 2,000 cycles, or from 2,000 to 3,000 cycles). This indicates that JP values of Fig. 3 depend neither on parameter convergence (as estimated with SSE values) nor on chain length (number of cycles).

3,000 – 6,000	3,000 - best SSE loglik	3,000 - worst SSE loglik	1,000 - 2,000	2,000 - 3,000
88	88	88	88	88
73	76	70	76	74
91	90	92	87	90
93	94	92	94	94
45	52	38	43	44
89	88	90	87	88
95	94	96	94	94
99	100	98	99	99
97	98	96	96	96
91	94	88	87	89
45	50	42	44	44
35	40	30	36	38
98	96	100	98	98
43	42	44	39	40
83	82	84	82	82

References :

Beaumont MA (2010) Approximate Bayesian Computation in Evolution and Ecology. Annual Review of Ecology, Evolution and Sytematics 41: 379-406. DOI: 10.1146/annurev-ecolsys-102209-144621

Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C (2013) Approximate Bayesian Computation. PLoS Comput Biol 9(1): e1002803. doi:10.1371/journal.pcbi.1002803