Additional note on the MCMC convergence approximation strategy

The computational time required by the (less sensitive to LBA) site-heterogeneous models for a phylogenetic analysis is extremely important. In particular, given the high combinatorial range of heterogeneity across sites, obtaining a full convergence of MCMC chains may be difficult. 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, Sunnâker *et al.* 2013):

1) we introduced multiple large-scale perturbations to the dataset by applying jackknifing, both on genes (100 or 10 replicates) and on species (4 different taxon samplings) and every analysis started with a different 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% of the whole data.

2) we ran a single chain on all these replicates until all the parameters reached a clear plateau and the size of the plateau was at least half of the burn-in. For ten jackknife replicates of the 90 species sample (Fig. 3), we ran the chains much longer in order to obtain an ESS > 50 for all parameters and observed that the topology was unchanged.

A careful observation of every individual jackknife replicate (as well as additional unpublished analyses of similar datasets) clearly shows that: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 JS% 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) tree topology converges much more rapidly than the Dirichlet process, so that the effective sample size (ESS) of some parameters (e.g. Nmode) can be relatively small (e.g. 20) while the topology already converged to its posterior distribution.

3) the phylogenetic position of sponges is highly supported. In our analysis with a 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 are sister to all other animals.

Overall, the chosen approximation in the convergence of the chains is likely to have no effects on our conclusions given the remarkable congruency we obtained between independent gene jackknifing replicates for 4 different taxon samplings (97, 90, 81 and 72 species), with identical topologies obtained, except for the LBA of ctenophores and hexactinellids in the case of 72 species, and an statistically unsupported NNI change of the position of *Salpingoeca dolichothecata*.

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

Sunnâker 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