CORRECTIONS & AMENDMENTS

Corrigendum: Resolving the evolutionary relationships of molluscs with phylogenomic tools

Stephen A. Smith, Nerida G. Wilson, Freya E. Goetz, Caitlin Feehery, Sónia C. S. Andrade, Greg W. Rouse, Gonzalo Giribet & Casey W. Dunn

In this Letter, we investigated the evolutionary relationships of molluscs with multigene data sets assembled from new transcriptome data and published genomes and transcriptomes. Since publishing these results, examination of our gene sequence matrix by others revealed that all instances of six amino acids (E, F, I, L, P and Q) were replaced by ambiguous characters in our super matrix. This led to the exclusion of data that should have been in the final analyses. The data exclusion was caused by incorrect handling of protein data at the final stage of matrix concatenation by the published program Phyutility (http://code.google.com/p/phyutility/). We have fixed the program, regenerated the final matrices, and re-run our analyses. There was minimal impact on our results, with no changes in the topology of the tree at deep nodes that had consistent strong support in our published analyses. There are minor variations in support values in the corrected analyses. The corrected matrices have been deposited at Dryad under the existing accession number (http://dx.doi.org/10.5061/dryad.24cb8). Figure 1 shows the corrected Fig. 2 of the original Letter, with corrected support values. In the text of the original Letter, the sentence "Bayesian analyses using the site-heterogeneous CAT model of protein evolution also place Scaphopoda as the sister group to Gastropoda, with a posterior probability of 89%" should read "Bayesian analyses using the site-heterogeneous CAT model of protein evolution also place Scaphopoda as the sister group to Bivalvia, with a posterior probability of 81%". In the Methods Summary, both instances of "27%" in the following phrase should be "41%": "27% character occupancy (that is, 27% of the matrix consists of unambiguous amino acid data, with the remainder being missing data or alignment gaps)" and "21%" should be "32%" in the sentence "This matrix has 40% gene occupancy, 21% character occupancy and is 216,402 sites long.". In the Methods, the following two sentences should be removed: "PhyloBayes misidentified the data type of our matrix as DNA, resulting in model misspecification and lack of convergence. We conducted the analyses presented here with a modified version that was forced to read all matrices as protein sequences." The next three sentences from the final paragraph of the Methods should now read "Five PhyloBayes runs under the fully parameterized CAT model were run, and each converged by 2,000 cycles based on time series plots of the likelihood scores and number of partitions. The runs were allowed to run, each for more than 3,500 cycles, and estimated about 300 categories for the model." instead of "Five PhyloBayes runs under the fully parameterized CAT model each converged at around 1,500 cycles (at least 86,000 generations) based on time-series plots of the

likelihood scores and number of partitions. The runs were allowed to run for 5,000 cycles for two runs and 2,500 cycles for three runs. The runs estimated 140 (±10) categories for the model." The original Letter's Supplementary Figures 2–9 have also been updated with the results of analyses based on the corrected matrices. Differences in results for these figures include increased support for some relationships highlighted in the original manuscript, and changes in some relationships within Bivalvia. These changes do not alter any of the conclusions of our manuscript. We thank Hervé Philippe, Raphael Poujol and Béatrice Roure for bringing this error to our attention.
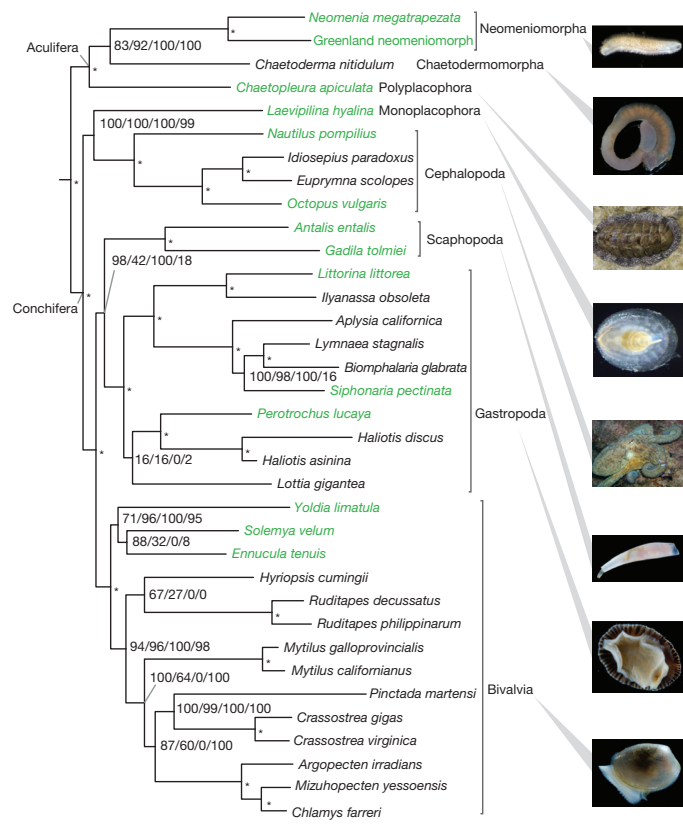

Figure 1 | This is the corrected Fig. 2 of the original Letter.