Developing and applying methods of molecular clock calibration to date then assembly of metabolisms and the deep species divergences in the tree of life

Supervisory team:
Main supervisor: Prof Philip Donoghue (University of Bristol)
Second supervisor: Prof Tim Lenton (University of Exeter)
Prof Ziheng Yang FRS (University College London), Dr Mario dos Reis (Queen Mary University of London), Prof Davide Pisani (University of Bristol), Dr Tom Williams (University of Bristol)

Host institution: University of Bristol

Project description:
The molecular clock provides the only viable means of establishing evolutionary timescales and measuring rates of molecular evolution. It works by calibrating molecular evolution to time, usually based on fossil evidence that constrains the ages of one or more clades present in a molecular phylogeny. Perversely, the molecular clock is most needed but least suited to evolutionary lineages that have a poor or no record, such as divergences deep within the tree of life, representing some of the most formative events in evolutionary history, from the origin of eukaryotes to the last universal common ancestor. This project proposes to overcome these challenges through the introduction of a new suite of calibrations that exploit gene duplication and horizontal transfer events, as well as endosymbioses, which impose relative constraints on the ages of clades, extending the utility of fossil calibrations and dramatically improving the accuracy and precision of evolutionary timescales where they are needed most – for the deepest (and shallowest) divergences in the tree of life.

You will development the theory and practical implementation these new approaches to the calibration of molecular clocks, as well as their application to real-world problems in evolutionary history. This will centre on development of MCMCtree (part of the PAML package) which is developed, maintained and freely distributed Supervisor Yang. You will develop MCMCtree to implement the novel classes of calibration strategy including cross-bracing and the relative age of clades that are not hierarchically related (via endosymbioses or HGTs). For example, cross-bracing will be implemented through the prior on node ages such that the same speciation nodes in different paralogues share the same ages; for relative calibrations the prior on node ages must ensure that donors are older than recipients of genes, genomes, or organelles. This will achieve coherent Bayesian inference of divergence times, integrating the information on times from fossil calibrations, on genetic distances from molecules, and on time-sharing or time-polarity due to gene duplications and gene-transfer, respectively. This is computationally and statistically superior to ad hoc approaches implementing geological age constraints on nodes age or post hoc approaches subsampling the posterior distribution of trees for topologies in which the relative age constraints are observed.

You will apply this software technology to (i) improving the evolutionary timescale for the emergence of the fundamental lineages of extant life (the gamut of the tree of life), and (ii) elucidating the evolutionary assembly of reaction pathways within the nitrogen metabolism, as characterised by the KEGG database. You will be combine these results in modelling the evolutionary assembly of the nitrogen cycle at an Earth System level.