

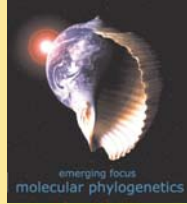
Nucleotide substitution strand bias and control region orientation in bivalve mitochondrial genomes

Gerhard Steiner, Hermann Dreyer & Martina Knapp

Molecular Phylogenetics, Dept. Evolutionary Biology, University of Vienna, Austria

Althanstr. 14, A-1090 Vienna, Austria

Email: gerhard.steiner@univie.ac.at



BACKGROUND

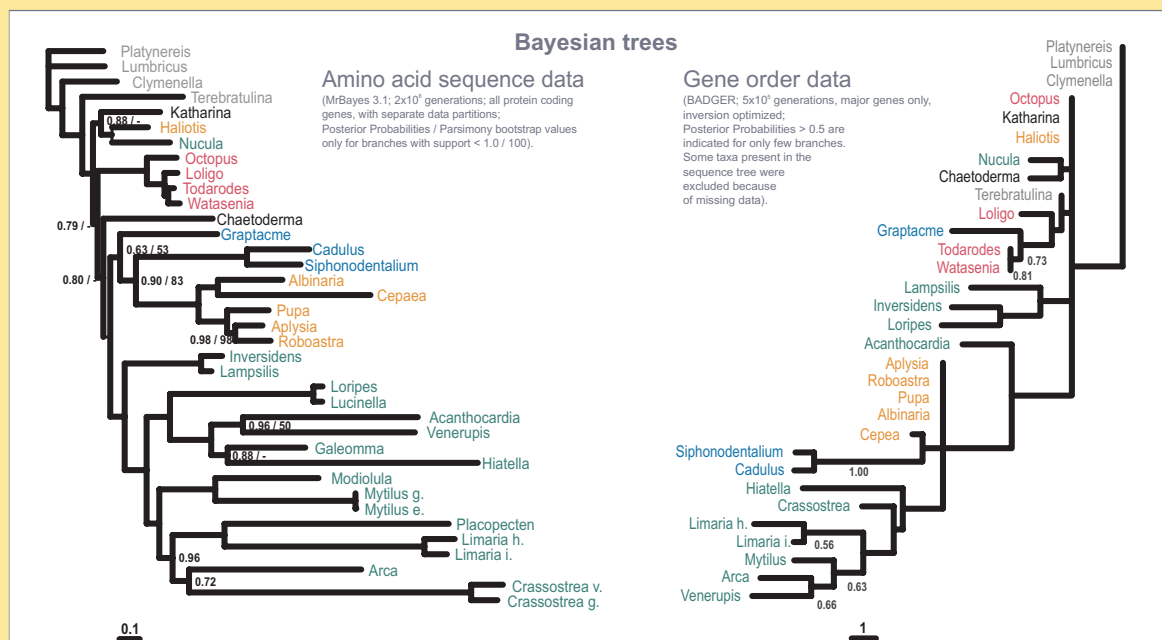
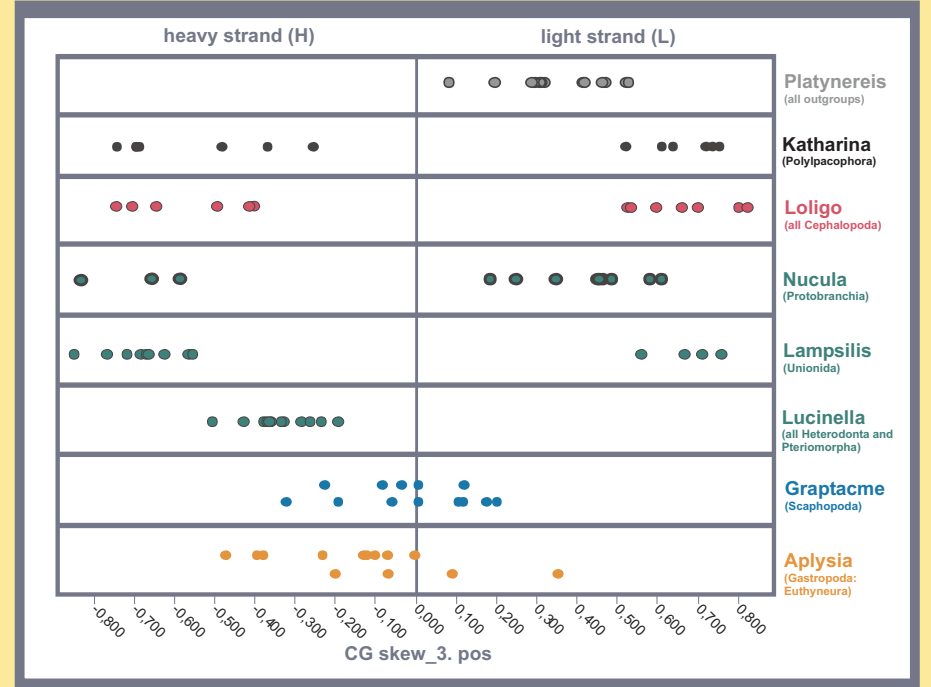
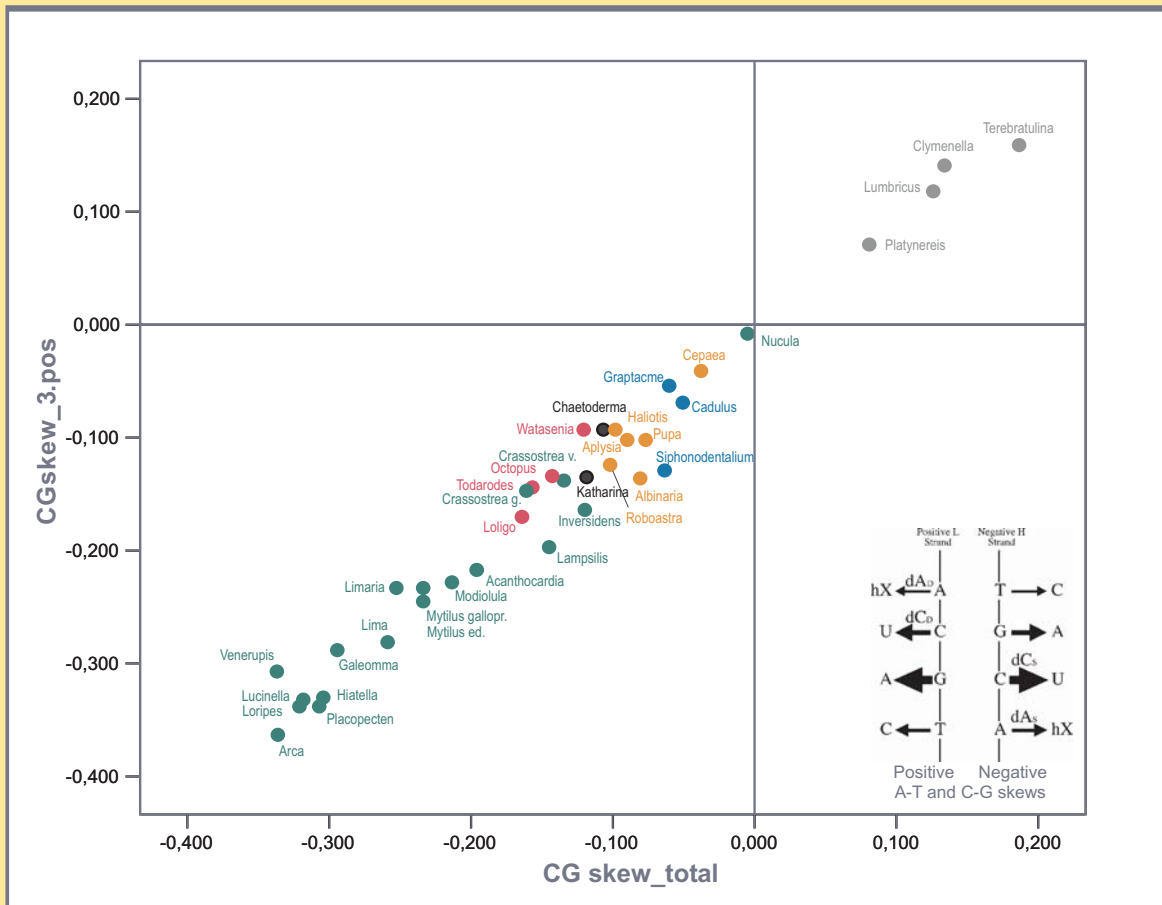
Different types of nucleotide substitutions occur in different frequencies in mitochondrial genes depending on the strand the genes are encoded on. As a result, the genes on the so-called light strand (L) are rich in A and C and have positive A-T and C-G skews. Genes encoded on the heavy strand (H) are rich in T and G and have negative A-T and C-G skews. This is most obvious in the weakly selected third codon positions. The mechanism considered responsible for this bias is the asymmetric replication process of the mitochondrial genome due to the asymmetry of the origin of replication or control region. The L-strand contains the "preferred side" of the control region and is replicated more rapidly than the H-strand.

We compare the protein coding mt-genes of 18 bivalve species to those of other molluscs, annelids, and a brachiopod to assess implications for phylogenetic inference and putative inversions of the control region.

RESULTS and DISCUSSION

All but three species of bivalves have their genes encoded on the same strand, the H-strand, whereas the outgroup species have all genes on the L-strand. Only the protobranch *Nucula nucleus* and the two unionids have genes on both strands. This is most likely the plesiomorphic situation also found in the polyplacophoran *Katharina tunicata*, the caudofoveate *Chaetoderma nitidulum*, the cephalopods, gastropods, scaphopods, and most other metazoans.

All Mollusca show negative C-G skews. Skews are strongest in the bivalves with all genes on the H-strand. *Nucula* shows no overall skew, although individual genes are strongly skewed. In contrast, scaphopods and euthyneuran gastropods have weaker skews that do not correlate to the strand they are encoded on.



Comparing the phylogenetic trees inferred from amino acid sequences and gene order data unexpected correlations of nucleotide substitution rates and gene rearrangement rates become apparent. The species with near-plesiomorphic gene order root at the base of the sequence tree with short branches. We conclude that the "mono-stranded" mt-genomes in the outgroup and in bivalves arose independently from "bi-stranded" genomes and cannot be explained by an inversion of the control region alone. However, mixed skews as in *Graptacme* and *Aplysia* may indicate a recent inversion of the control region and an incomplete subsequent skew shift.

Gene rearrangement rates increased independently in bivalves and gastropods because *Nucula* and *Haliotis* have near-plesiomorphic gene orders.

The "mono-stranded" bivalve species are monophyletic in the sequence tree. The unusual position of the "bi-stranded" species is likely due to different patterns and rates of nucleotide and amino acid substitutions (nucleotide sequence trees have the same topology!).

The apparent failure to recover correct trees from mt-genome data poses considerable challenges for bioinformatics. The strand-specific substitution rates need to be considered in the models of likelihood-based methods. For gene order data, algorithms simulating and optimizing rearrangement pathways are urgently needed.

Although mt-genome data may not serve for inferring the origin of Bivalvia, we are convinced that improved phylogenetic inference algorithms together with a dense taxon sample will resolve the relationships of the major bivalve lineages.

*) Hassanin et al. (2005) — Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of metazoa, and consequences for phylogenetic inferences. *Systematic Biology* 54: 277–298.