

Vertebrate Hox Gene Clusters and the Evolution of Phylogenetic Footprints

Peter F. Stadler¹, Claudia Fried¹, Sonja Prohaska¹, Christoph Flamm², Günther Wagner³,
Chi-Hua Chiu⁴, Chris Amemiya⁵

¹ University Leipzig, Department of Computer Science, ² University of Vienna; Institute for Theoretical Chemistry and Structural Biology, ³ Department of Ecology & Evolutionary Biology, Yale University, ⁴ Rutgers University, ⁵ Virginia Mason Research Center, Seattle

Abstract

Evolutionarily conserved non-coding genomic sequences represent a potentially rich source for the discovery of gene regulatory regions. Since these elements are subject to stabilizing selection they evolve much slower than adjacent non-functional DNA. These so-called phylogenetic footprints can be detected by comparison of the sequences surrounding orthologous genes in different species. Therefore the loss of phylogenetic footprints as well as the acquisition of conserved non-coding sequences in some lineages, but not others, can provide evidence for the evolutionary modification of cis-regulatory elements.

A statistical model of footprint evolution that allows us to estimate the loss of sequence conservation that can be attributed to gene loss and other structural reasons. This approach to studying the pattern of cis-regulatory element evolution, however, requires the comparison of very long sequences from many species. We have therefore developed an efficient software tool for the identification of corresponding footprints in long sequences from multiple species.

We apply this novel method, among other systems, to the evolution of the HoxA cluster in teleost fish species. For instance, we find that the loss of conservation after cluster duplication is more extensive than expected from structural reasons. This suggests that binding site turnover and/or adaptive modification may also contribute significantly to the loss of sequence conservation. Novel methods for estimating loss rates of tests to compare relative loss rates in different lineages can be used to detect enhanced rates of evolution in non-coding conserved DNA. We demonstrate, furthermore, that non-coding sequence conservation can be used successfully as a valuable source of phylogenetic information independent of protein coding sequences.

Publications can be found <http://www.bioinf.uni-leipzig.de/Publications/index.html>