Sonja J. Prohaska

Bioinformatics
Department of Computer Science
University of Leipzig
Hätelstr. 16-18
D-04107 Leipzig, Germany

sonja@bioinf.uni-leipzig.de
Phone: ++49 341 97 167 02
Fax: ++49 341 97 166 79

ABSTRACT

Evolution of
Conserved Non-Coding Sequences

Dissertation
Bioinformatics
October 7th, 2005

In many eukaryotic genomes only a small fraction of the DNA codes for proteins but the non-protein coding DNA harbors important genetic elements directing the development and the physiology of the organisms. Such elements like promoters, enhancers, insulators and micro-RNA genes and their modular organization is responsible for the complexity of metazoans. The molecular evolution of these genetic elements is difficult to study because their functional significance is hard to deduce from sequence information alone.

In this thesis, we will introduce you to the modes and relevance of regulatory control with a focus on transcriptional regulation. The little, that is known about the evolution of regulatory elements is summarized as well as the biological background of our study system, a cluster of very conserved developmental genes called Hox genes.

Detection of conserved regulatory regions by comparative genomics plays a key role for our work. We therefore propose improvements to our phylogenetic footprinting program, called `tracker`. Integration of `DIALIGN2` together with its special features refine footprint detection and enable us to find even more diverged footprints. This is of biological relevance since regulatory proteins may bind to rather short and unspecific motifs. With the `barbeque` approach we are able to break the main shortcoming of footprint detection via local alignment methods. Taking a set of motifs we are able to search for clusters of these sites independent of their order and orientation. This includes a biologists common knowledge about the organization of a regulatory element into the automated detection of regulatory regions. Even very different assemblies of the same set of motifs can result in identical functions.

Special interest goes to functional changes and novelties after sequence duplication. In particular, the Hox cluster duplications are supposed to play a role in major evolutionary transitions in chordate evolution. We therefore collected available data for Hox and ParaHox sequences of teleosts ('real fishes') and reconstructed the duplication history considering also phylogenetic signals from detected footprints. Our data are consistent with the assumption that these fishes 'recently' underwent a genome duplication. Therefore, teleost Hox cluster sequences resemble the ideal study system for research on the non-coding sequence evolution after duplication.

We have developed an approach to study the rate of evolution of functional non-coding sequences at a macro-evolutionary scale. This approach is specially designed for non-coding sequence evolution in contrast to other relative rate tests that use evolutionary models from coding sequence evolution. We have observed asymmetric evolution of the recently duplicated posterior Hox genes with one of the two copies evolving faster. Hence, there is evidence for a concerted asymmetric divergence of coding sequences on the same cluster as well as between non-coding sequences.

On the way to understand the evolution of non-coding sequences we have advanced in identifying a regulatory element that is likely to be involved in the transition from fin to limb. Furthermore, we have gotten some more insights in the evolutionary constraints on Hox clusters like exclusion of interspersed repeats from the inner core of gnathostome Hox clusters.