

Selective Constraint Beyond Apparent Sequence Conservation

Olga A. Vakhrusheva
*Department of
bioengineering and
bioinformatics, M.V.*

*Lomonosov Moscow State
University, Building B,
Moscow 119899, Russia;
Institute for Information
Transmission Problems
RAS, Bolshoi Karetny
pereulok 19, Moscow
127994, Russia
E-mail:
vakh57@gmail.com*

Georgii A. Bazykin
*Department of
bioengineering and
bioinformatics, M.V.*

*Lomonosov Moscow State
University, Building B,
Moscow 119899, Russia;
Institute for Information
Transmission Problems
RAS, Bolshoi Karetny
pereulok 19, Moscow
127994, Russia
E-mail: gbazykin@iitp.ru*

Alexey S. Kondrashov
*Department of
bioengineering and
bioinformatics, M.V.*

*Lomonosov Moscow State
University, Building B,
Moscow 119899, Russia;
Life Sciences Institute and
Department of Ecology
and Evolutionary Biology,
University of Michigan,
Ann Arbor, Michigan
48109, USA*

Abstract

Functional significance of non-coding sequences is commonly assessed through above-random conservation between genomes. However there is accumulating data that points to the rapid turnover of regulatory sequences. In the case of generality of this phenomenon direct similarity-based approaches would be insufficient to infer functional significance of non-coding sequences.

Here we address the question of whether functional conservation is possible without sequence conservation. We apply bioinformatics approach to this problem by tracing patterns of conservation in orthologous introns in two distant pair of species. Our findings suggest that orthologous non-coding genomic compartments may continue to play a functional role in distant species even after having diverged beyond recognition at the sequence level.

1. Introduction

At moderate evolutionary distances, functional

significance of non-coding sequences can be readily assessed through their above-random conservation between genomes. However, there is accumulating evidence that regulatory sequences are capable of rapid turnover, and, generally, conservation of sequence is not a necessary prerequisite for conservation of function.

For example, sequences at orthologous loci conserved among teleosts and mammals, but lacking significant sequence similarity between mammals and teleosts, can drive similar patterns of gene expression in zebrafish transgenic assays [1]. These findings may point to insufficiency of sequence similarity-based approaches to analysis of functional conservation. We used a bioinformatic approach to the question of whether functional conservation is possible without sequence conservation, and whether it can be inferred from genome wide comparative studies.

2. Results and Discussion

We considered two pairs of species chosen so that the evolutionary distance between species in a pair is significantly less than between species from two different pairs, though large enough to ensure that

conservation in non-coding regions is mainly due to their functional role. We studied two such quartets, of dipterans and vertebrates. We hypothesized that functional conservation between pairs of species should lead to above-average occurrence of conserved (within each pair) sequences at loci which are orthologous between pairs, even when no conservation between pairs is observed. Orthologous introns were chosen as sequence units within which conservation was analyzed, as their orthology can be inferred even at large phylogenetic distances through flanking exons.

We selected introns orthologous in all the four species (defined as the introns in orthologous positions of coding sequences in orthologous proteins). Then excluded from the analysis introns with significant local similarities between species from different pairs and, thus, considered only those introns that are unalignable between pairs.

Nevertheless, there was a 2-8-fold excess of 4-sets of orthologous introns in which each species pair carried a conserved non-coding element, compared to the random expectation expectation in dipteran and vertebrate species quartets.

We also have shown a strong correlation between intron sequence similarity (or intron length) within one species pair and its retention in the other pair (again taking into consideration only introns unalignable between different pairs) for both dipteran and vertebrate quartets. These results indicate that selective constraint, presumably caused by retention of the ancestral function, often persists even in unalignable DNA segments.

3. References

- [1] S. Fisher, E. A. Grice, R. M. Vinton, S. L. Bessling, and A. S. McCallion, *Conservation of RET regulatory function from human to zebrafish without sequence similarity*, *Science* (2006), v. 312, №. 5771, pp. 276-279.