Evolution of protein and non-coding RNA genes studied with comparative genomics

AKADEMISK AVHANDLING

som för avläggande av medicine doktorsexamen
vid Sahlgrenska Akademin vid Göteborgs universitet kommer att offentligen försvaras
i föreläsningssal Tor Bjurström, Medicinaregatan 3B, Göteborg,
tisdagen den 18 januari 2011, kl 09.00

Av

Marcela Davila

Fakultetsopponent:
Professor Paul P. Gardner
Wellcome Trust Sanger Institute, United Kingdom

Avhandlingen baseras på följande delarbeten:

I   Inventory and analysis of the protein subunits of the ribonucleases P and MRP provides
        further evidence of homology between the yeast and human enzymes
        Rosenblad, M.A., **Lopez, M.D**., Piccinelli, P. and Samuelsson, T.
        Nucleic Acids Res. 2006 34, 5145-5156.

II   Computational screen for spliceosomal RNA genes aids in defining the phylogenetic
        distribution of major and minor spliceosomal components
        Davila Lopez, M., Alm Rosenblad, M. and Samuelsson, T.
        Nucleic Acids Res. 2008 36, 3001-3010.

III   Early evolution of histone mRNA 3' end processing
        Davila Lopez, M. and Samuelsson, T.
        RNA. 2008 14, 1-10.

IV   Analysis of gene order conservation in eukaryotes identifies transcriptionally and
        functionally linked genes
        Davila Lopez, M. and Samuelsson, T.
        PLoS One. 2010 5(5):e10654.

V   eGOB: Eukaryotic Gene Order Browser
        Davila Lopez, M. and Samuelsson, T.
        Submitted for publication

GÖTEBORGS UNIVERSITET

Evolution of proteins and non-coding RNAs genes studied with comparative genomics

Marcela Davila

Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine
Sahlgrenska Academy at University of Gothenburg, Box 440, SE-405 30, Gothenburg.

Abstract

The identification of protein and non-coding RNA (ncRNA) genes is one important step in the analysis of a genome. This thesis focuses on the identification and analysis of proteins and ncRNAs homologues by exploiting a variety of computational methods in order to reach conclusions as to their structure, function, evolution and regulation. This work is composed of two different parts. One deals with computational prediction of protein and ncRNA homologues from different ribonucleoprotein (RNP) complexes and the other addresses problems related to non-random gene order in eukaryotes.

In the first part RNPs that were previously not well explored with respect to their phylogenetic distribution were examined. Thus, homology-based methods were employed to analyze the RNP complexes of RNase P, RNase MRP and the spliceosome as well as RNPs and RNA structures involved in the 3' end processing of histone mRNAs. We identified a large number of previously unrecognized homologues that improved our understanding of the evolution of the different RNPs. For example, homology relationships of the RNases P and MRP proteins were identified providing further evidence of homology between the human and the yeast RNPs. We presented evidence that the histone 3' end processing machinery is more ancient than previously anticipated and can be traced to the root of the eukaryotic phylogenetic tree. We presented a detailed map of the distribution of the spliceosomal U12-type RNA genes, supporting an early origin of the minor spliceosome and pointing to a number of occasions where it was lost during evolution.

In the second part we generated gene order maps to show the localization of both protein and ncRNA genes in a wide range of eukaryotic organisms. Non-random gene order was then examined to identify the most important determinants of gene order conservation. One important conclusion was that gene pairs that are evolutionarily conserved and that are divergently transcribed are much more likely to be related by function as compared to poorly conserved gene pairs. The genes of such pairs are likely to be related also in terms of transcriptional control. Moreover, we presented the eukaryotic Gene Order Browser (eGOB), where data related to this project is available and where researchers can visualize and compare the evolution of gene organization in different organisms. In addition, the browser may be used to identify pairs of adjacent genes that are evolutionarily conserved and likely to be transcriptionally linked. eGOB is available at http://egob.bioimedicine.gu.se.