

Alessandra Carbone, Laboratoire Génomique des microorganismes  
Université Pierre et Marie Curie - CNRS

*Coding of evolutionary pathways in proteins: from sequence to function*

Today, networks of protein interactions do not describe protein-protein partnership at a residue level. This information is necessary to control protein behavior though. We shall present an approach based on sequence analysis to detect important information on protein binding sites and on mechanical and allosteric properties at the residue level. We shall use a fine reading of the conservation and co-evolution signals between residues in the protein sequences. This information is encoded in the tree topology of the distance tree associated to evolved sequences observable today. We shall present two methods, one applicable to protein families of variable divergence and the other to very conserved protein families.

Gergely J. Szöllősi, ELTE-MTA Biophysics Research Group  
Eötvös University

*Efficient Exploration of the Space of Reconciled Gene Trees*

Gergely J. Szöllősi(1), Wojciech Rosikiewicz(2), Bastien Boussau(3), Eric Tannier(3)  
and Vincent Daubin(3)

(1) ELTE-MTA "Lendület" Biophysics Research Group H-1117, Pa'zma'ny P. stny.  
1A., Budapest, Hungary

(2) Laboratory of Bioinformatics, Faculty of Biology, Adam Mickiewicz University,  
Pozna'n, Poland;

(3) LBBE, CNRS, UMR 5558, Universit'e Lyon 1, F-69622 Villeurbanne, France;

As molecular phylogeneticists, we infer gene trees based on sequence information. Unfortunately, sequences alone contain limited signal, and as a result phylogenetic reconstruction almost always involves choosing between statistically equivalent or weakly distinguishable relationships. Although each homologous gene family has its own unique story, they are all related by a shared species history, which could be helpful for gene tree inference. We have recently published a probabilistic reconciliation model, which describes the relationships between a gene tree and a species tree as a series of events, such as duplication, transfer and loss, speciation and extinction (Szöllősi et al. Syst. Biol. 2013). We now propose an efficient way to integrate sequences and reconciliation information in the inference of gene trees.

To design a species tree aware method for reconstructing gene phylogenies, the space of reconciled gene trees must be explored using information from both a model of sequence evolution and a reconciliation model. Such an exploration can be tedious with classical approaches. To circumvent this problem, we present a general probabilistic approach to exhaustively explore all reconciled gene trees that can be amalgamated as a combination of clades observed in a sample of gene trees. For a sample derived from the posterior distribution of trees obtained from a bayesian MCMC analysis, this approach provides an accurate approximation of gene tree likelihood.

We demonstrate using both simulations and biological sequences that gene phylogenies reconstructed using the joint likelihood are dramatically more accurate than those reconstructed using sequences alone. In fact, we find that even using a simplistic model of sequence evolution, the joint reconstruction yields significantly more accurate gene trees than the sequence-based inference with the complex model used in simulations. Considering 1099 homologous gene families from 36 genomes of cyanobacteria we find that the majority of phylogenetic discord results from errors in sequence based reconstruction that can be corrected using information aggregated across gene families by a putative species tree. The result is

a striking reduction in apparent phylogenetic discord, with resp. 24%, 59% and 46% percent reductions in the mean numbers of duplications, transfers and losses per gene family.

Our probabilistic method overcomes a fundamental limitation of recent parsimony based methods to improve gene trees given a putative species tree (David and Alm Nature 2011, Wu et al. Syst. Biol. 2013) by not having to rely on any ad hoc assumption about statistical support, while at the same time deploying approximations that make it more efficient than methods that rely on a local search of tree space (Akerborg et al. PNAS 2009).

The open source implementation of the method is available from <https://github.com/ssolo/ALE.git>.

#### References:

Akerborg, O., B. Sennblad, L. Arvestad, and J. Lagergren (2009)  
Simultaneous bayesian gene tree reconstruction and reconciliation analysis.  
Proc Natl Acad Sci U S A 106:57149.

David, L. A. and E. J. Alm. (2011)  
Rapid evolutionary innovation during an archaean genetic expansion.  
Nature 469:93 6.

Szöllősi, G. J., E. Tannier, N. Lartillot, and V. Daubin. (2013)  
Lateral gene transfer from the dead.  
Systematic Biology 62:386.

Szöllősi G. J., Rosikewicz W., Boussau B., Tannier E. and Daubin V. (2013)  
Efficient Exploration of the space of reconciled trees  
Systematic Biology doi:10.1093/sysbio/syt054

Wu, Y.-C., M. D. Rasmussen, M. S. Bansal, and M. Kellis. (2013)  
Treefix: Statistically informed gene tree error correction using species trees.  
Systematic Biology 62:11020.