Mathematically and biologically sound model for HGT-reconciliations

Paweł Górecki[1], **Jerzy Tiuryn**[1]

Keywords: horizontal gene transfer, duplication-loss model

1 Introduction

We present a model based on reconciliation for the duplication, loss and horizontal gene transfers (HGT) [1, 3, 4]. We define an evolutionary scenario (called H-tree) which can represent an evolution of genes in the context of evolution of species.

We are interested in all scenarios - not only parsimonious ones. A rewrite system for transforming H-trees is defined (see Figure 2). We prove that the system is mathematically (confluency, soundness and strong normalization) and biologically (we model all possible evolutionary scenarios) well founded. We show that the scenario in normal form (i.e. non-reducible) is minimal and unique in the size and in the cost computed as the number of gene duplications, losses and transfers. We show also that, for a given gene tree $\mathcal{G}$ and a given species tree $\mathcal{S}$, there could be many normal form H-trees representing evolutionary scenarios which match $\mathcal{G}$ and $\mathcal{S}$. Finally, we show that the normal form scenarios can be easily transformed into the reconciled trees [1, 5]. Also, algorithms for generating scenarios and reconciliation are developed. Concepts of this model are similar to the model presented in [2] in the context of duplication-loss model.

2 Model and the system of rules

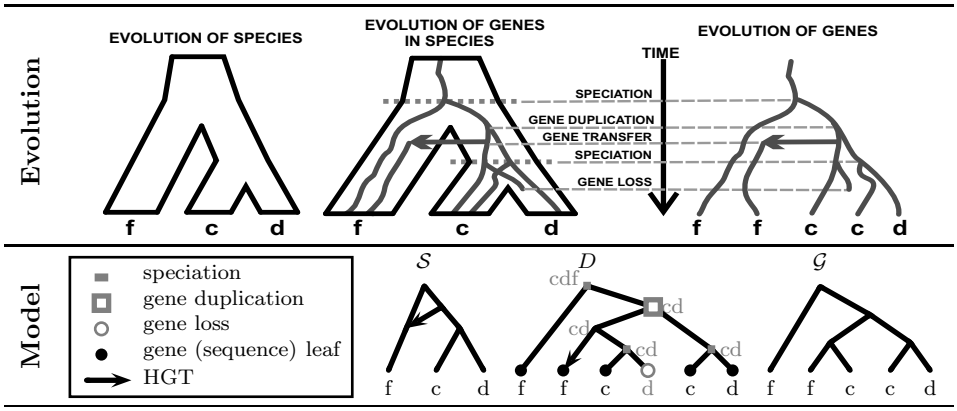

Figure 1: Evolution of genes in species with one horizontal gene transfer

In *Evolution* part of Figure 1, we show concepts of evolutionary gene and species trees and show an embedding of a gene family tree (rightmost) into a species tree (leftmost) with

[1]Institute of Informatics, Warsaw University, Banacha 2, 00-097 Warsaw, Poland, Earth. E-mail: {gorecki,tiuryn}@mimuw.edu.pl

one horizontal gene transfer. This picture can be represented in our model as shown in the bottom part of this figure. Thus, we have a gene tree $\mathcal{G}$ and a species tree with one transfer D and three species f, c and d. D is an H-tree, which represents an evolution of genes in species (we omit a formal definition).

Transfers in a species tree (and also in an H-tree) cannot form a cycle (see [1]). Moreover, each transfer can be associated with a time stamp determining when it happened. In formal approach the time dependencies between transfers can be modeled by partial orders.

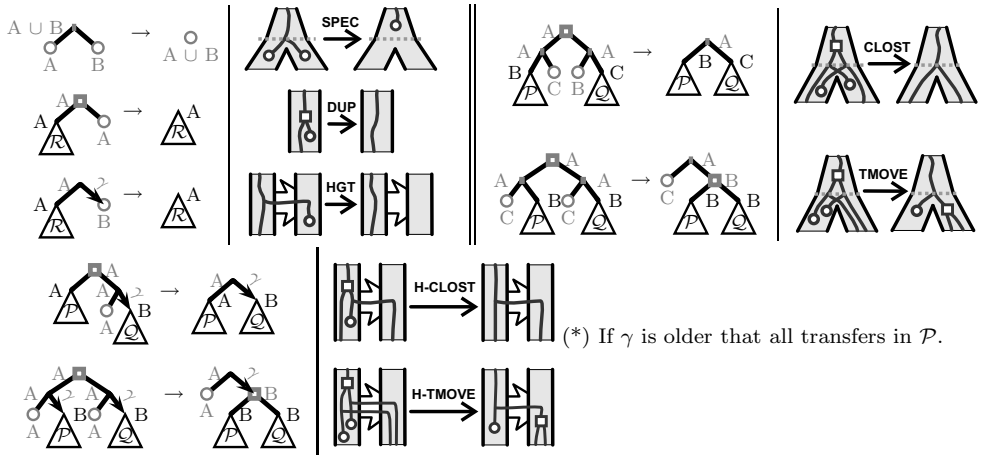

Figure 2: H-rules and their biological interpretations

The system of rules presented in Figure 2 can be used to transform H-trees (i.e., evolutionary scenarios). We claim that the system preserves biological properties of the trees, (that is, species trees, gene trees and relations between transfers associated with the transformed H-tree). The rules can be used to transform an evolutionary scenario into its equivalent normal form, which has the minimal cost and size. Thus, it is a natural and clear methodology for analyzing parsimonious and non-parsimonious scenarios in the transfer models.

References

[1] P. Górecki. Reconciliation problems for duplication, loss and horizontal gene transfer. In *RECOMB 2004*, pages 316–325, San Diego, 2004.

[2] P. Górecki and J. Tiuryn. On the structure of reconciliations. *Lecture Notes in Bioinformatics, RECOMB 2004 Workshop on Comparative Genomics (to appear soon)*, 2005.

[3] M. Hallett and J. Lagergren. Efficient algorithms for lateral gene transfer problems. In *Proceedings RECOMB 2001*, pages 149–156, New York, 2001.

[4] M. Hallett, J. Lagergren, and Ali Tofigh. Simultaneous identification of duplications and lateral transfers. In *RECOMB 2004*, pages 316–325, San Diego, 2004.

[5] R. D. M. Page. Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Systematic Biology*, 43:58–77, 1994.