

Ruriko Yoshida

Open Problems in Geometry of Cophylogeny

Ruriko Yoshida

Dept. of Statistics University of Kentucky

Joint work with C. Schardl, J. Jaromczyk, and P. Huggins

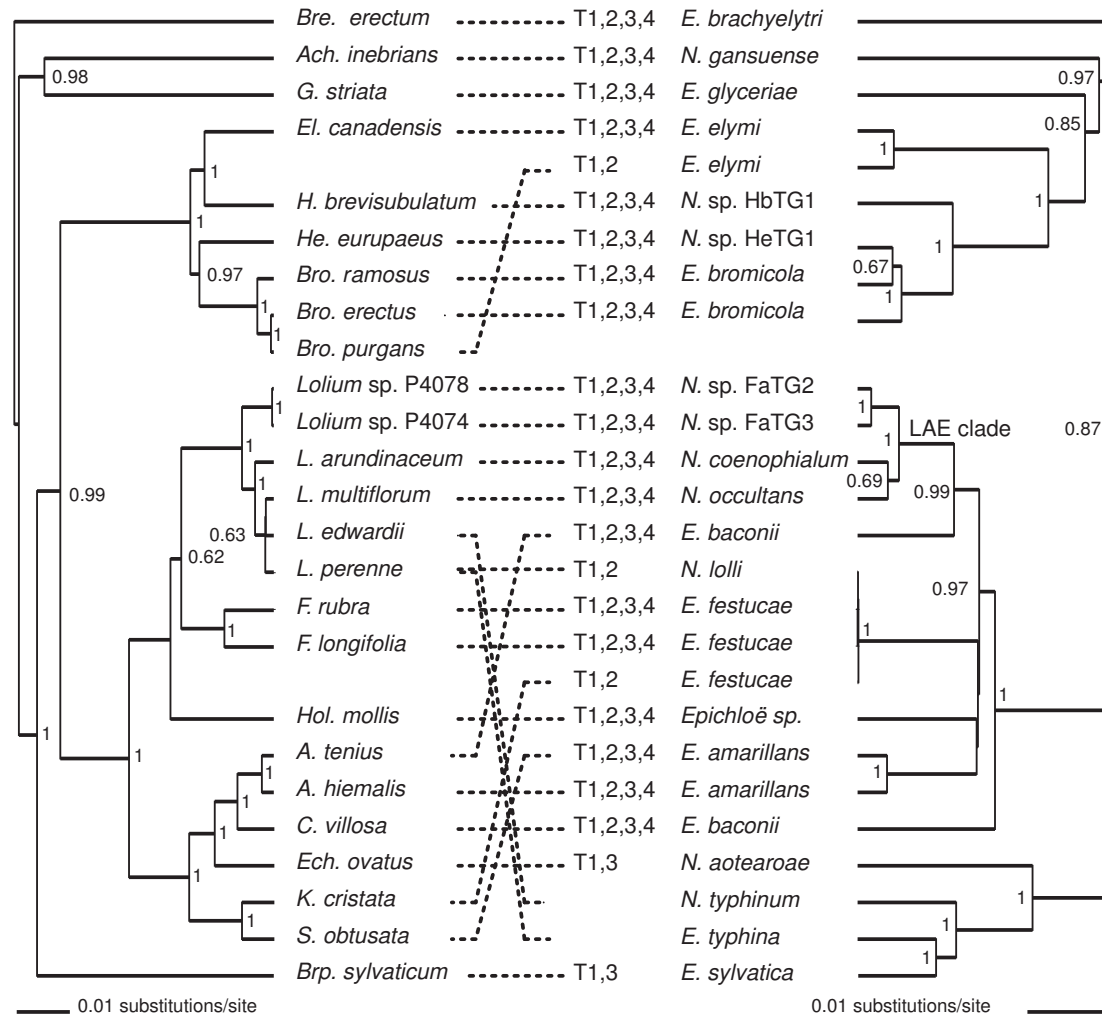


Figure 1: Ultrametric ML time trees for plant and endophyte data sets in [Schardl et al, 2008] constructed via BEAST. Hosts and their endophytes are indicated by dashed lines. Numeric values on nodes represent their posterior probabilities estimated by BEAST.

Cophylogeny

Suppose we have two sets of multi-species sequence data H and P . Let \mathcal{T}_H be the space of trees on H and \mathcal{T}_P be the space of trees on P .

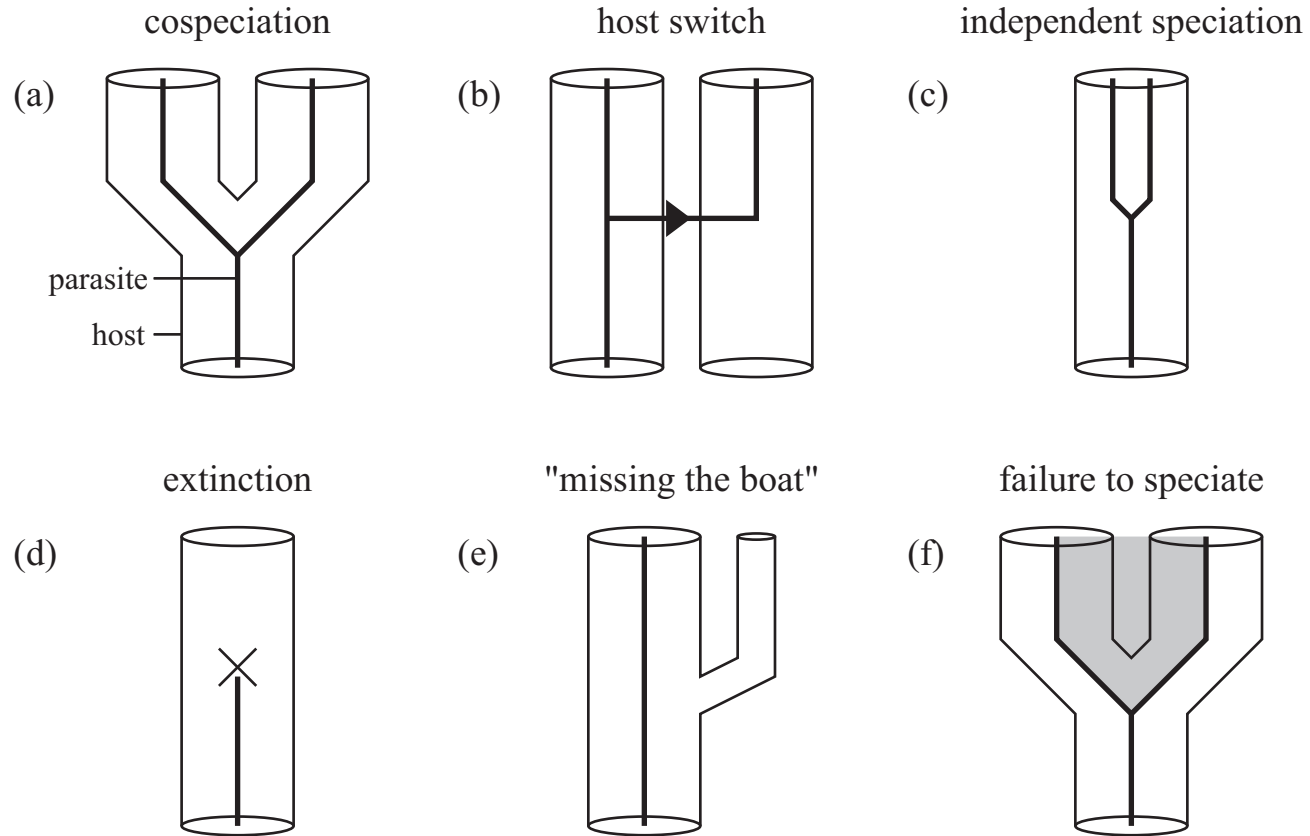
Assuming that there is a distribution on the cross product of tree spaces for T_H and T_P where T_H is a phylogenetic tree for H and T_P is a phylogenetic tree for P . A **cophylogeny** is a conditional joint distribution $P(T_H, T_P | H, P)$ on $\mathcal{T}_H \times \mathcal{T}_P$ which satisfies

$$P(T_H, T_P | H, P) \neq P(T_H | H, P) \cdot P(T_P | H, P)$$

Note: Even though two phylogenetic trees are correlated, tree topologies of T_H and T_P might differ. We can apply this to species and gene trees.

Assume: in the evolution history a divergence of a gene or a speciation occurs once at a time,

6 different processes in a host–parasite association



Geometry of Cophylogenetic trees

Definition: The support $S \subset \mathcal{T}_H \times \mathcal{T}_P$ of a cophylogeny is called a **space of cophylogenetic trees**.

Definition: Suppose the host or species tree T_H is given. The support of the conditional distribution $P(T_P|T_H, H, P)$, $S_{T_H} \subset \mathcal{T}_P$, is called the **space of cophylogenetic trees given T_H** .

Remark: In general $S_{T_H} \neq \mathcal{T}_P$ and $S \neq \mathcal{T}_H \times \mathcal{T}_P$.

Example: If we assume a perfect codivergence, that is, T_H and T_P are identical (for e.g., [Huelsenbeck et. al., 2000]), the space of cophylogenetic trees is

$$S = \{(D_H, D_P) : D_H \text{ is a tree metric for } T_H \text{ and } D_P \text{ is a tree metric for } T_P \text{ such that } T_H = T_P\}$$

The space of k -interval cospeciation

In evolution a speciation in host is likely to be followed by a reactionary speciation in parasite, and often vice versa. Combinatorially, this assumption can be made explicit by assuming that for each pair of host species A, B , and corresponding parasite species a, b , the number of edges between A, B is within k of the number of edges between a, b . We say such a cophylogeny satisfies **k -interval cospeciation**.

Proposition [Huggins and Y., 2008]

Under the 1-interval cospeciation with the given host tree T_H in taxa $\{1, 2, \dots, n\}$, if a tree T_P in taxa $\{1', 2', \dots, n'\}$ contains a quartet $[i'_1, i'_3; i'_2, i'_4]$ or $[i'_1, i'_4; i'_2, i'_3]$, and if the corresponding quartet in T_H generated by their hosts $\{i_1, i_2, i_3, i_4\}$ is $[i_1, i_2; i_3, i_4]$, then T_P cannot be the parasite tree for T_H .

Example

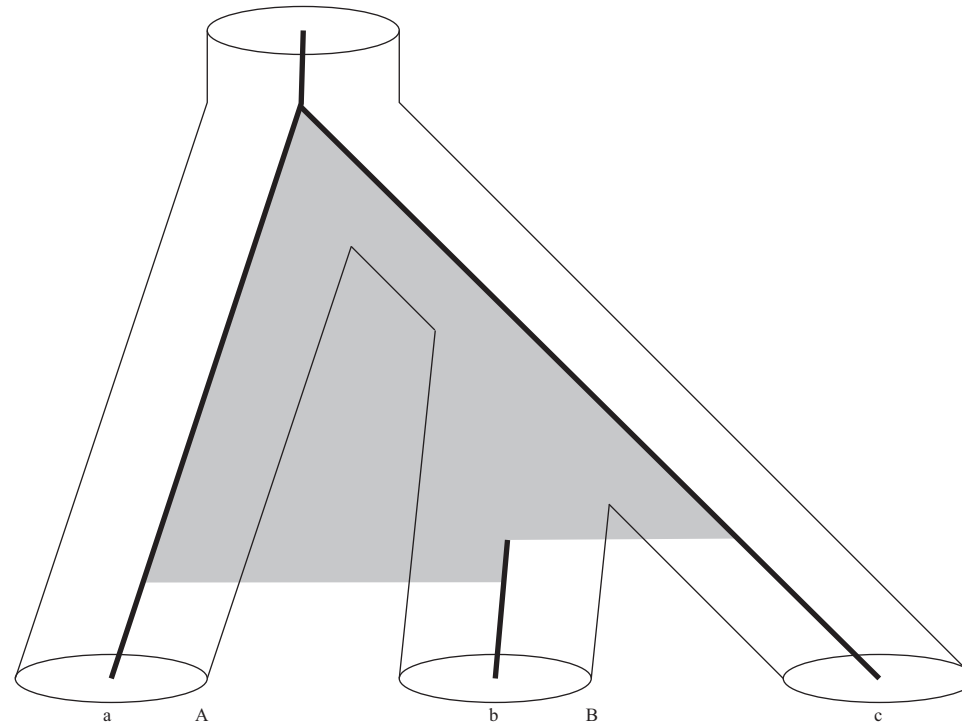


Figure 2: A parasite fails to speciate and then follows after host's speciation. These events are described with notation in [Pages, 2003].

Example

Example: $k = 1$ and $n = 4$

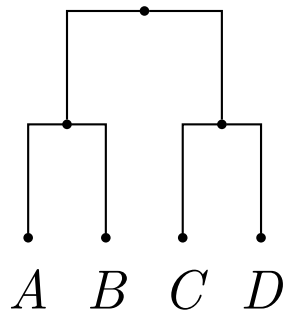
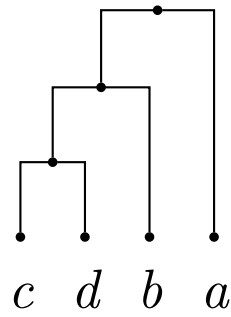
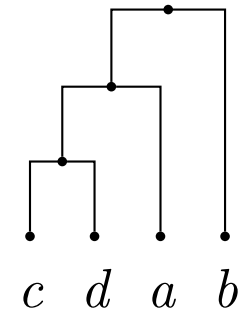
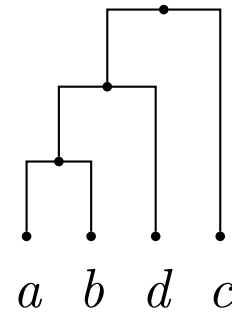
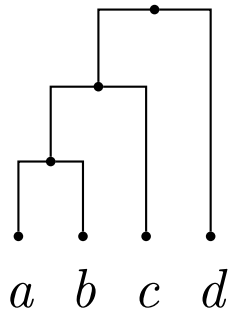
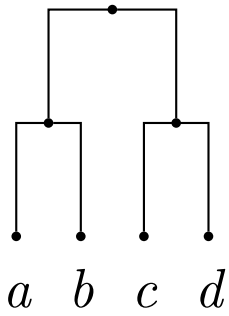


Figure 3: Host tree.

Example...

There are 5 possible parasite tree topologies.



Some open problems

- Given a host tree T_H , which tree topologies are possible for parasite tree, assuming k -interval cospeciation? How many parasite trees are possible?
- Are there interesting cophylogenies, such as k -interval cospeciation, for which the space of cophylogenetic trees admits a linear characterization analogous to the Four Point Condition?
- Given the host tree T_H and cophylogeny, for each parasite tree topology, how many ways to get the parasite tree topology with the host tree?
- Given the host tree T_H , consider the ideal of invariants I_{T_P} for each $T_P \in S_{T_H}$. Is there a nice characterization or algorithm to compute invariants in the intersection ideal $\cap I_{T_P}$?

Advertisement

Statistics postdoc position

We are looking for a new PhD in mathematics, statistics, or related field for a statistics Postdoc position in phylogenomics who is interested in algebra, combinatorics, statistics, and biology. Part of the postdoctoral work will focus on developing theory for analyzing co-divergence/co-evolution between genes or between hosts and their parasites.

For more details, please see

<http://polytopes.net/research/postdoc.php>

Mid-program on Molecular evolution and phylogenetic trees organized by Erick Matson, Peter Huggins, and Y. will be on April 2nd and 3rd.

Ruriko Yoshida

Thank you....

<http://arxiv.org/abs/0809.1908>