

BACKGROUND

Cophylogeny studies involve the analysis of the current interactions between organisms to estimate the historical or evolutionary interactions that occurred between the ancestors of these organisms. The input for such analysis requires a *tanglegram* that consists of a host and a parasite phylogenetic tree with their associations (Fig. 1). A phylogenetic tree is typically a binary tree that explains how a group of related organisms evolved over time. An association refers to a pair of organisms, one from each tree, which interact with one another. In general, the host tree is assumed to be independent of the parasite tree, but the parasite tree is dependent on the host tree to some degree.

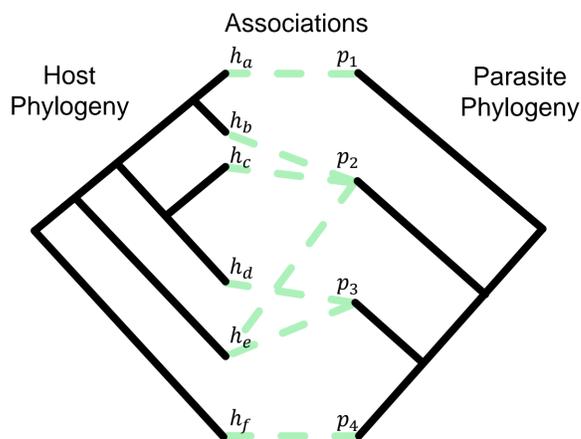


Figure 1. A tanglegram.

MOTIVATION

- Current cophylogeny tools are not good at solving problems containing widespread parasites (e.g. p_2 in Fig. 1 is on three hosts).
- Widespread parasites are common in nature.

AIM

Design a method to solve cophylogeny problems, where parasites can associate with more than one host (widespread parasites).

CO-EVOLUTIONARY EVENTS

There are five co-evolutionary events that are used in “event based” cophylogeny methods – *codivergence*, *loss*, *failure to diverge*, *duplication* and *host switch* (Fig. 2). In order to achieve the aim of this research, a new event, *spread*, is created. This event allows a widespread parasite to occupy another host without the need of diverging (unlike *host switch*).

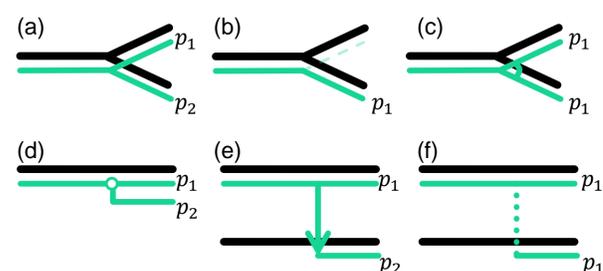


Figure 2. Co-evolutionary events: (a) codivergence, (b) loss, (c) failure to diverge, (d) duplication, (e) host switch and (f) spread.

METHOD

The *Split-and-Solve* method presented here follows a divide-and-conquer idea that solves small parts of the tanglegram first, then joins the solutions together. Currently, this process can only use timed phylogenies, where the ordering of the internal nodes is fixed.

Procedure

1. Split the tanglegram into smaller tanglegrams for each internal or multi-host leaf node in the parasite tree (P). (Fig. 3)
2. Solve the tanglegrams based on post order traversal for split nodes. (Fig. 4 & 5)
3. Join the solutions for each sibling pair of internal nodes. (Tables 1-3)
4. Save solutions and continue until the root of P is reached.

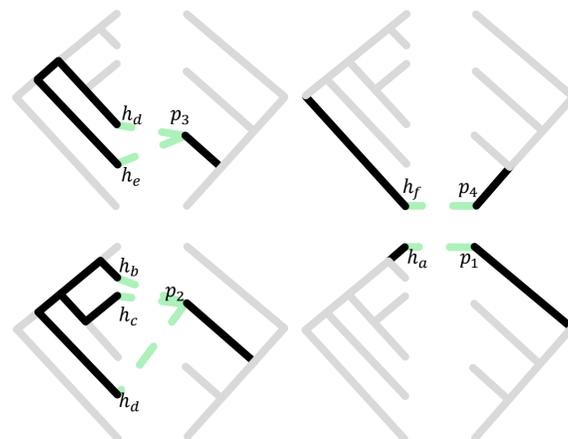


Figure 3. Individual tanglegrams.

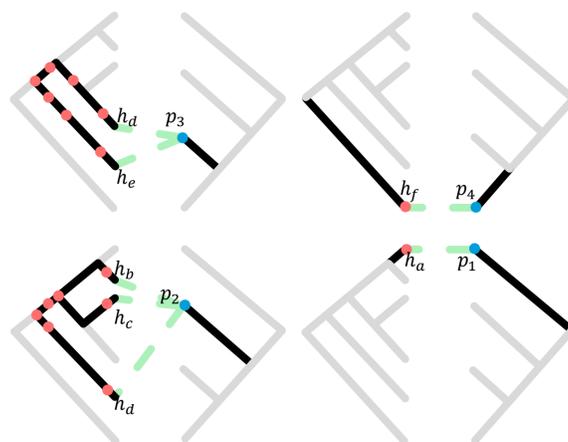


Figure 4. The possible locations of solution roots (red) for each parasite node (blue).

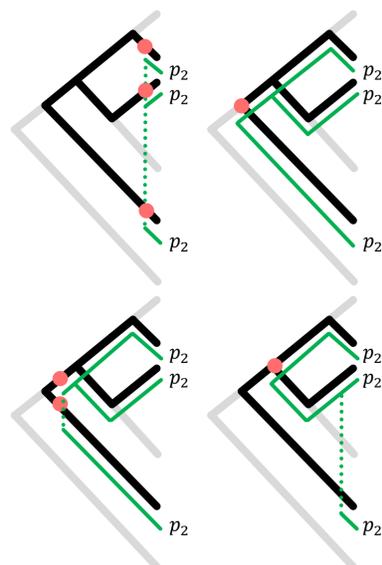


Figure 5. The solutions for mapping p_2 to various root locations on its split tanglegram.

		Map Edges Overlap	
		True	False
Both are on the Same Subtree	True	Duplication	Host switch or Codivergence
	False	X	Host switch

Table 1. The type of events used for joining a single lineage with a solution map based on if the edges in both maps overlap and/or share a common ancestor. The time of joint is earlier than the map's root time.

		Map Edges Overlap	
		True	False
Both are on the Same Subtree	True	X	X
	False	X	Host switch or Codivergence

Table 2. The type of events used for joining a single lineage with a solution map. The time of joint is older than the map's root time.

		Same Root Location	
		True	False*
Roots are on the Same Time	True	Duplication	Host switch or Codivergence
	False	X	Host switch or Codivergence

Table 3. The type of events used for joining a map with another map based on if the roots of both maps have the same time and/or location. The time of joint is older than the maps' root time. (* The maps are not embedded within one another)

RESULTS

The results produced by Split-and-Solve are compared to Jane^[1], a cophylogeny tool that can solve widespread parasite problems. The total number of events used for comparison is the sum of *spreads*, *host switches*, *losses* and *duplications*. The best solution from both methods for the tanglegram in Figure 1 is shown in Figure 6. There are a total of ten compared events in Jane's solution. The Split-and-Solve method obtained a much better solution with only five events. The difference in event number is mostly contributed by the use of *spread* in Split-and-Solve, which Jane does not have.

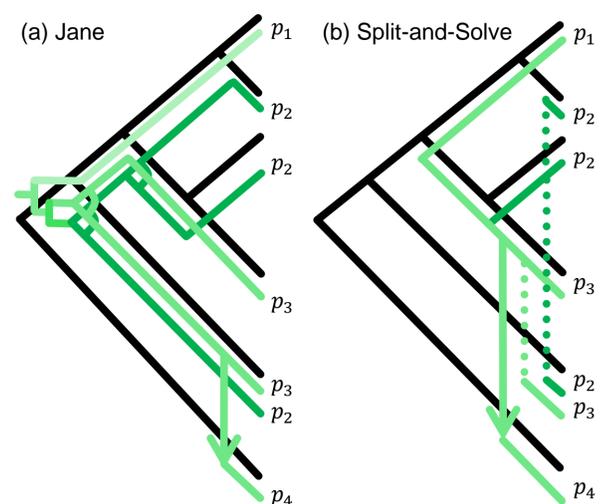


Figure 6. (a) Jane's solution contains zero codivergence, two duplications (at the root), one host switch, seven losses and three failure to diverges. (b) *Split-and-Solve* obtained two codivergences, one host switches, one loss and three spreads.

CONCLUSION

The Split-and-Solve algorithm is effective for solving cophylogeny problems that involve widespread parasites.

FUTURE WORK

- Integrate heuristic into Split-and-Solve.
- Reduce constraints on the joining algorithm to allow ancestor of widespread parasites to also be widespread.
- Incorporate Split-and-Solve into existing cophylogeny software or make a GUI for Split-and-Solve.

REFERENCES

[1] C. Conow, D. Fielder, Y. Ovadia, and R. Libeskind-Hadas. *Jane: a new tool for the cophylogeny reconstruction problem*. Algorithms for Molecular Biology, 5(1):16, 2010. ISSN 1748-7188. doi:10.1186/1748-7188-5-16.