A Polynomial Time Solution for the Widespread Cophylogeny Reconstruction Problem

Ben Drinkwater (A/Prof Michael Charleston)

School of Information Technologies

FACULTY OF ENGINEERING & INFORMATION TECHNOLOGIES

Background

A crucial factor in the study of coevolution is the specific evolutionary adaptations of a parasite to successfully inhabit its host [1].

- Often coevolutionary research has focused on cases where parasites are constrained to inhabiting a single host (*high host specify*) however this level of precision precludes many coevolutionary systems.
- We refer to systems which allow parasites to inhabit multiple hosts (*low host specify*) as the **widespread parasites problem**.
- The widespread parasites problem is known to be difficult as the underlying problem where parasites permitted only one association is **NP-Complete** [2].

Previous Research

Initial attempts at solving this problem have focused on solutions only involving failure to diverge events [3].

- To date the software tool Jane [4,5] is the only application capable of solving the widespread parasites problem, when the internal host nodes are fixed.
- This method is known to force widespread events deep within the cophylogenetic reconstructions which often **does not accurately represent biology**.

Introducing the Spread Event

To overcome the limitations caused by only using the *failure to diverge* event this research proposes leveraging the *spread* event [6].

- A spread event represents a switch between host lineages without a divergence event. This event matches known coevolutionary scenarios such as the case where the same species of lice often switch between different species of penguins at joint breeding sites [7].
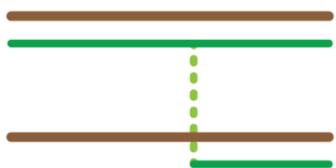

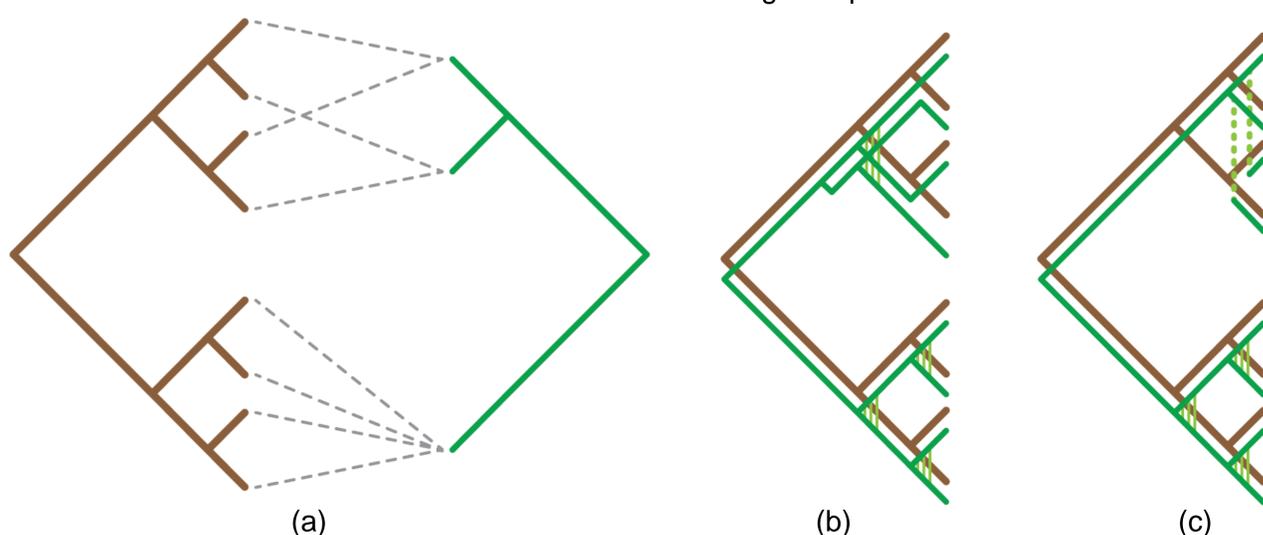**Fig 1:** The spread event

(a) (b) (c)

Fig 2: A tanglegram of a coevolving system including widespread parasites (a) and two Pareto optimal reconstructions. The solution recovered by Jane (b) has a cost of 10. The newly proposed algorithm recovers solution (c) with a cost of only 8.

Handling Widespread Parasites

This method reduces the widespread parasites problem to the case where parasites are only permitted one association through the construction of *association trees*.

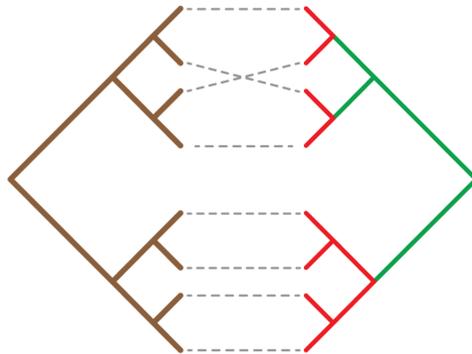- Association trees are constructed to **maximize** the number of solutions by mirroring the host trees.

Fig 3: The *association trees* generated for the tanglegram in Figure 2.

The association tress are solved using *failure to diverge* and *spread*.

- The solutions for each associated tree are mapped into the host tree.
- This step runs in $O(n^3)$ using a previously implemented **dynamic programming algorithm**. As there are $O(n)$ association pairs this process takes $O(n^4)$.
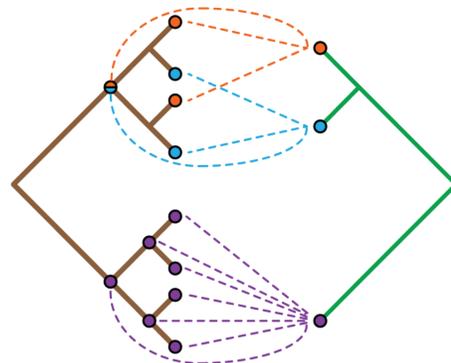

Fig 4: Set of parasite leaf mapping sites

The initial parasite tree is then solved using all $O(n)$ permutations of the parasite leaf nodes.

- This is completed using a Node Mapping algorithm I have previously implemented which runs in $O(n^3)$.
- Therefore the **total running time for this algorithm is $O(n^4)$** which is **the first polynomial solution** solving this problem using the spread event.

Results

Comparing algorithms.

- To fairly compare the latest release of Jane and this New Algorithm I have assembled a **wide range of biological data** featuring parasitism, mutualism and symbiosis, and host-pathogen systems.
- Each algorithm runs within a metaheuristic framework with the minimum cost solutions recorded in Table 1.

Table 1: The minimum cost scores recovered for both Jane 4 and the New Algorithm.

Data Sets	Jane's Best Solution	New Algorithm's Best Solution
Set 1 [8]	67	29
Set 2 [9]	33	25
Set 3 [10]	52	40
Set 4 [11]	148	78
Set 5 [12]	4	2
Set 6 [13]	9	8
Set 7 [14]	15	13
Set 8 [15]	122	55
Set 9 [16]	17	12
Set 10 [17]	8	8

- These results show a **reduction in the minimum cost by 43%.** This result strengthens the argument of the spread event as a viable method for reducing the cost of a cophylogenetic reconstruction where widespread parasites are permitted.

Further Work

An extension of this research is to consider the more **generalised case** where widespread parasitic events are not restricted to occurring before other biological events. This would allow for more biologically accurate solutions.

Conclusion

This research offers the **first polynomial solution** capable of solving the **widespread cophylogeny reconstruction problem** using spread events. This result is a significant step forward in the study of coevolving systems.

References

[1] Poulin, R., Evolutionary Ecology of Parasites, volume 2. Princeton University Press, 2nd edition, 2007. ISSN 0020-7519
[2] Ovadia, Y., et al, "The Cophylogeny Reconstruction Problem is NP-Complete", Journal of Computational Biology, vol.18, no.1, p.59-65, 2011.
[3] Charleston, M., Perkins, S, Traversing the tangle: Algorithms and Applications for Cophylogenetic Studies, Journal of Biomedical Informatics, vol. 39, no. 1, p. 67-71, 2006
[4] Conow, C., et al., "Jane: A new tool for the Cophylogeny Reconstruction Problem," Algorithms for Molecular Biology, vol.5, no.1, p.16, 2010.
[5] Conow, C., et al. Jane's Cost Models and Events, 2012. http://www.cs.hmc.edu/~hadas/jane/CostEvents.html.
[6] Qiao, A., A Novel Solution for Dealing with Widespread Parasites in Cophylogeny Reconstruction, Honours Thesis, University of Sydney, 2012.
[7] Banks, J., et al. Cophylogenetic relationships between penguins and their chewing lice. Journal of Evolutionary Biology, vol. 19, no. 1, p.156–166, 2006
[8] Gomez-Acevedo, S., et al, Neotropical mutualism between *acacia* and *pseudomyrmex* phylogeny and divergence times. Molecular Phylogenetics and Evolution vol. 56 no. 1 p. 393-408, 2010
[9] Refregier, G., et al., Cophylogeny of the anther smut fungi and their caryophyllaceous hosts: prevalence of host shifts and importance of delimiting parasite species for inferring cospeciation. BMC Evolutionary Biology vol. 8 no. 1, p. 100, 2008
[10] Sorenson, M.D., et al., Clade-limited colonization in brood parasitic finches. Systematic Biology, vol. 53, no. 1, p. 140-153, 2004
[11] McLeish, M., Van Noort, S., Codivergence and multiple host species use by fig wasp populations of the ficus pollination mutualism. BMC Evolutionary Biology, vol. 12, no.1, p. 1, 2012
[12] Paterson, et al., Have chondracanthid copepods co-speciated with their teleost hosts? Systematic Parasitology vol. 44 no. 2, p. 79-85, 1999
[13] Hafner, M., et al., Cophylogeny between pocket gophers and chewing lice. Tangled trees. University of Chicago Press, Chicago, 2003, p. 195-220
[14] Jackson, A., Charleston, M., A cophylogenetic perspective of RNA-Virus evolution. Molecular biology and evolution vol. 21, no. 1, p. 45-57, 2004
[15] Lockyer, A., et al, The phylogeny of the schistosomatidae based on three genes with emphasis on the interrelationships of schistosoma weinland, 1858. Parasitology vol. 126 no. 3, p. 203-224, 2003.
[16] Weckstein, J.D.: Biogeography explains cophylogenetic patterns in toucan chewing lice. Systematic Biology vol. 53 no. 1, p.154-164, 2004
[17] Dilcher, et al., Genetic characterization of Tribec virus and Kemerovo virus, two tick-transmitted human-pathogenic Orbiviruses. Virology vol. 423, no. 1, p. 68-76, 2012