

Final Report

Submitted by NIMBioS Sabbatical Visitor: James H. Degnan

Department of Mathematics and Statistics, University of Canterbury,
Christchurch, New Zealand

Sabbatical Dates: January – June 2012

My time at NIMBioS as a sabbatical fellow was very productive, and I am grateful for the support that NIMBioS provided both for my fellowship and for several visitors with whom I collaborated while at NIMBioS. This was particularly helpful since, being located in New Zealand, it is usually more difficult to collaborate with researchers from the US and Europe. Projects were related to problems in modeling gene trees in species trees and implications for inferring species trees (phylogenetic trees).

During my time at NIMBioS, I worked on several manuscripts that had been submitted prior to arriving at NIMBioS and resubmission occurred while at NIMBioS. These manuscripts cite NIMBioS support and grant EF-0832858:

[1] T Stadler, JH Degnan (2012) [A polynomial time algorithm for calculating the probability of a ranked gene tree given a species tree](#). *Algorithms for Molecular Biology* 7:7

[2] B Eldon, JH Degnan. (2012) [Multiple merger gene genealogies in two species: monophyly, paraphyly, and polyphyly for two examples of Lambda coalescents](#). *Theoretical Population Biology* 82:117-1230, doi:10.1016/j.tpb.2012.05.001

[3] Y Yu, JH Degnan, L Nakhleh. (2012) The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genetics* 8:e1002660

[4] JH Degnan, NA Rosenberg, T Stadler. A characterization of the set of species trees that produce anomalous ranked gene trees. (Minor revision decision for IEEE-ACM Transactions in Computational Biology and Bioinformatics, since resubmitted)

Manuscripts that were worked on primarily while at NIMBioS that were submitted either while at NIMBioS or shortly afterward are

[5] M DeGiorgio, JH Degnan. Using minimum coalescence times to infer species trees. Submitted to *Systematic Biology*.

[6] S Zhu, JH Degnan, B Eldon. Hybrid_sim: simulation of multiple merger and Kingman gene genealogies in species networks and species trees. Submitted to *Bioinformatics*.

[7] JH Degnan. Anomalous unrooted gene trees. Submitted to *Systematic Biology*.

[8] ES Allman, JH Degnan, JA Rhodes. Species tree inference by the STAR method, and generalizations. Submitted to *J. Computational Biology*.

Research visits from Mike DeGiorgio (UC Berkeley), John Rhodes (University of Alaska Fairbanks), and Elizabeth Allman (University of Alaska Fairbanks), were especially useful for completing projects. I have an ongoing project with John Rhodes and Elizabeth Allman on identifying species trees from split probabilities, which we worked on during their visit. This work was presented as a poster at the Society for Molecular Biology meeting in June 2012, and will be presented again at the Joint Statistical Meetings on 31 July 2012 in San Diego.

With Laura Kubatko (Ohio State), who visited me at NIMBioS we worked on a paper on deriving distributions on branch lengths of gene trees in species trees. We expect this to be submitted as a paper at some point, but it still needs a little work. Finally, a research visit with Alain Franc (INRA, France) is likely to result in two projects, one on a novel way to compute probabilities of gene trees in species trees when there are multiple “colors” on the gene lineages; and another on species delimitation, perhaps with applications to DNA barcoding. Although I had met Alain before, our collaboration began in earnest at NIMBioS and hopefully will result in more long-term projects. We also discussed the possibility on proposing a working group at NIMBioS on DNA barcoding based on this visit.

While at NIMBioS I also started working on a novel consensus tree algorithm which can reduce a large tree or set of trees with multiply labeled leaves to a uniquely leaf-labeled tree (or set of tied trees) and could be useful for gene duplication or incomplete lineage sorting problems. This is a long-term project, but something I could not have started without the free time available during a sabbatical.

Other activity during my time NIMBioS included giving an invited presentation and software tutorial at Mathematical Biosciences Institute in January, and giving a seminar at NIMBioS on phylogenetic networks, and a seminar at Sergey Gavrillet’s lab on some unsolved problems on tree balance. I very much enjoyed attending lab meetings for Brian O’Meara’s PDG group, Ben Fitzpatrick’s reading group on mutualism, and Sergey Gavrillet’s meetings. Brian O’Meara and I talked about possible future work together on investigating some properties of “anomalous gene trees.”

A final unplanned bit of activity is that I proposed a project for the Google Summer of Code in April, which was accepted, and have been working with a student, Islam Ismailov at the University of Illinois at Chicago, since then on a project to compute probabilities of ranked gene trees given species trees.