

**Working Group on Gene Tree and Species Tree Reconciliation**  
**Final Report, December 2012**

NIMBioS – Summary Report for the three meetings of the working group “Inferring Patterns and Processes of Gene Diversification by Reconciling Gene Trees and Species Trees (Gene Tree Reconciliation)”; Burleigh, Eulenstein, and Liberles (Organizers)

Meeting: December 16-18, 2010. Participants included: Cecilie Ane (University of Wisconsin), Gordon Burleigh (University of Florida), Oliver Eulenstein (Iowa State University), Pawel Górecki (University of Warsaw, Poland), Snehlata Huzurbazar (University of Wyoming), Tony Jhvueng (University of Tennessee), Anke Konrad (University of Wyoming), Jim Leebens-Mack (University of Georgia), David Liberles (University of Wyoming), Liang Liu (Delaware State University), Brian O’Meara (University of Tennessee), and Matt Rasmussen (Massachusetts Institute of Technology)

Meeting: August 10-12, 2011. Participants included: Lars Arvestad (Stockholm University, Sweden), Gordon Burleigh (University of Florida), Juanjuan Chai (University of Tennessee), Oliver Eulenstein (Iowa State University), Pawel Górecki (University of Warsaw, Poland), Stefanie Hartmann (University of Potsdam, Germany), Snehlata Huzurbazar (University of Wyoming), Tony Jhvueng (University of Tennessee), Jim Leebens-Mack (University of Georgia), David Liberles (University of Wyoming), Liang Liu (Delaware State University), Brian O’Meara (University of Tennessee), and Jessica Wu (Massachusetts Institute of Technology)

Meeting: November 26-29, 2012. Participants included: Ruchi Chaudhary (Iowa State University), Nadia El-Mabrouk (University of Montreal, Canada), Oliver Eulenstein (Iowa State University), Pawel Górecki (University of Warsaw, Poland), Tony Jhvueng (Feng-Chia University, Taiwan), Vaishali Katju (University of New Mexico), Laura Kubatko (Ohio State University), David Liberles (University of Wyoming), Brian O’Meara (Stanford University), Noah Rosenberg (Stanford University), and by skype, Lars Arvestad (Stockholm University, Sweden), Jim Leebens-Mack (University of Georgia), and Liang Liu (University of Georgia)

What exactly did you do during the meetings?

The meetings were composed of a combination of lectures and working group break-out sessions. As the meetings had different participants with different background, discussions centered on biological realism and biological data, on discrete algorithms and mathematical methods, and on statistical methods and models associated with phylogenetic tree reconciliation. There were several themes that included modeling of coalescent processes as well as duplication and loss under different evolutionary processes, and algorithm and software development for probabilistic and parsimony based tree reconciliation. Different groups with different trajectories emerged, as described in the individual meeting reports.

- a. Describe new collaborations that were established, if any (with/between whom? Provide names and affiliations). Include a brief description of what the collaborators are hoping to accomplish.

Liang Liu and David Liberles have a grant funded by NSF together, two researchers who met in the first working group.

The following research collaborations were established between researchers that have never worked as a group before.

Liang Liu, Pawel Gorecki, Oliver Eulenstein, and David Liberles are currently working on a grant application together to be sent to NIH in February, 2013.

Nadia El-Mabrouk, Pawel Górecki, and Oliver Eulenstein developed new algorithms for tree reconciliation. Currently the collaborators are in the process of implementing the algorithms and to publish these results.

Noah Rosenberg, Pawel Górecki, and Oliver Eulenstein initiated a new collaboration to design improved distance measures for tree reconciliation.

David Liberles and Nadia El-Mabrouk discussed simulating sequence evolution with explicit genotype-phenotype maps using protein structure and using this to evaluate the patterns and strategies associated with detecting phylogenetic patterns when duplicate genes neofunctionalize using parsimony.

Brian O'Meara, Tony Djwung, Liang Liu, and Snehalata Huzurbazar are

working on a collaboration that involves the use of AIC.

This is all preliminary, but was described in the Meeting 3 report.

b. What are the meeting deliverables, e.g., papers, software, database, including things that are in progress? Be specific, e.g., for papers, list the tentative titles and current co-authors.

1) Papers (published related to gene tree/species tree reconciliation by meeting attendees 2011-2012):

Konrad, A., Teufel, A.I., Grahnen, J.A., and Liberles, D.A. 2011. "Towards a general model for the evolutionary dynamics of gene duplicates." *Genome Biology and Evolution*, 3:1197-1209.

Konrad A., Yarunova, E., Tinta, T., Piskur, J., and Liberles, D.A. 2012. "The global distribution and evolution of deoxyribonucleoside kinases in bacteria." *Gene*, 492:117-120.

Konrad, A., Piskur, J., and Liberles, D.A. 2012. "The evolution of catalytic residues and enzyme mechanism within the bacterial nucleoside phosphorylase superfamily 1." *Gene*, 510:154-161.

André Wehe, J. Gordon Burleigh, and Oliver Eulenstein. Efficient Algorithms for Knowledge-Enhanced Supertree and Supermatrix Phylogenetic Problem. *IEEE/ACM Trans. Comput. Biology Bioinform.* in print 2012 (invited paper from ISBRA 2012).

Pawel Górecki and Oliver Eulenstein. Deep Coalescence Reconciliation with Unrooted Gene Trees: Linear Time Algorithms. 18th Annual International Computing and Combinatorics Conference (COCOON 2012), *Lecture Notes in Computer Science 7434*: 531-542, 2012.

Pawel Górecki and Oliver Eulenstein. A Robinson-Foulds Measure to Compare Unrooted Trees with Rooted Trees. *ISBRA 2012, Lecture Notes in Computer Science 7292*: 115-126.

Pawel Górecki, J. Gordon Burleigh, and Oliver Eulenstein. GTP Supertrees from Unrooted Gene Trees: Linear Time Algorithms for NNI Based Local Searches. *ISBRA 2012, Lecture Notes in Computer Science 7292*: 102-114.

André Wehe, J. Gordon Burleigh, and Oliver Eulenstein. Algorithms for Knowledge-Enhanced Supertrees. *ISBRA 2012, Lecture Notes in Computer Science 7292*: 263-274 (top ranked ISBRA paper on easychair).

Pawel Górecki and Oliver Eulenstein. Algorithms: Simultaneous Error-Correction and Rooting for Gene Tree Reconciliation and the Gene Duplication Problem. *BMC Bioinformatics*, 13 (Suppl 10):S14, 2012. A preliminary version of this paper appeared in ISBRA 2011.

Ruchi Chaudhary, J. Gordon Burleigh and Oliver Eulenstein. Efficient Algorithms for Rapid Error Correction for Gene Tree Reconciliation using Gene Duplications, Gene Duplication and Loss, and Deep Coalescence. *BMC Bioinformatics*, 13 (Suppl 10): S11, 2012. A preliminary version of this paper appeared in ISBRA 2011.

Harris T. Lin, J. Gordon Burleigh and Oliver Eulenstein. Consensus Properties for the Deep Coalescence Problem and Their Application for Scalable Tree Search. *BMC Bioinformatics* 13 (Suppl 10):S12, 2012. A preliminary version of this paper appeared in ISBRA 2011.

Pawel Górecki and Oliver Eulenstein. Maximum likelihood models and algorithms for gene tree evolution with duplications and losses. *APBC 2011; BMC Bioinformatics*, 12(Suppl 1):S15, 2011.

Wen-Chieh Chang, Gordon J Burleigh, David F. Fernández-Baca, and Oliver Eulenstein. An ILP solution for the gene duplication problem. *APBC 2011; BMC Bioinformatics*, 12(Suppl 1):S14, 2011.

Pawel Górecki and Oliver Eulenstein. A linear time algorithm for error-corrected reconciliation of unrooted gene trees. *ISBRA 2011*, LNCS6674: 148-159.

Ruchi Chaudhary, J. Gordon Burleigh and Oliver Eulenstein. Algorithms for Rapid Error Correction for the Gene Duplication Problem. *ISBRA 2011*, ; *Lecture Notes in Computer Science* 6674: 227-239.

Harris T. Lin, J. Gordon Burleigh and Oliver Eulenstein. The Deep Coalescence Consensus Tree Problem is Pareto on Clusters. *ISBRA 2011*, ; *Lecture Notes in Computer Science* 6674: 172-183.

Liu, L., L. Yu, V. Kalavacharla, Z. Liu (2011) A Bayesian model for gene family evolution. *BMC Bioinformatics* 2011, 12:426

## 2) Grants (Funded by meeting attendees):

Liu and Liberles have a grant on gene duplicate modeling for gene tree/species tree reconciliation in a Bayesian framework funded by NSF (DMS1222745) 9/12-8/15.

- c. What are the next steps in your group activities?

From the collaborations that were described, there are several individual efforts that will continue to work towards publication and towards funding from NSF and NIH. There is also a desire to keep a working group of this nature functioning. The next steps will be applications for future meetings internationally, in Europe (Poland) and in Quebec.

e. Are there any students/postdocs at your home institution who are involved in projects related to the group activities? List names and departmental affiliation.

- 1) André Wehe, Ph.D. student (earned his Ph.D. in Spring 2012), Departments of Computer Science and Computer Engineering, Iowa State University
- 2) Anke Konrad, Ph.D. student, Department of Molecular Biology, University of Wyoming
- 3) Ashley Teufel, Ph.D. student, Department of Molecular Biology, University of Wyoming
- 4) Jing Zhao, Ph.D. student, Department of Statistics, University of Georgia
- 5) Joel Sjöstrand, Ph.D. student, Stockholm University, Sweden
- 6) Raja Hashim Ali, Ph.D. student, Royal Institute of Technology, Sweden
- 7) Mehmood Alam Khan, Ph.D. student, Royal Institute of Technology, Sweden
- 8) Manuel Lafond, Ph.D. student, Department of Computer Science, University of Montreal, Canada
- 9) Ethan Jewett, Ph.D. student, Department of Biology, Stanford University
- 10) Wen-Chieh Chang, Ph.D. student (earned his Ph.D. in Fall 2012), Department of Computer Science, Iowa State University
- 11) Harris Lin, Ph.D. student, Department of Computer Science, Iowa State University
- 12) Ruchi Chaudhari, Ph.D. student, Department of Computer Science, Iowa State University
- 13) Nick Pappas, M.S. student (earned his M.S. in Spring 2012), Department of Computer Science, Iowa State University

- 14) Bryan Schrader, B.S. student, Department of Computer Science, Iowa State University
- 15) Matt Rasmussen, Ph.D. student (graduated), Massachusetts Institute of Technology.
- 16) Jessica Wu, Ph.D. student, Massachusetts Institute of Technology.