Phylogenetics of Heterogeneous Samples

(Keynote Talk)

Russell Schwartz

Carnegie Mellon University
Pittsburgh, PA 15213 USA

Phylogenetics, or the inference of evolutionary trees, is one of the oldest and most intensively studied topics in computational biology. Yet it remains a vibrant area of research, in part because advances in our ability to gather data for phylogenetic inference continue to create novel and more challenging variants of the phylogeny problem. In this talk, I will discuss a particular challenge underlying some important phylogenetic problems in the genomic era: reconstructing evolutionary histories from samples of heterogeneous populations, each of which may contain contributions from multiple evolutionary stages or pathways. This problem combines the challenges of identifying common population subgroups from large variation data sets and reconstructing a history of those subgroups. Methods for solving the problem thus end up drawing from a variety of computational techniques, including classic discrete algorithmic approaches to phylogenetics, and machine learning and statistical inference methods for finding robust structure within large, noisy data sets. In this talk, I will present two examples of the problem on very different scales. I will describe the use of phylogenetic methods for inferring evolutionary histories of cell lineages within tumors from genome-wide assays, where cell-to-cell heterogeneity within individual tumors complicates analysis. I will further describe the application of similar concepts for inferring histories of human population groups from genetic variation data, where variability within populations and admixture between them present similar difficulties. Collectively, these problems illustrate some of the special challenges of phylogenetic inference on heterogeneous samples and some of the breadth of techniques needed to address these challenges.

References

- [1] Schwartz, R., Schackney, S.: Applying unmixing to gene expression data for tumor phylogeny inference. *BMC Bioinf.* 11, 42 (2010)
- [2] Tsai, M.-C., Billelloch, G., Ravi, R., Schwartz, R.: A consensus tree approach for reconstructing human evolutionary history and detecting population substructure. In: Borodovsky, M., Gogarten, J.P., Przytycka, T.M., Rajasekaran, S. (eds.) *ISBRA 2010. LNCS (LNBI)*, vol. 6053, pp. 167–178. Springer, Heidelberg (2010)
- [3] Tolliver, D., Tsourakakis, C., Subramanian, A., Shackney, S., Schwartz, R.: Robust unmixing of tumor states in array comparative genomic hybridization data. *Bioinformatics Proc.* 26(12), i106–i114 (2010)
- [4] Tsai, M.-C., Billelloch, G., Ravi, R., Schwartz, R.: A consensus-tree approach for reconstructing human evolutionary history and identifying population substructure. *IEEE/ACM Trans. Comp. Biol. Bioinf.* (2011) (in press)