

1125-92-1348

Ruth Davidson* (redavid2@illinois.edu) and **James H. Degnan**. *When is it surprising that all gene trees are unique? An application of the generalized birthday problem.*

Invariants corresponding to probabilities of sequence patterns in the leaves of phylogenetic trees are a major area of research in algebraic statistics. In 2011, Allman, Degnan, and Rhodes derived invariants for probabilities of gene tree topologies evolving within a species tree under the multi-species coalescent (MSC) model. Yet gene invariants arise from a mathematical object—a probability distribution—instead of from frequencies of gene tree topologies in finite datasets. So it is essential to compare expected gene frequencies to phylogenomic datasets. In 2013 Salichos and Rokas recovered 1,070 topologically distinct gene trees from 23 yeast genomes, a result considered surprising by many. Through a large-scale data simulation and by deriving bounds on probabilities for gene tree discord, we show that this result is not surprising under the MSC model. This is joint work with James Degnan. (Received September 16, 2016)