ALGEBRAIC MODELS OF LARGE SCALE GENOME REARRANGEMENT EVENTS

SANGEETA BHATIA

(Received 14 October 2018; first published online 29 January 2019)

Variations in genome arrangements are an important source of phylogenetic information and have been used to inform phylogenetic studies since the 1930s. The arrangement of genes along chromosomes is roughly analogous to that of beads on a necklace. It is no surprise therefore that in designing algorithms for comparison of genome arrangements, permutations have constituted the construct of choice for representing genomes. However, notwithstanding the widespread use of permutations in genome rearrangement studies, the rich theory of permutation groups has not been exploited to tackle the problems in evolutionary genomics. Our thesis attempts to address this gap in the existing methods. We describe two algebraic models for the evolution of genomes through large scale rearrangement events.

Our first model is an algebraic translation of a universal genome rearrangement operator called double cut and join (DCJ). In going from a graph-theoretic to an algebraic setting, we provide new proofs for some results related to the DCJ and discuss the potential for adapting the DCJ operator to specific biological models.

The second part of our thesis makes use of the well-developed theory of rewriting systems to design a flexible framework for determining weighted rearrangement distances. The proposed framework can be adapted to a number of different rearrangement models and presents an important contribution to the study of weighted distances in genome rearrangement literature. Our work is the first to apply the theory of rewriting systems to a problem in phylogenetics, thereby linking these two separate fields.

Science flourishes best when scientific disciplines are not isolated in silos but, like teens at a party, freely mingle and converse with each other. Contributing to the conversation between algebra and biology is the primary goal of this thesis.

The significance of our work lies in bringing new methods and approaches to biology and raising new mathematical questions motivated from biology.

The work presented in the thesis has been published in [1] and [2].

References

[1]   S. Bhatia, A. Egri-Nagy and A. R. Francis, 'Algebraic double cut and join', *J. Math. Biol.* **71**(5) (2015), 1149–1178.

[2]   S. Bhatia, P. Feijão and A. R. Francis, 'Position and content paradigms in genome rearrangements: the wild and crazy world of permutations in genomics', *Bull. Math. Biol.* **80** (2018), 3227–3246.

SANGEETA BHATIA, Department of Infectious Disease Epidemiology,
Imperial College London, UK
e-mail: sangeetabhatia03@gmail.com