Exact dynamics of duplications and mutations in whole genome sequences

Maxim Koroteev, Jonathan Miller

Physics and Biology Unit

Okinawa Institute of Science and Technology (Graduate University), maxim.koroteev@oist.jp

We introduce a family of models of sequence duplications and point mutations that reproduces steady state duplication length distributions of natural genomes, both eukaryotic and prokaryotic.

For naturally occurring genomes, whole genome alignment and computation of identical sequences yield length distributions of sequence duplications that typically exhibit an algebraic form, with an exponent of around *-3*[1].

It has been demonstrated that a dynamics consisting solely of duplications can generate algebraic duplication length distributions[2]. Exponents and amplitudes of duplication length distributions from natural genomes are also reproduced by a model combining random duplication with point substitution.

Discrete dynamical equations for our models have been derived and solved analytically, demonstrating excellent correspondence with numerical simulations and a mapping onto fragmentation dynamics. Continuum generalizations of these models exhibit an asymptotic regime of duplication and point substitutions where the exponent of the duplications length distribution takes the value *-3* irrespective of details of the source distribution. The regime is observed for lengths of duplicates $m \ll E$, where E is the first moment of the underlying source of duplications.

1. K.Gao, J. Miller (2011) Algebraic distribution of segmental duplication lengths in whole-genome sequence self-alignment, *PLoS One,* **6***(7)***.** *130.1371/journal.pone.0018464*

2. M.V. Koroteev, J. Miller. (2011) Scale-free duplication dynamics: a model for ultraduplication, *PRE,* **84**(061919).