Probabilistic models for genome rearrangements

Asher Moshe, Einat Hazkani-Covo, Oren Avram, Itsik Pe’er, and Tal Pupko

Genome rearrangements

Genome rearrangement events can be divided into three categories:
- Order-changing – events that change the order of the genes i.e. inversion and transposition
- Chromosome-changing – events that change the number of chromosomes i.e. fusion and fission
- Repertoire-changing – events that change the gene content i.e. gene duplication and loss

Genome rearrangement events are a force that shapes the evolution and plays a role in speciation. Gaining a better understanding of these events can allow us to better understand their effect. Finding a correlation between genome rearrangement rates and speciation can help elucidate this effect. Inferring these rates may also give us tools to detect other interesting but more elusive events such as whole-genome-duplication.

Approximate Bayesian computation (ABC)

Even for the simplest models of genome rearrangements, the use of likelihood-based approach is impractical, as one cannot assume that each gene is independent in regard to genome rearrangement events. Thus, the use of the simulation-based ABC approach can help overcome this problem. In this approach we generate a large amount of simulation using parameters drawn from a prior, we compare our data to these simulation using a set of features and choose the closest simulations to assess the parameters of the data.

Simulation results

To test the method we generated 200 datasets with parameters drawn from the prior and aimed to infer them. Here (on the right) we see correlation graphs showing high correlation for two out of three parameters for the model (high $R^2$). Note simulation were based on a tree inferred from 73 E. coli strains, which were also used for testing real data (see tree on bottom right).

Basic model

We start with a basic model aimed to show that this approach is indeed feasible. The basic model assumes a single chromosome genome, a fixed gene repertoire, and allow for order changing events i.e. inversion and transposition.

The model contains three parameters, inversion and transposition rates which measure the rates in which blocks of genomes undergo the relevant events, and an $\alpha$-parameter which determines the block-size distribution (assuming a Zipfian distribution).

Real data results

We aimed to infer the genome rearrangement parameters of bacteria. We used data of 73 E. coli strains obtained from Microbializer$^\star$. Above is the posterior distribution of the inferred parameters. Note that although prior distribution are uniform in the ranges above, the posterior distributions show a clear single peak around a value.

Discussion

- We have shown that the use of the ABC method for inference of genome-rearrangement rates is indeed feasible
- We next aim to use more complex models that better encapsulate the nature of genome rearrangement by adding the events specified above.
- We aim to use the inference of genome rearrangement rates as a stepping stone toward identification of whole genome duplication events

https://microbializer.tau.ac.il


This study was supported by the Edmond J. Safra Center for Bioinformatics at Tel Aviv University.