Reconstruction of ancestral genomes in presence of gene gain and loss

Shuai Jiang[12*],   Pavel Avdeyev[3],   and   Max A. Alekseyev[1]

[1]Computational Biology Institute, George Washington University, Ashburn, VA, USA
[2]Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA
[3]Bioinformatics Institute, Academic University, St. Petersburg, Russia

Genome rearrangements (such as *reversals*, *translocations*, *fusions*, and *fissions*) are evolutionary events that shuffle genomic material without altering it otherwise. One of the key computational problems in comparative genomics is reconstruction of genomes of common ancestors for genomes of living species and the sequence of evolutionary events (*evolutionary history*) between them. In rearrangement-based approaches (particularly employed by the MGRA tool published by the third author in *Genome Res.* 2009), ancestral genomes are reconstructed by minimizing the number of rearrangements along the branches of the phylogenetic tree.

In algorithmic studies of genome rearrangements, genomes are traditionally idealized to have equal gene content. However, in reality, instances of the same gene in different lineages may independently mutate, making it impossible to recognize them as orthologs. While for two genomes the information about missing (*deleted*) genes may not be that beneficial, this situation changes as more genomes come into consideration. In particular, genomes that share some genes are likely to be evolutionarily closer to each other than to a genome where these genes are absent. But most importantly, with the growing number of input genomes, the number of genes shared across all the genomes drops substantially. Therefore, in comparative studies of multiple genomes, it becomes crucial to maintain not only information about orthologous genes but also about deleted/inserted genes across the genomes.

We present a tool called MGRA2 that extends MGRA to support gene insertion and deletion (*indel*) operations. Given a set of genomes and their phylogenetic tree, MGRA2 reconstructs ancestral genomes at the internal nodes of the tree. MGRA2 not only organicly incorporates indels into the rearrangement analysis of multiple genomes but also generalizes algorithms employed by MGRA and make them applicable to "hard" genomic datasets inaccessible for MGRA and similar tools. To evaluate the performance of MGRA2, we conducted two sets of experiments for real and simulated genomes and compared the results of MGRA2 with other existing tools such as GapAdj and PMAG$^+$. These experiments demonstrated supremacy of MGRA2 in all comparisons.

The MGRA2 software is distributed under GNU GPL v2 license. The MGRA2 sources are available at GitHub repository `http://github.com/ablab/mgra/`.

---

*Corresponding author. Email: `jiangs@email.sc.edu`