Adequate Subgraph Approach for Guided Genome Halving Problem

Pavel Avdeyev[1⋆]

The George Washington University, Washington DC, USA

Introduction Genome rearrangements are large-scale evolutionary events that shuffle genomic architectures. Since genome rearrangements are rare, the number of events between two genomes is used in phylogenomic studies to measure the evolutionary distance between them. Such measurement is often based on the *maximum parsimony assumption*, implying that the evolutionary distance can be estimated as the minimum number of rearrangements between genomes.

The maximum parsimony assumption enables addressing the *ancestral genome reconstruction problem*, which asks to reconstruct ancestral genomes from the given extant genomes, by minimizing the total distance between genomes along the branches of the phylogenetic tree. The basic case of this problem with just three given genomes is known as the *genome median problem* (GMP), which asks for a single ancestral genome (*median genome*) at the minimum total distance from the given genomes.

Whole genome duplications (WGDs) represent yet another type of dramatic evolutionary events, which simultaneously duplicate each chromosome of a genome. WGDs are known to have happened in the evolution of plants [1] and yeasts [2]. An analog of the GMP in presence of a WGD is known as the *guided genome halving problem* (GGHP). This problem is posed for input genomes A and B, where all genes in B are present in a single copy (*ordinary genome*), while all genes in A are present in two copies (*duplicated genome*). The GGHP asks for an ordinary ancestral genome R that minimizes the total evolutionary distance between genomes A and $2R$ (genome resulted from the WGD of R) and between genomes B and R.

Methods A major tool for analysis of genome rearrangements is the *breakpoint graph*, which encodes gene adjacencies in different genomes by edges of different colors. A median genome corresponds to a certain *optimal* perfect matching in the breakpoint graph of the given genomes. While the GMP is NP-hard [3], one of the prominent exact and practical solutions to the GMP is based on decomposition of the breakpoint graph into *adequate subgraphs* [4], i.e., induced subgraphs where any optimal matching can be extended to an optimal matching in the whole graph.

Contracted breakpoint graphs represent a generalization of the breakpoint graphs to genomes with duplicated genes. The GGHP can be posed as finding a certain optimal matching in the contracted breakpoint graph of the given

⋆ Corresponding author. Email: avdeyev@gwu.edu

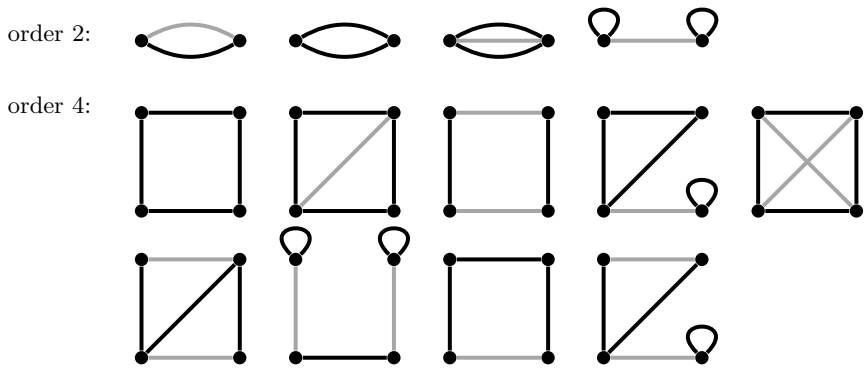

Fig. 1. For the given duplicated genome (*black color*) and ordinary genome (*gray color*), simple adequate subgraphs of order 2 and 4 that may appear in their contracted breakpoint graph.

genomes A and B. Since this problem is NP-hard [3], no efficient algorithm for the GGHP is currently known.

Results We extend the notion of adequate subgraphs for contracted breakpoint graphs and identify all simple adequate subgraphs of order 2 and 4 (Fig. 1). This further enables us to design an efficient divide-and-conquer algorithm for the GGHP. Our algorithm searches for adequate subgraphs in the given contracted breakpoint graph and combines optimal matchings in these subgraphs into an optimal matching (representing a solution to the GGHP) in the whole graph.

Conclusion In the present study we design an exact fast algorithm for the GGHP based on the generalized notion adequate subgraphs in contracted breakpoint graphs. In future research, we plan to further extend the notion of adequate subgraphs to other ancestral genome reconstruction problems with duplicated genes, such as the *guided genome aliquoting problem*.

References

1. Romain Guyot and Beat Keller. Ancestral genome duplication in rice. *Genome*, 47(3):610–614, 2004.
2. Manolis Kellis, Bruce W Birren, and Eric S Lander. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature*, 428(6983):617–624, 2004.
3. Eric Tannier, Chunfang Zheng, and David Sankoff. Multichromosomal median and halving problems under different genomic distances. *BMC Bioinformatics*, 10:120, 2009.
4. Andrew Wei Xu. A fast and exact algorithm for the median of three problem: A graph decomposition approach. *Journal of Computational Biology*, 16(10):1369–1381, 2009.