# Maximally distant genomes under the DCJ operation

*Manda Riehl* University of Wisconsin, Eau Claire

Abstract. We study combinatorial questions relating to the double cut and join model of genome rearrangements (DCJ). The DCJ distance is defined to be the fewest number of DCJ operations required to transform genome A into genome B. Using the distance formula and adjacency graph developed by Bergeron, Mixtacki, and Stoye [1], we present a formula for the number of genomes maximally distant from a given genome $A$ of length $N$, as a function of the number of telomeres and adjacencies in the starting genome. We also present an exponential generating function for this number with a fixed number of telomeres.

Any genome can be represented by a distinct arrangment of vertices, called *adjacencies*, and external vertices, called *telomeres*. The double cut and join model of genome rearrangement [2] is a general model acting on two vertices in the genome and includes the mutation models of inversions, interchanges, translocations, fusions, fissions, circularizations, linearizations, excisions, and integrations. To find the DCJ distance between genome $A$ and genome $B$, one constructs a bipartite adjacency graph with vertices corresponding to the sets of adjacencies and telomeres of the two genomes. Two vertices are connected with an edge for every head or tail they share [1].

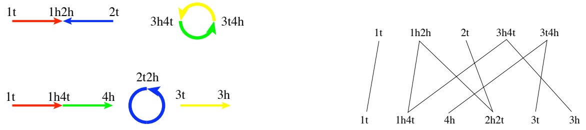

Figure 1: Two genomes and their associated adjacency graph.

Given that $A$ and $B$ are defined on the same N genes, the DCJ distance is given by

$$d_{DCJ}(A,B) = N - (C + I/2),$$

where $C$ is the number of cycles and $I$ is the number of odd paths in the adjacency graph of A and B [1]. Thus in the example, the distance is 2.

To calculate the number of maximally distant genomes, we first note that the maximum distance is $N$, and we therefore need to count all possible ways to construct adjacency graphs with no cycle and no odd paths, so that the $C + I/2$ term is zero. Given that the initial genome $A$ has $2m$ telomeres and $n$ adjacencies, the number of genomes maximally distant from $A$ is

$$G_{max}(m,n) = \frac{(2m-1)!}{2^{m-1}(m-1)!} \sum_{k=0}^{n} \binom{n+m-1}{k} \binom{n}{k} 2^k k!.$$

The exponential generating function for the sequence with fixed value of $m$ is given by

$$f_m(x) = (\frac{(2m-1)!}{2^{m-1}(m-1)!} \frac{e^{\frac{x}{1-2x}}}{(1-2x)^m}.$$

We are also investigating properties of the maximum distance graph $M$, defined with all possible genomes of length $N$ as vertices, and an edge between two vertices $a$ and $b$

if $a$ is maximally distant from $b$. We are examining this graph in both the signed and unsigned cases of DCJ. We are also working on modifying the method in [1] to create an analogue of the adjacency graph to obtain a formula for distance in the unsigned case.

*This is joint work with Jacqueline Christy, Joshua McHugh and Noah Williams.*

## References

[1] Bergeron, A., Mixtacki, J. and Stoye, J. (2006) A Unifying View of Genome Rearrangements. WABI 2006. pp. 163-173.

[2] Yancopoulos, S., Attie, O. and Friedberg, R. 2005. Efficient sorting of genomic permutations by translocation, inversion and block interchange. Bioinformatics 21, 3340 3346.

[3] Fertin, G., Labarre, A., Rusu, I., Tannier, E., and Vialette, S. (2009) Combinatorics of Genome Rearrangements. The MIT Press.