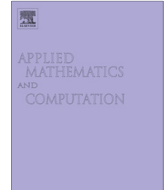




ELSEVIER

Contents lists available at ScienceDirect

## Applied Mathematics and Computation

journal homepage: [www.elsevier.com/locate/amc](http://www.elsevier.com/locate/amc)

## The evolution of the random reversal graph

Christian M. Reidys<sup>a,\*</sup>, Emma Yu Jin<sup>b</sup><sup>a</sup> Dept. of Mathematics and Computer Science, University of Southern Denmark, Denmark<sup>b</sup> Department of Computer Science, University of Kaiserslautern, Germany

## ARTICLE INFO

## Keywords:

Reversal

Permutation

Giant component

Threshold probability

## ABSTRACT

Genomes are typically represented as signed permutations, where each integer corresponds to a single gene and its sign indicates the orientation of the gene. In this paper we study genome rearrangements via reversals (Sankoff and Blanchette, 1999) [20], flipping entire segments of genes and changing their orientations. This abstraction leads to the notion of the reversal graph in which two signed permutations are neighbors if they differ by one reversal. We identify the evolution of multiple genomes within the random reversal graph, i.e. the probability space consisting of subgraphs over signed permutations, obtained by selecting edges (reversals) with independent probability  $\lambda_n$ , which is an important framework for the analysis of genome rearrangements in the sense of relating the reversal rate  $\lambda_n$  with the global evolution of genomes. Our main result shows that the structure of the random reversal graph changes dramatically at  $\lambda_n = 1/\binom{n+1}{2}$ . For  $\lambda_n = (1 - \epsilon)/\binom{n+1}{2}$ , the random graph consists of components of size at most  $O(n \ln n)$  a.s. and for  $(1 + \epsilon)/\binom{n+1}{2}$ , there emerges a unique largest component of size  $\sim \varphi(\epsilon) \cdot 2^n \cdot n!$  a.s., where  $\varphi(\epsilon)$  is the survival probability of a certain branching process. This “giant” component is furthermore dense in the reversal graph.

© 2013 Elsevier Inc. All rights reserved.

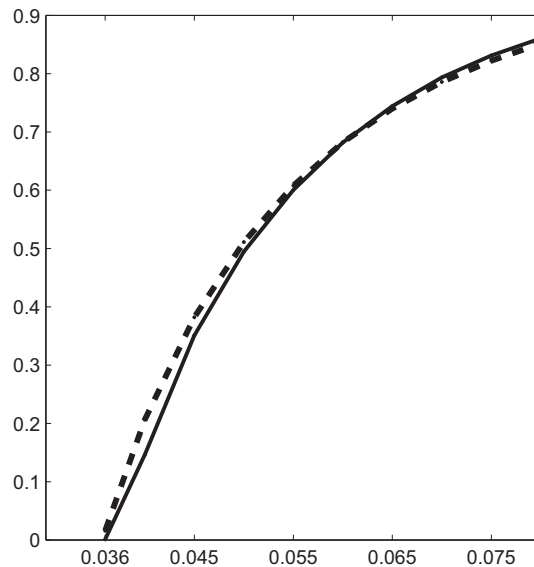
## 1. Introduction and background

Seventy years ago, Dobzhansky and Sturtevant [7] initiated the study of genome rearrangements. Due to recent progress in large-scale sequencing and comparative mapping, genome rearrangements have become increasingly popular [18,17,16,22]. Sequence data show that the genomes of different species, most of which being very similar on the letter-by-letter DNA code basis, differ mainly in gene order and orientation, which motivated Sankoff et al. [21] to model the evolution of genomes by representing order and orientation of genomes as signed permutations. In fact, comparing the two genomes provides an evolutionary history of the two species and traces out the diverging sequences of DNA.

For uni-chromosome genomes, we have [4,23] different rearrangement mechanisms: **(a)** reversals (inversions), of any number of consecutive terms in the ordered set, which, in case of signed permutations also reverses the polarity of each term within the scope of the inversion [11,4,20], **(b1)** transpositions and **(b2)** inverted transpositions, of any number of consecutive terms from their position to a new position between any other pair of consecutive genes in conserved (or inverse) order. We next display the actions **(a)**, **(b1)** and **(b2)** on the signed permutation  $(+2, +6, -3, -4, +1, -5)$ .

\* Corresponding author.

E-mail addresses: [duck@imada.sdu.dk](mailto:duck@imada.sdu.dk), [duck@santafe.edu](mailto:duck@santafe.edu) (C.M. Reidys), [jin@cs.uni-kl.de](mailto:jin@cs.uni-kl.de) (E. Yu Jin).



**Fig. 1.** The evolution of the largest component in the random reversal graph: we display the fraction of signed permutations contained in the largest component of the random reversal graph (y-axis) as a function of the probability of reversals,  $\lambda_n$ , (x-axis) for  $n = 7$ , where the dashed line represents theoretical results from [Theorem 1](#) and the solid line represents the results from computer experiments.

- (a) :  $(+2, +6, \boxed{-3, -4, +1}, -5) \rightarrow (+2, +6, -1, +4, +3, -5)$ ,
- (b1) :  $(+2, +6, \boxed{-3, -4, +1}, -5) \rightarrow (+2, -3, -4, +1, +6, -5)$ ,
- (b2) :  $(+2, +6, \boxed{-3, -4, +1}, -5) \rightarrow (+2, -1, +4, +3, +6, -5)$ .

In multi-chromosome genomes we find in addition reciprocal translocations. Here we refer the readers to the book [\[10\]](#) by Fertin et al. for more details concerning the multi-chromosome genome model.

In this paper we restrict our analysis to reversals as edit operation. Our framework is not limited to the reversals but remains valid if we have other genome rearrangements mechanisms since we study monotone graph properties.<sup>1</sup> Studying monotone properties allows us to derive conclusions that remain valid even if additional rearrangement operations are taken into account. Reversals change the order of the genes in a genome, and also the direction of transcription. For instance, the reversal  $\rho_{3,5}$  maps the signed permutation  $(+1, +4, \boxed{+2, +5, +3})$  to  $(+1, +4, \boxed{-3, -5, -2})$ . Reversals and signed permutations constitute the reversal graph in which two signed permutations are adjacent if they differ by a single reversal. The reversal graph is regular of degree  $\binom{n+1}{2}$ , has diameter  $(n+1)$  [\[15\]](#) and is connected, since any signed permutation can be generated via a sequence of reversals.

We follow the random graph paradigm and study random subgraphs of the reversal graph, obtained by selecting reversals with probability  $\lambda_n$ , having all signed permutations as vertices. In analogy to Erdős and Renyi’s result [\[8,9\]](#), we also observe a phase transition at  $\lambda_n = 1 / \binom{n+1}{2}$ . Selecting reversals with probability  $(1 - \epsilon) / \binom{n+1}{2}$  the random graph consists of small components. Increasing this probability to  $(1 + \epsilon) / \binom{n+1}{2}$ , these components merge into a unique giant of size precisely  $\sim \varphi(\epsilon) \cdot 2^n \cdot n!$ , where  $\varphi(\epsilon)$  denotes the survival probability of a certain branching process, see [Fig. 1](#). The main difficulty in localizing this phase transition consists in simulating a certain branching process within the Cayley graph of signed permutation, which is generated by the complicated generator set of reversals.

The paper is organized as follows: we first discuss some basic facts. Then we present a series of auxiliary lemmas and finally prove the main result.

## 2. Some basic facts

### 2.1. Random graphs

We consider the Cayley graph  $\Gamma(B_n, R_n)$ , having vertex set  $B_n$  and edges  $\{v, v'\}$ , where  $v^{-1}v' \in R_n$ . Let  $B_n$  denote the set of signed permutation of length  $n$  and  $R_n$  be the set of reversals  $\rho_{i,j}$ , where  $1 \leq i \leq j \leq n$ . For  $v, v' \in B_n$ , let  $d(v, v')$  be the minimal

<sup>1</sup> For instance “connectivity” is a monotone property: a graph remains connected when additional edges are inserted.

Download English Version:

<https://daneshyari.com/en/article/6421735>

Download Persian Version:

<https://daneshyari.com/article/6421735>

[Daneshyari.com](https://daneshyari.com)