BIOINFORMATICS

Fractionation, rearrangement and subgenome dominance
David Sankoff and Chunfang Zheng
Department of Mathematics and Statistics, University of Ottawa, Ottawa, Canada K1N 6N5

ABSTRACT
Motivation: Fractionation is arguably the greatest cause of gene order disruption following whole genome duplication, causing severe biases in chromosome rearrangement-based estimates of evolutionary divergence. Results: We show how to correct for this bias almost entirely by means of a ‘consolidation’ algorithm for detecting and suitably transforming identifiable regions of fractionation. We characterize the process of fractionation and the performance of the algorithm through realistic simulations. We apply our method to a number of core eudicot genomes, and we by studying the fractionation regions detected, are able to address topical issues in polyploid evolution. Availability and implementation: Code for the consolidation algorithm, and sample data, is available at: http://137.122.149.195/Software/Fractionation/Fractionation.html Contact: sankoff@uottawa.ca

1 INTRODUCTION
Fractionation (Eckard, 2001; Freeling, 2001) of excess non-coding DNA (Eckard, 2001; Freeling, 2001) involves the constant excision of excess non-coding DNA between two genomes, by and studying the fractionation regions detected, are able to address topical issues in polyploid evolution.

Availability and implementation: Code for the consolidation algorithm, and sample data, is available at: http://137.122.149.195/Software/Fractionation/Fractionation.html Contact: sankoff@uottawa.ca

2 METHODS
2.1 Methodological preamble
In this article, we focus on the evolutionary divergence of a genome after WGD compared with a related diploid. The mathematics of genome comparison at the level of gene order or conserved synteny blocks is well worked out (reviewed in [Eckard et al., 2003]) when the two genomes have identical complements of single-copy genes only, but the problem is more difficult when they have different gene complements or genes that occur two or more times in a genome. There are many approaches to the comparison of genomes with duplicates, as reviewed in [Eckard et al., 2003]. These range from ‘guided genome halving’ (Zheng et al., 2012), where single-copy genes may simply be excluded from the analysis, to ‘duplication-loss’ models.
where the ordering of loss and duplication events is the focus [Marron et al. 2004] to the 'exemplar method' where it is the duplicate genes that are excluded or rather reduced to single-copy status in an optimal way before the comparison and everything in between (Sankoff, 1999). Although all of these are well motivated in particular contexts, they all invoke objective functions that are largely irrelevant to fractionation.

The central datum in genome comparison algorithms is generally the 'gene adjacency'. Ideally, the genes are oriented, i.e. their 5'- and 3'-ends are distinguished, and when we speak of two genes being adjacent, we mean that the 3'-end of one is adjacent to the 5'-end of the next one, if they are on the same DNA strand or that their two 5'-ends (or two 3'-ends) are adjacent if they are on complementary strands. An adjacency is thus defined by two specific gene ends and not by two genes.

Decomposing a genome into its adjacencies, to take advantage of genome rearrangement algorithms, loses no information in the case of diploid genomes if all genes are present in only one copy, since then the genome is completely determined by its adjacencies. This equivalence between genomes and sets of adjacencies may break down when duplicate genes occur in large numbers. Nevertheless, in this article, we will use the adjacencies directly to bypass the arbitrary use of objective functions and other assumptions inherent in most comparison algorithms. In particular, it will be appropriate in the WGD-fractionation context to simply compare the sets of adjacencies that occur at least once in a genome as the relevant representation of that genome.

2.2 Excess adjacencies as a measure of rearrangement

As a doubled genome evolves through genome rearrangement and/or fractionation, there is a direct effect on the inventory of oriented gene adjacencies. Initially, a new tetraploid and an unduplicated sister genome will have the exact same adjacencies, although these will have multiplicity 2 in the tetraploid. Over time, rearrangement will change the adjacencies in the diploid, without appreciably changing the total number of adjacencies. (This number can change only as a result of the rare events of chromosomal fission or fusion, whereas it does not change after the more frequent events of inversion or reciprocal translocation.) At the same time, in the tetraploid, both rearrangement and fractionation will add new adjacencies without necessarily losing the original ones, since an evolutionary event need only affect one gene copy or one homeolog at a time, leaving the other intact. Thus, the total number of different adjacencies in the diploid and the ancient tetraploid \( T \), in excess of the number in the diploid \( D \) alone, measures the degree of evolutionary divergence.

We carried out simulations of the schema in Figure 2 to quantify the effect of rearrangement on the number of adjacencies, involving random chromosomal inversions and reciprocal translocations in the proportions 10:1. The simulations modeled a diploid with 24,000 genes randomly divided among 20 chromosomes. The two breakpoint points of each rearrangement were chosen at random on the same chromosome for inversions, and on two randomly chosen chromosomes, for translocations. A separate series of simulations modeled the inversions as short (random inversion length drawn from a geometric distribution, with mean length of five genes) and unrestricted, with two random end points, in the proportion 20:1. The differences in the results between the two kinds of model were negligible in all analyses in this article, including the one in Section 3.2 latter.

All simulations were run 20 times, and their average values were plotted. Because of the large number of genes in the simulations, variability among the replicates would be almost imperceptible on the scale of the diagrams to be presented here.

Figure 3 shows the tight linear relationship between the number of rearrangements simulated \( r \), sampled at \( s = 600, 1200, \ldots \), \( x \) to form the ancient tetraploid \( T \) and \( x \) to form the modern diploid \( D \) and the increase in the number of adjacencies, when the number of deletions \( d \) is zero.

Although we can invent examples where the number of adjacencies and the number of rearrangements are less predictive of each other, these would be encountered with negligible frequency in real data or in simulations. In the ensuing sections, therefore, we will use the number of excess adjacencies as a proxy for the amount of rearrangement.

2.3 The effects of fractionation on adjacencies

We carried out extensive simulation experiments to assess how the process of fractionation impacts the apparent amount of rearrangement. The genes deleted to form the ancient tetraploid were chosen entirely at random from the entire genome, under the constraint that once a gene was deleted, its duplicate could not be deleted in a later step. The solid lines in Figure 3 show the effect of increasing amounts of fractionation on the apparent amount of rearrangement, a gross distortion.

3 ALGORITHM

We developed the consolidation algorithm to assess and correct for this distortion. This focuses on detecting and accounting for two homeologous regions of single-copy genes in the ancient tetraploid which contain no genes in common (as the genes concerned are single copy) but whose combined [or consolidated, Langham et al. 2004] gene content is exactly the same as some contiguous region in the diploid. We call maximal regions like this Class 1 regions, in contrast to Class 2 regions, where there is only one region in the tetraploid which contains only, and exactly, the same single-copy genes as a region in the ancient tetraploid, but where there is sometimes clear adjacency evidence of where the duplicate copies
D. Sankoff and C. Zheng

3.1 Consolidation algorithm

Input: Triple \((L, D, T)\), where

- \(L\) is a set of genes.
- \(D\) is a diploid genome, i.e. \(L\) is partitioned among a number of chromosomes; the genes on a chromosome are ordered.
- \(T\) is an ‘ancient’ tetraploid, i.e. each gene of \(L\) occurs in either one or two copies; the genes are partitioned and ordered on a number of chromosomes, with no constraint on positions of two copies of the same gene.

- For each gene \(g \in L\), pointers \([C_1(g), \ldots, C_3(g)]\) indicate its chromosomal membership and its position in that chromosome’s order. The subscript ‘1’ refers to \(D\) and ‘2’ and ‘3’ refer to the occurrences of \(g\) in \(T\), where \(C_3(g) = \emptyset\) if \(g\) occurs only once in \(T\).

Output: Triple \((L', D', T')\), where

- \(D'\) is a diploid genome with gene set \(L'\), and \(T'\) is a ‘pure’ tetraploid (all genes occur exactly twice) on \(L'\).
- \(L' = \cup_{i=1}^3 L_i\). \(L_i\) is the set of genes with two copies in \(T\), and \(L'\) is a set of ‘virtual’ genes, each one representing a triple of intervals of genes (in \(L_i\)), \((I_1, I_2, I_3)\), the first, \(I_1\), on a chromosome in \(D\), the second and third on chromosomes in \(T\), such that \(I_2 \approx \emptyset\), \(I_2 \bigcup I_3 = I_1\). The \(I_1\) from different virtual genes are disjoint, and their union is \(L_1 \bigcup L_2\), the entire set of single-copy genes in \(T\). Moreover, the elements of \(L'\) are jointly maximal in that the union of no two of them can be a virtual gene.

Detect fractionation intervals

While there are still single-copy genes in \(T\) that have not been placed intervals, create new element of \(L'_V\) as follows:

- Initialize empty intervals \(I_1, I_2\), and \(I_3\). Find the first single-copy (in \(T\)) gene in \(D\), add it to \(I_1\) and \(I_2\).
- While the next gene in \(D\) is a single-copy (in \(T\)) gene on the same chromosome as the other genes in \(I_1\) and it is on the same chromosome as \(I_2\) or \(I_2\), add it to \(I_1\) and to \(I_2\) or \(I_3\). If it is not on the same chromosome as \(I_2\) and \(I_3\) is empty until the present time, the gene determines on which chromosome the interval \(I_3\) is to be located.
- While the genes in \(I_2\) (or \(I_3\)) are not contiguous, remove the last gene assigned to \(I_1\) and \(I_2\) or \(I_3\).
- If \(I_3 = \emptyset\), the virtual gene \(g\) representing the triple \((I_1, I_2, I_3)\) is added to \(L'_V\), and it replaces the genes in \(I_1\) on \(D\), and the genes in \(I_2\) and \(I_3\) on \(T\).

Else (if \(I_3 \neq \emptyset\)),

- place \(I_1\)

Consider the two genes (actually gene ends) \(a\) and \(c\) adjacent on the left and right of \(I_1\) in \(D\) and the two genes \(b\) and \(d\)
adjacent on the left and right of $I_2$ in $T$. Locate all further copies of $a, b, c$ and $d$ in $T$, and insert $I_3$, or the virtual gene representing $I_3$ into $T$ adjacent to one, or between two, of these gene ends, in such a way as to minimize the number of new adjacencies thus created. This involves trying all eight combinations or orientations for the virtual genes in its three locations

• Now the virtual gene $g$ representing the triple $(I_1, I_2, I_3)$ can be added to $L_g$, replacing the genes in $I_1$ on $D$, and the genes in $I_2$ and $I_3$ on $T$.

• There are minor details, which we shall omit, on handling ‘telomeric’, i.e. at the ends of chromosomes, genes.

Note that place $I_3$ contains an optimization step. The rest of the algorithm is deterministic, and it may very well not call place $I_3$ at all. Nevertheless, the consolidation algorithm may be reformulated as a combinatorial optimization problem, where the objective is to minimize the number of different adjacencies by ‘placing’ all the empty $I_g$ intervals at the same time.

The consolidation algorithm, as presented, has quadratic worst case running time, in the number of genes, but in practice, on large genomes both real and simulated, it takes seconds or minutes. Other consolidation algorithms, with linear worst case running time, have been devised by J. Kováč (personal communication) and K. Jahn (personal communication).

The consolidation algorithm treats the fractionation intervals as identical units, one in the diploid and two in the tetraploid. In this way, it accounts not only for rearrangements which includes a whole interval in its scope but also for rearrangements which disrupt an interval, in that a fractionation involving such an interval will generally be automatically counted as two intervals, resulting in two virtual units instead of one. What the consolidation algorithm does not account for, however, are rearrangements occurring completely within one of the fractionation intervals.

To correct for this, within each fractionation region, we first consider all the adjacencies in $I_1$ and compare each $I_2$ and, if necessary, $I_3$ with $I_1$. For the comparison, with $I_2$, for example we delete the genes in $I_1$ not in $I_2$. We count only those $a_g$ adjacencies in $I_2$ not in the reduced $I_1$. We add the number of these ‘rearranged’ adjacencies to the set of adjacencies produced by the consolidation algorithm.

4 RESULTS AND DISCUSSION

4.1 Applying the algorithm to simulated genomes

As can be seen in Figure 3 for moderate rearrangement distances, and for a wide range of fractionation rates, our procedure almost completely wipes out the distortion caused by the fractionation.

4.2 The study of fractionation regions

The consolidation algorithm was motivated as a way of correcting estimates of genomic divergence, but an important byproduct is its systematic identification of fractionation regions. This will be the focus of the remainder of this article, including its empirical aspect. First, we will explore fractionation regions in the context of our simulations and in the next section, discuss how to analyze this process in the core eudicots.

4.3 Why can not we just ignore fractionation regions?

The consolidation algorithm adapts genomic rearrangement assessment to the case of mixed duplicate and single-copy genes, through the device of virtual genes representing fractionation regions. We may ask, what if we just delete all single-copy genes and compare the diploid sister genome with whatever duplicates remain in the ancient tetraploid? This question is motivated by the original approach (since abandoned) to the ‘guided genome halving’ problem [Zheng et al. 2007].

The results, shown in Figure 4, show that this alternative approach is not a satisfactory way of handling single-copy genes, considering its substantial underestimate of the amount of rearrangement, which is especially severe when the fractionation process is far advanced.
When only a few of the homeologous gene pairs have been reduced to single copy, fractionation regions will tend to be small and dispersed across the genome. Most rearrangements will not change the contents of any fractionation region. When most of the gene pairs have been reduced to single copy, fractionation regions will tend to be longer, especially if there have been few rearrangements. Then each new rearrangement is likely to disrupt a fractionation region, usually breaking it into two. These tendencies can be seen in Figure 8, where the fractionation regions for a genome with 15,000 out of 24,000 gene pairs reduced to single copies, are little affected by rearrangements. Where 21,000 of the pairs are thus reduced, the size of the fractionation regions is much more sensitive to rearrangements.

### 4.4 Rearrangements and fractionation region size

When only a few of the homeologous gene pairs have been reduced to single copy, fractionation regions will tend to be small and dispersed across the genome. Most rearrangements will not change the contents of any fractionation region. When most of the gene pairs have been reduced to single copy, fractionation regions will tend to be longer, especially if there have been few rearrangements. Then each new rearrangement is likely to disrupt a fractionation region, usually breaking it into two. These tendencies can be seen in Figure 8, where the fractionation regions for a genome with 15,000 out of 24,000 gene pairs reduced to single copies, are little affected by rearrangements. Where 21,000 of the pairs are thus reduced, the size of the fractionation regions is much more sensitive to rearrangements.

### 4.5 Fractionation in the poplar genome

In applying our method to the genome of poplar (Populus trichocarpa) (Tuskan et al. 2004), an ancient tetraploid, compared with diploid sister genomes, grapevine (Vitis vinifera) (Juillot et al. 2002) and cacao (Theobroma cacao) (Avendaño et al. 2008), we discovered that the large majority, 70–90%, of the apparent rearrangements are actually attributable to fractionation. We also carried out matching simulations, with the same number of duplicate and single-copy genes, produced by random paralog deletion and a number of random rearrangements, mostly inversions, to produce the same number of excess adjacencies. Examining the consolidated regions detected by our algorithm, there are a number of regions much longer than those in the simulations (Fig. 6), suggesting a non-independence of deletion events affecting neighboring genes and a clear tendency for genes to be deleted in one of the two homeologs, as would be predicted by the recent theory of subgenome dominance (Schnable et al. 2008). "Genes are disproportionately lost from one parental subgenome, the subgenome that is less expressed in the parental subgenome, the subgenome that is less expressed in the"

---

1 Data were drawn from CoGe (Evans and Fisler, 2008; Evans et al. 2008) and processed using the SynMap procedure, with a 2:1 syntenic depth setting to account for the ancient tetraploidy of poplar, and minimum syntenic block length 2 (to avoid the excessive noise with minimum size 1 but capturing more orthologies than with higher minimum sizes).

---

**Fig. 9.** Size of Populus–Vitis fractionation regions compared with simulations matched for rearrangement and deletion rates. In Populus–Vitis, the single-copy genes are concentrated in fewer, larger regions.

**4.6 Effect of deletion event run size**

Motivated by the discrepancy between the poplar genome fractionation region sizes and the results of the simulations, we ran new simulations where instead of choosing duplicate genes to delete one at a time at random, we deleted one or more contiguous genes at a time, starting at a random position in the genome, according to a geometric distribution with mean $\mu$. To model the quantitative bias in the fractionation process, successive genes were deleted from the same homeolog with ‘continuation’ probability $p \geq \frac{1}{2}$ and from the opposing homeolog with probability $1 - p$.

To infer the best values of the parameters $\mu$, $p$, and the $r$ number of rearrangements to account for the poplar–Vitis and poplar–cacao comparison, we simulated the evolutionary process with a fixed number of genes in the diploid and tetraploid and a fixed number of single-copy genes in the tetraploid. We searched in a three-dimensional grid where $\mu \in [1, 2]$, $p \in [0.5, 1.0]$ and $r$ takes on a wide range of values depending on $\mu$, for the values of the parameters that could produce the same number of adjacencies before and after the consolidation algorithm and the same number of fractionation regions as the real data. The results for:

- The real genomes,
- A model where one gene is deleted at a time ($\mu = 1, p = 1$),
- The best model with geometrically distributed deletion lengths,
- A model with $\mu = 2$,

are compared in Table 1.

The results from the two diploids are similar with respect to $\mu$, namely 1.7 and 1.8, but the $p$ are rather different: 0.5 for Vitis and 0.7 for cacao. The most important fact is that in taking into account the sizes of the fractionation regions in the simulations requires us to increase the number of rearrangements substantially, by 50% in the case of Vitis and by over 100% for the cacao comparison.

As we built no provision into the simulations for fractionation to affect the two homeologs differently there is no reason to expect the kind of fractionation bias we observed in poplar to show up in the simulations. Indeed, Figure 11 shows that simulations with...
Inference (in boldface) of deletion event size parameter $\mu$, continuation parameter $p$ and number of rearrangements $r$, necessary to model poplar–Vitis and poplar–cacao comparisons. $\text{Adj}_1$: adjacencies before consolidation, $\text{Adj}_2$: adjacencies after consolidation, $f$: number of fractionation regions, $i$: average size of fractionation regions.

| Library | Number of genes, Single-copy, Genome size | Real genome data | $\mu$ | $p$ | $r$ | $f$ | $i$ |
|---------|-----------------------------------------|-----------------|-------|----|----|-----|-----|
| Populus | 18,309, 5,287 single-copy, Vitis: 11,798 genes | 16,877, 10,988 | 680  | 1  | 16,866 | 3005 | 1.76 |
|         |                                         | 10,970          |       | 1  | 3005 | 1.76 |
|         |                                         | 16,866          | 1.7   | 0.5 | 16,900 | 2355 | 2.23 |
|         |                                         | 10,957          | 2      | 0.5 | 16,888 | 2184 | 2.42 |
| Populus | 18,221 genes, S.557 single-copy, Cacao: 11,889 genes | 16,357, 10,279 | 320  | 1  | 16,618 | 3014 | 1.84 |
|         |                                         | 10,074          | 1.8    | 0.7 | 16,433 | 2297 | 2.42 |
|         |                                         | 10,074          | 2      | 0.7 | 16,371 | 2195 | 2.53 |

Fig. 10. Discrepancy in two Populus intervals in the number of genes, compared with random simulations. In Populus regions, the single-copy genes are concentrated in one of the two homeologous intervals. Jagged nature of graphs due to statistical fluctuation but to measurement of discrepancy from an ‘even’ split, which is necessarily calculated slightly differently for even and odd totals of genes in the fractionation intervals. Black dashed line represents simulations with deletion event run size geometrically distributed with mean 1.7.

or without deletion event run size variability do not differ. This is further confirmation of the reality of the observed effect that it is not an artifact of duplicate deletion regime nor of the consolidation algorithm.

4.7 Single-copy region syntenies with cacao and Vitis

Since the cacao and Vitis genomes have a long history of independent rearrangements, we can expect that fractionation regions induced by the consolidation algorithm by one of them in poplar will generally differ from the regions determined by the other. Nevertheless, we found many regions in the Vitis–poplar comparison that overlapped substantially in poplar with regions in the cacao–poplar comparison. In particular, we found eight regions that contained more than 15 genes in both comparisons; six of these pairs of comparisons had more than 15 genes in common; of most interest is that the common genes were 100% on only one homeolog in seven of the eight cases. This strongly suggests that there is a functional reason for these large sets of contiguous genes to be reduced to single-copy and for these copies to be retained on a single homeolog.

4.8 Subgenome dominance

Although we have detected and measured substantial bias in the fractionation of ancient core eudicot tetraploids, without a gene-by-gene reconstruction of the ancestral chromosomes, we cannot as yet be sure that the bias is systematically in favor of one or other of the diploids contributing to the original polyploid. This dominance of one of the subgenomes has been demonstrated in the case of maize (Schnable et al. 2011), where it was also correlated to gene expression, patterns which persisted despite a number of genome rearrangement events disrupting the original synteny relations among genes.

5 DISCUSSION

An analytically advantageous feature of our analysis is that it partitions the rearrangements that have affected a tetraploid into those that have operated within a fractionation interval and those that have left these intervals intact, either because the intervals are outside the scope of the rearrangement or the interval is affected as a whole, without any effect internally. Of course, in the the history of the tetraploid there will have been some rearrangements that have involved elements both within and outside fractionation intervals, but these are rendered irrelevant and invisible to our analysis because their effect is simply to fragment the intervals into two or more new intervals and hence perpetuate the within-outside partitioning of the rearrangements.

One of the problems remaining with this work is the large number of genes that do not appear in any synteny block in the diploid-tetraploid analysis of pairs of core eudicot genomes. Despite reducing the minimum block length as low as 2 in SynMap. Some of this is undoubtedly due to genes absent from one of the genomes, but much of it will be due to movement of genes out of their erstwhile homoeologous contexts, by various processes (Wicker et al. 2010; Woodhouse et al. 2011).

In our simulations of geometric deletion events, we ‘marked’ some duplicate genes for deletion before carrying out the rearrangement step, and only actually deleted them afterward. This was for purposes of comparability of the results with the previous simulations where deletions followed rearrangements. Our goal in this marking was to allow a degree of switching between homoeologous chromosomes during the deletion event, and no connection was assumed between two different events. However, this computational device could also be used to model subgenome dominance, by using two different $p$-values one for the dominant subgenome and a smaller value for the other subgenome. Marking becomes analogous to (lack of) methylation or to other epigenetic regulatory mechanism that can persist despite extensive genome rearrangement.

It will be of interest to explore the extensions of our treatment of fractionation to genomes descending from more complex polyploidization events than WGD. The plant genomes studied here, as well as all the other core eudicots, descend from an ancestral...
D. Sankoff and C. Zheng

hexaploidization event [Jaillon et al. 2007]. Extensive fractionation has left only a hundred or so triplicates intact and about a thousand duplicates, in Vitis and in cacao [Zheng et al. 2008]. Many core eudicot angiosperm genomes have been sequenced but as of yet no sequence of a dicot that diverged before the hexaploidization event has been published, so that extensions of our procedure could be tested.

ACKNOWLEDGEMENTS

The authors thank Jakub Kováˇc for a number of valuable comments and suggestions during the work reported in this study and Eric Lyons for guidance to current trends in subgenome research.

Conflict of Interest: none declared.

REFERENCES

Argout, X. et al. (2001) The genome of Ziziphus jujuba. Nat. Genet., 26, 101–108. Dietrich, F. S. et al. (2004) The Atriplex geoffroei genome as a tool for mapping the ancient Saccharomyces cerevisiae genome. Science, 304, 304–307.

El-Mabrouk, N. and Sankoff, D. (2012) Analysis of gene order evolution beyond single-copy genes. In Anisimova, M. (ed). Combinatorics of Genome Rearrangements. New York: Humana Press, pp. 397–429.

Fertin, G. et al. (1974) Combinatorics of Genome Rearrangements. Cambridge, MA: MIT Press.

Freeling, M. et al. (2012) Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. Curr. Opin. Plant Biol., 15, 131–139.

Jaillon, O. et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature, 449, 463–467.

Kellis, M. et al. (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature, 428, 637–642.

Langham, R. J. et al. (2004) A sense of self: the role of DNA sequence elimination in allopolyploidization. Plant Cell, 16, 1699–1704.

Marron, M. et al. (2004) Analysis of gene order evolution beyond single-copy genes. In Anisimova, M. (ed). Combinatorics of Genome Rearrangements. New York: Humana Press, pp. 397–429.

Dietrich, F. S. et al. (2004) The Atriplex geoffroei genome as a tool for mapping the ancient Saccharomyces cerevisiae genome. Science, 304, 304–307.

El-Mabrouk, N. and Sankoff, D. (2012) Analysis of gene order evolution beyond single-copy genes. In Anisimova, M. (ed). Combinatorics of Genome Rearrangements. New York: Humana Press, pp. 397–429.

Fertin, G. et al. (1974) Combinatorics of Genome Rearrangements. Cambridge, MA: MIT Press.

Freeling, M. et al. (2012) Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. Curr. Opin. Plant Biol., 15, 131–139.