

Restricted DCJ-Indel Model Revisited

Marília D. V. Braga¹ and Jens Stoye²

¹ Inmetro – Instituto Nacional de Metrologia, Qualidade e Tecnologia, Brazil

² Technische Fakultät and CeBiTec, Universität Bielefeld, Germany
mbraga@inmetro.gov.br, jens.stoye@uni-bielefeld.de

Abstract. The Double Cut and Join (DCJ) is a generic operation representing many rearrangements that can change the organization of a genome, but not its content. For comparing two genomes with unequal contents, in addition to DCJ operations, we have to allow insertions and deletions of DNA segments. The distance in the so-called general DCJ-indel model can be exactly computed, but allows circular chromosomes to be created at intermediate steps, even if the compared genomes are linear. In this case it is more plausible to consider the restricted DCJ-indel model, in which the reincorporation of a circular chromosome has to be done immediately after its creation. This model was studied recently by da Silva *et al.* (BMC Bioinformatics 13, Suppl. 19, S14), but only an upper bound for the restricted DCJ-indel distance was provided. Here we solve an open problem posed in that paper and present a very simple proof showing that the distance, that can be computed in linear time, is always the same for both the general and the restricted DCJ-indel models. We also present a simpler algorithm for computing an optimal restricted DCJ-indel sorting scenario in $O(n \log n)$ time.

1 Introduction

Genomes can be composed of one or more chromosomes, that can be linear or circular. A good estimate of evolutionary *distance* based on whole genome comparison can be obtained by asking for the minimum number of rearrangements that are necessary to transform one genome into another one. In the literature this transformation has also been referred to as *sorting* one genome *into* another genome. A sequence of rearrangements sorting a genome A into a genome B is called *scenario*, that is *optimal* when its length is minimum. Typical rearrangements that change the organization of genomes are inversions of chromosomal segments, translocations of the ends of two linear chromosomes, and chromosome fusions and fissions.

A polynomial algorithm was proposed by Hannenhalli and Pevzner in 1995 to compute the genomic distance between two genomes with equal contents considering all mentioned rearrangements. The paper [1], however, relies on the analysis of many particular cases and is full of technical details, making it susceptible to errors [2–6]. Later the same set of rearrangements were unified in the simple Double Cut and Join (DCJ) model [7], which has become very popular over the last few years due to its general applicability and mathematical elegance [8–12].

Computing the DCJ distance and finding one optimal DCJ sorting scenario can be done in linear time [8]. However, while sorting a genome into another by DCJ, circular chromosomes can appear in the intermediate steps [7, 8]. In the *general model* many circular chromosomes can coexist in some intermediate step, even if the compared genomes are composed of linear chromosomes only, such as eukaryotic nuclear genomes. To account for this fact, a restricted version of the DCJ model has been considered, where in the start and end genomes all chromosomes are linear, and whenever in an intermediate step a circular chromosome is created, it has to be reincorporated into a linear chromosome in the next step. These two consecutive DCJ operations, which create and reincorporate a circular chromosome, mimic a transposition or a block interchange [7, 13].

In Figure 1 we give examples of a general and a restricted DCJ sorting scenarios. While the general and the restricted DCJ distance are equal and can be computed in linear time [7, 8], the currently best known algorithm to find an optimal restricted sorting scenario runs in $O(n \log n)$ time [13], where n is the number of common DNA segments between the compared genomes.

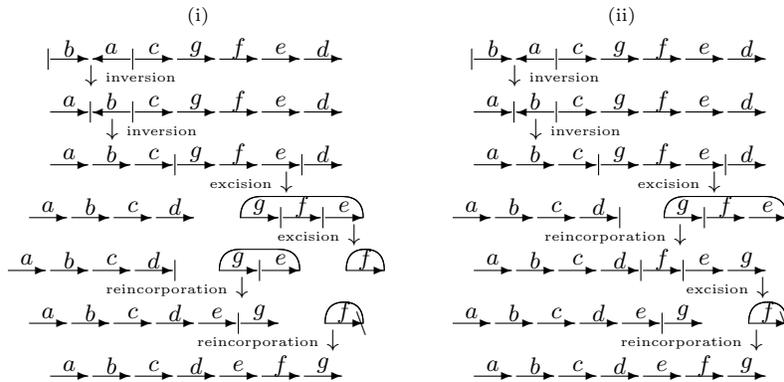


Fig. 1. (i) An optimal sorting sequence in the general DCJ model – many circular chromosomes can coexist in the intermediate species. (ii) An optimal sorting sequence in the restricted DCJ model – a circular chromosome is immediately reincorporated after its excision. The first excision-reincorporation mimics the interchange of segments d and g , while the second excision-reincorporation mimics the transposition of segment f .

Many variants of the general DCJ model have been proposed, including an extension to genomes with unequal contents, i.e. where DNA segments may be present in one, but not in the other genome. In order to transform one such genome into the other one, insertions and deletions (indels) of segments are necessary, giving rise to the so-called *DCJ-indel* model [14, 15], for which both the distance and one optimal sorting scenario can be obtained in linear time. However, like in the basic DCJ model, several circular chromosomes may coexist in intermediate steps.

A restricted version of the DCJ-indel model [16] has also been considered, but the question whether both the general and the restricted DCJ-indel distances were the same was not so easy to answer as it was for the general DCJ model. In fact, the paper by da Silva *et al.* [16] gives only an upper bound for the restricted DCJ-indel distance and an algorithm that achieves this bound. Deriving an exact distance formula and an optimal sorting algorithm were left as open problems. In Figure 2 we give examples of a general and a restricted DCJ-indel sorting scenarios.

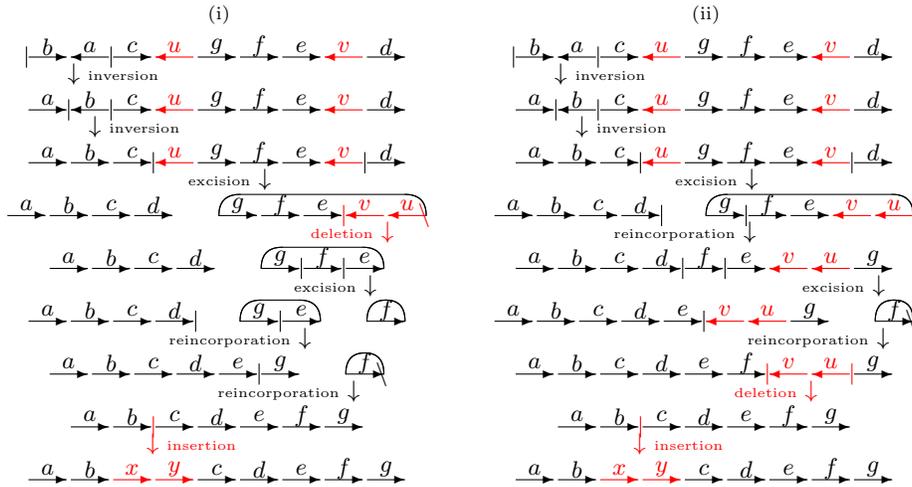


Fig. 2. (i) An optimal sorting sequence in the general DCJ-indel model – many circular chromosomes can coexist in the intermediate species. (ii) An optimal sorting sequence in the restricted DCJ-indel model – a circular chromosome is immediately reincorporated after its excision. The first excision-reincorporation mimics the interchange of segments d and $\bar{u}g$, while the second excision-reincorporation mimics the transposition of segment f .

In this work we prove that the distance is always the same for both the general and the restricted DCJ-indel models, as already conjectured in [17]. We also give a simple algorithm for computing an optimal sorting scenario under the restricted DCJ-indel model.

This paper is organized as follows. In Section 2 we give definitions and previous results used in this work. In Section 3 we show how to compute the distance and one optimal sorting scenario in the restricted DCJ-indel model. Section 4 concludes by relating this work to other genomic distance measures and pointing out open problems concerning their restricted versions.

2 Preliminaries

Each marker g in a genome is an oriented DNA fragment, represented by the symbol g , if it is read in direct orientation, or by the symbol \bar{g} , if it is read in reverse orientation. Each one of the two extremities of a linear chromosome is called a *telomere*, represented by the symbol \circ . Each chromosome in a genome can then be represented by a string that can be circular, if the chromosome is circular, or linear and flanked by the symbols \circ , if the chromosome is linear.

We deal with models in which duplicated markers are not allowed. Given two genomes A and B , possibly with unequal content, let \mathcal{G} , \mathcal{A} and \mathcal{B} be three disjoint sets, such that \mathcal{G} is the set of *common markers* that occur once in A and once in B , \mathcal{A} is the set of markers that occur only in A and \mathcal{B} is the set of markers that occur only in B . The markers in sets \mathcal{A} and \mathcal{B} are also called *unique markers*.

As an example, consider the linear genomes $A = \{\circ\bar{b}\bar{a}\bar{c}\bar{u}\bar{g}\bar{f}\bar{e}\bar{v}\bar{d}\circ\}$ and $B = \{\circ a b x y c d e f g \circ\}$, that are the top and the bottom genomes represented in both parts of Figure 2. Here we have $\mathcal{G} = \{a, b, c, d, e, f, g\}$, $\mathcal{A} = \{u, v\}$ and $\mathcal{B} = \{x, y\}$.

2.1 DCJ operations

A *cut* performed on a genome A separates two adjacent markers of A . A *double-cut and join* or *DCJ* applied on a genome A is the operation that performs cuts in two different positions in A , creating four open ends, and joins these open ends in a different way. As an example consider the first DCJ applied to genome $A = \{\circ\bar{b}\bar{a}\bar{c}\bar{u}\bar{g}\bar{f}\bar{e}\bar{v}\bar{d}\circ\}$ in Figure 2. This operation cuts before and after $\bar{b}\bar{a}$, creating the segments $\circ\bullet$, $\bullet\bar{b}\bar{a}\bullet$ and $\bullet\bar{c}\bar{u}\bar{g}\bar{f}\bar{e}\bar{v}\bar{d}\circ$, where the symbol \bullet represents the open ends. If we then join the first with the third and the second with the fourth open end, we obtain $A' = \{\circ\bar{a}\bar{b}\bar{c}\bar{u}\bar{g}\bar{f}\bar{e}\bar{v}\bar{d}\circ\}$. This DCJ corresponds to the inversion of contiguous markers $\bar{b}\bar{a}$. Indeed, a DCJ operation can correspond to several rearrangements, such as an inversion, a translocation, a fusion or a fission, and also to circular excisions and reincorporations [7].

Some additional rearrangements correspond to more than one DCJ operation. A *block interchange* occurs when two segments exchange their positions. A particular case is a *transposition*, in which one of the two segments is empty. When a block interchange or a transposition affects one single chromosome it is said to be *internal*, otherwise *external*. These rearrangements require at least three distinct cuts and cannot be represented by a single DCJ operation. Instead, they can be obtained by a composition of two DCJ operations. While external block interchanges and transpositions can always be mimicked by two consecutive translocations, internal ones can only be mimicked by two DCJs if the first is a circular excision and the second is a circular reincorporation. We call such a pair of operations an *ER composition* (see Figure 3).

2.2 The general DCJ model

In the general DCJ model the genomes have the same content and can be unichromosomal or multichromosomal, linear or circular. Given two genomes A and B

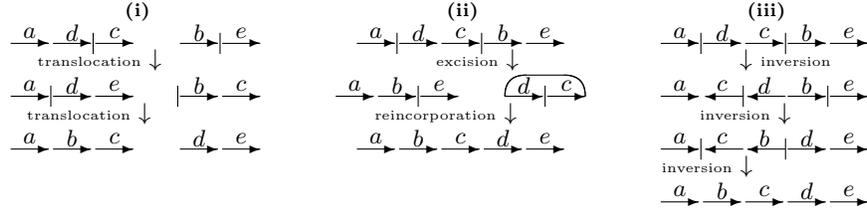


Fig. 3. (i) External block interchange of markers d and b mimicked by two translocations. (ii) Internal block interchange of markers d and b mimicked by an ER composition. (iii) Without a circular excision, the internal block interchange of markers d and b requires at least three inversions to be mimicked.

with equal contents, the DCJ distance of A and B , denoted by $d^{DCJ}(A, B)$, is the minimum number of DCJ operations that sort A into B and can be exactly computed in linear time [8]. Consider a DCJ ρ transforming the genome A into another genome A' . If $d^{DCJ}(A, B) = d^{DCJ}(A', B) + 1$, the operation ρ is said to be optimal. Under the general DCJ model, an optimal sorting scenario, composed of optimal DCJ operations, can also be obtained in linear time [8].

The restricted DCJ model. In the restricted DCJ model the genomes are linear, unichromosomal or multichromosomal. Given two linear genomes A and B with equal contents, the *restricted DCJ distance* of A and B , denoted by $d^{rDCJ}(A, B)$, is the minimum number of DCJ operations that sort A into B , with the restriction that a circular excision has to be immediately followed by a circular reincorporation, forming an ER composition. After an optimal circular excision, there is always an optimal circular reincorporation [7]. Such an ER composition is said to be optimal and guarantees that $d^{rDCJ}(A, B) = d^{DCJ}(A, B)$. The best algorithm to find a restricted DCJ sorting scenario runs in $O(n \log n)$ time [13], where n is the number of markers in A and B , respectively.

2.3 Indels

DCJ operations are able to change only the organization of the genomes, but not their contents. When the genomes have unequal contents, we need to consider *insertions* and *deletions* of blocks of contiguous markers [14, 18]. We refer to insertions and deletions as *indel* operations. Indels have two restrictions: (i) markers of \mathcal{G} cannot be deleted; and (ii) an insertion cannot produce duplicated markers [15]. At most one chromosome can be entirely deleted or inserted at once. We illustrate an indel with the following example: the insertion of markers xy into the genome $B' = \{ \circ abcdefg \circ \}$, that results into $B = \{ \circ abxy cdefg \circ \}$, as we can see in the last sorting step of both scenarios shown in Figure 2. The opposite operation would be a deletion.

The triangular inequality problem. Since indels can be applied to blocks of markers of arbitrary size, the triangular inequality does not hold for genomic distances

that consider this type of operation. Given any three genomes A , B and C and a distance measure d , consider without loss of generality that $d(A, B) \geq d(A, C)$ and $d(A, B) \geq d(B, C)$. Then the triangular inequality is the property that guarantees that $d(A, B) \leq d(A, C) + d(B, C)$.

Although this property holds for the classical models that consider only rearrangements, it does not hold for the approaches that allow indels. Consider for example the genomes $A = \{ \circ abcde \circ \}$, $B = \{ \circ ac\bar{d}be \circ \}$ and $C = \{ \circ ae \circ \}$ [14]. While A and B can be sorted into C with only one indel, the minimum number of inversions required to sort A into B is three. In this case the triangular inequality is disrupted. This is a problem if one intends to use this distance to compute the *median* of three or more genomes [11] and in phylogenetic reconstructions [19].

2.4 The general DCJ-indel model

In the DCJ-indel model the genomes can be unichromosomal or multichromosomal, linear or circular. We assign the cost of 1 to each DCJ operation and a positive cost w to each indel. Given two genomes A and B , the *DCJ-indel distance* of A and B , denoted by $d^{DCJ-id}(A, B)$, is the minimum cost of a sequence of DCJ and indel operations that sort A into B . If $w = 1$, the DCJ-indel distance corresponds exactly to the minimum number of steps required to sort A into B . For any positive $w \leq 1$, the DCJ-indel distance can be exactly computed in linear time [15, 20].

Let S be a rearrangement scenario with DCJ and indel operations, we denote by $\|S\|$ the cost of S . By definition, $\|S\| = n^{DCJ} + w(n^{ins} + n^{del})$, where n^{DCJ} is the number of DCJ operations and n^{ins} and n^{del} are, respectively, the number of insertions and deletions in S . If S is an optimal scenario sorting A into B , then $\|S\| = d^{DCJ-id}(A, B)$.

Establishing the triangular inequality. The triangular inequality does not hold for the DCJ-indel distance, but a correction can be applied *a posteriori*, as proposed in [15, 21]. It comprises summing to the distance a surcharge that depends on the number of unique markers. It has been shown that, given a positive constant $k = (w+1)/2$, for any $k' \geq k$ the triangular inequality holds for the function $m(A, B) = d^{DCJ-id}(A, B) + k'(|\mathcal{A}| + |\mathcal{B}|)$.

3 Restricted DCJ-indel model

In the restricted DCJ-indel model the genomes are linear, unichromosomal or multichromosomal. We assign the cost of 1 to each DCJ operation and a positive cost $w \leq 1$ to each indel. Given two such genomes A and B , the *restricted DCJ-indel distance* of A and B , denoted by $d^{rDCJ-id}(A, B)$, is the minimum cost of a scenario of DCJ and indel operations that sort A into B , with the restriction that a circular excision has to be immediately followed by a circular reincorporation, forming an *ER* composition.

In this section we first solve an open problem from [16], before we present a simple algorithm to compute an optimal rearrangement scenario under the restricted DCJ-indel model.

3.1 Computing the distance

Given two linear genomes A and B without duplicated markers, let S_1 be an optimal DCJ-indel scenario transforming A into B and let n^{DCJ} , n^{ins} and n^{del} be the number of DCJ operations, insertions and deletions in S_1 , such that we have $d^{DCJ-id}(A, B) = \|S_1\| = n^{DCJ} + w(n^{ins} + n^{del})$.

As shown in [16], the scenario S_1 can be transformed into another optimal scenario S_2 of the same cost, so that S_2 starts with n^{ins} insertions, followed by n^{DCJ} DCJ operations, followed by n^{del} deletions. We can represent S_2 as follows:

$$S_2 = S_2^{ins} ++ S_2^{DCJ} ++ S_2^{del}$$

where S_2^{ins} is the prefix of S_2 with only insertions, S_2^{del} is the suffix of S_2 with only deletions, and S_2^{DCJ} is the substring of S_2 with DCJ operations. The symbol $++$ denotes concatenation of rearrangement scenarios.

Let A' be the linear genome obtained after applying to A the insertions of S_2^{ins} and let B' be the linear genome obtained after applying to A' the DCJ operations of S_2^{DCJ} . Then the distance can be rewritten as

$$d^{DCJ-id}(A, B) = \|S_2^{ins}\| + d^{DCJ}(A', B') + \|S_2^{del}\|.$$

Thus, A' and B' are two linear genomes with the same set of markers and DCJ distance $d^{DCJ}(A', B') = |S_2^{DCJ}|$, where $|S_2^{DCJ}|$ denotes the number of operations in S_2^{DCJ} . From [7, 13] we know that there exists a *restricted* DCJ scenario R of the same cost as S_2^{DCJ} , sorting A' into B' . Hence there also exists a restricted DCJ-indel sorting scenario S_3 transforming A into B :

$$S_3 = S_2^{ins} ++ R ++ S_2^{del}.$$

Clearly, S_3 has the same cost as S_2 and thus as S_1 , being an optimal restricted DCJ-indel sorting scenario. These observations give rise to the following theorem:

Theorem 1. *Given two linear genomes A and B without duplicated markers, we have*

$$d^{rDCJ-id}(A, B) = d^{DCJ-id}(A, B).$$

Observe that Theorem 1 holds even if we assign the cost of 1 to each DCJ and a positive cost $w \leq 1$ to each indel operation.

Complexity. For any positive indel cost $w \leq 1$, the DCJ-indel distance can be computed in linear time [15, 20], and thus the same is true for the restricted DCJ-indel distance.

Establishing the triangular inequality. Obviously the correction proposed in [15, 21] to establish the triangular inequality for the DCJ-indel distance also holds for the restricted DCJ-indel distance.

3.2 Finding an optimal sorting scenario

It can be easily seen that the procedure described in the previous subsection implies a simple algorithm for finding a restricted DCJ-indel scenario sorting a linear genome A into a linear genome B (Algorithm 1).

Algorithm 1 Find a restricted DCJ-indel scenario sorting a linear genome A into a linear genome B

1. Compute an optimal DCJ-indel scenario S_1 sorting A into B using the algorithm from [15, 20].
 2. Modify S_1 by moving the insertions up and the deletions down, as shown in [16], obtaining a scenario $S_2 = S_2^{ins} ++ S_2^{DCJ} ++ S_2^{del}$.
 3. Use S_2^{ins} to transform A into a linear genome A' .
 4. Use S_2^{DCJ} to transform A' into a linear genome B' (A' and B' have the same content $\mathcal{G}' = \mathcal{G} \cup \mathcal{A} \cup \mathcal{B}$).
 5. Apply the restricted DCJ algorithm from [13] to obtain a restricted DCJ scenario R sorting A' into B' .
 6. Concatenate the three parts to obtain the scenario $S_3 = S_2^{ins} ++ R ++ S_2^{del}$, that is a restricted DCJ-indel scenario sorting A into B .
-

Complexity. In Algorithm 1, steps 1-4 and 6 can be implemented in linear time, while step 5 takes $O(n \log n)$ time, where $n = |\mathcal{G}'|$ is the number of markers in A' , respectively B' . Thus, a restricted DCJ-indel sorting scenario can be computed in $O(n \log n)$ time.

Implementation. While an implementation of the restricted DCJ sorting is available in [22], to the best of our knowledge there exists no implementation of the general DCJ-indel sorting algorithm. Given such an implementation would then make it rather straightforward to also implement the restricted DCJ-indel sorting algorithm.

4 Conclusions and Perspectives

In this paper we have solved an open problem, showing that, even if the indel cost is distinct from and upper bounded by the DCJ cost, the restricted DCJ-indel distance is equal to the DCJ-indel distance, that can be computed in linear

time. This allows the correction for establishing the triangular inequality in the DCJ-indel distance to be automatically extended to the restricted DCJ-indel distance.

We have also proposed an algorithm to generate an optimal restricted DCJ-indel sorting scenario in $O(n \log n)$ time. The most complicated parts of this algorithm are: (a) obtaining a general DCJ-indel sorting scenario between two genomes with unequal contents (step 1 of Algorithm 1) and (b) obtaining a restricted DCJ sorting scenario between genomes with equal contents (step 5 of Algorithm 1). An implementation of (b) is available in [22], but the implementation of (a) still has to be developed.

The inversion-indel distance. The inversion-indel is a related model that applies to unichromosomal (linear or circular) genomes only, and, instead of generic DCJ operations, allows only inversions of DNA segments, besides indels. An example is given in Figure 4. In [18] two algorithms were provided for this distance: an exact one for the case in which only one indel direction is allowed (i.e. when we have either only insertions or only deletions); and a heuristic for the symmetric case, in which both insertions and deletions are allowed. Recently, in a joint work with other authors [23], we proved that, for an important class of instances of the symmetric case, the inversion-indel distance equals the DCJ-indel distance. An exact solution for the general symmetric case remains an open problem.

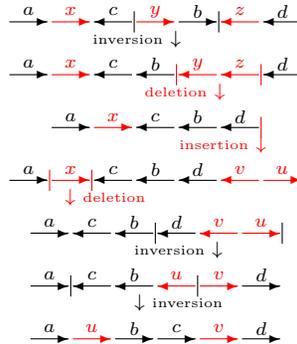


Fig. 4. An optimal sorting scenario in the inversion-indel model.

The restricted DCJ-substitution distance. The DCJ-substitution is another related model that applies to linear genomes, unichromosomal or multichromosomal. In this model we have generic DCJ operations, but, instead of indels, more powerful operations are considered: *substitutions* allow blocks of contiguous markers to be replaced by other blocks of contiguous markers [24]. In other words, a deletion and a subsequent insertion that occur at the same position of

the genome can be modeled as a substitution, counting together for one single step. In the DCJ-substitution model, indels are special cases of substitutions: if a block of markers is substituted by the empty string, we have a deletion; analogously, if the empty string is substituted by a block of markers, we have an insertion.

In the general DCJ-substitution model the results are very similar to the general DCJ-indel model. For a cost of 1 assigned to DCJ operations and any positive cost $w \leq 1$ assigned to substitutions, there is a formula to efficiently compute the distance [20, 24]. However, the general and the restricted DCJ-substitution distances are not the same, as we can see in the example given in Fig. 5. The restricted version of the DCJ-substitution distance is a complete open problem that we intend to study in the future.

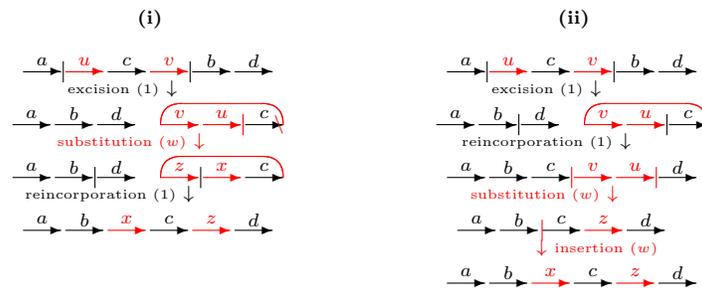


Fig. 5. (i) An optimal sorting scenario in the general DCJ-substitution model, with cost $2+w$. (ii) An optimal sorting scenario in the restricted DCJ-substitution model, with cost $2+2w$.

Acknowledgments

MDVB is funded by the Brazilian research agency CNPq grant PROMETRO 563087/10-2.

References

1. Hannenhalli, S., Pevzner, P.: Transforming men into mice (polynomial algorithm for genomic distance problem). In: Proc. of FOCS 1995. (1995) 581–592
2. Tesler, G.: Efficient algorithms for multichromosomal genome rearrangements. J. Comput. Syst. Sci. **65** (2002) 587–609
3. Ozery-Flato, M., Shamir, R.: Two notes on genome rearrangement. J. Bioinf. Comput. Biol. **1** (2003) 71–94
4. Jean, G., Nikolski, M.: Genome rearrangements: A correct algorithm for optimal capping. Inf. Process. Lett. **104** (2007) 14–20

5. Bergeron, A., Mixtacki, J., Stoye, J.: A new linear time algorithm to compute the genomic distance via the double cut and join distance. *Theor. Comput. Sci.* **410** (2009) 5300–5316
6. Erdős, P.L., Soukup, L., Stoye, J.: Balanced vertices in trees and a simpler algorithm to compute the genomic distance. *Appl. Math. Lett.* **24** (2011) 82–86
7. Yancopoulos, S., Attie, O., Friedberg, R.: Efficient sorting of genomic permutations by translocation, inversion and block interchange. *Bioinformatics* **21** (2005) 3340–3346
8. Bergeron, A., Mixtacki, J., Stoye, J.: A unifying view of genome rearrangements. In: *Proc. of WABI 2006*. Volume 4175 of LNBI. (2006) 163–173
9. Adam, Z., Sankoff, D.: The ABCs of MGR with DCJ. *Evol. Bioinform. Online* **4** (2008) 69–74
10. Mixtacki, J.: Genome halving under DCJ revisited. In: *Proc. of COCOON 2008*. Volume 5092 of LNCS. (2008) 276–286
11. Tannier, E., Zheng, C., Sankoff, D.: Multichromosomal median and halving problems under different genomic distances. *BMC Bioinformatics* **10** (2009) 120
12. Thomas, A., Varré, J.S., Ouangraoua, A.: Genome dedoubling by DCJ and reversal. *BMC Bioinformatics* **12** (2012) S20
13. Kováč, J., Warren, R., Braga, M.D.V., Stoye, J.: Restricted DCJ model (the problem of chromosome reincorporation). *J. Comput. Biol.* **18** (2011) 1231–1241
14. Yancopoulos, S., Friedberg, R.: DCJ path formulation for genome transformations which include Insertions, Deletions, and Duplications. *J. Comput. Biol.* **16** (2009) 1311–1338
15. Braga, M.D.V., Willing, E., Stoye, J.: Double cut and join with insertions and deletions. *J. Comput. Biol.* **18** (2011) 1167–1184
16. da Silva, P.H., Machado, R., Dantas, S., Braga, M.D.V.: Restricted DCJ-indel model: sorting linear genomes with DCJ and indels. *BMC Bioinformatics* **13** (2012) S14
17. Braga, M.D.V.: An overview of genomic distances modeled with indels. In: *Proc. of CiE*. Volume 7921 of LNCS. (2013) 22–31
18. El-Mabrouk, N.: Sorting signed permutations by reversals and insertions/deletions of contiguous segments. *Journal of Discrete Algorithms* **1** (2001) 105–122 A preliminary version appeared in *Proc. of CPM 2000*, LNCS 1848.
19. Swenson, K.M., Arndt, W., Tang, J., Moret, B.: Phylogenetic reconstruction from complete gene orders of whole genomes. In: *Proc. of Asia-Pacific Bioinformatics Conf.* Volume 6 of *Advances in Bioinformatics and Comp. Biology*. (2008) 241–250
20. da Silva, P.H., Machado, R., Dantas, S., Braga, M.D.V.: DCJ-indel and DCJ-substitution distances with distinct operation costs. To appear in *Alg. for Mol. Biol.* (2013) A preliminary version appeared in *Proc. of WABI 2012*, LNBI 7534.
21. Braga, M.D.V., Machado, R., Ribeiro, L.C., Stoye, J.: On the weight of indels in genomic distances. *BMC Bioinformatics* **12** (2011) S13
22. Hilker, R., Sickinger, C., Pedersen, C., Stoye, J.: UniMoG - a unifying framework for genomic distance calculation and sorting based on DCJ. *Bioinformatics* **28** (2012) 2509–2511
23. Willing, E., Zaccaria, S., Braga, M.D.V., Stoye, J.: On the inversion-indel distance. To appear in *BMC Bioinformatics* (2013)
24. Braga, M.D.V., Machado, R., Ribeiro, L.C., Stoye, J.: Genomic distance under gene substitutions. *BMC Bioinformatics* **12** (2011) S8