Compeau *Algorithms for Molecular Biology* 2013, **8**:6
http://www.almob.org/content/8/1/6

AMB ALGORITHMS FOR MOLECULAR BIOLOGY

RESEARCH Open Access

DCJ-Indel sorting revisited

Phillip EC Compeau

Abstract

Background: The introduction of the double cut and join operation (DCJ) caused a flurry of research into the study of multichromosomal rearrangements. However, little of this work has incorporated indels (i.e., insertions and deletions of chromosomes and chromosomal intervals) into the calculation of genomic distance functions, with the exception of Braga et al., who provided a linear time algorithm for the problem of DCJ-indel sorting. Although their algorithm only takes linear time, its derivation is lengthy and depends on a large number of possible cases.

Results: We note the simple idea that a deletion of a chromosomal interval can be viewed as a DCJ that creates a new circular chromosome. This framework will allow us to amortize indels as DCJs, which in turn permits the application of the classical breakpoint graph to obtain a simplified indel model that still solves the problem of DCJ-indel sorting in linear time via a more concise formulation that relies on the simpler problem of DCJ sorting. Furthermore, we can extend this result to fully characterize the solution space of DCJ-indel sorting.

Conclusions: Encoding indels as DCJ operations offers a new insight into why the problem of DCJ-indel sorting is not ultimately any more difficult than that of sorting by DCJs alone. There is still room for research in this area, most notably the problem of sorting when the cost of indels is allowed to vary with respect to the cost of a DCJ and we demand a minimum cost transformation of one genome into another.

Keywords: Genome rearrangements, DCJ, Indels, Sorting, Solution space

Background

In the simplest terms, DNA may mutate in two fundamentally different ways. On the one hand, single-nucleotide polymorphisms alter the base at a single position of the nucleic acid polymer; on the other hand, huge mutations called chromosomal rearrangements can move around, duplicate, insert, or delete huge blocks of DNA, often from one chromosome to another.

Chromosomal rearrangements were first observed by Dobzhansky and Sturtevant in 1938 ([1]), but extensive efforts to quantify their study did not take off until the early 1990s. In the last two decades, a number of discrete genomic models have been proposed and studied (see [2] for an overview of the combinatorics of genome rearrangements).

Having selected a genomic model and a collection of genome operations to consider, the standard algorithmic problem is the computation of the *distance* between two genomes Π and Γ, or the minimum number of allowable operations required to transform Π into Γ; the more difficult problem of *sorting* demands the operations themselves. The first historical example of such a discrete genomic distance is the *prefix reversal distance* for permutations (which model the order of genes along a single linear chromosome), introduced in [3] and bounded in [4-6]. The computation of prefix reversal distance has been proposed to be *NP*-Hard (see [7]).

More recent research has moved past permutations and toward multichromosomal genomic models that incorporate both linear and circular chromosomes. One of these models, which we will study in this paper, models the chromosomes of a genome with paths and cycles in a graph. For this model, the double cut and join operation (DCJ) was introduced in [8] and incorporates segment reversals with a number of other operations. Interestingly, a linear time greedy algorithm exists for DCJ sorting two genomes having equal gene content (see [9]).

The incorporation of insertions and deletions of chromosomes and chromosomal intervals (collectively called *indels*) into DCJ distance was discussed in [10] and quantified rigorously in [11]. The latter authors provided a linear

Correspondence: pcompeau@math.ucsd.edu
Department of Mathematics, UC San Diego, 9500 Gilman Drive 0112, San Diego, CA 92093, United States

time algorithm for the associated problem of *DCJ-indel sorting*, which gives a minimum collection of DCJ and indel operations required to transform one genome into another. Yet their argument is case-ridden, and so in this paper, which builds upon [12], we wish to provide a much simpler presentation of DCJ-indel sorting that still yields a linear-time solution to the problem.

Main text

Preliminaries

Say that we are given a perfect matching on $2N$ labeled vertices $\mathscr{V}$, forming a set $\mathcal{G}$ of N edges called *genes*; the vertices of each gene form its *head* and *tail*. We define a *genome* Π as the edge-disjoint union of two matchings. The *genes* of Π, denoted $g(\Pi)$, form a matching on $\mathscr{V}$ such that $g(\Pi) \subseteq \mathcal{G}$; the *adjacencies* of Π, denoted $a(\Pi)$, form a matching on $V(g(\Pi))$. We color the genes of Π black and the adjacencies of Π blue (see Figure 1(a)).

A consequence of these definitions is that Π comprises a disjoint collection of paths and cycles, where each connected component alternates between black genes and blue adjacencies. Each component of Π is called a *chromosome*; paths (cycles) of Π define *linear* (*circular*) chromosomes of Π. The endpoint v of a path in Π is called a *telomere* of Π; v is not incident to an adjacency, and so for clerical purposes, we say that v has the *null adjacency* $\{v, \emptyset\}$. A genome consisting of only circular (linear) chromosomes is called a *circular* (*linear*) *genome*. Note that Π is circular if and only if the edges of $a(\Pi)$ form a perfect matching on $V(\Pi)$.

Henceforth, we only consider genome pairs $\{\Pi, \Gamma\}$ such that $g(\Pi) \cup g(\Gamma) = \mathcal{G}$. A workhorse data structure encoding the relationship between Π and Γ is the *breakpoint graph* ([13]), denoted by $\mathrm{B}(\Pi, \Gamma)$ and defined as the edge-disjoint union[a] of $a(\Pi)$ and $a(\Gamma)$, where adjacencies of Γ will be colored red (Figure 1(b)). Observe that $B(\Pi, \Gamma)$ is also a collection of disjoint paths and cycles, which alternate between red and blue edges. The *length* of a connected component of $\mathrm{B}(\Pi, \Gamma)$ is its total number of edges; we consider an isolated vertex in $\mathrm{B}(\Pi, \Gamma)$ to be a path of length 0. The breakpoint graph is also the line graph of the *adjacency graph*, which was first defined in [9] and has also been used in rearrangement studies.

A *double cut and join* operation (DCJ) on Π (introduced in [8]) *uses* one or two adjacencies of Π via one of the following four operations to produce a new genome Π':

1. $\{v, w\}, \{x, y\} \longrightarrow \{v, x\}, \{w, y\}$
2. $\{v, w\}, \{x, \emptyset\} \longrightarrow \{v, x\}, \{w, \emptyset\}$
3. $\{v, \emptyset\}, \{w, \emptyset\} \longrightarrow \{v, w\}$
4. $\{v, w\} \longrightarrow \{v, \emptyset\}, \{w, \emptyset\}$

The DCJ incorporates a wide range of genome rearrangements, as shown in Figure 2.

For the particular case that Π and Γ have the same genes (i.e., $g(\Pi) = g(\Gamma) = \mathcal{G}$), the *DCJ distance* between Π and Γ, written $d_{\mathrm{DCJ}}(\Pi, \Gamma)$, is the minimum number of DCJs required to transform Π into Γ One can easily verify that d_{DCJ} forms a metric on the set of all genomes having gene

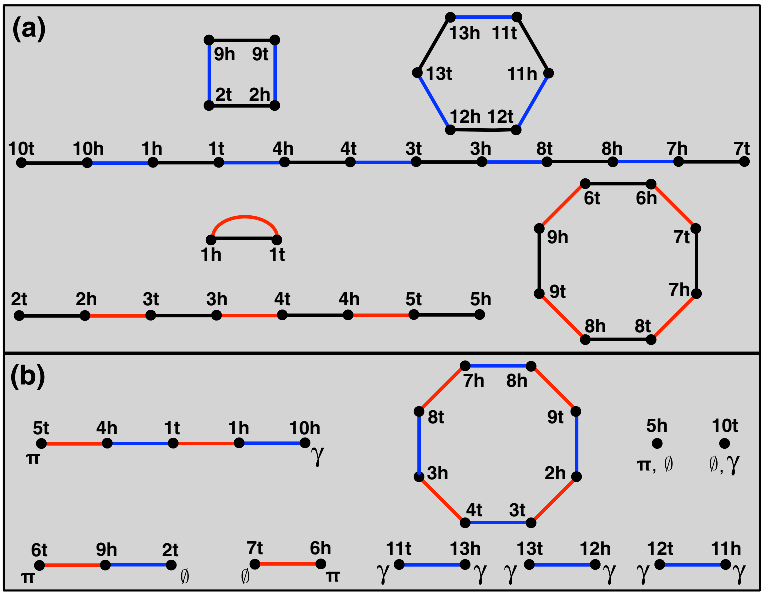

Figure 1 Two Genomes and their Breakpoint Graph. (a) Genomes Π and Γ on a collection of 12 genes. We use "h" and "t" to denote the head and tail of a gene. Π is drawn with blue adjacencies, and Γ is drawn with red adjacencies. **(b)** The breakpoint graph of Π and Γ. We have labeled the endpoint v of a path with π if v is π-open, with γ if v is γ-open, and with $\emptyset$ if v is a telomere of at least one genome.

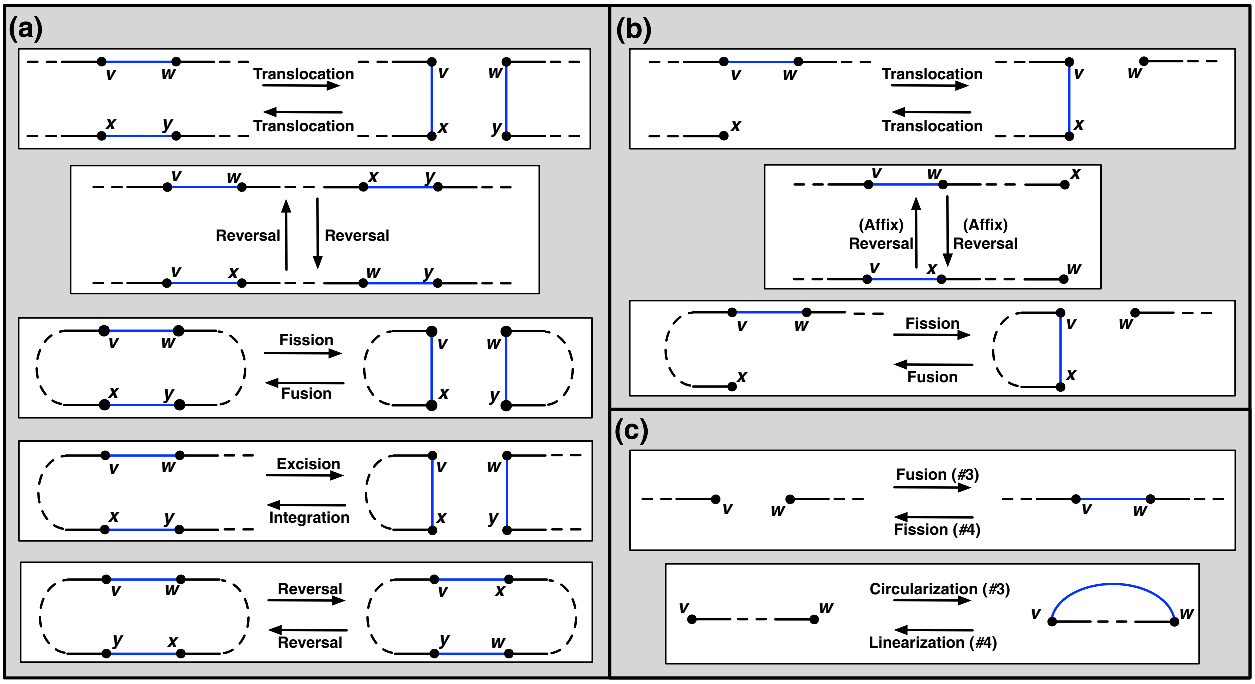

Figure 2 The Collection of All Possible DCJ Operations. The DCJ incorporates many operations, depending on the structure of the chromosomes involved and whether the adjacencies used belong to the same chromosome. **(a)** Operation 1 in the definition of the DCJ incorporates linear internal translocations, reversals, circular fusions/fissions, the excision of a circular chromosome from a linear chromosome, and the integration of a circular chromosome into a linear chromosome. **(b)** Operation 2 incorporates telomeric translocations, affix reversals (which involve the telomere of a linear chromosome), and the fission of a linear chromosome into a circular and linear chromosome (together with its inverse). **(c)** Operations 3 and 4 include linear fusions/fissions as well as the linearization/circularization of a single chromosome.

set $\mathcal{G}$. A closed formula for DCJ distance was derived in [9] and translated into breakpoint graph notation in [14]:

$$d_{\mathrm{DCJ}}(\Pi,\Gamma) = N - c(\Pi,\Gamma) - \frac{p_{\mathrm{even}}(\Pi,\Gamma)}{2} \qquad (1)$$

Here, $c(\Pi,\Gamma)$ and $p_{\mathrm{even}}(\Pi,\Gamma)$ denote the number of cycles and even-length paths in $\mathrm{B}(\Pi,\Gamma)$, respectively.

For the more general case that Π and Γ do not share the same genes, a *deletion* of a chromosomal interval of Π replaces adjacencies $\{v,w\}$ and $\{x,y\}$ (contained in the order (v,w,x,y) along a chromosome of Π) with the adjacency $\{v,y\}$ and removes the path connecting w to x. We also allow deletions of entire chromosomes; however, we must stipulate (following the lead of the authors in [11]) that every vertex removed from Π must belong to $\mathcal{V} - V(\Gamma)$.[b] The *insertion* of a chromosome or chromosomal interval into Π to obtain Π' is defined as the inverse of a corresponding deletion from Π' that yields Π. Note that a consequence of this definition is that we may not insert a gene unless it is contained in $\mathcal{G}$. Insertions and deletions are collectively called *indels*; thus, we define the *DCJ-indel distance* between Π and Γ, written $d_{\mathrm{DCJ}}^{\mathrm{ind}}(\Pi,\Gamma)$, as the minimum number of DCJs and indels required to transform Π into Γ.

Because insertions and deletions are inverse operations, it follows that $d_{\mathrm{DCJ}}^{\mathrm{ind}}(\Pi,\Gamma) = d_{\mathrm{DCJ}}^{\mathrm{ind}}(\Gamma,\Pi)$. However, although $d_{\mathrm{DCJ}}^{\mathrm{ind}}$ is symmetric, unlike d_{DCJ} it does not form a metric, as the triangle inequality does not hold; see [15] for a more complete discussion.

DCJ-Indel sorting

Handling circular singletons

We begin our discussion of DCJ-indel sorting by defining a *circular singleton* of Π (adapted from[11]) as a circular chromosome C such that $V(C) \cap V(\Gamma) = \emptyset$. Note that C is defined with respect to Γ as well as Π. Ideally, we could delete (insert) all circular singletons of Π and Γ immediately to simplify the problem of DCJ-indel sorting; fortunately, this is indeed the case, as shown by the following two results.

Proposition 1. *If Π' is formed by removing a circular singleton C from Π, then $d_{DCJ}^{ind}(\Pi',\Gamma) = d_{DCJ}^{ind}(\Pi,\Gamma) - 1$. Furthermore, when transforming Π into Γ via a minimum collection of DCJs and indels, no gene belonging to a circular singleton of Π can ever appear in the same chromosome as a gene of Γ.*

Proof. Any collection of k DCJs and indels transforming Π' into Γ can be supplemented by the deletion of C to yield $k+1$ DCJs and indels transforming Π into Γ; thus, $d^{\mathrm{ind}}_{\mathrm{DCJ}}(\Pi', \Gamma) \geq d^{\mathrm{ind}}_{\mathrm{DCJ}}(\Pi, \Gamma) - 1$.

To obtain the reverse bound, let us view a transformation $\mathbb{T}$ of Π into Γ as a sequence $(\Pi_0, \Pi_1, \dots, \Pi_n)$ $(n \geq 1)$, where $\Pi_0 = \Pi$, $\Pi_n = \Gamma$, and Π_{i+1} is obtained from Π_i as the result of a single DCJ or indel. Consider the sequence $(\Pi'_0, \Pi'_1, \dots, \Pi'_n)$, where Π'_i is constructed from Π_i by removing the subgraph of Π_i induced by the vertices of C under the stipulation that whenever we remove a path P connecting v to w, we replace adjacencies $\{v, x\}$ and $\{w, y\}$ in Π with $\{x, y\}$ in Π'_i. It is easy to see that $\Pi'_0 = \Pi'$, $\Pi'_n = \Gamma$, and for every i in range, either Π'_{i+1} is the result of a DCJ or indel applied to Π'_i or $\Pi'_{i+1} = \Pi'_i$; thus, $(\Pi'_0, \Pi'_1, \dots, \Pi'_n)$ encodes a transformation of Π' into Γ using at most n DCJs and indels. Furthermore, one can verify that $\Pi'_{i+1} = \Pi'_i$ only when an adjacency of C is used by a DCJ in $\mathbb{T}$ changing Π_i to Π_{i+1} or when Π_{i+1} is produced from Π_i by a deletion of vertices that all belong to C. At least one such operation must always occur in $\mathbb{T}$; hence, $d^{\mathrm{ind}}_{\mathrm{DCJ}}(\Pi', \Gamma) \leq d^{\mathrm{ind}}_{\mathrm{DCJ}}(\Pi, \Gamma) - 1$.

The proposition's second conclusion follows from the fact that if for some j $(1 \leq j \leq n-1)$, a chromosome of Π_j contains a gene g_1 of Π and a gene g_2 of C, then one DCJ was required to combine g_1 and g_2 into the same chromosome, and another will be needed to separate them, yielding two distinct values of i for which $\Pi'_{i+1} = \Pi'_i$. From the first part of the proof, we may conclude that $d^{\mathrm{ind}}_{\mathrm{DCJ}}(\Pi, \Gamma) < n$. □ □

Letting $\mathrm{sing}(\Pi, \Gamma)$ denote the total number of circular singletons of Π and Γ, we have an immediate corollary.

Corollary 2. *The DCJ-indel distance is given by the following:*

$$d^{ind}_{DCJ}(\Pi, \Gamma) = sing(\Pi, \Gamma) + d^{ind}_{DCJ}(\Pi^0, \Gamma^0) \tag{2}$$

where Π^0 (Γ^0) is formed by removing all circular singletons from Π (Γ).

With respect to DCJ-indel sorting, Corollary 2 allows us to assume without loss of generality that Π and Γ do not contain any circular singletons.

We next make an observation taken from [16], which is that the deletion of a chromosomal interval of Π connecting w to x may be viewed as a DCJ: $\{v, w\}, \{x, y\} \rightarrow \{v, y\}, \{w, x\}$; this operation produces a circular chromosome containing w and x that is scheduled for removal, including the case that v or y equals $\emptyset$ (the deletion of an entire linear chromosome is handled by $u = x = \emptyset$); see Figure 3. Because insertions are the inverses of deletions, we would like to conclude that indels may be placed in a one-to-one correspondence with the removal of circular chromosomes. Ironically, the apparent exception to this proposed rule is the deletion of an entire circular chromosome.

Yet if a deleted circular chromosome C is not produced as the result of a DCJ, then C must be a circular singleton of Π in order to be deleted. Otherwise, C has been produced as the result of a DCJ applied to a chromosomal interval; by the method we just described, we can amortize the deletion in this DCJ unless the DCJ also creates another circular chromosome C' that is scheduled for deletion. However, this sequence of operations cannot arise in a minimum collection of DCJs and indels transforming Π into Γ, as we could simply delete the original chromosome from which C and C' were produced by the DCJ in question, thus requiring a single operation instead of three.

Toward a new model of Indels

We will follow the observation made in [16] that the actual removal of deleted chromosomes can occur as a final step in the transformation of Π into Γ. As a result, we may view the transformation of Π into Γ as composed of three steps: inserting chromosomes into Π to yield a new genome Π' with $g(\Pi') = \mathcal{G}$; applying a sequence of DCJs to produce a genome Γ' having the same genes as Π'; and finally, deleting chromosomes from Γ' to produce Γ. Note that we can equivalently view the first step as the deletion of chromosomes from Π' to obtain Π. Combining this observation with our correspondence between indels and circular chromosomes above, we may introduce the following framework.

Define a *completion* of Π as a genome Π' having $g(\Pi') = \mathcal{G}$ and for which $a(\Pi')$ is composed of $a(\Pi)$ together with a perfect matching on $V(\Pi') - V(\Pi)$. We call the adjacencies of $a(\Pi') - a(\Pi)$ *new*. Note that the chromosomes of Π embed as chromosomes of Π' and that the components of $\Pi' - \Pi$ form cycles because the new adjacencies of Π' induce a perfect matching on $V(\Pi') - V(\Pi)$; we may now without ambiguity call these circular chromosomes of Π' the *indels* of Π'. A *completion* of a pair of genomes (Π, Γ) is simply a pair (Π', Γ') for which Π' and Γ' are completions of Π and Γ, respectively. The above discussion implies that for any minimum cost transformation of Π into Γ, the indels of Π' correspond bijectively to DCJ operations, so that we will amortize each unit indel cost by that of a DCJ operation. This amortization yields the following equation for DCJ-indel distance:

$$d^{\mathrm{ind}}_{\mathrm{DCJ}}(\Pi, \Gamma) = \min_{(\Pi', \Gamma')} \left\{ d_{\mathrm{DCJ}}(\Pi', \Gamma') \right\} \tag{3}$$

where the minimum is taken over all completions of (Π, Γ). A completion (Π^*, Γ^*) is *optimal* if it attains the minimum in (3). Applying the closed form equation for the DCJ distance in (1) to immediately produces the following result.

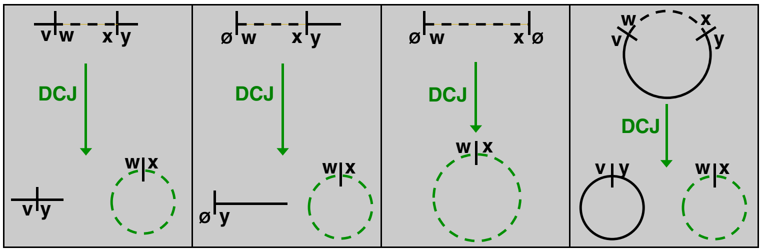

Figure 3 Encoding the Deletion of a Chromosomal Interval as a DCJ. The deletion of a chromosomal interval connecting *w* to *x* can be encoded by a DCJ that turns the interval connecting *w* to *x* into a circular chromosome. The four possible deletions of a chromosomal interval are shown in the above figure; this correspondence holds even when the interval in question is taken to be an entire linear chromosome.

Theorem 3. *The DCJ-indel distance is given by the following equation:*

$$d_{DCJ}^{ind}(\Pi,\Gamma) = N - \max_{(\Pi',\Gamma')} \left\{ c(\Pi',\Gamma') + \frac{p_{even}(\Pi',\Gamma')}{2} \right\} \tag{4}$$

where the maximum is taken over all completions of (Π, Γ).

Constructing an optimal completion

In light of Theorem 3, we have reduced DCJ-indel sorting to the problem of constructing indels intelligently to maximize a weighted sum of breakpoint graph components. Once we have produced an optimal completion (Π^*, Γ^*), we can simply invoke the $O(N)$-time sorting algorithm described in [9] to transform Π^* into Γ^* via a minimum collection of DCJs.

Our goal is to construct (Π^*, Γ^*) by direct analysis of $B(\Pi, \Gamma)$. Because Π and Γ do not necessarily share the same genes, $B(\Pi, \Gamma)$ may contain path endpoints that are not telomeres. Accordingly, we define a vertex v to be *π-open* (*γ-open*) if $v \notin \Pi$ ($v \notin \Gamma$). In other words, v must be matched to some other π-open vertex when constructing the indels of Π^*.[c] The paths of $B(\Pi, \Gamma)$ are therefore classified according to their endpoints: a *π-path* (*γ-path*) ends in one π-open (γ-open) vertex and one telomere (of either Π or Γ); a *$\{\pi, \gamma\}$-path* ends in a π-open vertex and a γ-open vertex (such a path must have even length at least 2); a *$\{\pi, \pi\}$-path* (*$\{\gamma, \gamma\}$-path*) ends in two π-open (γ-open) vertices and must therefore have odd length. We should also provide statistics for counting these different components. Define $p^{\pi,\gamma}$ as the number of $\{\pi, \gamma\}$-paths in $B(\Pi, \Gamma)$; p_{even}^{π} as the number of even-length π-paths in $B(\Pi, \Gamma)$; and p_{even}^{0} as the number of even-length paths in $B(\Pi, \Gamma)$ containing no open vertices (i.e., ending in two telomeres). Similar statistics counting odd-length paths can be defined analogously. We have dropped the genomes $\{\Pi, \Gamma\}$ from these statistics for the sake of simplicity; all component statistics will be taken with respect to $B(\Pi, \Gamma)$ unless otherwise noted.

We first present a proposition regarding the parity of the paths of $B(\Pi, \Gamma)$.

Proposition 4. *The component statistics of* $B(\Pi, \Gamma)$ *satisfy the following condition:*

$$p^{\pi,\gamma} \equiv \left| p_{odd}^{\pi} - p_{even}^{\pi} \right| \equiv \left| p_{odd}^{\gamma} - p_{even}^{\gamma} \right| \mod 2 \tag{5}$$

Proof. The total number of π-open vertices is equal to $V(\Pi') - V(\Pi)$ and must therefore be even. Of course, the same is the case for γ-open vertices, and counting π-open and γ-open vertices over the connected components of $B(\Pi, \Gamma)$ thus produces the following equivalences:

$$p_{\text{odd}}^{\pi} + p_{\text{even}}^{\pi} + p^{\pi,\gamma} \equiv 0 \mod 2 \tag{6}$$

$$p_{\text{odd}}^{\gamma} + p_{\text{even}}^{\gamma} + p^{\pi,\gamma} \equiv 0 \mod 2 \tag{7}$$

Adding $p^{\pi,\gamma}$ to both sides of (6) and (7) gives the following:

$$p^{\pi,\gamma} \equiv \left(p_{\text{odd}}^{\pi} + p_{\text{even}}^{\pi} \right) \equiv \left(p_{\text{odd}}^{\gamma} + p_{\text{even}}^{\gamma} \right) \mod 2 \tag{8}$$

The equivalence of (5) and (8) is an arithmetical fact. □

We next establish two necessary conditions on optimal completions by culling the set of possible adjacencies of any such completion. Our general strategy is to consider the addition of a new adjacency $\{v, w\}$ to a completion Π' as *linking* the component(s) of $B(\Pi, \Gamma)$ whose endpoints are the (π-open) vertices v and w. Our first result states that we must always link the endpoints of any $\{\pi, \pi\}$-path to each other.

Lemma 5. *If* (Π^*, Γ^*) *is an optimal completion of* (Π, Γ), *then every* $\{\pi, \pi\}$*-path (*$\{\gamma, \gamma\}$*-path) of length* $2k-1$ *in* $B(\Pi, \Gamma)$ *(*$k \geq 1$*) embeds into a cycle of length* $2k$ *in* $B(\Pi^*, \Gamma^*)$.

Proof. Let P be a path of length $2k - 1$ connecting π-open vertices v and w in $B(\Pi, \Gamma)$. Our claim is that we must link v and w in $B(\Pi^*, \Gamma^*)$. Suppose for the sake of contradiction that we have a completion (Π', Γ') such that

P does not embed into a cycle of length $2k$ in $\mathrm{B}(\Pi', \Gamma')$; in this case, we must have adjacencies $\{v, x\}$ and $\{w, y\}$ in $a(\Pi')$, where all four vertices are distinct.

Consider the completion Π'' that is identical to Π' except that $\{v, x\}$ and $\{w, y\}$ are replaced by $\{v, w\}$ and $\{x, y\}$. In $\mathrm{B}(\Pi'', \Gamma')$, we have closed P into a cycle of length $2k$, and at the same time, we have changed neither the parity nor the linearity/circularity of the component containing x and y. Because we have increased the number of breakpoint graph cycles by 1 without changing the total number of paths, it follows from (1) that $d_{\mathrm{DCJ}}(\Pi'', \Gamma') = d_{\mathrm{DCJ}}(\Pi', \Gamma') - 1$, and so (Π', Γ') cannot be optimal. □

Having dealt with $\{\pi, \pi\}$- and $\{\gamma, \gamma\}$-paths of $\mathrm{B}(\Pi, \Gamma)$, any remaining component of $\mathrm{B}(\Pi^*, \Gamma^*)$ must be either a *j-bracelet*, which is a cycle linking j $\{\pi, \gamma\}$-paths (where $j \geq 2$ and j is even), or a *k-chain*, in which two π-paths or two γ-paths are linked via an intermediate number of $\{\pi, \gamma\}$-paths to form a path containing k components from $\mathrm{B}(\Pi, \Gamma)$ ($k \geq 2$). Note that when k is even, a k-chain C must contain either two π-paths or two γ-paths, and when k is odd, C must contain one π-path and one γ-path.

For the sake of simplicity, we will represent a j-bracelet by $(P_1 : P_2 : \cdots : P_j)$ and a k-chain by $[P_1 : P_2 : \cdots : P_k]$, where every P_i is linked to P_{i+1}, and in the case of a j-bracelet, P_1 is linked to P_j. Because we wish to maximize a weighted sum of breakpoint graph components, we might guess that we should look for many short bracelets and chains. Indeed, the length of a bracelet or chain in $\mathrm{B}(\Pi^*, \Gamma^*)$ is heavily restricted by the following lemma.

Lemma 6. *If* (Π^*, Γ^*) *is an optimal completion, then a component* C^* *of* $\mathrm{B}(\Pi^*, \Gamma^*)$ *can only contain two or more* $\{\pi, \gamma\}$*-paths if* C^* *is a 2-bracelet.*

Proof. Again, say for the sake of contradiction that we have an optimal completion (Π', Γ') for which a component C' of $\mathrm{B}(\Pi', \Gamma')$ contains two or more $\{\pi, \gamma\}$-paths. If C' is not a 2-bracelet, then it must contain $\{\pi, \gamma\}$-paths P_1 and P_2 that are linked by precisely one new adjacency. Say that P_1 joins π-open vertex v to γ-open vertex w and that P_2 joins π-open vertex x to γ-open vertex y. To meet the assumption that P_1 and P_2 are linked by precisely one new adjacency, suppose that $\{v, x\} \in a(\Pi')$ but $\{w, y\} \notin a(\Gamma')$, where instead $\{w, w'\}$ and $\{y, y'\}$ are in $a(\Gamma')$. Replacing these two adjacencies with $\{w, y\}$ and $\{w', y'\}$ defines a different completion Γ'' for which $\mathrm{B}(\Pi', \Gamma'')$ contains $(P_1 : P_2)$. Viewed as an operation on $\mathrm{B}(\Pi', \Gamma')$ to yield $\mathrm{B}(\Pi', \Gamma'')$, we have two cases.

First, if C' was a bracelet, then we have formed two new bracelets from C', one of which is $(P_1 : P_2)$. Otherwise, C' was a chain, in which case we have formed a chain (of the same parity) in addition to $(P_1 : P_2)$. In either case, $d_{\mathrm{DCJ}}(\Pi', \Gamma'') < d_{\mathrm{DCJ}}(\Pi', \Gamma')$, and so (Π', Γ') cannot be optimal. □

Following Lemma 6, we may only have 2-bracelets, 2-chains, and 3-chains in $\mathrm{B}(\Pi^*, \Gamma^*)$. After a simple result about 2-chain components, we will be ready to state our main result on DCJ-indel sorting.

Proposition 7. *The breakpoint graph of an optimal completion cannot have one 2-chain joining two odd* π*-paths and another 2-chain joining two even* π*-paths. The same holds for* γ*-paths.*

Proof. Once again, proceed by contradiction and assume that (Π', Γ') is an optimal completion with such 2-chains $[P_1 : P_2]$ and $[P_3 : P_4]$. Replacing these 2-chains with $[P_1 : P_3]$ and $[P_2 : P_4]$ replaces two odd paths in $\mathrm{B}(\Pi', \Gamma')$ with two even paths; hence, (Π', Γ') cannot be optimal. □

Theorem 8. *Algorithm 9, given below, defines an* $O(N)$ *time algorithm for DCJ-indel sorting. For pairs* $\{\Pi, \Gamma\}$ *having* $\mathrm{sing}(\Pi, \Gamma) = 0$*, the DCJ-indel distance is given by the following equation:*

$$d^{\mathrm{ind}}_{\mathrm{DCJ}}(\Pi, \Gamma) = N - \left[\left(c + p^{\pi,\pi} + p^{\gamma,\gamma} + \left\lfloor \frac{p^{\pi,\gamma}}{2} \right\rfloor\right) + \frac{1}{2}\left(p^{0}_{\mathrm{even}} + \min\left\{p^{\pi}_{\mathrm{odd}}, p^{\pi}_{\mathrm{even}}\right\} + \min\left\{p^{\gamma}_{\mathrm{odd}}, p^{\gamma}_{\mathrm{even}}\right\} + \delta\right)\right] \quad (9)$$

Here, $\delta = 1$ *when* $p^{\pi,\gamma}$ *is odd and either* $p^{\pi}_{\mathrm{odd}} > p^{\pi}_{\mathrm{even}}$, $p^{\gamma}_{\mathrm{odd}} > p^{\gamma}_{\mathrm{even}}$ *or* $p^{\pi}_{\mathrm{odd}} < p^{\pi}_{\mathrm{even}}, p^{\gamma}_{\mathrm{odd}} < p^{\gamma}_{\mathrm{even}}$*; otherwise,* $\delta = 0$.

Proof. We aim to construct an optimal completion (Π^*, Γ^*) having

$$c(\Pi^*, \Gamma^*) = c + p^{\pi,\pi} + p^{\gamma,\gamma} + \left\lfloor \frac{p^{\pi,\gamma}}{2} \right\rfloor \quad (10)$$

$$p_{\mathrm{even}}(\Pi^*, \Gamma^*) = p^{0}_{\mathrm{even}} + \min\left\{p^{\pi}_{\mathrm{odd}}, p^{\pi}_{\mathrm{even}}\right\} + \min\left\{p^{\gamma}_{\mathrm{odd}}, p^{\gamma}_{\mathrm{even}}\right\} + \delta \quad (11)$$

First, we count the cycles of $\mathrm{B}(\Pi^*, \Gamma^*)$. By Lemma 5, every $\{\pi, \pi\}$-path or $\{\gamma, \gamma\}$-path of $\mathrm{B}(\Pi, \Gamma)$ must be closed into a cycle by adding a single new adjacency (Step 1 of Algorithm 9). We now claim that there exists an optimal completion containing $\left\lfloor \frac{p^{\pi,\gamma}}{2} \right\rfloor$ 2-bracelets. Note that we may always replace 3-chains $[P_1 : P_2 : P_3]$ and $[P_4 : P_5 : P_6]$ (where P_1 and P_4 are π-paths) with $[P_1 : P_4]$, $(P_2 : P_5)$, and $[P_3 : P_6]$, without increasing the DCJ distance of the associated completion because we have obtained a cycle from two paths. This argument implies

Step 2 of Algorithm 9 and produces the value of $c(\Pi^*, \Gamma^*)$ stated above.

As for the even paths of $B(\Pi^*, \Gamma^*)$, let us operate under the assumption that $p^{\pi,\gamma}$ is odd. Then after forming a maximal collection of 2-bracelets, we will be left with one additional $\{\pi, \gamma\}$-path P. We claim that (Π^*, Γ^*) will be optimal if we link as many π-paths (γ-paths) of opposite parity as possible. On the one hand, Proposition 7 states that we cannot have 2-chains $[P_1 : P_2]$ and $[P_3 : P_4]$, where P_1 and P_2 are even π-paths and P_3 and P_4 are odd π-paths. On the other hand, say that we have a 2-chain $[P_1 : P_2]$ and a 3-chain $[P_3 : P : P_4]$, where without loss of generality we assume that P_1 and P_2 are odd π-paths, P_3 is an even π-path, and P_4 is a γ-path. Replacing these chains with the chains $[P_1 : P_3]$ and $[P_2 : P : P_4]$ does not change the number of paths of even length in $B(\Pi^*, \Gamma^*)$, implying Step 3 of Algorithm 9.

As a result, all remaining π-paths must have the same parity, as must all the γ-paths; thus, we may choose any π-path and γ-path to link to P (Step 4 of Algorithm 9) and form a 3-chain. The length of this 3-chain may be even ($\delta = 1$) or odd ($\delta = 0$) depending on whether the length of its π-path and γ-path have equal parity or not. All remaining paths must therefore be 2-chains linking pairs of π-paths or pairs of γ-paths (Step 5 of Algorithm 9).

If instead $p^{\pi,\gamma}$ is even, then $\delta = 0$, and the argument for constructing an optimal completion proceeds similarly, except that no $\{\pi, \gamma\}$-paths will remain after forming a maximal collection of 2-bracelets, eliminating the need for Step 4. □

Algorithm 9. *Given genomes (Π, Γ), the following algorithm constructs an optimal completion (Π^*, Γ^*) in $O(N)$ time.*

```
0  Remove all circular singletons from $\Pi$ and $\Gamma$.
1  Close every $\{\pi,\pi\}$-path ($\{\gamma,\gamma\}$-path) into a cycle by
   adding a single new adjacency to $\Pi^*$ ($\Gamma^*$).
2  Form a maximum set of 2-bracelets.
3  Form a maximum set of even 2-chains by linking
   pairs of $\pi$-paths ($\gamma$-paths) having opposite parity.
4  If $p^{\pi,\gamma}$ is odd, then link the remaining $\{\pi,\gamma\}$-path
   with any remaining $\pi$-path and $\gamma$-path to form a
   3-chain.
5  Arbitrarily link pairs of remaining $\pi$-paths, all of
   which have the same parity, to form 2-chains. Do the
   same for remaining $\gamma$-paths.
```

The solution space of DCJ-Indel sorting

The problem of DCJ sorting is well understood, its solution space having been described in [17]. Thus, by Theorem 3, to identify the solution space of DCJ-indel sorting (an open problem), we simply need to enumerate the construction of indels of an optimal completion. We mentioned this enumeration in [12], but here we will explore the details of the calculation.

Handling circular singletons

By Proposition 1, we may consider the circular singletons of Π and Γ independently of other chromosomes; for that matter, because insertions and deletions are defined symmetrically, we may assume that Π contains k chromosomes and that Γ is the empty genome. Then by Corollary 2 and the trivial fact that any DCJ applied to Π changes the total number of chromosomes of Π by at most 1 (see [8]), we may obtain Γ from Π in k steps if and only if we perform j successive DCJs ($0 \leq j < k$), each of which fuses two circular chromosomes into one, followed by applying $k - j$ chromosome deletions.

Assuming that k is relatively small, the enumeration of all such transformations of Π into Γ poses a tedious but straightforward task, as a fusion of two circular chromosomes corresponds to a DCJ using two adjacencies from different chromosomes.

Genomes lacking circular singletons

Having handled circular singletons, we may assume that $\text{sing}(\Pi, \Gamma) = 0$. Fortunately, the lemmas presented before Theorem 8 have greatly reduced the collection of possible optimal completions, which we now continue to pare down.

Proposition 10. *Every π-path (γ-path) embedding into a 3-chain of an optimal completion must have the same parity.*

Proof. Say for the sake of contradiction that we have an optimal completion (Π', Γ') such that $B(\Pi', \Gamma')$ contains 3-chains $[P_1 : P_2 : P_3]$ and $[P_4 : P_5 : P_6]$, where P_1 and P_4 are π-paths of opposite parity. Consider the completion (Π'', Γ''), which is defined by rejoining adjacencies of (Π', Γ') to form $[P_1 : P_4]$, $(P_2 : P_5)$, and $[P_3 : P_6]$ in $B(\Pi'', \Gamma'')$. The 2-chain $[P_1 : P_4]$ must have even length, and $(P_2 : P_5)$ is a cycle; thus, $d_{\text{DCJ}}(\Pi'', \Gamma'') < d_{\text{DCJ}}(\Pi', \Gamma')$, and so (Π', Γ') cannot be optimal. □

Proposition 11. *If $p^{\pi,\gamma}$ is even, then the breakpoint graph of an optimal completion must contain a maximum set of even-length 2-chains.*

Proof. We proceed by contradiction. Say that (Π', Γ') is an optimal completion for which an odd π-path P_1 and an even π-path P_2 are contained in different components of $B(\Pi', \Gamma')$, neither of which is an even 2-chain. By Propositions 7 and 10, we may assume that P_1 and P_2 embed into an odd-length 2-chain $[P_1 : P_5]$ and a 3-chain $[P_2 : P_3 : P_4]$. Because $p^{\pi,\gamma}$ is even by Proposition 4, we must have at least one additional 3-chain $[P_6 : P_7 : P_8]$, where (again by Proposition 10) P_6 is an even-length π-path, and the γ-paths P_4 and P_8 have the same parity. With these assumptions in hand, we may rejoin adjacencies to form the four components $[P_1 : P_2]$ (even), $[P_5 : P_6]$ (even), $(P_3 : P_7)$, and $[P_4 : P_8]$ (odd), producing a cycle and two even 2-chains from our original three paths. Hence, by (4), (Π', Γ') cannot be optimal. □

We are now ready to fully describe the collection of optimal completions when $p^{\pi,\gamma}$ is even. To construct an optimal completion, after closing each $\{\pi,\pi\}$-path and $\{\gamma,\gamma\}$-path, which can be done uniquely, we must form a maximum collection of even 2-chains by Proposition 11. Recall that our aim is to maximize the statistic $c(\Pi^*,\Gamma^*) + \frac{p_{\text{even}}(\Pi^*,\Gamma^*)}{2}$, and consider the following two subcases.

Case 1. $p^{\pi,\gamma}$ is even, $p^{\pi}_{\text{odd}} \leq p^{\pi}_{\text{even}}$, and $p^{\gamma}_{\text{odd}} \geq p^{\gamma}_{\text{even}}$. First, a maximal collection of even-length 2-chains will total $p^{\pi}_{\text{odd}} + p^{\gamma}_{\text{even}}$ components, which requires simply choosing p^{π}_{odd} even-length π-paths, then matching them to odd-length π-paths. This can be achieved in A_1 ways, where

$$A_1 = \binom{p^{\pi}_{\text{even}}}{p^{\pi}_{\text{odd}}} \cdot \left(p^{\pi}_{\text{odd}}\right)! = \mathrm{P}\left(p^{\pi}_{\text{even}}, p^{\pi}_{\text{odd}}\right) \tag{12}$$

Next, we follow the same method for forming even-length 2-chains by linking γ-paths of opposite parity, yielding B_1 total matchings:

$$B_1 = \mathrm{P}\left(p^{\gamma}_{\text{odd}}, p^{\gamma}_{\text{even}}\right) \tag{13}$$

Here, we use $\mathrm{P}(n,k)$ to denote the partial permutation statistic: $\mathrm{P}(n,k) = \frac{n!}{(n-k)!}$. We will be left with $p^{\pi}_{\text{even}} - p^{\pi}_{\text{odd}}$ even π-paths and $p^{\gamma}_{\text{odd}} - p^{\gamma}_{\text{even}}$ odd γ-paths. It is impossible to create any more even-length paths in $\mathrm{B}(\Pi^*,\Gamma^*)$, and so we must form a maximum collection of $\frac{p^{\pi,\gamma}}{2}$ 2-bracelets from the $\{\pi,\gamma\}$-paths:

$$C_1 = \left(p^{\pi,\gamma}_{\text{even}} - 1\right)!! = \left(p^{\pi,\gamma}_{\text{even}} - 1\right)\left(p^{\pi,\gamma}_{\text{even}} - 3\right)\cdots(5)(3)(1) \tag{14}$$

Note the definition of double factorial. Finally, we link arbitrary remaining π-paths to each other and arbitrary remaining γ-paths to each other:

$$D_1 = \left(p^{\pi}_{\text{even}} - p^{\pi}_{\text{odd}} - 1\right)!! \cdot \left(p^{\gamma}_{\text{odd}} - p^{\gamma}_{\text{even}} - 1\right)!! \tag{15}$$

By the independence of these four procedures, the total number of optimal completions is simply given by the product $A_1 \cdot B_1 \cdot C_1 \cdot D_1$.

Case 2. $p^{\pi,\gamma}$ is even, $p^{\pi}_{\text{odd}} > p^{\pi}_{\text{even}}$, and $p^{\gamma}_{\text{odd}} > p^{\gamma}_{\text{even}}$. In this case, we first form a maximum set of 2-chains:

$$A_2 = \mathrm{P}\left(p^{\pi}_{\text{odd}}, p^{\pi}_{\text{even}}\right) \cdot \mathrm{P}\left(p^{\gamma}_{\text{odd}}, p^{\gamma}_{\text{even}}\right) \tag{16}$$

We then have $p^{\pi}_{\text{odd}} - p^{\pi}_{\text{even}}$ odd-length π-paths and $p^{\gamma}_{\text{odd}} - p^{\gamma}_{\text{even}}$ odd-length γ-paths remaining. Assume without loss of generality that $p^{\pi}_{\text{odd}} - p^{\pi}_{\text{even}} \geq p^{\gamma}_{\text{odd}} - p^{\gamma}_{\text{even}}$, and set $m = \min\{p^{\pi,\gamma}, p^{\gamma}_{\text{odd}} - p^{\gamma}_{\text{even}}\}$. We may attain the formula in (9) if and only if we form $2j$ even-length 3-chains for some integer j satisfying $0 \leq j \leq \frac{m}{2}$, then create $\frac{p^{\pi,\gamma}}{2} - j$ total 2-bracelets from the remaining $\{\pi,\gamma\}$-paths. Any remaining odd-length π-paths (γ-paths) must then be linked to each other to form (odd-length) 2-chains in $\mathrm{B}(\Pi^*,\Gamma^*)$. The number of such possibilities can be counted by the following statistic B_2:

$$B_2 = \sum_{j=0}^{m/2} \binom{p^{\pi}_{\text{odd}} - p^{\pi}_{\text{even}}}{2j} \binom{p^{\gamma}_{\text{odd}} - p^{\gamma}_{\text{even}}}{2j} \binom{p^{\pi,\gamma}}{2j} \left[(2j)!\right]^2 \cdot \left(p^{\pi}_{\text{odd}} - p^{\pi}_{\text{even}} - 2j - 1\right)!! \left(p^{\gamma}_{\text{odd}} - p^{\gamma}_{\text{even}} - 2j - 1\right)!! \left(p^{\pi,\gamma} - 2j - 1\right)!! \tag{17}$$

Again, the two statistics can be carried out independently, yielding $A_2 \cdot B_2$ total optimal completions.

In both of the first two cases, reversing the inequalities will lead to analogous arguments. For the next two cases, suppose instead that $p^{\pi,\gamma}$ is odd, and select a single $\{\pi,\gamma\}$-path P that must belong to a 3-chain.

Case 3. $p^{\pi,\gamma}$ is odd, $p^{\pi}_{\text{odd}} < p^{\pi}_{\text{even}}$, and $p^{\gamma}_{\text{odd}} > p^{\gamma}_{\text{even}}$. Note that there are $A_3 = p^{\pi,\gamma}$ total ways to select a $\{\pi,\gamma\}$-path P. Of the four possibilities for the parity of the paths to which P may be linked to form a 3-chain, one may wish to verify that the only way we cannot attain the maximum in (9) is if we link P to an odd-length π-path and an even-length γ-path. Thus, we arrive at three mutually exclusive subcases.

In our first subcase, P is linked to an even-length π-path and an odd-length γ-path:

$$B_3 = p^{\pi}_{\text{even}} \cdot p^{\gamma}_{\text{odd}} \tag{18}$$

We now have an even number of $\{\pi,\gamma\}$-paths remaining and have reduced our problem to a simpler one that falls under Case 1 above, from which we may obtain some number C_3 of optimal completions.

In the second subcase, we join P to an odd-length π-path and an odd-length γ-path. First, select two such paths:

$$D_3 = p^{\pi}_{\text{odd}} \cdot p^{\gamma}_{\text{odd}} \tag{19}$$

Again we have reduced the problem to a subproblem falling under Case 1, from which we may obtain E_3 total optimal completions. In our third and final subcase, we join P to an even π-path and an even γ-path:

$$F_3 = p^{\pi}_{\text{even}} \cdot p^{\gamma}_{\text{even}} \tag{20}$$

Say that applying Case 1 to the resulting subcase in which $p^{\pi,\gamma}$ is even yields G_3 total optimal completions. Then by independence, the total number of optimal completions over all three subcases will be given by $A_3 \cdot (B_3 \cdot C_3 + D_3 \cdot E_3 + F_3 \cdot G_3)$.

Case 4. $p^{\pi,\gamma}$ is odd, $p^{\pi}_{\text{odd}} > p^{\pi}_{\text{even}}$, and $p^{\gamma}_{\text{odd}} > p^{\gamma}_{\text{even}}$. Having selected P from the $A_4 = p^{\pi,\gamma}$ total $\{\pi, \gamma\}$-paths, one may verify that the only way we can achieve the maximum in (9) is by linking P to an odd-length π-path and an odd-length γ-path, of which there are $B_4 = p^{\pi}_{\text{odd}} \cdot p^{\gamma}_{\text{odd}}$ total choices. We have therefore reduced our problem of linking components of $\mathrm{B}(\Pi, \Gamma)$ to a smaller problem, falling under Case 2, for which $p^{\pi,\gamma}$ is even. If there are C_4 total solutions to this smaller problem, then the number of optimal completions is given by $A_4 \cdot B_4 \cdot C_4$.

As in the first two cases, reversing the inequalities defining Cases 3 and 4 will result in analogous arguments.

Conclusions

In this paper, we have demonstrated how the problem of DCJ-indel sorting, first solved in [11], can equally be handled via direct inspection of the breakpoint graph. Unfortunately, we still do not see a natural correspondence between the two approaches to DCJ-indel sorting, which appear to be at odds because their definitions of indels are equivalent but motivated differently.

Furthermore, modeling an indel as a circular chromosome resulting from a DCJ has uncovered the solution space of DCJ-indel sorting, thus resolving an open problem. We wonder if other operations could be adapted to a similar model to yield a straightforward calculation of other genomic distances involving indels. We are also curious whether this model applies to the case of finding a minimum-cost transformation of one genome into another as we vary the parameter associated with the (constant) indel cost.

Endnotes

[a] This definition allows $\mathrm{B}(\Pi, \Gamma)$ to contain cycles of length 2.

[b] In particular, this requirement bars the trivial transformation of Π into Γ in which every chromosome from Π is deleted, and then all the chromosomes of Γ inserted.

[c] Note that v cannot be simultaneously π- and γ-open, although it may be a telomere of both Π and Γ or be π-open and a telomere of Γ (in both cases, v is an isolated vertex of $\mathrm{B}(\Pi, \Gamma)$, i.e., a path of length 0).

Competing interests

The author declares that he has no competing interests.

Acknowledgements

The author would like to acknowledge the support of Pavel Pevzner (UC San Diego Department of Computer Science), who offered guidance during the drafting of the manuscript.

Received: 21 December 2012 Accepted: 7 February 2013
Published: 1 March 2013

References

1. Dobzhansky T, Sturtevant AH: **Inversions in the chromosomes of drosophila pseudoobscura.** *Genetics* 1938, **23:**28–64.
2. Fertin G, Labarre A, Rusu I, Tannier E, Vialette S: *Combinatorics of Genome Rearrangements*. Cambridge: MIT Press; 2009.
3. Harry Dweighter (pseudonym of Goodman J): **Problem E2569.** *Am Math Mon* 1975, **82:**1010.
4. Gates WH, Papadimitriou CH: **Bounds for sorting by prefix reversal.** *Discrete Math* 1979, **27:**47–57. [http://www.sciencedirect.com/science/article/pii/0012365X79900682]
5. Heydari MH, Sudborough I: **On the diameter of the pancake network.** *J Algorithms* 1997, **25:**67–94. [http://www.sciencedirect.com/science/article/pii/S0196677497908749]
6. Chitturi B, Fahle W, Meng Z, Morales L, Shields C, Sudborough I, Voit W: **An upper bound for sorting by prefix reversals.** *Theor Comput Sci* 2009, **410**(36):3372–3390. [http://www.sciencedirect.com/science/article/pii/S0304397508003575]. [Graphs, Games and Computation: Dedicated to Professor Burkhard Monien on the Occasion of his 65th Birthday].
7. Bulteau L, Fertin G, Rusu I: **Pancake flipping is hard.** *CoRR*. preprint, **abs/1111.0434**.
8. Yancopoulos S, Attie O, Friedberg R: **Efficient sorting of genomic permutations by translocation, inversion and block interchange.** *Bioinformatics* 2005, **21**(16):3340–3346. [http://bioinformatics.oxfordjournals.org/content/21/16/3340.abstract]
9. Bergeron A, Mixtacki J, Stoye J: **A unifying view of genome rearrangements.** *WABI 2006. LNCS (LNBI)* 2006, **4175:**163–173.
10. Yancopoulos S, Friedberg R: **DCJ path formulation for genome transformations which include insertions, deletions, and duplications.** *J Comput Biol* 2009, **16**(10):1311–1338.
11. Braga MDV, Willing E, Stoye J: **Genomic distance with DCJ and indels.** *Proc 10th Int Conf Algorithms Bioinformatics* 2010:90–101. [http://portal.acm.org/citation.cfm?id=1885783.1885793]
12. Compeau PEC: **A simplified view of DCJ-Indel distance.** In *WABI Volume 7534 of* Lecture Notes in Computer Science. Edited by Raphael BJ, Tang J: Springer; 2012:365–377. [http://dblp.uni-trier.de/db/conf/wabi/wabi2012.html#Compeau12]
13. Bafna V, Pevzner PA: **Genome rearrangements and sorting by reversals.** *SIAM J Comput* 1996, **25**(2):272–289.
14. Tannier E, Zheng C, Sankoff D: **Multichromosomal median and halving problems under different genomic distances.** *BMC Bioinformatics* 2009, **10:**120. [http://www.biomedcentral.com/1471-2105/10/120]
15. Braga M, Machado R, Ribeiro L, Stoye J: **On the weight of indels in genomic distances.** *BMC Bioinformatics* 2011, **12**(Suppl 9):S13. [http://www.biomedcentral.com/1471-2105/12/S9/S13]
16. Ma J, Ratan A, Raney BJ, Suh BB, Miller W, Haussler D: **The infinite sites model of genome evolution.** *Proc Natl Acad Sci USA* 2008, **105**(38):14254–14261. [http://dx.doi.org/10.1073/pnas.0805217105]
17. Braga MD, Stoye J: **The solution space of sorting by DCJ.** *J Comput Biol* 2010, **17**(9):1145–1165.

doi:10.1186/1748-7188-8-6
Cite this article as: Compeau: DCJ-Indel sorting revisited. *Algorithms for Molecular Biology* 2013 **8**:6.