**BMC
Bioinformatics**

PROCEEDINGS

Open Access

On pairwise distances and median score of three genomes under DCJ

Sergey Aganezov Jr[1,2], Max A Alekseyev[1,3]*

Abstract

In comparative genomics, the rearrangement distance between two genomes (equal the minimal number of genome rearrangements required to transform them into a single genome) is often used for measuring their evolutionary remoteness. Generalization of this measure to three genomes is known as the *median score* (while a resulting genome is called *median genome*). In contrast to the rearrangement distance between two genomes which can be computed in linear time, computing the median score for three genomes is NP-hard. This inspires a quest for simpler and faster approximations for the median score, the most natural of which appears to be the halved sum of pairwise distances which in fact represents a lower bound for the median score.

In this work, we study relationship and interplay of pairwise distances between three genomes and their median score under the model of Double-Cut-and-Join (DCJ) rearrangements. Most remarkably we show that while a rearrangement may change the sum of pairwise distances by at most 2 (and thus change the lower bound by at most 1), even the most "powerful" rearrangements in this respect that increase the lower bound by 1 (by moving one genome farther away from each of the other two genomes), which we call *strong*, do not necessarily affect the median score. This observation implies that the two measures are not as well-correlated as one's intuition may suggest.

We further prove that the median score attains the lower bound exactly on the triples of genomes that can be obtained from a single genome with strong rearrangements. While the sum of pairwise distances with the factor 2/3 represents an upper bound for the median score, its tightness remains unclear. Nonetheless, we show that the difference of the median score and its lower bound is not bounded by a constant.

Background

The number of large-scale rearrangements (such as reversals, translocations, fissions, and fusions) between two genomes is often used as a measure of their evolutionary remoteness. The minimal number of such rearrangements required to transform one genome into the other is called *rearrangement distance*. Computing rearrangement distances between the genomes of interest is often a prerequisite for their comparative analysis (e.g., phylogeny reconstruction).

Double-Cut-and-Join (DCJ) rearrangements [1,2] (also known as *2-breaks* [3]) represent a convenient model of reversals, translocations, fissions, and fusions, which allows one to compute the corresponding *DCJ distance* between two genomes in linear time.

Phylogeny reconstruction for three given genomes involves reconstruction of their *median genome* that minimizes the total distance from the given genomes. This minimal total distance, called the *median score* [4], represents a natural generalization of the DCJ distance to the case of three genomes. In contrast to DCJ distance between two genomes, computing the median score of three genomes is NP-hard [4,5]. While there exist exact [6,7] and heuristic [8-10] algorithms for this problem, they can hardly be used for large genomes. This inspires

* Correspondence: maxal@cse.sc.edu
[1]Algorithmic Biology Laboratory, St. Petersburg Academic University, St. Petersburg, Russia
Full list of author information is available at the end of the article

a quest for simpler and faster approximations for the median score.

The simplest and easily computable approximation for the median score of three genomes is given by the sum of their pairwise DCJ distances, which we call the *triangle score*. In this work, we study the tightness of this approximation. In particular, we show that with the factor 1/2 it represents a lower bound and with the factor 2/3 it represents an upper bound for the median score. We further prove that the lower bound is attained exactly for the triples that can be obtained from a single genome by *strong* rearrangements that increase the triangle score by 2 (by moving one genome farther away from each of the other two genomes). In other words, strong rearrangements are those that increase the lower bound by 1. From this perspective, it natural to expect that strong rearrangements always increase the median score. However, we disprove this expectation with a counterexample.

While tightness of the upper bound remains unclear, we remark that a better upper bound for the median score may improve performance of algorithms for computing median score based on the adequate subgraph decomposition [6,7]. Still, we make an initial step in this direction by proving that there is no upper bound equal the lower bound plus a constant.

Methods

Breakpoint graphs and genome rearrangements

In this work, we focus on circular genomes consisting of one or more circular chromosomes. A circular genome on a set of n genes (say, $\{1, 2,..., n\}$) can be represented as a perfect matching on $2n$ vertices [11,12] where each gene is represented with a pair of vertices, corresponding to the gene's extremities: "head" and "tail"; while each adjacency between two genes in the genome is represented with an edge between respective extremities. *Breakpoint* graph of genomes $A_1, A_2,..., A_k$, denoted BG $(A_1,..., A_k)$ is defined as the superposition of k perfect matchings representing given genomes, each of its own color [12]. We refer to edges representing adjacencies in the genome A_i as A_i-*edges* ($i = 1, 2,..., k$). When all genomes $A_1, A_2,..., A_k$ are identical, their breakpoint graph is called an *identity breakpoint graph*. Every identity breakpoint graph consists of *trivial multicycles* formed by k parallel edges of all colors.

A *DCJ* rearrangement in genome A replaces a pair of A-edges with another pair of A-edges forming matching on the same four vertices. The *DCJ distance* between genomes A and B on the same set of n genes, denoted with $d_{dcj}(A, B)$, is defined as the minimal number of DCJs required to transform one genome into the other. The DCJ distance $d_{dcj}(A, B)$ is closely connected with the number $c(A, B)$ of alternating cycles (i.e., cycles with

edges of alternating colors) in the breakpoint graph BG (A, B) by the formula: $d_{dcj}(A, B)= n - c(A, B)$. We remark that $c(A, B)$ may range from 1 to n, implying that $0 \le d_{dcj}(A, B) \le n - 1$. A single DCJ in A or B can change $d_{dcj}(A, B)$ by at most 1 [1,3,13].

Let A, B, C be genomes on a set of n genes. Their breakpoint graph BG(A, B, C) is formed by A-edges, B-edges, and C-edges so that each pair of genomes define alternating cycles, called respectively AB-cycles, AC-cycles, and BC-cycles (Figure 1). We further define the *triangle score* ts(A, B, C) as the sum of pairwise DCJ distances:

$$\text{ts}(A, B, C) = d_{dcj}(A, B) + d_{dcj}(A, C) + d_{dcj}(B, C).$$

Since a DCJ in one of the genomes can change each of the two corresponding distances by at most 1, it can change ts(A, B, C) by at most 2.

Lemma 1. *Let A, B, C be genomes on the same set of n genes such that $d_{dcj}(A, B)$ and $d_{dcj}(A, C)$ are less than $n - 1$. Then in BG(A, B, C) there exists a pair of A-edges that belong to two distinct AB-cycles and two distinct AC-cycles.*
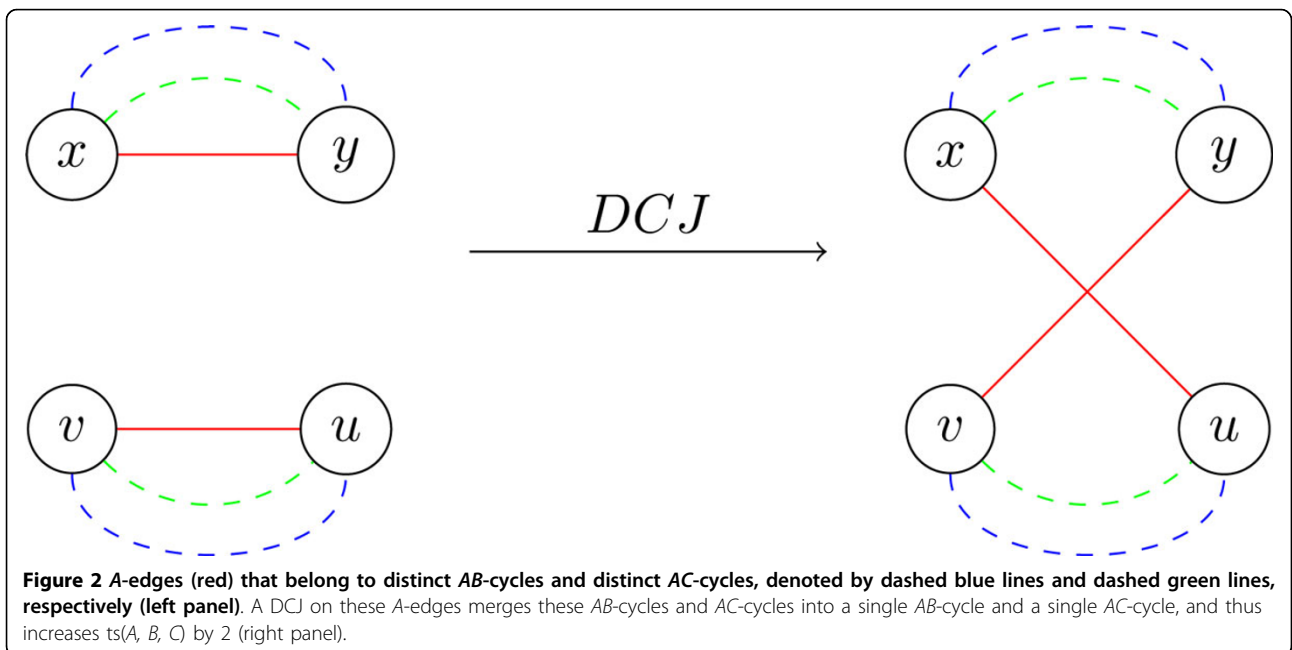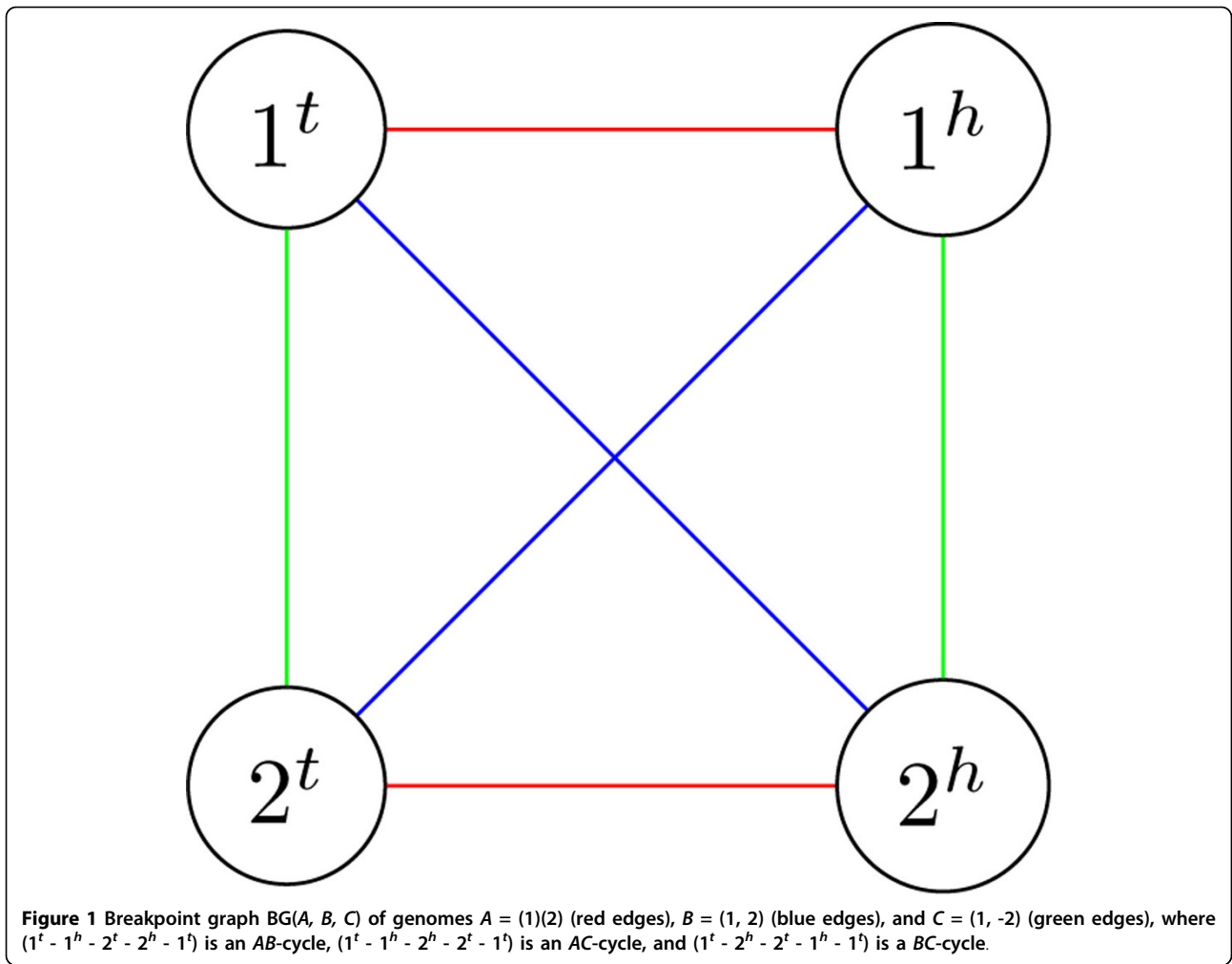
Proof. Since $d_{dcj}(A, B) < n - 1$, there exist at least two distinct AB-cycles in the breakpoint graph BG(A, B, C). Therefore, the AB-cycles define a partition of the set of A-edges S_A into two or more nonempty subsets: $S_A = P_1 \cup P_2 \cup...$

Similarly, since $d_{dcj}(A, C) < n - 1$, there exist at least two distinct AC-cycles in BG(A, B, C) so that the AC-cycles define a partition of S_A into two or more nonempty subsets: $S_A = Q_1 \cup Q_2 \cup....$ Intersecting the subsets in the two partitions, we get a partition of S_A into subsets, each consisting of A-edges that belong to the same AB-cycle and the same AC-cycle: $S_A = \cup_{i, j}(P_i \cap Q_j)$.

Suppose that there is no required pair of A-edges, implying that for any two non-empty intersections $P_i \cap Q_j$ and $P_{i'} \cap Q_{j'}$, we have either $i = i'$ or $j = j'$. Without loss of generality, assume that $P_1 \cap Q_1$ is non-empty. Then for every $i > 1$ and $j > 1$, $P_i \cap Q_j$ must be empty, implying that $P_i \subset Q_1$. In particular, $P_2 \cap Q_1 = P_2$ is non-empty and by the same reasoning, we have $P_1 \subset Q_1$. Therefore, $P_i \subset Q_1$ for all i, implying that $Q_1 = S_A$, a contradiction to non-emptiness of Q_2. This contradiction proves that a required pair of A-edges exists.

Theorem 2. *If between three genomes A, B, C on the same set of n genes at least two pairwise DCJ distances are less than $n - 1$, then there exists a DCJ (called strong) that increases ts(A, B, C) by 2.*

Proof. Without loss of generality, we assume that $d_{dcj}(A, B) < n - 1$ and $d_{dcj}(A, C) < n - 1$. By Lemma 1, there are A-edges (x, y) and (u, v) that belong to distinct AB-cycles and distinct AC-cycles in BG(A, B, C). Using any DCJ on these edges (Figure 2), we decrease the number of AB-cycles as well as the number of AC-cycles by 1.

Figure 1 Breakpoint graph BG(A, B, C) of genomes A = (1)(2) (red edges), B = (1, 2) (blue edges), and C = (1, -2) (green edges), where $(1^t - 1^h - 2^t - 2^h - 1^t)$ is an *AB*-cycle, $(1^t - 1^h - 2^h - 2^t - 1^t)$ is an *AC*-cycle, and $(1^t - 2^h - 2^t - 1^h - 1^t)$ is a *BC*-cycle.

Figure 2 A-edges (red) that belong to distinct AB-cycles and distinct AC-cycles, denoted by dashed blue lines and dashed green lines, respectively (left panel). A DCJ on these *A*-edges merges these *AB*-cycles and *AC*-cycles into a single *AB*-cycle and a single *AC*-cycle, and thus increases ts(*A, B, C*) by 2 (right panel).

Since *BC*-cycles remain intact, this DCJ increases ts(*A, B, C*) by 2.

Theorem 3. *Let (p, q, r) be a triple of integers from the interval* [0, *n* - 1], *satisfying the triangle inequality. There exist three genomes on a set of n genes whose pairwise DCJ distances are (p, q, r). Moreover, these genomes can be obtained with $\lfloor \frac{p+q+r}{2} \rfloor$ strong DCJs and possibly one other DCJ (when p + q + r is odd) from a single genome.*

Proof. Without loss of generality, we assume that $p \leq q \leq r$ and that $p = d_{dcj}(A, B)$, $q = d_{dcj}(A, C)$, and $r = d_{dcj}(B, C)$ where *A, B, C* are the genomes being constructed.

If $p + q + r$ is even, we start with *A, B, C* being the same genome and notice that

$$(p, q, r) = \frac{p+q-r}{2} \cdot (1, 1, 0) + \frac{p+r-q}{2} \cdot (1, 0, 1) + \frac{q+r-p}{2} \cdot (0, 1, 1)$$

where by the triangle inequality all coefficients are non-negative. This identity instructs us to apply $\frac{p+q-r}{2}$ strong DCJs to the genome *A* (increasing both $d_{dcj}(A, B)$ and $d_{dcj}(A, C)$), $\frac{p+r-q}{2}$ strong DCJs to genome *B*, and $\frac{q+r-p}{2}$ strong DCJs to genome *C*. Existence of such strong DCJs is guaranteed by Theorem 2.

If $p + q + r$ is odd, then we have $p > 0$ as otherwise the triangle inequality would imply $q = r$ and thus even $p + q + r$. In this case we start with *A, B, C* being three genomes with pairwise DCJ distances (1, 1, 1) such that BG(*A, B, C*) consists of trivial multicycles, except for the vertices 1^t, 1^h, 2^t, and 2^h connected as in Figure 1. It is easy to see that these genomes can be obtained from the same

genome by one strong DCJ and one non-strong DCJ. We further increase the pairwise DCJ distances between genomes *A, B, C* by (*p', q', r'*) = (*p* - 1, *q* - 1, *r* - 1) with $\frac{p'+q'+r'}{2}$ strong DCJs as above (notice that $p' + q' + r'$ is even and (*p', q', r'*) satisfies the triangle inequality). Therefore, the total number of strong DCJs in this case is

$$1 + \frac{p' + q' + r'}{2} = \frac{p + q + r - 1}{2} = \left\lfloor \frac{p+q+r}{2} \right\rfloor.$$

While all triples of pairwise DCJ distances are achievable with strong DCJs, not all breakpoint graphs of three genomes can be constructed from an identity breakpoint graph this way. In particular, Figure 3 gives an example of breakpoint graph BG(*A, B, C*) such that ts(*A, B, C*) cannot be decreased by 2 with a DCJ. In this example, we have ts(*A, B, C*) = 6 but there is no sequence of three DCJs that would produce BG(*A, B, C*) from an identity breakpoint graph.
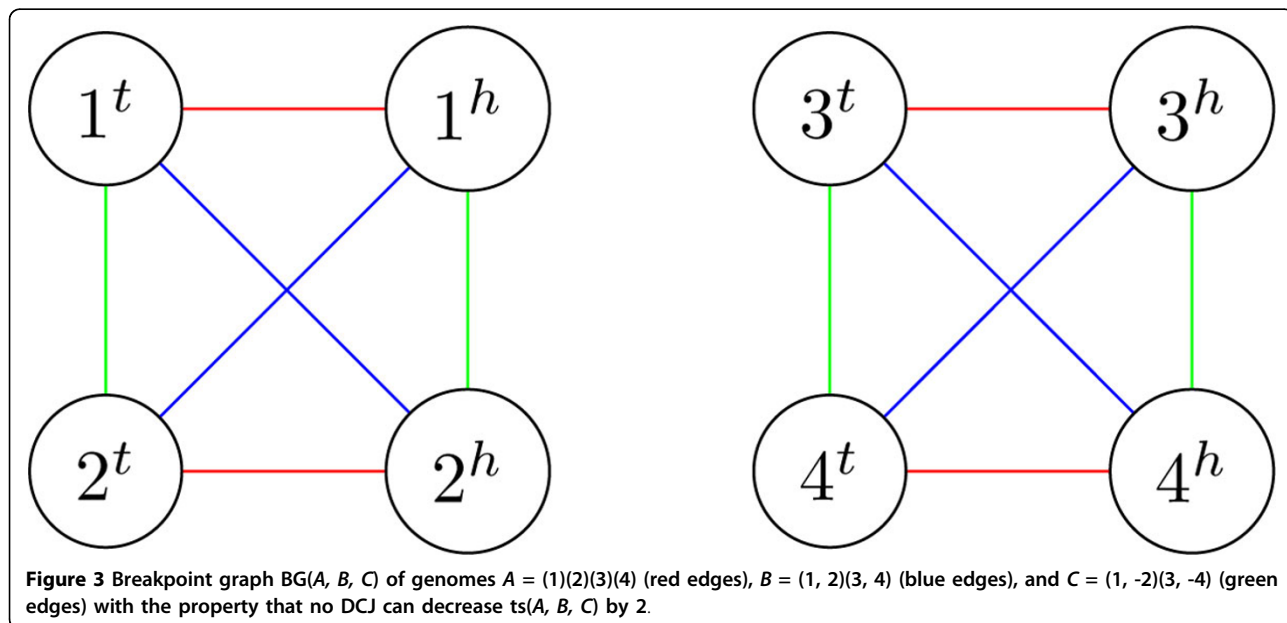
In the next section we demonstrate how DCJs on three genomes can affect their median score.

Strong rearrangements and median score

Median genome problem for given genomes *A, B, C* is to find a genome (which is called *median genome* and may not be unique) that attains the *median* score [4]:

$$\text{ms}(A, B, C) = \min_M d_{dcj}(A, M) + d_{dcj}(B, M) + d_{dcj}(C, M).$$

The median problem can be alternatively posed as finding the minimal number (equal ms(*A, B, C*)) of DCJs required to transform the genomes *A, B, C* into a single (median) genome (or, vice versa, to obtain *A, B, C*

Figure 3 Breakpoint graph BG(*A, B, C*) of genomes *A* = (1)(2)(3)(4) (red edges), *B* = (1, 2)(3, 4) (blue edges), and *C* = (1, -2)(3, -4) (green edges) with the property that no DCJ can decrease ts(*A, B, C*) by 2.

from a single genome). In fact, this formulation further generalizes and becomes particularly useful for phylogeny reconstruction of a larger number of genomes [12].

From perspective of this formulation, it becomes important to realize what triples of genomes can be obtained from a single genome with strong DCJs. We start with proving a helpful lemma and bounds on the median score in terms of the triangle score.

Lemma 4. *For three genomes on the same set of genes, a DCJ in one of the genomes may change their median score by at most 1.*

Proof. Let A, B, C be genomes of the same set of genes. Consider a DCJ in a genome A and denote the resulting genome by A'. Let M be a median genome of the genomes A, B, C so that

$$\mathrm{ms}\,(A, B, C) = \mathrm{d}_{dcj}\,(A, M) + \mathrm{d}_{dcj}\,(B, M) + \mathrm{d}_{dcj}\,(C, M)\,.$$

Clearly, $\mathrm{d}_{dcj}(A, M) \geq \mathrm{d}_{dcj}(A', M) - 1$ and thus

$$\begin{aligned}
\mathrm{ms}\,(A, B, C) &= \mathrm{d}_{dcj}\,(A, M) + \mathrm{d}_{dcj}\,(B, M) + \mathrm{d}_{dcj}\,(C, M) \\
&\geq \mathrm{d}_{dcj}\,(A', M) + \mathrm{d}_{dcj}\,(B, M) + \mathrm{d}_{dcj}\,(C, M) - 1 \geq \mathrm{ms}\,(A', B, C) - 1,
\end{aligned}$$

i.e., $\mathrm{ms}(A', B, C) - \mathrm{ms}(A, B, C) \leq 1$. Symmetrically, we also have $\mathrm{ms}(A, B, C) - \mathrm{ms}(A', B, C) \leq 1$ and thus $|\mathrm{ms}(A, B, C) - \mathrm{ms}(A', B, C)| \leq 1$.

Theorem 5. *For genomes A, B, C on the same set of genes, we have*

$$\frac{1}{2} \cdot \mathrm{ts}\,(A, B, C) \leq \mathrm{ms}\,(A, B, C) \leq \frac{2}{3} \cdot \mathrm{ts}\,(A, B, C)\,.$$

Proof. Consider a transformation of each of the genomes A, B, C into a median genome with DCJs. The total number of DCJs in this transformation is $\mathrm{ms}(A, B, C)$. Since each DCJ decreases $\mathrm{ts}(A, B, C)$ by at most 2, we have $\mathrm{ts}(A, B, C) \leq 2 \cdot \mathrm{ms}(A, B, C)$, implying that $1/2 \cdot \mathrm{ts}(A, B, C) \leq \mathrm{ms}(A, B, C)$.

On the other hand, the number of DCJs in any transformation of the genomes A, B, C into the genome A is at least $\mathrm{ms}(A, B, C)$, implying that $\mathrm{d}_{dcj}(B, A) + \mathrm{d}_{dcj}(C, A) \geq \mathrm{ms}(A, B, C)$. Similarly, we have $\mathrm{d}_{dcj}(A, B) + \mathrm{d}_{dcj}(C, B) \geq \mathrm{ms}(A, B, C)$ and $\mathrm{d}_{dcj}(A, C) + \mathrm{d}_{dcj}(B, C) \geq \mathrm{ms}(A, B, C)$. Summing up these three inequalities, we get $2 \cdot \mathrm{ts}(A, B, C) \geq 3 \cdot \mathrm{ms}(A, B, C)$ and thus $\mathrm{ms}(A, B, C) \leq 2/3 \cdot \mathrm{ts}(A, B, C)$.

We remark that the lower bound and a slightly better upper bound $\mathrm{ms}(A, B, C) \leq \min\{\mathrm{d}_{dcj}(A, B) + \mathrm{d}_{dcj}(A, C), \mathrm{d}_{dcj}(A, B) + \mathrm{d}_{dcj}(B, C), \mathrm{d}_{dcj}(A, C) + \mathrm{d}_{dcj}(B, C)\}$ was also used in [7].

The following theorem classifies all triples of genomes for which the median score coincides with its lower bound and links them with the genomes constructed in Theorem 3.

Theorem 6. *For genomes A, B, C on the same set of genes, $\mathrm{ms}(A, B, C) = 1/2 \cdot \mathrm{ts}(A, B, C)$ if and only if A, B, C can be obtained from a single genome with strong DCJs.*

Proof. Suppose that $\mathrm{ms}(A, B, C) = 1/2 \cdot \mathrm{ts}(A, B, C)$. Let M be a median genome of the genomes A, B, C. Then A, B, C can be obtained from M with $\mathrm{ms}(A, B, C) = 1/2 \cdot \mathrm{ts}(A, B, C)$ DCJs. This transformation increases the triangle score from $\mathrm{ts}(M, M, M) = 0$ to $\mathrm{ts}(A, B, C)$. Since each of $1/2 \cdot \mathrm{ts}(A, B, C)$ DCJs can increase the triangle score by at most 2, they all must be strong.

Vice versa, suppose that A, B, C are obtained from a single genome with strong DCJs. Lemma 4 implies that a strong DCJ does not increase the difference between the median score and its lower bound. Since the transformation starts with the median score equal its lower bound (i.e., their difference is 0), they further remain equal along the whole transformation, resulting in $\mathrm{ms}(A, B, C) = 1/2 \cdot \mathrm{ts}(A, B, C)$.

It remains unclear how tight is the upper bound given in Theorem 5, while a better upper bound may improve performance of algorithms for computing median score based on the adequate subgraph decomposition [6,7]. Below we prove however that the upper bound cannot be equal to the lower bound plus a constant.

Theorem 7. *The difference $\mathrm{ms}(A, B, C) - 1/2 \cdot \mathrm{ts}(A, B, C)$ of the median score and its lower bound is not bounded from above by a constant.*

Proof. To prove the theorem, for every $n = 1, 2, \ldots$, we will construct three genomes A_n, B_n, C_n on the same $4n$ genes for which $\mathrm{ms}(A_n, B_n, C_n) - 1/2 \cdot \mathrm{ts}(A_n, B_n, C_n) = n$.

We start with genomes $A_1 = (1)\,(2)\,(3)\,(4)$, $B_1 = (1, 2)\,(3, 4)$, and $C_1 = (1, -2)\,(3, -4)$. The breakpoint graph BG(A_1, B_1, C_1) consists of two strongly adequate subgraphs (Figure 3). We have $\mathrm{ms}(A_1, B_1, C_1) = 4$ and $\mathrm{ts}(A_1, B_1, C_1) = 6$, resulting in $\mathrm{ms}(A_1, B_1, C_1) - 1/2 \cdot \mathrm{ts}(A_1, B_1, C_1) = 1$.

To construct BG(A_n, B_n, C_n) we take n copies of BG(A_1, B_1, C_1) and relabel their vertices appropriately. In particular, for $n = 2$ we get genomes $A_2 = (1)\,(2)\,(3)\,(4)\,(5)\,(6)\,(7)\,(8)$, $B_2 = (1, 2)\,(3, 4)\,(5, 6)\,(7, 8)$, and $C_2 = (1, -2)\,(3, -4)\,(5, -6)\,(7, -8)$. Since edges of a median genome do not connect strongly adequate subgraphs of the breakpoint graph [6,7], every copy of BG(A_1, B_1, C_1) in BG(A_n, B_n, C_n) contributes 4 to the median score. It is also clear that every copy of BG(A_1, B_1, C_1) contributes 6 to the triangle score, implying that $\mathrm{ms}(A_n, B_n, C_n) - 1/2 \cdot \mathrm{ts}(A_n, B_n, C_n) = 4n - 3n = n$.

We conclude our analysis with the last but not the least observation about the lower bound $1/2 \cdot \mathrm{ts}(A, B, C) \leq \mathrm{ms}(A, B, C)$. According to Lemma 4, a DCJ in one of the genomes A, B, C can either increase/decrease the right hand side of this inequality (i.e., the median score) by 1, or keep it intact. For a strong DCJ (moving one genome farther away from each of the other two genomes), the left hand side of the inequality is increased by 1. From this perspective, it is very natural to expect that a strong DCJ should also increase the median score (e.g., it was so in the proof of Theorem 6). Surprisingly,

this intuition fails: Figure 4 gives a counterexample of a breakpoint graph of three genomes with a strong DCJ that does not increase the median score.
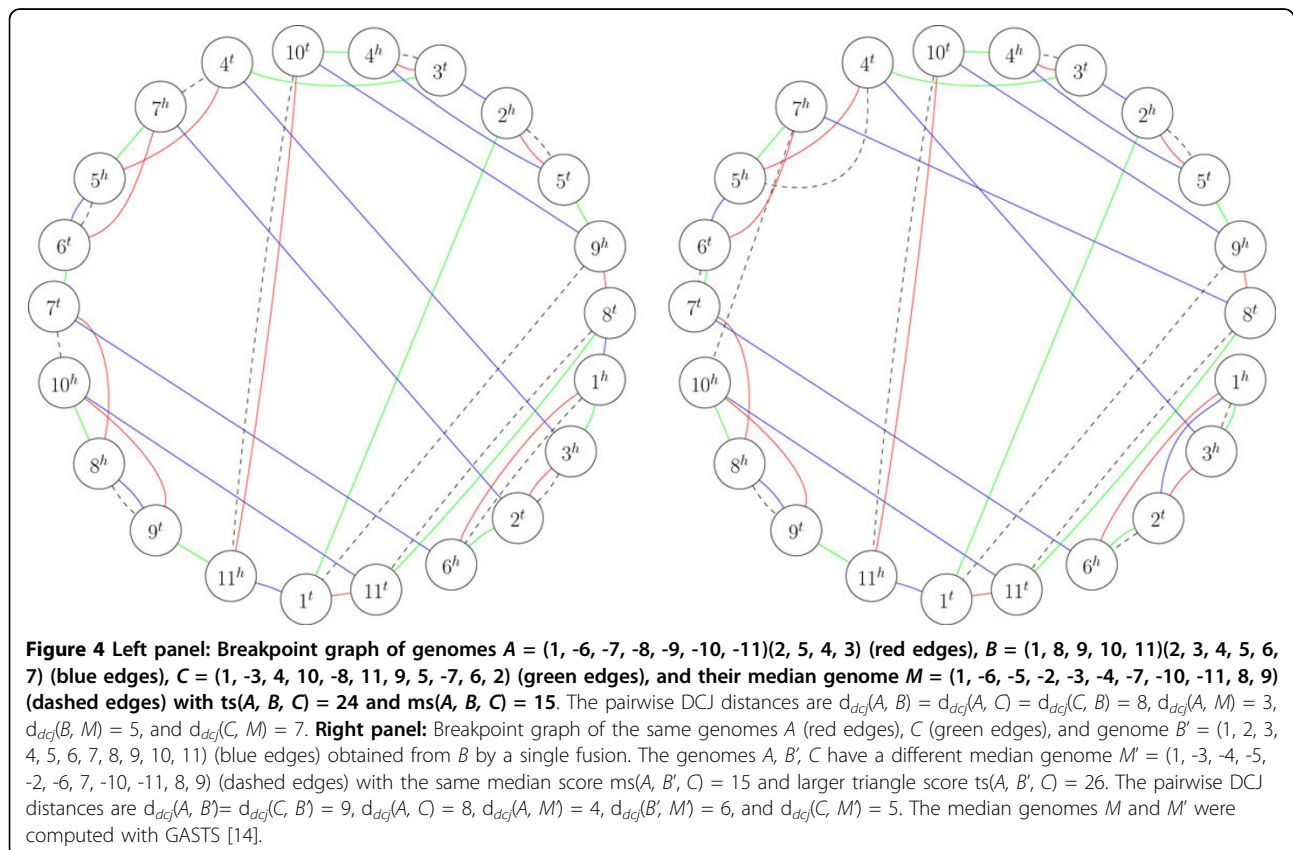
Results and discussion

We studied two measures of evolutionary remoteness of three genomes A, B, C: the triangle score $ts(A, B, C)$ (equal the sum of the pairwise rearrangement distances) and the median score $ms(A, B, C)$ (equal the minimum total rearrangement distance from a single genome). While computing $ts(A, B, C)$ takes linear time and computing $ms(A, B, C)$ is NP-hard, they are connected by the inequality $1/2 \cdot ts(A, B, C) \leq ms(A, B, C) \leq 2/3 \cdot ts(A, B, C)$ (Theorem 5) giving the lower and upper bounds for the median score in terms of the triangle score.

In view of the median genome problem as finding a transformation of the given genomes into a single genome (or a reverse transformation of a single genome into the given genomes) with the smallest number of genome rearrangements, it is important to understand how rearrangements can change the median score and its bounds. When A, B, C equal the same genome M, the median score trivially coincides with its lower and upper bounds as $ts(M, M, M) = ms(M, M, M) = 0$. Since each rearrangement may change the triangle score by at most 2 and

the median score by at most 1 (Lemma 4), we are particularly interested in strong rearrangements which increase the triangle score by 2 (and thus increase the lower bound by 1).

We showed that the median score attains its lower bound (i.e., $ms(A, B, C) = 1/2 \cdot ts(A, B, C)$) exactly on the triples of genomes that can be obtained from a single genome with strong rearrangements (Theorem 6). We proved that strong rearrangements are common enough to exist for any triple of genomes as soon as at least two of their pairwise distances are smaller than the maximum (Theorem 2) and to produce a triple of genomes with the prescribed pairwise distances (Theorem 3). From this perspective, it comes as a total surprise that strong rearrangements are not "powerful" enough to always increase the median score as illustrated by the counterexample in Figure 4. This counterexample implies that the median score and the triangle score are not as well-correlated as one's intuition may suggest.

It remains unclear how tight is the upper bound for the median score, while a better upper bound may improve performance of existing algorithms for computing the median score. Nonetheless, we made an initial step in this direction by proving that there is no upper bound equal the lower bound plus a constant (Theorem 7).

Figure 4 Left panel: Breakpoint graph of genomes A = (1, -6, -7, -8, -9, -10, -11)(2, 5, 4, 3) (red edges), B = (1, 8, 9, 10, 11)(2, 3, 4, 5, 6, 7) (blue edges), C = (1, -3, 4, 10, -8, 11, 9, 5, -7, 6, 2) (green edges), and their median genome M = (1, -6, -5, -2, -3, -4, -7, -10, -11, 8, 9) (dashed edges) with $ts(A, B, C)$ = 24 and $ms(A, B, C)$ = 15. The pairwise DCJ distances are $d_{dcj}(A, B) = d_{dcj}(A, C) = d_{dcj}(C, B)$ = 8, $d_{dcj}(A, M)$ = 3, $d_{dcj}(B, M)$ = 5, and $d_{dcj}(C, M)$ = 7. **Right panel:** Breakpoint graph of the same genomes A (red edges), C (green edges), and genome B' = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11) (blue edges) obtained from B by a single fusion. The genomes A, B', C have a different median genome M' = (1, -3, -4, -5, -2, -6, 7, -10, -11, 8, 9) (dashed edges) with the same median score $ms(A, B', C)$ = 15 and larger triangle score $ts(A, B', C)$ = 26. The pairwise DCJ distances are $d_{dcj}(A, B') = d_{dcj}(C, B')$ = 9, $d_{dcj}(A, C)$ = 8, $d_{dcj}(A, M')$ = 4, $d_{dcj}(B', M')$ = 6, and $d_{dcj}(C, M')$ = 5. The median genomes M and M' were computed with GASTS [14].

Author details

[1]Algorithmic Biology Laboratory, St. Petersburg Academic University, St. Petersburg, Russia. [2]Department of Further Mathematics, National Research University of Information Technologies, Mechanics and Optics, St. Petersburg, Russia. [3]Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA.

Competing interests

The authors declare that they have no competing interests.

References

1. Yancopoulos S, Attie O, Friedberg R: **Efficient sorting of genomic permutations by translocation, inversion and block interchange.** *Bioinformatics* 2005, **21(16)**:3340-3346.
2. Bergeron A, Mixtacki J, Stoye J: **A Unifying View of Genome Rearrangements.** *Lecture Notes in Computer Science* 2006, **4175**:163-173.
3. Alekseyev MA, Pevzner PA: **Multi-Break Rearrangements and Chromosomal Evolution.** *Theoretical Computer Science* 2008, **395(2-3)**:193-202.
4. Tannier E, Zheng C, Sankoff D: **Multichromosomal median and halving problems under different genomic distances.** *BMC Bioinformatics* 2009, **10**:120.
5. Caprara A: **The Reversal Median Problem.** *INFORMS Journal on Computing* 2003, **15**:93-113.
6. Xu AW, Sankoff D: **Decompositions of Multiple Breakpoint Graphs and Rapid Exact Solutions to the Median Problem.** *Lecture Notes in Computer Science* 2008, **5251**:25-37.
7. Xu AW: **A Fast and Exact Algorithm for the Median of three Problem: a Graph Decomposition Approach.** *Journal of Computational Biology* 2009, **16**:1369-1381.
8. Lenne R, Solnon C, Tannier E, Birattari M: **Effective Stochastic Local Search Algorithms for the Genomic Median Problem.** *Doctoral Symposium on Engineering: Stochastic Local Search Algorithms (SLSDS). IRIDIA Technical Report Series* 2007, 1-5.
9. Rajan V, Xu A, Lin Y, Swenson K, Moret B: **Heuristics for the inversion median problem.** *BMC Bioinformatics* 2010, **11(Suppl 1)**:S30.
10. Goeffon A, Nikolski M, Sherman DJ: **An efficient probabilistic population-based descent for the median genome problem.** *Proceedings of the 10th annual conference on Genetic and evolutionary computation (GECCO)* 2008, 315-322.
11. Alekseyev MA, Pevzner PA: **Are There Rearrangement Hotspots in the Human Genome?** *PLoS Computational Biology* 2007, **3(11)**:e209.
12. Alekseyev MA, Pevzner PA: **Breakpoint Graphs and Ancestral Genome Reconstructions.** *Genome Research* 2009, **19(5)**:943-957.
13. Alekseyev MA: **Multi-Break Rearrangements and Breakpoint Re-uses: from Circular to Linear Genomes.** *Journal of Computational Biology* 2008, **15(8)**:1117-1131.
14. Xu AW, Moret BME: **GASTS: Parsimony Scoring under Rearrangements.** *Lecture Notes in Computer Science* 2011, **6833**:351-363.