# On the number of breakpoint medians of random genomes

Poly H. da Silva[1,2], Arash Jamshidpey[3*], David Sankoff[4]

[1]Department of Statistics, Columbia University, 1125 Amsterdam Ave., New York, 10027, NY, USA.
[2]Irving Institute for Cancer Dynamics, Columbia University, 1190 Amsterdam Ave., New York, 10027, NY, USA.
[3]Department of Statistics, University of California, Berkeley, 367 Evans Hall, Berkeley, 94720, CA, USA.
[4] Department of Mathematics and Statistics, University of Ottawa, 150 Louis Pasteur Pvt., Ottawa, K1N 6N5, ON, Canada.

*Corresponding author(s). E-mail(s): arash.jamshidpey@berkeley.edu;
Contributing authors: phd2120@columbia.edu; sankoff@uottawa.ca;

**Abstract**

Consider $\boldsymbol{k}$ genomes $\boldsymbol{\mathcal{X}} = \{\boldsymbol{\xi_1}, \cdots, \boldsymbol{\xi_k}\}$ that are chosen uniformly and independently at random from the space of all genomes of a specified size with identical gene content. Let $\boldsymbol{\mathcal{M}}$ be the set of their breakpoint medians. The mathematical study of the size and distribution of $\boldsymbol{\mathcal{M}}$ is rather complex. In this paper, we initiate the study of the distribution of $\boldsymbol{\mathcal{M}}$, and introduce the notion of "*median inverse*" to estimate the chance of a given genome being an "*approximate*" breakpoint median of $\boldsymbol{\mathcal{X}}$. As a result, we investigate the expected number of approximate breakpoint medians of $\boldsymbol{\mathcal{X}}$, which also provides an upper bound for the expected number of medians $\mathbb{E}|\boldsymbol{\mathcal{M}}|$.

**Keywords:** Random genomes, breakpoint distance, median inverse.

## 1 Introduction

A simple but effective way to measure gene-order dissimilarity between two genomes is to count the number of their breakpoints. A breakpoint of a genome $G$ with respect to another genome $G'$ is a pair of genes adjacent in $G$ but not in $G'$. Their breakpoint distance is then defined by $d(G, G') := |\mathcal{A}_G \Delta \mathcal{A}_{G'}|/2$, where $\mathcal{A}_G$ and $\mathcal{A}_{G'}$ represent the sets of gene adjacencies of $G$ and $G'$, and $\Delta$ denotes the symmetric difference operation for sets. We also denote by $\mathcal{A}_{G,G'}$ the set of common gene adjacencies of $G$ and $G'$. Although the number of breakpoints of $G$ with respect to $G'$ may be different from that of $G'$ with respect to $G$, they are equal if both genomes have the same gene contents, where in this case, the breakpoint distance formula reduces to $d(G, G') = |\mathcal{A}_G| - |\mathcal{A}_{G,G'}|$.

Introduced in Sankoff and Blanchette (1997), the breakpoint median of $k \geq 3$ genomes $G_1, \cdots, G_k$, is a genome that minimizes the total distance function $\sum_{i=1}^{k} d(\ .\ , G_i)$. Often employed to reconstruct the gene-order evolutionary history in the small phylogeny problem, the medians are proved to be rather useful for inferring the ancestral genomes under certain conditions (see Jamshidpey and Sankoff (2013); Jamshidpey (2016); Jamshidpey and Sankoff (2017) for the mathematical study). The medians are also aimed to capture the genetic similarities shared among $k \geq 3$ genomes. However despite their importance, finding and constructing medians can be rather challenging (da Silva et al, 2024; Larlee et al, 2014). In fact the median problem is NP-hard for various genomic distances including the breakpoint distance (Bryant, 1998; Caprara, 2003; Tannier et al, 2009; Fertin et al, 2009). Moreover, multiple median genomes may exist for a given set of genomes, some of which may be distant from each other, and some may be far from the true ancestor, rendering them unsuitable for accurate inference. Therefore, studying the size and distribution of the set of medians of $k \geq 3$ random genomes is important for advancing our understanding of genomic evolution.

To be more precise, labeling genes by numbers, a linear unichromosomal genome of size $n$ can be represented by a permutation on $[n] := \{1, \cdots, n\}$. Letting $S_n$ denote the set of all permutations on $[n]$, or equivalently the set of all genomes with identical set of genes $1, \cdots, n$, it is of particular interest to find the probability $\mathbb{P}(A \subset \mathcal{M})$ for any set $A \subset S_n$, where $\mathcal{M}$ is the set of medians of some genomes (permutations) $\xi_1^{(n)}, \cdots, \xi_k^{(n)}$ chosen uniformly and independently at random from $S_n$. Solving this problem however is difficult and the probability depends on various parameters including the pairwise distances of genomes in $A$. In this paper, we address the problem for the simplest case of singletons $A = \{x\}$. The symmetry of the permutation group indicates that this probability is the same for all choices of $x$. However, it is not very simple to find the exact value of $\mathbb{P}(x \in \mathcal{M})$ which depends on $k$ and $n$. Computing this probability, one can easily derive $\mathbb{E}|\mathcal{M}|$, the expected number of breakpoint medians of random permutations $\xi_1, \cdots, \xi_k$.

To establish our results, we introduce the notion of the "*median inverse*" of $\pi \in S_n$ as the set of all $k$-subsets $\{y_1, \cdots, y_k\}$ of $S_n$ for which $\pi$ is a median. We define the "*approximate breakpoint medians*" of $k$ genomes to be a genome whose adjacencies, except a few of them, are selected from the union of adjacencies of the $k$ genomes. For a given genome $\pi$, we construct and count all $k$-subsets $\{y_1, \cdots, y_k\}$ of genomes in $S_n$ for which $\pi$ is an approximate median. This gives an asymptotic upper bound for the size of the median inverse set. For any given genome, this also determines the probability of being an approximate median of a set of random genomes which itself is an upper bound for the probability of being their exact median. Finally, we derive an explicit expression for the expected number of the approximate medians of a set of random genomes. For any given $k \in \mathbb{N}$ and $\pi \in S_n$, our approach provides an algorithm to count the number of all possible $k$-subsets $\{y_1, \cdots, y_k\} \subset S_n$ for which $\pi$ is a median.

The paper is organized as follows. Section 2 introduces some important concepts such as *median inverse* and *segment sets*. In Section 3, we provide a brief description of the main results. Section 4 discusses the main results on the number of approximate medians of a set of random genomes. We first recall some results about the behaviour of the median of $k$ genomes with (almost) maximum pairwise breakpoint distances. In particular, for any small $k \in \mathbb{N}$ and $\pi \in S_n$, we establish a theory for counting the number of $k$-subsets $\{y_1, \cdots, y_k\} \subset S_n$ for which $\mathcal{A}_\pi \subset \cup_{i=1}^k \mathcal{A}_{y_i}$. This motivates us to build a theory in the general case in Theorems 1-4. More precisely, for any set of permutations (genomes) $\{x_1, \cdots, x_k\} \subset S_n$ and any median of this set, namely $x$, Theorem 1 provides an upper bound for $|\mathcal{A}_x \setminus \cup_{i=1}^k \mathcal{A}_{x_i}|$, the total number of adjacencies of $x$ which are not adjacencies of $x_1, \cdots, x_k$. Using this, Theorem 2 gives the asymptotic behaviour of the corresponding upper bound for the total number of adjacencies of any median of random genomes $\xi_1, \cdots, \xi_k$ which are not taken from the adjacencies of $\xi_1, \cdots, \xi_k$. Finally, Theorems 3 and 4 provide an explicit expression of the probability for any given permutation to be an approximate median of $\xi_1, \cdots, \xi_k$. This in addition computes the expected number of approximate breakpoint medians of $\xi_1, \cdots, \xi_k$.

## 2 Definitions

We assume that there are no duplicated genes. This means linear unichromosomal genomes with $n$ genes or markers labelled by $1, \cdots, n$ are represented by permutations on $[n] := \{1, \cdots, n\}$. Let $S_n$ denote the group of all permutations on $[n]$, endowed with function composition as the multiplication operation, and denote by $id := id^{(n)}$ the identity permutation $1\ 2\ 3\ ...\ n$. For a permutation $\pi := \pi_1\ ...\ \pi_n$, any unordered pair $\{\pi_i, \pi_{i+1}\} = \{\pi_{i+1}, \pi_i\}$, for $i = 1, ..., n-1$, is called a (gene) adjacency of $\pi$. The set of all adjacencies of $\pi$ is denoted by $\mathcal{A}_\pi$. Also, $\mathcal{A}_{x_1,...,x_k}$ denotes the set of all common adjacencies of $x_1, ..., x_k \in S_n$. The *breakpoint distance* (bp distance) between $x, y \in S_n$ is defined by

$$d(x,y) = d^{(n)}(x,y) := n - 1 - |\mathcal{A}_{x,y}| = |\mathcal{A}_x \Delta \mathcal{A}_y|/2.$$

Note that the bp distance is a left-invariant pseudometric on $S_n$, that is, for any $x, y, z \in S_n$, $d(x,y) = d(zx, zy)$.

Define the total breakpoint distance of $x \in S_n$ to a set of permutations $A \subset S_n$ by $d_T(x, A) := \sum_{y \in A} d(x,y)$. A *median* of $A \subseteq S_n$ is a permutation in $S_n$ (not necessarily unique) whose total distance to $A$ takes the minimum value, i.e. it is a permutation $x \in S_n$ such that $d_T(x, A) = \min_{y \in S_n} d_T(y, A)$. Furthermore, the median value of $A$ is the minimum value of the total distance function to $A$. We denote by $\mathcal{M}_n(A)$ the set of all breakpoint medians of $A \subset S_n$. Note that that $\mathcal{M}_n(A)$ is always non-empty, but not necessarily a singleton.

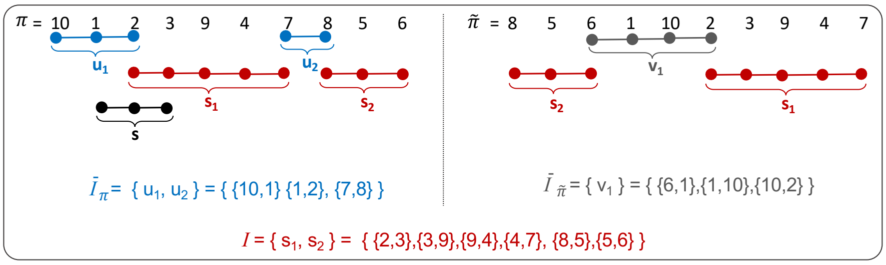
π = 10 1 2 3 9 4 7 8 5 6
u₁
u₂
s₁
s₂
s
Ī_π = { u₁, u₂ } = { {10,1} {1,2}, {7,8} }
π̃ = 8 5 6 1 10 2 3 9 4 7
v₁
s₂
s₁
Ī_π̃ = { v₁ } = { {6,1},{1,10},{10,2} }
I = { s₁, s₂ } = { {2,3},{3,9},{9,4},{4,7}, {8,5},{5,6} }

**Fig. 1** The segment set $I$ is contained in permutations $\pi$ and $\tilde{\pi}$. As $\pi \neq \tilde{\pi}$, the complement segment sets of $I$ with respect to $\pi$ and $\tilde{\pi}$ are different, i.e., $\bar{I}_\pi \neq \bar{I}_{\tilde{\pi}}$. The segments $s_1$ and $s_2$ are strongly disjoint, $s_1$ and $u_1$ are disjoint, and $s_1$ and $s$ intersect. The segments $s$ and $u_1$ are consistent and their union is $\{\{10,1\}\{1,2\}\{2,3\}\}$, while the segments $s$ and $v_1$ are inconsistent.

It is not hard to see that any median of $x, y \in S_n$ is a permutation $z \in S_n$ for which $d(x,y) = d(x,z) + d(z,y)$ and vice versa. Any such permutation $z \in S_n$ is called a geodesic point of $x, y$ (Jamshidpey et al, 2014). The set of all geodesic points of $x, y \in S_n$ is denoted by $\overline{[x,y]}$.

A segment of $S_n$ is a set of consecutive (gene) adjacencies of a permutation (genome) in $S_n$. More precisely, a segment of length $k \in [n-1]$ is a set of adjacencies $\{\{n_0, n_1\}, \{n_1, n_2\}, ..., \{n_{k-2}, n_{k-1}\}, \{n_{k-1}, n_k\}\}$, for $k+1$ distinct numbers (genes) $n_0, n_1, ..., n_k \in [n]$. In particular, the empty set $\emptyset$ is regarded as a segment of length 0. We denote by $|s|$ the length of a segment $s$. We say two segments $s$ and $s'$ are strongly disjoint if there is no common genes (numbers) shared between them. They are disjoint if they do not share any common adjacencies. Otherwise, we say they intersect each other.

A segment set $I$ of a permutation $\pi \in S_n$ is a subset of the gene adjacencies of $\pi$, i.e. $I \subset \mathcal{A}_\pi$. Alternatively, we say $\pi$ contains $I$. Note that a segment set can be contained in more than one permutation. Denote by $\mathcal{I}^{(n)}$ the set of all segment sets of $S_n$. By a segment of a segment set $I$ we mean a maximal segment contained in $I$. Also we denote by $\|I\| := k$, the number of segments of $I$. We emphasize that "$|\ .\ |$" is used for both cardinality of a set and absolute value of a real number. So for a segment set $I$ with segments $s_1, \cdots, s_{\|I\|}$, $|I| = \sum_{i=1}^{\|I\|} |s_i|$ counts the number of adjacencies of $I$. Two segment sets $I$ and $J$ (in particular two segments $s$ and $s'$, respectively) are said to be consistent, if their union is contained in $\mathcal{A}_\pi$, for some permutation $\pi$. When we speak of union of two or more segment sets (respectively, two or more segments) we always assume that they are pairwise consistent. We say segment sets $I_1, ..., I_k$ completes each other if there exists a permutation $\pi$ such that $\cup_{i=1}^k I_i = \mathcal{A}_\pi$. In particular, the complement of a segment set $I$ with respect to $\pi \in S_n$ is the unique segment set $\bar{I}_\pi := \mathcal{A}_\pi \setminus I$. An example is given in Figure 1.

## 3 Brief description of results

In this paper, we introduce the *"Median Inverse Problem"*, to study the expected number of medians for a set of random permutations. We find an upper bound for the probability that a permutation $\pi$ is a median of $k$ genomes $\xi_1, \cdots, \xi_k$, chosen independently and uniformly at random from $S_n$. Given a set $A \subset S_n$, we look for all $k$-tuples $(x_1, ..., x_k) \in S_n^k$ with $A \subset \mathcal{M}_n(x_1, \cdots, x_k)$, that is, the $k$-median inverse of $A$ is defined by

$$\mathcal{M}_{n,k}^{-1}(A) := \{(x_1, ..., x_k) \in S_n^k : \ A \subset \mathcal{M}_n(\{x_1, ..., x_k\})\}.$$

We find an upper bound for the cardinality of $\mathcal{M}_{n,k}^{-1}(A)$ when $A$ is a singleton. We approximate the asymptotic probability that $k$ independent random permutations in $S_n$ have a given median $\pi \in S_n$, and find an upper bound for that. In general, for $k$ given permutations $X = \{x_1, ..., x_k\}$, there may exist a median $\pi$ such that $\mathcal{A}_\pi \setminus \cup_{i=1}^k \mathcal{A}_{x_i} \neq \emptyset$. We find an upper bound, denoted by $\mathcal{O}_n(X)$, for $\max\{|\mathcal{A}_\pi \setminus \cup_{i=1}^k \mathcal{A}_{x_i}| : \ \pi \in \mathcal{M}_n(X)\}$, which is the maximum number of adjacencies a median $\pi$ may have, that do not belong to any permutation $x_1, \cdots, x_k$. For random permutations $\xi_1^{(n)}, \cdots, \xi_k^{(n)}$ chosen independently from $S_n$ and for any sequence of real numbers $(a_n)_{n\in\mathbb{N}}$ tending to $\infty$, we show that $|\mathcal{O}_n(\{\xi_1^{(n)}, ..., \xi_k^{(n)}\})|/a_n \to 0$, in probability, as $n$ goes to $\infty$. For $c \geq 0$, we define a $c$-approximate median of $X$ to be a permutation $\pi$ for which $|\mathcal{A}_\pi \setminus \bigcup_{x\in X} \mathcal{A}_x| \leq c$. Motivated by this, the $c$-approximate median set and $c$-approximate median inverse of $\pi \in S_n$ are defined as

$$\mathscr{L}_{n,c}(X) := \{\pi \in S_n : \ |\mathcal{A}_\pi \setminus \bigcup_{x\in X} \mathcal{A}_x| \leq c\}, \tag{1}$$

and

$$\mathscr{L}_{n,k,c}^{-1}(\pi) = \{(y_1, ..., y_k) \in S_n^k : |\mathcal{A}_\pi \setminus \bigcup_{i=1}^k \mathcal{A}_{y_i}| \leq c\}. \tag{2}$$

We show that with high probability, as $n \to \infty$, we have

$$\mathcal{M}_n(\xi_1^{(n)}, \cdots, \xi_k^{(n)}) \subset \mathscr{L}_{n,a_n}(\xi_1^{(n)}, \cdots, \xi_k^{(n)}),$$

implying $|\mathcal{M}_{n,k}^{-1}(\pi^{(n)})| \leq |\mathscr{L}_{n,k,a_n}^{-1}(\pi^{(n)})|$ for any arbitrary sequence of $(\pi^{(n)})_{n\in\mathbb{N}}$, $\pi^{(n)} \in S_n$. Finally, we construct the elements of $\mathscr{L}_{n,k,c}^{-1}(\pi^{(n)})$, and give an exact expression for its cardinality (see Theorems 3 and 4). Letting $\mathcal{O}_n^* = \mathcal{O}(\xi_1, \cdots, \xi_k)$, the inequality

$$\mathbb{E}|\mathcal{M}(\xi_1, \cdots, \xi_k)| \leq |\mathscr{L}_{n,k,\mathcal{O}_n^*}^{-1}(\pi^{(n)})|/(n!)^{k-1}$$

provides an upper bound for the expected number of medians of $\xi_1, \cdots, \xi_k$. Among their various advantages, the methods discussed in this paper provide an algorithmic perspective on both constructing and counting approximate medians of a set of random genomes.

# 4 Results

For a permutation $x \in S_n$, it is clear that $\mathcal{M}_{n,1}^{-1}(x) = \{x, y\}$, where $y \neq x$ is the unique permutation for which $d(x, y) = 0$. For instance, $\mathcal{M}_{n,1}^{-1}(id) = \{id, y\}$, where $y =$

$n\ n-1\ \cdots\ 2\ 1$. Also, $\mathcal{M}_{n,2}^{-1}(x) = \{(x_1, x_2) \in S_n^2 :\ x \in \overline{[x_1, x_2]}\}$. It is well known that the expected number of common adjacencies of two random permutations is $O(1)$; see Proposition 4. This means that, with high probability, the random permutations are at approximately maximum distance from each other. Hence, to study the median of $k$ random permutations, we need to review some results for the median of $k$ permutations with pairwise maximum distance $n-1$ from each other. In particular recall that a permutation $\pi$ is the median of $k$ permutations with pairwise maximum distance $n-1$ from each other, if and only if every adjacency of $\pi$ is an adjacency of at least one of those $k$ permutations. In other words we have

**Proposition 1** (Jamshidpey et al (2014)). *Let $X \subset S_n$ such that $d(x, y) = n-1$ for any $x, y \in X$. Then $\pi$ is a median of $X$ if and only if $\mathcal{A}_\pi \subset \bigcup_{x \in X} \mathcal{A}_x$.*

This is not true in general when the distances of input genomes are not the maximum value $n-1$. For example, if $n = 9$, $x = 2\ 7\ 5\ 6\ 8\ 3\ 9\ 4\ 1$ and $\pi = 6\ 8\ 9\ 3\ 4\ 1\ 2\ 7\ 5$, then every adjacency of $\pi$ is either an adjacency of $id = id^{(n)}$ or an adjacency of $x$, but $d^{(9)}(id, x) = 7 < d^{(9)}(id, \pi) + d^{(9)}(\pi, x) = 8$. In fact, this happens because all common adjacencies of $id$ and $x$ must be adjacencies of $\pi$ in order to have $\pi \in \overline{[id, x]}$ as stated in the next proposition.

**Proposition 2** (Jamshidpey et al (2014)). *Let $x, y \in S_n$. Then $z \in \overline{[x, y]}$ if and only if $\mathcal{A}_{x,y} \subset \mathcal{A}_z \subset \mathcal{A}_x \cup \mathcal{A}_y$.*

Recall from (1) and (2) that, for any $X \subset S_n$

$$\mathscr{L}_{n,0}(X) := \{\pi \in S_n : \mathcal{A}_\pi \subset \bigcup_{x \in X} \mathcal{A}_x\},$$

and

$$\mathscr{L}_{n,k,0}^{-1}(\pi) = \{(x_1, ..., x_k) \in S_n^k : \pi \in \mathscr{L}_{n,0}(\{x_1, ..., x_k\})\} = \{(x_1, ..., x_k) \in S_n^k : \mathcal{A}_\pi \subset \bigcup_{x \in X} \mathcal{A}_x\}. \quad (3)$$

In fact, Proposition 1 implies that for $X \subset S_n$ with $d(x, y) = n-1$, for $x, y \in X$, we have $\mathcal{M}_n(X) = \mathscr{L}_{n,0}(X)$. Letting

$$V_{n,k} = \{(x_1, ..., x_k) \in S_n^k :\ d(x_i, x_j) = n-1,\ \text{for}\ \ i \neq j\},$$

we get, for any $\pi \in S_n$, $\mathscr{L}_{n,k,0}^{-1}(\pi) \cap V_{n,k} = \mathcal{M}_{n,k}^{-1}(\pi) \cap V_{n,k}$. In other words, when we restrict ourselves to $V_{n,k}$, elements of $\mathcal{M}_{n,k}^{-1}(\pi)$ are identical to those of $\mathscr{L}_{n,k,0}^{-1}(\pi)$. We continue this paper with studying $|\mathscr{L}_{n,k,0}^{-1}(\pi)|$. To this end, we try to find an ordered $k$-tuple $(J_1, ..., J_k)$ where every $J_i$, for $i = 1, ..., k$, is a segment set of $S_n$ such that $\bigcup_{i=1}^k J_i = \mathcal{A}_\pi$, and then find all possible $k$-tuples $(x_1, ..., x_k) \in S_n^k$ such that, for $i = 1, ..., k$, permutation $x_i$ contains exactly segment set $J_i$ from $\pi$, not anything more. We will see that this gives us a way to count $|\mathscr{L}_{n,k,0}^{-1}(\pi)|$ exactly. More precisely, for a segment set $I$ in a permutation $\pi$, we define $\mathcal{H}_\pi^{(n)}(I)$ to be the set of all permutations having exactly segment set $I$ from $\pi$, that is $\mathcal{H}_\pi^{(n)}(I) = \{x \in S_n : \mathcal{A}_{\pi,x} = I\}$. Observe that, for permutations $x, y \in S_n$, $x \in \mathcal{H}_y^{(n)}(I)$ if and only if $y \in \mathcal{H}_x^{(n)}(I)$. Also, one can see that, for two non-identical segment sets of $\pi \in S_n$, namely $I \neq I'$, we have $\mathcal{H}_\pi^{(n)}(I) \cap \mathcal{H}_\pi^{(n)}(I') = \emptyset$, since if $x \in \mathcal{H}_\pi^{(n)}(I)$ and $y \in \mathcal{H}_\pi^{(n)}(I')$, then $\mathcal{A}_{\pi,x} = I \neq I' = \mathcal{A}_{\pi,y}$.

Let

$$\mathscr{P}_{k,0}^{(n)}(\pi) = \{(J_1, ..., J_k) \in (\mathcal{I}^{(n)})^k : \bigcup_{i=1}^{k} J_i = \mathcal{A}_\pi\}. \tag{4}$$

Note that the $J_i$ in the definition of $\mathscr{P}_{k,0}^{(n)}(\pi)$ may intersect each other. If $\mathcal{J} = (J_1, ..., J_k), \mathcal{J}' = (J_1', ..., J_k') \in \mathscr{P}_{k,0}^{(n)}(\pi)$, such that $\mathcal{J} \neq \mathcal{J}'$, then

$$(\mathcal{H}_\pi^{(n)}(J_1) \times ... \times \mathcal{H}_\pi^{(n)}(J_k)) \cap (\mathcal{H}_\pi^{(n)}(J_1') \times ... \times \mathcal{H}_\pi^{(n))}(J_k')) = \emptyset.$$

Now, if $(x_1, ..., x_k) \in \mathscr{L}_{n,k,0}^{-1}(\pi)$, then $\mathcal{A}_\pi \subset \cup_{i=1}^{k} \mathcal{A}_{x_i}$. Therefore, there exists $(J_1, ..., J_k) \in \mathscr{P}_{k,0}^{(n)}(\pi)$ such that, for any $i = 1, ..., k$, $\mathcal{A}_{\pi,x_i} = J_i$ implying that $(x_1, ..., x_k) \in \mathcal{H}_\pi^{(n)}(J_1) \times ... \times \mathcal{H}_\pi^{(n)}(J_k)$. On the other hand, if

$$(x_1, ..., x_k) \in \bigcup_{(\tilde{J}_1, ..., \tilde{J}_k) \in \mathscr{P}_{k,0}^{(n)}(\pi)} \mathcal{H}_\pi^{(n)}(\tilde{J}_1) \times ... \times \mathcal{H}_\pi^{(n)}(\tilde{J}_k),$$

then there exists $(J_1, ..., J_k) \in \mathscr{P}_{k,0}^{(n)}(\pi)$ such that $x_i \in \mathcal{H}_\pi^{(n)}(J_i)$, for $i = 1, ..., k$, which means by itself $\mathcal{A}_{\pi,x_i} = J_i$. Thus

$$\mathcal{A}_\pi = \bigcup_{i=1}^{k} J_i = \bigcup_{i=1}^{k} \mathcal{A}_{\pi,x_i} \subset \bigcup_{i=1}^{k} \mathcal{A}_{x_i}.$$

Hence, $(x_1, ..., x_k) \in \mathscr{L}_{n,k,0}^{-1}(\pi)$. We have proved the following proposition.

**Proposition 3.** *Let $n, k$ be natural numbers, and $\pi$ be a permutation in $S_n$. Then*

$$\mathscr{L}_{n,k,0}^{-1}(\pi) = \bigcup_{(\tilde{J}_1, ..., \tilde{J}_k) \in \mathscr{P}_{k,0}^{(n)}(\pi)} \mathcal{H}_\pi^{(n)}(\tilde{J}_1) \times ... \times \mathcal{H}_\pi^{(n)}(\tilde{J}_k),$$

*and*

$$|\mathscr{L}_{n,k,0}^{-1}(\pi)| = \sum_{(\tilde{J}_1, ..., \tilde{J}_k) \in \mathscr{P}_{k,0}^{(n)}(\pi)} \prod_{i=1}^{k} |\mathcal{H}_\pi^{(n)}(\tilde{J}_i)|.$$

So, knowing the number of elements of $\mathcal{H}_\pi^{(n)}(\tilde{J}_i)$ has an important role in counting the number of elements of $\mathscr{L}_{n,k,0}^{-1}(\pi)$.

What can we say when the permutations are chosen uniformly and independently at random from $S_n$? One can see that the situation in this case is similar to that in the case of permutations with pairwise maximum distance from each other. The following classic result can shed light on this.

**Proposition 4.** *Let $\xi = \xi^{(n)}$ be a permutation chosen uniformly at random from $S_n$, and let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers tending to $\infty$. Then*

*i)* $\mathbb{E}[d(id,\xi)] = n - 1 - \frac{2(n-1)}{n},\ n \in \mathbb{N}$.

*ii)* $var(d(id,\xi)) = (2 - \frac{2}{n})(-1 + \frac{2}{n}) + \frac{4(n-2)^2}{n(n-1)},\ n \in \mathbb{N}$.

*iii) For any* $\varepsilon > 0$, $\mathbb{P}(n - 1 - d(id,\xi) \geq \varepsilon a_n) \to 0,$ *as* $n \to \infty$.

*Proof.* For simplicity, we drop the superscript of $id^{(n)}$ in the following computations. For $i = 1, ..., n-1$, let $\chi_i = 1$ if the $i$-th adjacency of $\xi^{(n)}$, namely $\{\xi_i^{(n)}, \xi_{i+1}^{(n)}\}$ is an adjacency of $id$, and let $\chi_i = 0$, otherwise. For $i = 1, ..., n-1$, we have

$$\mathbb{E}[\chi_i] = \frac{2(n-1)}{n(n-1)} = \frac{2}{n}.$$

We can write

$$\mathbb{E}|\mathcal{A}_{id,\xi^{(n)}}| = \sum_{i=1}^{n-1} \mathbb{E}[\chi_i] = \frac{2(n-1)}{n}.$$

Also,

$$\begin{aligned} var(|\mathcal{A}_{id,\xi^{(n)}}|) &= \sum_{i=1}^{n-1} \mathbb{E}[\chi_i^2] + 2\sum_{i<j} \mathbb{E}[\chi_i\chi_j] - \mathbb{E}[|\mathcal{A}_{id,\xi^{(n)}}|]^2 \\ &= \mathbb{E}[|\mathcal{A}_{id,\xi^{(n)}}|](1 - \mathbb{E}[|\mathcal{A}_{id,\xi^{(n)}}|]) + 2\sum_{j-i>1} \mathbb{E}[\chi_i\chi_j] + 2\sum_{j-i=1} \mathbb{E}[\chi_i\chi_j]. \end{aligned}$$

But, for $j - i > 1$, we consider two cases. First, the $i$-th adjacency of $\xi^{(n)}$ can be one of the $n-3$ adjacencies of $id$, namely $\{u, u+1\}$ for $u = 2, ..., n-2$, each with two different directions, i.e. either $\xi_i^{(n)} = u$ and $\xi_{i+1}^{(n)} = u+1$, or $\xi_{i+1}^{(n)} = u$ and $\xi_i^{(n)} = u+1$. In this case there are $n-4$ adjacencies of $id$ (exclude $\{u, u+1\}$ and both of its neighbouring adjacencies) that $j$-th adjacency of $\xi_i^{(n)}$ can be identical to, each with two directions. The second case, is when the $i$-th adjacency of $\xi^{(n)}$ is either $\{1, 2\}$ or $\{n-1, n\}$, each with two possible directions, and in this case there are $n-3$ adjacencies of $id$ (exclude the one chosen for $i$-th adjacency of $\xi^{(n)}$ and its unique neighbouring adjacency in $id$), that can be picked for the $j$-th adjacency of $\xi^{(n)}$, again each with two directions. In summary, for $j - i > 1$,

$$\begin{aligned}\mathbb{E}[\chi_i\chi_j] &= \mathbb{P}(\chi_i = 1, \chi_j = 1)\\ &= \frac{(2(n-3)\times 2(n-4)) + ((2\times 2)\times(2(n-3))}{n(n-1)(n-2)(n-3)} = \frac{4}{n(n-1)}. \end{aligned} \tag{5}$$

But the number of ways one can choose $i, j$ such that $j - i > 1$ is $1 + ... + (n-3) = (n-2)(n-3)/2$. So

$$2\sum_{j-i>1}\mathbb{E}[\chi_i\chi_j] = \frac{4(n-2)(n-3)}{n(n-1)}.$$

Similarly, for $j - i = 1$, the $i$-th and $i+1$-st adjacencies of $\xi^{(n)}$ should be identical to two consecutive adjacencies of $id$. Considering the direction, this gives $2(n-2)$ possible ways in which $\chi_i = 1$ and $\chi_{i+1} = 1$. This implies that, for $j - i = 1$,

$$\mathbb{E}[\chi_i\chi_j] = \mathbb{P}(\chi_i = 1, \chi_j = 1) = \frac{2(n-2)}{n(n-1)(n-2)} = \frac{2}{n(n-1)}$$

As the number of ways one can choose $i, j$ such that $j - i = 1$ is $n - 2$, we can write

$$2\sum_{j-i=1}\mathbb{E}[\chi_i\chi_j] = \frac{4(n-2)}{n(n-1)}.$$

Putting all this together, we get

$$var(|\mathcal{A}_{id,\xi^{(n)}}|) = \frac{2(n-1)}{n}(1 - \frac{2(n-1)}{n}) + \frac{4(n-2)(n-3)}{n(n-1)} + \frac{4(n-2)}{n(n-1)},\ n \in \mathbb{N}$$

which concludes the second part of the proposition. Finally, letting $n \to \infty$, we have $\mathbb{E}[|\mathcal{A}_{id,\xi^{(n)}}|] \to 2$ and $var(|\mathcal{A}_{id,\xi^{(n)}}|) \to 2$. Therefore, for any arbitrary sequence $(a_n)_{n\in\mathbb{N}}$, satisfying the conditions mentioned in the statement of the proposition, Chebyshev's inequality implies convergence in probability of $|\mathcal{A}_{id,\xi^{(n)}}|/a_n$ to 0, as $n \to \infty$, and therefore, $(iii)$ is proved. □

As we mentioned before, when the pairwise distances of permutations in $X \subset S_n$ take the maximum value $n - 1$, a permutation $\pi$ is a median of $X$ if and only if each of its adjacencies is an adjacency of exactly one of the permutations in $X$. This is not true in general. In fact, for a general $X \subset S_n$, a median need not to take all of its adjacencies from $\cup_{x\in X}\mathcal{A}_x$. Can we find an upper bound for the number of adjacencies of any median of $X$ which are not from $\cup_{x\in X}\mathcal{A}_x$? In other words, can we find a good uniform upper bound for $|\mathcal{A}_\pi \setminus \cup_{x\in X}\mathcal{A}_x|$ , for any median $\pi$ of $X$? The next theorem answers this question. Before stating the theorem we introduce some notations as follows.

Denote by $\mathcal{P}(S)$ the set of all subsets of a set $S$. Let $X = \{x_1, ..., x_k\} \subset S_n$ and let $\mathcal{B}^X_X = \mathcal{B}^X_{x_1,...,x_k} := \mathcal{A}_{x_1,...,x_k}$. Then, for any $j = 1...k$, let

$$\mathcal{B}^X_{x_1,...,x_{j-1},x_{j+1},...x_k} := \mathcal{A}_{x_1,...,x_{j-1},x_{j+1},...x_k} \setminus \mathcal{B}^X_{x_1,...,x_k}$$

Continuing this, for any $U = \{x_{i_1}, ..., x_{i_r}\} \subset X$, we set

$$\mathcal{B}_U^X = \mathcal{B}_{x_{i_1},...,x_{i_r}}^X := \mathcal{A}_U \setminus \bigcup_{U \subsetneq V} \mathcal{B}_V^X.$$

Also, for a permutation $\pi$ and $r \leq k$, let $\bar{\varepsilon}_{i_1,...,i_r}^X(\pi) := |\mathcal{A}_\pi \cap \mathcal{B}_{x_{i_1},...x_{i_r}}^X|$. For $x \in S_n$ and a subset $X \subset S_n$, recall that the bp total distance of $x$ to $X$ is given by

$$d_T(x, X) = d_T^{(n)}(x, X) := \sum_{y \in X} d(x, y).$$

**Theorem 1.** *Let $X = \{x_1, ..., x_k\} \subset S_n$, and let $x \in \mathcal{M}_n(X)$. We assume the labels of elements of $X$ are such that $d_T(x_k, X) = \min_{i=1,...,k} d_T(x_i, X)$. Then*

$$\begin{aligned} &|\mathcal{A}_x \setminus \bigcup_{i=1}^k \mathcal{A}_{x_i}| \\ &\quad \leq \sum_{r=2}^k (r-1)\{\sum_{1 \leq i_1 < ... < i_r \leq k} \bar{\varepsilon}_{i_1,...,i_r}^X(x) - \sum_{1 \leq i_1 < ... < i_{r-1} < k} |\mathcal{B}_{x_{i_1},...,x_{i_{r-1}},k}^X|\} \\ &\qquad \leq \sum_{r=2}^{k-1} (r-1) \sum_{1 \leq i_1 < ... < i_r < k} |\mathcal{B}_{x_{i_1},...,x_{i_r}}^X|. \end{aligned} \tag{6}$$

*In particular, for $k = 3$, for any $x \in \mathcal{M}_n(X)$, $|\mathcal{A}_x \setminus \bigcup_{i=1}^3 \mathcal{A}_{x_i}| \leq |\mathcal{B}_{x_1,x_2}^X|$.*

*Proof.* To ease the notation, when there is no risk of ambiguity, we let $\mathcal{B}_{i_1,\cdots,i_\ell} = \mathcal{B}_{x_{i_1},...,x_{i_\ell}}$. Let $\eta = |\mathcal{A}_x \setminus \cup_{i=1}^k \mathcal{A}_{x_i}|$. Then

$$\eta + \sum_{r=1}^k \sum_{1 \leq i_1 < ... < i_r \leq k} \bar{\varepsilon}_{i_1,...,i_r}^X(x) = n - 1.$$

As $x$ is a median of $X$, we have

$$\begin{aligned} &d_T(x, X) = k(n-1) - \sum_{r=1}^k [r \sum_{1 \leq i_1 < ... < i_r \leq k} \bar{\varepsilon}_{i_1,...,i_r}^X(x)] \\ &= (k-1)(n-1) + \eta - \sum_{r=2}^k [(r-1) \sum_{1 \leq i_1 < ... < i_r \leq k} \bar{\varepsilon}_{i_1,...,i_r}^X(x)] \\ &\leq d_T(x_k, X) = (k-1)(n-1) - (\sum_{1 \leq i_1 < k} |\mathcal{B}_{i_1,k}^X| + 2 \sum_{1 \leq i_1 < i_2 < k} |\mathcal{B}_{i_1,i_2,k}^X| \\ &\qquad + \cdots + (k-2) \sum_{1 \leq i_1 < ... < i_{k-2} < k} |\mathcal{B}_{i_1,...,i_{k-2},k}^X| + (k-1)|\mathcal{B}_{1,...,k}^X|). \end{aligned}$$

Hence,

$$\eta \leq$$

$$(\sum_{r=2}^{k}(r-1)\sum_{1\leq i_1<...<i_r\leq k}\bar{\varepsilon}^X_{i_1,...,i_r}(x)) - (\sum_{r=2}^{k}(r-1)\sum_{1\leq i_1<...<i_{r-1}<k}|\mathcal{B}^X_{i_1,...,i_{r-1},k}|)$$

$$\leq \sum_{r=2}^{k-1}(r-1)\sum_{1\leq i_1<...<i_r<k}|\mathcal{B}^X_{i_1,...,i_r}|, \quad (7)$$

where the last inequality holds because $\bar{\varepsilon}^X_{i_1,...,i_r}(x) \leq |\mathcal{B}^X_{i_1,...,i_r}|$, for any $r \leq k$ and $1 \leq i_1 < ... < i_r \leq k$. □

For $X = \{x_1, ..., x_k\} \subset S_n$, let $\sigma$ be an arbitrary permutation on $\{1, .., k\}$ such that

$$d_T(x_{\sigma(k)}, X) = \min_{i=1...k} d_T(x_i, X),$$

Consider the relabelling $x^\sigma_i := x_{\sigma(i)}$, for $i = 1, ..., k$, for elements of $X$. So for $k \geq 3$, we can denote $\mathcal{O}_n = \mathcal{O}_{n,k} : (S_n)^k \to \mathbb{R}_+$,

$$\mathcal{O}_n(X) := \sum_{r=2}^{k-1}(r-1)\sum_{1\leq i_1<...<i_r<k}|\mathcal{B}^X_{x^\sigma_{i_1},...,x^\sigma_{i_r}}|.$$

**Remark 1.** *Of course for $X = \{x_1, ..., x_k\} \subset S_n$ with maximum distance $d(x_i, x_j)$ $= n - 1$, for $i \neq j$, $\mathcal{O}_n(X) = 0$ and therefore, as we have seen in Proposition 1, every median picks its adjacencies from the union of adjacencies of $k$ permutations. But the converse is not true, namely, there exist sets of permutations $X$ such that $\mathcal{O}_n(X) = 0$, but adjacency sets of permutations have intersections with each other. So Theorem 1 gives a stronger statement when permutations of $X$ are not located at the maximum distance of each other but $\mathcal{O}_n(X) = 0$ and in this case we still have the same property. For example, consider three permutations $id = id^{(6)} = 1\ 2\ 3\ 4\ 5\ 6$ , $x = 4\ 6\ 5\ 1\ 3\ 2$, and $y = 4\ 2\ 6\ 5\ 1\ 3$, and let $X = \{id, x, y\}$. We have $\mathcal{A}_{id,x} = \{\{2, 3\}, \{5, 6\}\}$, $\mathcal{A}_{id,y} =$ $\{\{5, 6\}\}$, $\mathcal{A}_{x,y} = \{\{5, 6\}, \{1, 5\}, \{1, 3\}\}$, and $\mathcal{A}_{id,x,y} = \{\{5, 6\}\}$. Then $d_T(id, X) = 7$, $d_T(x, X) = 5$, and $d_T(y, X) = 6$, and thus $\mathcal{O}_n(X) = |\mathcal{B}_{id,y}| = |\mathcal{A}_{id,y} \setminus \mathcal{A}_{id,x,y}| = 0$.*

Motivated by Theorem 1 and rewriting

$$\mathscr{L}_{n,0}(X) := \{\pi \in S_n : |\mathcal{A}_\pi \setminus \bigcup_{x\in X}\mathcal{A}_x| \leq 0\},$$

for $c \geq 0$, the set of $c$-approximate median set and $c$-approximate median inverse for $X$ is defined in (1) and (2)

$$\mathscr{L}_{n,c}(X) := \{\pi \in S_n : |\mathcal{A}_\pi \setminus \bigcup_{x\in X}\mathcal{A}_x| \leq c\},$$

and

$$\begin{aligned}\mathscr{L}^{-1}_{n,k,c}(\pi) &:= \{(x_1,...,x_k) \in S_n^k : \pi \in \mathscr{L}_{n,c}(\{x_1,...,x_k\})\} \\ &= \{(x_1,...,x_k) \in S_n^k : |\mathcal{A}_\pi \setminus \bigcup_{x\in X} \mathcal{A}_x| \leq c\}.\end{aligned}$$

Note that the left-invariance property of the breakpoint distance implies $\mathcal{A}_{x,y} = \mathcal{A}_{\pi x,\pi y}$, for $\pi, x, y \in S_n$. Therefore, for any $\pi, x \in S_n$ and $X = \{x_1,...,x_k\} \subset S_n$, $d_T(x, X) = d_T(\pi x, \pi X)$, where $\pi X = \{\pi x_1,...,\pi x_k\}$. This yields $\pi\mathcal{M}_n(X) = \mathcal{M}_n(\pi X)$, and therefore, for any $x, y \in S_n$, $|\mathcal{M}^{-1}_{n,k}(x)| = |\mathcal{M}^{-1}_{n,k}(y)|$. Also, denoting the bp median value of $X$ by $\mu_n(X)$, we can write $\mu_n(X) = \mu_n(\pi X)$. On the other hand, for $\pi \in S_n$ and $k$-tuple $(x_1,...,x_k) \in S_n^k$, denote $\pi(x_1,...,x_k) = (\pi x_1,...,\pi x_k)$. Similarly to the median inverse case, write

$$\begin{aligned}\mathscr{L}^{-1}_{n,k,c}(\pi x) &= \{(\pi x_1,...,\pi x_k) : |\mathcal{A}_{\pi x} \setminus \cup_{i=1}^k \mathcal{A}_{\pi x_i}| \leq c\} \\ &= \{\pi(x_1,...,x_k) : |\mathcal{A}_x \setminus \cup_{i=1}^k \mathcal{A}_{x_i}| \leq c\} = \pi\mathscr{L}^{-1}_{n,k,c}(x),\end{aligned}$$

and thus, for any $x, y \in S_n$, $|\mathscr{L}^{-1}_{n,k,c}(x)| = |\mathscr{L}^{-1}_{n,k,c}(y)|$, since $(x_1,...,x_k) \mapsto (\pi x_1,...,\pi x_k)$ is a bijection for any given $\pi \in S_n$.

Let $\xi_1^{(n)},...,\xi_k^{(n)}$ be $k$ permutations chosen uniformly and independently at random from $S_n$. Roughly speaking, Proposition 1 implies that for any sequence $(a_n)_{n\in\mathbb{N}}$, for which $a_n \to \infty$ and $a_n/n \to 0$, as $n \to \infty$, and any sequence of permutations $(\pi_n)_{n\in\mathbb{N}}$, with $\pi_n \in S_n$, $|\mathscr{L}^{-1}_{n,k,a_n}(\pi_n)|$ somehow gives an upper bound for $|\mathcal{M}^{-1}_{n,k}(\pi_n)|$. This is formalized in the next theorem.

**Theorem 2.** *Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of real numbers diverging to $\infty$, as $n \to \infty$. Let $\xi_1^{(n)},...,\xi_k^{(n)}$ be $k$ permutations chosen uniformly and independently at random from $S_n$. Then, as $n \to \infty$,*

$$\frac{\mathscr{O}_n(\xi_1^{(n)},...,\xi_k^{(n)})}{a_n} \to 0, \text{ in probability,}$$

*and*

$$\mathbb{P}(\mathcal{M}_n(\xi_1^{(n)},...,\xi_k^{(n)}) \subseteq \mathscr{L}_{n,a_n}(\xi_1^{(n)},...,\xi_k^{(n)})) \to 1. \tag{8}$$

*Proof.* Let $X = \{x_1,...,x_k\} \subset S_n$, and consider sets $U_2,...,U_k \subset X$, such that for a fixed pair $i_1 \neq i_2$ in $\{1,...,k\}$,

$$U_2 = \{x_{i_1}, x_{i_2}\} \subsetneqq U_3 \subsetneqq ... \subsetneqq U_k = X,$$

that is for any $l = 2,...,k-1$, $|U_{l+1} \setminus U_l| = 1$. Then by definition

$$\bigcup_{i=2}^k \mathcal{B}^X_{U_i} = \mathcal{A}_{x_{i_1},x_{i_2}}.$$

This yields

$$|\mathcal{O}_n(X)| \le \sum_{i<j<k} |\mathcal{A}_{x_i,x_j}|.$$

So if $|\mathcal{O}_n(X)| \ge c$, for $c \ge 0$, then for at least one pair of points in $X$, namely $x, y$,

$$|\mathcal{A}_{x,y}| \ge \frac{c}{\binom{k-1}{2}}.$$

Letting $\tau$ be the minimum index $i \in \{1, ..., k\}$ such that

$$d_T^{(n)}(\xi_i^{(n)}, \{\xi_1^{(n)}, ..., \xi_k^{(n)}\}) \le d_T^{(n)}(\xi_j^{(n)}, \{\xi_1^{(n)}, ..., \xi_k^{(n)}\}),$$

for $j = 1, ..., k$, the last inequality implies that, for any $\varepsilon > 0$

$$\begin{aligned}\mathbb{P}(\mathcal{O}_n(\xi_1^{(n)}, ..., \xi_k^{(n)}) \ge \varepsilon a_n) &\le \mathbb{P}\Big( \bigcup_{\substack{i<j \\ i,j \ne \tau}} \Big\{ |\mathcal{A}_{\xi_i^{(n)},\xi_j^{(n)}}| \ge \frac{\varepsilon a_n}{\binom{k-1}{2}} \Big\} \Big) \\ &\le \sum_{\substack{i<j \\ i,j \ne \tau}} \mathbb{P}\Big( |\mathcal{A}_{\xi_i^{(n)},\xi_j^{(n)}}| \ge \frac{\varepsilon a_n}{\binom{k-1}{2}} \Big) \to 0,\end{aligned}$$

as $n \to \infty$, by Proposition 1, hence the first part is proved. To prove (8), note that $\mathcal{M}_n(\xi_1, ..., \xi_k) \nsubseteq \mathscr{L}_{n,a_n}(\xi_1, ..., \xi_k)$ implies that

$$a_n < |\mathcal{A}_m \setminus \cup_{i=1}^k \mathcal{A}_{\xi_i}| \le |\mathcal{O}_n(\xi_1, ..., \xi_k)|,$$

for some median $m \in \mathcal{M}_n(\xi_1, ..., \xi_k)$. Therefore as $n \to \infty$,

$$\mathbb{P}(\mathcal{M}_n(\xi_1, ..., \xi_k) \nsubseteq \mathscr{L}_{n,a_n}(\xi_1, ..., \xi_k)) \le \mathbb{P}\left( \frac{|\mathcal{O}_n(\xi_1, \cdots, \xi_k)|}{a_n} > 1 \right) \to 0,$$

which completes the proof. □

**Remark 2.** *Although the theorem is true for any arbitrary diverging sequence* $(a_n)_{n\in\mathbb{N}}$*, it is useful to pick a small order sequence such that* $a_n/n \to 0$*, as* $n \to 0$*. In fact, the smaller the order of* $a_n$*, the better the upper bound.*

As discussed before, the bp distance is left invariant, and hence

$$\mathbb{P}(x \in \mathcal{M}_n(\xi_1^{(n)}, \cdots, \xi_k^{(n)})) = \mathbb{P}(y \in \mathcal{M}_n(\xi_1^{(n)}, \cdots, \xi_k^{(n)}))$$

for any $x \ne y \in S_n$. It is not hard to see that, for any permutation, the probability to be a median of $\xi_1, \cdots, \xi_k$ is too small, and this probability indeed converges to 0, as $n$ tends to $\infty$.

**Proposition 5.** *Let $\xi_1^{(n)}, \cdots, \xi_k^{(n)}$ be $k$ permutations chosen independently at random from $S_n$. For any $\pi_n \in S_n$, as $n \to \infty$,*

$$\mathbb{P}(\pi_n \in \mathcal{M}_n(\xi_1^{(n)}, \cdots, \xi_k^{(n)})) \to 0.$$

*Proof.* We denote by $\tilde{\mu}_{n,k}$ the median value of $\xi_1^{(n)}, \cdots, \xi_k^{(n)}$. From Jamshidpey et al (2014), for any arbitrary sequence $(a_n)_{n=1}^{\infty}$ with $a_n \to \infty$ we have

$$\frac{\tilde{\mu}_{n,k} - (k-1)n}{a_n} \to 0$$

in probability. This also follows from Theorem 2. On the other hand, for any $\pi_n \in S_n$, Proposition 4 implies that

$$\sum_{i=1}^{k} \frac{d(\pi_n, \xi_i^{(n)}) - n}{a_n} = \frac{d_T(\pi_n, \{\xi_1^{(n)}, \cdots, \xi_k^{(n)}\}) - kn}{a_n} \to 0$$

in probability. Thus $\pi_n$ cannot be a median for $\{\xi_1^{(n)}, \cdots, \xi_k^{(n)}\}$, with high probability, and the proposition follows. □

Although, the probability to be a median is too small, to study the expected number of medians of $\xi_1, \cdots, \xi_k$, it is important to find its exact value. It is clear that $\mathbb{P}(x \in \mathcal{M}(\xi_1)) = 1/n!$ for any $x \in S_n$. However as $\mathcal{M}_n(\xi_1, \xi_2) = \overline{[\xi_1, \xi_2]}$, we obtain

$$\mathbb{P}(x \in \mathcal{M}_n(\xi_1^{(n)}, \xi_2^{(n)})) = \frac{|\{(y_1, y_2) \in S_n^2 : x \in \overline{[y_1, y_2]}\}|}{(n!)^2} \gg \frac{2}{n!} + \left(\frac{1}{n!}\right)^2.$$

For $k$ random permutations one can use the notion of accessible points introduced in (Jamshidpey, 2016, Chapter 4) and (Jamshidpey et al, 2014, Theorem 2) to get a similar crude lower bound

$$\binom{k}{2} \frac{|\{(y_1, y_2) \in S_n^2 : x \in \overline{[y_1, y_2]}\}|}{(n!)^k}$$

for the probability to be an $a_n$-approximate median for any permutation $x$, $a_n \to \infty$ as $n \to \infty$.

So far, we have seen that

$$\mathbb{P}(|\mathcal{M}_n(\xi_1, \cdots, \xi_k)| \leq |\mathscr{L}_{n,a_n}(\xi_1, \cdots, \xi_k)|) \to 1,$$

and

$$\mathbb{E}|\mathcal{M}_n(\xi_1, \cdots, \xi_k)| \leq \mathbb{E}|\mathscr{L}_{n,\mathcal{O}_n^*}(\xi_1, \cdots, \xi_k)|,$$

where $\mathcal{O}_n^* = \mathcal{O}_n(\xi_1, \cdots, \xi_k)$. The next theorem indicates how the approximate median inverse set helps to find the expected number of $c$-approximate medians for a set of random permutations, for any given $c \geq 0$. To state the result, for $c > 0$, let

$$\mathscr{P}_{k,c}^{(n)}(\pi) = \{(J_1, ..., J_k) \in (\mathcal{I}^{(n)})^k : \ |\mathcal{A}_\pi \setminus \bigcup_{i=1}^{k} J_i| \leq c\},$$

and note that for $c = 0$, this definition is identical to the one in (4).

**Theorem 3.** *Let $n, k$ be natural numbers, and let $c \geq 0$ be a real number. Also let $\pi$ be a permutation in $S_n$. Then*

$$\mathbb{E}|\mathscr{L}_{n,c}(\xi_1^{(n)}, \cdots, \xi_k^{(n)})| = \frac{|\mathscr{L}_{n,k,c}^{-1}(\pi)|}{(n!)^{k-1}},$$

*where*

$$\mathscr{L}_{n,k,c}^{-1}(\pi) = \bigcup_{(\tilde{J}_1,...,\tilde{J}_k)\in \mathscr{P}_{k,c}^{(n)}(\pi)} \mathcal{H}_\pi^{(n)}(\tilde{J}_1) \times ... \times \mathcal{H}_\pi^{(n)}(\tilde{J}_k),$$

*and*

$$|\mathscr{L}_{n,k,c}^{-1}(\pi)| = \sum_{(\tilde{J}_1,...,\tilde{J}_k)\in \mathscr{P}_{k,c}^{(n)}(\pi)} \prod_{i=1}^{k} |\mathcal{H}_\pi^{(n)}(\tilde{J}_i)|.$$

*Proof.* From the definition

$$\mathbb{P}(\pi \in \mathscr{L}_{n,c}(\xi_1^{(n)}, \cdots, \xi_k^{(n)})) = \frac{|\mathscr{L}_{n,k,c}^{-1}(\pi)|}{(n!)^k}.$$

For $x \in S_n$, let the indicator random variable $\delta_x^c = 1$ if $x \in \mathscr{L}_{n,c}(\xi_1^{(n)}, \cdots, \xi_k^{(n)})$ and let $\delta_x^c = 0$ otherwise. It is then clear that

$$\mathbb{E}|\mathscr{L}_{n,c}(\xi_1^{(n)}, \cdots, \xi_k^{(n)})| = \sum_{x\in S_n} \mathbb{E}\delta_x^c = \frac{n!|\mathscr{L}_{n,k,c}^{-1}(\pi)|}{(n!)^k}.$$

We must now count the number of elements in $\mathscr{L}_{n,k,c}^{-1}(\pi)$ in terms of $\mathcal{H}_\pi^{(n)}(J_i)$, for $c > 0$. As in the case of $c = 0$, let $\mathcal{J} = (J_1, ..., J_k), \mathcal{J}' = (J_1', ..., J_k') \in \mathscr{P}_{k,c}^{(n)}(\pi)$, such that $\mathcal{J} \neq \mathcal{J}'$, then

$$(\mathcal{H}_\pi^{(n)}(J_1) \times ... \times \mathcal{H}_\pi^{(n)}(J_k)) \cap (\mathcal{H}_\pi^{(n)}(J_1') \times ... \times \mathcal{H}_\pi^{(n)}(J_k')) = \emptyset.$$

Now, if $(x_1, ..., x_k) \in \mathscr{L}_{n,k,c}^{-1}(\pi)$, then there exist at most $c$ adjacencies of $\pi$ that are not in $\cup_{i=1}^{k} \mathcal{A}_{x_i}$. Therefore, there exists $(J_1, ..., J_k) \in \mathscr{P}_{k,c}^{(n)}(\pi)$ such that, for any $i = 1, ..., k$, $\mathcal{A}_{\pi,x_i} = J_i$ implying that $(x_1, ..., x_k) \in \mathcal{H}_\pi^{(n)}(J_1) \times ... \times \mathcal{H}_\pi^{(n)}(J_k)$. On the other hand, if

$$(x_1, ..., x_k) \in \bigcup_{(\tilde{J}_1,...,\tilde{J}_k)\in \mathscr{P}_{k,c}^{(n)}(\pi)} \mathcal{H}_\pi^{(n)}(\tilde{J}_1) \times ... \times \mathcal{H}_\pi^{(n)}(\tilde{J}_k),$$

then there exists $(J_1, ..., J_k) \in \mathscr{P}_{k,c}^{(n)}(\pi)$ such that $x_i \in \mathcal{H}_\pi^{(n)}(J_i)$, for $i = 1, ..., k$, and so $\mathcal{A}_{\pi,x_i} = J_i$. Thus

$$\mathcal{A}_\pi \setminus \bigcup_{i=1}^{k} J_i = \mathcal{A}_\pi \setminus \bigcup_{i=1}^{k} \mathcal{A}_{\pi,x_i} = \mathcal{A}_\pi \setminus \bigcup_{i=1}^{k} \mathcal{A}_{x_i}.$$

Therefore, $|\mathcal{A}_\pi \setminus \bigcup_{i=1}^{k} \mathcal{A}_{x_i}| \le c$, and thus, $(x_1, ..., x_k) \in \mathscr{L}_{n,k,c}^{-1}(\pi)$. □

**Remark 3.** *We have*

$$|\mathscr{P}_{k,c}^{(n)}(\pi)| = \sum_{i=0}^{c} \binom{n-1}{i} (2^k - 1)^{n-1-i}.$$

*To see this, for any segment set $\bar{J}$ of $\pi$ with $i$ adjacencies, for $0 \le i \le c$, we partition $\mathcal{A}_\pi \setminus \bar{J}$ into $2^k - 1$ segment sets $\tilde{J}_A$, $\emptyset \ne A \subset [k]$, and let $J_i = \cup_{A:i\in A} \tilde{J}_A$. Each $(J_1, \cdots, J_k) \in \mathscr{P}_{k,c}^{(n)}(\pi)$ with $\mathcal{A}_\pi \setminus (\cup_{i=1}^{k} J_i) = \bar{J}$ is determined by one and only one such partition of $\mathcal{A}_\pi \setminus \bar{J}$. Having $(2^k - 1)^{n-1-i}$ such partitions, the claim is proved.*

We have so far seen that to estimate $|\mathscr{L}_{n,k,c}^{-1}(\pi)|$, for $c \ge 0$, we need to count the number of elements of $\mathcal{H}_\pi^{(n)}(J_i)$ for $(J_1, ..., J_k) \in \mathscr{P}_{k,c}^{(n)}(\pi)$. In addition to its use for counting the approximate medians as indicated in Theorem 3, for given $\pi$ and $I$, $\mathcal{H}_\pi(I)$ can help to understand the mechanism under which a permutation $\pi$ becomes a median of $k$ other permutations. Roughly speaking, for any $\pi \in S_n$, we first partition $\mathcal{A}_\pi$ into $2^k$ disjoint segment set $\tilde{J}_A$, $A \subset [k]$ with $|\tilde{J}_\emptyset| \le c$. Let $J_i = \cup_{A:i\in A} \tilde{J}_A$ and $\bar{J} = \tilde{J}_\emptyset$. In fact $\tilde{J}_A$ is the set of all adjacencies of $\pi$ included only in $J_i$, $i \in A$ and not included in $J_j$, $j \notin A$. For each $i = 1, \cdots, k$, we then construct all permutations $x \in \mathcal{H}_\pi(J_i)$ that do not have any adjacency in common with $\pi$ except those in $J_i$. Then for any choice of $x_1, \cdots, x_k$ with $x_i \in \mathcal{H}_\pi(J_i)$, we have $\pi \in \mathscr{L}_{n,|\bar{J}|}(x_1, \cdots, x_k)$ and $d(x_i, \pi) = n - 1 - |J_i|$. This, in addition, will provide an efficient way to generate a $c$-approximate median $\pi$ of $k$ random permutations $\xi_1, \cdots, \xi_k$ with breakpoint distances $d(\xi_i, \pi) = d_i$, for any $d_1, \cdots, d_k \ge 0$ with $d_1 + \cdots d_k - (k-1)(n-1) \le c$. More precisely, sample a random permutation $\xi$ and sample $k$ independent random segment sets $I_1, \cdots, I_k$ from $\xi$; $I_1, \cdots, I_k$ may intersect. We construct random permutations $\xi_i$ from $I_i$, $i = 1, \cdots, k$ by, first, assigning a random direction to the segments of $I_i$, in one of the $2^{\|I_i\|}$ different ways, and then, rearranging the segment sets and points of $[n]$ which are not present in $I_i$, such that $\mathcal{A}_{\pi,\xi_i} = I_i$. As a result we have random permutations $\xi_1 \cdots, \xi_k$ with a $c$-approximate median $\xi$ such that $d(\xi, \xi_i) = d_i$.

The rest of the paper is devoted to a more detailed analysis of $\mathcal{H}_\pi(I)$. We give an explicit representation for the number of elements in $\mathcal{H}_\pi^{(n)}(J_i)$; see Theorem 4. To establish this, for $J \in \mathcal{I}^{(n)}$, let

$$\mathcal{R}_n(J) = \{x \in S_n : \; J \subset \mathcal{A}_x\}.$$

Then, by inclusion-exclusion principle we have

$$|\mathcal{H}_\pi^{(n)}(I)| = \sum_{I\subset J\subset \mathcal{A}_\pi} (-1)^{|J\setminus I|}|\mathcal{R}_n(J)|. \tag{9}$$

To simplify this further, we introduce the type of a segment set $J$, $I \subset J \subset \mathcal{A}_\pi$, and we will see that the value of $|\mathcal{R}_n(\cdot)|$ is identical for two segment sets of the same type. Formally, let $\bar{I}_\pi := \mathcal{A}_\pi \setminus I$, and denote by $\bar{I}_\pi^{(i)}$, for $i = 1, ..., \|I\|+1$, the $i$-th segment of $\bar{I}_\pi$ ($i$-th from the left when considered as a segment of $\pi$). Note that $\bar{I}_\pi^{(1)}$and $\bar{I}_\pi^{(\|I\|+1)}$ may be empty segments. The type of a segment set $J \in \mathcal{I}^{(n)}$, where $I \subset J \subset \mathcal{A}_\pi$, with respect to $\pi$ and $I$, is identified by $\lambda := (\lambda_1, ..., \lambda_{\|I\|+1})$, where, for $i = 1, ..., \|I\|+1$, $\lambda_i$ is identified by the quadruple $\lambda_i := (\lambda_i^{(1)}, \lambda_i^{(2)}, \lambda_i^{(3)}, \lambda_i^{(4)}) \in \mathbb{N}\times\mathbb{N}\times\{0,1\}\times\{0,1\}$, where $\lambda_i^{(1)} := |J\cap\bar{I}_\pi^{(i)}|$ is the number of common adjacencies of $J$ and $\bar{I}_\pi^{(i)}$; $\lambda_i^{(2)} := \|J\cap\bar{I}_\pi^{(i)}\|$ is the number of segments of intersection of $J$ and $\bar{I}_\pi^{(i)}$; $\lambda_1^{(3)} = 0$ and, for $i = 2, ..., \|I\|+1$, $\lambda_i^{(3)} = 1$ if the leftmost adjacency of $\bar{I}_\pi^{(i)}$ is also in $J$ and otherwise $\lambda_i^{(3)} = 0$; and finally $\lambda_{\|I\|+1}^{(4)} = 0$, and for $i = 1, ..., \|I\|$, $\lambda_i^{(4)} = 1$ if the rightmost adjacency of $\bar{I}_\pi^{(i)}$ is also in $J$ and otherwise $\lambda_i^{(4)} = 0$. The next theorem counts the elements of $\mathcal{H}_\pi^{(n)}(I)$.

**Theorem 4.** *Let $\pi$ be a permutation in $S_n$, and $I \in \mathcal{I}^{(n)}$ be a segment set contained in $\pi$. Then*

$$|\mathcal{H}_\pi^{(n)}(I)| = \sum_\lambda \left\{ (-1)^{\sum_{i=1}^{\|I\|+1}\lambda_i^{(1)}} \prod_{i=1}^{\|I\|+1} \left[ \binom{\lambda_i^{(1)}-1}{\lambda_i^{(2)}-1} \binom{|\bar{I}_\pi^{(i)}|-\lambda_i^{(1)}-1}{\lambda_i^{(2)}-\lambda_i^{(3)}-\lambda_i^{(4)}} \right] \times \right. \\ \left. 2^{\{|I|+\sum_{i=1}^{\|I\|+1}\lambda_i^{(2)}-\sum_{i=1}^{\|I\|+1}(\lambda_i^{(3)}+\lambda_i^{(4)})\}} (n-|I|-\sum_{i=1}^{\|I\|+1}\lambda_i^{(1)})! \right\}, \tag{10}$$

*where the summation is over all $\lambda = (\lambda_i)_{i=1}^{\|I\|+1}$ with $\lambda_i = (\lambda_i^{(1)}, \lambda_i^{(2)}, \lambda_i^{(3)}, \lambda_i^{(4)})$ belongs to $\{1, ..., |\bar{I}_\pi^{(i)}|\} \times \{1, ..., \min\{\lambda_i^{(1)}, |\bar{I}_\pi^{(i)}| + 1 - \lambda_i^{(1)}\}\} \times \{0,1\} \times \{0,1\}$.*

To prove Theorem 4, we need the following lemmas.

**Lemma 1.** *Let $I$ be a segment set of $S_n$ with $m$ adjacencies and $k$ segments. Then the number of permutations in $S_n$ containing $I$ is equal to $2^k(n-m)!$.*

*Proof.* As the segment set $I$ has $m$ adjacencies and $k$ segments, each permutation containing $I$ has $n-m-k$ points (genes) which are not used in $I$. Noting that segments have two directions, we then have $2^k(k+(n-m-k))!$ permutations containing $I$. □

**Lemma 2.** *Let $I, J \in \mathcal{I}^{(n)}$ and $\pi \in S_n$, such that $I \subset J \subset \mathcal{A}_\pi$. Let $\lambda = (\lambda_i)_{1\leq i\leq \|I\|+1}$, with $\lambda_i = (\lambda_i^{(1)}, \lambda_i^{(2)}, \lambda_i^{(3)}, \lambda_i^{(4)})$, for $i = 1, ..., \|I\|+1$ be the type of $J$ with respect to $\pi$*

*and $I$. Then*

$$|\mathcal{R}_n(J)| = 2^{\{|I|+\sum\limits_{i=1}^{\|I\|+1}\lambda_i^{(2)}-\sum\limits_{i=1}^{\|I\|+1}(\lambda_i^{(3)}+\lambda_i^{(4)})\}}(n-|I|-\sum_{i=1}^{\|I\|+1}\lambda_i^{(1)})!.$$

*Proof.* We have

$$\|J\| = \|I\| + \sum_{i=1}^{\|I\|+1}\lambda_i^{(2)} - \sum_{i=1}^{\|I\|+1}(\lambda_i^{(3)}+\lambda_i^{(4)}),$$

and also the number of adjacencies of $J$ is equal to

$$|J| = |I| + \sum_{i=1}^{\|I\|+1}\lambda_i^{(1)}.$$

Therefore, Lemma 1 finishes the proof. □

**Lemma 3.** *Let $\pi \in S_n$ and $I \in \mathcal{I}^{(n)}$. The number of segment sets $J$, with $I \subset J \subset \mathcal{A}_\pi$ and with type $\lambda = (\lambda_i)_{1\le i\le \|I\|+1}$ with respect to $\pi$ and $I$, where $\lambda_i = (\lambda_i^{(1)}, \lambda_i^{(2)}, \lambda_i^{(3)}, \lambda_i^{(4)})$ for $i = 1, ..., \|I\|+1$, is*

$$\prod_{i=1}^{\|I\|+1}\binom{\lambda_i^{(1)}-1}{\lambda_i^{(2)}-1}\binom{|\bar{I}_\pi^{(i)}|-\lambda_i^{(1)}-1}{\lambda_i^{(2)}-\lambda_i^{(3)}-\lambda_i^{(4)}}.$$

*Proof.* The idea is to consider segment $\bar{I}_\pi^{(i)}$ as a permutation and count the number of possible ways one can choose a segment set $\tilde{J}_i$ from it with $\lambda_i^{(1)}$ number of adjacencies and $\lambda_i^{(2)}$ number of segments. More explicitly, for $i = 1, ..., \|I\|+1$, if $(\lambda_i^{(3)}, \lambda_i^{(4)}) = (1,1)$, then the number of ways we can do this is equal to the number of solutions of two independent equations

$$\mathbb{X}_1 + ... + \mathbb{X}_{\lambda_i^{(2)}} = \lambda_i^{(1)},$$

with $\mathbb{X}_i \ge 1$, for $i = 1, ..., \lambda_i^{(2)}$, and $\mathbb{Y}_2 + ... + \mathbb{Y}_{\lambda_i^{(2)}} = |\bar{I}_\pi^{(i)}| - \lambda_i^{(1)}$, with $\mathbb{Y}_i \ge 1$, for $i = 2, ..., \lambda_i^{(2)} - 1$, which is equal to

$$\binom{\lambda_i^{(1)}-1}{\lambda_i^{(2)}-1}\binom{|\bar{I}_\pi^{(i)}|-\lambda_i^{(1)}-1}{\lambda_i^{(2)}-2} = \binom{\lambda_i^{(1)}-1}{\lambda_i^{(2)}-1}\binom{|\bar{I}_\pi^{(i)}|-\lambda_i^{(1)}-1}{\lambda_i^{(2)}-\lambda_i^{(3)}-\lambda_i^{(4)}}$$

since $\lambda_i^{(3)} + \lambda_i^{(4)} = 2$. Similarly, for the cases $(\lambda_i^{(3)}, \lambda_i^{(4)}) = (0,0), (0,1), (1,0)$, we can prove that the number of ways one can choose a segment set $\tilde{J}_i$ from it with $\lambda_i^{(1)}$

number of adjacencies and $\lambda_i^{(2)}$ number of segments is

$$\binom{\lambda_i^{(1)}-1}{\lambda_i^{(2)}-1}\binom{|\bar{I}_\pi^{(i)}|-\lambda_i^{(1)}-1}{\lambda_i^{(2)}-\lambda_i^{(3)}-\lambda_i^{(4)}}.$$

Multiplying all possibilities for $i = 1, ..., \|I\| + 1$ yields the result. □

*Proof of Theorem 4.* The proof is a direct application of Lemmas 2 and 3, and (9), the inclusion-exclusion principle. □

## 5 Discussion

We introduced the notion of median inverse and used it to study the probability that any given permutation $x \in S_n$ is an approximate median of $k$ permutations $\mathcal{X} = \{\xi_1^{(n)}, \ldots, \xi_n^{(n)}\}$ chosen uniformly and independently at random from $S_n$. Due to the left-invariance of the breakpoint distance, this probability is the same for all permutations in $S_n$. Consequently, we computed the expected number of approximate medians of $\mathcal{X}$. The key was to observe that any median of $\mathcal{X}$ can have at most $\mathcal{O}_n(\mathcal{X})$ adjacencies not from the set $\cup_{i=1}^k \mathcal{A}_{\xi_i}$. We showed that $\mathcal{O}_n(\mathcal{X})$ is relatively small, and as $n \to \infty$, $\mathcal{O}_n(\mathcal{X})/a_n \to 0$ in probability for any diverging sequence $(a_n)_{n\in\mathbb{N}}$.

To determine the probability of a permutation $x$ being an approximate median of $\mathcal{X}$, we required to find the size of the approximate median inverse set $\mathscr{L}_{n,k,c}$. Our counting technique relies on partitioning the set of adjacencies of any given permutation $x$ into $2^k$ parts. Each part $\tilde{J}_A$, indexed by a unique subset $A \subseteq [k]$, specifies the adjacencies of $x$ that should be present in a permutation $x_i$ for $i \in A$ and absent in all other permutations $x_j$ for $j \notin A$. We define $J_i = \cup_{A\subseteq[k]:i\in A}\tilde{J}_A$ as the set of adjacencies of $x$ present in $x_i$. By completing the segment set $J_i$, it becomes straightforward to count the number of possible ways to construct $x_i$ such that $\mathcal{A}_{x,x_i} = J_i$. The count of such constructions, denoted by $\mathcal{H}_x(J_i)$, is computed in Theorems 3 and 4. As discussed after Remark 3, this approach also offers an efficient means to generate a $c$-approximate median $\pi$ from $k$ random permutations $\xi_1, \cdots, \xi_k$ with breakpoint distances $d(\xi_i, \pi) = d_i$, for any $d_1, \cdots, d_k \geq 0$ where $d_1 + \cdots d_k - (k-1)(n-1) \leq c$.

Geodesic points, which represent the intermediate permutations between two given permutations and serve as their medians, play a crucial role as they are instrumental in constructing accessible median genomes (Jamshidpey et al, 2014). By employing the techniques outlined in this study, we can establish an upper bound for the expected number of geodesic points between two randomly selected permutations. This bound, in turn, is very useful in analyzing geodesic point density, providing valuable insights into its implications for comparative genomics.

In summary, not only does our analysis computes the chance of a permutation serving as an approximate median of random genomes, but it also establishes a systematic method for generating such medians efficiently. Our findings provide a foundation for algorithmic approaches to quantify these probabilities effectively. Although the computations presented in this paper focus on unsigned linear unichromosomal genomes, it is important to emphasize that our methodology readily extends to all genome types. Therefore, analogous results can be obtained for signed, unsigned, unichromosomal and multichromosomal genomes with linear and/or circular chromosomes. This work holds significant promise for advancing our understanding of breakpoint median genomes, and offers a robust framework for future exploration and application.

# Statements and Declarations

## Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

## Data availability

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

# References

Bryant D (1998) The complexity of the breakpoint median problem. Centre de recherches mathematiques

Caprara A (2003) The reversal median problem. INFORMS Journal on Computing 15(1):93–113

Fertin G, Labarre A, Rusu I, et al (2009) Combinatorics of genome rearrangements. The MIT Press

Jamshidpey A (2016) Population dynamics in random environment, random walks on symmetric group, and phylogeny reconstruction. PhD thesis, Université d'Ottawa/University of Ottawa

Jamshidpey A, Sankoff D (2013) Phase change for the accuracy of the median value in estimating divergence time. BMC bioinformatics 14(15):S7

Jamshidpey A, Sankoff D (2017) Asymptotic medians of random permutations sampled from reversal random walks. Theoretical Computer Science

Jamshidpey A, Jamshidpey A, Sankoff D (2014) Sets of medians in the non-geodesic pseudometric space of unsigned genomes with breakpoints. BMC genomics 15(6):S3

Larlee CA, Zheng C, Sankoff D (2014) Near-medians that avoid the corners; a combinatorial probability approach. BMC genomics 15(6):S1

Sankoff D, Blanchette M (1997) The median problem for breakpoints in comparative genomics. In: International Computing and Combinatorics Conference, Springer, pp 251–263 da Silva PH, Jamshidpey A, Sankoff D (2024) Sampling gene adjacencies and geodesic points of random genomes. In: RECOMB International Workshop on Comparative Genomics, Springer, pp 189–210

Tannier E, Zheng C, Sankoff D (2009) Multichromosomal median and halving problems under different genomic distances. BMC bioinformatics 10(1):120