

Molecular Evolution

- The occurrence of inaccuracy.

Point mutation: *Mutation at the level of nucleotides.*

1. Substitutions
2. Insertions
3. Deletions

Large scale mutations (Rearrangements of structural variations)

Genome Rearrangement Problems: Detecting rearrangements of structural variations.

1. Deletions: a segment of the genome is lost.
2. Transpositions: move a segment to another location.



Continued

3. Inversions (or Reversals)

A segment of the genome is reversed and the strands are exchanged to its double helix pair.

Example

CCGTGCGTACTGC

CCGTCATGCGACTGC

CCGTGTACGCACTGC (After reversal of GCGTAC)

4. Duplications

A segment of the genome is copied and inserted in the genome.



Continued

- There are three types of duplications, namely (a) **Tandem** duplications (as above), (b) **Retro-transposition** (insert the copy in an arbitrary location) and (c) **Whole genome duplication**.

5. Reciprocal translocation

A segment of a chromosome that contains a **telomere** (端粒) is exchanged with a segment of another chromosome that contains a **telomere**.



6. **Fusion**: two chromosomes are joined into one.

7. **Fission**: one chromosome splits into two chromosomes.

(*) **All the above seven operations act on a genome at the level of DNA segments rather than on nucleotides.**

Homologous

Definition

Two segments are said to be **homologous** if they derived from a common ancestor and are distinguished by a **replication event (end up in two different genomes)** or by a **duplication event** both belong to the same genome.

Note. Genes are often taken as those homologous segments because, due to their functional utility, they are less subject to small mutations and are **rarely cut into smaller pieces** by rearrangements.

(*) A genome is often represented by a sequence of segments.



Combinatorics of Genome Rearrangements

- In 1936, Dobzhansky and Sturtevant proposed an indicator of an **evolutionary distance** between organisms. They proposed a scenario of “inversions” to explain chromosomal differences between 17 groups of flies.
- In 1941, Sturtevant and Novitski formulated the problem of **minimizing the number of inversions** that may explain the differences in rearrangement between two species.
- Enter the DNA sequencing era: in 1988, Palmer and Herbon point out the importance of rearrangements.
- They are able to examine **the differences in the gene order of the mitochondrial genomes of cabbage and turnip**. That is to say, they have similar sequences, but different structure.

Relation with Permutations

- In 1982, Watterson proposed the idea of representing the relative positions of genes in different genomes as “**Permutations**”.

Reference: Watterson, The chromosome inversion problem, J. Th. Biology, 99(1982), 1 – 7.

Idea

Transforming one permutation into another by means of a minimum number of “allowed operations” is in general equivalent to “sorting” a permutation by using the same operation.

The Genome Rearrangement Problem

Given a set of genomes and a set of possible evolutionary events, find a shortest set of events transforming those genomes into another.

Reversal of a Segment

- This is the most common rearrangement problem.

Definition (Reversal of a segment)

$$\text{Let } \pi = (a_1, a_2, \dots, a_n) = \begin{bmatrix} 1 & 2 & 3 & \dots & n \\ a_1 & a_2 & a_3 & \dots & a_n \end{bmatrix}.$$

The reversal of the segment $a_i a_{i+1} a_{i+2} \dots a_j$, $i < j$ is the permutation $(a_1, a_2, \dots, a_{i-1}, a_j, a_{j-1}, \dots, a_i, a_{j+1}, \dots, a_n)$, denoted by $\pi^{i,j}$.

(*) $\pi^{i,j}$ can be written as $\pi \circ \rho$ where

$$\rho = (1, 2, \dots, i-1, j, j-1, j-2, \dots, i, j+1, \dots, n).$$

Example of Reversals

- $\pi = (11\ 4\ 8\ 9\ 7\ 6\ 5\ 1\ 3\ 2\ 10)$
- $\pi^{4,7} = (11\ 4\ 8\ 5\ 6\ 7\ 9\ 1\ 3\ 2\ 10) = \pi \circ (1\ 2\ 3\ 7\ 6\ 5\ 4\ 8\ 9\ 10\ 11)$
- The most popular study of the genome rearrangement problem is sorting permutations by reversals.
- It is believed that “reversal operation” plays the most important role in evolution, from one spicity to another.



Important Comments

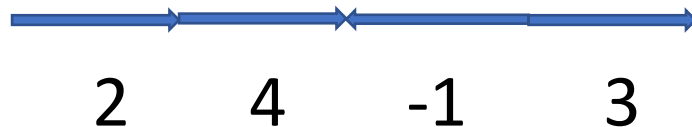
- The basic objects (genomes) we shall consider are “homologous markers” which are segments of genomes that can be found in many species. This is the reason why we say “genes” for homologous markers. **Therefore, we have the following simplified model.**
- **Assumptions**
 1. The order of genes in each genome is known (with no duplication).
 2. All genomes share the same set of genes.
 3. All genomes contain a single copy of each gene.
 4. All genomes consist of a single chromosome.

Review of Permutations

- A permutation π of the set $[n] = \{1, 2, 3, \dots, n\}$ is a bijection from $[n]$ onto $[n]$. We call π a permutation of order n .
- The set of all permutations of order n is denoted by S_n . It is known that $\langle S_n, \circ \rangle$ is a group, a symmetric group of order n . More over, S_n has $n!$ elements.
- Every permutation has a cycle decomposition. A cycle of length 2 is called a transposition.
- Any permutation can be represented by a product of transpositions. If a permutation is represented by an even (resp. odd) number of transpositions, it call an even (resp. odd) permutation.

Signed Permutation

- Signed permutations model the organization of genomes better than unsigned ones, because they take into consideration of the double helix structure of DNA.
- Each DNA strand has an orientation, and by complementary of the orientation of one strand is the reverse of the orientation of the other.
- For example, $(2, 4, -1, 3) = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & -1 & 3 \end{bmatrix}$ represents



Continued

- The reverse orientation of the above strand is (-2, -4, 1, -3).

Definition (Signed permutation)

A signed permutation of $[-n, n]$ $\pi = (a_{-n}, a_{-n+1}, \dots, a_{-1}, a_1, a_2, \dots, a_n)$ is a permutation of $[-n, n]$ that satisfies $a_{-i} = -a_i$ for each $i \in [n]$.

Example
$$\begin{bmatrix} -4 & -3 & -2 & -1 & 1 & 2 & 3 & 4 \\ 3 & -1 & 4 & 2 & -2 & -4 & 1 & -3 \end{bmatrix}$$

Definition (Hyperoctahedral group)

The set of all signed permutations of $[-n, n]$ forms a group under the operation composition of functions. The group is known as a Hyperoctahedral group, denoted by $S_n^\pm$.

Exercise 7

- Find an algorithm to sort a permutation into identity permutation by using only reversal operation as defined in page 7.

