

Idea of "Molecular Evolution"(1) 最生不正確的轉變 (Inaccuracy!)Point Mutations: Mutation at the level of nucleotides

1. Substitutions (取代 - 5 字母)
2. Insertions (插入 - 5 字母)
3. Deletions (刪去 - 5 字母)

Larger scale mutations (Rearrangements or Structural variations)

(1) Genome rearrangement problems: Detecting ↑ .)

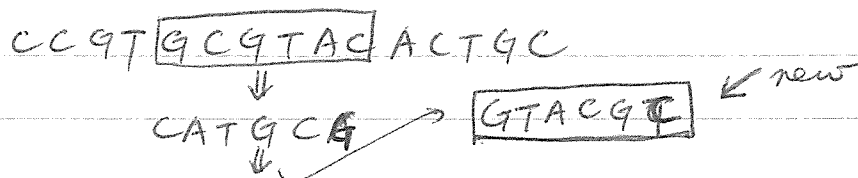
1. Deletions. A segment of the genome is lost.

2. Transpositions, moves to another location



3. Inversions (or reversals).

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



4. Duplications. A segment of the genome is copied and
(B)
inserted in the genome.



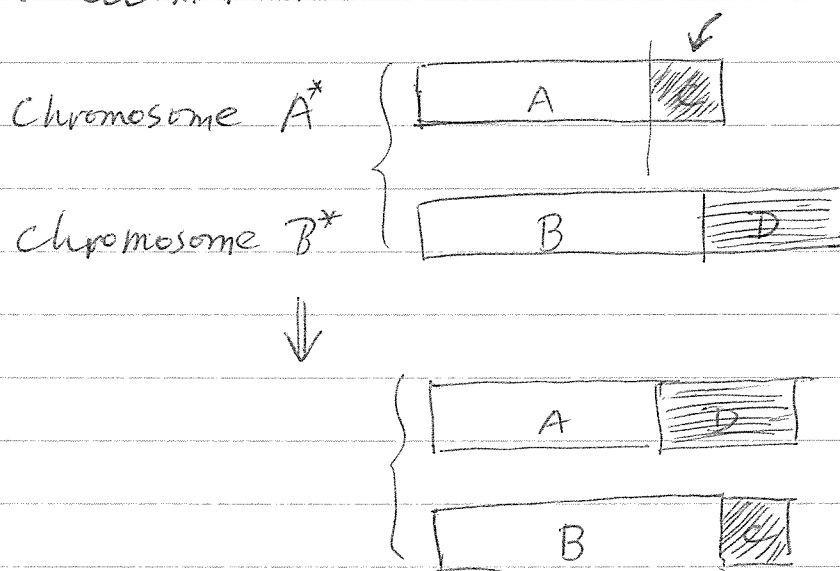
(a) Tandem duplications ↑ (複製, 然後插入在旁邊)

(b) Retrotransposition: Insert the copy in an arbitrary location.

(c) Whole genome duplications as title.

5. Reciprocal translocation.

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



6. Fusion. Two chromosomes A^* and B^* are joined into one.



7. Fission. One chromosome splits into two chromosomes.

↳ The inverse of a "Fusion".

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

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

Definition (Homologous). 同型, 同源

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

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 by rearrangements.

(比较不会再分成小的片段。)

Birth of the Combinatorics of Genome Rearrangements

- (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.

- (1941) Sturtevant and Novitski

Formulated the problem of minimizing the number of inversions that may explain the differences in rearrangement between two species

Entered the genome sequencing era

↓
Point out the importance of rearrangements
(Palmer and Herbon) (1988)

↪ Examined the differences in the gene order of the mitochondrial genomes of cabbage and turnip.

↪ 具有相似的 sequences, 不同的结构

Watterson, 1982, The chromosome inversion problem,
et.al

J. Th. Biology, 99(1982), 1-7.

proposed to represent the relative positions of genes
in different genomes as "permutations".

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.

For example, the reversal of a segment.

Definition (Reversal of a segment)

Let $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ be a permutation $\begin{pmatrix} 1 & 2 & \dots & n \\ \pi_1 & \pi_2 & \dots & \pi_n \end{pmatrix}$.

The reversal of the segment $\pi_i, \pi_{i+1}, \dots, \pi_j$, $i < j$ is the
permutation $\pi^{<i,j>} = (\pi_1, \pi_2, \dots, \pi_{i-1}, \pi_j, \pi_{j-1}, \dots, \pi_i, \pi_{j+1}, \dots, \pi_n)$.

(*) $\pi^{<i,j>}$ can be written as $\pi \circ \rho$ where $\rho = (1, 2, \dots, i-1, \underline{j, j-1, \dots, i}, j+1, \dots, n)$.

e.g. $\pi = (11\ 4\ 8\ \underline{9\ 7\ 6\ 5}\ 1\ 3\ 2\ 10)$

$$\pi^{<4,7>} = (11\ 4\ 8\ \underline{5\ 6\ 7\ 9}\ 1\ 3\ 2\ 10)$$

$$(11\ 4\ 8\ 9\ 7\ 6\ 5\ 1\ 3\ 2\ 10) \circ (1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11)$$

$$= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 11 & 4 & 8 & 5 & 6 & 7 & 9 & 1 & 3 & 2 & 10 \end{pmatrix}$$

$$= (11\ 4\ 8\ \underline{5\ 6\ 7\ 9}\ 1\ 3\ 2\ 10)$$

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 one another.

The most popular study : Sorting permutations by "reversals".

- (1) The basic objects we shall consider are "homologous markers" which are the segments of genomes that can be found in many species.

(因此,我们相信这些 Markers 应该是它(他,她)们祖先所
共同拥有 Genes!)

(*) This is the reason we say "genes" for "homologous markers."

Model (Simplified)

Assumptions

1. The order of genes in each genome is known. (With no duplications!)
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

$$[n] =_{\text{def}} \{1, 2, \dots, n\}$$

(*) A permutation π is a bijection of $[n]$ onto $[n]$.
(itself)

(*) $\begin{pmatrix} 1 & 2 & \dots & n \\ \pi_1 & \pi_2 & \dots & \pi_n \end{pmatrix}$ is denoted by $(\pi_1, \pi_2, \dots, \pi_n)$.

(*) S_n : Symmetric group of order n with function "composition" $\circ$.

$$|S_n| = n!,$$

(*) Each permutation has a cycle decomposition.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 4 & 8 & 9 & 7 & 6 & 5 & 1 & 3 & 2 & 10 \end{pmatrix} = (1, 4, 7)(2, 8, 3, 9)(5, 6)(10)$$

(cycle 有逗美)

- (*) The cycle decomposition of a permutation can be viewed as a 2-factor defined on $[n]$.
- (*) If the 2-factor corresponding to π has an even number of even cycles, then π is an even permutation.
- (*) A cycle with two elements (vertices) is also called a (2-cycle) transposition.
- (*) A permutation can be written as a product of transpositions.

e.g. $(2413) = (1, 2, 4, 3) = (1, 2) \circ (2, 4) \circ (4, 3)$

||

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}$$

- (*) An even permutation is the composition of an even number (odd) of transpositions.

e.g. $(48976513210) = (1, 4, 7)(\underline{2, 8, 3, 9})(\underline{5, 6})(10)$

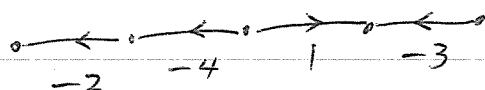
even cycles

$$= (1, 4) \circ (4, 7) \circ (2, 8) \circ (8, 3) \circ (3, 9) \circ (5, 6).$$

- (**) The alternating group A_n is the subgroup of S_n formed by all even permutations.

(*) 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 complementarity, the orientation of one strand is the reverse of the orientation of the other.



$$\left(\begin{array}{cccc|cccc} 1 & 2 & 3 & 4 & -4 & -3 & -2 & -1 \\ -2 & -4 & 1 & -3 & 3 & -1 & 4 & 2 \end{array} \right)$$

↓

$$\left(\begin{array}{cccccccc} -4 & -3 & -2 & -1 & 1 & 2 & 3 & 4 \\ 3 & -1 & 4 & 2 & -2 & -4 & 1 & -3 \end{array} \right) \text{ denoted by } (\underline{-2} \ \underline{-4} \ \underline{1} \ \underline{-3}).$$

Definition (Signed Permutation)
of $[-n, n]$,

A signed permutation $\pi = (\pi_n, \pi_{n+1}, \dots, \pi_{-1}, \pi_1, \pi_2, \dots, \pi_n)$

is a permutation of $[-n, n]$ that satisfies $\pi_{-i} = -\pi_i$ for each $i \in [n]$.

e.g. $\uparrow$ In above example, $\pi_{-4} = 3$, $\pi_4 = -3$, $\pi_{-2} = 4$, $\pi_2 = -4$.

Definition (Hyperoctahedral group)

The hyperoctahedral group, denoted by $S_n^\pm$, is the set of all signed permutations of $[-n, n]$ with composition " $\circ$ ".