Date Lecture 6 No. 1

Idea of "Molecular Evolution"

(•) 發生不正確的轉變 (Inaccuracy!)

Point Mutations : Mutation at the level of nucleotides

1. Substitutions (取代一个字母)

2. Insertions (插入一个字母)

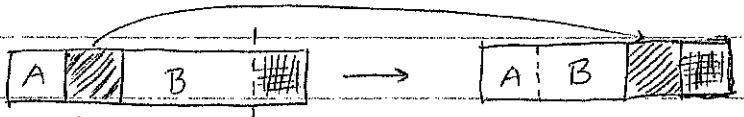
3. Deletions (刪去一个字母)

Larger scale mutations (Rearrangements or Structural variations)

(•) Genome rearrangement problems : Detecting ↑ .)

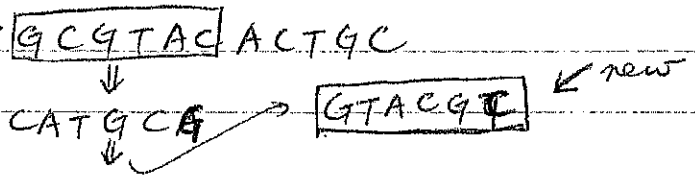
1. Deletions. A segment of the genome is lost.

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).



Genome Rearrangements

From one cell to another, one individual to another, one species to another, the content of the DNA molecules is often similar, but their organization often differs dramatically. The "mutations" that affect this organization are called genome rearrangements, and the structural differences between molecules in two genomes motivate the study of their combinatorics.

Indeed, the inference of the evolutionary events that can explain the multiple combinations of observed genomes can often be formalized as "combinatorial optimization problems".

✓ (•) A DNA molecule may evolve by point mutations (the level of nucleotide).

Three different kinds of point mutations:

(1) Substitutions, (2) Insertions and (3) Deletions.
(取代) (加入) (去掉)

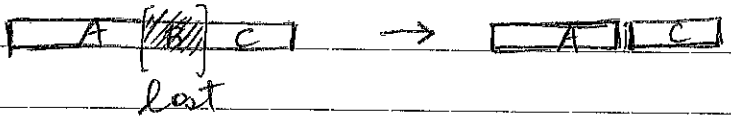
Page
Date

(*) Most of the time, a sequence may evolve by modifying its organization at a larger scale.

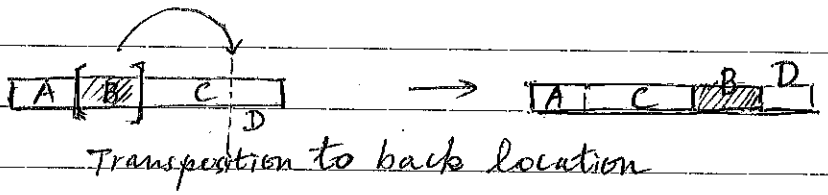
These large-scale mutations are called rearrangements (or structural variations), and detecting them is the goal of "genome rearrangement" problems.

(**) Here are types of rearrangements.

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



(2) Transpositions. A segment moves to another location



Transposition to back location

(3) Inversions (or Reversals). A segment of the genome is reversed and the strands are exchanged.

(倒过来之后，用配对那个来接下去。)

CCGTGCGTACACTGC

↓

CATGCG

↓ change strand. 配对

GTACGC

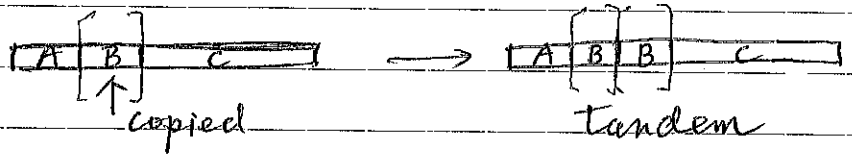
↓

CCGTGTACGCACTGC

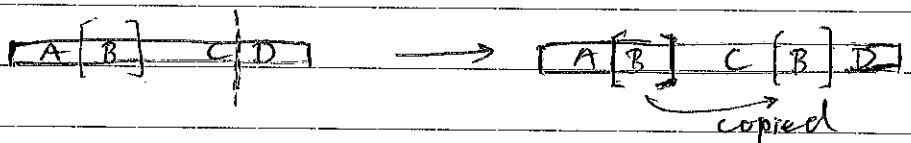
NAN PAO

(4) Duplications. A segment is copied and inserted.

(a) tandem duplications (把 copy 的部份直接接在後面)



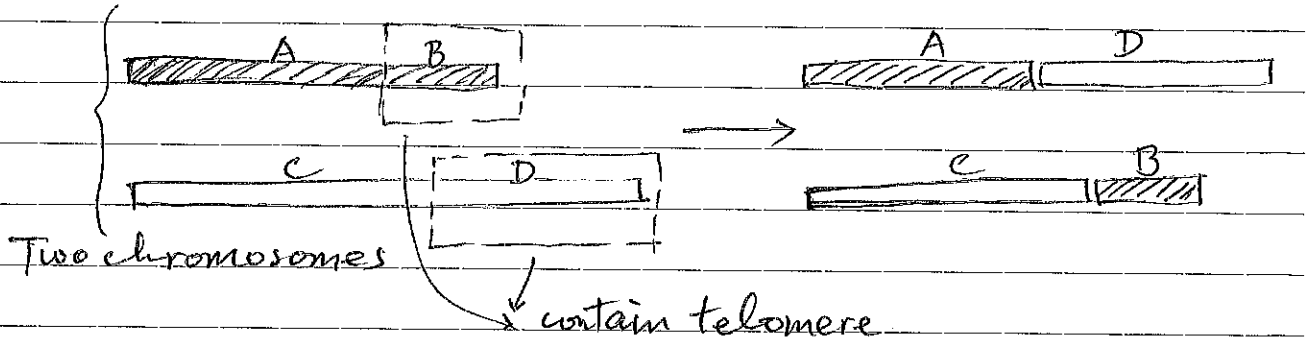
(b) retrotranspositions (把 copy 的部份插入在後面某處)



(c) whole genome duplications (整組 copy).

(5) Reciprocal translocation

A segment of a chromosome that contains a telomere 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.

(8) Horizontal (or lateral) transfer

A segment of the genome is copied from one genome to another. (常見於單核細胞, unicellular organisms)

(*) 上述的八種改變都是出現"Genome"層次上, 改變都是"片斷"的形式而是單一字母的形態。

Genome Rearrangement 發展的歷史介紹(簡要)

1936, {Dobzhansky, Sturtevant} Proposed to use the degree of disorder between the organization of genes in two different genomes as an indication of an evolutionary distance between organisms.

1941, {Sturtevant, Novitski} minimizing the number of rearrangements, for example inversions.

1982, Watterson et al. proposed to represent the relative positions of genes in different genomes as permutations.

NAN PAO

⇓ 產生Models

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.

Models (Simplest) (•) 要求滿足以下的規定,若放寬一個,就比較複雜。

1. The order of genes in each genome is known.
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.

⇓

(*) Genomes are modeled by permutations and each gene is then assigned a unique number which can found exactly once in a genome.

(**) Genomes as Permutations !

(•) A permutation π is a bijection over the set $\{1,2,\cdots,n\}$ $=_{def}[1,n]$. The image of $i \in [1,n]$ is denoted by π_i.

(The elements π_i of the permutation π are called genes.

(•) Two row notation : $\begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ \pi_1 & \pi_2 & \pi_3 & & \pi_n \end{pmatrix}$.

(*) One row notation $\pi = (\!(\pi_1\ \pi_2\ \pi_3 \cdots \pi_n)\!)$.

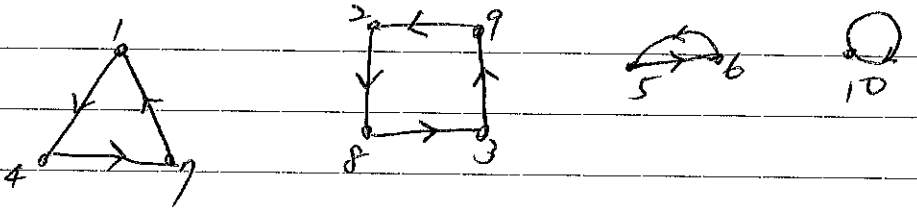
For example, $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 4 & 8 & 9 & 7 & 6 & 5 & 1 & 3 & 2 & 10 \end{pmatrix}$ is denoted by $(\!(4\,8\,9\,7\,6\,5\,1\,3\,2\,10)\!)$.

(•) It is well-known that a permutation π can also be represented by cycles (disjoint).

$(\!(4\,8\,9\,7\,6\,5\,1\,3\,2\,10)\!) =_{def} (147)(2839)(56)(10)$

(fixed point can be deleted)

Graph notation



(*) A permutation can be decomposed into disjoint cycles.

(**) The set of all permutations over $[1,n]$ "generates" a group by using function composition as its binary operation.

NAN PAO

(•) The group is known as the symmetric group of order n, S_n.

(•) A permutation is even if the number of even cycles in its cycle decomposition is even, otherwise the permutation is odd.

(*) The set of all even permutations over $[1,n]$ forms a group, alternating group, A_n. $|A_n| = \frac{1}{2}|S_n|$.

(•) $\langle\langle 4\ 8\ 9\ 7\ 6\ 5\ 1\ 3\ 2\ 10\rangle\rangle$ is an even permutation on $[1,10]$.

Signed Permutations

(•) A signed permutation on $[1,n]$ is a permutation of the set $\{-n, -n+1, \cdots, -2, -1, 1, 2, \cdots, n\}$ that satisfies $\pi_{-i} = -\pi_i$.

$$\pi =_{def} \begin{pmatrix} -n & -n+1 & \cdots & -2 & -1 & 1 & 2 & \cdots & n \\ \pi_{-n} & \pi_{-n+1} & \cdots & \pi_{-2} & \pi_{-1} & \pi_1 & \pi_2 & \cdots & \pi_n \end{pmatrix}$$

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

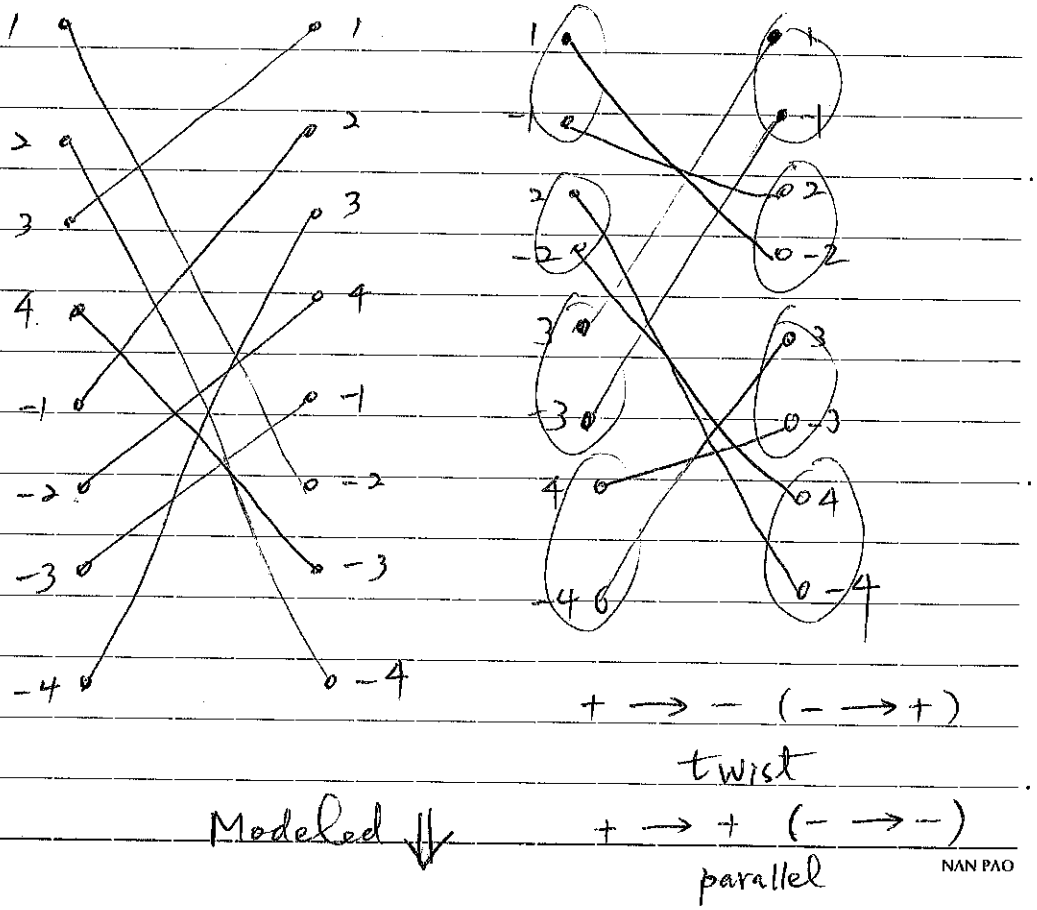
$=_{def}$ $(-2\ -4\ 1\ -3)$ (In short.) ——(✓)

$\pi = (3\ -1\ 4\ 2\ -2\ -4\ 1\ -3)$

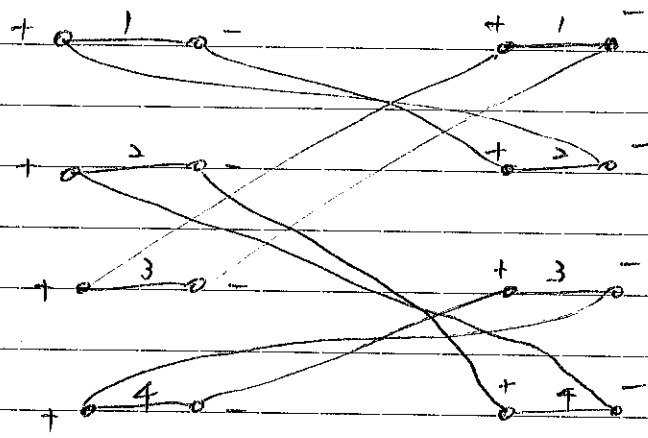
(•) See next page for graph representation.

Definition The set of all permutations (signed) on $[1,n]$ forms a group, known as the hyperoctahedral group, denoted by $S_n^{\pm}$. (合成的數為二元運算)

Note. Signed permutations model the organization of genomes better than permutations, because they take into account the double helix structure of DNA.



NAN PAO



Distance of permutations

(•) Let π and σ be two genomes (permutations). The distance of π and σ is defined by the minimum number of "operations" transfer π into σ.

(*) For example, the number of reversals or transpositions.

We shall explain more details in next lectures.