7. The Algebraic Theory of Genome Rearrangements

Literature:

Meidanis J., & Dias Z. (2000). *An Alternative Algebraic Formalism for Genome Rearrangements.* In: Sankoff D., Nadeau J.H. (eds) Comparative Genomics. Computational Biology, Springer, Vol. 1, pp. 213-223.

Dias, Z., & Meidanis, J. (2001). *Genome Rearrangements Distance by Fusion, Fission, and Transposition is Easy.* Proceedings of SPIRE 2001. pp. 250-253.

Note that although this chapter is confined to the study of unsigned permutations, the discussed theory can be extended to signed permutations. For further details, see Meidanis and Diaz (2000).

7.1. Composition of permutations

Have you ever wondered what the "$\circ$" operator does? What does $\pi \circ \rho$ mean?

The symbol "$\circ$" denotes the composition (multiplication) operator of permutations which is, just as the composition of functions, applied from right to left, i.e. in composition $\pi \circ \sigma$, we first apply sigma (to the identity permutation $(1 \cdots n)$), and then we apply π, resulting in the permutation $(\pi_{\sigma_1} \cdots \pi_{\sigma_n})$.

Example 11. *Two permutations* $\pi = (2\ 1\ 4\ 3)$, $\sigma = (1\ 4\ 2\ 3)$*;*

$$\pi \circ \sigma = (\pi_1\ \pi_4\ \pi_2\ \pi_3) = (2\ 3\ 1\ 4)$$

Then a reversal $\rho(i, j)$ corresponds to the identity permutation in which the elements $i..j$ are reversed, i.e., $\rho(i, j) := (1 \cdots i - 1 \cdot j \cdot j - 1 \cdots i \cdot j + 1 \cdots n)$.

Example 12. $\rho(2, 3)$ *corresponds to permutation* $(1\ 3\ 2\ 4)$ *and* $\pi \circ \rho(2, 3) = (2\ 4\ 1\ 3)$.

The set of all permutations of n elements, denoted Π_n, under composition operator $\circ$ forms a group $(\Pi_n, \circ)$. Its neutral element is $(1 \cdots n)$ and the inverse of a permutation $(\pi_1 \pi_2 \cdots \pi_n)$ is obtained by exchanging positions and elements in π, i.e. $\pi^{-1}_{\pi_i} = i$.

Note that in general $\pi \circ \sigma \neq \sigma \circ \pi$.

7.2. Cycles of a permutation

A permutation can also represented by a *composition of one or more cycles*:

Definition 25. *A* cycle *of a permutation π, denoted by $C = (i_1, \ldots, i_k)$ is a set of elements such that for $1 \leq j \leq k-1$ holds that $\pi_{i_j} = i_{j+1}$ and $\pi_{i_k} = i_1$.*

Important: we distinguish between a permutation and its cycle decomposition by delimiting elements in the latter by ",".

Theorem 7. *Every permutation has a unique disjoint cycle decomposition*[1].

Example 13. *The disjoint cycle composition of σ is:*

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

Note:

- $(1\ 2\ 3) \neq (1,2,3)$ but $(1,2,3) = (2,3,1) = (3,1,2)$
- the inverse of a cycle is the cycle in which elements have reversed order, the neutral element is $(1) \cdots (k)$
- cycles with a single element are usually omitted from the cycle representation e.g. $\pi^1 = (1,2)(3,4)(6,7,8)$

Permutations can also be decomposed into multiple, non-disjoint cycles. Alternative representations of the cycle $\pi^2 = (1,2,3,4,5)$ include

$$\begin{aligned} &(1,2,3)(3,4,5) \\ &(1,2)(2,3)(3,4)(4,5) \\ &(1,5)(1,4)(1,3)(1,2)\,. \end{aligned}$$

The order of non-disjoint cycles is no longer arbitrary e.g. $(1,2)(2,3) \neq (2,3)(1,2)$.

Algorithm 3 describes the procedure of obtaining the disjoint cycle decomposition of sequence of cycles.

Example 14. $\pi^3 = (1,5,3,2)$, $\pi^4 = (1,4)(3,5)$*:*

$$\pi^3\pi^4 = (1,5,3,2)(1,4)(3,5) = (1,4,5,2)$$

[1] up to the order of cycles and the rotation of elements within cycles

Algorithm 3 Construction of a disjoint cycle decomposition from a sequence of cycles

Input: Collection of cycles $\mathcal{C} = C^1, C^2, \ldots$

Output: Collection of disjoint cycles $\mathcal{C}' = C'^1, C'^2, \ldots$

$i_1 = 1$
let $\mathcal{C}'$ be an empty collection of cycles
add new cycle $C' = (i_1,)$ to $\mathcal{C}'$.
for $j = 2..n$ **do**
  $i_j = i_{j-1}$
  traverse from right to left through each cycle C of $\mathcal{C}$, update $i_j \leftarrow C_{k+1}$ if and only if $C_k = i_j$
  if $i_j = C'_1$ **then**
    let i_j be the next higher number not contained in any cycle of $\mathcal{C}'$
    associate variable C' with a new cycle $(i_j,)$
    add C' to $\mathcal{C}'$
  else
    append i_j to cycle C'
  end if
end for
return $\mathcal{C}'$

7.3. Algebraic theory meets genome rearrangements (or: going beyond reversals)

Definition 26. *A* genome *is an (unsigned) permutation π in disjoint cycle notation where each cycle corresponds to a circular chromosome.*

The following discussion is restricted to circular chromosomes, but the algebraic theory can be extended to linear chromosomes. For further details, see Feijão, P., & Meidanis, J. (2013). *Extending the Algebraic Formalism for Genome Rearrangements to Include Linear Chromosomes.* IEEE/ACM Transactions on Computational Biology and Bioinformatics, 10(4), pp. 819–831.

Example 15. $\pi^5 = (1, 4, 5)(2, 3)$ *is a genome with two chromosomes.*

A rearrangement in a genome π can be modeled by a product with a permutation ρ.

Example 16. $\rho = (2, 4, 5)$

$$\rho\pi^2 = (2, 4, 5)(1, 2, 3, 4, 5) = (1, 4, 2, 3, 5)$$

Permutation ρ is a transposition of the blocks $[2, 3]$ *and* $[4]$.

Applying a 2-cycle $\rho = (a, b)$ to a genome has the following effect:

- Fission: if a and b are in the same cycle, this cycle is split in two, separating a and b
- Fusion: if a and b are in different cycles, the cycles are joined into one

Example 17. *Fission:* $(4,5)\ \pi^5 = (1,5)(2,3)(4)$*; fusion:* $(2,5)\ \pi^5 = (1,4,2,3,5)$

7.4. The rearrangement power of permutations

Definition 27. *The* norm *of a permutation* π*, denoted by* $\|\pi\|$*, is the minimum number of 2-cycles needed to decompose* π*.*

The norm of a permutation can be seen as a measure of its rearrangement power.

Observation 11. *The norm of a cycle with* k *elements is* $k-1$*.*

Observation 12. *The norm of a permutation* π *with* n *elements and* c *disjoint cycles is* $\|\pi\| = n - c$*.*

Example 18. $\|\pi^5\| = 5 - 2 = \|(1,4)(4,5),(2,3)\| = 3$

Problem 6 (Sorting by Algebraic Rearrangements)**.** *Given genomes* π *and* σ*, find permutations* $\rho_1, \rho_2, \dots \rho_d$ *that transform* π *into* σ *such that the algebraic distance* $ad(\pi,\sigma) := \sum_{i=1}^{d} \|\rho_i\|$ *is minimum.*

How to compute $ad(\pi,\sigma)$?

$$
\begin{aligned}
\rho_d \cdots \rho_2\rho_1\pi &= \sigma && \Leftrightarrow \\
\rho_d \cdots \rho_2\rho_1 &= \sigma\pi^-1 && \Rightarrow \\
\|\rho_d \cdots \rho_2\rho_1\| &= \|\sigma\pi^-1\| && \\
\sum_{i=1}^{l} \|p_i\| &\geq \|\sigma\pi^-1\| && \text{(norm property)} \\
ad(\pi,\sigma) &\geq \|\sigma\pi^{-1}\| &&
\end{aligned}
$$

Observation 13. *We can obtain the rearrangement operations by decomposing* $\sigma\pi^{-1}$ *and* $\|\sigma\pi^{-1}\|$ *serves as lower bound of* $ad(\pi,\sigma)$*.*

Thus, a solution to Problem 6 is to find a minimal 2-cycle decomposition $\rho_d \cdots \rho_2\rho_1$ of $\sigma\pi^{-1}$. Then, by definition, $ad(\pi,\sigma) = \|\sigma\pi^{-1}\| = d-1$. Each 2-cycle corresponds to a fusion, fission.

A fission followed by a fusion can model a transposition:

Example 19. $\pi^6 = (1,3,4,2,5,6)$*:*

$$(2,4)(4,6)\ \pi^6 = (2,4,6)\ \pi^6 \quad \textit{// transposition of interval } [2..5]$$

Observation 14. *If elements* a*,* b *and* c *are in the same cycle in* π *and appear in this very order, then* $\rho = (abc)$ *is a transposition in* π*.*

Therefore, the algebraic distance $ad(\pi, \sigma) = \|\sigma\pi^{-1}\|$ is also known as *fusion, fission and transposition* (FFT) distance. But, in terms of costs, a transposition will still cost 1 fusion and 1 fission.