AN ALGEBRAIC FORMALISM FOR GENOME REARRANGEMENTS

João Meidanis

Zanoni Dias

We propose a way of thinking about genome rearrangement problems in terms of permutation groups. Every one that has worked with genome rearrangements has certainty noted a strong group theoretic flavor, but the exact relationship between the two areas has never deeply investigated. Here we relate the recent theory of genome rearrangements to the (older) theory of permutation groups, and set the ground for further advances in the area. This work has motivated by the fact that many arguments in genome rearrangements are of the form "look at the figure", and lack more formal algebraic derivation. In a way, we intend to give the area an algebraic formalism, much as analytic geometry replaced geometric arguments based on pictures.

1 Introduction

In this paper we are concerned with the genome rearrangement problem viewed as a combinatorial problem. In the general formulation of this problem we are given two genomes (or parts of genomes), viewed as ordered lists of genes (or others markers), and a set of allowed mutation events (reversal, transpositions, etc). To solve the problem we must find a minimum number of events that lead from one genome to another. In general the selection is symmetric, that is, the same series of events, taken backward, will transform the second genome into the first. We will also restrict ourselves to the case of conservative events, that is, events that do not change the available gene pool. Thus events such as duplications or deletions will not be considered in our study.

Recent developments in this field include the polynomial solution to the signed reversal case (Hannenhalli and Pevzner, 1995), the NP-hardness of unsigned reversal distance (Caprara, 1996), and partial results for transposition distance (Bafna and Pevzner, 1995; Meidanis et al., 1997b). Many doctoral dissertations were devoted to this theme (see, for instance, the theses of Vergara (Vergara, 1997), Christie (Christie, 1998), and Walter (Walter, 1999)). Transposition distance seems to be a harder problem, that has eluded researchers for many years now. Its computational complexity is still unknown.

The mathematical formalization of this problem usually begins by representing genomes as *permutations*. Thus, a genome π consisting of genes π_1, π_2, π_3, ..., π_n in this order is written as:

$$\pi = \pi_1 \pi_2 \pi_3 \dots \pi_n \tag{1}$$

meaning that π is the function (permutations are functions):

$$[1 \Rightarrow \pi_1, 2 \Rightarrow \pi_2, 3 \Rightarrow \pi_3, \dots, n \Rightarrow \pi_n]$$

that is, π maps 1 into π_1, and so on.

In this paper we will propose a different view of a genome as a permutation, namely, that π is the function:

$$[\pi_1 \Rightarrow \pi_2, \pi_2 \Rightarrow \pi_3, \dots, \pi_{n-1} \Rightarrow \pi_n, \pi_n \Rightarrow \pi_1] \tag{2}$$

that is, π maps π_1 into π_2, π_2 into π_3, and so on. Note that the last gene π_n is mapped into the first gene π_1. This is necessary, became permutations are functions that map each element into some other, and they cannot repeat images. However, this implies a circular character to our genome. But circular genomes do exist, and, as we will see in subsequent sections, the study of rearrangement of linear genomes is really not much different from circular ones.

Our goal in this note is to convince the reader that interpretation 2 is much more sensible, for a number of reasons. First, it allows us to directly apply many long know results from permutation group theory. Important tools such as breakpoints, the breakpoint graph, cycles, good cycles, bad cycles, grey edges, black edges, which served as basic building blocks for most of the advances in the field can be algebraically defined instead of graphically defined as they have been until now. Therefore, arguments that relied on pictures can now be expressed completely in algebraic terms. We consider this a powerful step towards a massive attack on such problems, much like analytic geometry is a powerful way of looking into geometric problems.

2 Permutation Groups

Permutations groups have been studied at least since the eighteenth century, when Galois wrote his much acclaimed theory for solving algebraic equations. Here we briefly recall a few classical results that are useful in genome rearrangements. For more information see references (Jacobson, 1985; MacLane and Birkhoff, 1971).

Given a base set E, a *permutation* on E is a one-to-one function from E onto itself. Permutations are composed of one or more *cycles*. A cycle involving elements a, b, c, for instance, is written:

$$(a\ b\ c)$$

meaning that a is mapped into b, which is mapped into c, which in turn is mapped back into a. Cycles can be of any length. Cycles of size 1 are not explicitly written. Thus, if we write:

$$\alpha = (a\ b\ c)$$

we implicity mean that all others elements are left in place by α, that is, $\alpha(x) = x$ for $x \neq a, b, c$. Note: $(a\ b\ c) = (b\ c\ a) = (c\ a\ b)$.

The *product* or *composition* of two permutations α, β is denoted by $\alpha\beta$. In general $\alpha\beta \neq \beta\alpha$, but when α and β are disjoint cycles they commute: $\alpha\beta = \beta\alpha$. Every permutation can be written in an unique way as a product of disjoint cycles (apart from the order of the factors. We refer to this as the *cycle decomposition* of a permutation.

The identity permutation, that maps every element into itself, will be denoted by 1. Every permutation α has an inverse α^{-1} such that $\alpha\alpha^{-1} = \alpha^{-1}\alpha = 1$. For cycles, the inverse is obtained reverting the order of the elements: $(a\ b\ c)$ is the inverse of $(c\ b\ a)$. For a general permutation, invert every cycle in its cycle decomposition.

To compute the product of α and β, $\alpha\beta$, we must keep in mind that β will be applied first, and then α, as in $\alpha\beta(x) = \alpha(\beta(x))$. Therefore, to compute a product of nondisjoint cycles we need to proceed as follows. For instance:

$$(a\ b\ c)(a\ b\ d)(c\ d\ b)$$

To compute this, we start with any element, say a, and compute its image. The element a is fixed by the rightmost cycle, then is mapped into b by the second cycle, and b is mapped into c by the leftmost cycle. So, the final destination of a is c. We then write:

$$(a\ b\ c)(a\ b\ d)(c\ d\ b) = (a\ c\ \ldots$$

and then proceed finding out the image of c: c goes to d, d goes to a, a goes to b, respectively, by the rightmost, middle, and leftmost cycle, so c is finally mapped into b. And so on. We reach the result:

$$(a\ b\ c)(a\ b\ d)(c\ d\ b) = (a\ c\ b)(d) = (a\ c\ b)$$

since singleton cycles do not need to be explicitly indicated.

One important operation is the *conjugation*. The conjugation of β by α is the permutation $\alpha\beta\alpha^{-1}$. This results in a permutation with the same structure of β but the elements are changed by α. For instance, if $\beta = (\beta_1\ \beta_2\ \ldots\ \beta_k)$ then:

$$\alpha\beta\alpha^{-1} = (\alpha(\beta_1)\ \alpha(\beta_2)\ \ldots\ \alpha(\beta_k))$$

If β is a product of disjoint cycles, each one will be affected by α in the same way to form $\alpha\beta\alpha^{-1}$. Conjugations are so important that we will have a special notation for them: $\alpha \cdot \beta$ means the same as $\alpha\beta\alpha^{-1}$.

2.1 Two- and Three-Cycles

A two-cycle, or 2-cycle, is a cycle of size 2. A three-cycle, or 3-cycle, is a cycle of order 3. It is important to know how products by 2- or 3-cycles affect an arbitrary permutation. It is simple, too.

Let $\alpha = (a\ b)$ be a 2-cycle. Its affect on an arbitrary permutation β can be described as follows. If a and b are in the same cycle in β, this cycle is broken in two in $\alpha\beta$. If a and b are in two distinct cycles in β, these two cycles become one in $\alpha\beta$. This is valid when we talk about the unique cycle decompositions of β and $\alpha\beta$.

The same results are valid for $\beta\alpha$. Notice that $\beta\alpha$ and $\alpha\beta$ are conjugates: $\alpha(\beta\alpha)\alpha^{-1} = \alpha\beta$, and therefore have the same cycle structure.

Now take an arbitrary 3-cycle $\alpha = (a\ b\ c)$ and an arbitrary permutation β. Three cases appear:

1. If a, b, and c are in three different cycles in β, these three cycles become a single cycle in $\alpha\beta$.
2. If two of a, b, c, are in the same cycle, and the third element is in a different cycle in β, then these two cycles recombine into another two cycles in $\alpha\beta$. Thus, the total number of cycles is maintained.
3. If a, b, and c are all in the same cycle in β, the result depends on the orientation they have in this cycle can have the form $(a\ \ldots\ b\ \ldots\ c\ \ldots)$ or $(a\ \ldots\ c\ \ldots\ b\ \ldots)$. In the first case, the cycle becomes $(a\ \ldots\ c\ \ldots\ b\ \ldots)$ in $\alpha\beta$. In the second case, the cycle breaks into $(a\ \ldots)(b\ \ldots)(c\ \ldots)$ in $\alpha\beta$.

The same results (except for the exact format of the results cycles in case 3) are valid for $\beta\alpha$.

3 Genome Rearrangements

To formalize genome rearrangement problems we will use as base set for the permutations the set $E_n = \{-1, +1, -2, +2, \ldots, -n, +n\}$, where n é the number of genes. Thus, we will be modeling both strands of the underlying DNA molecule. Each element $+i$ or $-i$ represents a marker on the i^{th} gene, with its opposite meaning a marker in the same location in the opposite strand. We will first model circular genomes, which conform more naturally to the formalism, and will later comment on the necessary adaptations for linear genomes.

To begin with, let γ be the permutation that takes each elements into its counterpart on the other strand. The permutation γ can be written as:

$$\gamma = (-1\ \ +1)(-2\ \ +2)\ldots(-n\ \ +n)$$

that is, a product of n disjoint 2-cycles.Notice that $\gamma(a) \neq a$ for all $a \in E_n$, and $\gamma^2(a) = \gamma(\gamma(a)) = a$ for all $a \in E_n$. In other hands, $\gamma^2 = 1$ or, equivalently, $\gamma^{-1} = \gamma$.

A cycle is *admissible* when it does not contain $-i$ and $+i$. Thus, γ is far from being an admissible cycle. An admissible cycle of size n is called a *genome strand*, because it models a strand of a genome formed by these n genes in some order. If

α	reverse complement
$(+3\ \ -1\ \ +7\ \ +5)$	$(-5\ \ -7\ \ +1\ \ -3)$
$(+2\ \ +4\ \ +6)$	$(-6\ \ -4\ \ -2)$

Table 1: Examples of admissible cycles α and its reverse complement.

we have an admissible cycle α, we can compute its reverse complement, as in the examples of the Table 1.

These is an algebraic way of obtaining the complement, though. If α is an admissible cycle, α^{-1} is its reverse; $\gamma \cdot \alpha = \gamma\alpha\gamma$ is its complement. The reverse complement is when we do both: $(\gamma \cdot \alpha)^{-1}$ or $\gamma \cdot (\alpha^{-1})$, which results in the same expression $\gamma\alpha^{-1}\gamma$.

Given a genome strand π_1, its reverse complement $\pi_2 = \gamma\pi_1^{-1}\gamma$ forms the complementary strand of the same genome. We represent this genome as the product of the two strands: $\pi = \pi_1\pi_2$. Since the strands form two disjoint cycles it doesn't matter in which order we take the product: $\pi_1\pi_2 = \pi_2\pi_1$. Also, it doesn't matter which strand we call π_1: had we started with π_2 we would have computed its reverse complement π_1 and find genome would have been the same. This is just as DNA should be: no matter which strand you pick, when you let it pair with its reverse complement, you get the same DNA molecule.

Formally, we define a *genome* as a permutation that can be written as $\pi_1\gamma\pi_1^{-1}\gamma$, for some genome strand π_1. Notice that $\gamma\pi\gamma = \pi^{-1}$ for every genome π. The general genome rearrangement problem then becomes: given two genomes π and σ and a class of operations, find the minimum number of events (operations) that transform π into σ. This minimum number is called the *distance* between π and σ.

We will talk about several classes of operations later, but for any of the several kinds of problems obtained by choosing a different set of operations, the *breakpoint graph* plays an important role. Classically, the breakpoint graph is constructed as in the Figure 1.

Details of the construction have been described previously several times and will not repeated here (Hannenhalli and Pevzner, 1995; Meidanis and Setubal, 1997). Our objective is to obtain this graph, or an equivalent structure, by algebraic manipulations. The breakpoint graph is used when we want to transform π into a constant permutation $\sigma = (+1\ \ +2\ \ +3\ \ +4\ \ +5)$. It then depends on both π and σ. In fact, it has been defined for linear genomes, and to adapt to that we need consider "extended" versions of π and σ: $\pi_1 = (+0\ \ -3\ \ +2\ \ -5\ \ -4\ \ +1)$ and identify -0 with -6.

The breakpoint graph is composed of black edges, which depend only on π, and of grey edges, which depend only on σ. Its turns out that $\gamma\pi$ is a product of 2-cycles, which composed exactly to the black edges. And $\gamma\sigma$ corresponds to $\gamma\sigma$

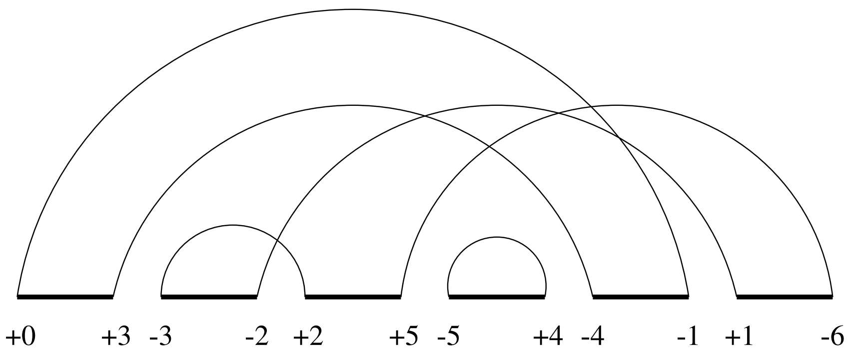

Figure 1: Breakpoint Graph for genomes $\pi = (-3 \ +2 \ -5 \ -4 \ +1)$ and $\sigma = (+1 \ +2 \ +3 \ +4 \ +5)$.

in the same way. In the preceding example, we have:

$$\begin{aligned} \gamma\pi &= (-0 \ +0)(-1 \ +1)(-2 \ +2)(-3 \ +3)(-4 \ +4)(-5 \ +5) \\ &\quad\ (+0 \ -3 \ +2 \ -5 \ -4 \ +1)(-1 \ +4 \ +5 \ -2 \ +3 \ -0) \\ &= (+0 \ +3)(-3 \ -2)(+2 \ +5)(-5 \ +4)(-4 \ -1)(+1 \ -0) \end{aligned}$$

exactly the black edges. And $\gamma\sigma$ will give the grey edges:

$$\gamma\sigma = (+0 \ -1)(+1 \ -2)(+2 \ -3)(+3 \ -4)(+4 \ -5)(+5 \ -0)$$

In the classical theory of genome rearrangements the cycle structure of the breakpoints graph plays an important role. Although we could not obtain the cycles themselves of the breakpoint graph, we derived an algebraic expression for the *square* of each cycles. This expression is just the product $(\gamma\pi)(\gamma\sigma) = \gamma\pi\gamma\sigma$. In the example, we have:

$$\gamma\pi\gamma\sigma = (+0 \ -4)(+3 \ -1)(-3 \ +5 \ +1)(-2 \ -0 \ +2)(-5)(+4)$$

For each cycle of breakpoint graph we have two cycles in $\gamma\pi\gamma\sigma$. If the cycle in the breakpoint graph is $(a_1 \ a_2 \ \ldots \ a_{2k})$, we have $(a_1 \ a_3 \ \ldots \ a_{2k-1})$ and $(a_{2k} \ a_{2k-2} \ \ldots \ a_2)$ in $\gamma\pi\gamma\sigma$. Therefore, this is not exactly the square of breakpoint cycle, because one of them is reversed. Strictly speaking, we could not talk about cycles (in the permutation group sense) of the breakpoint graph, since they have no orientation. This in part explains why one cycle in the square is reversed. Had we taken $\gamma\pi\gamma\sigma$ the other cycle would have been reversed.

In any case, these constructions allow us to rephrase all the technical properties of breakpoint graphs in algebraic terms. For instance, how many breakpoints π has with respect to σ? This is just the number of elements not fixed by $\gamma\pi\gamma\sigma$, divided by 2:

$$b(\pi, \sigma) = \frac{|Supp(\gamma\pi\gamma\sigma)|}{2}$$

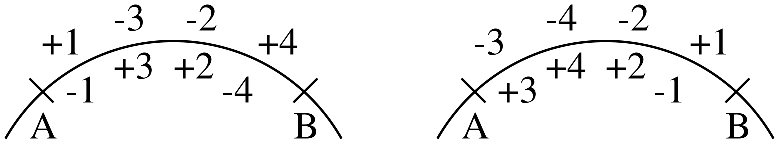

Figure 2: Linear genomes with fixed extremes.

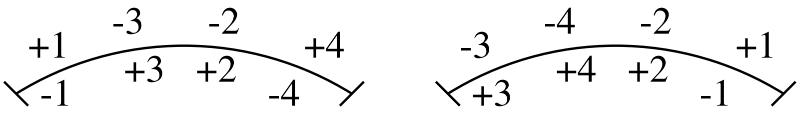

Figure 3: Linear genomes without fixed extremes.

where the support $Supp(\alpha)$ of a permutation α is the set of elements not fixed by α:

$$Supp(\alpha) = \{x \in E_n | \alpha(x) \neq x\}$$

Likewise, the number of cycle in the breakpoint graph is half the number of cycles in $\gamma\pi\gamma\sigma$. We can define also the length (size) of the cycles, good or bad cycles, cycle that can be broken by certain operations, interleaving cycles, hurdles, fortress, all in algebraic terms.

3.1 Linear Genomes

Up to now, the theory fits nicely with circular genomes. In this section we will treat the case of linear genomes.

First, we must recognize that there are actually two kinds of linear genome: with and without fixed extremes. Let us define each kind, starting with the one with fixed extremes.

When we compare two regions of two different genomes, and these regions are flanked by conserved parts, we need to use the fixed-extreme case (Figure 2). In this case, we add an extra dummy gene BA, which represents the fixed extremities of the regions, and proceed as in the circular case.

When we compare two entire linear genomes, we need to take into account that there is a free reversal that can be applied, so the distance in this case becomes:

$$d_{free} = \min(d_{fixed}(\pi, \sigma), d_{fixed}(\gamma \cdot \pi, \sigma))$$

More details on the relationship between linear and circular genome rearrangement problems can be found in the references (Walter, 1999; Meidanis et al., 1997a).

3.2 Operations

We will define in this section the events (operations) of reversal, transposition (or block move), and block interchange, some in the their signed and unsigned

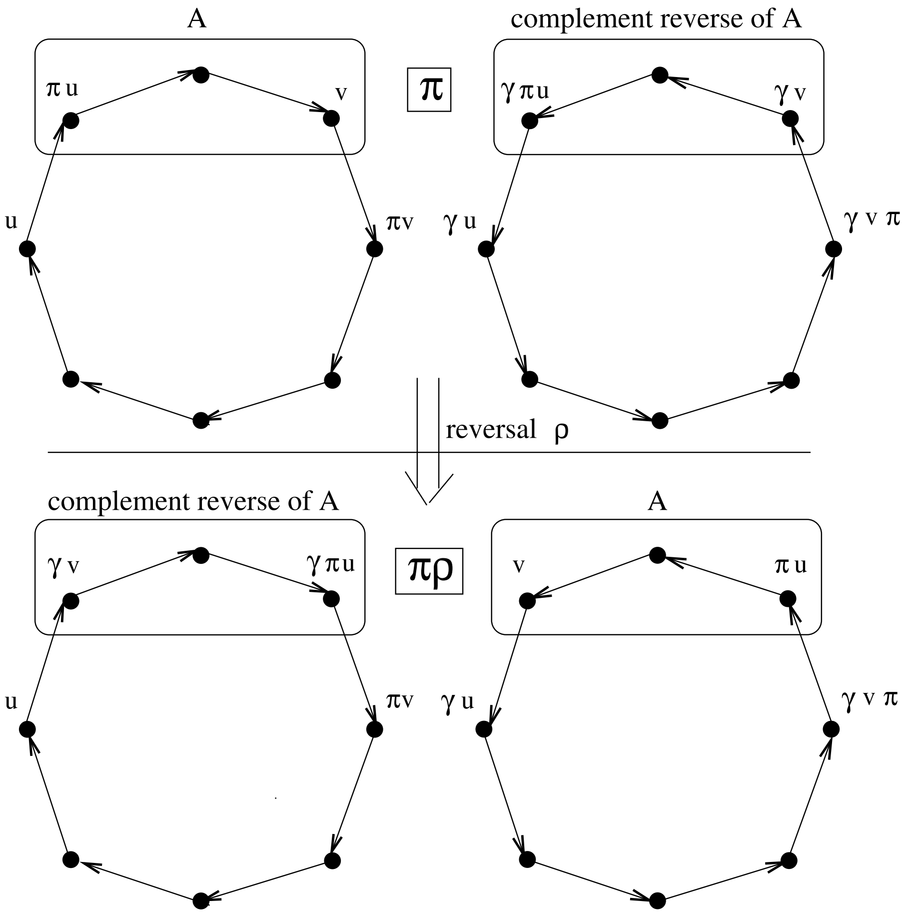

Figure 4: A reversal ρ applied on genome π.

version. We will do a very detailed job for (unsigned) reversals, and then just state the results for the others, to save space.

Given a genome π, to perform a reversal (signed) on it we need to choose two distinct markers u and v, in the same strand of π, and then replace the path from u to v (excluding u but including v) by its reverse complement in this strand. Of course, a similar operation will be performed on the other strand, to make sure the final result is still a valid genome. Figure 4 shows what is meant.

We want to write the resulting genome σ as $\pi\rho$, where ρ is a permutation that will represent the reversal. With some work, we see that σ differs from π only in the following mappings:

$$\sigma u = \gamma v, \;\; \sigma v = \gamma u, \;\; \sigma\gamma\pi u = \pi v, \;\; \sigma\gamma\pi v = \pi u$$

therefore, $\rho = \pi^{-1}\sigma$ maps:

$$\rho u = \pi^{-1}\gamma v,\ \ \rho v = \pi^{-1}\gamma u,\ \ \rho\gamma\pi u = v,\ \ \rho\gamma\pi v = u$$

with all other elements fixed by ρ. Noting that $\pi^{-1}\gamma = \gamma\pi$ for every genome π, we arrive at:

$$\rho u = \gamma\pi v,\ \ \rho v = \gamma\pi u,\ \ \rho\gamma\pi u = v,\ \ \rho\gamma\pi v = u$$

or, written otherwise:

$$\rho = (u\ \gamma\pi v)(v\ \gamma\pi u)$$

This is then the general formula of a reversal applicable to π, where u and v are two elements in same strand of π. We say that *π and σ differ by a reversal* when there is such a reversal ρ with $\sigma = \pi\rho$. notice that the definition of a reversal depends on π. There is no way to define a class of permutations that will be "the reversals", valid for all genomes. Each genome has a particular set of reversals that can be applied to it, and this sets varies from one genome to another.

For this reason, we had to define when two genomes differ by a reversal, rather than what is a reversal. The former is a global concept, the latter a local concept, because it depends on the genome.

The reversal distance problem is: given two genomes π and σ, find the minimum k such that there are genomes δ_0,δ_1,...,δ_k with $\pi = \delta_0$, $\sigma = \delta_k$ and δ_i differs from δ_{i+1} by a reversal, for $i = 0\ldots(k-1)$.

An *unsigned reversal* is defined similarly, but has the form:

$$\bar{\rho} = (u\ v)(\gamma\pi v\ \gamma\pi u)$$

where, as before, u and v are distinct elements of the same strand in π.

A *transposition* (unsigned) is defined as:

$$\tau = (u\ v\ w)(\gamma\pi w\ \gamma\pi v\ \gamma\pi u)$$

where u, v and w are three distinct elements in the same strand in π, appearing in this order (u, v, w) in the strand.

A *signed transposition* is defined as:

$$\bar{\tau} = (u\ v\ \gamma\pi w)(w\ \gamma\pi v\ \gamma\pi u)$$

where u, v and w are distinct elements in the same strand in π, appearing in this order (u, v, w) in this strand. A signed transposition models the event in which a block detaches itself from a genome and reappears elsewhere, in the same strand but the block is reversed.

A *block interchange* is defined as:

$$\beta = (u\ w)(\gamma\pi w\ \gamma\pi u)(v\ x)(\gamma\pi x\ \gamma\pi v)$$

where u, v, w and x are four distinct elements in the same strand of π, appearing in the order (u, v, w, x) in this strand.

Each one (or a group of) of these types of events can be used to define a genome rearrangement problem: given two genomes π and σ, find the minimum k such that there are genomes δ_0,δ_1,...,δ_k with $\pi = \delta_0$, $\sigma = \delta_k$ and δ_i differs from δ_{i+1} by the specific operation (or a group of), for $i = 0\ldots(k-1)$.

4 Using the Theory

We will use the theory developed to show two results whose proof was based on pictorial representations. The first result appears in Christie's proof that a block interchange cannot create three cycles (Christie, 1997). The other result is that there is only one way for a transposition to break a cycle, proved by Walter & colleagues by reference to a picture (Walter et al., 1998).

Theorem 1 *Let π and σ be two genomes, and β a block interchange on π. Then the number of cycles in $(\pi\beta)^{-1}\sigma$ is not higher than 4+ number of cycles in $\pi^{-1}\sigma$.*

Proof: We have $(\pi\beta)^{-1}\sigma = \beta^{-1}\pi^{-1}\sigma$, and differs from $\pi^{-1}\sigma$ by a multiplication by β^{-1}. The cycle structure of β^{-1} is the same as β's: four 2-cycles. By the classical results about products by 2-cycles it is immediate that multiplying 4 2-cycles we cannot create more then 4 extra cycles. □

Theorem 2 *Let π and σ be two genomes, and $\tau = (u\ v\ w)(\gamma\pi w\ \gamma\pi v\ \gamma\pi u)$ a (unsigned) transposition on π, where u,v,w appear in this order in the same cycle of π. Then $(\pi\tau)^{-1}\sigma$ has four more cycles than $^{-1}\sigma$ if and only if u,v,w are in the same cycle in $\pi^{-1}\sigma$ and appear in the order (u, v, w) in this cycle.*

Proof: Transposition do not mix genome strands (at least unsigned transpositions, which is the kind we are using here), and therefore we know that the elements of a strand of π will form a strand in $\pi\tau$ (possibly in different order). Let π_1 be the strand that contains u,v and w, and σ_1, the corresponding strand in σ. We then have $\pi = \pi_1\gamma\pi_1^{-1}\gamma$, $\sigma = \sigma_1\gamma\sigma_1^{-1}\gamma$, and $\pi^{-1}\sigma = \pi_1^{-1}\sigma_1(\gamma\pi_1\gamma)(\gamma\sigma_1^{-1}\gamma)$. Then $(\pi\tau)^{-1}\sigma = \tau^{-1}\pi^{-1}\sigma$ will be the product of disjoint permutations $(w\ v\ u)\pi_1^{-1}\sigma_1$, and $(\gamma\pi u\ \gamma\pi v\ \gamma\pi w)(\gamma\pi_1\gamma)(\gamma\sigma_1^{-1}\gamma)$.

In the first component $(w\ v\ u)\pi_1^{-1}\sigma_1$ we have a product of a 3-cycle by $\pi_1^{-1}\sigma_1$. We know from classical permutation group theory that this produces two extra cycles if and only if u,v,w appear in the same cycle of $\pi_1^{-1}\sigma_1$ in the order (u, v, w) as stated. □

Acknowldgments

Research supported by grants from FAPESP.

References

Bafna, V. and Pevzner, P. (1995). Sorting by transpositions. In *Proc. 6th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 614–623.

Caprara, A. (1996). Sorting by reversals is difficult. Technical report, DEIS,University of Bologna.

Christie, D. A. (1997). Sorting permutations by block-interchange. *Information Processing Letters*, 60(4):165–169.

Christie, D. A. (1998). *Genome rearrangement problems*. PhD thesis, Glasgow University.

Hannenhalli, S. and Pevzner, P. (1995). Transforming cabbage into turnip (polynomial algorithm for sorting signed permutations by reversals). In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing*, pages 178–189.

Jacobson, N. (1985). *Basic Algebra*. W. H. Freeman and Company, New York.

MacLane, S. and Birkhoff, G. (1971). *Algebra*. The Macmillan Company, London, sixth printing edition.

Meidanis, J. and Setubal, J. C. (1997). *Introduction to Computational Molecular Biology*. PWS Publishing Co.

Meidanis, J., Walter, M. E., and Dias, Z. (1997a). Distancia de reversao de cromossomos circulares. In *Proc. XXIV Seminario Integrado de Software e Hardware - SEMISH'97*, pages 70–79.

Meidanis, J., Walter, M. E., and Dias, Z. (1997b). Transposition distance between a permutation and its reverse. In *Proc. 4th South American Workshop on String Processing - WSP'97*, pages 70–79.

Vergara, J. P. (1997). *Sorting by Bounded Permutations*. PhD thesis, Virginia Polytechnic Institute and State University.

Walter, M. E. (1999). *Algoritmos para Problemas em Rearranjo de Genomas*. PhD thesis, University of Campinas. In Portuguese.

Walter, M. E., Dias, Z., and Meidanis, J. (1998). Reversal and transposition distance of linear chromosomes. In *String Processing and Information Retrieval: A South American Symposium - SPIRE'98*. Submitted to Journal of Computational Biology.

Instituto de Computacao, Universidade Estadual de Campinas, Campinas-SP, Brasil.
E-mail: meidanis@dcc.unicamp.br

Instituto de Computacao, Universidade Estadual de Campinas, Campinas-SP, Brasil.
E-mail: zanoni@dcc.unicamp.br