On maximal instances for the original syntenic distance

Cedric Chauve

LaCIM et Département d'Informatique
Université du Québec à Montréal

Guillaume Fertin

LINA, Université de Nantes

Theoretical Computer Science, 326(1-3):29-43, 2004.

On maximal instances for the original syntenic distance

Cedric Chauve[1] and Guillaume Fertin[2]

[1] LaCIM et Département d'Informatique,
Université du Québec à Montréal
Case Postale 8888, Succursale Centre-Ville, H3C 3P8, Montréal (QC), Canada
chauve@lacim.uqam.ca
[2] LINA FRE 2729, Université de Nantes
2 rue de la Houssinière, BP 92208 - F44322 Nantes Cedex 3, France
fertin@lina.univ-nantes.fr

Abstract. The syntenic distance between two genomes has been introduced by Ferretti, Nadeau and Sankoff as an approximation of the evolutionary distance between genomes for which the gene order is not known. This distance is the minimum number of fusions, fissions and translocations required to transform a genome into an other. Kleinberg and Liben-Nowell, as well as Pisanti and Sagot, proved independently that for genomes with n chromosomes the diameter for this distance is $2n-4$. Pisanti and Sagot also generalized this result, showing that the maximal distance between a genome with m chromosomes and a genome with n chromosomes is $n+m-4$. Kleinberg and Liben-Nowell asked for a characterization of maximal instances for the syntenic distance (pairs of n-chromosome genomes at a distance of $2n-4$). In this paper, we give such a characterization, and show that we can extend it to pairs of genomes with respectively n and m chromosomes that are at maximal distance.

1 Introduction

In the past few years, the interest in determining evolutionary distance between species, in the framework of genome rearrangement, has been continuously growing. In this context, the evolution models are at a genomic level, where mutations take place between large pieces of DNA, thus affecting the order of one or several genes within chromosomes. If one considers two genomes $\mathcal{G}_1$ and $\mathcal{G}_2$, the distance between $\mathcal{G}_1$ and $\mathcal{G}_2$ is usually defined as the minimum number of mutations needed to transform one genome into the other. Depending on the mutations that are considered (and thus allowed in the model), this gives rise to several distinct problems. One can cite for instance the *reversal distance* [1, 3, 7], where mutations are described in terms of reversals of chromosomes segments, or the *transposition distance* [2], where a transposition is a mutation that takes a segment out of a chromosome and inserts it at another location in this chromosome. Several other models or variants also exist, that either take or do not take into account the order of the genes within chromosomes.

In this paper, we are interested in the *syntenic distance*, where the order of genes within chromosomes is not considered; hence, every chromosome is seen as an unordered set of genes. This notion has been introduced by Ferretti, Nadeau and Sankoff [6]. In this model, the mutations that are allowed are threefold : (i) *fusion*: two chromosomes are joined to form one, (ii) *fission*: a chromosome is split into two, and (iii) *translocation*: two chromosomes exchange arbitrary subsets of their genes (one of these subsets can be empty). Among others, it has been shown that computing the syntenic distance between two arbitrary genomes is *NP*-hard [5]; moreover, an approximation algorithm with approximation ratio 2 is known [5], and some other variants of such an approximation algorithm have been given [6, 11]. We also mention that some specific subclasses of instances have been considered, such as *linear synteny*, *exact synteny* or *nested synteny* [5, 10, 13]. In some of these cases, the distance can be computed in polynomial time.

However, the algorithmic issue of approximating precisely the original syntenic distance between two genomes still asks for a better solution, and thus for a better understanding of the combinatorial nature of this problem. Among the natural combinatorial notions related to distances over a set of objects is the *diameter*, that is the maximum distance between two of these objects (here two genomes over the same set of genes). For the syntenic distance, the diameter for pairs of n-chromosome genomes has been shown to be equal to $2n-4$ [9, 13], and one particular instance reaching this value has been given. As a natural extension of these results, Kleinberg and Liben-Nowell asked for a characterization of those instances that are maximal (that is, pairs of n-chromosome genomes at a syntenic distance of $2n-4$) [9]. The computation of the syntenic diameter was later generalized by Pisanti and Sagot [13] who proved that the maximal distance between an n-chromosome genome and an m-chromosome genome, called the *bidimensional syntenic diameter*, is $n+m-4$. The main result of the present article is an answer to the question of Kleinberg and Liben-Nowell about maximal instances, that we generalize to the bidimensional case. Our characterization will moreover allow us to decide in polynomial time whether a pair of genomes is at maximal distance.

In Section 2, we formally state the problem and its model, we recall some known properties and we introduce some notations and definitions. In Section 3, we give necessary and sufficients conditions for square instances (pairs of n-chromosome genomes) to be maximal. Finally, in Section 4, we fully characterize those instances that are maximal, and we extend this characterization to bidimensional instances.

2 Preliminaries

Syntenic distance. Following [6], we define a *genome* $\mathcal{G}$ over a given set of *genes* as a *partition* of this set of genes into an unordered collection of *chromosomes* (sometimes called *synteny sets*). Hence the order among chromosomes and the order of genes on a chromosome are not taken into account, and a given gene cannot appear into several chromosomes. In the syntenic distance model, the mutations considered are the *fusions* of two chromosomes (they are joined to form one chromosome), the *fissions* of a chromosome (it is split into two chromosomes) and the *translocations* between two chromosomes (they exchange arbitrary subsets of their genes, where one of these subsets can be empty). These mutations never involve, either as an input or as a result, empty chromosomes and do not cause the duplication of a gene.

For example let $\{a,b,c,p,q,r,x,y\}$ be a set of genes, and $\mathcal{G}_1 = \{\{a,b,c\},\{p,q,r\},\{x,y\}\}$ be a genome with 3 chromosomes. The genomes $\mathcal{G}_2 = \{\{a,b\},\{c\},\{p,q,r\},\{x,y\}\}$, $\mathcal{G}_3 = \{\{a,b,c,x,y\}, \{p,q,r\}\}$ and $\mathcal{G}_4 = \{\{a,p\},\{b,c,q,r\},\{x,y\}\}$ result respectively from a fission of the first chromosome of $\mathcal{G}_1$, a fusion of the first and third chromosomes of $\mathcal{G}_1$, and a translocation between the first and second chromosomes of $\mathcal{G}_1$.

Given two genomes $\mathcal{G}_1$ and $\mathcal{G}_2$ over the same set of genes (there is no gene that appears in only one of the two genomes), the *syntenic distance* is the minimum number of mutations (fusions, fissions and translocations) needed to transform $\mathcal{G}_1$ into $\mathcal{G}_2$. This distance will be denoted $\mathcal{D}(\mathcal{G}_1, \mathcal{G}_2)$.

Compact representation of an instance. Let $[n] = \{1,2,\ldots,n\}$ and let an instance of the syntenic distance be specified by two genomes, $\mathcal{G}_1 = \{\mathcal{T}_1,\ldots,\mathcal{T}_m\}$ (where $\mathcal{T}_i$ is the set of genes of the i^{th} chromosome) and $\mathcal{G}_2 = \{\mathcal{U}_1,\ldots,\mathcal{U}_n\}$, on the same set of genes. The *compact representation* of $\mathcal{G}_1$ with respect to $\mathcal{G}_2$, first introduced in [6], is an unordered collection of m subsets $\{\mathcal{S}_1,\ldots,\mathcal{S}_m\}$ of $[n]$ obtained by replacing in the sets $\mathcal{T}_1,\ldots,\mathcal{T}_m$ every gene g by the indices of the chromosomes of $\mathcal{G}_2$ containing g. We should immediately notice that in the compact representation, every element of the underlying set $[n]$ can appear in several of the sets $\mathcal{S}_1,\ldots,\mathcal{S}_m$.

For example, let $\{a,b,c,p,q,r,x,y\}$ be a set of genes, and $\mathcal{G}_1 = \{\{p,q,x\},\{a,b\},\{c,r\},\{y\}\}$ and $\mathcal{G}_2 = \{\{a,b,c\},\{p,q,r\},\{x,y\}\}$ be two genomes. The compact representation of $\mathcal{G}_1$ with respect to $\mathcal{G}_2$ $\{\{2,3\},\{1\},\{1,2\},\{3\}\}$.

Given a compact representation of a genome, one can perform on it fusions, fissions and translocations, provided that one does not duplicate an element in the compact representation of a chromosome. For example the fusion of $\{1,2\}$ and $\{2,3\}$ gives $\{1,2,3\}$. Now let $\mathcal{G}_1$ and $\mathcal{G}_2$ be an instance of the syntenic distance, where $\mathcal{G}_1$ has m chromosomes and $\mathcal{G}_2$ has n chromosomes, $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_m\}$ be the compact representation of $\mathcal{G}_1$ with respect to $\mathcal{G}_2$ and $\overline{\mathcal{G}}_n = \{\{1\}, \dots, \{n\}\}$. It has been proved that $\mathcal{D}(\mathcal{G}_1, \mathcal{G}_2)$ is the minimum number of fusions, fissions and translocations needed to transform $\mathcal{S}$ into $\overline{\mathcal{G}}_n$ [5, 6]. We call $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_m\}$ the compact representation of the instance $(\mathcal{G}_1, \mathcal{G}_2)$ for the syntenic distance.

Remark 1. It should be noticed that in several papers about the syntenic distance between genomes, authors consider a genome as a *collection* of subsets of the set of genes, and not a *partition* of the set of genes. Following this definition, they allow the repetition of genes in the same genome, and in this framework the equivalence with compact instances does not hold anymore. The only published proof of the equivalence between general instances and compact instances in [5] does not make this point clear. However, it is not difficult to see that minor modifications make this proof hold in the case where no gene is repeated.

From now on, we consider only the compact representation of instances of the syntenic distance, that we call simply *syntenic instances*, that is collections[1] $\{\mathcal{S}_1, \dots, \mathcal{S}_m\}$ of subsets of $[n]$ (the two considered genomes $\mathcal{G}_1$ and $\mathcal{G}_2$ have respectively m and n chromosomes) and the fusions, fissions and translocations for compact instances. For such a syntenic instance, we call an *optimal mutation sequence* any minimal (in terms of number of mutations) sequence of mutations (fusions, fissions and translocations) that transform $\{\mathcal{S}_1, \dots, \mathcal{S}_m\}$ into $\overline{\mathcal{G}}_n$. One says that such a sequence *solves* the syntenic instance $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_m\}$ and one denotes by $\mathcal{D}(\mathcal{S})$ the length of such an optimal sequence. A syntenic instance is said to be an *n-square syntenic instance* if $n = m$. If $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_m\}$ is the compact representation of the instance $(\mathcal{G}_1, \mathcal{G}_2)$, we denote by $\mathrm{Dual}(\mathcal{S})$ the compact representation of the instance $(\mathcal{G}_2, \mathcal{G}_1)$ (that is the compact instance where $\mathcal{S}' = \{\mathcal{S}'_1, \dots, \mathcal{S}'_n\}$ is the compact representation of $\mathcal{G}_2$ with respect to $\mathcal{G}_1$ and is compared to $\overline{\mathcal{G}}_m$).

Structural properties of the syntenic distance. We now recall some properties that will be used in the proofs of the characterization of maximal syntenic instances. We refer the reader to [5, 11] for the proofs of these properties.

Proposition 1. (Duality). *For every syntenic instance $\mathcal{S}$, $\mathcal{D}(\mathcal{S}) = \mathcal{D}(\mathrm{Dual}(\mathcal{S}))$.*

Proposition 2. (Canonicity). *For every syntenic instance, if there is an optimal mutation sequence solving it with k_1 fusions, k_2 translocations and k_3 fissions, then there is also an optimal mutation sequence solving it and* starting *by k_1 fusions,* followed *by k_2 translocations and* finishing *by k_3 fissions.*

Let $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_m\}$ and $\mathcal{S}' = \{\mathcal{S}'_1, \dots, \mathcal{S}'_m\}$ be two syntenic instances, and π be a permutation on $[n]$. One denotes by $\pi.\mathcal{S}'$ the syntenic instance obtained when π acts on the elements of $\mathcal{S}'$ (every element $j \in [n]$ is replaced by its image $\pi(j)$ in π). One says that $\mathcal{S}$ *dominates* $\mathcal{S}'$ if there is a permutation π on $[n]$ such that for all $i \in [m]$ we have $(\pi.\mathcal{S}')_i \subseteq \mathcal{S}_i$.

Proposition 3. (Monotonicity). *Let $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_m\}$ and $\mathcal{S}' = \{\mathcal{S}'_1, \dots, \mathcal{S}'_m\}$ be two syntenic instances. If $\mathcal{S}$ dominates $\mathcal{S}'$, then $\mathcal{D}(\mathcal{S}) \geq \mathcal{D}(\mathcal{S}')$.*

The syntenic diameter. The diameter of the syntenic distance was studied by Kleinberg and Liben-Nowell [9], and Pisanti and Sagot [13]. Let us denote by $\mathcal{SD}(m, n)$ the maximal syntenic distance over all the syntenic instances composed of m subsets of $[n]$. In [9, 13], the diameter for n-square syntenic instances ($m = n$) was computed and shown to be equal to $2n - 4$. The following theorem, proved in [13], generalizes the above result to the bidimensional case.

Theorem 1. (Bidimensional syntenic diameter). *For all $n, m \geq 4$, $\mathcal{SD}(n, m) = n + m - 4$.*

From now on, a syntenic instance $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_m\}$ on $[n]$ will be called an *(m, n)-maximal syntenic instance* if it cannot be solved in less than $n + m - 4$ mutations. If $n = m$, a maximal syntenic instances will be called a *maximal n-square syntenic instances.*

[1] It should always be kept in mind that a genome is an unordered collection of chromosomes and that the indices of the sets $\mathcal{S}_i$ do not represent an order on these subsets: two collections of subsets of $[n]$, that differ only by a permutation of the subsets they contain, represent the same instance.

Gossiping and synteny. The main tool that we will use to characterize maximal syntenic instances is a relationship between the transformation of a genome into another by translocations and a problem of gossiping, introduced by Kleinberg and Liben-Nowell in [9] and developed by Liben-Nowell in [12].

More precisely, the *gossiping* problem is used to model information dissemination in communication networks. It has been introduced in the 50s, and has received considerable attention since, leading to a wide literature. For more information on the subject, we refer to the survey [8]. The gossiping problem is defined as follows: we start with a set of n people, denoted by integers $1, \ldots, n$, each knowing a single piece of information, denoted by $\{i\}$ for the person labeled by i. Those people communicate between them, using the *telephone model*, that is a communication takes place between two people only, and both exchange all the information they have. One of the first questions that arose from this model is to determine the minimum number of calls to be made in order that everyone knows everything, that is $[n]$.

Kleinberg and Liben-Nowell pointed out an equivalence between translocations and a variant of gossiping, that they called *incomplete* gossiping. In this problem, people do not necessarily want to know all the pieces of information, but each person wants to know a *specific subset* of the total information (that is a subset of $[n]$). Hence, in particular, during a call a person is allowed to give to the other caller only an arbitrary subset of his/her information. Let us denote by $\mathcal{S}_i$ the information that the person i wants to know. In this model, for a given information configuration $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_n\}$ (where $\mathcal{S}_i$ represents the information the person i wants to know), one says that it can be *disseminated in k calls* if there is a sequence of k calls that leads to the situation where, for $i = 1, \ldots, n$, the person i knows at least the information $\mathcal{S}_i$ (if every person i knows exactly $\mathcal{S}_i$, one says that $\mathcal{S}$ is *exactly disseminated*).

The relationship with translocations in the syntenic distance problem is as follows (see [12, Section 4] for more details): if I_j and I_k are the sets of information that two people j and k know before calling each other, after this call they respectively know I'_j and I'_k where $I_j \subseteq I'_j$, $I_k \subseteq I'_k$ and $I_j \cup I_k = I'_j \cup I'_k$; which leads to the remark that one could evolve from sets I'_j and I'_k to the sets I_j and I_k by a translocation. The following theorem follows from this relationship (see [12, Section 4]).

Theorem 2. *Let $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_n\}$ be an n-square syntenic instance. $\mathcal{D}(\mathcal{S})$ is the minimum, among all permutations π_S of $\mathcal{S}$, of the minimum numbers of calls necessary to disseminate π_S.*

Notation. When during a call two people, say i and j, exchange totally their available information, we say that this call is *complete* and we denote it by (i, j).

Reduction of a syntenic instance. Finally, before starting the description of our characterization of maximal syntenic instances, let us introduce the notion of reduction of a syntenic instance, that will be used intensively in our proofs. Let X be a subset of $[n]$ ($X = \{x_1, \ldots, x_k\}$, with $x_1 < \ldots < x_k$) and $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_m\}$ be a syntenic instance. We call the *X-reduction of $\mathcal{S}$* the syntenic instance $\mathcal{S}^X = \{\mathcal{S}^X_1, \ldots, \mathcal{S}^X_m\}$ obtained from $\mathcal{S}$ as follows: in each $\mathcal{S}_j$, with $1 \leq j \leq m$, every element $x_i \in X$ is replaced by a single occurrence of x_1 and the resulting collection $\mathcal{S}' = \{\mathcal{S}'_1, \ldots, \mathcal{S}'_m\}$ of m subsets of $[n]$ are normalized on $[n-k+1]$ to give $\mathcal{S}^X$. Hence $\mathcal{S}^X$ is a syntenic instance on $[n-k+1]$. For example, let $\mathcal{S} = \{\{1,2,3,4,5\}, \{1,2,3,4\}, \{1,2,5\}, \{1,2\}, \{1\}\}$ and $X = \{1,2,4\}$. Then $\mathcal{S}' = \{\{1,3,5\}, \{1,3\}, \{1,5\}, \{1\}, \{1\}\}$ and $\mathcal{S}^X = \{\{1,2,3\}, \{1,2\}, \{1,3\}, \{1\}, \{1\}\}$ is a syntenic instance with $n = 3$.

3 The case of maximal square syntenic instances

We recall that an n-square syntenic instance is composed of n subsets of $[n]$. In this section, we first describe some conditions that an n-square syntenic instance should satisfy in order to be a maximal n-square syntenic instance. We then show that this conditions are sufficient to be a maximal n-square syntenic instance.

3.1 Necessary conditions

The conditions introduced here formalize the quite intuitive idea that a maximal syntenic instance does not involve small subsets of $[n]$.

Lemma 1. *Let $n \geq 5$ and $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_n\}$ be an n-square syntenic instance. If there exists $i \in [n]$ such that $|\mathcal{S}_i| \leq n-3$ then $\mathcal{S}$ is not a maximal n-square syntenic instance.*

Proof. We consider the gossiping problem for $\mathcal{S}$ and we show that $\mathcal{S}$ can be disseminated in $2n-5$ complete calls, thus proving by Theorem 2 that $\mathcal{D}(\mathcal{S}) \leq 2n-5$. $\mathcal{S}$ is dominated by the syntenic instance $\mathcal{S}' = \{\mathcal{S}'_1, \ldots, \mathcal{S}'_n\}$, defined by $\mathcal{S}'_i = [n]$ for all $i \in [n]\backslash\{1\}$, and $\mathcal{S}'_1 = [n]\backslash\{n-2, n-1, n\}$. We will show that it is possible to disseminate $\mathcal{S}'$ in a sequence of $2n-5$ complete calls. Before displaying this sequence, we introduce the following notation : for any p, q, (p,q) means that p calls q. The sequence of $2n-5$ calls that disseminates $\mathcal{S}'$ is then as follows: first, for i from $n-4$ downto 1, $(i, n-3)$; then $(n-2, n-3)$, $(n-1, n)$, $(n-2, n-1)$ $(n-3, n)$ (after this, n calls have been made); and finally, for j from 2 to $n-4$, (j, n). At the end of the process, every person i, for $i \in [n]\backslash\{1\}$, knows $[n]$, while the person 1 knows $[n-3]$. Since this process uses $2n-5$ calls, we conclude by Proposition 3 and Theorem 2 that $\mathcal{D}(\mathcal{S}) \leq \mathcal{D}(\mathcal{S}') \leq 2n-5$, and thus $\mathcal{S}$ is not maximal. □

Hence, we can now restrict our study to syntenic instances where every set $\mathcal{S}_i$ is of size at least $n-2$. The next three lemmas give necessary conditions on these syntenic instances that will be proved to be sufficient in the following section.

Lemma 2. *Let $n \geq 4$ and $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_n\}$ be an n-square syntenic instance. If there exists $i, j \in [n]$ and $x, y \in [n]$ ($i \neq j$ and $x \neq y$) such that $x \notin \mathcal{S}_i$, $y \notin \mathcal{S}_i$ and $x \notin \mathcal{S}_j$, then $\mathcal{S}$ is not a maximal n-square syntenic instance.*

Proof. Here again we consider the gossiping problem associated to this syntenic instance, and show that $\mathcal{S}$ can be disseminated in $2n-5$ calls. $\mathcal{S}$ is dominated by the square syntenic instance $\mathcal{S}'$ defined by $\mathcal{S}'_i = [n]$ for all $i \in [n]\backslash\{n-3, n-2\}$, $\mathcal{S}'_{n-3} = [n-2]$ and $\mathcal{S}'_{n-2} = [n-1]$. $\mathcal{S}'$ can be disseminated by the following sequence of $2n-5$ complete calls: for i from 1 to $n-1$, $(i, i+1)$ (after this, n and $n-1$ know $[n]$, $n-2$ knows $[n-1]$ and $n-3$ knows $[n-2]$); then for j from 1 to $n-4$, (j, n) (after this, for any $i \in \{1, 2, \ldots, n-4, n-1, n\}$, i knows $[n]$; $n-2$ knows $[n-1]$ and $n-3$ knows $[n-2]$). Thus, $\mathcal{D}(\mathcal{S}) \leq \mathcal{D}(\mathcal{S}') \leq 2n-5$ by Proposition 3 and Theorem 2, and $\mathcal{S}$ is not a maximal n-square syntenic instance. □

Lemma 3. *Let $n \geq 5$ and $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_n\}$ be an n-square syntenic instance. If there exists three distinct integers $i, j, k \in [n]$ and $x \in [n]$ such that $|\mathcal{S}_i| \leq n-2$, $x \notin \mathcal{S}_j$ and $x \notin \mathcal{S}_k$, then $\mathcal{S}$ is not a maximal n-square syntenic instance.*

Proof. Due to Lemmas 1 and 2 above, we just need to focus on syntenic instances dominated by instances $\mathcal{S}'(n)$ such that $\mathcal{S}'_i = [n]$ for all $1 \leq i \leq n-3$, $\mathcal{S}'_{n-2} = [n-2]$ and $\mathcal{S}'_{n-1} = \mathcal{S}'_n = [n]\backslash\{n-2\}$. We now distinguish two cases: $n = 5$ and $n \geq 6$.

When $n = 5$, it is possible to disseminate $\mathcal{S}'(5)$ with the following five complete calls: $(4,5)$, $(1,2)$, $(2,5)$, $(1,3)$, $(3,4)$. Thus, in 5 translocations, it is possible to solve $\mathcal{S}'(5)$. Thus it is not a maximal 5-square syntenic instance.

Now suppose $n \geq 6$. In that case, we first perform on $\mathcal{S}'(n)$ $n-5$ fusions over $n-4$ copies of $[n]$. Let us denote by $\mathcal{S}''$ the resulting syntenic instance, that is $\mathcal{S}'' = \{[n], [n], [n-2], [n]\backslash\{n-2\}, [n]\backslash\{n-2\}\}$. As $\mathcal{S}'(5)$ is nothing else that the X-reduction of $\mathcal{S}''$ for $X = [n-4]$, we can deduce from the sequence of translocations described above to solve $\mathcal{S}'(5)$ a sequence of 5 translocations that can be applied to $\mathcal{S}''$. Hence, after these n mutations ($n-5$ fusions and 5 translocations) one has a syntenic instance $\{[n-4], \{n-3\}, \{n-2\}, \{n-1\}, \{n\}\}$, that can be transformed into $\overline{\mathcal{G}}_n$ by $n-5$ fissions on the subset $\{1, 2, \ldots, n-4\}$ producing the $n-4$ subsets $\{1\}, \{2\}, \ldots, \{n-4\}$. Altogether, we used $2n-5$ mutations, and thus $\mathcal{S}'$ is not maximal, which implies, by monotonicity (Proposition 3) that $\mathcal{S}$ is not a maximal n-square syntenic instance. □

Lemma 4. *Let $n \geq 6$ and $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_n\}$ be an n-square syntenic instance. If there exists $i, j \in [n]$ $(i \neq j)$ such that $|\mathcal{S}_i| \leq n-2$ and $|\mathcal{S}_j| \leq n-2$, then $\mathcal{S}$ is not a maximal n-square syntenic instance.*

Proof. Thanks to Lemma 2, we know that if there exists an x such that $x \notin \mathcal{S}_i$ and $x \notin \mathcal{S}_j$, then $\mathcal{S}$ is not maximal. Now let us focus on the remaining cases, that correspond to syntenic instances dominated by $\mathcal{S}'(n)$, where $\mathcal{S}'(n)$ is such that $\mathcal{S}'_i = [n]$ for all $1 \leq i \leq n-2$, $\mathcal{S}'_{n-1} = [n-2]$ and $\mathcal{S}'_n = [n]\backslash\{1,2\}$. We follow the same method as in the proof of Lemma 3, and notice first that for $n = 6$, the sequence of 7 complete calls $(5,6), (3,4), (1,2), (4,5), (2,3), (2,6), (1,4)$ disseminates $\mathcal{S}'(6)$. Now, if $n \geq 7$, we first perform on $\mathcal{S}'(n)$ $n-6$ fusions over $n-5$ copies of $[n]$. Let us denote by $\mathcal{S}''$ the resulting syntenic instance, that is $\mathcal{S}'' = \{[n], [n], [n], [n], [n-2], [n]\backslash\{1,2\}\}$. As $\mathcal{S}'(6)$ is nothing else that the X-reduction of $\mathcal{S}''$ for $X = \{4, 5, \ldots, n-2\}$, we can deduce from the sequence of translocations described above to solve $\mathcal{S}'(6)$ a sequence of 7 translocations that can be applied to $\mathcal{S}''$. Hence, after these $n+1$ mutations ($n-6$ fusions and 7 translocations), one has a syntenic instance $\{\{1\}, \{2\}, \{3\}, \{4, 5, \ldots, n-2\}, \{n-1\}, \{n\}\}$, that can be transformed into $\overline{\mathcal{G}}_n$ by $n-6$ fissions on the subset $\{4, 5, \ldots, n-2\}$ producing the $n-5$ subsets $\{4\}, \{5\}, \ldots, \{n-2\}$. Altogether, we can solve $\mathcal{S}'(n)$ with $(n-6) + 7 + (n-6) = 2n-5$ mutations. Hence, by monotonicity, $\mathcal{S}$ is not a maximal n-square syntenic instance. $\square$

Altogether, the four previous lemmas (Lemmas 1 to 4) lead to the following proposition.

Proposition 4. *Let $n \geq 6$, $\mathcal{A}(n) = \{\mathcal{A}_1, \ldots, \mathcal{A}_n\}$ be the n-square syntenic instance defined by $\mathcal{A}_i = [n]\backslash\{i\}$ for $i = 1, \ldots, n$, and $\mathcal{B}(n) = \{\mathcal{B}_1, \ldots, \mathcal{B}_n\}$ be the n-square syntenic instance defined by $\mathcal{B}_i = [n]\backslash\{i\}$ for $i = 1, \ldots, n-2$, $\mathcal{B}_{n-1} = [n]\backslash\{n-1, n\}$ and $\mathcal{B}_n = [n]$. Then every maximal n-square syntenic instance dominates $\mathcal{A}(n)$, $\mathcal{B}(n)$ or $\mathrm{Dual}(\mathcal{B}(n))$.*

3.2 Sufficient conditions

In the previous section we proved that every maximal square syntenic instance dominates $\mathcal{A}(n)$, $\mathcal{B}(n)$ or $\mathrm{Dual}(\mathcal{B}(n))$. In this section, we prove that the first two are maximal n-square syntenic instances (Proposition 5). By duality (cf. Proposition 1), this proves that $\mathrm{Dual}(\mathcal{B}(n))$ is also a maximal n-square syntenic instance. In order to prove that $\mathcal{A}(n)$ and $\mathcal{B}(n)$ are maximal n-square syntenic instances, it follows immediately from Theorem 2 that we need to consider only the gossiping instances obtained as permutations of these syntenic instances. We first prove a general lemma on gossiping, that we apply later to $\mathcal{A}(n)$ and $\mathcal{B}(n)$.

Lemma 5. *Let $n \geq 2$ and $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_n\}$ be an instance of gossiping composed of n subsets of $[n]$. If there exists $i, j \in [n]$, with $i \neq j$, such that*

(1) $\mathcal{S}_i = [n]\backslash\{j\}$,
(2) for every $k \in [n]$, $k \neq i$ and $k \neq j$, then $j \in \mathcal{S}_k$,

then every sequence of calls C that disseminates $\mathcal{S}$ in such a way that

(3) i knows exactly $\mathcal{S}_i$ at the end of the calls of C,

can not be shorter than $2n-4$ calls.

Proof. From the hypothesis (2), the information $\{j\}$ has to be known by all other people than i, and C contains are at least $n-2$ calls disseminating j. Let $c_1, \ldots, c_\ell$ the subsequence of calls of C that do not imply j. It follows from the hypothesis (1) and (3) that this sequence of calls should disseminate exactly $[n]\backslash\{j\}$ to i, and so does not involve any call disseminating j. If one considers the graph whose vertices are the integers of $[n]\backslash\{j\}$ (that is all people other than j) and edges are given by the calls $c_1, \ldots, c_\ell$, this graph should be connected, and then have at least $n-2$ edges. Hence there are at least $n-2$ calls implying j and $n-2$ calls that do not disseminate j, which proves the lemma. $\square$

Lemma 6. *Let $n \geq 4$ and $\mathcal{S}$ be a permutation of $\mathcal{A}(n)$. Then $\mathcal{S}$ can not be disseminated in less than $2n-4$ calls.*

Proof. Let $\mathcal{S}$ be a permutation of the sets of $\mathcal{A}(n)$, and C a sequence of calls disseminating $\mathcal{S}$. If, at the end of the calls of C, there exists i and j such that i knows exactly $[n]\backslash\{j\}$, by definition of $\mathcal{A}(n)$, j should be known by any other person than i, and then the conditions (1), (2) and (3) of Lemma 5 are verified. Hence by Lemma 5 C contains at least $2n-4$ calls.

Otherwise, if there does not exist i and j such that i knows exactly $[n]\backslash\{j\}$, we can deduce from the definition of $\mathcal{A}(n)$ that after the sequence C of calls, everybody knows $[n]$, which requires $2n-4$ calls. $\square$

Lemma 7. *Let $n \geq 4$ and $\mathcal{S}$ be a permutation of $\mathcal{B}(n)$. Then $\mathcal{S}$ can not be disseminated in less than $2n-4$ calls.*

Proof. Let $\mathcal{S}$ be a permutation of the sets of $\mathcal{B}(n)$. Let C be a sequence of calls disseminating $\mathcal{S}$, and $\mathcal{S}' = \mathcal{S}'_1, \dots, \mathcal{S}'_n$ the subsets exactly known by all the people at the end of C.

Here again, if, at the end of C the conditions (1), (2) and (3) of Lemma 5 are verified for $\mathcal{S}'$, or $\mathcal{S}'_k = [n]$ for all $k \in [n]$, then C contains at least $2n-4$ calls.

Now we consider that, at the end of C, not everybody knows $[n]$ and the conditions (1), (2) and (3) of Lemma 5 are not verified together.

First, let suppose that a subset $\mathcal{S}'_k$ has size $n-1$, say $\mathcal{S}'_k = [n]\backslash\{j\}$ for some $j \in [n]$. In this case, the conditions (1) and (3) of Lemma 5 are verified. So the condition (2) should not be verified and there should be $\ell \in [n]$ such that $j \notin \mathcal{S}'_\ell$, that is j is not present in two subsets of $\mathcal{S}'$, which is in contradiction with the structure of $\mathcal{B}(n)$. So no element of $\mathcal{S}'$ has size $n-1$, and by definition of $\mathcal{B}(n)$, if not everybody knows $[n]$ at the end of C, there is $i \in [n]$ such that $|\mathcal{S}'_i| \leq n-2$ and for every $k \neq i$ $\mathcal{S}'_k = [n]$. Finally, if $|\mathcal{S}'_i| = n-2$, by definition of $\mathcal{B}(n)$ it implies that $\mathcal{S}'_i = [n]\backslash\{n-1, n\}$, with $i \neq n$ and $i \neq n-1$.

Then, following the proof of Lemma 5, one can claim that C should contain a subsequence C' of at least $n-3$ calls that do not involve n or $n-1$ (in order that i learns exactly all its information), and a subsequence C'' of at least $n-2$ calls disseminating n (resp. $n-1$) among all people but i and n (resp. $n-1$).

Let us consider the case where the first call involving either n or $n-1$, that is the first call of C'', is not the call $(n-1, n)$. After all the calls of C, the sum of the number of copies of n and $n-1$ should be $2n-2$ (because $n-1$ people want to know $[n]$), while there are 2 copies before the calls of C''. Moreover, each call of C'' increases this number of copies by at most two. Hence, if the first call of C'' does not involve both n and $n-1$, it increases the number of copies by one, and there are $2n-5$ copies to disseminate, which asks for at least $n-2$ additional calls, and then C'' contains at least $n-1$ calls, and then, with C', to $2n-4$ calls.

Finally, consider the case where the first call of C'' is $(n-1, n)$ and suppose that $|C''| = n-2$. All calls of C'' are such that $\{n, n-1\}$ is known by the two involved persons after the call while it was not the case before the call. Each call $c = (j_1, j_2)$ among these $n-2$ calls belongs to one of the three following subsequences of C'' :

C_1. neither j_1 nor j_2 knows $[n]$ after the call c ;
C_2. either j_1 or j_2, but not both, knows $[n]$ before the call c ;
C_3. neither j_1 nor j_2 knows $[n]$ before the call c, but both do after c.

We define a mapping from C_3 to C_1 as follows: every call of C_3, say $c = (j_1, j_2)$, where j_1 knows the information n and $n-1$ before the call (but not j_2 by definition of C'') is mapped to a call (j_1, j_3) of C_1 involving the person j_1 and such that j_3 does not know the information pieces n and $n-1$ before this call but does after the call. This mapping is well defined because if j_1 knows $\{n-1, n\}$ before a call, it learned it during another call from C_1 (it can not be a call from C_2 because j_1 does not know $[n]$ before the call c). It is an injective mapping due to the fact that if the call (j_1, j_2) belongs to C_3, it is the only one of C_3 involving one of these two people (by definition of C_3). Hence we have $|C_1| \geq |C_3|$, which implies that $|C_1| + |C_2| + |C_3| \geq |C_2| + 2|C_3|$. As, by hypothesis, $|C''| = |C_1| + |C_2| + |C_3| = n-2$, we then have

$$n - 2 \geq |C_2| + 2|C_3|. \tag{1}$$

However, each call of C_2 (resp. C_3) can increase the number of people who knows $[n]$ by exactly one (resp. two), and each call of C_1 does not change this number. So, as $[n]$ should be disseminated among $n-1$ people, we have $|C_2| + 2|C_3| = n-1$, which leads to a contradiction with (1), and then with the hypothesis $|C''| = n-2$. Hence $|C''| \geq n-1$, and then $\mathcal{S}$ can not be disseminated in less than $2n-4$ calls. □

Proposition 5. *If $n \geq 6$, then $\mathcal{A}(n)$ and $\mathcal{B}(n)$ are maximal n-square syntenic instances.*

Proof. This is an immediate consequence of Lemma 6, Lemma 7 and Theorem 2. □

4 Characterization of maximal syntenic instances

We can now state our main characterization results and their algorithmic consequences. Our first result is a characterization of maximal n-square syntenic instances for $n \geq 6$.

Theorem 3. *Let $n \geq 6$, $\mathcal{A}(n) = \{\mathcal{A}_1, \ldots, \mathcal{A}_n\}$ be the n-square syntenic instance defined by $\mathcal{A}_i = [n] \backslash \{i\}$ for $i = 1, \ldots, n$, and $\mathcal{B}(n) = \{\mathcal{B}_1, \ldots, \mathcal{B}_n\}$ be the n-square syntenic instance defined by $\mathcal{B}_i = [n] \backslash \{i\}$ for $i = 1, \ldots, n-2$, $\mathcal{B}_{n-1} = [n] \backslash \{n-1, n\}$ and $\mathcal{B}_n = [n]$. An n-square syntenic instance is a maximal n-square syntenic instance if and only if it dominates at least one of instances $\mathcal{A}(n)$, $\mathcal{B}(n)$ or* $\mathrm{Dual}(\mathcal{B}(n))$.

Proof. This result follows immediately from Propositions 1, 4 and 5 and the fact that $\mathrm{Dual}(\mathcal{A}(n)) = \mathcal{A}(n)$. □

We can now extend Theorem 3 to the case of general syntenic instances, that is we consider syntenic instances composed of m subsets of $[n]$, with $n \neq m$.

Remark 2. We can limit our study to the case $n > m$, and the case $m > n$ follows by duality (Proposition 1).

The next result gives a characterization of (m, n)-maximal syntenic instances in terms of reduced syntenic instances.

Lemma 8. *A syntenic instance $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_m\}$ on $[n]$, where $n > m \geq 4$, is an (m, n)-maximal syntenic instance if and only if for every subset $\{a, b\}$ of $[n]$, $\mathcal{S}^{\{a,b\}}$ is an $(m, n-1)$-maximal syntenic instance.*

Proof. Suppose that $\mathcal{S}$ is not an (m, n)-maximal syntenic instance: there is a sequence of strictly less than $n+m-4$ mutations solving it such that the last mutation is a fission that splits a subset $\{a, b\}$ of $[n]$ (by canonicity and the fact that $n > m$). Hence there is a subset $\{a, b\}$ of $[n]$ such that $\mathcal{S}^{\{a,b\}}$ is not an $(m, n-1)$-maximal syntenic instance. Now, suppose that there exists $\{a, b\} \subset [n]$ such that $\mathcal{S}^{\{a,b\}}$ is not an $(m, n-1)$-maximal syntenic instance. This syntenic instance can be solved by a sequence of strictly less than $n+m-5$ mutations. If we complete this sequence by a fission of $\{a, b\}$ we solve $\mathcal{S}$ in strictly less than $n+m-4$ mutations and $\mathcal{S}$ is not an (m, n)-maximal syntenic instance. □

We now generalize the notion of reduction of a syntenic instance. Let $\mathcal{S}$ be a syntenic instance composed of m subsets of $[n]$ ($n > m$), $P = p_1, \ldots, p_m$ be a set partition of $[n]$ into m non-empty subsets (the p_i's) and $S^0, \ldots, S^m$ be the unique sequence of syntenic instances defined by $S^0 = \mathcal{S}$ and $S^{i+1} = (S^i)^{p_i}$ (the p_i-reduction of S^i) for $i = 0, \ldots, m-1$. The P-reduction of $\mathcal{S}$, denoted by $\mathcal{S}^P$, is defined as the m-square syntenic instance $\mathcal{S}^P = S^m$. The following lemma is a a generalization of Lemma 8 to the case of P-reductions of syntenic instances.

Lemma 9. *A syntenic instance $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_m\}$ on $[n]$, where $n > m \geq 4$, is an (m, n)-maximal syntenic instance if and only if for every set partition $P = p_1, \ldots, p_m$ of $[n]$ into m non-empty subsets, $\mathcal{S}^P$ is an m-square maximal syntenic instance.*

Proof. If, for a set partition $P = p_1, \ldots, p_m$ of $[n]$, $\mathcal{S}^P$ can be solved with less than $2m - 4$ mutations, then by adding to these mutations a sequence of $n - m$ fissions one can define a sequence of less than $n + m - 4$ mutations solving $\mathcal{S}$.

So we just have to concentrate on the case where for every set partition P, $\mathcal{S}^P$ is an m-square maximal syntenic instance, and prove that in this case $\mathcal{S}$ is maximal. If $n - m = 1$, it follows immediately from Lemma 8. Otherwise, suppose that $n - m > 1$ and that $\mathcal{S}$ is not an (m, n)-maximal syntenic instance. From Lemma 8 one can say that there is $\{a, b\} \subset [n]$ (with $a < b$) such that $\mathcal{S}^{\{a,b\}}$ (denoted by $\mathcal{T}$) is not an $(m, n-1)$-maximal syntenic instance. Hence by induction, we know that there is a set partition Q of $[n-1]$ into m non-empty sets such that $\mathcal{T}^Q$ is not an m-square maximal syntenic instance. We now define a set partition P of $[n]$ as follows: every element x of Q greater than or equal to b is replaced by $x + 1$ and b is added to the set containing a. It is immediate to verify that every sequence of k mutations solving $\mathcal{T}^Q$ induces a sequence of k mutations solving $\mathcal{S}^P$, which is not an m-square maximal syntenic instance. □

Thus, if we want a characterization of (m, n)-maximal syntenic instances that will be easier to translate in a decision algorithm, one just has to find which syntenic instances cannot be reduced to a non m-square maximal syntenic instance. Such syntenic instances are described in the following result.

Theorem 4. *Let $n > m \geq 6$, $\mathcal{C}(m, n) = \{\mathcal{C}_1, \ldots, \mathcal{C}_m\}$ and $\mathcal{D}(m, n) = \{\mathcal{D}_1, \ldots, \mathcal{D}_m\}$ be the syntenic instances on $[n]$ defined by: $\mathcal{C}_i = [n]\backslash\{i\}$ for $i = 1, \ldots, m-1$ and $\mathcal{C}_m = [n]\backslash\{m-1\}$, and $\mathcal{D}_i = [n]\backslash\{i\}$ for $i = 1, \ldots, m-1$ and $\mathcal{D}_m = [n]\backslash\{m, m+1\}$. A syntenic instance $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_m\}$ on $[n]$ is an (m, n)-maximal syntenic instance if and only if it dominates at least one of the two syntenic instances $\mathcal{C}(m, n)$ and $\mathcal{D}(m, n)$.*

Proof. Let us consider a syntenic instance $\mathcal{S}$ that does not dominate one of the two syntenic instances $\mathcal{C}(m, n)$ and $\mathcal{D}(m, n)$.

If $\mathcal{S}$ contains a set $\mathcal{S}_i$ such that $|\mathcal{S}_i| \leq n - 3$, let $x, y, z \in [n]$ that do not appear in $\mathcal{S}_i$ and $P = p_1, \ldots, p_m$ a partition of $[n]$ with $p_1 = \{x\}$, $p_2 = \{y\}$ and $p_3 = \{z\}$. Then it follows from Lemma 1 that $\mathcal{S}^P$ is not an m-square maximal syntenic instance, and Lemma 9 implies that $\mathcal{S}$ is not an (m, n)-maximal syntenic instance. Otherwise, let $\mathcal{S}$ be such that there exists a set $\mathcal{S}_i$ with $|\mathcal{S}_i| \leq n-2$. From the definition of $\mathcal{C}(m, n)$ and $\mathcal{D}(m, n)$, there are three possible cases where such a syntenic instance can dominate $\mathcal{C}(m, n)$ or $\mathcal{D}(m, n)$.

- Suppose that there are $j \in [n]$ and $x, y \in [n]$ ($j \neq i$ and $x \neq y$) such that $x \notin \mathcal{S}_i$, $y \notin \mathcal{S}_i$ and $x \notin \mathcal{S}_j$. Let P be a set partition of $[n]$ with $p_1 = \{x\}$ and $p_2 = \{y\}$. Then it follows from Lemma 2 that $\mathcal{S}^P$ is not an m-square maximal syntenic instance.
- Suppose there exists integers $j, k \in [n]$ distinct from i, and distinct integers $x, y, z \in [n]$ such that $x \notin \mathcal{S}_j$, $x \notin \mathcal{S}_k$, $y \notin \mathcal{S}_i$ and $z \notin \mathcal{S}_i$. Let P be a set partition of $[n]$ with $p_1 = \{x\}$, $p_2 = \{y\}$ and $p_3 = \{z\}$. Then it follows from Lemma 3 that $\mathcal{S}^P$ is not an m-square maximal syntenic instance.
- Suppose there exists $j \in [n]$ ($j \neq i$) and distinct integers $x, y, z, t \in [n]$ such that $|\mathcal{S}_i| \leq n-2$, $|\mathcal{S}_j| \leq n-2$, $x \notin \mathcal{S}_i$, $y \notin \mathcal{S}_i$, $z \notin \mathcal{S}_j$ and $t \notin \mathcal{S}_j$. Let P be a set partition of $[n]$ with $p_1 = \{x\}$, $p_2 = \{y\}$, $p_3 = \{z\}$ and $p_4 = \{t\}$. Then it follows from Lemma 4 that $\mathcal{S}^P$ is not an m-square maximal syntenic instance.

Hence we just need to verify that every syntenic instance $\mathcal{S}$ that dominates $\mathcal{C}(m, n)$ and $\mathcal{D}(m, n)$ is an (m, n)-maximal syntenic instance, (that is, by Lemma 9, can not be reduced to an m-square syntenic instance that dominates $\mathcal{A}(n)$ or $\mathcal{B}(n)$). This point can be shown by induction on $n - m$. If $n - m = 1$, for every subset $\{a, b\}$ of $[n]$, it is immediate to verify that $\mathcal{S}^{\{a,b\}}$ dominates one of the syntenic instances $\mathcal{A}(n)$ or $\mathcal{B}(n)$ defined in Theorem 3. If $n - m > 1$, suppose that $\mathcal{S}$ is not an (m, n)-maximal syntenic instance. By Lemma 8, there exists a subset $\{a, b\}$ of $[n]$ such that $\mathcal{S}^{\{a,b\}}$ is not an $(m, n-1)$-maximal syntenic instance. By induction, it implies that $\mathcal{S}^{\{a,b\}}$ does not dominate one of the syntenic instances $\mathcal{C}(m, n-1)$ and $\mathcal{D}(m, n-1)$. This leads to a contradiction with the fact that $\mathcal{S}^{\{a,b\}}$ is a reduction of a syntenic instance that dominates one the two syntenic instances $\mathcal{C}(m, n)$ and $\mathcal{D}(m, n)$. □

Remark 3. Theorems 3 and 4 give a characterization of maximal syntenic instances $(\mathcal{G}_1, \mathcal{G}_2)$ with $n \geq 6$ (where n is the size of $\mathcal{G}_2$). However, the syntenic diameter is known to be equal to $2n-4$ for any $n \geq 4$, thus the cases $n = 4$ and $n = 5$ also need to be considered. This work has been done using a computer program based on Liben-Nowell's algorithm [12], and is described in [4].

Remark 4. Our two main results (Theorems 3 and 4) relate the maximality of a syntenic instance in terms of domination of this syntenic instance over a small number of simple syntenic instances. Moreover, testing that a given syntenic instance dominates either $\mathcal{C}(m,n)$ or $\mathcal{D}(m,n)$ is easy: for $\mathcal{C}(m,n)$ for syntenic instance, it is sufficient to test that each set of the syntenic instance is of size at least $n-1$ and that there are at most one $i \in [n]$ that does not belong to two different sets of the syntenic instance. This can clearly be achieved in polynomial time. The same holds for $\mathcal{D}(m,n)$, which implies immediately that for $n, m \geq 6$ and $\mathcal{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_m\}$ a syntenic instance, one can decide in time polynomial in $n+m$ whether $\mathcal{S}$ is a (m,n)-maximal syntenic instance.

5 Conclusion

In this paper, we answered two open questions from Kleinberg and Liben-Nowell [9] about the syntenic distance - we described a characterization of square maximal syntenic instances that induces a polynomial time algorithm deciding whether a square syntenic instance is maximal - and we extended our results to the bidimensional case.

An interesting point is the usefulness of the relationship between translocations and calls in the gossiping problem. This fact is central in the proofs of our results.

Thanks to our study, it is also possible to confirm the fact that maximal square syntenic instances can be solved by translocations only (that is, fusions and fissions do not necessarily help for syntenic instances that are at distance equal to the diameter). This extends a similar result in the case of the n-square syntenic instance containing n copies of $[n]$ due to [9]. However, it has been pointed out in [12] that for some syntenic instances, fusions and fissions are necessary in order to get the minimum distance between two genomes. Hence, among several open problems, we would like to point out the following one: is it possible to characterize those syntenic instances that can be solved by translocations only ?

References

1. V. Bafna and P. Pevzner. Genome rearrangements and sorting by reversals. *SIAM Journal on Computing*, 25(2):272-289, 1996.
2. V. Bafna and P. Pevzner. Sorting by transpositions. *SIAM Journal on Discrete Mathematics*, 11(2):224-240, 1998.
3. A. Bergeron. A very elementary presentation of the Hannenhalli-Pevzner theory. In *Combinatorial Pattern Matching (CPM'01)*, volume 2089 of *Lecture Notes in Computer Science*, pages 106–117. Springer, 2001.
4. C. Chauve and G. Fertin On maximal instances for the original syntenic distance. Technical Report 03.03, IRIN, Nantes, 2003.
5. B. DasGupta, T. Jiang, S. Kannan, M. Li and E. Sweedyk. On the complexity and approximation of syntenic distance. *Discrete Applied Mathematics*, 88(1-3):59–82, 1998.
6. V. Ferretti, J. H. Nadeau and D. Sankoff. Original synteny. In *Combinatorial Pattern Matching (CPM'96)*, volume 1075 of *Lecture Notes in Comput. Sci.*, pages 159–167. Springer, 1996.
7. S. Hannenhalli and P. Pevzner. Transforming cabbage into turnip: polynomial algorithm for sorting signed permutations by reversals. *Journal of the ACM*, 46(1):1-27, 1999.
8. S. Hedetniemi, S. Hedetniemi and A. Liestman. A survey of gossiping and broadcasting in communication networks. *Networks*, 18:319-349, 1988.
9. J. Kleinberg and D. Liben-Nowell. The syntenic diameter of the space of N-chromosome genomes. In *Conference on Gene Order Dynamics, Comparative Maps, and Multigene Families (DCAF)*, pages 185–197. Kluwer Academic Press, 2000.
10. D. Liben-Nowell and J. Kleinberg. Structural properties and tractability results for linear synteny. In *Combinatorial Pattern Matching (CPM'00)*, volume 1848 of *Lecture Notes in Computer Science*, pages 248–263. Springer, 2000.

11. D. Liben-Nowell. On the structure of syntenic distance. *Journal of Computational Biology*, 8(1):53–67, 2001.
12. D. Liben-Nowell. Gossip is synteny: incomplete gossip and syntenic distance between genomes. *Journal of Algorithms*, 43(2):264–283, 2002.
13. N. Pisanti and M. F. Sagot. Further thoughts on the syntenic distance between genomes. *Algorithmica*, 34(2):157–180, 2002.