Survey On Bioinformatics And Computational Biology

Pranav Kumar[1]

[1]*Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi-835215, India*

G. Sahoo[2]

[2]*Department of Information Technology, Birla Institute of Technology, Mesra, Ranchi-835215, India*

Abstract

The abstract is to be in fully-justified italicized text, at the top of the left-hand column as it is here, below the author information. Use the word "Abstract" as the title, in 12-point Times, boldface type, centered relative to the column, initially capitalized. The abstract is to be in 10-point, single-spaced type, and up to 150 words in length. Leave two blank lines after the abstract, then begin the main text.

1. Introduction

A Bioinformatics and computational biology are rooted in life sciences as well as computer and information sciences and technologies. Both of these interdisciplinary approaches draw from specific disciplines such as mathematics, physics, statistics, computer science and engineering, biology, and behavioral science. Bioinformatics and computational biology maintain close interactions with life sciences to realize their full potential. Bioinformatics involves information theory and data management applying principles of information sciences and technologies to make the vast, diverse, and complex life sciences data more understandable and useful. Computational biology on the other hand, uses mathematical and computational approaches to directly address biological hypotheses, both theoretical and experimental in nature. Although bioinformatics and computational biology are distinct .There is also significant overlap and activity at their interface and they are consequently often confused by those outside of the discipline. The bioinformaticians build software while computational biologists use and tweak the software to tackle their specific biological questions. Nicholas M. Luscombe, Dov Greenbaum and Mark Gerstein proposed the following definition of bioinformatics:

"Bioinformatics is conceptualizing biology in terms of molecules and applying information techniques (derived from disciplines such as applied maths, computer science and statistics) to understand and organize the information associated with these molecules on a large scale". The term "Bioinformatics" seems to have been first used in mid 1980s in order to describe the application of information science and technology in life sciences the supporting information management in field like heath care. The closely related field of computational biology also provides support for computation. The computation biology generally is concerned with development of novel and efficient algorithms that can be proven to work on a difficult problem such as multiple sequence alignment or genome fragment assembly. Bioinformatics is now involved in these activities by organizing biological data related to genomes with a view to applying this information in agriculture, pharmacology and commercial application. The computers can be used to retrieve, process, analyse and simulate biologically information.

2. Problem domain in Bioinformatics

A fundamental challenge in bioinformatics is transforming a biological problem into a computational problem. The computational problem must on one hand be abstract enough to be solved by algorithms and on the other hand account for practical constraints of biological problem to be applicable in practice.

3. Inspiration

Bioinformatics is a scientific discipline that applies computer science and information technology to help understand biological processes. It is the science of how information is generated, transmitted, received, and clarified in biological system. The handing and analysis of entire genome sequence of organism is one of the prime tasks of the bioinformatics. With new techniques for sequencing the entire genome of organisms, biology is rapidly moving towards a data intensive computational science. The great challenge we face now is how to process this huge amount of data and

extract from it relevant biological information where we analyze data coming from distinct species and learn from the similarities and differences in related genomes .This is motivated from an observation that closely related species mostly have homologous genes but the genes occur in different orders. The order of genes in the genomes of species can change during evolution and can provide information about their phylogenetic relationship.

4. Aim of Bioinformatics

The primary goal of bioinformatics is to increase the understanding of biological processes and uncover the wealth of biological information hidden in the mass of data and obtain a clearer insight into the fundamental biology of organisms. It strives to determine what information is biologically important and to decipher how it is used to control the chemical environment within the living organism. The new knowledge could have profound impacts on vibrant research fields as varied as human health agriculture, environment, energy and biotechnology.

5. Challenges in Bioinformatics and Computational biology

In bioinformatics the current research and computation are limited by the available of computer hardware. However this problem can be solved high performance computing resources. There are several reasons for the increased focus on high performance computing, larger data set increased computational requirements stemming from more sophisticated methodologies and latest developments in computer chip production. Bioinformatics demonstrates the growing data problem. It is considering that just a few years ago, experimentally generated data sets often fit on a single CD-ROM. Today data sets from high-throughput data sources as for examples next generation DNA sequencing no longer fit on a single DVD-ROM. In genome research, data sets appear to be growing at a rate that is faster than the corresponding increase in hardware performance. At the same time, methodological advances have led to computationally more demanding solutions. A common approach among these recent methods is the use of simulation and resampling techniques. Markov Chain Monte Carlo and Gibbs Sampling, Bootstrapping, and Monte Carlo Simulation are examples of increasingly popular methods that are important in various areas of statistical analysis investigating geographic, econometric, and biological data. For example in microarray data, building classification or prognostic rules from large microarray sets will be very time consuming. Here, preprocessing must become part of the cross-validation and resampling strategy which is necessary to estimate the rule's prediction quality

correctly. Both increased data size and increased simulation demands have been approached by researchers by means of parallel computing. Data sets can often be subdivided into chunks that can undergo analysis in parallel. This is particularly the case when the analysis of a given data 'chunk' does not depend on other chunks. Similarly, simulation runs that are not dependent on other simulations can be conducted in parallel. Hence both reasons (data size and simulation demands) for an increased interest in high performance and parallel computing can be seen as so called embarrassingly parallel problems that are suitable for execution in parallel.

6. Past history

Bafna and Pevzner[21] introduced the concept of breakpoint graph extends naturally to signed permutations and devised an approximation algorithm for signed permutation (π). It provides performance guaranteed error bound of 3/2 and performance ratio 7/4 for unsigned permutation. This algorithm has time complexity of $O(n^2)$. They introduced the notion of breakpoint graph of a permutation and revealed important links between the maximum cycle decomposition of graph and reversal distance. The approximation algorithm for signed permutation deals with the concept of Break Point Graph and the transformation from a signed permutation π of order n to an unsigned permutation π '. There are following procedure to change signed permutation in to unsigned permutation.

For signed permutation value of +i or −i to replace +i by 2i - 1,2i and -i by 2i, 2i-1. They observe that the identity signed permutation maps to the identity (unsigned) permutation. The effect of a reversal on π can be mimicked by a reversal on r'. They perform reversals only across breakpoints so that any reversal on the unsigned permutation can be mimicked by the signed permutation. They use the concept of cycle decomposition of G(r) and a reversal $\rho = [i, j]$, where ρ is a reversal on a cycle, if the breakpoints (ri-1, ri) and (rj, rj+l) belong to the same cycle.

A cycle is oriented if there exists a 1 or 2-reversal on it. Two cycles are crossing if some of the breakpoints corresponding to their black edges are interleaved in the permutation. They consider order of genes in two organisms is represented by permutations $\pi = (\pi 1, \pi 2 \ \pi 3, \ \dots.. \ \pi \ n)$ and $\alpha = (\alpha 1, \alpha 2 \ \alpha 3 \dots \alpha n)$. A reversal ρ of an interval [i, j] is the permutation $\rho = (\ 1 \ , 2 \ , .. ., i - 1 \ ,j \ ,j - 1,... , i+ 1 \ , i \ ,j + 1 \ , . . . ,n)$. They find $\pi \ o \ \rho$ has the effect of reversing genes $\pi i, \ \pi i+1, \ .,.,., \ \pi j$ .Where given permutations π and α the reversal distance problem is to find a series of reversals $\rho 1, \rho 2 \ \dots\dots. \ \rho t$ such that $\pi . \ \rho 1 . \ \pi . \ \rho 2 \ \dots\dots.. \ \rho \ t = \alpha$, t is minimum. t is the reversal distance between π and α. Sorting π by reversals is the problem of finding reversal distance d(π), between 2i-1and 2i.

Bader[2] invented the algorithms of transfer the one genome to other with minimum number of reversals distance d (π).The algorithm takes O (n^2) time complexity. Bader revealed linear algorithms to find out connected component which is more efficient than Hannenhali and pevzner proposed algorithm for breakpoint graph.

Hannenhalli and Pevzner[19] gave first polynomial time algorithm for a model of genome rearrangements. Based on Bafna and Pevzner bound(1) approximates the reversal distance extremely well for both simulated and biological data but Kececioglu and Sankoff gave an average difference between the bound (1) and the exact distance is less than 1 for random permutations. This intriguing performance raises a question whether the bound (1) overlooked a third parameter (in addition to the number of breakpoints and the size of maximum cycle decompositions) that would allow closing the gap between d (π) and b (π)-c (π). They proposed formula for finding reversal distance which is given by d (π) =b (π)-c (π) +h (π) based on this third hidden parameter h (π) (number of hurdles in π). They observe that every study of genome rearrangements involves solving a combinatorial "puzzle" to find a shortest series of reversals to transform one genome into another.

Siepel[17] stated the problem of estimating evolutionary distance from differences in gene order has been distilled to the problem of finding the reversal distance between two signed permutations. He invented an algorithm to finding all minimum sequence of reversals for one genome with respect to other. In these algorithms, there is no auxiliary information available in sorting by reversals of a signed permutation. Siepel introduced the fastest algorithm that generates the list of all sorting reversals for signed permutation (π) for next reversals. It is derived to solve the problem of finding all sorting reversals and experimental results are presented to indicate that while the new algorithm does not represent a significant improvement in asymptotic terms(it takes O (n^3)) time complexity for permutations of size n. The problem can now be solved by brute force in Θ (n^3) times. It performs dramatically better in practice than the best known alternative. He devised an algorithm to enumerate all sorting reversals with concept of Breakpoint graph B of permutation (π).

Setubal and Meidanis[13] have presented an alternative distinction oriented and unoriented gray edge based on black edges. They define two black edges of the same cycle are convergent if divergent sign of black edges are different. They have shown that a reversal splits a cycle if and only if it acts on divergent black edge of the cycle in breakpoint graph. The Breakpoint graph has all connected components, cycles and black edges within the cycles. M = {mi} to be the set of

components in overlap graph O, Ci = {ci, j} be the set of cycles that belong to mi, and bi, j= {bi, j, k} be the set of black edges that belong to ci, j.

Bergeron[3,6,8] gave a concept of a way to group all possible set of solution of singed permutation into an equivalence class (A). An equivalence class of optimal sequences of reversals over this equivalence relation is called a trace. Each class is identified by identity element known as representative element of class. However, no algorithm exists so far to compute this set of representatives except through the enumeration of all solutions, which takes too much time even for small permutations. Bergeron et al. [4] state as an open problem the design of such an algorithm. It has all solutions denoted by a representative element. There is no need to enumerate all the solution. Trace is a set of optimal solution composed by the same reversals but in different orders. The set of all optimal solutions is a Union of traces. Many optimal solutions are equivalent and same reversals. The set of all optimal sequences of reversals sorting π is a union of trace.

M.D.V[7] adopted methodology to grouping the optimal sequences of reversals into traces. They have developed an algorithm to enumerate all traces of a singed permutation (π).The elements of traces are arranged into lexicographic order. The algorithm presents all optimal sequence of solutions of a signed permutation, without enumerating all solutions. The main advantage of trace is that it reduces the solution space of sorting by reversals. Traces have enhanced the usage of the space (memory) for optimal sequence of reversals. However no algorithmic study was performed, and in particular the problem of giving one element in each class without enumerating all the solutions was mentioned open. They introduced a solution to this problem. Their algorithm gives one representative element per class of solutions, and counts the number of solutions in each class. The complete enumeration of the solutions is not needed, and the theoretical complexity, as well as the practical execution time is lower than in any other current method for the enumeration of the solutions.

They introduce a commutation rule to grouping the optimal sequence of reversals into equivalence class. The commutation rule is as follows: Identifying a sequence of reversals with a word on the alfaphabet A of reversals, the authors of [4] define an equivalence relation on these words: if α and β are reversals (intervals) and do not overlap, then the words $\alpha\beta$ and $\beta\alpha$ are equivalent. We say that α and β commute. Under this relation, any two words containing $\alpha\beta$ as a sub word are equivalent to the same word, replacing the sub word $\alpha\beta$ by $\beta\alpha$.

Braga[1] designed algorithms for enumeration of traces and its implementation based on Siepel and Bergeron's algorithm. After that Braga developed an algorithm to generate the normal form of traces. The traces are used to store the enumerated number of optimal sequence of solutions of a permutation. They provided a method to count the number of normal form of traces by algorithm. This paper explained algorithm to count the number of solution in each class with better practical and theoretical approach. This approach has been used in enumeration of optimal sequence of reversals. They also provided a way to make normal form of generated trace. Braga gave the concept of normalization of the traces into equivalent and lexicographical order (dictionary approach).she used a better sorted structure to handle the data processing and to store the equivalent traces.

7. Biological background

For understanding genome rearrangement problems, certain knowledge of biology and molecular genetics is necessary. Therefore, we will give a brief introduction to molecular genetics. This introduction only covers the subjects that are relevant for understanding genome rearrangement problems.

7.1 Assembly of Genome

7.1.1 Nucleic Acids

Deoxyribonucleic acid (DNA) is a molecular repository for genetic information and is responsible for encoding protein sequences in living organisms. A strand of DNA is a chemically linked chain of nucleotide bases that include the pyrimidines, thymine (T) and cytosine (C), and the purines, guanine (G) and adenine (A). The start of a nucleotide strand is denoted as the 5' terminus and the end of the strand is the 3' terminus. A pair of anti-parallel nucleotide strands forms the DNA double helix, as illustrated in Figure 2. Each base in the plus (+) or non template strand is linked via a hydrogen bond to a base in the minus or template strand to form a base pair. Each pair of linked bases complement one another such that t is always paired with a and c is always paired with g.

The size of a genome is measured in million base pairs (M.b.p). A segment of a chromosome is encoded into proteins. Such a segment is called a gene. Depending on which strand is encoding the protein. It can assign an orientation to each gene. On an even higher level of abstraction than before, a chromosome is a string of oriented genes. This orientation is only relative to the other genes on the same chromosome because the strand which encodes genes with positive orientation can be chosen arbitrarily. Thus a chromosome is

equivalent to its reverse complement which can be obtained by reading the chromosome backwards and inverting the orientation of every gene.
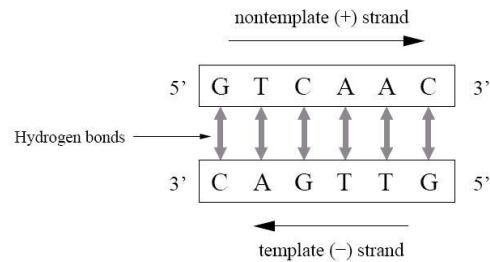
Figure 1: A pair of Nucleotide Strands forms DNA double helix structure.

7.1.2 DNA

A DNA or nucleotide sequence is represented by a string composed of the letters A, C, G, and T that are used to represent each of the four nucleotide bases. Nucleotide sequences range widely in length and can be up to several million bases long. It is shown in figure 1.

7.1.3 Genome Dynamics

The genome of an organism is stored in each of its cells. Genome can change due to replication errors and chemical or physical influences. Some of these changes destroy the functionality of the cell and let the cell die. Others change the behavior of the cell and lead to malfunctions of the cell. This happens for example in cancer cells. However most of the Changes have no influence at all and some can even result in benefits for the organism. These changes are inherited by child cells.

8. Genome Reordering

Reordering is a process to transform the genes order observed in one into that observed in the other. The reordering distance between single chromosome genomes can be estimated as minimum number of reversals required to transform one gene's ordering into other. Genome rearrangements are much less frequently than point Mutations, among vertebrates. It is often still possible to reconstruct these evolutionary events that happened between two highly diverged species to organisms with slightly modified genome information. One of the two strands of the DNA double helix will suffice to describe this information. This is due to complementary base pairing, where an A on one strand always binds to a T on the other and a C always binds to a G. The information stored in DNA allows the organization of molecules into functioning, living cells.
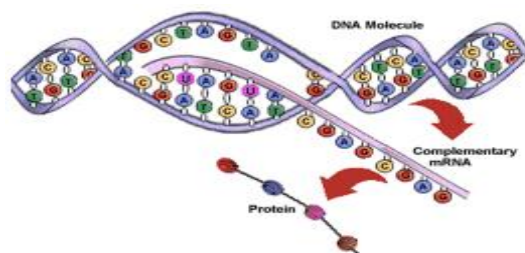
Figure 2: DNA molecule (Deoxyribonucleleic acid) is Genetic material.

8.1 Type of DNA sequence

DNA sequence can be divided into two parts
1. Comparative genomics.
2. Functional genomics.

8.1.1 Comparative Genomics

It deals the sequence from different organisms or even different individuals are compared in order to determine ancestries and co-relations' with dieses. There are different algorithms that sort the given DNA sequences (called permutation) and it is often used to propose evolutionary scenarios of the large scale genomic mutations between different organisms.

8.1.2 Functional Genomics

Functional genomics is a field of molecular biology that attempts to make use of the vast wealth of data produced by genomic projects (such as genome sequencing projects) to describe gene (protein) functions and interactions. It focuses on the dynamic aspects such as gene transcription, translation, and protein–protein interactions, as opposed to the static aspects of the genomic information such as DNA sequence or structures. It seeks to determine the roll of the sequence in living organism cell.

9. Genome model

The entire complement of genetic material carried by an individual is called the genome. Each genome contains one or more DNA molecules one per chromosome. Genomes can be modeled by permutation (π) .Each gene can be assigned a unique number and is exactly found once in the genome.

9.1 Genome

Genome contains all the biological information needed to build and maintain a living organism. It has the entire hereditary information of an organism. Genome is encoded in its deoxyribonucleic acid (DNA) molecule. This term genome was discovered in 1920 by Hans Winkler, professor of botany at the University of Hamburg, Germany. Genomes are partitioned into chromosomes. A chromosome can be linear (eukaryotes) or circular (prokaryotes).Comparison

of genome sequences through sorting techniques can reveal significant information about their ancestry and evolution. Assuming the genomes share the same set of genes with no duplications, a genome can be represented by a signed permutation. It takes two genome sequences (signed permutations) have almost identical gene sequences. By applying various rearrangement operations we try to transform one sequence into another. Changes in Genomic Sequences of different species (even of closely related individuals) differ from one another. These differences are caused by following:

1. Point mutation.
2. Genome rearrangement.

9.1.1 Point mutation

Point mutation is random mutation in deoxyribonucleic acid (DNA) that occurs at one point. This mutation may be insertion and deletions and substitution causes local changes during a genes evolution and only one nucleotide is changed in the gene sequence. There are following operations have applied into gene sequence. Point mutation usually takes splice during DNA replication. DNA replication occurs when one double-stranded DNA molecule creates two single strands of DNA that is a template for the creation of the coinciding strand.

Insertion	ATGGCG → ATGTGCG
Deletion	ATGTGCG → ATGGCG
Substitution	ATGTGCG → ATGCGCG

9.1.2 Genome rearrangement

Watterson et al. first proposed the minimum number of chromosomal reversals necessary to transform one ordering into other. The arrangement distance between single chromosome genomes can be estimated as the minimum number of reversals required to transform the gene ordering observed in one into that observed in the other. This measure, known as "reversals distance" .It can be computed as the reversal distance between signed permutations (π).Where multiple nucleotides are modified.

10. Genome rearrangement operation

The arrangement distance between single chromosome genomes can be estimated as the Minimum number of inversions required to transform the gene ordering observed in one into that observed in the other. This measure, known as "reversals distance" can be computed as the reversal distance between signed permutations. Genome rearrangement operations affect gene order and gene content. There are following example to illustrate the genome rearrangement.
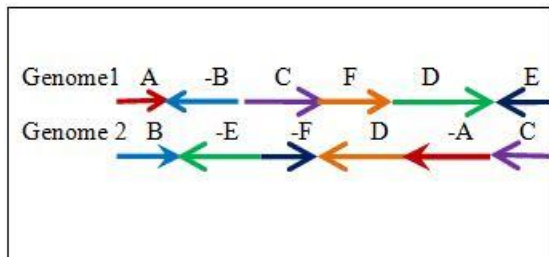
Figure 3: The Rearrangement of Genes of one Genome1 to Genome2.

10.1 Type of genome rearrangement operation

Genome rearrangement operation is classified for following chromosomal genome.

10.1.1 Single chromosomes genome (mitochondria, chloroplasts)

There are following Genome rearrangement operations:
1. Reversal.
2. Transpositions.
3. Reverse transpositions.
4. Gene Duplications.
5. Gene loss.

10.1.1.1 Reversal

In reversals operation, one region of a chromosome is flipped and reinserted. Reversals are the most common mutations acting on the order and orientation of genes in a genome. The minimum number of reversals is required to transform one gene ordering into another is called reversals distance. It is denoted by d (π).

Example: A signed permutation π = (-3, +2, +1,-4) and d (π) =4.

|	
(-3, +2, +1, -4)	r_1 = {1, 2, 4}
(-3, +4, -1, -2)	r_2 = {1, 3, 4}
(+1, -4, +3, -2)	r_3 = {2, 3, 4}
(+1, +2, -3, +4)	r_4 = {3}

In the above reversals operation a given signed permutation π = (-3, +2, +1,-4) sorted into permutation (π) = (+1, +2, +3, +4) after four reversals operations r1,r2,r3,r4.

10.1.1.2 Translocation

A region from one chromosome is aberrantly attached to another chromosome.

10.1.1.3 Transposition

A segment of genes moves from one place to another in the chromosome. A transposition is said to happen when two consecutive markers of a genome exchange their positions. It is always possible to produce the same result as a transposition with a sequence of three reversals. Thus a sequence of m transpositions can always be transformed in a sequence of 3m reversals.
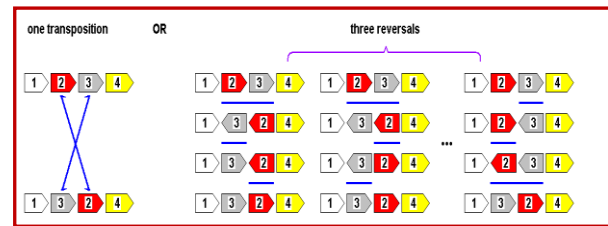
Figure 4: Transposition or a sequence of three Reversals may produce the same Rearrangement in a Genome and observe that the three Reversals can be applied in different Orders.

10.1.1.4 Deletion

A region of a chromosome is lost, resulting in the absence of all the genes in that area.

10.1.1.5 Duplication

A region of a chromosome is repeated, resulting in an increase in dosage from the genes in that area

10.1.2 Multiple-chromosomes genomes

There are following Genome rearrangement operations:

1. Translocation.
2. Fusion.
3. Fission.

10.1.2.1 Translocation

A chromosome translocation is a chromosome abnormality caused by rearrangement of parts between non homologous chromosomes. Reciprocal translocations are usually an exchange of material between non homologous chromosomes. Estimates of incidence range from about 1 in 500 to 1 in 625 human newborns.

10.1.2.2 Fusion and Fission

A gene fusion may be created when the translocation joins two otherwise separated genes.Two chromosomes are merging into one chromosome. The fission and fusion mechanisms are use in circular permutation. Two separate genes arise (potentially from the fission of a single gene). If the genes fuse together in different orders in two chromosomes a circular permutation occurs. There are following genome rearrangement operations.

11. Conclusion
At last we would like to conclude by saying that

12. Acknowledgment

13. References

[1] Braga M.D.V., Sagot M.-F., Scornavacca C. and Tannier E., "Exploring space of sorting by reversals with experiments and an application to evolution",Transactions on Computational Biology and Bioinformatics, volume 5, number 3, pp. 348–356, 2008 (A preliminary version appeared in ISBRA 2007, Lecture Notes in Bioinformatics vol. 4463, pp. 293–304).

[2] Bader D. A., Moret, B. M. E. and Yan, M.,"A linear-time algorithm for computing inversion distances between signed permutations with an experimental study", J. Computational Biology 8, 5 (2001), pp. 483–491.

[3] Bergeron A., Chauve C., Hartmann T. and St-Onge K., "On the properties of sequences of reversals that sort a signed permutation", JOBIM, pp. 99–108, 2002.

[4] Kaplan, H., Shamir, R., and Tarjan, R."A faster and simpler algorithm for sorting signed permutations by reversals". SIAM J. Computing 29, 3, pp. 880–892, first appeared in Proceedings, 8th Annual Symposium on Discrete Algorithms, New Orleans, LA. ACM Press, New York, pp. 344–351.

[5] Braga, M.D.V., Sagot, M., Scornavacca, C., Tannier, E.:"The solution space of sorting by reversals". In Mandoiu,I.I., Zelikovsky,A. (eds.) ISBRA 2007. LNCS (LNBI), vol. 4463, pp. 293–304. Springer, Heidelberg (2007).

[6] Berard S., Bergeron A., Chauve C. and Paul C., "Perfect sorting by reversals is not always difficult", IEEE / ACM Transactions on Computational Biology and Bioinformatics, Vol. 4, No. 1, pp. 4–16, 2007.

[7] Braga M. D. V., Gautier C. and Sagot M. F.,"An asymmetric approach to preserve common intervals while sorting by reversals", submitted to Algorithms for Molecular Biology, 2009.

[8] Bergeron A.,"A very elementary presentation of the Hannenhalli-Pevzner theory", Discrete Applied Mathematics, vol. 146, pp. 134–145, 2005.

[9] Blin, G., Fertin, G. and Chauve, C.,"The breakpoint distance for signed sequences", Texts in Algorithms, vol. 3, pp. 3-16, Computational Biology Nets 2004.

[10] Zheng, C., Lenert, A. and Sankoff, D.,"Reversal distance for partially ordered Genomes", Bioinformatics, Vol. 21, No. 1, Jun. 2005, pp. 502-508.

[11] M. D. V. Braga baobabLuna:"The solution space of sorting by reversals". Bioinformatics, 25(14):1833{1835, April 2009. Applications Note 11, issue 2, pp. 224–240, 1998.

[12] P. Berman and S. Hannenhalli,"Fast Sorting by Reversal". In D.S. Hirschberg and E.W. Myers, editors, Proceedings of the 7th Annual Symposium on Combinatorial Pattern Matching, Laguna Beach, Lecture Notes in Computer Science, 1075, Springer-Verlag, pp. 168–185, June 1996

[13] Setubal, J., and Meidanis, J. "Introduction to Computational Molecular Biology". PWS, Boston. 1997.

[14] Sankoff, D., and Blanchette, M. "Multiple genome rearrangement and breakpoint phylogeny. J. Computational Biology 1998. 5, pp. 555–570.

[15] Han Y.," Improving the Efficiency of Sorting by Reversals", Proceedings of The 2006 International Conference on Bioinformatics and Computational Biology, CSREA Press, Las Vegas, Nevada, USA, Jun 2006, pp. 406-409.

[16] P. Berman, S. Hannenhalli," Fast sorting by reversals", in: Proceedings of the CPM' 96, Lecture Notes in Computer Science, Vol. 1075, Springer, Berlin, Jun. 1996, pp. 168-185.

[17] Siepel, A.C.:"An algorithm to find all sorting reversals". In: Proceeding 6th Annual International' Conference on Computational Mol. Biology (RECOMB 2002), ACM Press, New York, April 2002, pp. 281–290.

[18] H. Kaplan, E. Verbin," Efficient data structures and a new randomized approach for sorting signed permutations by reversals", in 'Proceedings of the CPM' 03,Lecture Notes in Computer Science, Vol. 2676, Springer, Berlin, Jun. 2003, pp. 170–185.

[19] Hannenhalli, S. and Pevzner, P., " Transforming men into mice (polynomial algorithm for genomic distance problem)" , In Proceedings of the IEEE 36th Annual Symposium on Foundations of Computer Science, Oct. 1995, pp. 581-592.

[20] Caprara, A., "Sorting by reversals is difficult", RECOMB, Jan. 1997, pp. 75-83.

[21] V. Bafna and P. Pevzner, "Genome Rearrangement and Sorting by reversals", In 34th Annual Symposium on Foundation of Computer Science, Nov. 1993, pp. 148-157.

[22] Roy, Rahman and Thakur,"Sorting Primitives and Genome Rearrangement in 18.Bioinformatics: A Unified Perspective", World Academy of Science, Engineering and Technology, Vol. 38, Feb. 2008, pp. 363-368.

[23] Yin and Zhu, "Sorting Genome by Reversals and Translocations", Asia-Pacific Conference on Information Processing, Aug. 2009, pp. 391-394.

[24] Ajana, Y., Lefebvre, J.-F., Tillier, E.R.M., El. Mabrouk, N.: "Exploring the set of all minimal sequences of reversals and application to test the replication-directed reversal hypothesis". In: Guig´o, R.,Gusfield, D. (eds.) WABI 2002. LNCS, vol. 2452, Springer, Heidelberg, pp. 300–315. (2002).

[25] Swenson, K.M., Badr, G., Sankoff, D.: Listing all sorting reversals in quadratic time. In: Singh, M. (ed.) WABI 2010. LNCS, vol. 6293, Springer, Heidelberg, pp. 102–110, 2010.