# Resumo das publicações

Augusto Fernandes Vellozo

28 de dezembro de 2006

Neste texto, apresento a seguir o resumo das duas publicações que tenho até o momento.

## 1 Alignment with non-overlapping inversions in $O(n^3)$-time [18]

Alignments of sequences are widely used for biological sequence comparisons. Only biological events like mutations, insertions and deletions are usually modeled and other biological events like inversions are not automatically detected by the usual alignment algorithms.

Alignment with inversions does not have a known polynomial algorithm and a simplification to the problem that considers only non-overlapping inversions were proposed by Schöniger and Waterman [17] in 1992 as well as a corresponding $O(n^6)$ solution[1]. An improvement to an algorithm with $O(n^3 \log n)$-time complexity was announced in an extended abstract [1] and, in this present paper, we give an algorithm that solves this simplified problem in $O(n^3)$-time and $O(n^2)$-space in the more general framework of an edit graph.

Inversions have recently [5, 4, 12, 16] been discovered to be very important in Comparative Genomics and Scherer et al. in 2005 [10] experimentally verified inversions that were found to be polymorphic in the human genome. Moreover, 10% of the 1,576 putative inversions reported overlap RefSeq genes in the human genome. We believe our new algorithms may open the possibility to more detailed studies of inversions on DNA sequences using exact optimization algorithms and we hope this may be particularly interesting if applied to regions around known rearrangements boundaries. Scherer report 29 such cases and prioritize them as candidates for biological and evolutionary studies.

[1] In this case, $n$ denotes the maximal length of the two aligned sequences.

# 2 Alignment with non-overlapping inversions in $O(n^3 \log n)$-time (extended abstract) [1]

Alignment of sequences is widely used for biological sequence comparisons and can be associated with a set of edit operations that transform one sequence to the other. Usually, the only edit operations that are considered are the *substitution* (mutation) of one symbol by another one, the *insertion* of one symbol and *deletion* of one symbol. If costs are associated with each operation, there is a classic $O(n^2)$ dynamic program[2] that computes a set of edit operations with minimal total cost and exhibit the associated alignment, which has good quality and high likelihood for realistic costs.

Other important biological events like inversions are not automatically detected by the usual alignment algorithms and we can define a new edit operation, the *inversion* operation, which substitutes any segment by its *reverse complement* sequence. We can define a new alignment problem: given two sequences and fixed costs for each kind of edit operation, the *alignment with inversions* problem is an optimization problem that queries the minimal total cost of an edit operations set that transforms one sequence to the other. Moreover, one may also be interested in the exhibition of its correspondent alignment and/or edit operations. To the best of our knowledge, the computational complexities of alignment with inversions problem is unknown.

Some simplifications of this problem have been studied and were proved to be NP-complete [19, 3]. Many approximation algorithms were also proposed [15, 6]. Another important simplification is the problem known as *sorting signed permutations by reversals* and many polynomial algorithms have been obtained [13, 14, 2].

Another important approach was introduced in 1992, by Schöniger and Waterman [17]. They introduced a *simplification hypothesis*: *all regions involving in the inversions do not overlap.* This led to the *alignment with non-overlapping inversions* problem and they presented a $O(n^6)$ solution for this problem and also introduced a *heuristic* for it that reduced the running-time to something between $O(n^2)$ and $O(n^4)$.

Recently, indepent works [11, 8, 7, 9] gave exact algorithms for alignments with non-overlapping inversions with $O(n^4)$-time and $O(n^2)$-space complexity. In this present extended abstract, we announce an algorithm that solves this simplified problem in $O(n^3 \log n)$-time and $O(n^2)$-space.

[2] In this paper, $n$ denotes the maximal length of the two aligned sequences.

# Referências

[1] Carlos E. R. Alves, Alair Pereira do Lago, and Augusto F. Vellozo. Alignment with non-overlapping inversions in $O(n^3 \log n)$-time. *Electronic Notes in Discrete Mathematics*, 19:365–371, Jun 2005.

[2] David A. Bader, Bernard M. E. Moret, and Mi Yan. A linear-time algorithm for computing inversion distance between signed permutations with an experimental study. In *Algorithms and data structures (Providence, RI, 2001)*, volume 2125 of *Lecture Notes in Comput. Sci.*, pages 365–376. Springer, Berlin, 2001.

[3] Alberto Caprara. Sorting permutations by reversals and Eulerian cycle decompositions. *SIAM J. Discrete Math.*, 12(1):91–110 (electronic), 1999.

[4] Cáceres, Ranz, Barbadilla, Long, and Ruiz. Generation of a widespread Drosophila inversion by a transposable element. *Science*, 285(5426):415–418, Jul 1999.

[5] Cerdeño-Tárraga, Patrick, Crossman, Blakely, Abratt, Lennard, Poxton, Duerden, Harris, Quail, Barron, Clark, Corton, Doggett, Holden, Larke, Line, Lord, Norbertczak, Ormond, Price, Rabbinowitsch, Woodward, Barrell, and Parkhill. Extensive DNA inversions in the B. fragilis genome control variable gene expression. *Science*, 307(5714):1463–1465, Mar 2005.

[6] David A. Christie. A 3/2-approximation algorithm for sorting by reversals. In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (San Francisco, CA, 1998)*, pages 244–252, New York, 1998. ACM.

[7] A. P. do Lago, C. A. Kulikowski, E. Linton, J. Messing, and I. Muchnik. Comparative genomics: simultaneous identification of conserved regions and their rearrangements through global optimization. In *The Second University of Sao Paulo/Rutgers University Biotechnology Conference*, Rutgers University Inn and Conference Center, New Brunswick, NJ, August 2001.

[8] Alair Pereira do Lago, Ilya Muchnik, and Casimir Kulikowski. An $O(n^4)$ algorithm for alignment with non-overlapping inversions. In *Second Brazilian Workshop on Bioinformatics, WOB 2003*, Macaé, RJ, Brazil, 2003. http://www.ime.usp.br/~alair/wob03.pdf.

[9] Alair Pereira do Lago, Ilya Muchnik, and Casimir Kulikowski. A sparse dynamic programming algorithm for alignment with non-overlapping inversions. *Theor. Inform. Appl.*, 39(1):175–189, 2005.

[10] Feuk, MacDonald, Tang, Carson, Li, Rao, Khaja, and Scherer. Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. *PLoS Genet*, 1(4):e56, Oct 2005.

[11] Yong Gao, Junfeng Wu, Robert Niewiadomski1, Yang Wang, Zhi-Zhong Chen, and Guohui Lin. A space efficient algorithm for sequence alignment with inversions. In *Computing and Combinatorics, 9th Annual International Conference, COCOON 2003*, volume 2697 of *Lecture Notes in Computer Science*, pages 57–67. Springer-Verlag, 2003.

[12] Graham and Olmstead. Evolutionary significance of an unusual chloroplast DNA inversion found in two basal angiosperm lineages. *Curr Genet*, 37(3):183–188, Mar 2000.

[13] Sridhar Hannenhalli and Pavel Pevzner. Transforming cabbage into turnip: polynomial algorithm for sorting signed permutations by reversals. In *STOC '95: Proceedings of the twenty-seventh annual ACM symposium on Theory of computing*, pages 178–189, New York, NY, USA, 1995. ACM Press.

[14] Haim Kaplan, Ron Shamir, and Robert E. Tarjan. A faster and simpler algorithm for sorting signed permutations by reversals. *SIAM J. Comput.*, 29(3):880–892 (electronic), 2000.

[15] J. Kececioglu and D. Sankoff. Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement. *Algorithmica*, 13(1-2):180–210, 1995.

[16] Kuwahara, Yamashita, Hirakawa, Nakayama, Toh, Okada, Kuhara, Hattori, Hayashi, and Ohnishi. Genomic analysis of Bacteroides fragilis reveals extensive DNA inversions regulating cell surface adaptation. *Proceedings of the National Academy of Sciences U S A*, 101(41):14919–14924, Oct 2004.

[17] M. Schöniger and M. S. Waterman. A local algorithm for DNA sequence alignment with inversions. *Bulletin of Mathematical Biology*, 54(4):521–536, Jul 1992.

[18] Augusto F. Vellozo, Carlos E. R. Alves, and Alair Pereira do Lago. Alignment with non-overlapping inversions in $o(n^3)$-time. In *6th Workshop on Algorithms in Bioinformatics*. Springer, 2006. Lecture Notes in Bioinformatics 4175.

[19] R. Wagner. On the complexity of the extended string-to-string correction problem. In *Seventh ACM Symposium on the Theory of Computation*. Association for Computing Machinery, 1975.