

Evolutionary Two-Dimensional DNA Sequence Alignment

Edgar E. Vallejo¹ and Fernando Ramos²

¹ Computer Science Dept., Tecnológico de Monterrey, Campus Estado de México
Carretera Lago de Guadalupe Km 3.5 Col. Margarita Maza de Juárez, 52926
Atizapán de Zaragoza, Eestado de México, México

`vallejo@itesm.mx`

² Computer Science Dept., Tecnológico de Monterrey, Campus Cuernavaca
Ave. Paseo de la Reforma 182 Col. Lomas de Cuernavaca, 62589

Cuernavaca, Morelos, México

`fernando.ramos@itesm.mx`

Abstract. This article presents a model for DNA sequence alignment. In our model, a finite state automaton writes two-dimensional maps of nucleotide sequences. An evolutionary method for sequence alignment from this representation is proposed. We use HIV as the working example. Experimental results indicate that structural similarities produced by two-dimensional representation of sequences allow us to perform pairwise and multiple sequence alignment efficiently using genetic algorithms.

1 Introduction

The area of bioinformatics is concerned with the analysis of molecular sequences to determine the structure and function of biological molecules [2]. Fundamental questions about functional, structural and evolutionary properties of molecular sequence can be answered using sequence alignment.

Research in sequence alignment has focused for many years on the design and analysis of efficient algorithms that operate on linear character representation of nucleotide and protein sequences. The intractability of multiple sequence alignment algorithms evidences limitations for the analysis of molecular sequences from this representation. Similarly, due to the extension of typical genomes, this representation is also inconvenient from the human perception perspective.

2 The Model

In our model, a finite state automaton writes a two-dimensional map of DNA sequences [3]. The proposed alignment method is based on the overlapping of a collection of these maps. We overlap a two-dimensional map over another to discover coincidences in character patterns. Sequence alignment consists of the sliding of maps over a reference plane in order to search for the optimum overlapping. We use genetic algorithms to evolve the cartesian positions of a collection of maps that maximize coincidences in character patterns.

3 Experiments and Results

We performed several runs using HIV nucleotide sequences. Figure 1 shows the results of a typical run. We performed comparisons using conventional sequence alignment methods that operate on linear sequences. We found that our method yields similar results to those produced by the SIM local alignment algorithm.

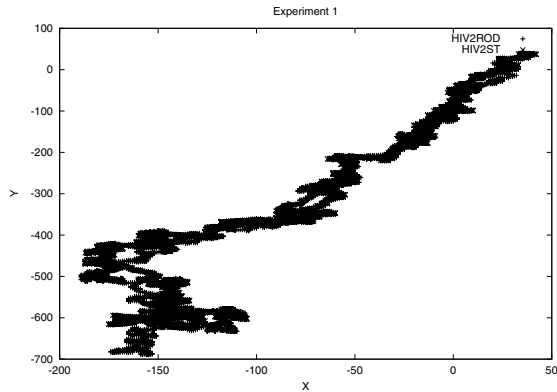


Fig. 1. Results. Pairwise DNA sequence alignment

4 Conclusions and Future Work

We present a sequence alignment method based on two-dimensional representation of DNA sequences and genetic algorithms. An immediate extension of this work is the consideration of protein sequences and the construction of phylogenies from two-dimensional alignment scores. Finally, a more detailed comparative analysis using evolutionary [1] and conventional [2] alignment methods could elucidate the significance of evolutionary two-dimensional sequence alignment.

References

1. Fogel, G. E., Corne, D. W. (eds.) 2003. *Evolutionary Computation in Bioinformatics*. Morgan Kaufmann Publishers.
2. Mount, D. 2000. *Bioinformatics. Sequence and Genome Analysis*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
3. Vallejo, E. E., Ramos, F. 2002. Evolving Finite Automata with Two-dimensional Output for Biosequence Recognition and Visualization In W. B. Langton, E. Cantú-Paz, K. Mathias, R. Roy, R. Poli, K. Balakrishnan, V. Honovar, G. Rudolph, J. Wegener, L. Bull, M. A. Potter, A. C. Schultz, J. E. Miller, E. Burke, N. Jonoska (eds.) *Proceedings of the Genetic and Evolutionary Computation Conference GECCO 2002*. Morgan Kaufmann Publishers.