

UVWORD-WEB: a web-based tool for DNA word analysis

Vicente Arnau*, Jordi Llorens*, Miguel Gallach* and Ignacio Marín*

*Departamento de Informática. *Departamento de Genética.

Universidad de Valencia. Spain



<http://www.uv.es/~genomica/>

To whom correspondence should be addressed:
Vicente.Arnau@uv.es

INTRODUCTION

We have developed a novel web-based tool for DNA word analysis: UVWORD-WEB. It is freely accessible for registered users at <http://protegeno.uv.es/>, and it has been optimised for Internet Explorer 4.0 or higher. The program has been written in a combination of HTML, Javascript, PHP and C languages. This last language has been used for increasing the speed of execution of the algorithms of analysis.

UVWORD scans for words of a particular size a given sequence and present four different types of analysis:

- Frequency Analysis.
- Frequency Comparative Analysis.
- DNA Comparative (Source-Target Analysis).
- Smoothing Analysis.



Fig1. Initial Page.

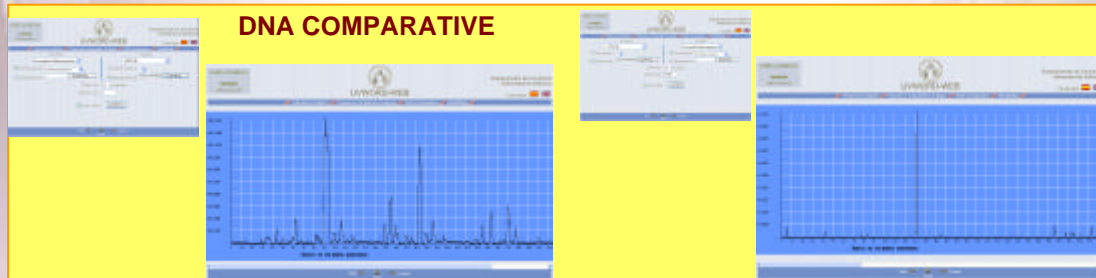
FREQUENCY ANALYSIS

word	frequency
AAAAAAAAAAAA	25217
CTCTAATCCAGC	6875
CTGGATTACAGC	5077
GCCTTAATCCAGC	5167
GCCTGGATTACAGC	6799
TTTTTTTTTTTT	25786

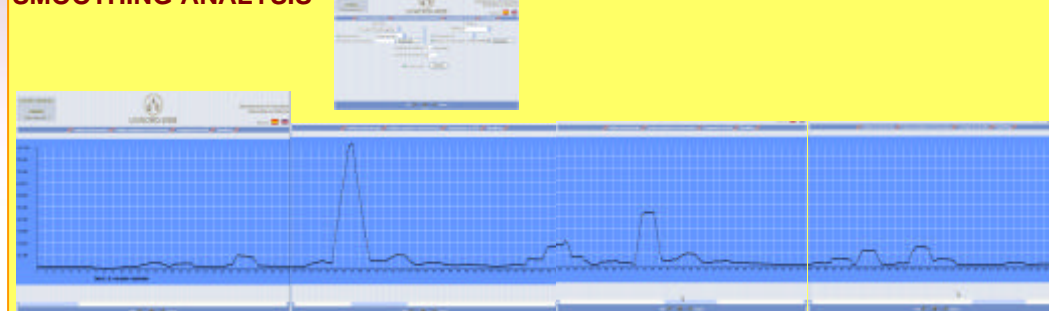
FREQUENCY COMPARATIVE ANALYSIS

WORD	FILE_1	FILE_2
ACTTAGAGAGATC	0	60
ATCTCTAGATGC	1	50
CTCTAGAGATGSA	0	45
GCAGTCTAGAGSA	0	41
CTCTAGAGATGCC	0	60
CTCTAGATGATG	0	46
GCAGTCTAGATCT	0	40
GCCTGGATTCAGC	1	41
GCAGTCTAGAG	0	51
GCAGTCTAGAT	0	54
GCAGTCTAGAT	1	51
GCAGTCTAGAT	0	42
GCAGTCTAGAT	0	45
GCAGTCTAGAT	0	57
GCAGTCTAGAT	0	47
GCAGTCTAGAT	0	54
GCAGTCTAGAT	0	42

DNA COMPARATIVE



SMOOTHING ANALYSIS



RESULTS AND DISCUSSION:

Algorithms have been optimized to allow a speed of at least 16 millions of words per second for $k = 8$ and at least 3.5 millions of words per second for larger k s.

REFERENCES:

Arnau, V and Marín, I. (2003) A fast algorithm for the exhaustive analysis of 12-nucleotide-long DNA sequences: application to human genomics. *Proceedings of the 17th International Parallel and Distributed Processing Symposium*. IEEE Computer Society, p. 153