

SNUFER: A software for localization and presentation of single nucleotide polymorphisms using a Clustal multiple sequence alignment output file

Marco A. B. Mansur¹, Giovana P. Cardozo¹, Elaine V. Santos¹ and Mozart Marins^{1,2,*}

¹Unidade de Biotecnologia, UNAERP, Ribeirão Preto, SP, Brazil; ²Heranza - Biotecnologia, Ribeirão Preto, SP, Brazil; Mozart Marins* - Email: marins@heranza.com.br; * Corresponding author

received August 13, 2008; accepted September 17, 2008; published October 17, 2008

Abstract:

SNUFER is software for the automatic localization and generation of tables used for the presentation of single nucleotide polymorphisms (SNPs). A multiple sequence alignment is generated using ClustalW ran inside SNUFER using an input of a FASTA file containing the sequences to be analyzed. The ClustalW output file is then used to generate a table which displays the SNPs detected in the aligned sequences and their degree of similarity. This table can be exported to Microsoft Word, Microsoft Excel or as a single text file, permitting further editing for publication. The software was written using Delphi 7 for programming and FireBird 2.0 for sequence database management. It is freely available for non-commercial use and can be downloaded from the URL <http://www.heranza.com.br/bioinformatica2.htm>.

Keywords: software; single nucleotide polymorphisms; multiple sequence alignment; ClustalW

Background:

In pathogenic microorganisms, single nucleotide polymorphisms (SNP) are used to define strains and are analyzed for possible correlations with geographic origin and differences in pathogenesis [1, 2]. After DNA sequencing of gene fragments from different individuals of the same species, SNPs are identified by multiple alignment of DNA sequences. One of the most popular software used for this purpose is Clustal, which employs a fast pairwise alignment algorithm [3], helping the user to identify identical and non-identical nucleotides along sequences. The most popular way of presenting SNPs based on these results is to manually build a table which shows the nucleotide changes between a selected reference sequence and the other aligned sequences. SNUFER was developed to automatize this task by integrating the steps of multiple sequence alignment, SNP identification and data presentation in a publication format.

Software description

After sequences are uploaded in SNUFER as a fasta file, the program provides the option to select all or a group of the sequences for SNP analysis. Using ClustalW [4], SNUFER performs a multiple sequence alignment and the result is shown in a table. The user then has three options. The first option is to run an SNP detection and to display a table of results which can be exported to Microsoft Word, Microsoft Excel or as a single text file (Figure 1). Any of the sequences can be selected as a reference for scoring nucleotide change and position. The exported table can be built with all SNPs found or with only those selected by the user. In the first two top rows of the table, the nucleotides (represented by A, T, C, G) and their position in the

reference sequence is shown for each SNP detected. The underlying rows will display, for each SNP position, the nucleotide change in each sequence, represented by A, T, C or G. Identical nucleotides between the sequence analyzed and the reference sequences will be indicated by an asterisk (*). The first column of the table will display sequence information such as the accession number and the second column will display the percentage of similarity between the reference sequence and each of the sequences analyzed. The second user option is to build a Microsoft Word or Excel table to show the similarity between two selected sequences or between all sequences. Finally, the third option is to display a consensus strand based on the Clustal alignment. SNUFER output files can also be exported as a text-only file which can later be imported into other applications.

Concluding remarks

SNUFER was written using Delphi 7 for programming and FireBird 2.0 for sequence database management. It is freely available for non-commercial use. The URL is <http://www.heranza.com.br/bioinformatica2.htm> for download. The program runs under Windows and not in Linux or Mac. SNUFER is a useful tool for the localization and presentation of SNPs detected, for example, in genes used for phylogenetic studies [5, 6].

Acknowledgment:

This study was supported by funds from Fundação de Amparo à Pesquisa do Estado de São Paulo (grants 04/14196-5, 06/068904). We thank Dr S.C. França and the

staff of the Biotechnology Unit UNAERP for general support and K. Markendorf for revision of the English text.

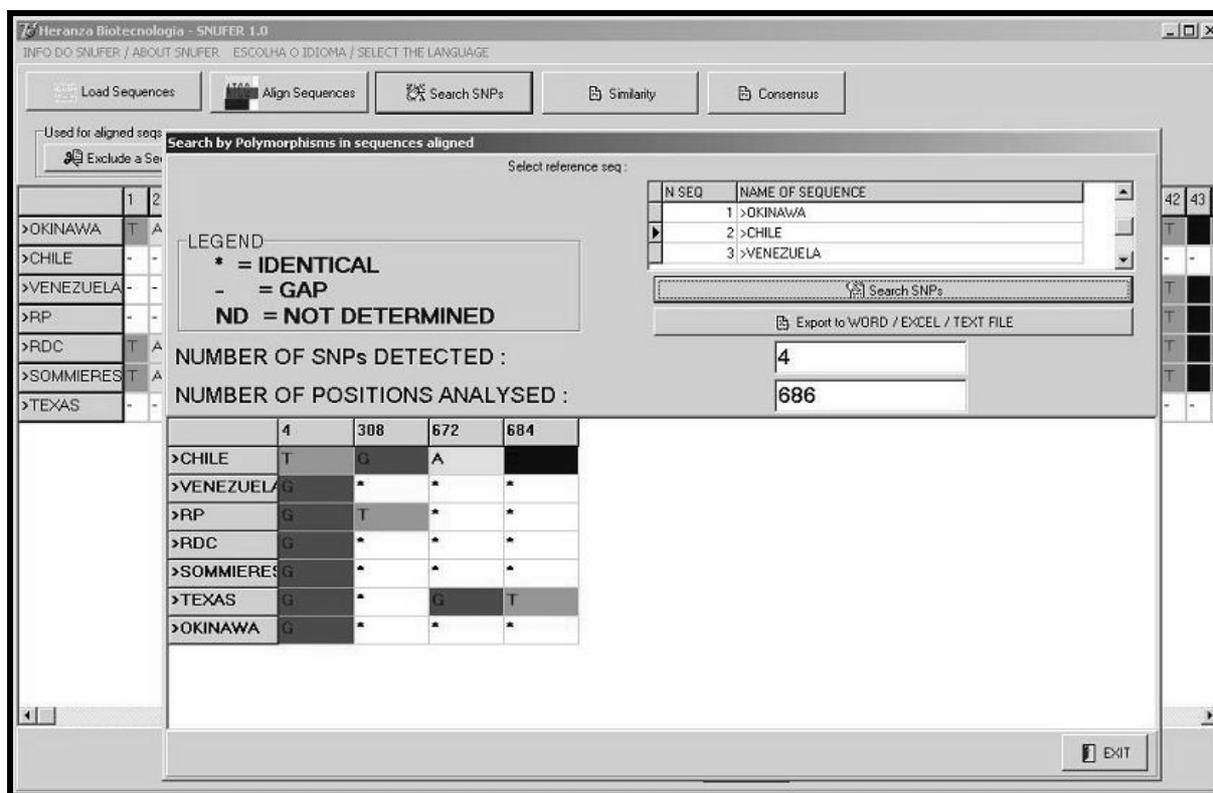


Figure 1: Representative screen of an SNP table built from a multiple sequence alignment performed by ClustalW. The table can be exported to Word, Excel or as a text file.

References:

- [01] K. Miura and Y. Rikihisa, *Infect Immun.*, 75: 3604 (2007) [PMID: 17438035]
- [02] S. K. Shukla *et al.*, *J Clin Microbiol.*, 45: 2312 (2007) [PMID: 17507511]
- [03] D. G. Higgins and P.M. Sharp, *Gene*, 73: 237 (1988) [PMID: 3243435]
- [04] M. A. Larkin *et al.*, *Bioinformatics*, 23: 2947 (2007) [PMID: 17846036]
- [05] G. P. Cardozo *et al.*, *Brazilian Journal of Microbiology*, 38: 478 (2007)
- [06] H. Inokuma *et al.*, *Clin Diagn Lab Immunol.*, 9: 1132 (2002) [PMID: 12204973]

Edited by P. Kanguane

Citation: Mansur *et al.*, *Bioinformatics* 3(2): 63-64 (2008)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.