

## Universal Fingerprinting Chip Server

Janet Casique-Almazán, Violeta Larios-Serrato, Gabriela Edith Olguín-Ruíz, Carlos Javier Sánchez-Vallejo, Rogelio Maldonado-Rodríguez & Alfonso Méndez-Tenorio\*

Laboratory of Biotechnology and Genomic Bioinformatics, Department of Biochemistry, National School of Biological Sciences, National Polytechnic Institute, CP 11340, Mexico City, Mexico; Alfonso Méndez-Tenorio – Email: amendezt@ipn.mx; Phone: +52 55-5729-6000 ext 62322; \*Corresponding author

Received May 23, 2012; Accepted June 16, 2012; Published June 28, 2012

### Abstract:

The Virtual Hybridization approach predicts the most probable hybridization sites across a target nucleic acid of known sequence, including both perfect and mismatched pairings. Potential hybridization sites, having a user-defined minimum number of bases that are paired with the oligonucleotide probe, are first identified. Then free energy values are evaluated for each potential hybridization site, and if it has a calculated free energy of equal or higher negative value than a user-defined free energy cut-off value, it is considered as a site of high probability of hybridization. The Universal Fingerprinting Chip Applications Server contains the software for visualizing predicted hybridization patterns, which yields a simulated hybridization fingerprint that can be compared with experimentally derived fingerprints or with a virtual fingerprint arising from a different sample.

**Availability:** <http://bioinformatica.homelinux.org/UFCVH/>

**Keywords:** Fingerprinting, universal, microarray, phylogenetic, virtual hybridization.

### Background:

Several mathematical models have been widely used to predict thermal stability of oligonucleotide duplexes. Whereas empirical methods based on nucleotide composition alone can be used to calculate the thermal stability with reasonable precision, nearest-neighbor (NN) models are better considered for describing thermal stability of oligonucleotide duplexes because this is strongly correlated with sequence variations [1, 2]. Complete sets of NN parameters for energy contributions due to perfect base-pairings, single mismatches, and dangling ends have been published recently [3-5]. These thermodynamic parameters provide a basis for predicting the stability of both perfectly paired and mismatch-containing duplex structures. Several computer tools have been developed that predict thermal stability of probes in order to select the most appropriate probes for microarray design [6, 7], but there are comparatively few of them that incorporate such analysis, using the most recent thermodynamic parameters, for probe design,

and even fewer of them incorporate analytic tools for both designing and interpreting microarray experiments. Furthermore, none of them analyze the probe set for the existence of stable mismatched-duplexes with the target. This server allows web access to the tool called Virtual Hybridization (VH) [8], which is a simulation of the hybridization reaction between the probes and DNA targets (including complete genomes), considering thermodynamic data for predicting the stability of the double chain and finding probable sites for hybridization as would happen in real conditions. This simulation takes into account not only the stability of perfect matched-duplexes, but also the stability of mismatched target-probe duplexes.

### Web server design:

All algorithms for DNA stability prediction have been implemented in the VH, which was written using the programming software Borland Delphi 7.0 (Borland Software

Corporation). Linux versions of the VH software have been developed using Borland Kylix Desktop 1.0 (Borland International), and they have been successfully tested in Linux Red Hat 7.3, Fedora Core 16 and Mandrake 10. User interface of the web server has been created with PHP 5.2 and the system has been currently tested in an Apache 2.2 server running on Linux Fedora Core 16.

### Software input:

The registration is free for both academic and non-commercial applications. Registered users can upload target and custom probe sets to perform the virtual hybridization. Alternatively users can select previously designed universal probe sets. Next, define the hybridization parameters: i) Cut-off free energy for stable hybridization, ii) Allowed number of mismatches and iii) Select target strand (direct, complementary or both). Finally, select the format mode of results according to their needs for the detailed information about hybridization of each probe and summary (global) table of results, (Figure 1a).

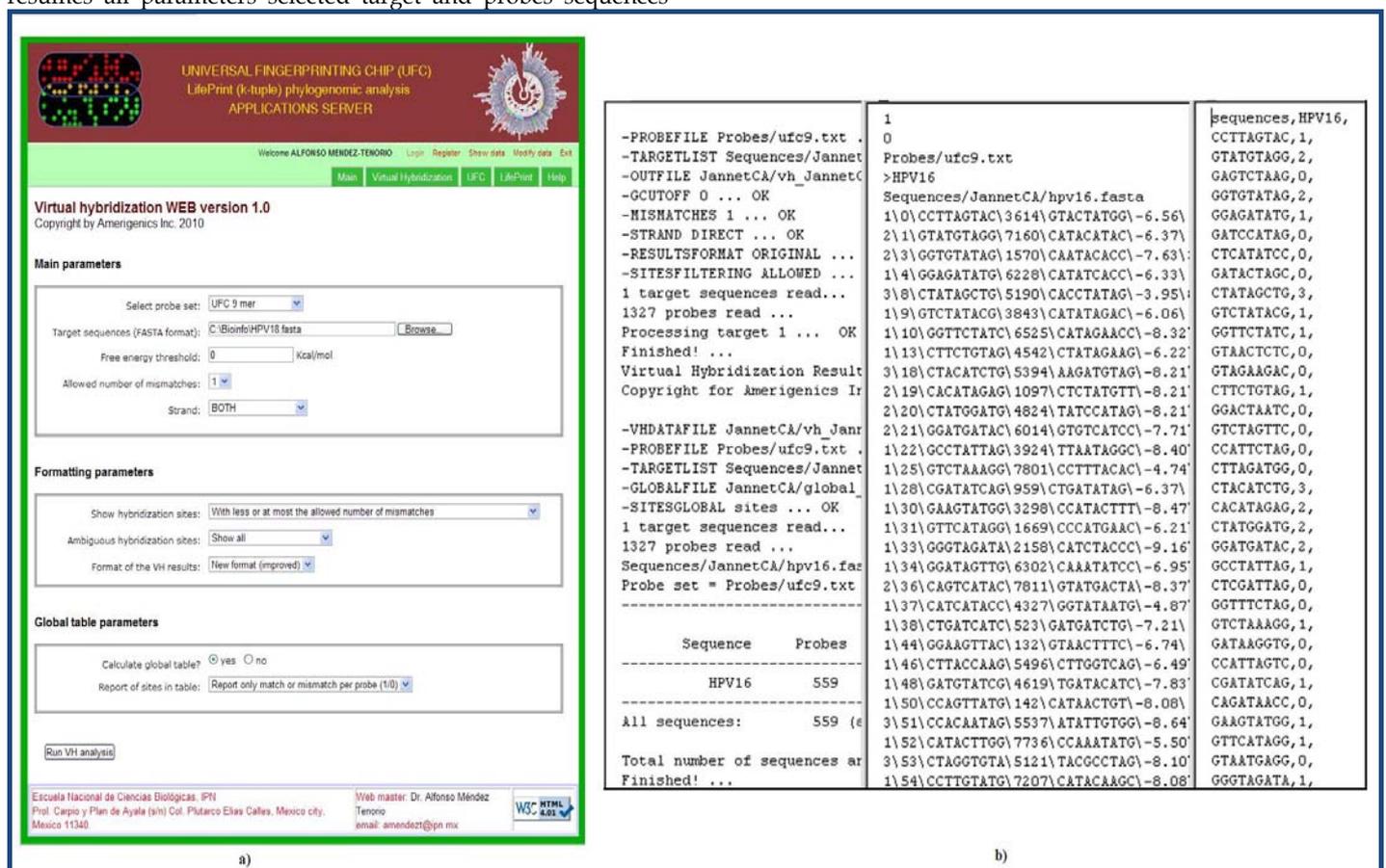
### Software output:

Once the virtual hybridization process is done, the software generates three output files (Figure 1b). A log file which resumes all parameters selected target and probes sequences

tested and summary of the results. The vh file shows the details of the results of the hybridization reporting total number of targets and probes, name of probe sets, description of target files and for each hybridized probe: total number of sites, number and sequence of the probe, target hybridization positions, recognized target sequence and free energy value. The global file shows one of the three alternative options: the overall results of the probes hybridized with each genome in binary code, absence (0) and presence (1), the total number of probe hits/target or, the total number of paired bases per probe. Global file data can be used for comparing virtual hybridization patterns which in turn can be analyzed for building phylogenetic trees using third-party software packages such as Phylip or MEGA.

### Future Development:

Current aim of this server is the prediction of hybridization probes against known target sequences. Hybridization patterns can be used for comparing sequences which can be used for both phylogenetic and phylogenomic studies. Future development will include specialized analytical tools for such applications.



**Figure 1:** Universal Fingerprinting Chip web server. a) Main parameters to perform virtual hybridization between a probe set of definite length and a set of user-defined sequences. b) Display of VH results file.

### Acknowledgement:

The web server design was supported by a grant from the National Council for Science and Technology (CONACyT project #105833).

### References:

[1] von Ahlsen N *et al. Clin Chem.* 2001 **47**: 1956 [PMID: 11673362]

- [2] SantaLucia J Jr, *Proc Natl Acad Sci U S A*. 1998 **95**: 1460 [PMID: 9465037]
- [3] Allawi HT & SantaLucia J Jr, *Biochemistry*. 1998 **37**: 2170 [PMID: 9485363]
- [4] Johnson T *et al.* *Biochemistry*. 1998 **37**: 12343 [PMID: 9724548]
- [5] Peyret N *et al.* *Biochemistry*. 1999 **38**: 3468 [PMID: 10090733]
- [6] Emrich SJ *et al.* *Nucleic Acids Res*. 2003 **31**: 3746 [PMID: 12824409]
- [7] Pozhitkov AE & Tautz D, *BMC Bioinformatics*. 2002 **3**: 9 [PMID: 11882251]
- [8] Reyes-López MA *et al.* *Nucleic Acids Res*. 2003 **31**: [PMID: 12527788]

**Edited by P Kanguane**

**Citation:** Casique-Almazán *et al.* *Bioinformation* 8(12): 586-588 (2012)

**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