Multi-Harmony:

detecting functional specificity from sequence alignment

K. Anton Feenstra Bernd W. Brandt Jaap Heringa

Centre for Integrative Bioinformatics VU (IBIVU), VU University Amsterdam

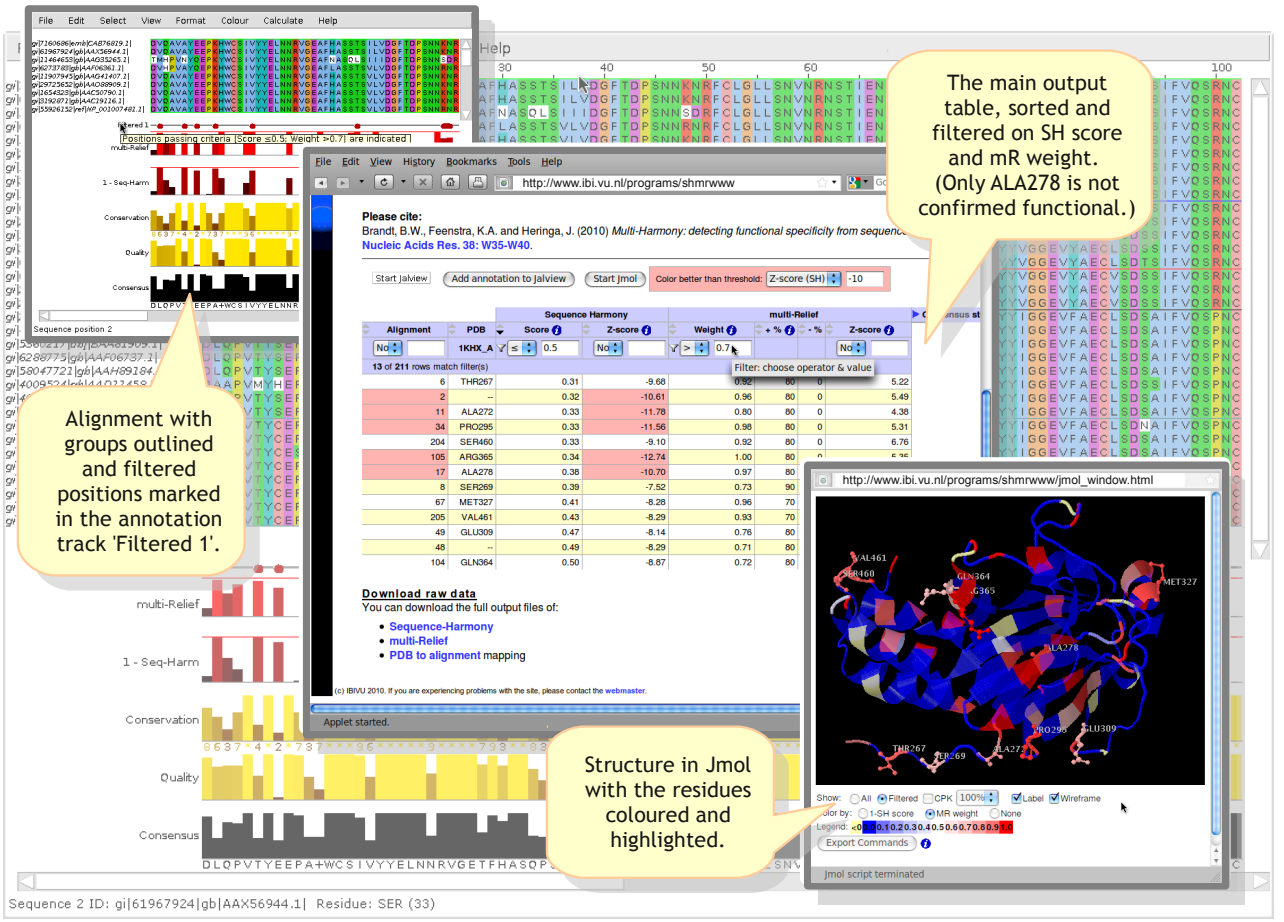
Contact: feenstra@few.vu.nl, bwbrandt@few.vu.nl, heringa@few.vu.nl Download this poster at: www.few.vu.nl/~feenstra/posters.html

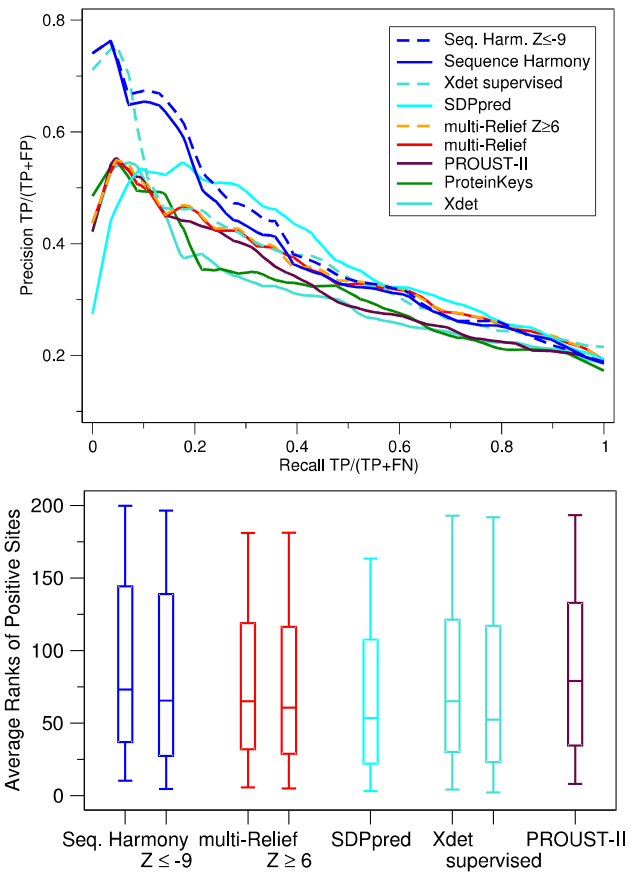
Introduction

Many protein families contain sub-families with functional specialization, such as binding different ligands or being involved in different protein-protein interactions. A small number of amino acids generally determine functional specificity. The identification of these residues can aid the understanding of protein function and help finding targets for experimental analysis. Our new multi-Harmony web server reliably detects these subtype-specific sites from an alignment and sub-grouping. It can be used as an interactive tool for detecting and exploring subtype-specific sites in the protein sequence and structure.

Results

Multi-Harmony is benchmarked on a test set comprising 23 protein families with known specificity sites and functional sub-grouping [4]. Performance at highest precision continues up to higher recall compared to all other methods tested (below, upper panel), and average ranks are comparable to other methods (below, lower panel). Overall performance is shown to be comparable to or better than the best of alternative available methods, such as SDPpred, ProteinKeys, PROUST-II and Xdet.





Methods

Combining our Sequence Harmony [1,2] and multi-Relief [3] methods in one web server [4] allows simultaneous analysis and comparison of specificity residues. Our methods have been significantly improved and extended:

- Sequence Harmony now handles multiple (more than two) subgroups
- multi-Relief has been changed from a sampling implementation to a deterministic one, making it more consistent and user friendly
- both methods now report Z-scores (see figure on the right)

The input is a multiple sequence alignment (FASTA, ClustalW, MSF, Stockholm or SELEX) and a list of group sizes. Optionally, a PDB file or ID can be provided or searched on the fly with BLAST. A remote REST-like access script is available on request for expert users.

The output table can be interactively sorted or filtered and the annota-ted alignment can be analysed in Jalview and Jmol (above); it includes:

- Sequence Harmony results (multi-group)
- multi-Relief results
- PDB to alignment mapping (if PDB ID was provided)

Multi-Harmony is available at:

http://www.ibi.vu.nl/programs/shmrwww/

References

[1] Pirovano W, Feenstra KA and Heringa J. (2006) Sequence comparison by Sequence Harmony identifies subtype specific functional sites. *Nucleic Acids Res* 34: 6540-6548.
[2] Feenstra KA, Pirovano W, Krab K and Heringa J. (2007) Sequence Harmony: detecting functional specificity from alignments. *Nucleic Acids Res* 35: W495-W498.
[3] Ye K, Feenstra KA, Heringa J, IJzerman AP and Marchiori E. (2008) Multi-RELIEF: a method to recognize specificity determining residues from multiple sequence alignments using a Machine Learning approach for feature weighting. *Bioinformatics* 24:18-25.
[4] Brandt BW, Feenstra AF and Heringa J. (2010) Multi-Harmony: detecting functional specificity from sequence alignment. *Nucleic Acids Res* 38: W35-W40.

ENFIN Enabling Systems Biology
CENTRE FOR INTEGRATIVE BIOINFORMATICS VU
nbic Netherlands Bioinformatics Centre