fLPS 2.0: rapid annotation of compositional biases in biological sequences

fLPS2 is a program that rapidly annotates compositionally-biased regions in biological sequences (both protein and DNA). The algorithm picks out 'Low Probability Subsequences' (LPSs) through a process of probability minimization. fLPS2 is explained in detail in the citations below. Using fLPS2, compositionally-biased regions of two types can be automatically annotated:

(i) **single-residue:** These occur when there are many residues of a single type in a short span. For example, PPPPPEPPPPAPPPEPPPIP is a protein stretch biased for P (proline). This is given a signature {P}.
(ii) **multiple-residue:** These occur when two or more residue types predominate in a short span. For example, HQHQHQQQQHHQQQQHHHHAHHHQQHHQIQQQQ is a protein region biased for H and Q (histidine and glutamine). This is given a signature {QH}.

The latest version is available at:

https://github.com/pmharrison/flps2/releases/latest .

Annotation of 'low-complexity' regions:

To label 'low-complexity' regions, the following run parameters are suitable:
./fLPS2 -t 3.5e-05 -m7 -M17 INPUT.FASTA > OUTPUT.FASTA
These were derived in the 2024 article in *Scientific Reports*, cited below. The software for choosing parameters for both fLPS2 and SEG (segmasker), that is a feature of this article, is available at:
https://github.com/pmharrison/parameters/releases/latest .
In this article, the user can find **a procedure for picking out similar compositionally-biased regions to a query**. The M domain from the Sup35p prion-forming protein from *S. cerevisiae* is used as an example.
The fLPS2 program is featured as a part of the ***Patterny*** package for analyzing intrinsic disorder, low-complexity and/or compositionally-biased regions in proteins. This is available at: https://github.com/pmharrison/patterny .

Citation and contact

Please cite these, if you use fLPS2:

P. M. Harrison. (2017). "fLPS: fast discovery of compositional biases for the protein universe", *BMC Bioinformatics*, **18(1):** 476, *doi: 10.1186/s12859-017-1906-3*.
P.M. Harrison. (2021). “fLPS 2.0: rapid annotation of compositional biases in biological sequences”, *PeerJ,* Oct 28:9:e12363. doi: 10.7717/peerj.12363. eCollection 2021.
P. M. Harrison (2024). “Optimizing strategy for the discovery of compositionally-biased or low-complexity regions in proteins”, Sci Rep. 2024 Jan 5;14(1):680. doi: 10.1038/s41598-023-50991-8.

For further inquiries and comments about fLPS2, please contact Paul Harrison