



HAL
open science

HH-MOTiF: de novo detection of short linear motifs in proteins by Hidden Markov Model comparisons

Roman Prytuliak, Michael Volkmer, Markus Meier, Bianca Habermann

► To cite this version:

Roman Prytuliak, Michael Volkmer, Markus Meier, Bianca Habermann. HH-MOTiF: de novo detection of short linear motifs in proteins by Hidden Markov Model comparisons. *Nucleic Acids Research*, 2017, 45 (18), pp.10921 - 10921. 10.1093/nar/gkx810 . hal-01692042

HAL Id: hal-01692042

<https://hal.science/hal-01692042>

Submitted on 5 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Corrigendum

HH-MOTiF: *de novo* detection of short linear motifs in proteins by Hidden Markov Model comparisons

Roman Prytuliak¹, Michael Volkmer¹, Markus Meier² and Bianca H. Habermann^{1,3,*}

¹Computational Biology Group, Max Planck Institute of Biochemistry, Martinsried, Germany, ²Research Group Quantitative Biology and Bioinformatics, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany and ³Computational Biology Group, Developmental Biology Institute of Marseille (IBDM) UMR 7288, CNRS, Aix Marseille Université, Marseille 13288 Cedex 9, France

Nucleic Acids Res (2017) 45 (W1): W470–W477. <https://doi.org/10.1093/nar/gkx341>

The authors wish to make the following correction to their article.

In the Supplementary Information, the following equation is incorrect:

Regex generation and evaluation: The p-value of a regex with highly or moderately conserved columns C_{cons} and cumulative frequencies $freq_j$ of amino acids in the input dataset listed in the column j is calculated according to the formula:

$$p = \frac{\sum_{i=1}^{N_{corr}} \left(1 - \left(1 - \prod_{j=1}^{C_{cons}} freq_j \right)^M \right)}{N_{corr}}$$

This formula is the averaged version of the Šidák correction formula (1). N_{corr} is the effective number of non-homologous proteins for a given motif tree.

The correct equation is provided below and has been corrected in the published article:

Regex generation and evaluation: The p-value of a regex with highly or moderately conserved columns C_{cons} and cumulative frequencies $freq_j$ of amino acids in the input dataset listed in the column j is calculated according to the formula:

$$p = \frac{\sum_{i=1}^{N_{tree}} \left(1 - \left(1 - \prod_{j=1}^{C_{cons}} freq_j \right)^M \right)}{N_{tree}}$$

This formula is the averaged version of the Šidák correction formula (1). N_{tree} is the number of proteins a given motif tree locates to.

1. Wright, S.P. (1992) Adjusted P-values for simultaneous inference. *Biometrics*, 48, 1005–1013.

*To whom correspondence should be addressed. Tel: +33 4 91 26 92 36; Email: bianca.habermann@univ-amu.fr