

A powerful HMMER for data mining

June 3 2015



HMMER, a fast, sensitive search tool, helps you find sequence relationships deep in evolutionary time. Credit: EMBL-EBI

HMMER software implements a powerful new generation of mathematical techniques for identifying hundreds of thousands of

related sequences. HMMER results help researchers infer the function of a protein and its evolutionary history. A new, open-source web interface at EMBL-EBI offers fast, easy-to-use search and visualization.

- HMMER software implements a powerful new generation of mathematical techniques for identifying hundreds of thousands of related sequences;
- HMMER results help researchers infer the function of a protein and its evolutionary history;
- New, open-source web interface offers fast, easy-to-use search and visualisation.

Researchers looking to understand the function and evolutionary history of a protein can now use HMMER algorithms through a dedicated website. HMMER, which uses sophisticated probability models and searches large sequence databases in seconds, has been incorporated into protein data services at EMBL-EBI including Pfam and InterPro.

HMMER was originally developed by Sean Eddy, currently at the Howard Hughes Medical Institute's Janelia Research Campus in the US, and the project turned into a 10-year collaboration with Rob Finn, now head of Protein Family resources at EMBL-EBI. The most recent version of the website is faster and more robust, and is widely available with a full suite of tools to help researchers interpret protein data.

"We wanted HMMER to be more accessible, because it is an incredibly powerful tool for looking at protein function," explains Finn. "Five years ago, the same kind of search would take hours to perform - now you can search more interactively, and the iterative search lets you start with a single sequence and pull in hundreds of thousands of related sequences. That takes you straight to these very distant relationships, which let you infer function and [evolutionary history](#) of a protein. So in my own research, I can now follow my train of thought without being interrupted

by annoying holding pages saying that my search is running."

The HMMER website lets users filter the input and output sequences by taxonomy or domain architecture, making it easier to interpret results. It also provides an intuitive visual interface that lets users navigate seamlessly between analyses just by clicking icons or figure elements such as histogram bars, table entries and taxonomy trees.

"Here at Janelia we're focused on the probability theory and computer engineering that power HMMER's algorithms," says Eddy. "What Rob and his team have done with the new EMBL-EBI web site is to make HMMER far more useful and accessible to biologists. The new website is super fast, but even more than that, it incorporates a lot of creative data visualisation and interactive tools, based on Rob's long experience with protein sequence analysis. Rob's work takes the HMMER project to a new level."

HMMER is both fast and extremely sensitive, detecting distant relationships and identifying fragments of sequences. The deep homologous relationships HMMER can find enable many aspects of computational biology research, from characterising metagenomic communities to understanding the evolution of development.

What's next? The HMMER web tools are already used as a curation platform for Pfam, and can be used to add expert community knowledge to public data resources. This will potentially expand the number of freely available, high-quality, annotated [protein](#) family entries available to researchers worldwide. The next step for the collaborators is to extend the software to accommodate DNA searches, which involves far larger datasets.

More information: europepmc.org/abstract/MED/21593126

Provided by European Molecular Biology Laboratory

Citation: A powerful HMMER for data mining (2015, June 3) retrieved 19 April 2024 from <https://phys.org/news/2015-06-powerful-hmmer.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.