

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

Exploring protein family relationships

ArticleInfo		
ArticleID	:	3920
ArticleDOI	:	10.1186/gb-2001-2-5-reports2003
ArticleCitationID	:	reports2003
ArticleSequenceNumber	:	17
ArticleCategory	:	Web report
ArticleFirstPage	:	1
ArticleLastPage	:	5
ArticleHistory	:	RegistrationDate : 2001-4-12 Received : 2001-4-12 OnlineDate : 2001-5-4
ArticleCopyright	:	BioMed Central Ltd2001
ArticleGrants	:	

David Chambers

Abstract

The Blocks server allows access to multiply aligned ungapped sequence segments corresponding to the most highly conserved regions of proteins, and aids detection and verification of protein sequence homologies using both DNA and protein as a starting material.

Content

The Blocks server allows access to multiply aligned ungapped sequence segments corresponding to the most highly conserved regions of proteins, and aids detection and verification of protein sequence homologies using both DNA and protein as a starting material. Blocks are generated automatically from most highly conserved regions in groups of proteins found in the [PROSITE](#) database of protein families and domains. In essence, the Blocks server represents a compilation of conserved protein domains, some with a known function but many more without. Some blocks define clear structural characteristics of a protein family (for example, a series of T-box-containing proteins) whereas others simply represent a series of conserved amino acid residues with no known function. Blocks searches differ in principle from BLAST searches, which are designed to detect end-to-end similarities or similarities across a broader sequence with individual submissions to the database.

Navigation

The home page is a straightforward, hyperlinked, entry point to the tools available. This is fine if you already have an idea of what you are doing or how you would like to do it, but it is somewhat sparse for the first-time user. Help is at hand, however - every analysis tool has an associated 'Help' or 'About' link. The background information on each of the Blocks-orientated tools is minimal but explanatory. It is assumed that the user already has basic experience of the mechanics of investigating molecular relationships. The Blocks server is also particularly well set up to provide an entry point from which other similar bioinformatics servers can be accessed. The facilities of the Blocks tools are already coordinated with those of other specialized protein analysis programs to ensure maximum coverage of databases and blocks. Thus the user can move very rapidly between sites dedicated to proteomics.

Reporter's comments

Timeliness

The present release is Blocks Database Version 12.0, June 2000, consisting of 4,071 blocks representing 998 groups documented in InterPro 1.0 keyed to SWISS-PROT 38 and TrEMBL.

Best feature

The Block Searcher compares a sequence of a newly identified DNA or protein against the current database of protein blocks. The advantage of searching a database of blocks is that information from multiply aligned sequences is present in a concentrated unbiased form, reducing background 'noise' and increasing sensitivity to distant relationships. Such a tool can give an indication of the evolutionary origins of a protein, and hence an indication of its function, without the query protein containing well-characterized functional motifs *per se*. In general, a group of related proteins have more than one region in common and their relationship is represented as a series of blocks separated by unaligned regions. If other blocks from a group also score highly in the search against the query sequence, this further reinforces the relationship of the query sequence to the proteins used to compile the block. As with most search and alignment tools the sensitivity can be altered to rule out chance alignments. This is a very powerful tool with which to attain the first clue to a protein's role.

Another very useful feature of the Blocks server is the 'codehop' (consensus-degenerate hybrid oligonucleotide primers) function (see related report - [Genome Biology 1\(1\):reports240](#)). This program allows retrieval of a protein block direct from the database (for example, using the keyword 'homeobox' from the GetBlocks function) followed by a user-driven automated process for the design of degenerate oligonucleotide primers. These primers are useful for both degenerate reverse-transcription PCR, and for other methods such as targeted differential display analysis.

Worst feature

The degree to which Blocks covers all conserved protein domains is entirely dependent on the algorithm used to generate the alignments; any tool of this type tends not to be exhaustive when used in isolation. This has been addressed by the authors of the site who have clearly laid out the limitations of the analysis as well as providing links to other sites dedicated to the same task.

Wish list

The output from a Block Search query is represented in a classic sequence-alignment format. It would very convenient if the same information were represented graphically so that the location of the identified block could be immediately placed on the query protein. Similarly, when homology to a block has been suggested, it would be invaluable if the locations of the block sequence in the proteins from which it was derived were displayed. This might give further clues to the function of that block in disparate family members.

Related websites

There are many websites aimed at recognizing local regions of similarity between the entire pool of protein sequences held on a given database, and the Blocks server will automatically screen through three major databases [Pfam](#), [ProDom](#) and [Domo](#), all of which are linked from the SearchBlocks page. Access to a fourth major site, the [PRINTS](#) database, is optional through the Blocks server and it is recommended that the two protein analysis tools are used independently. PRINTS is a collection of protein fingerprints, where a fingerprint is a group of conserved motifs used to characterize a protein family. This is a valuable addition as the identification of a protein fingerprint in a query sequence can be more informative than the discovery of an individual block. A wider selection of related sites are provided by the [ExPasy proteomics tools](#) listing.

Table of links

[Blocks WWW server](#)

[PROSITE](#)

*Genome **Biology** 1(1):reports240*

[Pfam](#)

[ProDom](#)

[Domo](#)

[PRINTS](#)

[ExPasy proteomics tools](#)

References

1. [Blocks WWW server.](#)