

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

Mining the chicken genome

ArticleInfo		
ArticleID	:	4341
ArticleDOI	:	10.1186/gb-2002-3-5-reports2013
ArticleCitationID	:	reports2013
ArticleSequenceNumber	:	23
ArticleCategory	:	Web report
ArticleFirstPage	:	1
ArticleLastPage	:	5
ArticleHistory	:	RegistrationDate : 2002-2-22 Received : 2002-2-22 OnlineDate : 2002-4-25
ArticleCopyright	:	BioMed Central Ltd2002
ArticleGrants	:	

David Chambers

Abstract

The chicken EST repository gives access to a BLAST server dedicated to mining recently released sequence information from the chick large-scale expressed sequence tag (EST) project

Content

The chicken EST repository gives access to a BLAST server dedicated to mining recently released sequence information from the chick large-scale expressed sequence tag (EST) project. The project, funded by the Biology and Biotechnology Science Research Council (BBSRC), provides a comprehensive resource of chicken ESTs for the scientific community. Such a facility is timely as the chick has lagged behind other model organisms such as mouse, rat and *Caenorhabditis elegans* with respect to tools for investigating the genome. Using a variety of technologies, a consortium of scientists has generated ESTs from 21 different embryonic and adult tissues, ranging from a complete early developmental stage (embryonic day (E2); Hamburger and Hamilton stage 10) to a single adult body part, such as the limb. The individual cDNA libraries constructed from each point were poly(T)-primed (that is, generated from the 3' end of an mRNA transcript), directionally cloned and sequenced from the 5' end. Preliminary analysis of the EST database reveals approximately 312,000 usable EST sequences with an average read length of approximately 700 base-pairs (bp), where the average cDNA insert size is 1.5 kilobases (kb). The sequences have already been passed through checks for vector-clipping and for contamination by, for example, ribosomal and bacterial sequences. Using homologies to known 5' ends, bioinformatic analysis at UMIST has shown that approximately 30% of the ESTs in each library represent putative full-length reads. The database therefore represents the most up-to-date and complete compilation of the chicken transcriptome, and an opportunity to access this information through a simple BLAST interface makes it an extremely valuable tool for biologists. If the proposed chicken genome project gets underway, the EST project will be an extremely valuable partner for gene identification and expression studies.

Navigation

Anyone familiar with the standard format of BLAST search engines (for example, at [NCBI BLAST](#) and [ExPASy](#)) will be immediately at home with the chick EST BLAST page. It is a simple search engine, constructed upon BLASTN, where query sequences are copied into the web page and then

compared directly with the sequence databases. A few options exist for more advanced similarity searching, such as alteration of filtering and expectation values, but no explanation of these appears at this site. However, a comprehensive BLAST user guide can be found at the [NCBI BLAST](#) website. As a default, the whole EST database will be scanned against the submitted sequence, but there is an option to search each developmental stage or tissue singly for an EST match.

Reporter's comments

Timeliness

The chick EST database server was first released on 14 December 2001 and is presently in a preliminary form.

Best feature

The comprehensive nature of the construction of the EST sequences is the foundation of this site. As the sequences have been generated from 21 different tissues across a range of developmental stages there is a strong probability that a large proportion of the chicken transcriptome is present. In addition, analysis of the sequence information has revealed a large amount of redundancy in ESTs between different stages and/or tissues. This allows a cDNA-walking strategy to be adopted for any given hit. For example, where a given EST match is detected for the query sequence, one can then use the most 5' sequence of that EST (approximately 50 bases or so) to re-screen the database and identify a different EST that is longer and contains further 5' sequence. Because of the number of cDNA libraries used and the 5' sequencing, this process can be reiterated several times to move progressively up into coding (and thus more informative) regions of transcripts. This effectively allows very rapid *in silico* cloning of chick cDNAs. Using this strategy and the chick EST database, it is possible to obtain more than 3 kb of contiguous sequence from short EST reads. Despite the large amount of sequence information previously available, this is the first time that such an approach has been easily and publicly possible for the chicken. This approach is also more feasible as a result of the availability of the identified cDNA clones from the Human Genome Mapping Project at low cost. Details of registration for this service and clone availability can be found at [ARK-Genomics](#) hosted by the Roslin Institute.

Worst feature

Because the sequence information derives from the 5' origin of the EST sequence, 3'-end sequences (adjacent to the poly(A) tail) of the mRNA transcripts are under- or un-represented (depending on the

length of cDNA insert compared to sequence read). While this does not necessarily affect large-scale expression analysis, many cDNA screening methodologies are 3' biased (for example, PCR-based subtractive hybridization and differential display reverse transcription PCR) and may be unlikely to produce sequence information sufficiently 5' to find a match with sequences deposited at the server. This is an unfortunate restriction on the usefulness of this resource for a relatively common means of scientific investigation.

Wish list

Completion of the 3' sequencing of the EST clones would greatly improve the usefulness of the database for some applications.

Related websites

There are numerous sites that operate BLAST servers to scan through sequence databases: the most comprehensive are probably those on the [NCBI BLAST](#) site and the website of the [European Bioinformatics Institute](#). These sites provide comprehensive outposts for accessing and similarity searching sequence databases, including sequences from many different organisms. For chick sequences, however, the UMIST server is unparalleled.

Table of links

[Chicken EST repository](#)

[NCBI BLAST](#)

[ExPASy](#)

[ARK-Genomics](#)

[European Bioinformatics Institute](#)

References

1. [Chicken EST repository](#).