Available online at www.sciencedirect.com
ScienceDirect
Experimental Parasitology 117 (2007) 229–235
Experimental Parasitology
www.elsevier.com/locate/yexpr

Minireview

The *Schistosoma mansoni* transcriptome: An update

Guilherme Oliveira *

Laboratory of Cellular and Molecular Parasitology, Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Av. Augusto de Lima 1715, Belo Horizonte, MG 30190-002, Brazil
Programa de Pós-Graduação e Pesquisa, Santa Casa de Belo Horizonte, Av. Francisco Sales 1111, 9° andar, Ala C, Belo Horizonte, MG 30150-221, Brazil

Received 1 February 2007; received in revised form 29 May 2007; accepted 1 June 2007
Available online 12 June 2007

Abstract

Large scale EST sequencing projects have been carried out for *Schistosoma mansoni* and *Schistosoma japonicum*. This update will briefly review the most recent accomplishments in the area and discuss the use of EST data for the purposes of gene discovery, gene model development, genome annotation and SNP analysis. In addition, the use of ESTs for studying other features of the transcriptome such as splice site and transcription initiation variants will be discussed as well as approaches to assigning function to unknown transcripts. Although EST sequencing has contributed much for schistosome research, other data mining possibilities exist, including the identification of putative drug and vaccine targets.

Index Descriptors and Abbreviations: Trematode; Schistosomiasis; Expressed sequence tags; Transcriptome; cDNA, complementary DNA; mRNA, messenger RNA; STS, sequence-tagged sites; ORESTES, Open Reading Frame Expressed Sequence Tags; RGMG, Minas Gerais Genome Network; CDD, Conserved Domain Database; KEGG, Kyoto Encyclopedia of Genes and Genomes; RNA, ribonucleic acid; DNA, deoxyribonucleic acid; SNP, single nucleotide polymorphism; MR4, Malaria Research and Reference Reagent Resource; SR3, Schistosome Related Reagent Repository; SAGE, Serial Analysis of Gene Expression

The transcriptome can be defined as a collection of the genes transcribed in an organism, tissue, or cell. Transcriptomic information has been extremely useful for gene discovery, the elaboration of gene models, training of gene finding algorithms (once a genome is available) the design of microarrays and has tremendously facilitated the cloning of genes of interest by the availability of cDNA clones. However, the mRNA transcribed is not expressed in equal amounts even in a single cell and they also differ in length. Therefore, experimental efforts towards obtaining the transcriptome of an organism generally will provide partial mRNA sequences and will be enriched for genes that are more highly transcribed. The partial cDNA sequences generated are called expressed sequence tags, ESTs. Despite some technical difficulties for producing full length cDNAs and incomplete views of the transcriptome provided, the approach has proven to be a powerful tool for the study of genes and their expression pattern.

The original and main motivation for the study of transcriptomes was the possibility of discovering the gene content of an organism and generating STSs that pointed to a gene, without the need to obtain the complete genome sequence (Adams et al., 1991). Although the unfinished genome sequences of *Schistosoma mansoni* and *Schistosoma japonicum* have been produced (El-Sayed et al., 2004; McManus et al., 2004), and are currently in the annotation stage, the transcriptome of both species have been studied on a large scale and have contributed to the discovery of genes and genome annotation, among other uses that will be further discussed in this update (genome sequencing will be discussed in this issue). Currently, Gen-

* Address: Laboratory of Cellular and Molecular Parasitology, Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Av. Augusto de Lima 1715, Belo Horizonte, MG 30190-002, Brazil. Fax: +55 31 3295 3115.
E-mail address: oliveira@cpqrr.fiocruz.br

0014-4894/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.exppara.2007.06.001

Bank contains information for over 650 species with at least 1000 ESTs (Table 1).

There are two main approaches for the large scale production of cDNA sequences. The conventional method is based on reverse-transcribing mRNA, cloning and sequencing the extremities of the inserted cDNA (Fulton et al., 1995). The alternative approach named ORESTES is to sequence short randomly amplified and cloned cDNAs (Dias et al., 2000) (Fig. 1). Both methods can potentially result in large numbers of redundant sequences of cDNAs. The sequences need to be assembled so that transcripts from one gene are grouped together. The assembly process yields, in many cases, full length virtual cDNA sequences. The methods are complimentary. The ORESTES tends to accumulate at the central region of the original mRNA and the conventional method at the 5′ and 3′ ends. Therefore, by combining both methods, the chance of assembling a full virtual cDNA sequence increases. The ORESTES method allows sampling of a larger number of small cDNA libraries, which is useful especially in cases of difficult to obtain mRNA. The method also normalizes the mini cDNA libraries, making the relative number of ESTs similar irrespective of the different gene transcription levels in the sampled tissue. Normalization permits the identification of rare transcripts. The conventional method, on the other hand, permits an in depth sampling of much larger cDNA libraries. In addition, the conventional method permits the estimation of the relative numbers of each transcript by not normalizing the libraries. Also importantly, the conventional cDNA sequencing method produces frozen stocks of clones, many of which are complete or near complete cDNAs. These cDNA clones are an important resource to the research community.

The study of the transcriptome can provide many different types of information about an organism. The main type of information generated is the discovery of new genes which can be accomplished efficiently and quickly with a transcriptomic approach, especially for organisms with large genomes, such as schistosomes. In the pre-genome era of schistosome research, this approach was used and yielded a

Table 1
Online resources related to *Schistosoma* ESTs available

Resource	Web site	Reference
ORESTES and annotation	http://verjo18.iq.usp.br/schisto/	Verjovski-Almeida et al. (2003)
TIGR gene index	http://compbio.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=s_mansoni	Merrick et al. (2003)
ESTs and KEGG pathways	http://rgmg.cpqrr.fiocruz.br/	Oliveira et al., unpublished
EST of *S. mansoni*, *S. japonicum* and *S. haematobium*	www.sanger.ac.uk/Projects/S_mansoni/	Wellcome Trust Sanger Institute, Unpublished
S. mansoni SNPs	http://bioinfo.cpqrr.fiocruz.br/snp/	Simões et al. (2007)
Schistosome Related Reagent Repository—SR3	www.afbr-bri.com/SR3/	Unpublished
Genome sequence of *S. mansoni*	www.genedb.org/genedb/smansoni/	Unpublished
Genome database for *S. mansoni*	www.shistodb.net	Unpublished, under construction
Genome sequence of *S. japonicum*	http://lifecenter.sgst.cn/sj.do#	Unpublished
CDD database	http://130.14.29.110/Structure/cdd/cdd.shtml	Marchler-Bauer et al. (2007)
Pfam database	www.sanger.ac.uk/Software/Pfam/	Finn et al. (2006)
PIR and UNIPROT	http://pir.georgetown.edu/	Wu et al. (2003, 2006a)
Swiss-Prot	www.expasy.org/sprot/	Wu et al. (2006a)
InterPro	www.ebi.ac.uk/interpro/	Mulder et al. (2002)
KEGG	www.genome.jp/kegg/	Kanehisa et al. (2004)

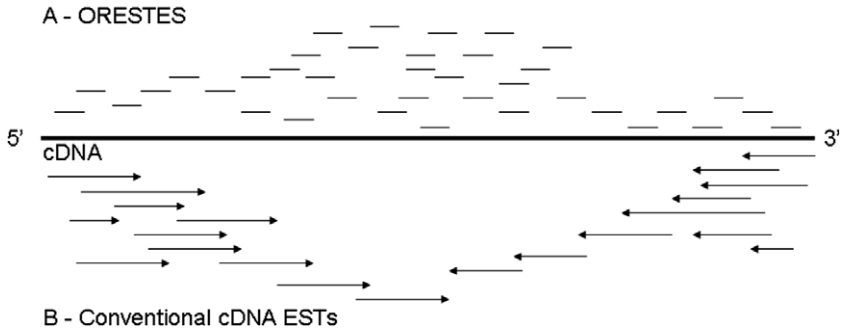

Fig. 1. Methods for producing expressed sequence tags (ESTs). The complete cDNA clone is represented by the central line. (A) ORESTES. (B) Conventional cDNA sequencing. Conventional cDNAs are usually 100–500 bp sequences produced from cDNA cloned and sequenced directionally in respect to the 5′or 3′ end, and tend to accumulate at the end of the representing mRNA. On the other hand, ORESTES are shorter sequences that tend to accumulate on the central region of the representing mRNA.