News

The Filarial Genome Project

T.R. Unnasch

Beverly, USA February – March 1994

Recently, the Strategic Research Branch of the Special Programme for Research and Training in Tropical Diseases (TDR) of the World Health Organization (WHO) has identified the study of the genomic structure of parasites as a research goal. In keeping with this goal, TDR sponsored a meeting to discuss the study of the genome of the filarial parasites of humans. The purpose of the meeting was to develop a framework for the study of the genome of filarial parasites. It was expected that the participants would identify a single species to be examined, and that they would develop a specific approach to the study of the parasite's genome. It was also necessary to develop a plan for collating and disseminating data and reagents produced by individuals involved in the project, and to ensure that investigators in endemic countries would be involved in the project.

Aims and Approaches

The program of the meeting was designed to assist the participants in developing a framework for the filarial genome project. Initially, the strategies that may be used in a genome mapping initiative were summarized by B. Slatko (New England Biolabs, Beverly, USA). In the 'top down' approach, a coarse genetic or physical map of the genome is first generated. Individual DNA sequences are then mapped with respect to one another, and to the previously determined map. In the 'bottom up' approach, individual clones are isolated and characterized. Attempts are then made to derive a linkage map by identifying clones that share overlapping sequences. These strategies have often been combined, as has been the case with the *Caenorhabditis elegans* genome-sequencing initiative.

A second series of presentations dealt with the advantages and limitations of undertaking a genome project on each of the filarial parasites of humans, ie. *Onchocerca volvulus*, *Wuchereria bancrofti* and *Brugia malayi*. Factors considered included the current state of knowledge about each parasite's genome, the availability of material from each organism, and the relative public health importance of each of the parasites. It was noted that molecular phylogenetic studies have shown that divergence in the nematode lineage is an ancient phenomenon. Thus, any attempt to extrapolate findings from free-living nematodes, such as *C. elegans*, to filarial parasites should be done with the understanding that the organisms may not be closely related evolutionarily. Despite this caveat, it is clear that parasitic nematodes share many common features with *C. elegans*. Thus, a filarial parasite genome project should attempt to utilize the *C. elegans* resource to the greatest extent possible.

A third series of presentations provided an update on *Plasmodium falciparum* and *Schistosoma mansoni* genome projects, which are already underway. The malaria project was summarized by A. Craig (John Radcliffe Hospital, Oxford, UK). The *P. falciparum* genome is being mapped, utilizing a library of yeast artificial chromosomes (YACs). The YACs are being connected by saturation mapping and hybridization, using sequence tags developed from previously characterized YACs. The data developed from this project are centralized and are being made available to all interested investigators. The *S. mansoni* project was summarized by M. Adams (Institute for Genomic Research, Gaithersburg, USA). In contrast to the malaria project, the *S. mansoni* genome project is taking a strictly 'bottom up' approach of developing a survey of expressed sequence tags (ESTs). To accomplish this, randomly selected cDNA clones present in a cDNA library prepared from adult mixed-sex mRNA are being sequenced. This approach has the advantage of providing a rapid payoff in the production of new data, but the level of redundancy in the sequencing will rise as the project progresses.

Caenorhabditis elegans Database

A major concern of any genome project involves the manner in which the information collected will be organized and disseminated to interested investigators. To this end, A. Coulson (Sanger Centre, Cambridge, UK) presented a hands-on demonstration of a *C. elegans* database (ACeDB), the database that is central to the *C. elegans* genome project. He emphasized that ACeDB has been developed with an open architecture. This makes it possible to incorporate data produced by other genome projects into ACeDB, or into a separate but related database.

Recommendations

Based upon the presentations given at the meeting, the participants agreed that *B. malayi* should be the organism of choice for a filarial genome project. *Brugia malayi* has several advantages over *W. bancrofti* and *O. volvulus*, the most important being that parasite material from all life cycle stages is more readily available from *B. malayi* than the other human filaria, and that cDNA libraries from most stages of the parasite life cycle have been produced. Because of the difficulty in generating a genetic or physical map for the filarial parasites, it was the consensus that a 'bottom up' approach would be the most practical. It was thought that the best strategy would be to follow the example of the *S. mansoni* project and embark on a study of ESTs in *B. malayi*. This approach should produce a high initial payoff in terms of data production. Furthermore, by including cDNAs derived from all the life cycle stages, information about develop mental expression of various genes would also be obtained.

Much of the DNA sequencing necessary for the project might be undertaken by laboratories located in endemic countries. The sequencing effort could be spread over a number of laboratories worldwide, by preparing grids from each cDNA library, and by providing different sets of grids to each laboratory. It was recommended that a high priority be placed on the establishment of a central facility, which would be responsible for preparing and maintaining the grids, and for collating

the data produced by the individual laboratories. To ensure that the data produced by the project were widely available, it was felt that ACeDB should be exploited, by distributing the filarial data either as part of ACeDB, or as a supplement to it. Furthermore, it was felt that distribution of the information should not be a passive process, and that the existence of the database and its availability should be publicized as widely as possible.

Acknowledgements
A meeting to develop a framework for the WHO/TDR Filarial genome project was sponsored by WHO/TDR. The meeting was held at New England Biolabs, Beverly, MA, from 27 February – 1 March 1994. The local organizing committee for the meeting consisted of Larry McReynolds, Barton Slatko and Claude Maina of New England Biolabs. The meeting chair was Steven Williams of Smith College, Northampton, MA, USA.

Thomas R. Unnasch is at the Division of Geographic Medicine, BBRB 206, The University of Alabama at Birmingham, Birmingham, AL 35294, USA.

Comment

Molecular Biology of Natural Resistance-associated Macrophage Protein

J. Ivanyi

Genetic factors have been the focus of current research in many common diseases caused by infections, autoimmunity or cancer. Unlike the rare metabolic disorders which can be attributed to single gene defects and are already targeted for genetic therapies, it has been more difficult to advance with new insights on diseases which are under complex, multigenic control. In the case of infections, lymphocyte-mediated immune responses are controlled via antigen presentation by major histocompatibility complex (MHC) gene products. While the role of non-MHC genes has been generally acknowledged, their identification and characterization are less advanced. This gap has narrowed by the recent cloning and sequencing study of the *Ity/Lsh/Bcg* gene[1].The three designations reflect the original demonstration of genetic influence in murine infections of such diverse nature as salmonellosis, leishmaniasis and mycobacteriosis. These are all intracellular infections and the gene is acting on the macrophage, which is the site of microbial replication. The product of the newly cloned candidate gene has been designated as 'Natural resistance-associated macrophage protein' (Nramp).

The fundamental feature of the *Nramp* gene action is the pleiotropic and conditional phenotypic expression. The extent of genetic influence varies from a decision over survival following *Salmonella typhimurium* infection, to influence merely over the early proliferation rate of *Mycobacterium bovis* [strain BCG (bacillus Calmette-Guérin)] after intravenous infection. However, even that effect is not apparent following aerosol infection with pathogenic *M. tuberculosis*[2]. The pleiotropic gene effects have been monitored on the basis of: (1) bacterial growth in explanted macrophages; (2) granuloma formation; (3) MHC class II expression and antigen-presenting efficacy; (4) production of reactive oxygen and nitrogen metabolites; (5) suppressor macrophage activities; (6) respiratory burst and hexose monophosphate shunt; or (7) interleukin 1 production[3–5]. Macrophages from innately resistant mice respond more vigorously, but paradoxical results arise due to the overriding effect of the bacterial load in susceptible strains in respect of both macrophage activation and immune responsiveness.

Localization of the *Nramp* gene on mouse chromosome 1 has been advanced using positional cloning analysis of a high-resolution linkage map of the proximal chromosomal segment[1]. Following exon amplification of a cloned domain, the critical option for one transcription unit (proximal to *Vil* and distal to λMmIC165 loci) as the *Nramp* candidate gene was based on its restricted expression in macrophages. Most strikingly, transition of a single nucleotide G → A at position 783 of the *Nramp* mRNA transcript, leading to the substitution Gly105 → Asp in that protein, was found to differentiate between several resistant and susceptible strains of mice. Further evidence of the Nramp function will require experiments involving gene transfer and gene disruption.

On the basis of amino acid sequence, the Nramp protein was found to be homologous to membrane-associated transporter molecules. It was predicted to contain several membrane-spanning helical domains, with conserved charged and hydrophilic residues which occur rarely in integral membrane proteins. Furthermore, ATP energy-dependent transport function was proposed on the basis of a structural motif. Vidal *et al.* envisage that the allelic difference represented by the change of glycine by the bulkier Asp residue within the second transmembrane helix could influence the transport function within the cell membrane lipid bilayer. The authors apparently favour the opinion that Nramp facilitates the transport of charged reactive nitrogen compounds. Their view is based on the structural homology of Nramp with CrnA, which is responsible for the membrane transport of nitrate in *Aspergillus nidulans*. This mechanism is attractive, since it could account for differences in the cytocidal activity against intracellular microbial pathogens, mediated by nitrogen intermediates within the phagolysosomes of macrophages[5]. However, the possibility of transporter function of other or multiple substrates remains open and would fit well with the pleiotropic nature of differences between the resistant and susceptible Nramp phenotypes.

The role of genetic factors in tuberculosis (TB) has been indicated from the finding of individual and racial differences in resistance during epidemics or in disease-endemic areas[6]. Genetic factors have been suggested also from concordance analysis in mono- and dizygotic twins and from segregation and pedigree analyses in multigeneration families. The nature of genes is not known, with the exception of human leukocyte antigen (HLA) DR2, which has been associated with the development of multibacillary forms