

The *Ixodes scapularis* Genome Project: an opportunity for advancing tick research

Catherine A. Hill¹ and Stephen K. Wikel²

¹Department of Entomology, Purdue University, West Lafayette, IN 47907, USA

²Center for Microbial Pathogenesis, School of Medicine, University of Connecticut Health Center, 263 Farmington Avenue, MC3710, Farmington, CT 06030, USA

The *Ixodes scapularis* Genome Project (IGP), the first to sequence a tick genome, will provide an unparalleled resource for studying tick biology and tick–host–pathogen relationships, and identifying novel targets for tick and tick-borne disease control. The IGP will be the first genomic analysis of a member of the subphylum Chelicerata and will accelerate the pace of tick research. The challenge for scientists is to translate IGP data into public health benefits.

Tick research enters the genomic era

Genomics research is revolutionizing biology and the understanding of a diverse range of organisms. Genomics has become a central and cohesive discipline of biomedical research [1]. Genomics approaches are increasingly being applied to study bloodfeeding arthropods – an important and diverse group of insects and arachnids – because of their importance as vectors of infectious agents and the need to develop novel methods for both suppressing vectors and blocking pathogen transmission.

The National Institute of Allergy and Infectious Diseases (NIAID) has made a significant commitment to sequencing the genomes of (i) microorganisms considered to be potential agents of bioterrorism; (ii) emerging and re-emerging infectious diseases; and (iii) the arthropod vectors of human disease agents. The Microbial Sequencing Centers (MSCs) were established recently by the NIAID as contract sequencing centers for providing high-throughput cDNA and genomic sequencing, genome assembly and automated genome annotation [Box 1(i)]. MSC-supported projects are required to provide rapid, unrestricted access to sequence data through public databases.

The need for a tick genome project to advance tick and tick-borne disease research has been recognized for many years. In 2004, an international consortium of tick researchers initiated the first effort to sequence a medically relevant tick. A white-paper proposal to sequence the tick *Ixodes scapularis* was submitted to the NIAID [Box 1(ii)], and the project was approved in April 2004. The *Ixodes scapularis* Genome Project (IGP) is a collaborative effort of the international tick-research community, the National Institutes of Health [Box 1(iii)] and

the MSCs. This initiative represents a new era in tick and tick-borne disease research.

Rationale for the IGP

Ixodid ticks (subphylum Chelicerata, class Arachnida) transmit the most diverse array of infectious agents of medical and veterinary importance of any bloodfeeding arthropods; they are second only to mosquitoes as vectors of human pathogens and they are the most important vectors of infectious agents of other animal species; they have a global distribution and they infest a wide range of vertebrate hosts [2–4]. In the USA, *I. scapularis* transmits the causative agents of Lyme disease, babesiosis, human granulocytic anaplasmosis and, possibly, the flavivirus agent of Powassan encephalitis. Despite its medical importance, there are major gaps in the knowledge of ixodid tick biology. Methods to control ticks and the diseases they transmit are currently limited. The IGP offers new ways to confront these challenges in disease control and to enhance the understanding of tick biology, particularly in areas not previously amenable to study.

The IGP builds upon an existing wealth of knowledge [5–7], expertise and ongoing tick research. As a community effort, analysis of *I. scapularis* genome data will be conducted by researchers studying aspects of tick biology, tick–host–pathogen interactions, and their control and evolutionary biology. The IGP will provide the first genomic overview of the taxonomically diverse subphylum Chelicerata and, thus, will greatly expand the scope of comparative and evolutionary eukaryotic analyses. Ticks are evolutionarily distant from insects, with the origin of the family Ixodidae estimated to be 120 million years ago [8]. The *I. scapularis* genome is predicted to be unique

Box 1. Website links

- (i) NIAID Microbial Sequencing Centers: <http://www.niaid.nih.gov/dmid/genomes/mscs/>
- (ii) IGP White Paper: <http://www.entm.purdue.edu/igp/overview.html>
- (iii) NIH: <http://www.nih.gov>
- (iv) Wikel Laboratory: http://cmp.uchc.edu/Wikel_Lab/wikel_lab.html
- (v) Genome Size: <http://www.genomesize.com/>
- (vi) Human Genome Project: http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml
- (vii) NCBI GenBank: <http://www.ncbi.nlm.nih.gov/Genbank>
- (viii) VectorBase: <http://www.vectorbase.org/>
- (ix) IGP: <http://www.entm.purdue.edu/igp/default.html>

Corresponding author: Wikel, S.K. (swikel@up.uchc.edu).

compared with the other sequenced invertebrate genomes such as those of *Drosophila melanogaster* and *Anopheles gambiae*, the mosquito vector of the malaria parasite. Gene prediction among such distantly related organisms will be challenging [9]; the IGP will require extensive expressed sequence tag (EST) sequencing and manual annotations, and will rely heavily on the expertise of the tick-research community. Genome data will provide information about many aspects of tick biology and will enable the identification of unique tick genes and physiological processes that could be exploited for tick control and, thus, the control of tick-borne disease. Genetic information will facilitate studies of tick phylogenetics, population biology, ecology and behavior: research areas that have been hindered by a lack of molecular data. The IGP will complement other vector genome projects such as the completed *An. gambiae* and the well-advanced *Aedes aegypti* (yellow-fever mosquito) projects, the ongoing *Culex pipiens* (West Nile virus mosquito) project and the tsetse-fly initiative [10], and will provide a powerful catalyst for genomic studies of other invertebrates.

An overview of the IGP

The goal of the IGP is to undertake whole-genome random shotgun sequencing (WGS) to a draft level of sixfold coverage, which is sufficient to assemble and annotate large sections of the genome. The *I. scapularis* colony maintained by S.K. Wikel at the University of Connecticut Health Center [Box 1(iv)] has been selected for the IGP. This colony, established in 1996 using ticks collected from Lyme disease endemic areas, has been continuously inbred and used extensively for experimental purposes. Little is known about the *I. scapularis* genome, which seems to be large (~2 Gbp) compared with sequenced insect genomes [11] (C.A. Hill *et al.*, unpublished); for example, the genome of *An. gambiae* is estimated to be

278 Mbp [12]. Genome sizes can be obtained at the website listed in [Box 1(v)]. The IGP will be conducted in two phases (Figure 1). The results of phase I will be essential for determining the feasibility of phase II and for guiding phase II studies. Like all genome projects, the IGP will function as an independent research endeavor and will pave the way for other large invertebrate genome projects.

Phase I activities are in progress and involve community-based development of several project resources. Extensive sequencing of normalized cDNA libraries composed of various *I. scapularis* tissues and developmental stages will be conducted to generate EST data for the research community from ~100 000 clones. EST sequences will also be crucial for annotating the genome in phase II. Approximately 40 randomly chosen large-insert bacterial artificial chromosome (BAC) clones containing *I. scapularis* genomic DNA (~110–120-Kbp average length) will be completely sequenced end to end. BACs generate raw sequence data for the community and will provide a detailed preliminary analysis of regions of the genome. BAC clone ends will also be sequenced to provide 'sequence-tagged sites' for genome assembly after WGS. BACs are also useful for linking assembled genome sequences to chromosomes [13] by physical mapping techniques such as fluorescent *in situ* hybridization (FISH). Phase I will reveal information about *I. scapularis* genes, gene density, repetitive elements, AT content and level of polymorphism in the sequenced colony.

In phase II, WGS to a draft level of coverage is proposed. The aim is to produce a highly accurate, ordered sequence that spans the euchromatic *Ixodes* genome (Figure 1). WGS involves randomly breaking the genome into segments of various sizes and cloning these fragments into plasmid and, possibly, fosmid vectors. Small-, medium- and large-insert clones are then sequenced in both directions to provide shotgun-sequence information

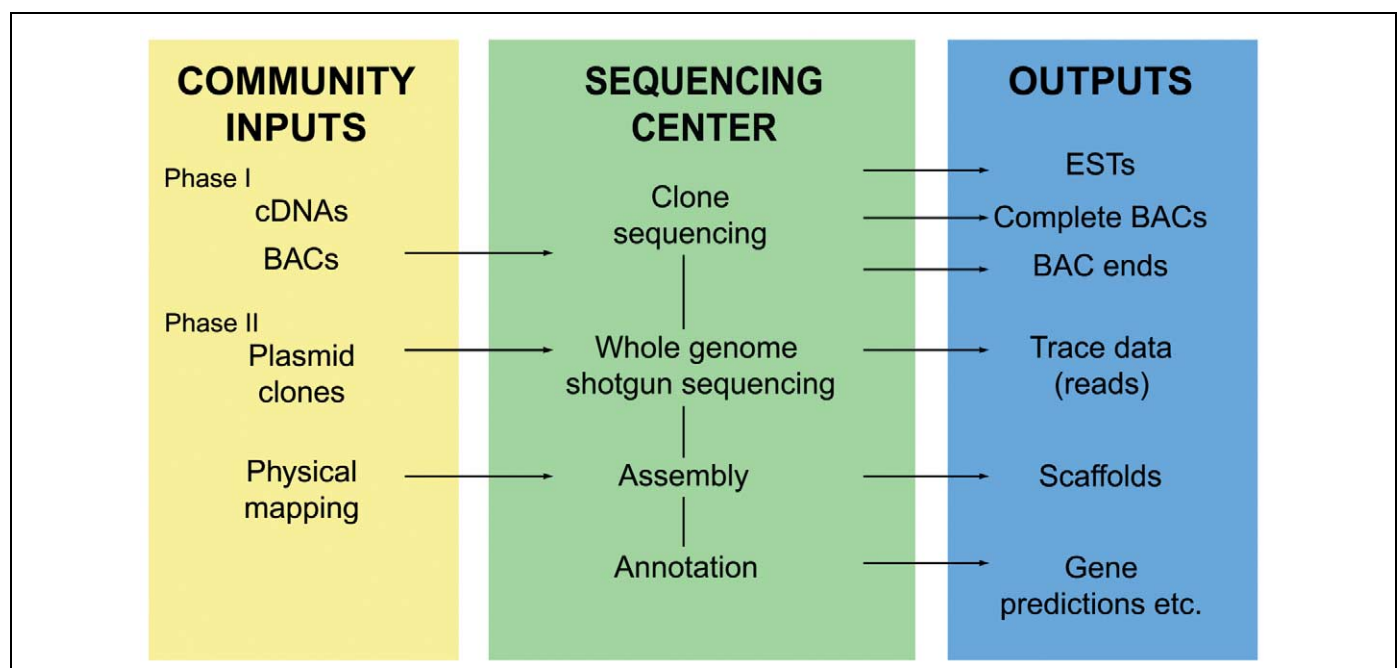


Figure 1. Overview of the IGP. The activities conducted by the tick-research community and the MSC in phases I and II of the project, and the expected project inputs and outputs are summarized.

(trace reads). These sequences also function as 'mate pairs' and are used in the assembly of the genome [14]. Consensus sequences (contigs) are then built from overlapping DNA segments, assembled into larger segments called 'scaffolds' using BAC-end sequence data and, subsequently, assigned or physically mapped to regions of the genome. Finished genomes such as that of the publicly funded Human Genome Project [Box 1(vi)] are highly desirable but are cost prohibitive and not justified for an organism such as *I. scapularis*. Draft genome coverage is expected to be sufficient to identify genes and elements of interest from *I. scapularis*. *Ixodes* scaffolds will contain gaps representing lack of clone coverage or sequencing difficulties. After assembly, genes will be predicted using a combination of existing EST data and automated *ab initio* (in the absence of biological data) gene predictions. Research community expertise will be crucial for confirming and refining gene predictions (also known as 'third-party annotations') to resolve genome regions and to close gaps through dedicated BAC sequencing.

All sequence data will be made available in a timely manner through regular downloads to public databases such as the NCBI GenBank database [Box 1(vii)]. An important new resource on the horizon for the research community is VectorBase [Box 1(viii)]. VectorBase is a relational database for invertebrate vectors of human pathogens that will provide access to vector genome data, and associated data types and bioinformatics analysis tools, including data generated by the IGP.

A community effort and new frontiers in tick research

Genome projects are inherently community-based efforts driven by a team of researchers. An important community role is the analysis of the massive quantities of various forms of genomic sequence data (Figure 1). The tick-research community is now challenged to develop new tools and techniques to exploit genomic data. This will include the production of microarrays, additional gene-knockout technologies [15,16], dedicated bioinformatics analysis tools and databases. Typically, these resources are not part of a core genome project.

As data filter through the sequencing pipeline, the tick-research community will be responsible for the use of the *Ixodes* genome resource to address the most important and pressing issues in tick research. *I. scapularis* genome data will be used to unravel the complex genetic basis of tick-host-pathogen relationships, disease transmission and vector competence. The IGP will advance the understanding of tick physiological processes such as blood-meal digestion and osmoregulation, and tick behaviors such as host seeking and mating that could be exploited for tick control. The genome data will facilitate the extension of knowledge of novel biopharmaceuticals in tick saliva. In essence, the data will result in the identification of new molecules.

A quantum leap in research

Although the IGP might provide the impetus for a quantum leap in research, it is ultimately the researchers themselves that dictate the nature of that leap. The tick-research community must be prepared to use the vast amount of data that will be generated. Genomic data will impact all areas of tick-related research. However, it is the unexpected outcomes of the genome project that could provide the most exciting prospects for future research and generate the necessary advances for understanding these important vectors of disease at a new level and at an accelerated pace {further information about the IGP and project updates can be found at the IGP website [Box 1(ix)]}.

Acknowledgements

We thank Kenneth Bourell for his assistance with the figure. The research of S.K.W. is supported, in part, by National Institutes of Health grant R01-AI46676, and the United States Army Medical Research and Materiel Command. C.A.H. is supported by an Indiana 21st Century Research and Technology Fund and by a USDA-ARS Cooperative Research Agreement 58-6205-4-008.

References

- Collins, F.S. *et al.* (2003) A vision for the future of genomics research. *Nature* 422, 835-847
- Balashov, Y.S. (1972) Bloodsucking ticks (Ixodidae) – vectors of disease of man and animals. *Misc. Pub. Entomol. Soc. Am.* 8, 161-376
- Fivaz, B. *et al.*, eds (1992). *Tick Vector Biology: Medical and Veterinary Aspects*, Springer-Verlag
- Sonenshine, D.E., ed. (1991) *Biology of Ticks* (Vol. 1), Oxford University Press
- Chen, C. *et al.* (1994) Cytogenetic characteristics of cell lines from *Ixodes scapularis* (Acari: Ixodidae). *J. Med. Entomol.* 31, 425-434
- Oliver, J.H., Jr. *et al.* (1993) Conspecificity of the ticks *Ixodes scapularis* and *I. dammini* (Acari: Ixodidae). *J. Med. Entomol.* 30, 54-56
- Ullmann, A.J. *et al.* (2002) A preliminary linkage map of the tick, *Ixodes scapularis*. *Exp. Appl. Acarol.* 28, 107-126
- Klompen, J.S.H. *et al.* (1996) Evolution of ticks. *Annu. Rev. Entomol.* 41, 141-161
- Mongin, E. *et al.* (2004) The *Anopheles gambiae* genome: an update. *Trends Parasitol.* 20, 49-52
- Butler, D. (2004) African labs win major role in tsetse-fly genome project. *Nature* 427, 384
- Ullmann, A.J. *et al.* Genome size and organization in the black-legged tick, *Ixodes scapularis* and the southern cattle tick, *Boophilus microplus*. *Insect Mol. Biol.* (in press)
- Holt, R.A. *et al.* (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298, 129-149
- Hong, Y.S. *et al.* (2003) Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito, *Anopheles gambiae*. *Mol. Genet. Genomics* 268, 720-728
- Venter, J.C. *et al.* (1998) Shotgun sequencing of the human genome. *Science* 280, 1540-1542
- Aljamali, M.N. *et al.* (2002) RNA interference: applicability in tick research. *Exp. Appl. Acarol.* 28, 89-96
- Narasimhan, S. *et al.* (2004) Disruption of *Ixodes scapularis* anticoagulation by using RNA interference. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1141-1146