The Tribolium Genome Project: An International Collaboration

Susan Brown
Professor of Biology
Kansas State University

Global collaborations come in all shapes and sizes. Some are mandated federally, others as new initiatives of institutional administration, and still other by researchers linked by a common need or interest. Large-scale projects are the mainstream in high-energy physics, but are not so common in the field of biology. The large data sets produced by genome sequencing projects, which requires many different types of expertise for comprehensive analysis, have spurred the formation of global collaborations that are highly interdisciplinary. The Tribolium Genome sequencing consortium is an example of such a collaboration. As large scale data analysis enters the mainstream in the biological sciences, more such global, interdisciplinary groups will form and strengthen all the institutions involve. These international collaborations also strengthen the interactions between our regional institutions, and raise all involved to a new level of competitiveness on a global scale.

Introduction

The genomic sequence of a eukaryotic organism provides a wealth of information, but, to be useful, the sequence must be annotated with additional information such as the location of genes, and chromosomal landmarks. In today's research world, it takes an international consortia of scientists to organize their efforts: first to justify a genome sequencing project, and then to coordinate the annotation efforts once the sequence is in hand. Computational and manual annotations are combined to provide an initial analysis of the genome, which is released to the public in several forms: published reports, specialized databases and websites, and national databases. In the following narrative, I describe the efforts of the International Tribolium Genome Sequencing consortium, from white paper to publication in the journal, *Nature*, to sequence, assembly and analyze the genome of the red flour beetle, an insect model for developmental genetics and pest biology. Interactions between consortium members have lead to several federally and internationally supported projects, some of which continue today, past the formal conclusion of the genome sequencing project.

Why sequence the genome of a flour beetle?

With the completion of the first draft of the human genome in 2001, the

National Human Genome Research Institute (NHGRI) considered white paper requests to sequence additional genomes that would provide insight into the function and evolution of the human genome. The relatively small genome of the fruit fly, *Drosophila melanogaster* had already been sequenced as a proof-of-concept for the whole genome shotgun approach to genome sequencing. We proposed sequencing the genome of the red flour beetle *Tribolium castaneum*, a world-wide pest of stored grains. Sequencing of the honey bee and the silkworm moth genomes were already underway; with the addition of Tribolium, we would have representative genomes from the four largest orders of holometabolous insects, those that develop as worm-like larva and pupate into winged adults, both of which can be agricultural pests and/or beneficials. Over the past two decades, we have developed several molecular and genomic tools for Tribolium including balancer chromosomes and genetic maps, as well as transformation and RNAi methodologies. As a result, the red flour beetle is now the third best invertebrate model organism for genetic studies of development, physiology and toxicology after Drosophila and the free-living nematode, *C. elegans*. In addition, Tribolium is the first mandibulate insect (a chewing rather than sucking insect) recommended for genome sequencing. Furthermore, sequencing the Tribolium genome provides our first insight into a Coleopteran genome, and there are more species in this order than in any other.

Several research groups, predominantly in the US and Europe, use Tribolium as a model system in which to study the genetic regulation of development; Evo-Devo studies. Our understanding of insect development is largely based on genetic studies conducted in fruit flies. However, their development has become highly specialized as they adapted to the specialized niche of rotting fruit. The red flour beetle is also specialized in its own right, but displays many traits shared by lower insects and other arthropods. It is the supposition of Evo-Devo researchers that these traits as well as their genetic regulation are likely to be ancestral features. For example, the fly larva hatches from the egg as a headless, limbless maggot, while the beetle larva emerges with a true head equipped with eating appendages adapted for chewing and antennae and a thorax equipped with three pair of walking limbs. Development of fruit fly body plan is quite specialized; all the segments of the body are produced simultaneously by a hierarchy of regulatory gene interactions. In most other insects and arthropods, segments are added on at a time at the posterior end of the embryos, more like somite development in the vertebrate backbone. Analysis of the Tribolium genome was expected to provide insight into developmental studies in both fruit flies and vertebrates.

Genome sequencing projects require funding from multiple sources

Academic, industrial and federal agencies contributed to the Tribolium Genome project. Our rationale for sequencing the genome was explained in a white paper to the NHGRI, which included letters of support from across the breadth of the scientific community.

The white paper was formally approved by the NHGRI in Sept 2003. Soon thereafter, we started working with the Human Genome Sequencing Center at Baylor College of Medicine to generate the sequence data and assemble it. With the completion of the human genome sequence, this center has focused, in part, on insect genomes including two other Drosophilid species and Honey bees. The USDA provided funds to jump-start the sequencing efforts. As part of the Tree of Life project, the National Science Foundation supports distribution of Tribolium BAC library (Bacterial Artificial Chromosomes contain large fragments, ~350 kb, of genomic DNA), which was constructed by Exelixis, an integrated drug discovery and development company in South San Francisco and is currently archived at the Clemson University Genomics Institute. The Kansas INBRE and the KSU plant Biotech Center supported our efforts to obtain Expression Tagged Sequences (ESTs) and the KSU Arthropod Genomics Center supports Beetlebase, the community resource for information about the Tribolium genome.

We supplied a few milligrams of beetle eggs, and after several small sequencing runs to verify sample quality, the HGSC at Baylor required less than one month in the Fall of 2004 to deposit 1.8 Gb of Tribolium genomic sequence in the Trace Archives at the National Center for Biomedical Information (NCBI). As with every new genome sequencing project, the raw, unassembled sequences provided researchers around the globe with a rich resource from which to piece together specific genes of interest to them. However, once the raw reads were assembled into contiguous sequences representing large regions of the genome and organized into scaffolds representing the chromosomes, (which required several months) it was time to annotate the genome, associating gene structure and function with different regions of the genome.

It takes a global village to annotate a genome

Computational analysis of the genome revealed more than 16,000 gene models. A subset of these needed to be manually evaluated to determine the quality of the genome sequence and the value of the computer generated gene models. More than 100 scientists from 67 institutions world-wide provided the initial analysis of the Tribolium genome. Some of the scientists in the group who used Tribolium as a model system analyzed genes or pathways directly relevant to their research. Others, interested in genome architecture or gene families and gene evolution, were delighted have another dataset to complement their previous work and joined the foray. Manual curation efforts were largely voluntary as federal agencies have chosen (wisely) not to fund additional genome annotation projects beyond those already established for the most important model organisms. The enthusiasm of the group waxed and waned as we discovered just how tedious genome annotation can be. However, excitement was maintained throughout by weekly conference calls during which one group would report its progress to the others. The difficulty of scheduling conference

calls that span the globe was surmounted by holding the calls on Wednesday mornings at 9 AM. This turned out to be a convenient time during morning coffee break for those of us in the Midwest and late afternoon tea for our European counterparts, but colleagues in Japan, India and Australia had to forego a good night's sleep to join in, and even our California colleagues had to wake with the dawn to participate.

The final report, published in the April 24 issue of the journal *Nature*, was truly a collaborative effort. Most of the detailed information in the first draft was relegated to more than 100 pages of supplementary data as we were required to restrict the paper to briefly describing the highlights. More than 25 companion papers were written and complete issues of two different journals were dedicated to our description of the Tribolium genome. There were many surprises to be found in the genome sequence. For example, comparing a large set of conserved proteins resulted in a new phylogenetic tree, placing honey bees, instead of Tribolium, at the base of the holometabolous insects. Second, several genes, conserved in beetles and vertebrates, but not found in fruit flies, were added to the list ancestral genes. We also found that beetles contain more p450 detoxifying proteins and orderant receptors than any most other insect analyzed to date. These finding have brought up new perplexing questions, such as trying to imagine why a beetle that prefers caches of grain stored in dark, dry environments, needs a more finely-tuned sense of "smell" than a honey bee foraging in a meadow.

The future of genome sequencing projects

The first wave of genome projects was federally funded and their progress was followed in detail by the entire research community, as befitting a new research paradigm. The second wave of projects was also justified by white papers. With the advent of new sequencing technology that greatly reduces the price of obtaining the sequence data, genome sequencing projects are now in the realm of individual research grants. Soon a genome sequence may be considered preliminary data for a research project grant, and some even speculate that in the not too distant future, it may rest within the purview of a Master's level research project. Even when it reaches this stage, sequencing the average eukaryotic genome will be an international collaboration, uniting researchers world-wide, through their interest in the next genome.