C. elegans: Sequence to Biology

This special issue of *Science* celebrates a landmark in biology: determination of the essentially complete DNA sequence of an animal genome. The animal is a small invertebrate, the nematode (or roundworm) *Caenorhabditis elegans*, and the sequence consists of about 97 million base pairs of DNA, approximately one-thirtieth the number in the human genome. Nonetheless, the information content is enormous—eight times that of the budding yeast *Saccharomyces cerevisiae*, the only other eukaryote with a sequenced genome.

The *C. elegans* sequence presents scientists with a treasure trove of information and an extremely powerful analytical tool. *Caenorhabditis elegans* is known for its advanced genetic information and complete descriptions of its cellular anatomy, cell lineage, and neuronal wiring diagram. Furthermore, recently developed methods for studying the effects of inhibiting any endogenous gene and for isolating deletions in any gene will greatly facilitate sequence-based gene analysis. The combination of the complete genome sequence with these and other methods will allow investigation of a host of problems at new levels of sophistication.

The history, technology, and culmination of the sequencing project are summarized in a Review by the *C. elegans* Genome Sequencing Consortium. Led by visionaries Sulston and Waterston, the consortium describes its experiences and provides insights concerning the strategies needed to tackle larger genomes. The Review also describes some of the general features and various surprises that emerge from this first panoramic view. One expected feature is the prevalence of gene families, which dominate all plant and animal genomes. Perhaps more surprising is the clustering of many groups of similar genes: 402 such gene clusters were identified, many involving novel gene families.

As an example of the kind of analysis now possible, Clarke and Berg focus on one important family of regulatory genes, those encoding DNA-binding zinc finger proteins, and investigate the potential for using their informational properties to explore regulation over an entire genome. Similar analyses are obviously feasible for other kinds of regulators—homeodomain proteins, nuclear hormone receptors, kinases, and proteases.

The other Reviews in this issue concentrate on a few selected aspects of the sequence. Can one define the essence of being a eukaryote? What is involved in progressing from unicellular to multicellular life? These major questions are addressed by Chervitz *et al.* in a comprehensive comparison between the genomes of *S. cerevisiae* and *C. elegans*.

Bargmann (neurobiology) and Ruvkun and Hobert (development) look at the constellation of genes and pathways that are present and absent in *C. elegans*. Despite its humble 302-cell nervous system, the nematode possesses genes for most of the known molecular components of vertebrate brains, which suggests that it provides a valid minimal model for the nervous systems of all animals. The complete sequence reveals the full extent of gene families and shows which developmental control genes are universal and which are missing in different phylogenetic lineages. This will lead to a better understanding of features that make multicellular organisms similar and different.

Nematodes belong to a major phylum that is represented in extraordinary numbers on this planet and includes many important plant and animal parasites. Blaxter discusses the position of *C. elegans* in the animal kingdom and the significance of its DNA sequence as an aid to understanding and controlling nematode parasites.

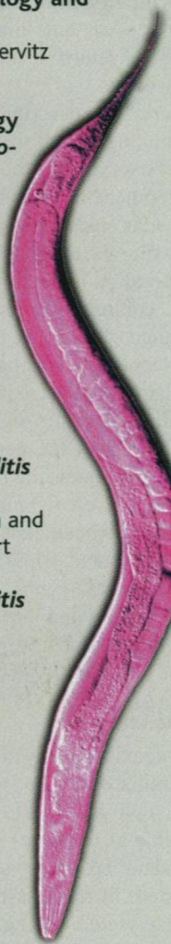
These analyses, fascinating as they are, barely scratch the surface of what will be possible with more extensive annotation, exploration, and exploitation of the *C. elegans* sequence. However, we are not yet ready to begin computing a worm. This first phase of sequencing is over, but the real research is only beginning.

—**JONATHAN HODGKIN, H. ROBERT HORVITZ, BARBARA R. JASNY, JUDITH KIMBLE**

J. Hodgkin is at the Medical Research Council Laboratory of Molecular Biology, Cambridge, CB2 2QH, UK. H. R. Horvitz is in the Department of Biology, Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. J. Kimble is in the Department of Biochemistry, Howard Hughes Medical Institute, University of Wisconsin, Madison, WI 53706, USA.

CONTENTS

REVIEW



See also related Editorial on p. 1989 and related News story on p. 1972.

Science