SCIENCE ONLINE SCIENCE MAGAZINE HOME SCIENCE NOW NEXT WAVE SCIENCE'S STKE SCIENCE CAREERS E-MARKETPLACE

DAVID S LEAF | Change Password | Change User Info | CiteTrack Alerts | Subscription Help | Sign Out

Science magazine

HELP SUBSCRIPTIONS FEEDBACK SIGN IN

SEARCH BROWSE ORDER THIS ARTICLE

GENOMICS:

The End of the Beginning

Sydney Brenner[*]

In classical experimental genetics, where many of us began, we could not assert the existence of a wild-type gene until a mutant version with an altered function had been isolated. For Mendel to say that there was a factor for tallness, he first had to find heritable dwarf variants that suffered from a lack of tallness. This genetics began with inherited changes in phenotype that provided, if not knowledge, then at least a classification of the functions of genes, and it used genetic complementation experiments to discover how many genes were involved in dictating each phenotype. But, if one asked how many genes were required to make a bacteriophage or a bacterium or a fly or a mouse, no answer could be given.

A quarter of a century ago, the advent of new methods to analyze genomes directly changed the field of genetics. When the genome of bacteriophage lambda was first sequenced, it allowed the enumeration of all of the open reading frames (DNA sequences that potentially can be translated into protein). Some of these were in genes that encoded proteins whose functions had been thoroughly explored, whereas others encoded new proteins that were not essential for the growth of bacteriophage in the laboratory.

Summary of this Article

dEbates: Submit a response to this article

Similar articles found in:
SCIENCE Online
ISI Web of Science
PubMed

PubMed Citation

Search Medline for articles by:
Brenner, S.

Search for citing articles in:
ISI Web of Science (3)

Alert me when:
new articles cite this article

Download to Citation Manager

Collections under which this article appears:
Genetics

Proteins are the workhorses of biological systems, and deepening functional analyses of organisms requires that their proteins be purified and characterized. By sequencing the genome of a complex organism, the amino acid sequences of all of the proteins are obtained, so to speak, in one blow, thus avoiding the terrifying prospect of separating and purifying all of the proteins, and sequencing them by laborious methods. Cloning the genes into expression vectors allows us to make large amounts of the proteins for study and, what is more, we can make mutations in them and study the consequences without ever going back to the original genome.

As time progressed, and methods of DNA sequencing improved, sequencing moved to larger and larger genomes. Although sequencing the human genome was contemplated quite early on, and sequencing of the *Caenorhabditis elegans* worm genome was begun as a continuation of the mapping program, what emerged next were the sequences of bacterial genomes, and the sequence of yeast, the latter accomplished by a European group effort. The sequence of the yeast genome was published in 1997, that of *C. elegans* in 1998 and, now, in three reviews in this issue (pages 2185, 2196, and 2204), we have the complete sequence of the 125-megabase genome of the fruit fly *Drosophila* (1-3).

When large-scale sequencing projects were first discussed in the mid-1980s, it was clear that a resource much larger

than the average research laboratory, as well as improvements in technology, would be required. Walter Gilbert was the first to suggest a sequence factory--I seem to remember the number of 250 for the technicians that would be needed--but most of our colleagues were bitterly opposed to this idea. One, I remember, advocated the cottage industry model, hoping that the sequence of the genomes of organisms would be accomplished by many scientists working on individual genes. However, building factories with increased automation and very large computer resources has provided an answer to large-scale genome sequencing.

In their review, Adams *et al*. (1) provide a list of gene functions in *Drosophila* , classified by the proteins deduced from the genomic sequences into the now familiar classes (of which "unknown" and "hypothetical" are the most common). Rubin and colleagues (3) compare the *Drosophila* genome sequence with that of yeast and *C. elegans*, the only other eukaryote genomes sequenced so far. It should be noted that the fly has fewer genes than the worm; the genome sequence predicts about 14,200 proteins for *Drosophila* as opposed to 18,400 for *C. elegans*.

Old geneticists knew what they were talking about when they used the term "gene", but it seems to have become corrupted by modern genomics to mean any piece of expressed sequence, just as the term algorithm has become corrupted in much the same way to mean any piece of a computer program. I suggest that we now use the term "genetic locus" to mean the stretch of DNA that is characterized either by mapped mutations as in the old genetics or by finding a complete open reading frame as in the new genomics. In higher organisms, we often find closely related genes that subserve closely related, but subtly different, functions. Thus, vertebrate genomes contain three different genetic loci specifying three different aldolase enzymes. In *Drosophila* , we have one aldolase "genetic locus" that produces three different aldolases by variable splicing of the messenger RNA. Indeed, this is clearly part of the genomic style of the fly. *Drosophila* has one myosin locus that produces all of the different heavy chains by variable splicing; in contrast, *C. elegans*, with simpler muscle systems, has four different myosin genes. We have to appreciate this before we can make sense of gene numbers. It also leads one to be cautious about the commonly accepted generalization that it takes four invertebrate genomes to make a vertebrate. The science of genomics is still in its infancy, and we will have to acquire far more sophisticated views of genomes and their evolution before we can answer such questions as why the fly has 352 zinc-finger genes but the worm has only 132.

The analysis of genome sequences gives us a comprehensive protein parts list and it short-cuts the massive amount of work that would have been required to characterize each protein individually. But there is one important piece of information that is almost totally missing: the sequence information that specifies when and where and for how long a gene is turned on or off. This switching information--which I call the left-hand value of the gene by analogy with the address of a computer location--cannot be deduced from the sequence. It is absolutely essential information because in complex organisms, evolution does not proceed by enlarging the protein inventory but rather by modulating the expression of genes.

The functional properties assigned to the protein products of genes are centered on what might be called molecular functions. These can be specified whenever a protein is found to be similar to one for which the function has been determined by conventional biochemical methods. It is the way one learns to speak a natural language by listening to other speakers, and is not the result of some elaborate computation. Quite often there is little connection between these molecular functions and the classical assignments of function by phenotype (which are at the level of the organism). The middle ground--that is, the participation of proteins in the physiology of cells and how cells contribute to the function of the organism--is a gap that still remains to be closed.

The problems faced by pre- and post-genomic genetics are therefore much the same--they all involve bridging the chasm between genotype and phenotype. Genome sequencing represents only a new beginning and not an end in itself. It is useful to have and will help us to answer the many questions that still lie ahead. Yeast, *C. elegans*, and *Drosophila* have large constituencies of researchers who will make good use of the genome sequence and, in the coming years, will tell us what it all means.

References

1. M. D. Adams *et al*., *Science* **287**, 2185 (2000).
2. E. W. Myers *et al*., *Science* **287**, 2196 (2000).
3. G. M. Rubin *et al*., *Science* **287**, 2204 (2000).

The author is at the Molecular Sciences Institute, Berkeley, CA 94704, USA. E-mail: brenner@molsci.org

- **Summary** of this Article
- **dEbates: Submit a response to this article**
- Similar articles found in:
  SCIENCE Online
  ISI Web of Science
  PubMed
- PubMed Citation
- Search Medline for articles by:
  Brenner, S.
- Search for citing articles in:
  ISI Web of Science (3)
- Alert me when:
  new articles cite this article
- Download to Citation Manager
- Collections under which this article appears:
  Genetics

Volume 287, Number 5461, Issue of 24 Mar 2000, pp. 2173-2174.
Copyright © 2000 by The American Association for the Advancement of Science.

PAGE TOP