

[What's a Genome?](#)
[What's a Chromosome?](#)
[What's a Gene?](#)
[What's DNA?](#)
[What's Genome Sequencing?](#)
[What is genome sequencing?](#)
[Why is genome sequencing so important?](#)
[How do you sequence a genome?](#)
[How does DNA sequencing work?](#)
[How does the sequencing machine know whether a base is an A, C, G, or T?](#)
[What happens after DNA sequences come out of the sequencing machines?](#)
[How do you assemble a genome?](#)
[How do scientists know if a genome sequence is right?](#)
[What makes sequencing the human genome different from sequencing other genomes?](#)
[When is a genome sequence done?](#)
[What's a genome map?](#)
[What are genome variations?](#)
[GNN Home](#)

- GENOME SEQUENCING

Genome sequencing is figuring out the order of DNA nucleotides, or bases, in a genome—the order of As, Cs, Gs, and Ts that make up an organism's DNA. The human genome is made up of over 3 billion of these genetic letters.

Today, DNA sequencing on a large scale—the scale necessary for ambitious projects such as sequencing an entire genome—is mostly done by high-tech machines. Much as your eye scans a sequence of letters to read a sentence, these machines "read" a sequence of DNA bases.

A DNA sequence that has been translated from life's chemical alphabet into our alphabet of written letters might look like this:

AGTCCGCGAATACAGGCTCGGT

That is, in this particular piece of DNA, an adenine (A) is followed by a guanine (G), which is followed by a thymine (T), which in turn is followed by a cytosine (C), another cytosine (C), and so on.

What is genome sequencing?

By itself, not a whole lot. Genome sequencing is often compared to "decoding," but a sequence is still very much in code. In a sense, a genome sequence is simply a very long string of letters in a mysterious language.

When you read a sentence, the meaning is not just in the sequence of the letters. It is also in the words those letters make and in the grammar of the language. Similarly, the human genome is more than just its sequence.

Imagine the genome as a book written without capitalization or punctuation, without breaks between words, sentences, or paragraphs, and with strings of nonsense letters scattered between and even within sentences. A passage from such a book in English might look like this:

**lgvrldlmnqvttthequickababcmfxlqbrownfoxjulrvsmped
overthelazyyyzplfdogjjjiurttiythedoglayhbbeldquietly
dreaminghwwiqldnsofdinnerplwosiucnd**

Pass your mouse over the letters to see the hidden words.

Even in a familiar language it is difficult to pick out the meaning of the passage: *The quick brown fox jumped over the lazy dog. The dog lay quietly dreaming of dinner.* And the genome is "written" in a far less familiar language, multiplying the difficulties involved in reading it.

So sequencing the genome doesn't immediately lay open the genetic secrets of an entire species. Even with a rough draft of the human genome sequence in hand, much work remains to be done. Scientists still have to translate those strings of letters into an understanding of how the genome works: what the various genes that make up the genome do, how different genes are related, and how the various parts of the genome are coordinated. That is, they have to figure out what those letters of the genome sequence *mean*.

Why is genome sequencing so important?

Sequencing the genome is an important step towards understanding it.

At the very least, the genome sequence will represent a valuable shortcut, helping scientists find genes much more easily and quickly. A genome sequence does contain some clues about where genes are, even though scientists are just learning to interpret these clues.

Scientists also hope that being able to study the entire genome sequence will help them understand how the genome as a whole works—how genes work together to direct the growth, development and maintenance of an entire organism.

Finally, genes account for less than 25 percent of the DNA in the genome, and so knowing the entire genome sequence will help scientists study the parts of the genome outside the genes. This includes the regulatory regions that control how genes are turned on and off, as well as long stretches of "nonsense" or "junk" DNA—so called because we don't yet know what, if anything, it does.

How do you sequence a genome?



Lab technician working with sequencing machines
Courtesy of Celera Genomics

The quick answer to this question is: in pieces. The whole genome can't be sequenced all at once because available methods of DNA sequencing can only handle short stretches of DNA at a time.

So instead, scientists must break the genome into small pieces, sequence the pieces, and then reassemble them in the proper order to arrive at the sequence of the whole genome. Much of the work involved in sequencing lies in putting together this giant biological jigsaw puzzle.

There are two approaches to the task of cutting up the genome and putting it back together again. One strategy, known as the "clone-by-clone" approach, involves first breaking the genome up into relatively large chunks, called clones, about 150,000 base pairs (bp) long. Scientists use genome mapping techniques (discussed in further detail later) to figure out where in the genome each clone belongs. Next they cut each clone into smaller, overlapping pieces the right size for sequencing—about 500 BP each. Finally, they sequence the pieces and use the overlaps to reconstruct the sequence of the whole clone.

The other strategy, called "whole-genome shotgun" method, involves breaking the genome up into small pieces, sequencing the pieces, and reassembling the pieces into the full genome sequence.



Room filled with sequencing machines
Courtesy of Celera Genomics

Each of these approaches has advantages and disadvantages. The clone-by-clone method is reliable but slow, and the mapping step can be especially time-consuming. By contrast, the whole-genome shotgun method is potentially very fast, but it can be extremely difficult to put together so many tiny pieces of sequence all at once.

Both approaches have already been used to sequence whole genomes. The whole-genome shotgun method was used to sequence the genome of the bacterium *Haemophilus influenzae*, while the genome of baker's yeast, *Saccharomyces cerevisiae*, was sequenced with a clone-by-clone method. Sequencing the human genome was done using both approaches.

.

[Previous](#)

[Top of page](#)

[Continue](#)

Updated on January 15, 2003