Genomics for Energy and Environmental Science

What is Genomics?
A genome is the complete set of genetic material encoding the biological instructions for developing and maintaining an organism. The genome for any free-living system—ranging from a single bacterial cell to a complex multicellular organism such as a human or a tree—is made up of DNA (deoxyribonucleic acid). Genomics is a scientific field focused on sequencing and analyzing genomes. Insights from genomic information are catalyzing an extraordinary transformation in our understanding of biological systems and how we use their capabilities to address challenges in the environment, energy, industry, biomedicine, and other application areas.

DNA: Life’s Molecular Archive of Genetic Information
Nucleotides are chemical subunits strung together to form DNA’s double-stranded, twisted ladder structure. DNA is made up of four kinds of nucleotides, each containing a different nitrogenous base: adenine (A), cytosine (C), guanine (G), and thymine (T). Each base on one DNA strand pairs with a base on the other strand—A’s pair with T’s, and C’s with G’s. These base pairs form the “rungs” of the DNA ladder and hold the two DNA strands together. The long, parallel “side rails” of the ladder, forming the backbone of each DNA strand, consist of sugar and phosphate molecules. DNA molecules can be hundreds, thousands, or millions of base pairs in length.

DNA molecules can be hundreds, thousands, or millions of base pairs in length. DNA is made up of four kinds of nucleotides, each containing a different nitrogenous base: adenine (A), cytosine (C), guanine (G), and thymine (T). Each base on one DNA strand pairs with a base on the other strand—A’s pair with T’s, and C’s with G’s. These base pairs form the “rungs” of the DNA ladder and hold the two DNA strands together. The long, parallel “side rails” of the ladder, forming the backbone of each DNA strand, consist of sugar and phosphate molecules. DNA molecules can be hundreds, thousands, or millions of base pairs in length.

Metagenomics: Sequencing and Analyzing Microbial Communities in Diverse Environments
Microbes and their communities make up the foundation of the biosphere and sustain all life on Earth. Although largely unexplored, the dimensions of the microbial world are immense, comprising more than half the planet’s biomass. The number of bacterial cells on Earth is estimated to be a billion times greater than the number of stars in the universe. The unique biochemicals that enable microbes to thrive in every niche on the planet represent a diverse and virtually limitless wealth of capabilities that can meet many national challenges.

Metagenomics is the sequencing and analysis of DNA extracted directly from microbial communities in the environment. Metagenomic studies are allowing the discovery of an array of previously unknown genes and proteins, thousands of species, and innumerable variations in critical biochemical capabilities. Pictured at right are some metagenome projects that are providing important insights to address challenges related to energy production and environmental science.

Tormenta gut. More than 200 species of microbes make up the community residing in this termite hindgut. Together they produce a bounty of wood-degrading enzymes that could be used by industry to make biofuels from woodchips and other forms of fibrous cellulosic biomass.

Thermal pools. To identify new microbes and enzymes resistant to the heat and stresses of industrial processing, researchers are investigating microbial communities living at near-boiling temperatures in pools at Yellowstone National Park.

EXPRESSIONING THE GENOME: DNA → RNA → PROTEIN

Translation is the process of making a temporary RNA copy of a gene’s DNA sequence. This RNA copy of a gene is called messenger RNA (mRNA). Like DNA, RNA consists of a long chain of nucleotides, but RNA is typically single stranded and uses a nucleotide called uracil (U) in place of thymine (T). The enzyme RNA polymerase binds and separates the double strands of DNA and then uses one DNA strand as a template for assembling mRNA. mRNA is typically single stranded and uses a nucleotide called uracil (U) in place of thymine (T). The enzyme RNA polymerase binds and separates the double strands of DNA and then uses one DNA strand as a template for assembling mRNA.

Translation is a process that uses the nucleotide base sequence of mRNA to direct the synthesis of a protein’s amino acid sequence. A large molecular complex called a ribosome binds the mRNA to initiate translation. With the help of another RNA called transfer RNA (tRNA), the ribosome adds amino acids one at a time to build a linear protein chain. A set of three mRNA bases (called a codon) specifies a particular amino acid or signals the ribosome to stop or start protein synthesis. Once the linear chain of amino acids is complete, the chain folds into a specific three-dimensional protein structure that performs a particular biological function.

For the ribosome to add an amino acid to a growing polypeptide chain, the amino acid must be attached to the correct tRNA. The correct tRNA can be identified by comparing the mRNA codon to a chart of codon-anticodon pairing that shows which tRNAs recognize which codons. Once the correct tRNA is identified, the ribosome adds the amino acid to the polypeptide chain.

The Genetic Code. Universal among all life forms, the genetic code is the language used to write the mRNA instructions for building proteins. Each three-letter codon in the mRNA specifies a particular amino acid. Since there are 64 different codons and 20 different amino acids, an amino acid can be represented by more than one codon. This redundancy allows for alternate pathways in the process of initiating protein synthesis. Three stop codons signify the end of a protein sequence.

The Genetic Code: Universal among all life forms, the genetic code is the language used to write the mRNA instructions for building proteins. Each three-letter codon in the mRNA specifies a particular amino acid. Since there are 64 different codons and 20 different amino acids, an amino acid can be represented by more than one codon. This redundancy allows for alternate pathways in the process of initiating protein synthesis. Three stop codons signify the end of a protein sequence.

Expressing the Genome: DNA → RNA → Protein

A gene is turned “on” when it is actively being expressed into protein, and turned “off” when expression is stopped. Producing proteins from genes occurs in two basic steps: (1) transcription and (2) translation.

Transcription is the process of making a temporary RNA copy of a gene’s DNA sequence. This RNA copy of a gene is called messenger RNA (mRNA). Like DNA, RNA consists of a long chain of nucleotides, but RNA is typically single stranded and uses a nucleotide called uracil (U) in place of thymine (T). The enzyme RNA polymerase binds and separates the double strands of DNA and then uses one DNA strand as a template for assembling mRNA.

Translation is a process that uses the nucleotide base sequence of mRNA to direct the synthesis of a protein’s amino acid sequence. A large molecular complex called a ribosome binds the mRNA to initiate translation. With the help of another RNA called transfer RNA (tRNA), the ribosome adds amino acids one at a time to build a linear protein chain. A set of three mRNA bases (called a codon) specifies a particular amino acid or signals the ribosome to stop or start protein synthesis. Once the linear chain of amino acids is complete, the chain folds into a specific three-dimensional protein structure that performs a particular biological function.

Expressing the Genome in Bacterial Cells
Bacteria are prokaryotes—single-celled organisms that lack a nucleus. Because no nuclear membrane separates a bacterium’s genome from ribosomes and other cellular components, protein synthesis can start before an mRNA transcript is complete. Bacterial genes do not have introns interrupting regions of DNA. Thus editing mRNA transcripts, which involves removing introns prior to protein synthesis, is not needed.

Expressing the Genome in Plant Cells
Plants are eukaryotes—organisms with cells that contain a membrane-bound nucleus. A eukaryote’s DNA is in the nucleus where mRNA is transcribed. The genes in plants and other eukaryotic organisms, such as humans and animals, contain noncoding regions called introns. In eukaryotic cells, introns are removed from mRNA transcripts, and the remaining coding regions (called exons) are spliced back together. Once edited, the mRNA is transported outside the nucleus for translation into proteins by ribosomes.
Still growing...