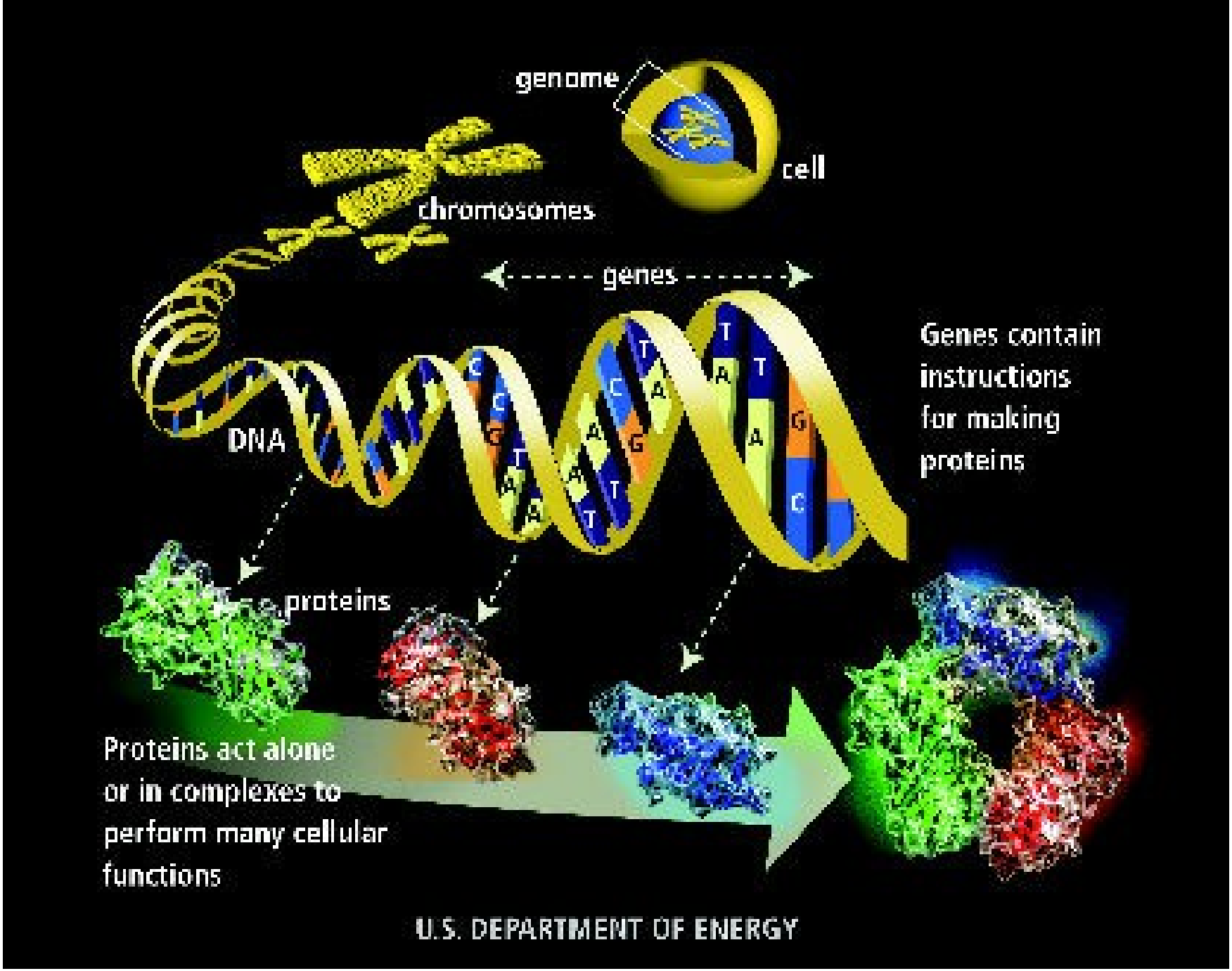
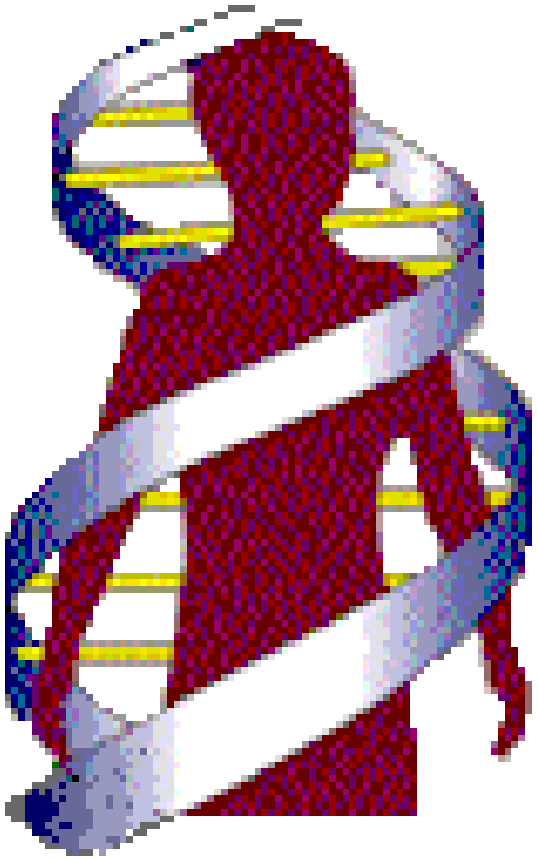
Genomics and Its Impact on Science and Society:

The Human Genome Project and Beyond

U.S. Department of Energy Genome Programs
http://doegenomes.org



Human Genome Program, U.S. Department of Energy, Genomics and Its Impact on Medicine and Society: A 2001 Primer, 2001



Human Genome Project

Human Genome Project

Goals:

- identify all the approximate 30,000 genes in human DNA,
- determine the sequences of the 3 billion chemical base pairs that make up human DNA,
- store this information in databases,
- improve tools for data analysis,
- transfer related technologies to the private sector, and
- address the ethical, legal, and social issues (ELSI) that may arise from the project.

Milestones:

- 1990: Project initiated as joint effort of U.S. Department of Energy and the National Institutes of Health
- June 2000: Completion of a working draft of the entire human genome
- February 2001: Analyses of the working draft are published
- April 2003: HGP sequencing is completed and Project is declared finished two years ahead of schedule

U.S. Department of Energy Genome Programs, Genomics and Its Impact on Science and Society, 2003

What does the draft human genome sequence tell us?

By the Numbers

- The human genome contains 3 billion chemical nucleotide bases (A, C, T, and G).
- The average gene consists of 3000 bases, but sizes vary greatly, with the largest known human gene being dystrophin at 2.4 million bases.
- The total number of genes is estimated at around 30,000--much lower than previous estimates of 80,000 to 140,000.
- Almost all (99.9%) nucleotide bases are exactly the same in all people.
- The functions are unknown for over 50% of discovered genes.

What does the draft human genome sequence tell us?

How It's Arranged

• The human genome's gene-dense "urban centers" are predominantly composed of the DNA building blocks G and C.

• In contrast, the gene-poor "deserts" are rich in the DNA building blocks A and T. GC- and AT-rich regions usually can be seen through a microscope as light and dark bands on chromosomes.

• Genes appear to be concentrated in random areas along the genome, with vast expanses of noncoding DNA between.

• Stretches of up to 30,000 C and G bases repeating over and over often occur adjacent to gene-rich areas, forming a barrier between the genes and the "junk DNA." These CpG islands are believed to help regulate gene activity.

• Chromosome 1 has the most genes (2968), and the Y chromosome has the fewest (231).

What does the draft human genome sequence tell us?

The Wheat from the Chaff

• Less than 2% of the genome codes for proteins.

• Repeated sequences that do not code for proteins ("junk DNA") make up at least 50% of the human genome.

• Repetitive sequences are thought to have no direct functions, but they shed light on chromosome structure and dynamics. Over time, these repeats reshape the genome by rearranging it, creating entirely new genes, and modifying and reshuffling existing genes.

• The human genome has a much greater portion (50%) of repeat sequences than the mustard weed (11%), the worm (7%), and the fly (3%).

What does the draft human genome sequence tell us?

How the Human Compares with Other Organisms

- Unlike the human's seemingly random distribution of gene-rich areas, many other organisms' genomes are more uniform, with genes evenly spaced throughout.

- Humans have on average three times as many kinds of proteins as the fly or worm because of mRNA transcript "alternative splicing" and chemical modifications to the proteins. This process can yield different protein products from the same gene.

- Humans share most of the same protein families with worms, flies, and plants; but the number of gene family members has expanded in humans, especially in proteins involved in development and immunity.

- Although humans appear to have stopped accumulating repeated DNA over 50 million years ago, there seems to be no such decline in rodents. This may account for some of the fundamental differences between hominids and rodents, although gene estimates are similar in these species. Scientists have proposed many theories to explain evolutionary contrasts between humans and other organisms, including those of life span, litter sizes, inbreeding, and genetic drift.

What does the draft human genome sequence tell us?

Variations and Mutations

• Scientists have identified about 3 million locations where single-base DNA differences (SNPs) occur in humans. This information promises to revolutionize the processes of finding chromosomal locations for disease-associated sequences and tracing human history.

• The ratio of germline (sperm or egg cell) mutations is 2:1 in males vs females. Researchers point to several reasons for the higher mutation rate in the male germline, including the greater number of cell divisions required for sperm formation than for eggs.

How does the human genome stack up?

Organism	Genome Size (Bases)	Estimated Genes
Human (*Homo sapiens*)	3 billion	30,000
Laboratory mouse (*M. musculus*)	2.6 billion	30,000
Mustard weed (*A. thaliana*)	100 million	25,000
Roundworm (*C. elegans*)	97 million	19,000
Fruit fly (*D. melanogaster*)	137 million	13,000
Yeast (*S. cerevisiae*)	12.1 million	6,000
Bacterium (*E. coli*)	4.6 million	3,200
Human immunodeficiency virus (HIV)	9700	9

Future Challenges: What We Still Don't Know

- Gene number, exact locations, and functions
- Gene regulation
- DNA sequence organization
- Chromosomal structure and organization
- Noncoding DNA types, amount, distribution, information content, and functions
- Coordination of gene expression, protein synthesis, and post-translational events
- Interaction of proteins in complex molecular machines
- Predicted vs experimentally determined gene function
- Evolutionary conservation among organisms
- Protein conservation (structure and function)
- Proteomes (total protein content and function) in organisms
- Correlation of SNPs (single-base DNA variations among individuals) with health and disease
- Disease-susceptibility prediction based on gene sequence variation
- Genes involved in complex traits and multigene diseases
- Complex systems biology including microbial consortia useful for environmental restoration
- Developmental genetics, genomics

U.S. Department of Energy Genome Programs, Genomics and Its Impact on Science and Society, 2003

Anticipated Benefits of Genome Research

Molecular Medicine

- improve diagnosis of disease
- detect genetic predispositions to disease
- create drugs based on molecular information
- use gene therapy and control systems as drugs
- design “custom drugs” (pharmacogenomics) based on individual genetic profiles

Microbial Genomics

- rapidly detect and treat pathogens (disease-causing microbes) in clinical practice
- develop new energy sources (biofuels)
- monitor environments to detect pollutants
- protect citizenry from biological and chemical warfare
- clean up toxic waste safely and efficiently

U.S. Department of Energy Genome Programs, Genomics and Its Impact on Science and Society, 2003

Anticipated Benefits of Genome Research-cont.

Risk Assessment

• evaluate the health risks faced by individuals who may be exposed to radiation (including low levels in industrial areas) and to cancer-causing chemicals and toxins

Bioarchaeology, Anthropology, Evolution, and Human Migration

• study evolution through germline mutations in lineages
• study migration of different population groups based on maternal inheritance
• study mutations on the Y chromosome to trace lineage and migration of males
• compare breakpoints in the evolution of mutations with ages of populations and historical events

Anticipated Benefits of Genome Research-cont.

DNA Identification (Forensics)

- identify potential suspects whose DNA may match evidence left at crime scenes
- exonerate persons wrongly accused of crimes
- identify crime and catastrophe victims
- establish paternity and other family relationships
- identify endangered and protected species as an aid to wildlife officials (could be used for prosecuting poachers)
- detect bacteria and other organisms that may pollute air, water, soil, and food
- match organ donors with recipients in transplant programs
- determine pedigree for seed or livestock breeds
- authenticate consumables such as caviar and wine

Anticipated Benefits of Genome Research-cont.

Agriculture, Livestock Breeding, and Bioprocessing

- grow disease-, insect-, and drought-resistant crops
- breed healthier, more productive, disease-resistant farm animals
- grow more nutritious produce
- develop biopesticides
- incorporate edible vaccines incorporated into food products
- develop new environmental cleanup uses for plants like tobacco

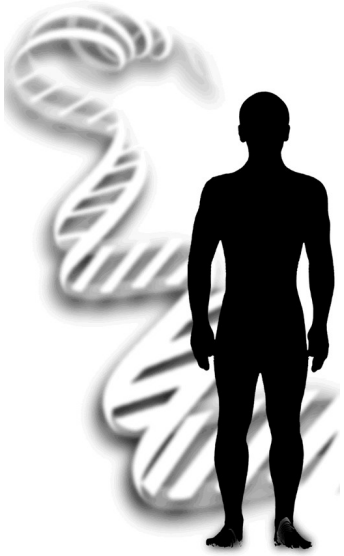
Medicine and the New Genetics

Gene Testing - Pharmacogenomics - Gene Therapy

Anticipated Benefits:

- improved diagnosis of disease
- earlier detection of genetic predispositions to disease
- rational drug design
- gene therapy and control systems for drugs
- personalized, custom drugs

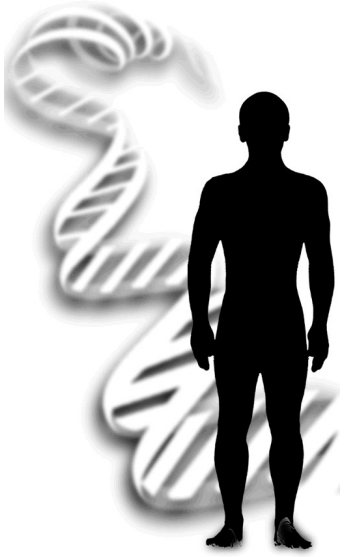
U.S. Department of Energy Genome Programs, Genomics and Its Impact on Science and Society, 2003



ELSI: Ethical, Legal, and Social Issues

- **Privacy and confidentiality of genetic information**.
- **Fairness in the use of genetic information** by insurers, employers, courts, schools, adoption agencies, and the military, among others.
- **Psychological impact, stigmatization, and discrimination** due to an individual's genetic differences.
- **Reproductive issues** including adequate and informed consent and use of genetic information in reproductive decision making.
- **Clinical issues** including the education of doctors and other health-service providers, people identified with genetic conditions, and the general public about capabilities, limitations, and social risks; and implementation of standards and quality- control measures.

U.S. Department of Energy Genome Programs, Genomics and Its Impact on Science and Society, 2003



ELSI Issues (cont.)

- **Uncertainties associated with gene tests for susceptibilities and complex conditions** (e.g., heart disease, diabetes, and Alzheimer's disease).
- **Fairness in access to advanced genomic technologies**.
- **Conceptual and philosophical implications** regarding human responsibility, free will vs genetic determinism, and concepts of health and disease.
- **Health and environmental issues** concerning genetically modified (GM) foods and microbes.
- **Commercialization of products** including property rights (patents, copyrights, and trade secrets) and accessibility of data and materials.

Beyond the HGP: What's Next?

HapMap

Chart genetic variation within the human genome

GENOMES to LIFE

ACCELERATING BIOLOGICAL DISCOVERY

U.S. DEPARTMENT OF ENERGY

Systems Biology

Exploring Microbial Genomes for Energy and the Environment

GENOMES to LIFE

ACCELERATING BIOLOGICAL DISCOVERY

U.S. DEPARTMENT OF ENERGY

Genomes to Life:
A DOE Systems Biology Program

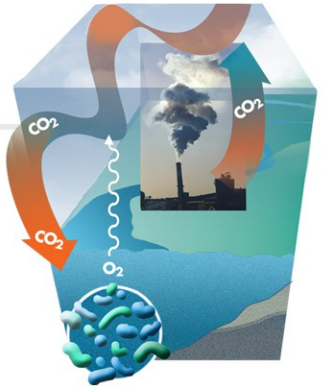
Exploring Microbial Genomes for Energy and the Environment

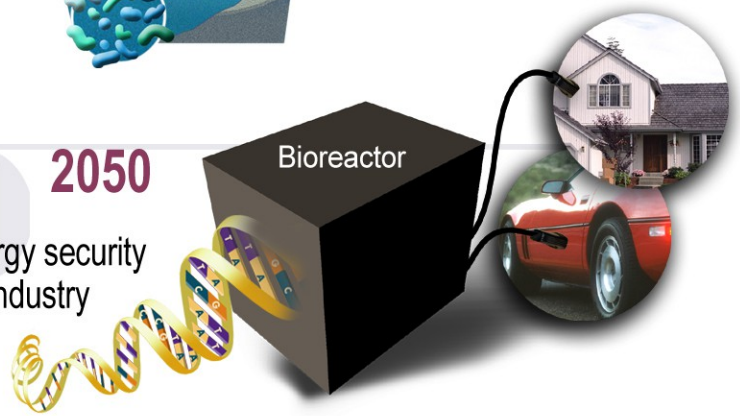
Goals

• identify the protein machines that carry out critical life functions
• characterize the gene regulatory networks that control these machines
• characterize the functional repertoire of complex microbial communities in their natural environments
• develop the computational capabilities to integrate and understand these data and begin to model complex biological systems

GTL Applications in Energy Security and Global Climate Change

Within a Decade	*Long Term*
Develop knowledge base for cost-effective cleanup strategies	**2020** Save billions of dollars in toxic waste cleanup and disposal
Understand earth's natural carbon cycle and design strategies for enhanced carbon capture	**2040** Help stabilize atmospheric carbon dioxide to counter global warming
Increase biological sources of fuels and electricity	**2050** Contribute to U.S. energy security • Biohydrogen-based industry in place





YGG-01-0521R7

HapMap

An NIH program to chart genetic variation

within the human genome

- Begun in 2002, the project is a 3-year effort to construct a map of the patterns of SNPs (single nucleotide polymorphisms) that occur across populations in Africa, Asia, and the United States.
- Consortium of researchers from six countries
- Researchers hope that dramatically decreasing the number of individual SNPs to be scanned will provide a shortcut for identifying the DNA regions associated with common complex diseases
- Map may also be useful in understanding how genetic variation contributes to responses in environmental factors