Chapter 21

Genomes and Their Evolution

Lectures by
Erin Barley
Kathleen Fitzpatrick

© 2011 Pearson Education, Inc.
Chapter 21

Genomes and Their Evolution

Reading the Leaves from the Tree of Life

• Complete genome sequences exist for a human, chimpanzee, E. coli, brewer’s yeast, corn, fruit fly, house mouse, rhesus macaque, and other organisms

• Genome comparisons provide information about evolutionary history
• **Genomics**
  study of whole sets of genes and their interactions

• **Bioinformatics**
  application of computational methods to the storage and analysis of biological data
• **Human Genome Project – The sequencing of human DNA**

• Started in 1990, completed 2003

• The project had three stages:
  • Genetic (or linkage) mapping
  • Physical mapping
  • DNA sequencing
Three-Stage Approach to Genome Sequencing

- A **linkage map** (genetic map) maps the location of several thousand genetic markers on each chromosome.
- A **genetic marker** is a gene or other identifiable DNA sequence.
- **Recombination frequencies** are used to determine the order and relative distances between genetic markers.
Cytogenetic map

Genes located by FISH

Chromosome bands
Cytogenetic map

Genes located by FISH

Chromosome bands

Genetic markers

1 Linkage mapping
A **physical map** expresses the distance between genetic markers, usually as the number of base pairs along the DNA.

It is constructed by cutting a DNA molecule into many short fragments and arranging them in order by identifying overlaps.
• Sequencing machines are used to determine the complete nucleotide sequence of each chromosome
• A complete haploid set of human chromosomes consists of 3.2 billion base pairs

1. Linkage mapping

2. Physical mapping

3. DNA sequencing

© 2011 Pearson Education, Inc.
Whole-Genome Shotgun Approach to Genome Sequencing

- The whole-genome shotgun approach was developed by J. Craig Venter in 1992
- This approach skips genetic and physical mapping and sequences random DNA fragments directly!
- Powerful computer programs are used to order fragments into a continuous sequence.
1. Cut the DNA into overlapping fragments short enough for sequencing.

2. Clone the fragments in plasmid or phage vectors.
1 Cut the DNA into overlapping fragments short enough for sequencing.

2 Clone the fragments in plasmid or phage vectors.

3 Sequence each fragment.

Figure 21.3-2
1. Cut the DNA into overlapping fragments short enough for sequencing.

2. Clone the fragments in plasmid or phage vectors.

3. Sequence each fragment.

4. Order the sequences into one overall sequence with computer software.
• At first many scientists were skeptical about the whole-genome shotgun approach
• *Now the sequencing method of choice!*  
• The development of newer sequencing techniques has resulted in *massive increases in speed* and *decreases in cost*
Scientists use bioinformatics to analyze genomes and their functions

- The Human Genome Project established databases and refined analytical software to make data available on the Internet

- **Bioinformatics resources** are provided by a number of sources
  - National Library of Medicine and the National Institutes of Health (NIH) created the National Center for Biotechnology Information (NCBI)
  - European Molecular Biology Laboratory
  - DNA Data Bank of Japan
  - BGI in Shenzhen, China
Genbank, the NCBI database of sequences, doubles its data approximately every 18 months.

Software is available that allows online visitors to search Genbank for matches to:

- A specific DNA sequence
- A predicted protein sequence
- Common stretches of amino acids in a protein

The NCBI website also provides 3-D views of all protein structures that have been determined.
Identifying Protein-Coding Genes and Understanding Their Functions

• Using available DNA sequences, geneticists can study genes directly in an approach called reverse genetics.

• **Gene Annotation** - The identification of protein coding genes within DNA sequences in a database.
  
  - largely an automated process.

• Comparison of sequences of previously unknown genes with those of known genes in other species may help provide clues about their function.

© 2011 Pearson Education, Inc.
How Systems Are Studied: An Example

- A systems biology approach can be applied to define gene circuits and protein interaction networks
- Researchers working on the yeast *Saccharomyces cerevisiae* used sophisticated techniques to disable pairs of genes one pair at a time, creating double mutants
- Computer software then mapped genes to produce a network-like “functional map” of their interactions
- The systems biology approach is possible because of advances in bioinformatics
Figure 21.5a

Translation and ribosomal functions

Mitochondrial functions

RNA processing

Transcription and chromatin-related functions

Nuclear-cytoplasmic transport

Nuclear migration and protein degradation

Mitosis

DNA replication and repair

Cell polarity and morphogenesis

Protein folding, glycosylation, and cell wall biosynthesis

Peroxisomal functions

Metabolism and amino acid biosynthesis

Secretion and vesicle transport

© 2011 Pearson Education, Inc.
Application of Systems Biology to Medicine

- A systems biology approach has medical applications
- **Cancer Genome Atlas** project
  - seeking common mutations in three types of cancer by comparing gene sequences and expression in cancer versus normal cells
  - has been so fruitful, it will be extended to ten other common cancers
  - Silicon and glass “chips” have been produced that hold a microarray of most known human genes
Multicellular eukaryotes have much noncoding DNA and many multigene families

- Most eukaryotic genomes neither encodes proteins nor functional RNAs.
- Much evidence indicates that noncoding DNA (previously called “junk DNA”) plays important roles in the cell.
- For example, genomes of humans, rats, and mice show high sequence conservation for about 500 noncoding regions.
• **Sequencing of the human genome reveals that 98.5% does not code for proteins, rRNAs, or tRNAs**

• **About a quarter of the human genome codes for introns and gene-related regulatory sequences**
• **Intergenic DNA** is noncoding DNA found between genes
  
  – **Pseudogenes** are former genes that have accumulated mutations and are nonfunctional
  
  – **Repetitive DNA** is present in multiple copies in the genome

• About three-fourths of repetitive DNA is made up of *transposable elements* and sequences related to them
Figure 21.7 Types of DNA sequences in the human genome.
Transposable Elements and Related Sequences

- mobile DNA segments
- Discovered by geneticist Barbara McClintock’s breeding experiments with Indian corn
- McClintock identified corn kernel color changes that showed some genetic elements move from other genome locations into the genes for kernel color
- These transposable elements move from one site to another in a cell’s DNA.
- Present in both prokaryotes and eukaryotes
How Transposable Elements Contribute to Genome Evolution

- Multiple copies of similar transposable elements may facilitate recombination, or crossing over, between different chromosomes.
- Insertion of transposable elements within a protein-coding sequence may block protein production.
- Insertion of transposable elements within a regulatory sequence may increase or decrease protein production.
• Transposable elements may carry a gene or groups of genes to a new position
• Transposable elements may also create new sites for alternative splicing in an RNA transcript
• In all cases, changes are usually detrimental but may on occasion prove advantageous to an organism
Other Repetitive DNA, Including Simple Sequence DNA

- About 15% of the human genome consists of duplication of long sequences of DNA from one location to another.
- In contrast, simple sequence DNA contains many copies of tandemly repeated short sequences.
• A series of repeating units of 2 to 5 nucleotides is called **a short tandem repeat (STR)**
• The repeat number for **STRs** can vary among sites (within a genome) or individuals
• **Simple sequence DNA is common in centromeres and telomeres, where it probably plays structural roles in the chromosome**
Duplication, rearrangement, and mutation of DNA contribute to genome evolution

- The basis of change at the genomic level is mutation

- The earliest forms of life had minimal number of genes

- The size of genomes has increased over evolutionary time, with the extra genetic material providing raw material for gene diversification
Duplication of Entire Chromosome Sets

- Accidents in meiosis can lead to **Polyploidy**
- The genes in one or more of the extra sets can diverge by accumulating mutations.
Alterations of Chromosome Structure

- Duplications and inversions result from mistakes during meiotic recombination
- Humans have 23 pairs of chromosomes, while chimpanzees have 24 pairs
- Following the divergence of humans and chimpanzees from a common ancestor, two ancestral chromosomes fused in the human line
Figure 21.12

(a) Human and chimpanzee chromosomes

(b) Human and mouse chromosomes
(a) Human and chimpanzee chromosomes
Figure 21.12b

(b) Human and mouse chromosomes

© 2011 Pearson Education, Inc.
Duplication and Divergence of Gene-Sized Regions of DNA

- Unequal crossing over during prophase I of meiosis can result in one chromosome with a deletion and another with a duplication of a particular region.

- Transposable elements can provide sites for crossover between nonsister chromatids.
Figure 21.13

Incorrect pairing of two homologs during meiosis

Nonsister chromatids

Gene

Transposable element

Crossover point

and