Universal Rules of Biology

- All cells store genetic (hereditary) information in *DeoxyriboNucleic Acid (DNA)*
- All cells replicate their genetic information by using the original DNA as a template and *enzymes* (biological catalysts that speed the process)
- All cells transcribe DNA into *RiboNucleic Acid (RNA)* and translate RNA into proteins in the same way
- There are exceptions to every rule in biology, including this one
Other Rules (and Exceptions)

- Mammals bear live young
- Human cells have *mitochondria* (small organs that produce energy for the cell)
- Cells contain DNA
- Sex of an individual is determined at fertilization by presence or absence of X and Y *chromosomes* (DNA structures)
DNA Molecular Structure

- DNA strand is a polymer (molecule consisting of a number of similar units bonded together) of nucleotides held together by strong, covalent (shared electron pair) bonds.
- Each nucleotide consists of a sugar (deoxyribose), a phosphate (phosphorus+oxygen combination), and a base (molecule that can donate an electron pair).
- The sugars are not symmetrical -- at one end they are joined to the phosphate by the 5th carbon (the 5' end) and at the other by the 3rd carbon (3'); thus, DNA strands are directional.
Four bases encode information in DNA: adenine (A), guanine (G), cytosine (C), and thymine (T).

A and T are complementary, link via weak hydrogen bonds; same with G and C.

Double-stranded DNA links complementary bases; exactly complementary strands have opposite directions.

The two strands twist into a double helix.
DNA Replication

Process 1: a DNA helicase (enzyme) splits the double strand into two single strands

Processes 2,3 (one for each strand): a DNA polymerase (enzyme) uses the single strand nucleotides as a template to match up complementary free nucleotides and build the other strand

- Strands replicated differently as DNA polymerases only work in one direction
- DNAReplication.flv - [http://www.youtube.com/watch?v=5VefaI0LrgE&feature=grec_index](http://www.youtube.com/watch?v=5VefaI0LrgE&feature=grec_index)
Replication Errors

- Low error rate \((10^{-9})\) due to error correcting processes
- Dissociation during conformation -- for new nucleotide to be covalently bound to growing polymer, DNA polymerase must undergo conformational change -- incorrect nucleotide more likely to dissociate
- Exonucleolytic proofreading -- a mismatched nucleotide at the extension end of the new strand prevents further extension -- it will be clipped off by an exonucleolytic enzyme
- Strand-directed mismatch repair -- on the new strand, mismatches are recognized, excised, and resynthesized
- Yet, replication a significant source of mutations
## Genome Sizes

<table>
<thead>
<tr>
<th>Species</th>
<th>Common name</th>
<th>Million bases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protopterus aethiopicus</td>
<td>lungfish</td>
<td>139000</td>
</tr>
<tr>
<td>Fritillaria assyriaca</td>
<td>butterfly</td>
<td>124900</td>
</tr>
<tr>
<td>Triticum aestivum</td>
<td>wheat</td>
<td>16000</td>
</tr>
<tr>
<td>Nicotiana tabacum</td>
<td>tobacco</td>
<td>4400</td>
</tr>
<tr>
<td><strong>Homo sapiens</strong></td>
<td><strong>human</strong></td>
<td><strong>3200</strong></td>
</tr>
<tr>
<td>Mus musculus</td>
<td>mouse</td>
<td>2400</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>fruit fly</td>
<td>120</td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>mouse-ear cress</td>
<td>100</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td>roundworm</td>
<td>78</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>yeast</td>
<td>12</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>bacteria</td>
<td>5</td>
</tr>
</tbody>
</table>

- Lungfish genome would need a 38 bit address
- Data requires challenging Computer Science
Human Genome Composition

**Complete Composition**
- Repeated sequence: 27%
- Regulatory region: 12%
- Genes: 8%
- Heterochromatin (genetically inactive): 53%

**Repeated Sequence**
- Long repeats: 15%
- Short repeats: 34%
- Retrovirus: 51%
Genes

- A gene is a DNA segment corresponding to a protein

- Human genes made up of regulatory region, *exons* (coding DNA), and *introns* (noncoding DNA)

<table>
<thead>
<tr>
<th>Species</th>
<th>Common name</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Triticum aestivum</td>
<td>wheat</td>
<td>107K-334K</td>
</tr>
<tr>
<td><em>Homo sapiens</em></td>
<td>human</td>
<td>22K-30K</td>
</tr>
<tr>
<td>Mus musculus</td>
<td>mouse</td>
<td>20K</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>fruit fly</td>
<td>14K</td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>mouse-ear cress</td>
<td>21K</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td>roundworm</td>
<td>20K</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>yeast</td>
<td>6K</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>bacteria</td>
<td>4K</td>
</tr>
</tbody>
</table>
Gene Regulation

- All cells in an individual have the same genome.
- Differences due to gene regulation -- activating or deactivating gene(s) under various conditions.
- Development during gestation is also due to gene regulation.
- Gene regulation makes complex, multicellular organisms possible.
- An *operon* is a group of genes controlled by a single, regulatory region.
Gene Regulation

- Gene regulatory proteins switch genes on (activators) or off (repressors) by binding to an area of the DNA regulatory region for the gene.
- Genes similar to content addressable memory, with activator similar to tag.
- Gene regulatory region also contains a promoter at the start of the gene’s coding DNA.
- Regulatory region for a Drosophila gene is 20K bases long and has sites for > 20 regulatory proteins.
Proteins

- An amino acid is a molecule that contains an acid group, an amine (nitrogen+hydrogen) group, and a side chain that varies with the specific acid.
- A protein is an amino acid polymer.
- 20 different amino acids are used to make proteins.
- Proteins typically fold into complex, 3D shapes.
- Proteins typically have reactive sites on their surface.
- Structural proteins make up a cell’s structure -- e.g. keratins for skin, hair, nails.
- Enzymes catalyze (increase the speed of) chemical reactions -- e.g. lactase converts lactose to glucose.
- Nucleic acid binding -- e.g. gene regulation.
- Signal transduction -- extracellular stimuli causing intracellular change; e.g. photon hitting retina cell.
Hemoglobin

- Protein complex of 2 alpha globins and 2 beta globins
- Carries oxygen to tissues
- Releases oxygen where cell metabolism is high
- Highly conserved
Protein Synthesis - Transcription

- Gene activation begins protein synthesis

**Transcription:**

- RNA polymerase (also called transcriptase) binds to DNA at promoter
- RNA polymerase reads single strand of DNA and synthesizes corresponding single-stranded messenger RNA (mRNA)
- RNA is similar to DNA
  - uses *ribose* sugar instead of deoxyribose
  - uses *uracil* (U) base instead of thymine (T) base
Protein Synthesis - RNA Processing

- **RNA processing:**
  - introns spliced out of RNA
  - allows alternate splicing (more than one protein from a given gene)
  - Code reuse / hacking new function
**Protein Synthesis - Translation**

- **Translation:**
  - **Ribosome** (RNA/protein complex) reads mRNA
  - mRNA bases are read in triples known as *codons*
  - Each codon corresponds to one of the 20 amino acids or the stop code which halts translation
  - $4^3$ codons map to 20 amino acids + stop => code is redundant, robust
  - Ribosome synthesizes corresponding protein by polymerizing appropriate amino acids
Protein Synthesis

- ProteinSynthesis.flv
- http://www.youtube.com/watch?v=D3fOXt4MrOM
Ribosome $\neq$ Turing Machine

- Protein synthesis within a cell highly parallel -- many ribosomes / cell
- Ribosome only moves 1 direction on mRNA
- Ribosome doesn’t write mRNA
Cell Division

- Asexually reproducing organisms
  - unpaired chromosomes
  - mitosis cell division only
- Sexually reproducing organisms
  - *diploid* genome -- chromosomes are paired, one in each pair from each parent
  - mitosis cell division
  - meiosis cell division -- producing *gametes* (egg and sperm cells) with *haploid* (unpaired) genomes
  - haploid egg and haploid sperm combine during fertilization to produce new, diploid cell
Mitosis

1. DNA replicated
2. Chromosome(s) created from replicated DNA
3. Nucleus divides to produce daughter nuclei, each with a complete genome
4. Cell divides to produce daughter cells
Meiosis

1. DNA replicated in diploid cells
2. Chromosomes created -- crossover, if any, occurs during this phase
3. Nucleus divides to produce daughter nuclei, each with a diploid genome
4. Cell divides to produce daughter cells
5. 2 nuclei divide again to produce 4 daughter nuclei, each with a haploid genome
6. 2 cells divide to produce 4 daughter cells
Crossover in Meiosis

- Crossover occurs sometimes during meiosis chromosome duplication.
- New chromosomes are produced that consist of the corresponding part of each parent’s chromosome.
- Adaptability produced is significant advantage to sexual reproduction.
- Analogous to two different coders dividing up methods for a class.
Gene Variations

- A given gene may have variations known as alleles.
- The specific allele (or pair of alleles in diploid genomes) an individual has is its genotype.
- The appearance or behavior it has due to its genotype is its phenotype.
- If the two alleles in a pair are identical, the individual is homozygous; otherwise, it is heterozygous.
- Heterozygous sickle cell genotype (AS) helps prevent malaria.
Viruses

- *Viruses* are non-living parasitic complexes that depend on the mechanisms of their hosts, e.g. ribosomes, to reproduce.
- Viruses consist of genetic material (DNA or RNA) and an enclosing protein coating.
- *Retroviruses* use a reverse transcriptase to put their genes into their host’s genome.
  - Can be deadly, e.g. HIV.
  - Can be useful for gene therapy.
- Some viruses, called *oncoviruses*, can cause cancer.

Influenza virus
References
