CLASS 1.3: 01/25/07
GENOME ORGANIZATION I

A. The Content of the Genome:

1. Terms used in genomic analyses
   a. Genome =
   b. Transcriptome =
      - Defined in terms of RNA molecules
      - Usually more RNAs than genes because of
   c. Proteome =
      - can be more than the transcriptome because of
   d. Interactome =

2. Mapping Genomes:
   a. Linkage maps
      - Make rough estimates about
   b. Restriction Maps
      - Linkage maps can be generated using
      - Restriction fragments more accurately measured
   c. Sequence maps:
      - Ultimate map based
      - Can now compare wildtype DNA with that
      - With current technology (computer algorithms) can identify
3. Polymorphisms
   - Originally identified because of multiple alleles at a locus =
   - Histocompatibility (HLA) antigens
   - Polymorphisms not limited
   - When restriction fragments and DNA sequences are compared =

   a. Type of polymorphisms:
      i. Single nucleotide polymorphism (SNP) = a change that is a single nucleotide when DNA sequence of alleles is compared
      ii. Restriction-fragment length polymorphism (RFLP) = a type of SNP that is located at a restriction enzyme recognition site
         - results in an altered fragment when the DNA is cut and displayed on a gel.
b. Polymorphisms as linkage markers
   - Can be identified
   - Can be used like any genetic marker

- SNPs and RFLPs occur very often in the genome =

- Each individual has a unique constellation of

  i.
  ii.

- Haplotype is the combination of polymorphisms
4. Genome sizes
   a. C value
      - Total amount of DNA in haploid genome =
      - Genomes range

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Species</th>
<th>Genome (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algae</td>
<td>Pyrenomas salina</td>
<td>$6.6 \times 10^5$</td>
</tr>
<tr>
<td>Mycoplasma</td>
<td>M. pneumoniae</td>
<td>$1.0 \times 10^6$</td>
</tr>
<tr>
<td>Bacterium</td>
<td>E. coli</td>
<td>$4.2 \times 10^6$</td>
</tr>
<tr>
<td>Yeast</td>
<td>S. cerevisiae</td>
<td>$1.3 \times 10^7$</td>
</tr>
<tr>
<td>Slime mold</td>
<td>D. discoideum</td>
<td>$5.4 \times 10^7$</td>
</tr>
<tr>
<td>Nematode</td>
<td>C. elegans</td>
<td>$8.0 \times 10^7$</td>
</tr>
<tr>
<td>Insect</td>
<td>D. melanogaster</td>
<td>$1.8 \times 10^8$</td>
</tr>
<tr>
<td>Bird</td>
<td>G. domesticus</td>
<td>$1.2 \times 10^9$</td>
</tr>
<tr>
<td>Amphibian</td>
<td>X. laevis</td>
<td>$3.1 \times 10^9$</td>
</tr>
<tr>
<td>Mammal</td>
<td>H. sapiens</td>
<td>$3.3 \times 10^9$</td>
</tr>
</tbody>
</table>

- Generally
- Not always
- **C-value paradox** = lack of direct correlation
Some amphibians have
- others have sizes
- Likely to need the same

b. Types of eukaryotic genomic sequences
   i. Non-repetitive
   -
   -
   ii. Moderately repetitive
   - multiple copies
   - can be members
   iii. Highly repetitive
   - found in
   -
   - often organized
   - satellite DNA =
   - minisatellite or minisatellite =
- Proportion of genome that is
- Much of the repetitive DNA =
- Because repetitive DNA has same/similar sequence,

c. Transposons
   - Can move
   - Can make
   - Often called selfish DNA =
     - Because they represent a repetitive DNA sequence =

   - Can be a
5. Identifying disease-causing genes
   - Often it is clear that disease

   - Can follow the mutant allele (disease-causing gene) through a pedigree and show

   - Problem is finding the

   - If you know what protein is altered or missing -

   - In the old days (5-10 years ago) had to establish strong linkage with disease and

   - Easier today

Example: Duchenne’s muscular dystrophy

- Affects
- Steps:
  a. Find region

b. Perform restriction analysis of DNA and look for DNA in patients that is missing or rearranged =
c. Perform a “zoo” blot to see if similar gene expressed in other mammals =

d. Use probe to screen cDNA library =

e. Conceptionally translate cDNA

f. Compare cDNA with genomic DNA

g. Sequence DNA of all DMD patients to

6. Not all DNA is found in the nucleus
   - First clues =
   - Non-Mendelian inheritance =

   - Occurs because gene controlling trait

   - In a fertilized zygote =

   - Examples =

   - Organelle DNA evolves at rate

   - Usually no

   - Using mit genome comparisons =
a. Mitochondrial genomes (mtDNA)
   - Small circular molecules =
   - Usually have several
   - Several 100
   - Mitochondrial genomes
   - Types of genes
     i. Components of the
     ii. Components of protein

   - Not all proteins that make up the mitochondria

<table>
<thead>
<tr>
<th>Species</th>
<th>Size (kb)</th>
<th>Protein-coding genes</th>
<th>RNA-coding genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fungi</td>
<td>19–100</td>
<td>8–14</td>
<td>10–28</td>
</tr>
<tr>
<td>Protists</td>
<td>6–100</td>
<td>3–62</td>
<td>2–29</td>
</tr>
<tr>
<td>Plants</td>
<td>186–366</td>
<td>27–34</td>
<td>21–30</td>
</tr>
<tr>
<td>Animals</td>
<td>16–17</td>
<td>13</td>
<td>4–24</td>
</tr>
</tbody>
</table>

b. Chloroplast genomes (ctDNA)
   - Circular molecules =
   - Multiple copies of DNA/organelle;

   - Types of genes:
     i. All rRNA and tRNA genes
     ii. Genes for 50-100 proteins
        - about 1/2 are needed for
        - others involved

<table>
<thead>
<tr>
<th>Genes</th>
<th>Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA-coding</td>
<td></td>
</tr>
<tr>
<td>16S rRNA</td>
<td>1</td>
</tr>
<tr>
<td>23S rRNA</td>
<td>1</td>
</tr>
<tr>
<td>4.5S rRNA</td>
<td>1</td>
</tr>
<tr>
<td>5S rRNA</td>
<td>1</td>
</tr>
<tr>
<td>tRNA</td>
<td>30–32</td>
</tr>
<tr>
<td>Gene expression</td>
<td></td>
</tr>
<tr>
<td>r-proteins</td>
<td>20–21</td>
</tr>
<tr>
<td>RNA polymerase</td>
<td>3</td>
</tr>
<tr>
<td>Others</td>
<td>2</td>
</tr>
<tr>
<td>Chloroplast functions</td>
<td></td>
</tr>
<tr>
<td>Rubisco and thylakoids</td>
<td>31–32</td>
</tr>
<tr>
<td>NADH dehydrogenase</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>105–113</td>
</tr>
</tbody>
</table>
c. Organelles evolved by endosymbiosis
   - Sequence comparisons between mtDNA and ctDNA and contemporary bacteria

   - Organellar genomes have lost many genes =

   - Transfer of proteins encoded by nuclear genes

B. Genome sequences and gene numbers:
   1. Introduction
      - First genomes
      - Since then, technology has improved and

      - Check out www.genomanewsnetwork.org
      - Types of organisms being sequenced
        a. Medically and economically

        b. Evolutionarily relevant

      - Most useful information with current algorithms =
- Some generalizations are possible (Fig 5.3)
- Many organisms have more genes than
- Organisms with more genes usually have redundant copies =
- Number of gene families may
- Three Kingdoms of Life

2. Prokaryotic (bacterial) genomes
   - Most of the DNA
   - Good correlation between

   - Average gene size =

   - Types of bacteria
     a. Obligate parasites
        - Have smallest genomes

        - 1500-2700 genes
        - Missing many genes
        - Get their metabolic products
b. Free-living bacteria
   - Have genome sizes ranging
   - Smallest need
   - Largest (Nitrogen fixing bacteria) need
   - *E. coli* (middle) =
   - Most genes fall equally into 2 categories
     i. Involved in
     ii. Involved in

2. Archaea genomes
   - Separate
   - Have biological properties
     - gene expression apparatus
     - cell division
   - Have genome sizes

3. Eukaryotic genomes
   - Unicellular eukaryotic genomes are
   - Higher eukaryotes have more genes,