Genetics: The Study of Biological information

- All organisms must do the following with vast amounts of bio-information:
  - store it
  - replicate it
  - transmit it during cell division
    - one cell -> two daughter cells
  - use it to develop, grow, reproduce and survive
  - transmit it to their progeny

- the book emphasizes 6 overarching themes:
  - biological info is stored as DNA
  - biological functions are dictated primarily from the shape and charge of proteins
  - all living things are related in the mechanisms by which biological information is stored, replicated etc.
  - modular construction of genomes (i.e., one set of the chromosomes) allows rapid evolution of inter-specific (among species) bio-complexity
  - genetic techniques allow dissection of biological complexity within and among species
  - the focus of the book is on humans, but ours is on plants
Biological Information is Encoded in the DNA Molecule

- single strand of DNA comprised of a non-branching chain comprised of alternating deoxyribose (a sugar) and phosphate molecules
- the two ends of one strand are different- one end has a P atom (5’ end), the other does not (3’ end)
- the two strands are in opposite orientation
- one of 4 nitrogenous bases is attached to each sugar molecule
- two bases have two rings and are called purines (G and A), other two are pyrimidines (C and T)
- the two strands align so that each base on one strand is directly opposite and weakly bound to a base on the other strand
- base pairing rules are A:T, G:C [purine:pyrimidine always]
- “A” is said to complement “T”, and “G” complements “C”
- two strands are therefore complementary, and knowledge of the base sequence of one strand precisely predicts the base sequence of the second strand
- double stranded, complementary DNA forms a double helix (see fig. 5.9) with the bases on the inside and the sugar-phosphate backbones of the two strands on the outside
Double Stranded DNA is composed of protein-encoding and noncoding regions

• One chromosome has either one or two double stranded DNA molecules, each with many hundreds of thousands of base pairs (bp, or nucleotide pairs or just nucleotides)
• Arabidopsis thaliana (see chapt 19) has a very small genome comprised of 5 chromosomes- their length in nucleotide pairs is:

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Apx. Chromosome Length (bp)</th>
<th># OF BAC’S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29,105,111</td>
<td>373</td>
</tr>
<tr>
<td>2</td>
<td>19,646,945</td>
<td>257</td>
</tr>
<tr>
<td>3</td>
<td>23,172,617</td>
<td>333</td>
</tr>
<tr>
<td>4</td>
<td>17,549,867</td>
<td>193</td>
</tr>
<tr>
<td>5</td>
<td>25,953,409</td>
<td>407</td>
</tr>
</tbody>
</table>

• above information derived from this website http://www.tigr.org/tdb/ath1/htmls/ath1.html
• Wheat has a genome of at least 1.6 billion bp and 21 chromosomes, i.e., an average length of 76 million bp per chromosome!
• wheat and arabidopsis probably have a similar number of functional genes- the extra DNA in wheat is primarily noncoding sequences between genes
• genes also have noncoding regions (introns) interspersed among the coding regions (exons)
Proteins are comprised of one or more chains of amino acids
- there are 20 distinct amino acids, each with its own shape and charge

Enzymes are almost exclusively proteins (usually with several polypeptides)

Enzymes have immense catalytic power for one or more chemical reactions
- i.e., Rate of Substrate - product conversion accelerated

Exquisitely specific
Functionality derived from conformation (shape) and binding affinities
Amino acids vary in shape and affinities to bind other molecules
AA sequence of a protein therefore dictates conformation and binding affinities
- therefore specificity and catalytic power of enzymes

Enzymes preferentially bind other molecules including proteins and DNA
- Binding at an enzyme’s “active site” leads to catalysis
- Binding at an enzyme’s “regulatory sites” leads to changes in catalytic power

Ribulose-1,5-bisphosphate carboxylase from Spinacia oleracea is featured in this slide. (detail below will not be on quizzes or tests)

Ribulose-1,5-bisphosphate carboxylase contains 2992 hydrogen bonds, which force the 16 protein chains to assume 208 helices, 248 beta-strands and 456 turns. Just in case the pattern shown here is somewhat confusing, the structural elements of this enzyme will be demonstrated using a single subunit each.

The small subunit is made up of 123 amino acids. The protein contains a four-stranded antiparallel sheet, which is flanked by two helices. Some turns stabilize the loops.

The large subunit contains 475 amino acids (amino acids 1-8 and 464-475 are not visible in the X-ray determined structure). Two domains are to be discerned in the protein. The N-terminal domain figures the first 150 amino acids. This domain contains a four-stranded antiparallel sheet as well as some helices and turns.

The carboxy terminal domain is the catalytically important part of the molecule. It's structure is an eight-stranded alpha/beta barrel. The barrel arises from the alternating order of beta-strands and helices. The bottom of the barrel is closed by a helix. The opposite side of the barrel features the active site of the enzyme. Amino acids in the loops between the helices and strands making up the barrel coordinate a magnesium ion. To the other side of the magnesium ion the substrate is bound.

In the complete enzyme complex the active site is situated in the interface of two adjacent large subunits. In a spacefilling view of the protein the accessibility of the active center for the substrate is visible within the interface region (if you use the mouse to rotate the protein, you may get the channels straight to focus).


9-98 © R Bergmann
All living things are closely related

- Figure P.3 (above) shows the incredible similarity of the amino acid sequence (each letter represents one of the 20 amino acids) for the critical enzyme cytochrome C among humans, a plant, yeast, flies, mice, and nematodes.
- same 20 amino acids are used by all organisms, and all organisms use essentially the same genetic code relating DNA nucleotide sequence to protein amino acid sequence.
- in many instances, genes from one organism can be fully functional in another organism
  - genes for insect and herbicide resistance in transgenic or Genetically Modified Organisms (GMO) plants come from bacteria, where they have the same function.
Relatedness is evident in gene/marker maps: an example

- This figure shows “synteny” or conservation of order of DNA sequences on chromosomes of rice, wheat (and relatives) and oats.
- These species diverged millions of years ago.
Relationships among plants at a gene controlling development.

- Wheat, barley, rye and rice diverged several millions of years ago.
  - In other words, they have been reproductively isolated for millions of years, but still retain substantial “synteny” which means conservation of the order of sub-sequences.
- The figure above (from TAG, 1998, 107, page 103.) shows RLFP maps for chromosomes from these four species.
- Lines that connect chromosomes (vertical lines) show where “orthologous” loci reside.
- Orthologous means similar because of common origin- but now present in different species.
*eukaryotic vs prokaryotic- students should know the difference between these phyla*

- Eukaryotes have a membrane bound nucleus, mitosis, meiosis, semi-autonomous organelles (chloroplasts and mitochondria for plants, mitochondria only for animals)

*many protein encoding regions have multiple coding regions (exons) separated by noncoding regions (introns)*

- Exons often code for a subsequence of a protein (a domain) that has a particular functionality

*fig p.4 a (above) shows how gene duplication followed by divergence leads to a family of related genes by shuffling of exons and introns.*