In Eukaryotes, Genes are Often Much Larger than the Coding Region

<table>
<thead>
<tr>
<th>Gene Product</th>
<th>mRNA Size</th>
<th>Number of Introns</th>
</tr>
</thead>
<tbody>
<tr>
<td>b-Globin</td>
<td>1.5</td>
<td>2</td>
</tr>
<tr>
<td>Insulin</td>
<td>1.7</td>
<td>2</td>
</tr>
<tr>
<td>Protein kinase C</td>
<td>1.1</td>
<td>7</td>
</tr>
<tr>
<td>Albumin</td>
<td>25</td>
<td>14</td>
</tr>
<tr>
<td>Catalase</td>
<td>34</td>
<td>12</td>
</tr>
<tr>
<td>LDL receptor</td>
<td>45</td>
<td>17</td>
</tr>
<tr>
<td>Factor VIII</td>
<td>186</td>
<td>25</td>
</tr>
<tr>
<td>Thymoglobulin</td>
<td>300</td>
<td>36</td>
</tr>
<tr>
<td>Dystrophin*</td>
<td>more than</td>
<td>more than</td>
</tr>
<tr>
<td>2000</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

*An altered form of this gene causes Duchenne muscular dystrophy.
The size specified here for a gene includes both its transcribed portion and nearby regulatory DNA sequences. (Compiled from data supplied by Victor McKusick.)

Considerations for Genome Sequencing

1. Satellite DNA is very difficult to sequence, as there are few markers to help order subclones; hence centromeric regions of the chromosomes are usually left unsequenced.

2. Middle repetitious DNA also causes difficulties; because one finds nearly identical sequences located in different regions of the genome, mistakes can be made in assembling sequence data. High quality discrepancies can identify these.

3. Much of the repetitious DNA is packaged in heterochromatin, which maintains these regions in a compact and transcriptionally silent form.

4. However, in many higher organisms, protein-coding genes are found embedded in repetitious DNA. Check out your favorite human gene on the UCSC Browser by taking off RepeatMasker!

Eukaryotic Cells - Coping with Large Genomes

Mammals: 3,000,000,000 bp
2 meters of DNA /cell

Only 2% codes for proteins!

Much of the DNA is repetitious.

Assembly into chromatin leads to effective packaging and can lead to gene silencing.

CHROMATIN DNA and its associated proteins

Nucleosome
- the basic subunit of chromatin
  - DNA: 167 bp + linker
  - Histone core: 1 H3/H4 tetramer + 2 H2A/H2B dimers
  - Linker histone H1 or H5

Histone protein core


Electron Micrograph of Chromatin Fibers (rat thymus nucleus)

Olins et al., 1975 J. Cell Biol, 64:528

“A eukaryotic chromosome made out of self-assembling 70A units, which could perhaps be made to crystallize, would necessitate rewriting our basic textbooks on cytology and genetics! I have never read such a naïve paper purporting to be of such fundamental significance. Definitely it should not be published anywhere!”


Quoted in "Chromatin" by K.D. van Holde, 1989

The Structure of the Nucleosome Core

Resolution: 2.8 Å
Half of the nucleosome structure is shown
One turn of the DNA helix is visible (73 bp)
View is down the superhelix axis
Protein - DNA contact: white hooks


The Histone Octamer

The complete histone octamer in the absence of DNA.
The unstructured N-terminal tails are extensively modified.
Color code:
- H2A
- H2B
- H3
- H4

Rhodes, 1997 Nature 389:231

DNA Packaging Domains

Euchromatin
- Less condensed
- Chromosome arms
- Unique sequences; gene rich
- Replicated throughout S
- Recombination during meiosis

Heterochromatin
- Highly condensed
- Centromeres and telomeres
- Repetitious sequences; gene poor
- Replicated in late S
- No meiotic recombination

Transcriptional activators
Hyper-acetylated histone tail
Heterochromatin Protein 1 complex
Hypo-acetylated histone tail; methylated H3/K9

Fruit Flies are Inexpensive and Easy to Culture

- Short life cycle
- Easily visible phenotypes
- Polytenic chromosomes
- Simple genome
- Fully sequenced genome
- Metazoan useful for behavioral, developmental and human disease research

Created by K.A. Haynes

Copyright © 2006, Washington University in St. Louis
Evolution of genes and genomes on the Drosophila phylogeny

Drosophila 12 Genomes Consortium

Comparative analysis of multiple genomes in a phylogenetic framework dramatically improves the precision and sensitivity of evolutionary inference, producing results that single-genome analyses cannot provide. The genomes of 12 Drosophila species, representing phylogenetic lineages, are sequenced and assembled and the genome sequences are compared with those of 32 other diverse species, representing major eukaryotic lineages, to reveal genetic and evolutionary mechanisms underlying the organismal properties that define Drosophila as a model organism. These genome sequences represent the foundation of genetic resources that have made Drosophila melanogaster a pre-eminent model for animal genetics, and will further refine fundamental research on mechanisms of development, cell biology, genetics, disease, neurobiology, behavior, echonomyology, and evolution. Despite remarkable similarities among these Drosophila species, we identified many potentially non-neutral changes in protein-coding genes, non-coding RNA genes, and cis regulatory regions. These may prove to underlie differences in the ecology and behavior of these diverse species.

Bio 4342: Research Explorations in Genomics

GOAL
Students work as a research team through a large-scale sequencing project to finish and annotate dot chromosome

PROCESS
Data generation, finishing and quality control in collaboration with the WU Genome Sequencing Center; complete annotation and analysis collaborating with WU Computer Science faculty.
Almost all of the same genes are present (27/28), but rearrangements within the chromosome are common!

Dot Chromosomes Genes Have Larger Introns Due to Repetitious DNA

Summing up…..

- Eukaryotic genomes are:
  - unexpectedly large, complex
  - contain a high percentage of repetitious sequences

- Much of the genome is packaged in heterochromatin:
  - formation may be targeted by repetitious sequences triggering an RNAi response
  - leads to alternative chromatin packaging
  - results in gene silencing
  - can impact nearby genes

- Organization of the genome, patterns of repetitious DNA are critical for genome function
  - can be studied through analysis of the dot chromosome