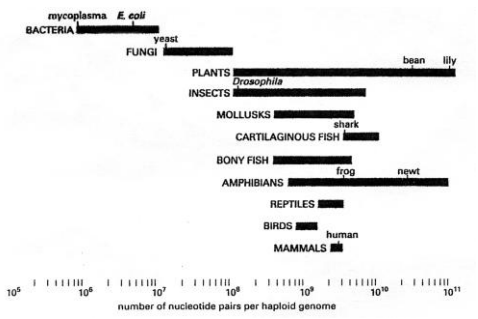
The Genomics Education Partnership

Genomes, Heterochromatin and the Dot Chromosome:

Comparative Genomics in *Drosophila*

Funded by the Howard Hughes Medical Institute

Lack of Relationship Between Amount of DNA and Organism Complexity



Alberts et. al., Molecular Biology of the Cell, Third Edition, 1994.

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

Table 8–1 The Size of Some Human Genes in Thousands of Nucleotides

	Gene Size	mRNA Size	Number of Introns
β-Globin	1.5	0.6	2
Insulin	1.7	0.4	2
Protein kinase C	11	1.4	7
Albumin	25	2.1	14
Catalase	34	1.6	12
LDL receptor	45	5.5	17
Factor VIII	186	9	25
Thyroglobulin	300	8.7	36
Dystrophin*	more than 2000	17	more than 50

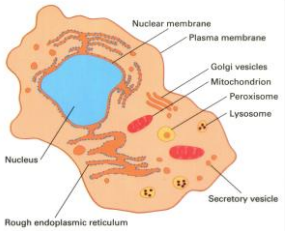
*An altered form of this gene causes Duchenne muscular dystrophy.
The size specified here for a gene includes both its transcribed portion and nearby regulatory DNA sequences. (Compiled from data supplied by Victor McKusick.) Alberts et al #3

Alberts et. al., Molecular Biology of the Cell, Third Edition, 1994.

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 repetitous 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.

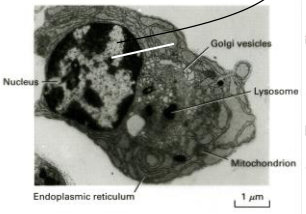
Assembly into chromatin leads to effective packaging and can lead to gene silencing.

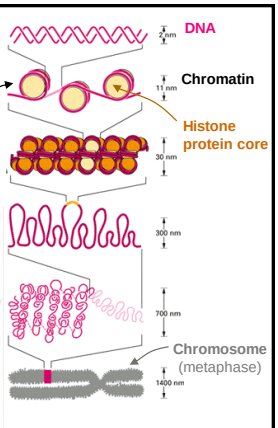
Lodish et.al., Molecular Cell Biology, 4th Edition, 2000

CHROMATIN

DNA and its associated proteins

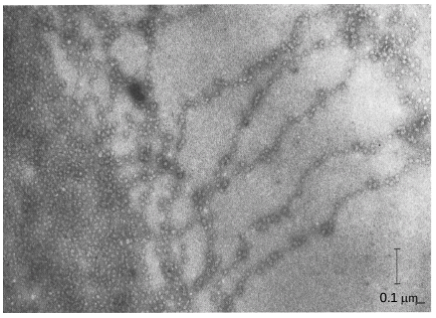
Nucleosome
- the basic subunit of chromatin
  - **DNA**: 167 bp + linker
  - **Histone core**:
    1 H3/H4 tetramer
    + 2 H2A/H2B dimers
  - **Linker** histone H1 or H5





Lodish et.al., Molecular Cell Biology, 4th Edition, 2000 Felsenfeld et al., 2003 Nature, 421:448

Electron Micrograph of Chromatin Fibers (rat thymus nucleus)



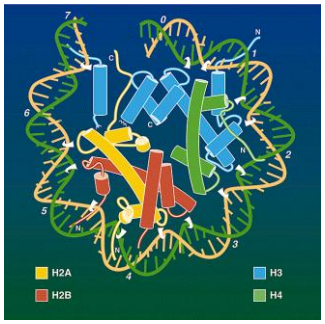
Olins et. al., 1975 J. Cell Biol, 64:528

"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!"

Anonymous review of paper submitted by C.F.L. Woodcock, 1973, showing EM pictures of nucleosome arrays.

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

The Structure of the Nucleosome Core



Rhodes, 1997 Nature 389:231, after Luger et. al., 1997 Nature 389:251

Resolution: 2.8 Å

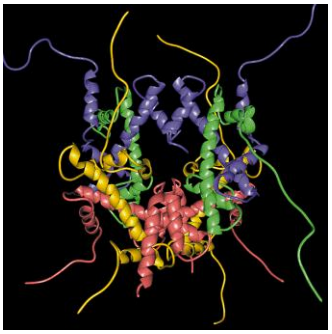
Half of the nucleosome structure is shown

One turn of the DNA helix is visible (73 bp)

View is down the superhelix axis

Protein - DNA contact: white hooks

The Histone Octamer



Rhodes, 1997 Nature 389:231

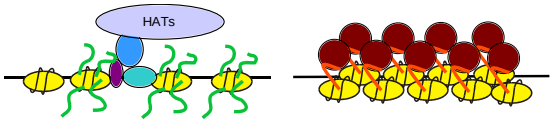
The complete histone octamer in the absence of DNA.

The unstructured N-terminal tails are extensively modified.

Color code:

- **H2A**
- **H2B**
- **H3**
- **H4**

DNA Packaging Domains



Euchromatin

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

Heterochromatin

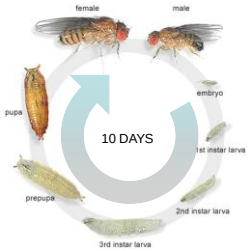
- Highly condensed
- Centromeres and telomeres
- Repetitious sequences; gene poor
- Replicated in late S
- No meiotic recombination

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
- Polytene chromosomes
- Simple genome
- Fully sequenced genome
- Metazoan useful for behavioral, developmental and human disease research

Assembled by K.A. Haynes

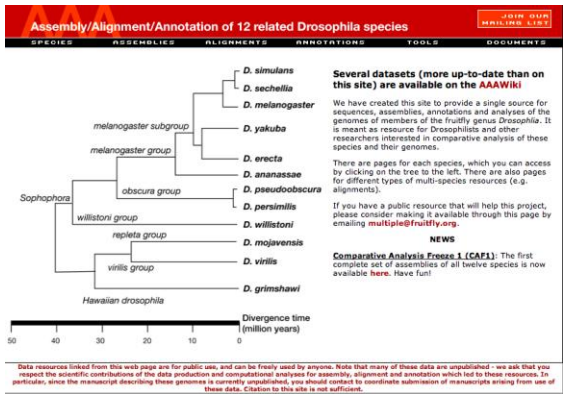
Vol 450|8 November 2007|doi:10.1038/nature06341 nature

ARTICLES

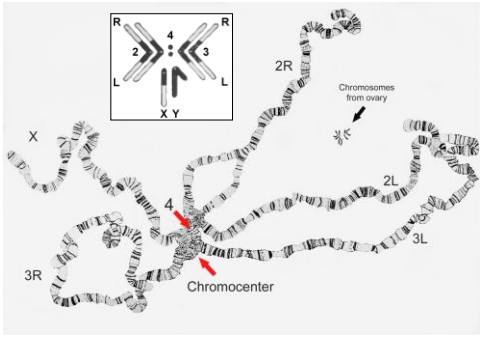
Evolution of genes and genomes on the *Drosophila* phylogeny

Drosophila 12 Genomes Consortium*

Comparative analysis of multiple genomes in a phylogenetic framework dramatically improves the precision and sensitivity of evolutionary inference, producing more robust results than single-genome analyses can provide. The genomes of 12 *Drosophila* species, ten of which are presented here for the first time (*sechellia*, *simulans*, *yakuba*, *erecta*, *ananassae*, *persimilis*, *willistoni*, *mojavensis*, *virilis* and *grimshawi*), illustrate how rates and patterns of sequence divergence across taxa can illuminate evolutionary processes on a genomic scale. These genome sequences augment the formidable genetic tools that have made *Drosophila melanogaster* a pre-eminent model for animal genetics, and will further catalyse fundamental research on mechanisms of development, cell biology, genetics, disease, neurobiology, behaviour, physiology and evolution. Despite remarkable similarities among these *Drosophila* species, we identified many putatively 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 behaviour of these diverse species.

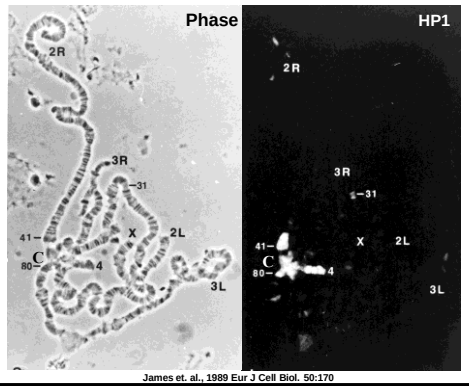


Drosophila melanogaster Chromosomes



modified from TS Painter, 1934, J. Hered 25:465

Distribution of Heterochromatin Protein 1 in *D. melanogaster*



James et. al., 1989 Eur J Cell Biol. 50:170

Bio 4342: Research Explorations in Genomics

GOAL

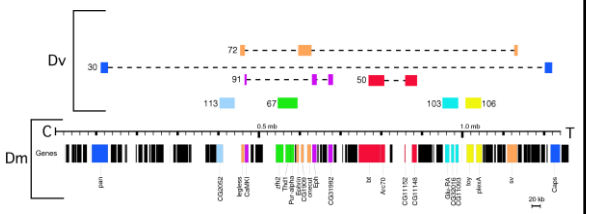
Students work as a research team through a large-scale sequencing project to finish and annotate dot chromosome

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.

Common techniques, individual projects, whole assembled

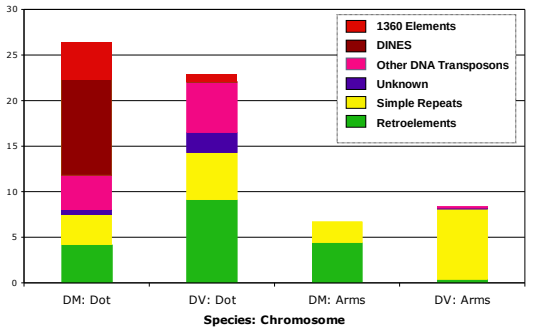
Initial Analysis of *Drosophila virilis* Dot Chromosome Fosmids



Almost all of the same genes are present (27/28), but rearrangements within the chromosome are common!

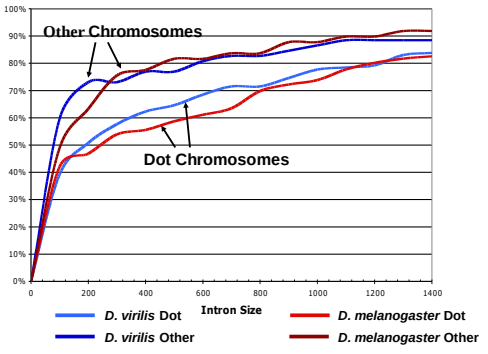
Slawson et. al., 2006 Genome Biology, 7(2):R15.

Dot Chromosomes of Both Species Have a High Density of Repetitious Elements; Types of Elements Differ



Slawson et. al., 2006 Genome Biology, 7(2):R15.

Dot Chromosomes Genes Have Larger Introns Due to Repetitious DNA

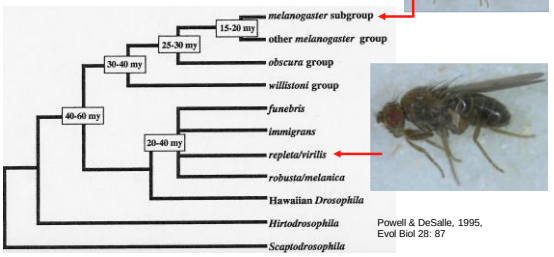


Slawson et. al., 2006 Genome Biology, 7(2):R15.

Our Research Goal

To compare finished sequence from the dot chromosomes of different Drosophila species.

Next targets: *D. grimshawi* and *D. mojavensis* *(in Hawaiian subgroup)*



Powell & DeSalle, 1995, Evol Biol 28: 87

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**