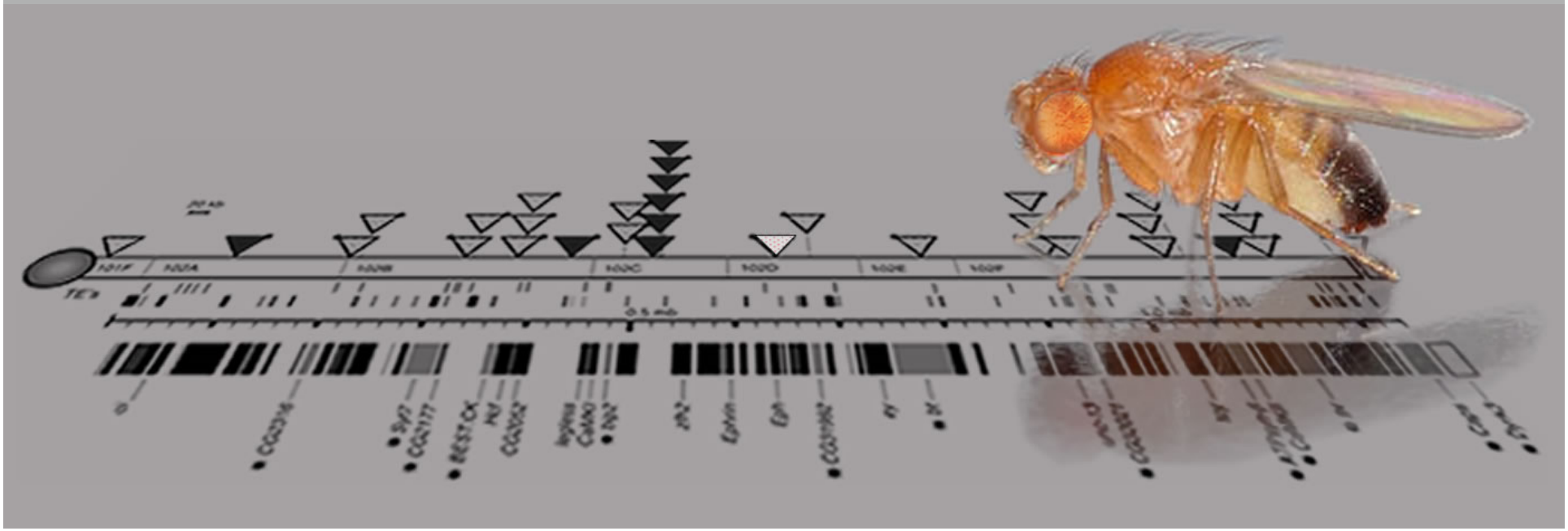
Finding Genes in a New Fly Genome: Teaching about Genes/Genomes via Bioinformatics Research

Sarah Elgin, Anya Goodman, Wilson Leung
Eric Tsoi, Charlene Emerson, David Carranza

January 2012

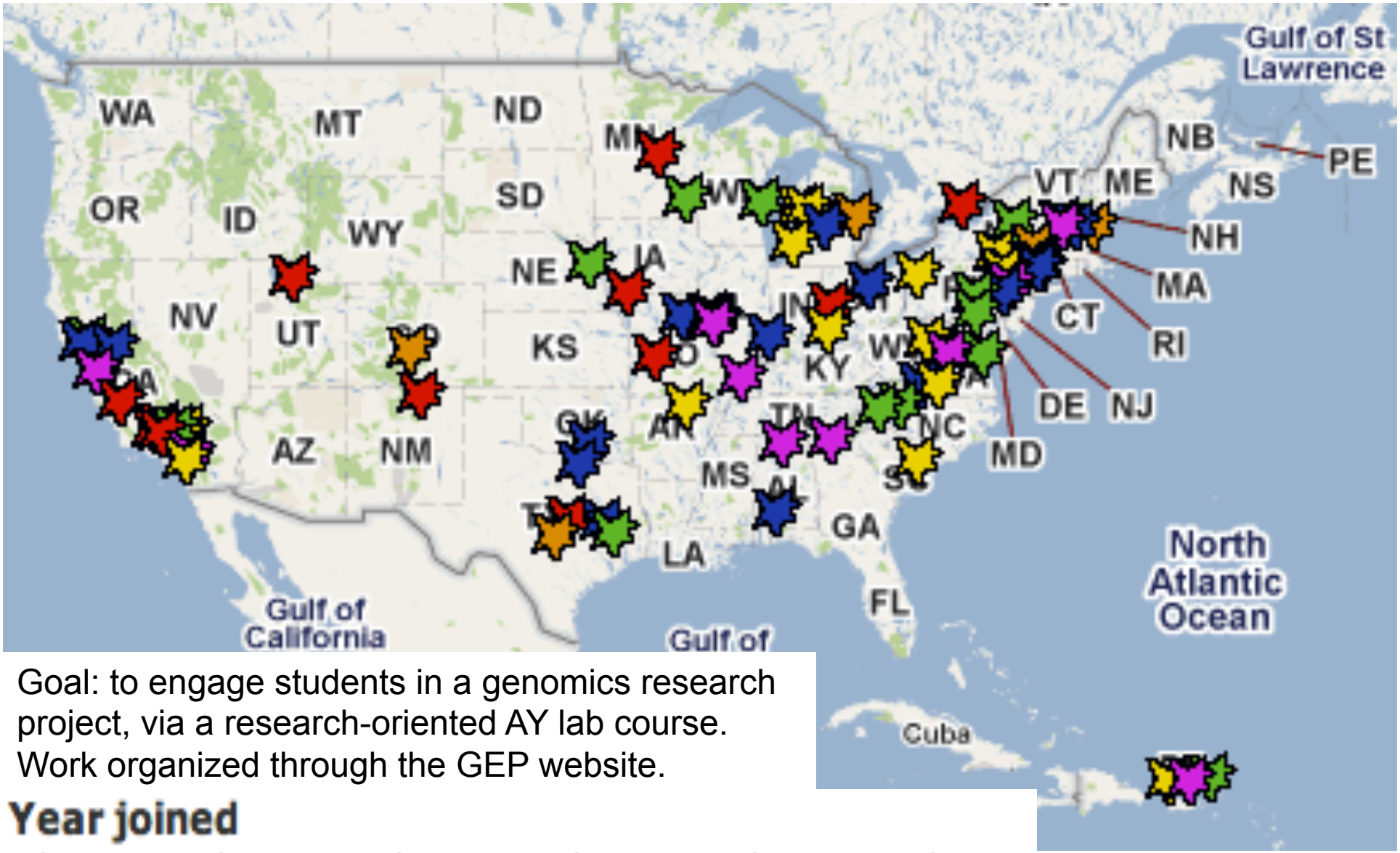


Workshop Goals

- Introduce Genomics Education Partnership
- Hands-on practice with genome annotation
- Discussion of curriculum options
  3-week module (~10 hr: 1 lecture, 3X3 lab)
  5-week – add a more difficult project
  10-week - real research!
- Scientific background on Drosophila genome

- http://gep.wustl.edu selgin@biology.wustl.edu
- Next workshops: June 24-26, August 19-22
  (HHMI supported)

Genomics Education Partnership (GEP)



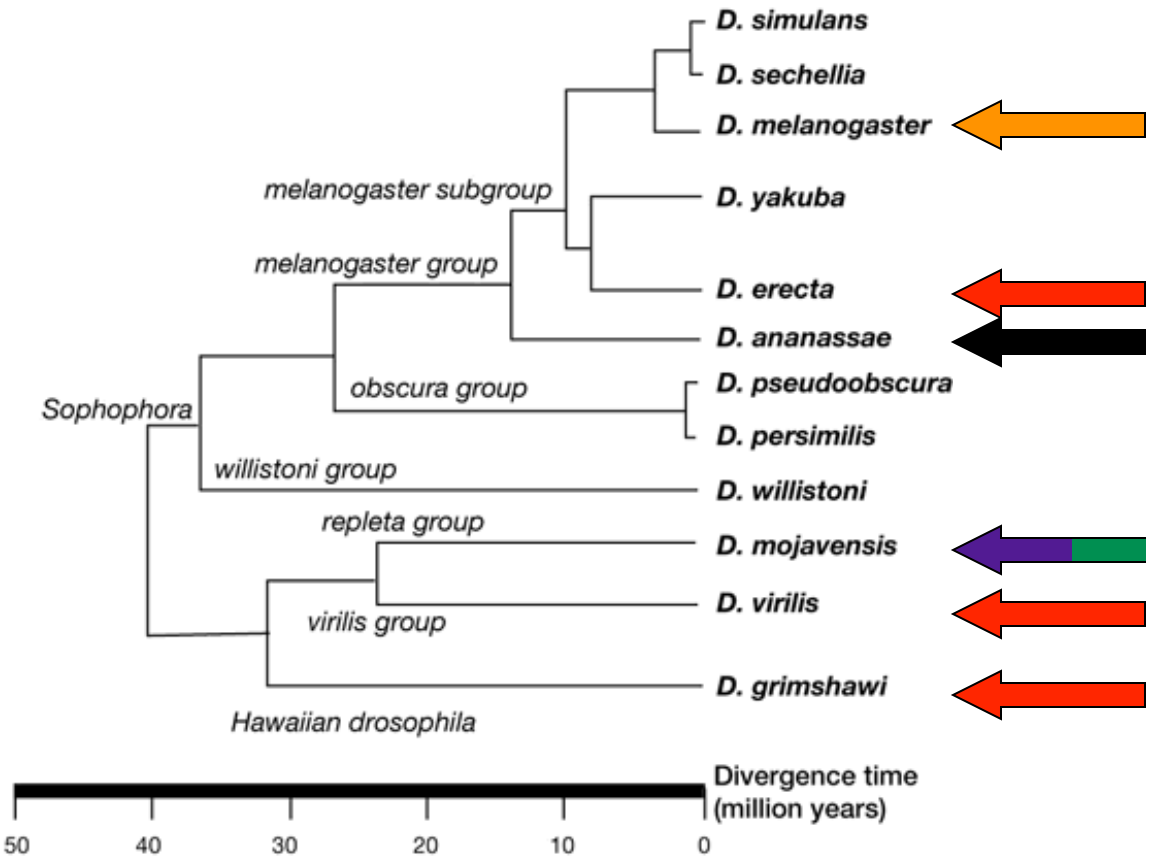
Goal: to engage students in a genomics research project, via a research-oriented AY lab course. Work organized through the GEP website.

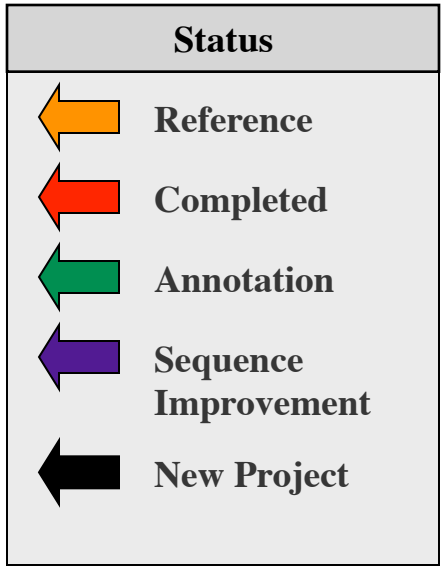
Year joined

2006 2007 2008 2009 2010 2011

Our GEP research goal:

Use comparative genomics to learn more about heterochromatic domains, analyzing the dot chromosomes and a control euchromatic region of Drosophila genomes





FlyBase: http://flybase.org

I hear and I forget.

I see and I remember.

I do and I understand

Confucius

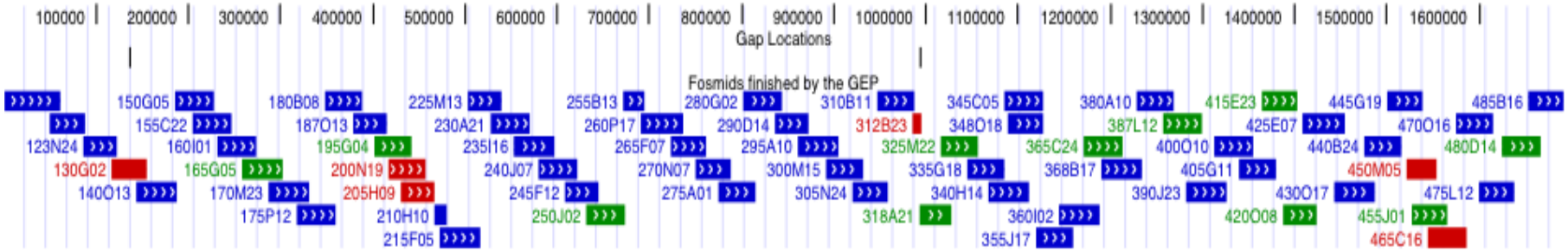
The scientific method allows ordinary people to do extraordinary things.

Francis Bacon

Genomics provides terrific opportunities to engage undergraduates in research!

Strategy: divide and conquer!
The *D. mojavensis* dot chromosome

- Students completed 68 projects covering 1.7 Mb closing 26/28 gaps, adding ~15,000 bp and improving ~5000 bp.
- Each project finished and annotated (all isoforms) twice; reconciliation for quality control done at Wash U

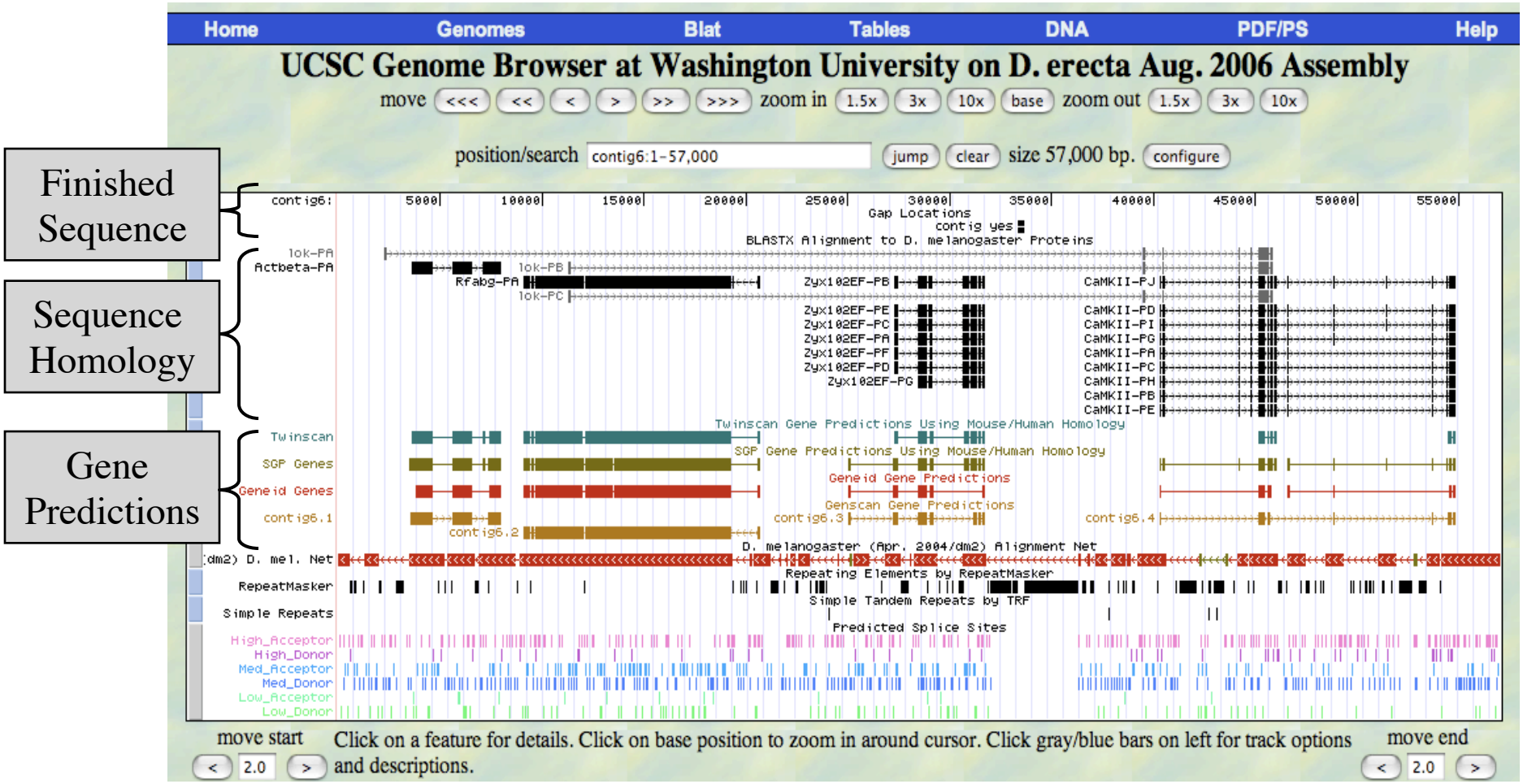


Fosmid sequence matches consensus sequence

Putative polymorphisms

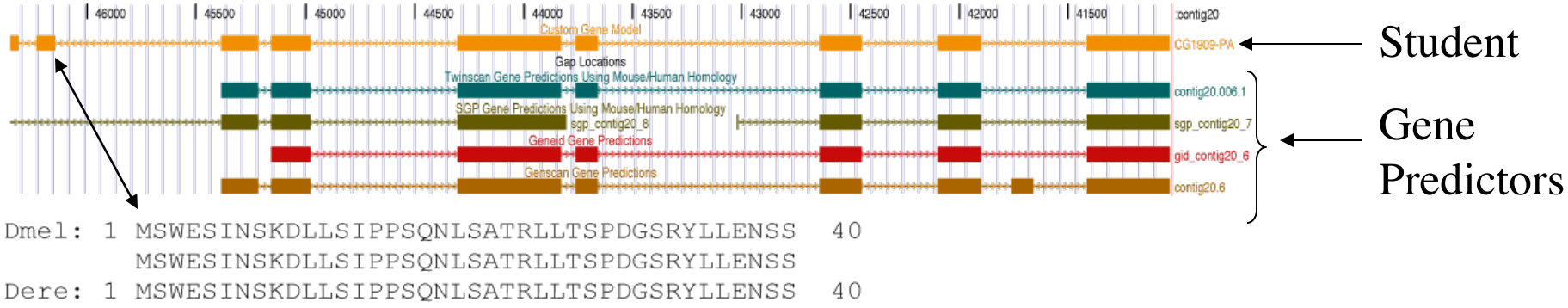
Finished sequences submitted to Genbank, annotations to Flybase.

Annotation: Create gene models using sequence homology and computational predictions



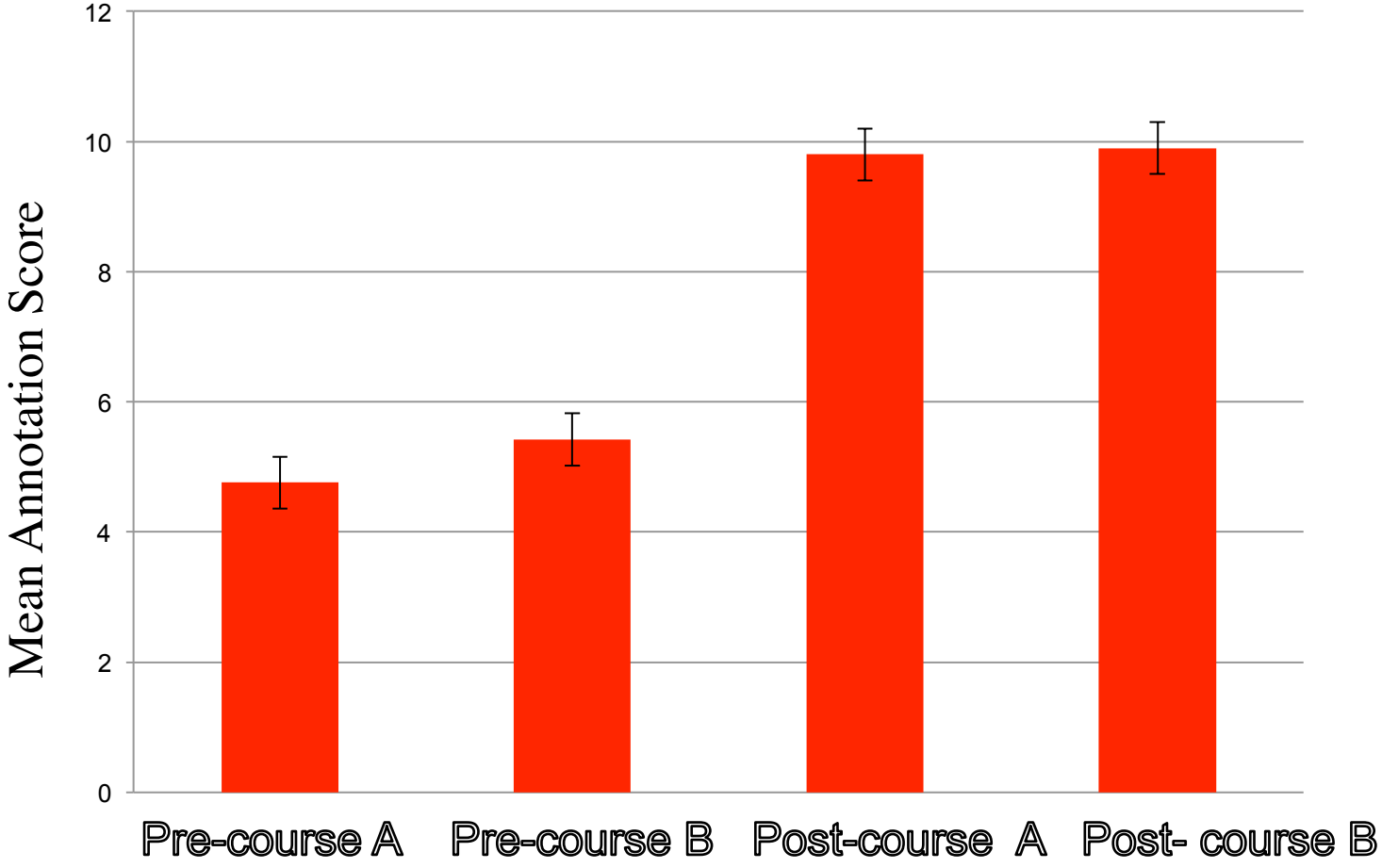
New! RNA seq tracks!

Learning from the annotation process…



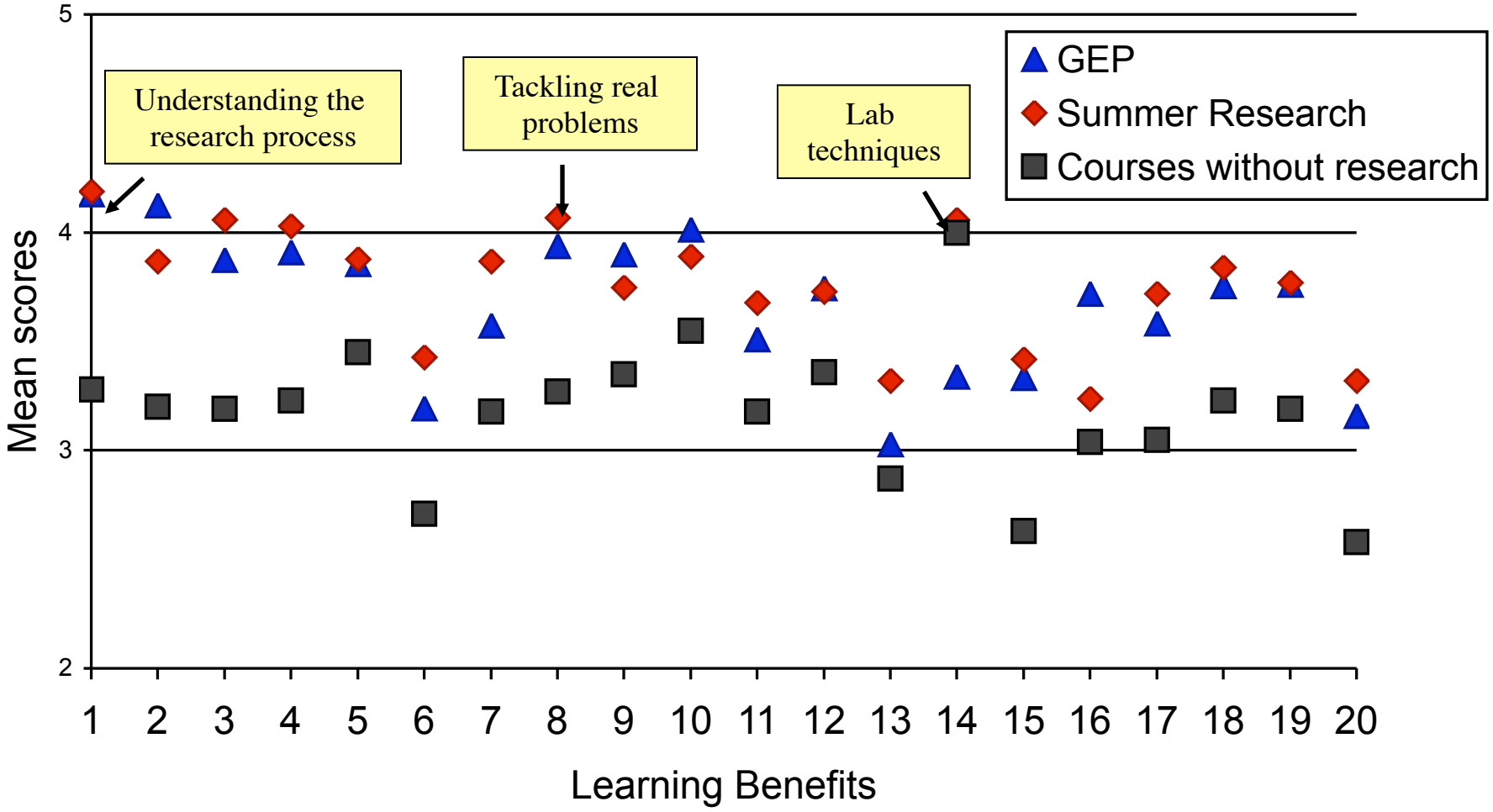
- Become familiar with tools available for finding genes; manage data
- Identify genes, create best models (start, exons, stop etc.)
- Requires synthesis of multiple lines of evidence
- Use power of comparative genomics; reference *D. melanogaster*
- Analyze genome organization (synteny), repeats
- Address questions of evolution
- Experience presenting data, supporting conclusions based on available evidence
- Students are making an original contribution
- Each project done twice independently, 60% - 75% congruence

But are they learning?
GEP annotation quiz results, 2010 - 2011



We see a positive correlation between quiz scores and self-reported gains.

GEP assessment: CURE survey (D Lopatto)



For most gains, a GEP semester course is as effective as a summer research experience!

Selected quotes…..

- The class was very intellectually challenging for me. It taught me to think in a way that I had never thought before.

- The thing that sparked me the most was the fact that I was able to perform the BLAST searches on my own and was able to explain to my instructor what I had done.

- I *seriously* think this should be the model for all biology courses.

- I guess we learned about genomics through doing the tasks…. it was sort of a self-teaching class….

- Everything we gained from the class…was either found by desperately messing around on the various websites …or by talking with other students.

- I know if I could survive this class then I could survive just about anything.

Some things to look for while annotating your dot chromosome genes….

- Is there a homologous gene in *D. melanogaster*?
- Is it on the dot chromosome?
- Are all of the isoforms found in *D melanogaster* present?
- How many exons?
- Any unusual splice sites?
- Can you identify the TSS?
- What is the order and orientation of genes compared to *D. melanogaster*?
- Are there repetitious elements nearby?

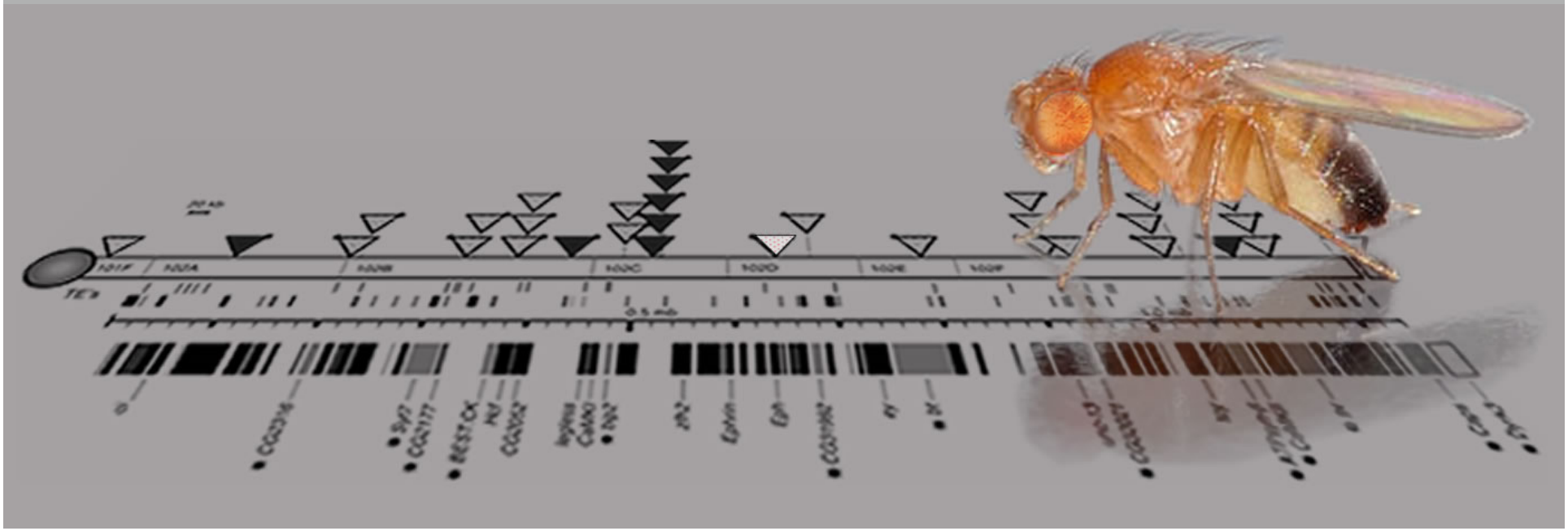
Check out your gene on FlyBase – what is the pattern of expression in *D. melanogaster*? Has a function been described?

Pooling our data, can we see any common 5’ or 3’ motifs unique to dot chromosome genes?

The Genomics Education Partnership: Investigating the and Function of the Dot Chromosome Genes in *Drosophila*

Sarah C R Elgin

January 2012



A collaborative investigation involving:

- former members of the Elgin Lab: Lee Silver, Carl Wu, TC James, Joel Eissenberg, Lori Wallrath, Fang Lin Sun, Karmella Haynes

- current members of the Elgin Lab: Nicole Riddle, Tingting Gu, Chris Shaffer, Wilson Leung

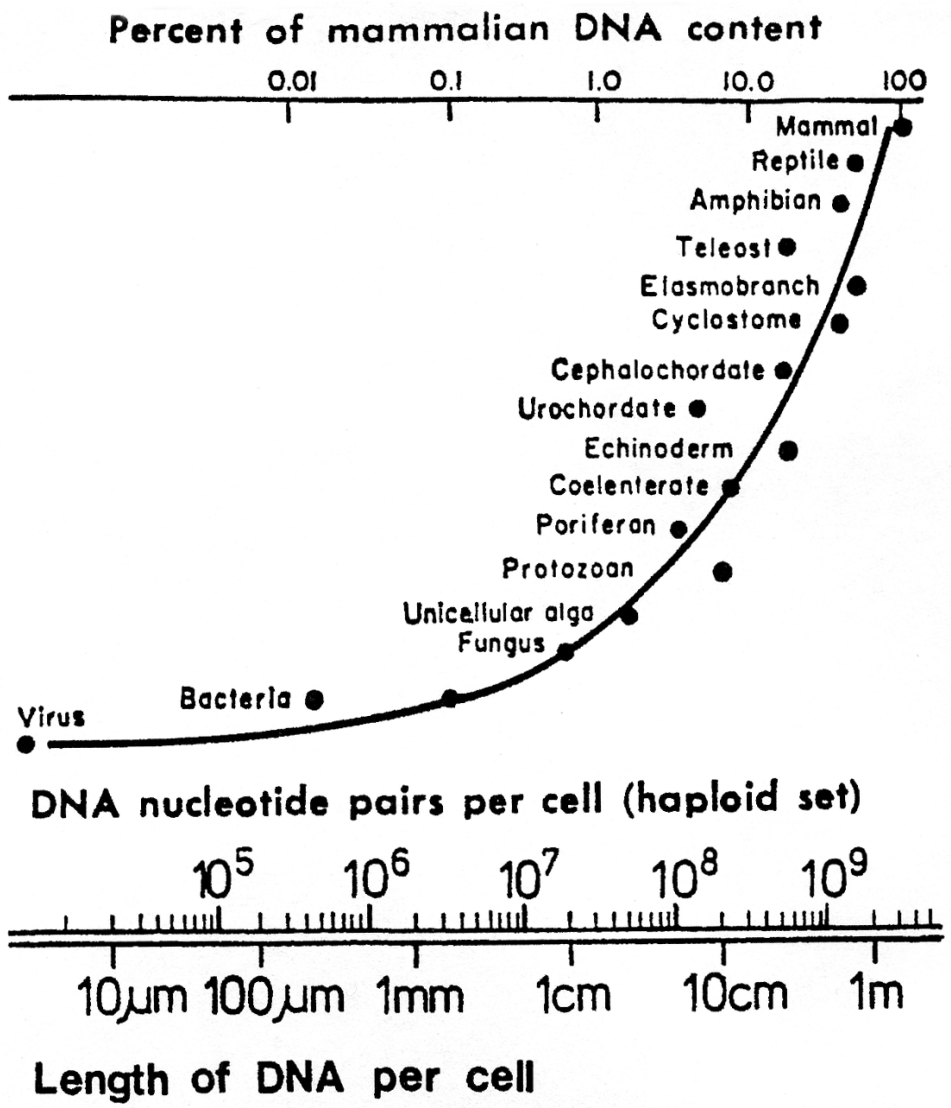
- modENCODE: Gary Karpen, Mitzi Kuroda, Vincenzo Pirrotta, Peter Park, and their colleagues

- Faculty and students of the Genomics Education Partnership

Goal: to understand the organization and functioning of the dot chromosome in Drosophila, an unusual heterochromatic domain.

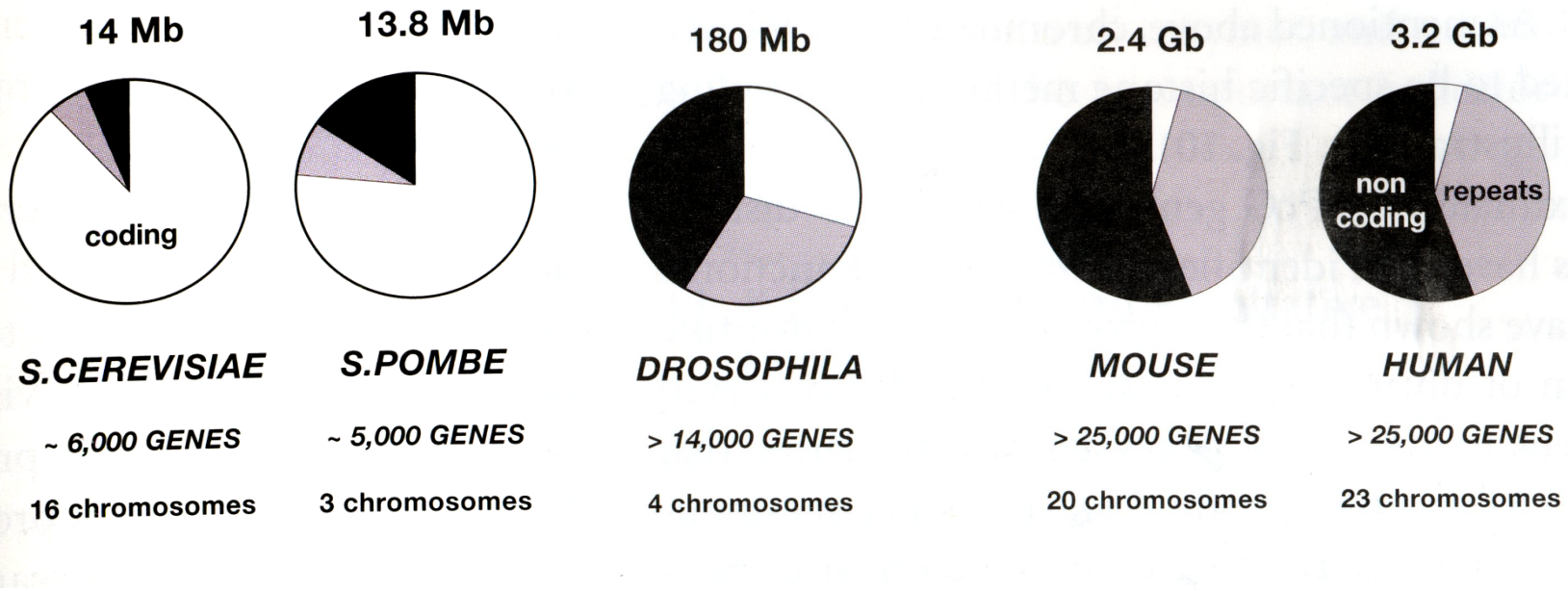
Funding: HHMI Professors Program
NIH General Medical Sciences, National Human Genome Research Institute

Minimum haploid DNA content - the C Value Paradox



Britten and Davidson, 1969 Science 165:349

Larger genomes reflect high levels of repeats - retroviral and DNA transposon remnants (TEs)

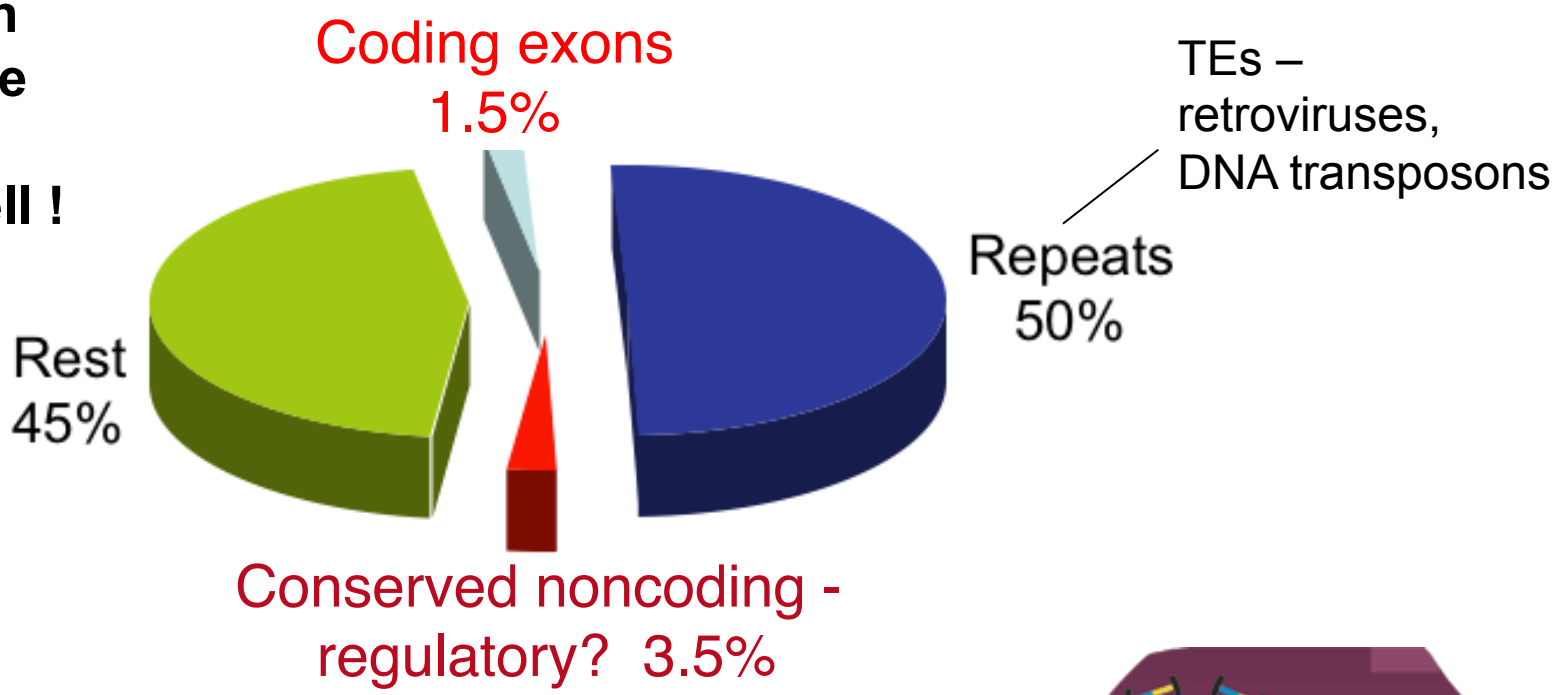


This introduces many complications when assembling a genome sequence!

Allis et al: Epigenetics 2007

Eukaryotic genomes are very large – and most of that DNA is non-coding!

Human Genome 3 Gb ~2 m/cell !



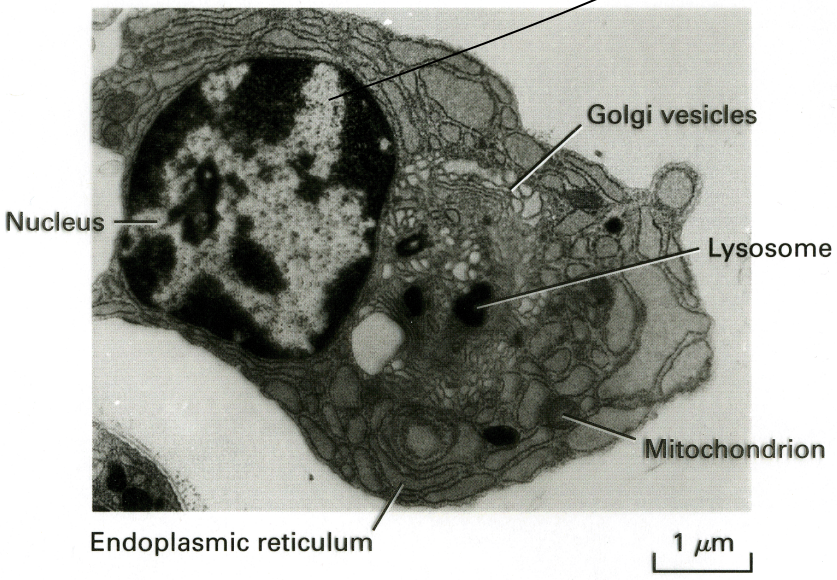
Key Questions:

- Is it junk or garbage?
- How is DNA packaged into a nucleus?
- How is silencing maintained – while allowing appropriate transcription ?

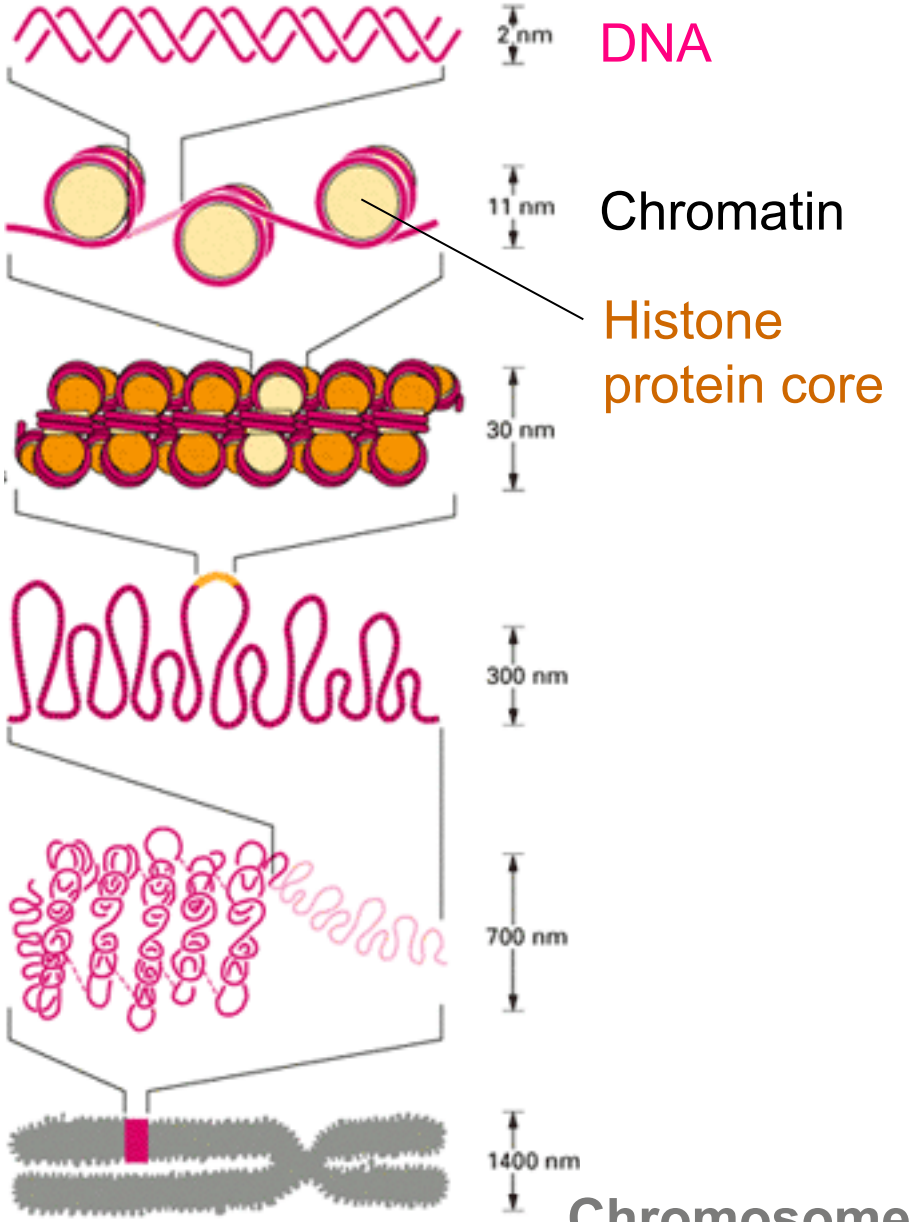
Chromatin formation:

First step - packaging in a nucleosome array

Second - differential packaging into heterochromatin & euchromatin

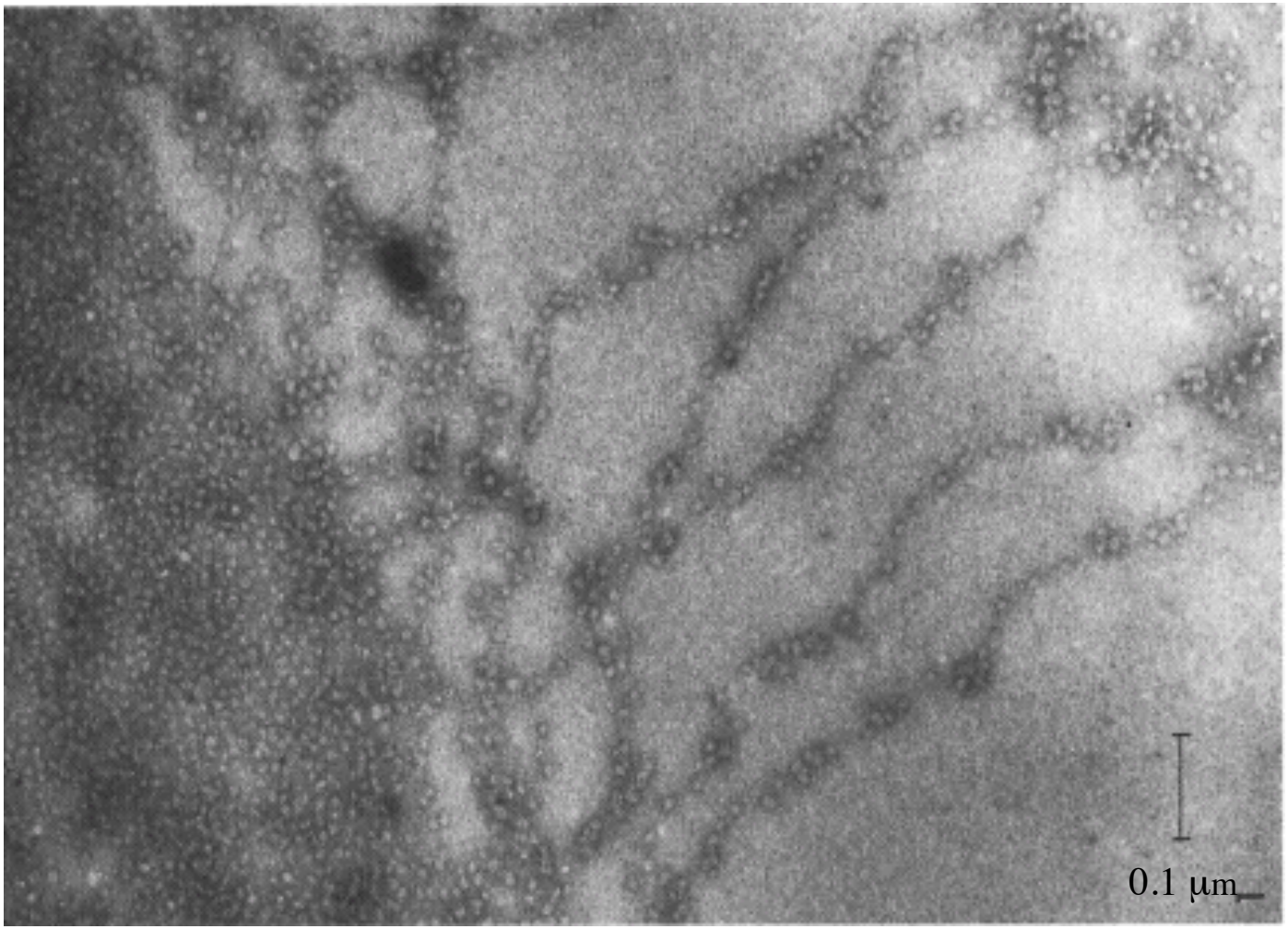


Lodish et.al., Molecular Cell Biology, 4th Edition



Felsenfeld et al. Nature 2003, 421: 448

Electron micrograph of chromatin fibers (rat thymus nucleus)



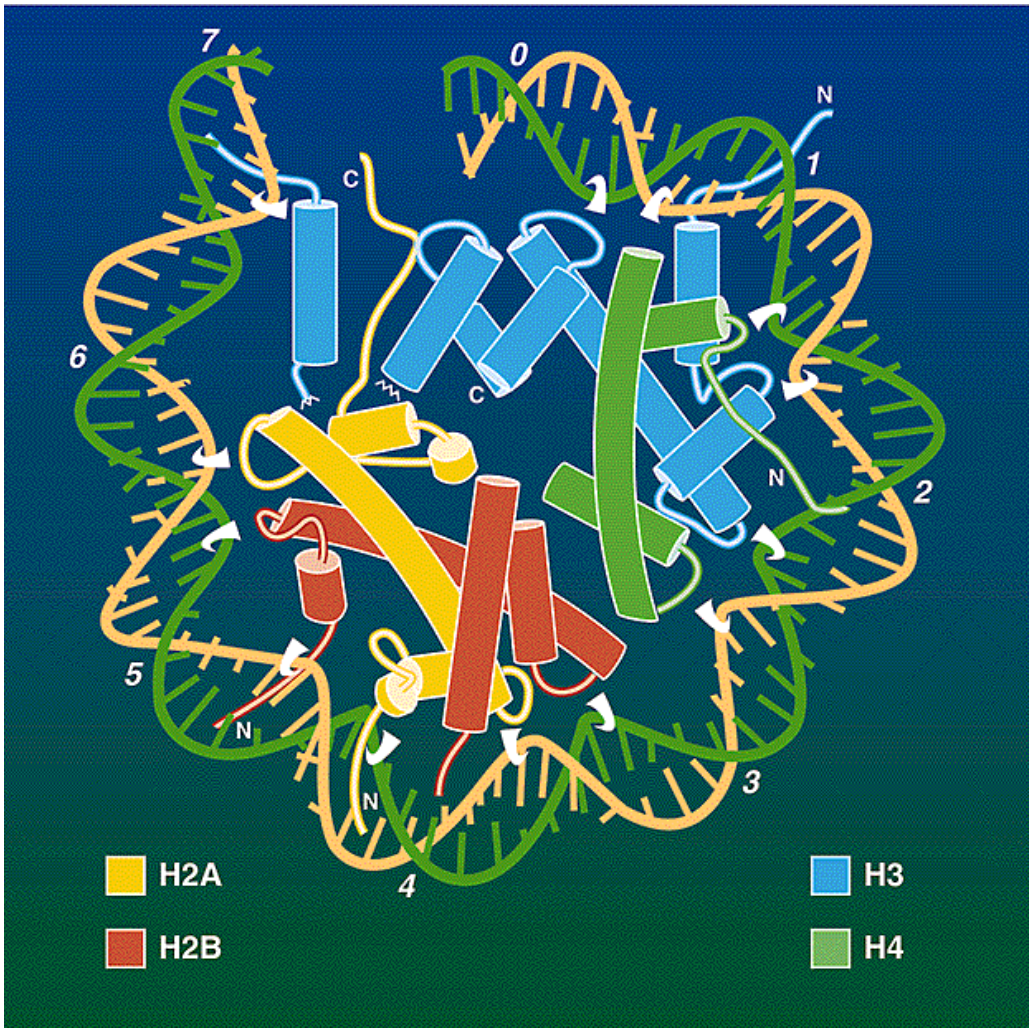
Olins et. al., 1975 J. Cell Biol, 64:528

“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

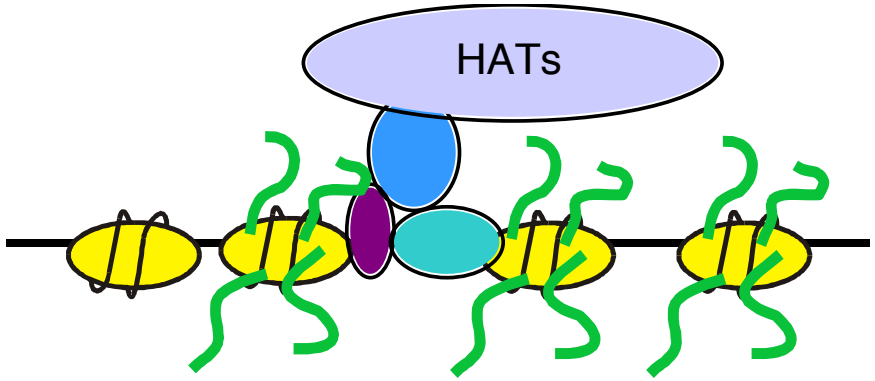
Half of the nucleosome structure is shown

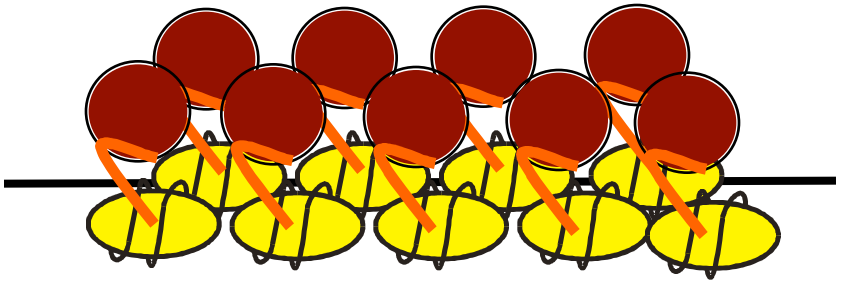
One turn of the DNA helix is visible (73 bp)

View is down the superhelix axis

Protein - DNA contact: white hooks

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

Chromatin structure = epigenetics !

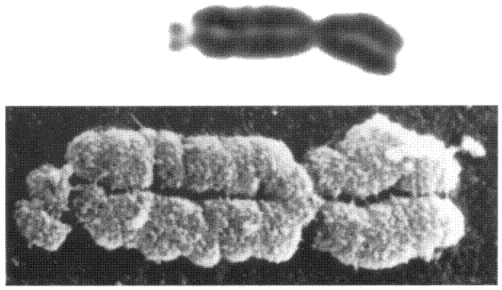
What sets and maintains tissue-specific gene expression patterns? Differences are heritable through mitosis, but independent of DNA sequence.

- DNA modification (mC)
- Chromatin structure
- Nuclear localization

Zoghbi and Beaudet 2007

It's all about silencing!
How is chromatin assembled?
When, where and how does gene silencing occur?

Incorrect silencing can lead to genetic disability, as seen in Fragile X syndrome



Fragile X Foundation

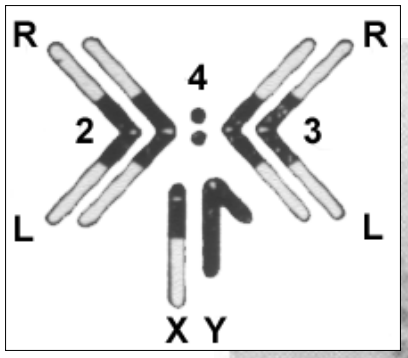
Heterochromatin formation – silencing counts!

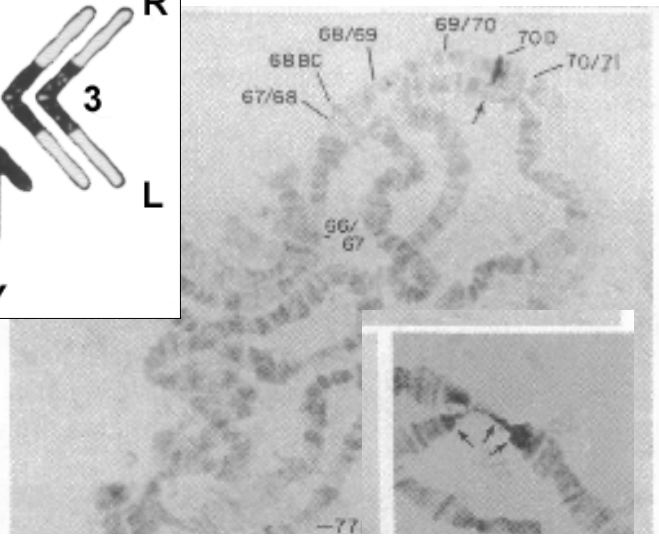
1

How is heterochromatin organized and packaged to promote silencing?

2

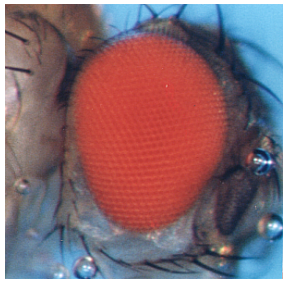
The fourth chromosome appears heterochromatic but has ~80 genes:
-- do these genes have unusual characteristics?
-- how has the chromosome evolved?
-- how do these genes function?





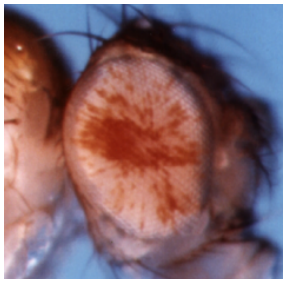
Mary Lou Pardue, MIT

euchromatin



expressed

heterochromatin



silenced

Fruit Flies!

Short life cycle, easily maintained: good genetic tools

Polytene chromosomes: excellent cytology

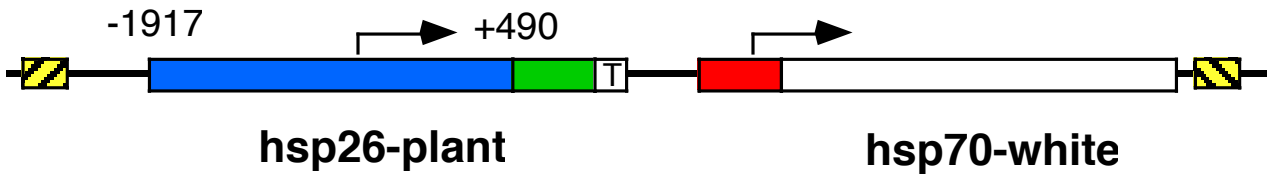
Biochemical approaches

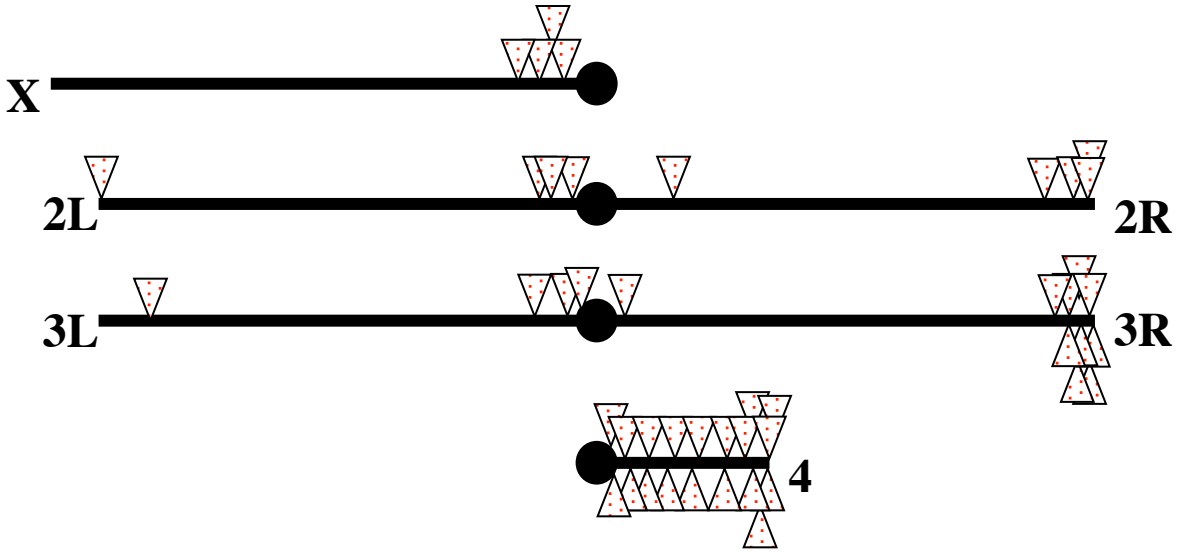
Simple genome, good reference sequence

PEV – reporter for gene silencing, heterochromatin formation

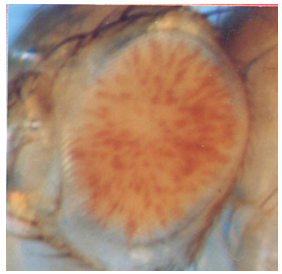
Metazoan useful for behavioral, developmental and human disease research

Transposition of a P element reporter allows sampling of euchromatic and heterochromatin domains

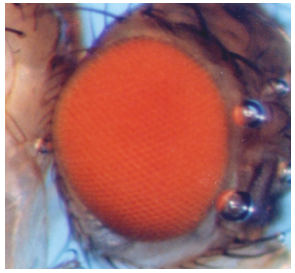




And the Y chromosome



Silenced
1%



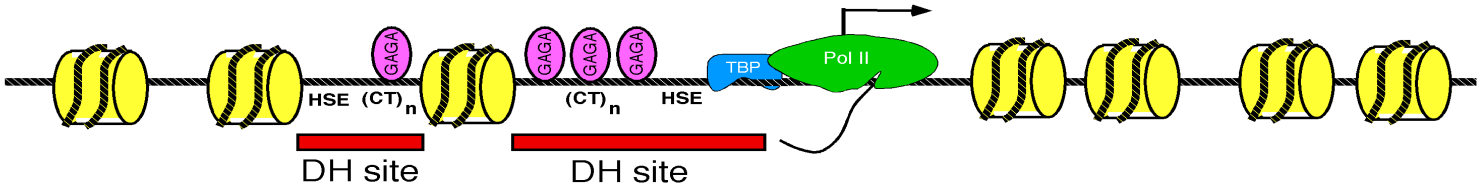
Active
99%

Wallrath and Elgin, 1995

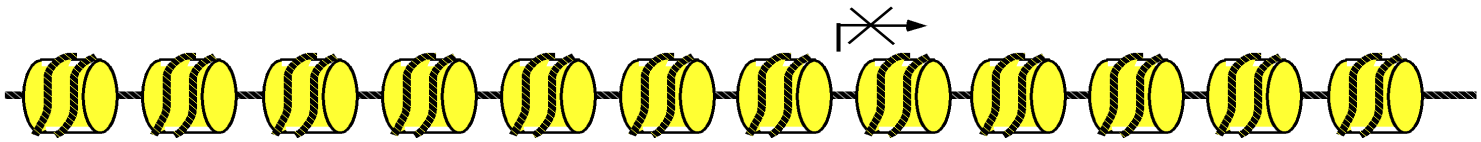
Assessing chromatin structure-
same gene, different environments

Analysis based on nuclease digestion of chromatin

Euchromatin



Heterochromatin



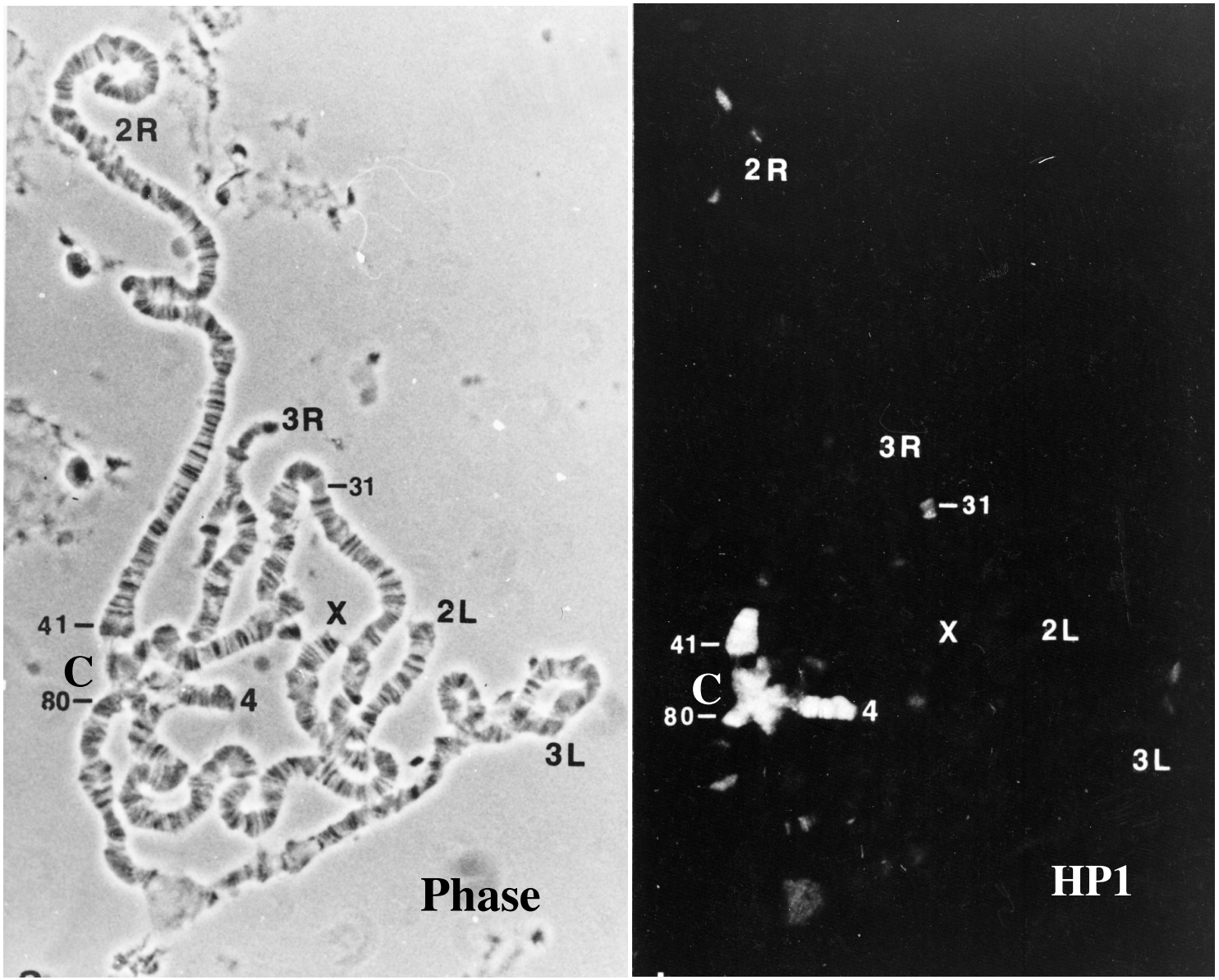
The euchromatic *hsp26* transgene:

- DH sites: accessibility at the TSS, upstream regulatory region
- irregular nucleosome array

The heterochromatic *hsp26* transgene:

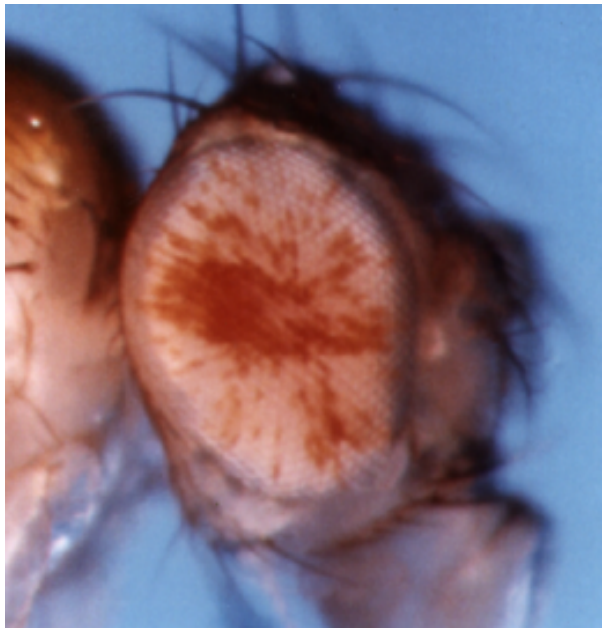
- loss of DH sites
- regular nucleosome array

Looking for heterochromatic proteins by immunofluorescent staining of the polytene chromosomes: discovery of HP1a



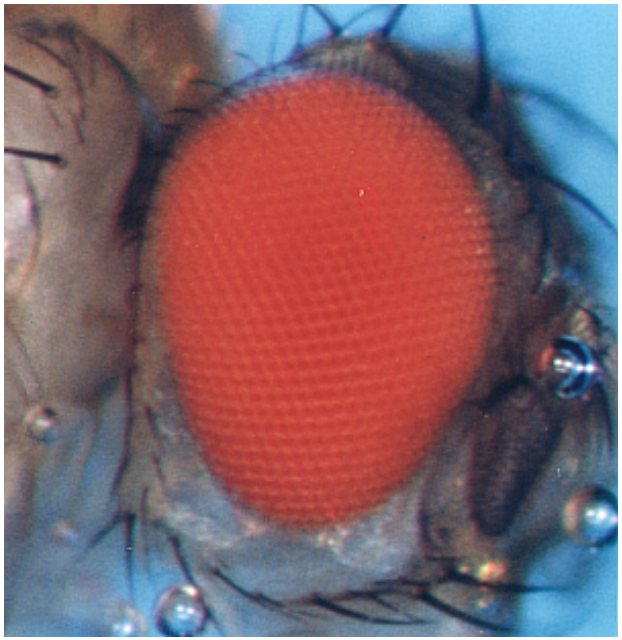
James & Elgin, 1986; James et al 1989

Heterochromatin-associated gene silencing is dependent on HP1



Mutations in

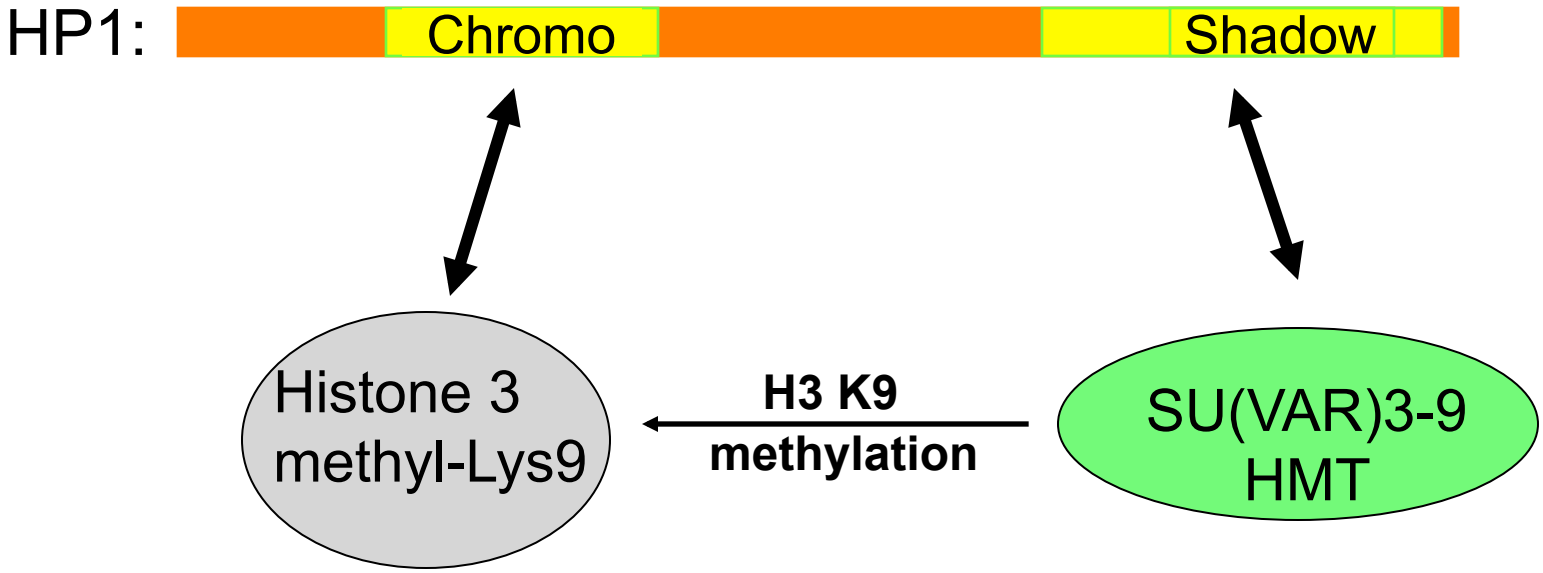
gene for HP1a



Mutations recovered by T Grigliatti as suppressors of PEV.
Dosage dependent response.

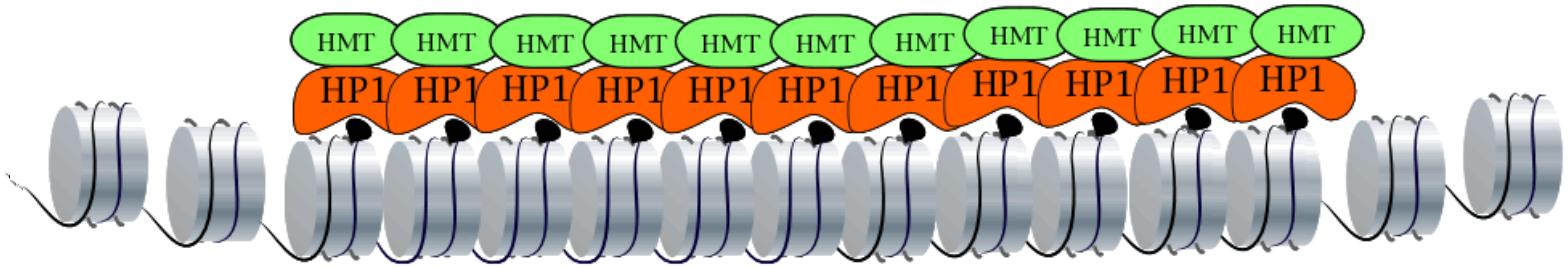
Eissenberg et al, 1990, PNAS 87: 9923

HP1 interacts with both the modified histone H3K9me2/3 *and* the modifying enzyme



[(SU(VAR)3-9 identified in screen by Reuter;
H3 interaction first shown from work in mammals – Jenuwein, Kouzarides;
demonstrated in flies by Imhof.]

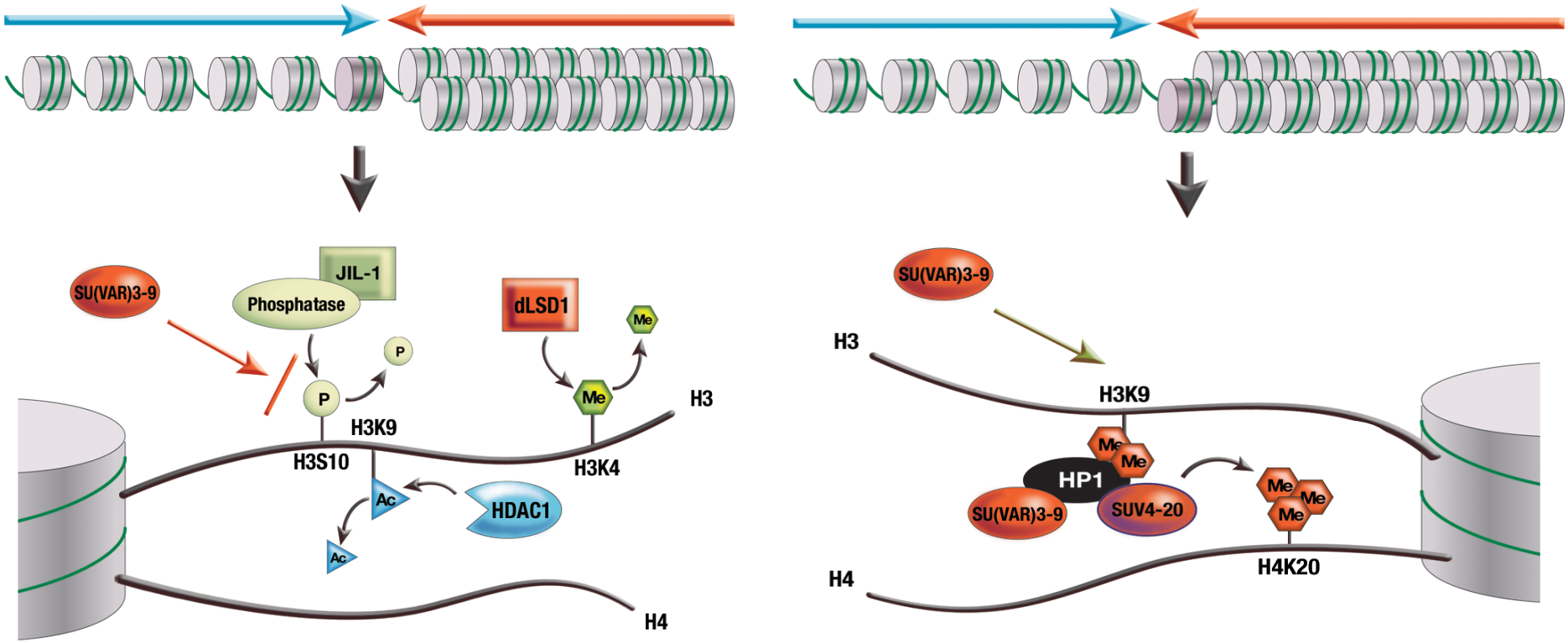
Model for spreading of heterochromatin



Establishing silencing: a multi-step process

Loss of euchromatin marks

Gain of heterochromatin marks



w^{m4} reporter (screens by Reuter, Grigliatti, others)

Heterochromatin formation on the dot chromosome…

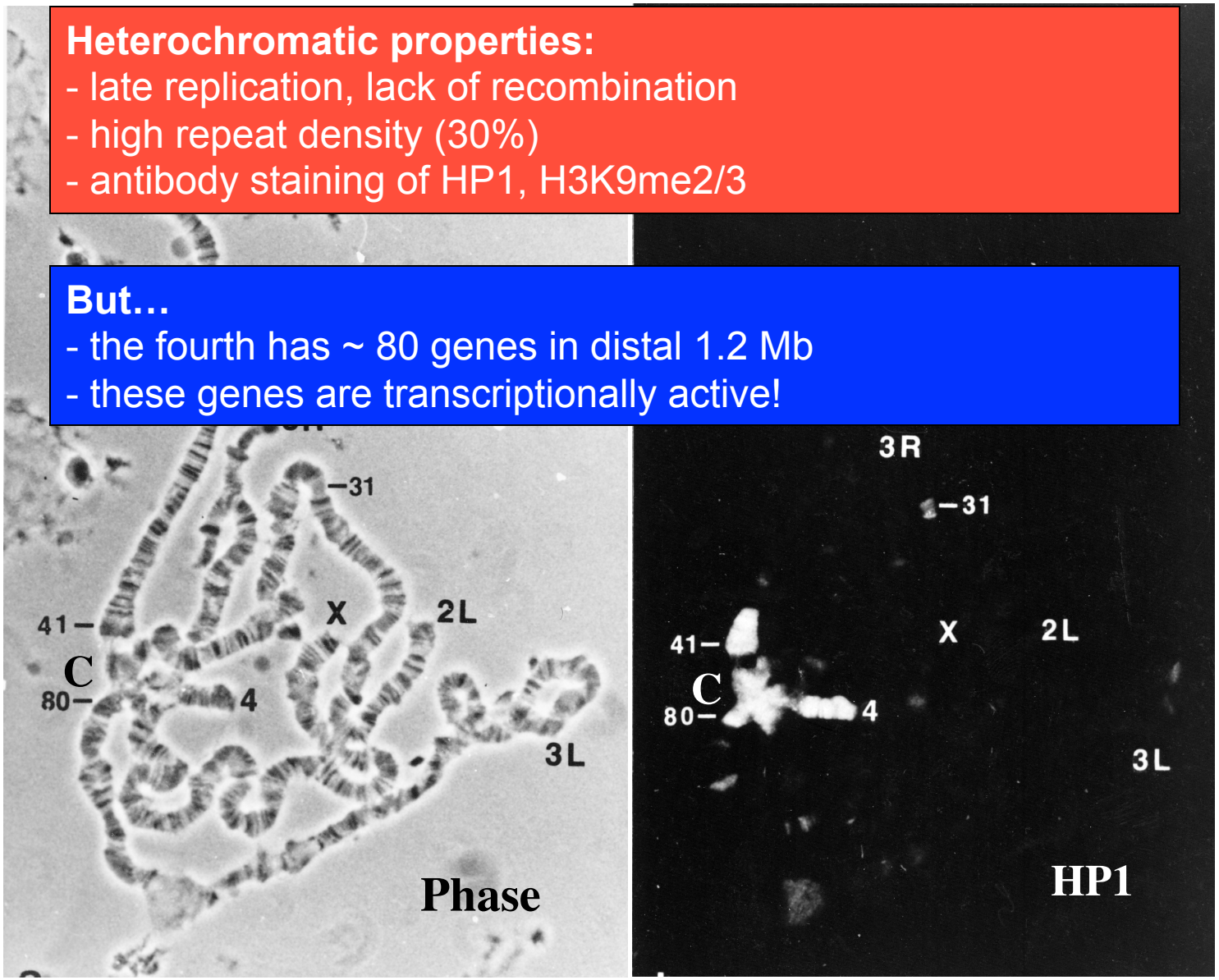
2

The fourth chromosome appears heterochromatic
but has ~80 genes:
-- do these genes have unusual characteristics?
-- how has the chromosome evolved?
- how do these genes function?

The *Drosophila melanogaster* fourth chromosome exhibits an amalgam of euchromatic and heterochromatic properties (HP1a association)

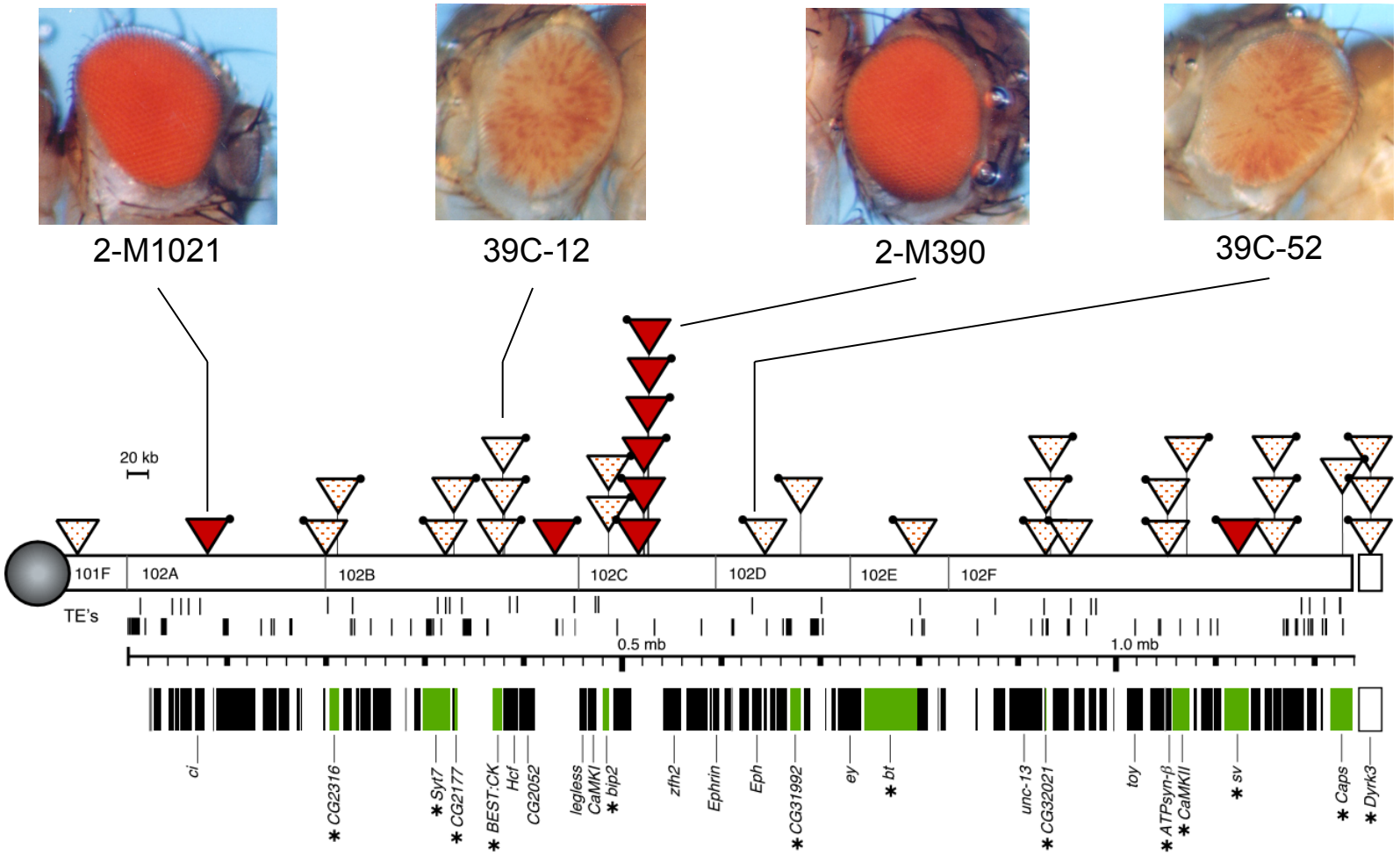
Heterochromatic properties:
- late replication, lack of recombination
- high repeat density (30%)
- antibody staining of HP1, H3K9me2/3

But…
- the fourth has ~ 80 genes in distal 1.2 Mb
- these genes are transcriptionally active!



James & Elgin, 1986; James et al 1989

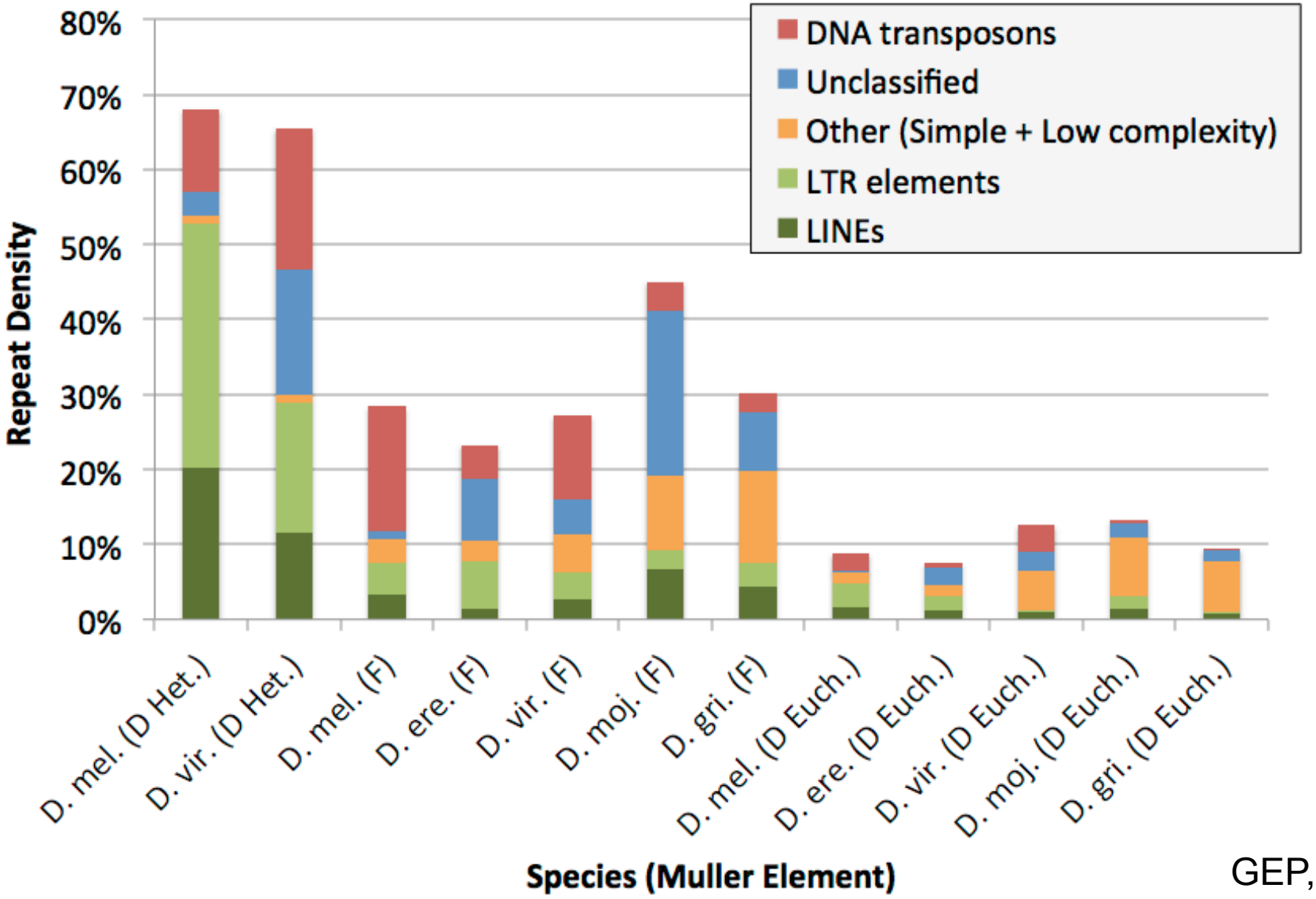
Most *hsp70-white* reporters exhibit variegation on insertion into the fourth chromosome



Sun et al 2004; Riddle et al 2007

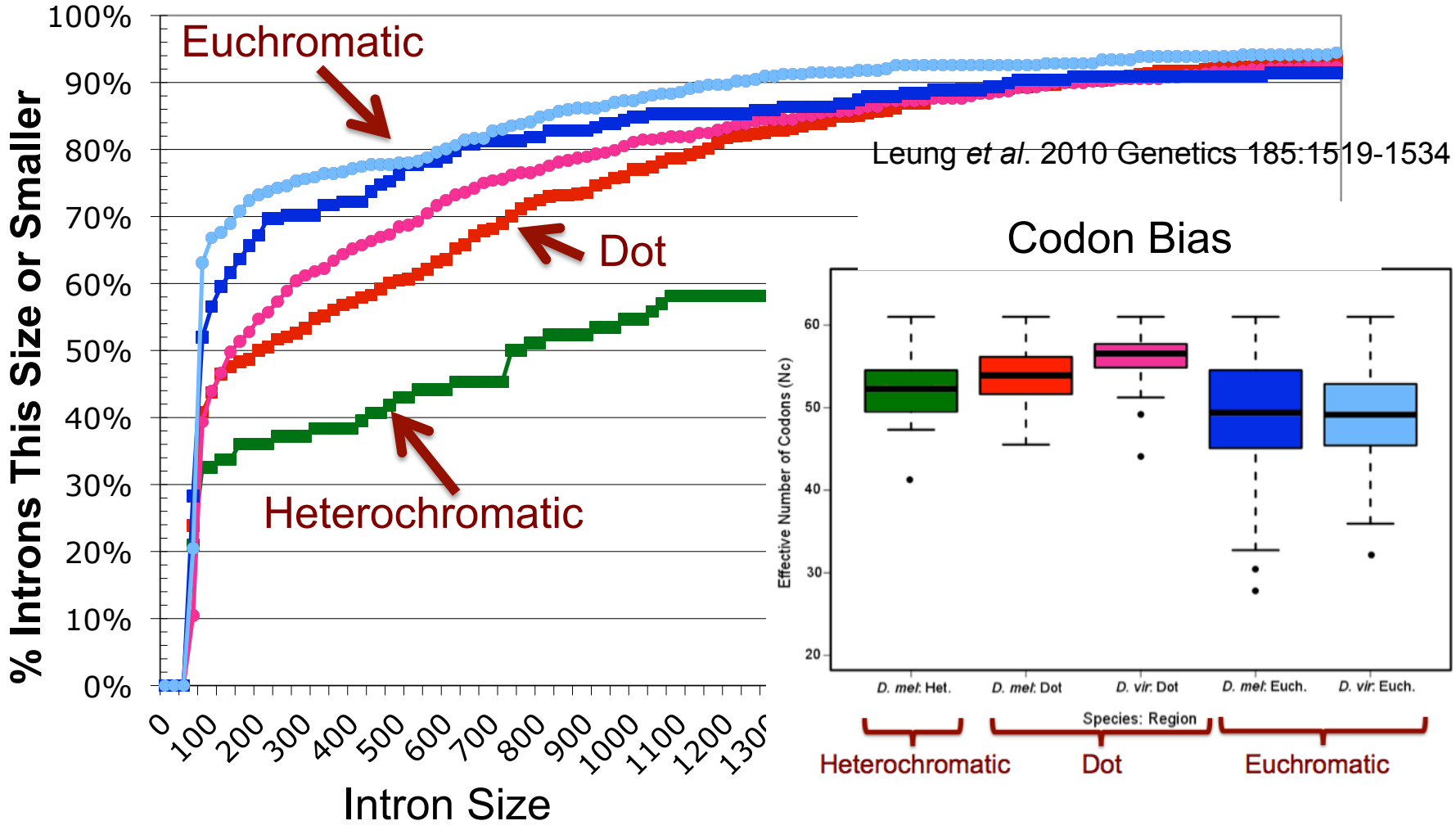
The *Drosophila* dot chromosomes are typically 25% - 30% repetitious DNA

(– but up to 80% in *D. ananassae*!)



GEP, 2011

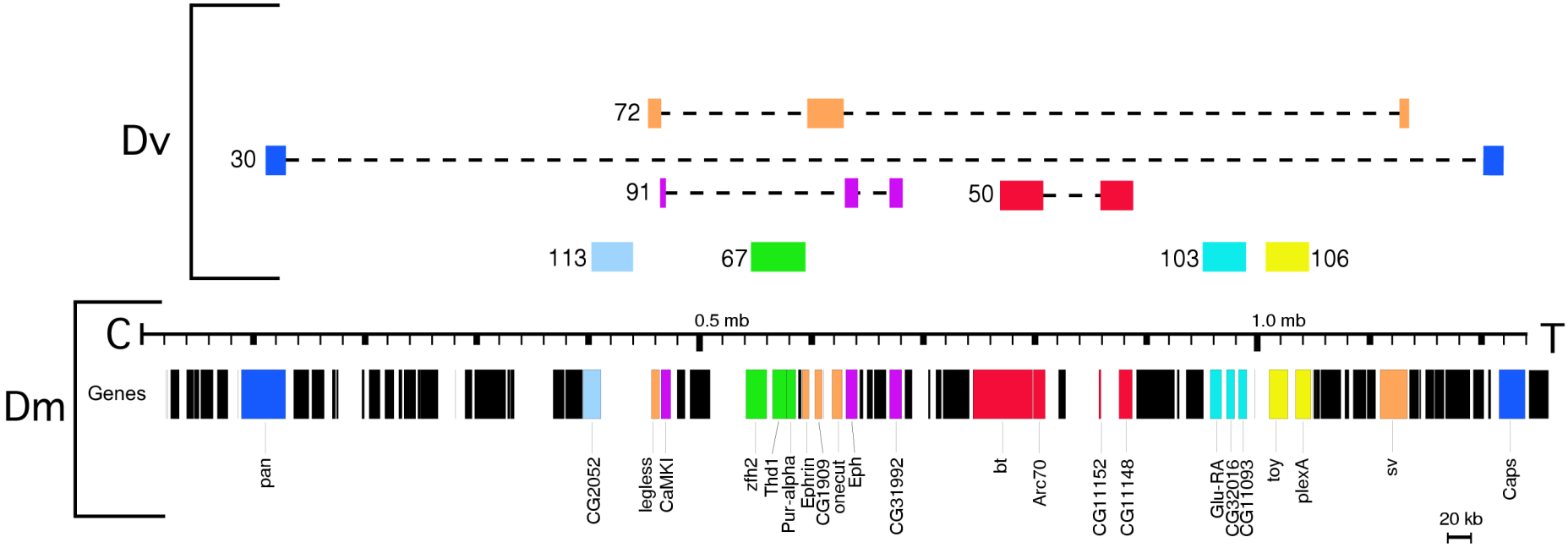
Dot chromosome genes: introns are larger, exons show less codon bias



Leung *et al*. 2010 Genetics 185:1519-1534

D. melanogaster Dot
D. virilis Dot
D. melanogaster Euch.
D. virilis Euch.
D. melanogaster Het.

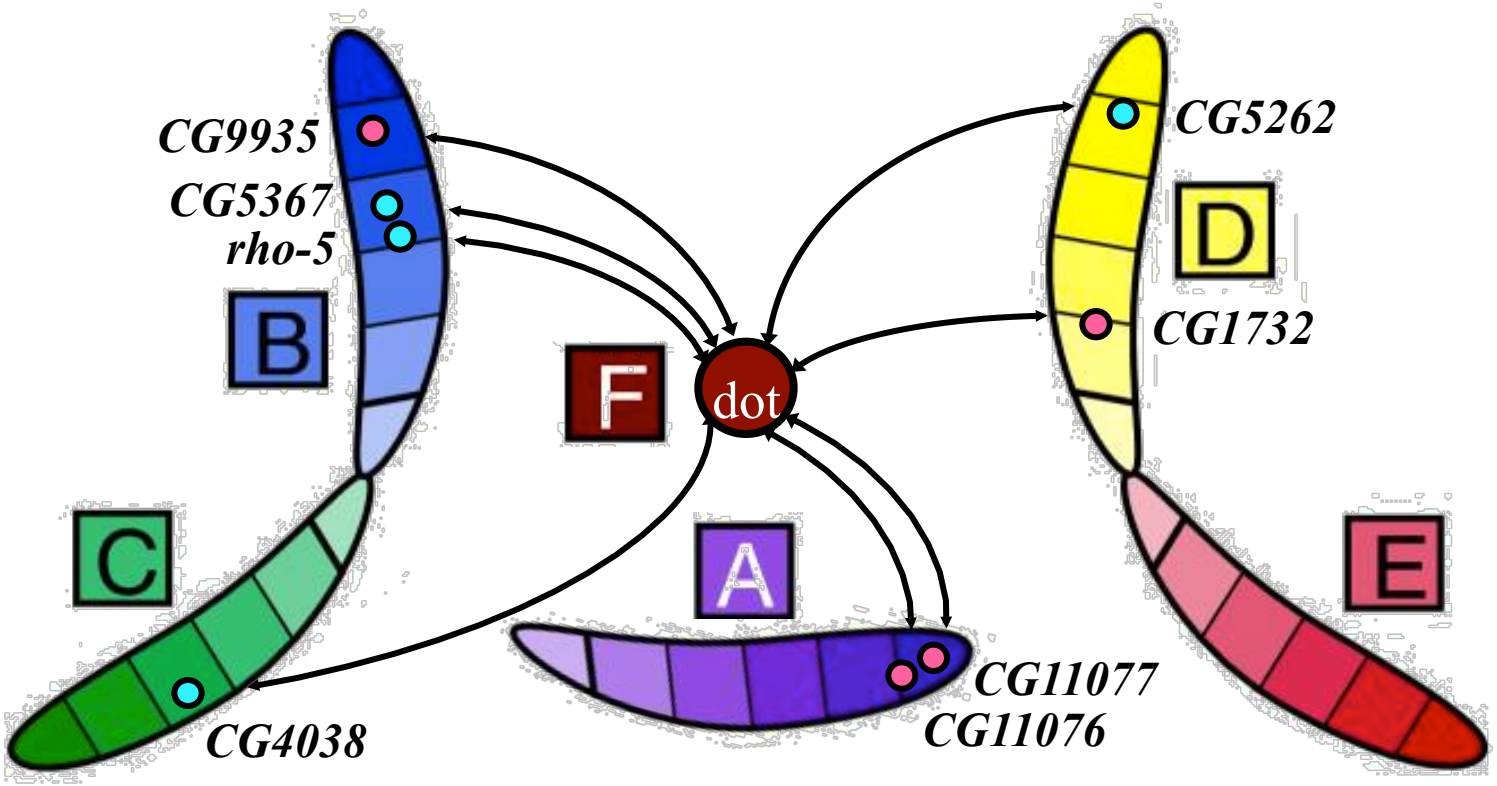
Initial analysis of *Drosophila virilis* dot chromosome fosmids



Almost all of the same genes are present (~90%), but rearrangements within the chromosome are common – a minimum of 33 inversions are needed to convert the order and orientation from *D virilis* to *D melanogaster*!

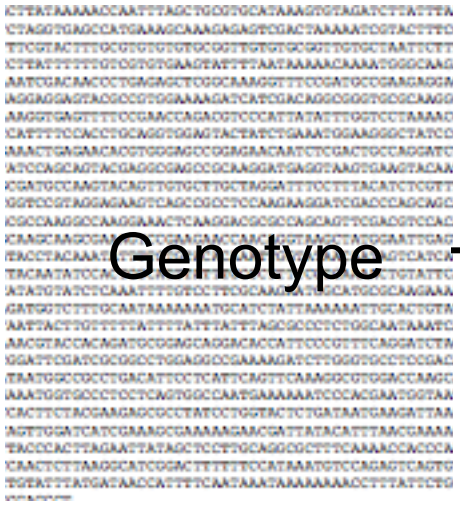
Slawson et. al., 2006 Genome Biology, 7(2):R15.

"Wanderer" genes move between the dot chromosome and a euchromatic site in the long arms; they adopt the properties (gene size, codon bias) of their local environment



Leung et al 2010

But remember - what determines phenotypes? It's not just your DNA….it's how it is packaged!

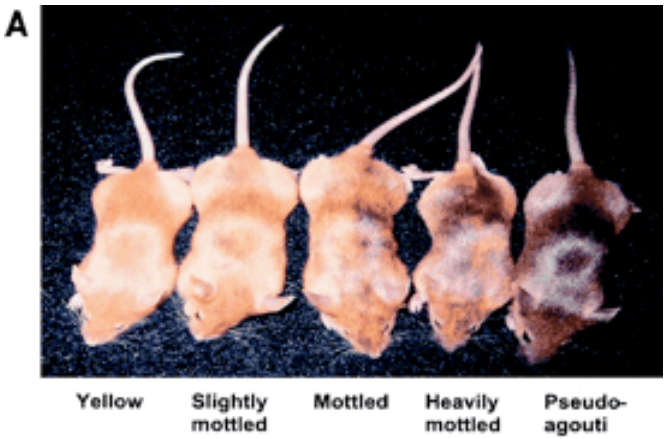


Genotype

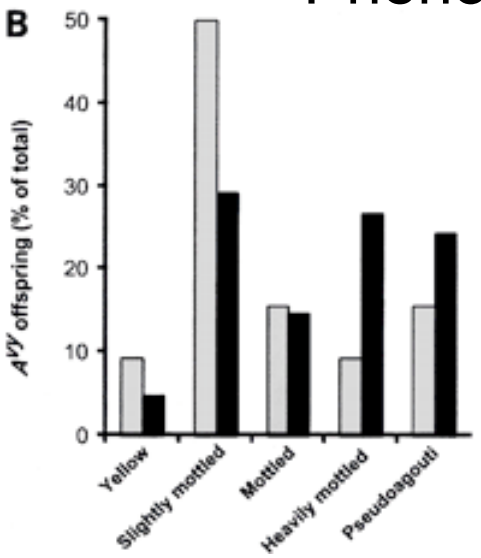
Environment (diet)
(grey bars = folate)

Epigenetics
?

Development



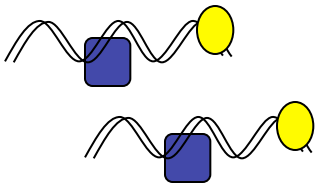
Phenotype



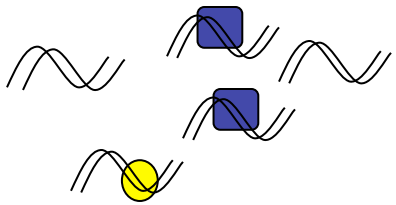
(Waterland and Jirtle 2003)

Higher resolution mapping: Chromatin ImmunoPrecipitation – ChIP (cells or nuclei)

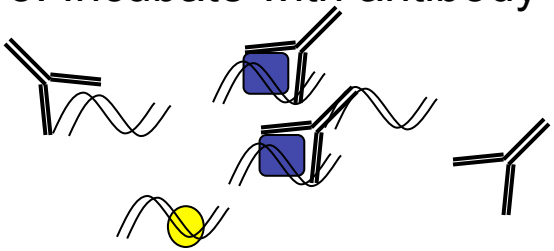
1. Crosslink proteins to DNA



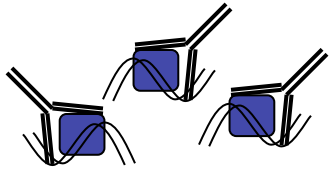
2. Isolate chromatin and sonicate



3. Incubate with antibody

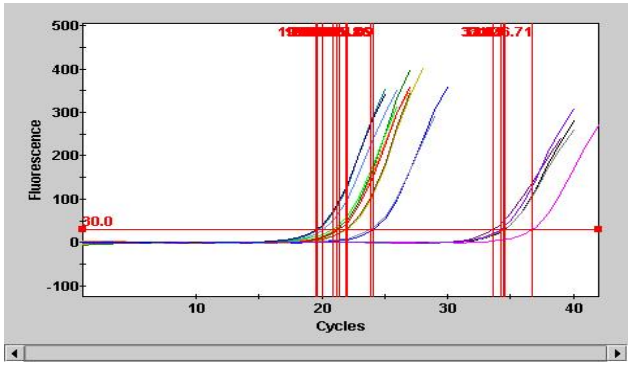


4. Isolate AB/chromatin complexes

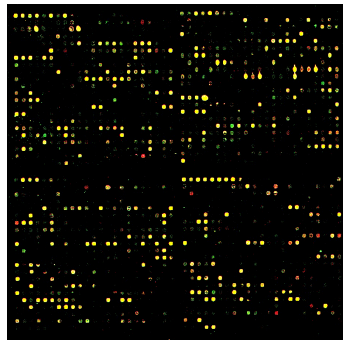


5. Isolate DNA from complexes

qPCR

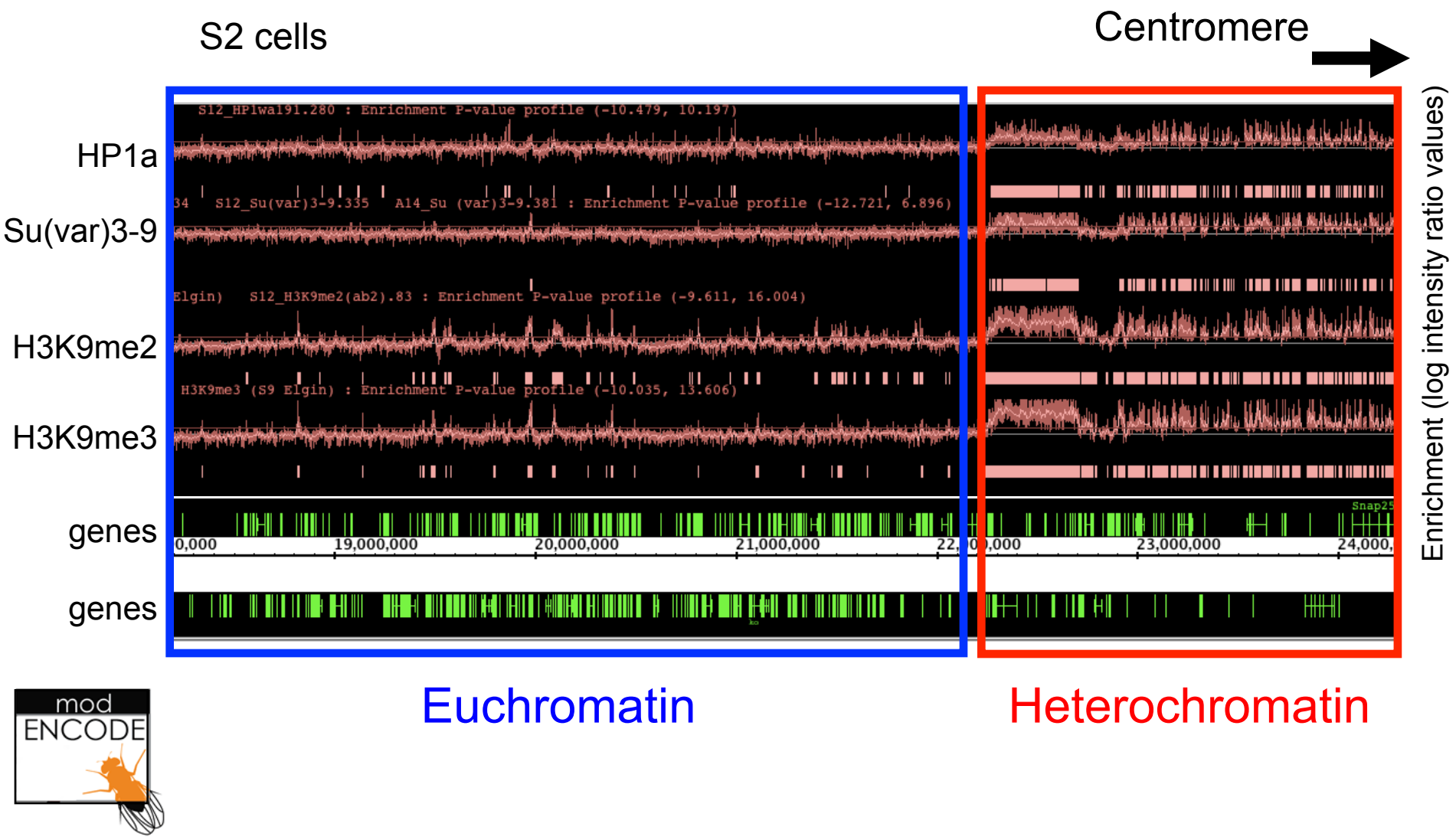


ChIP-chip*

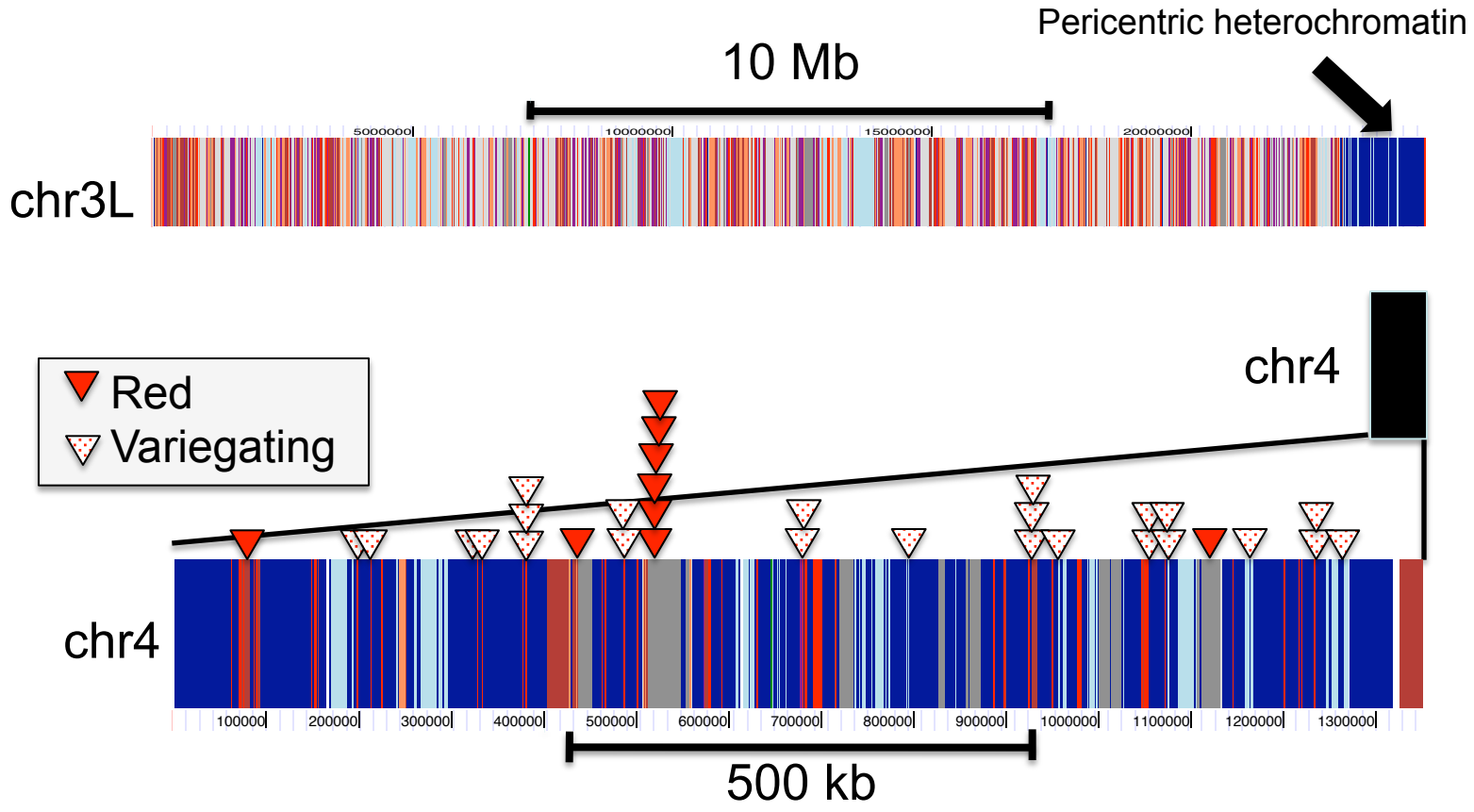


ChIP-seq

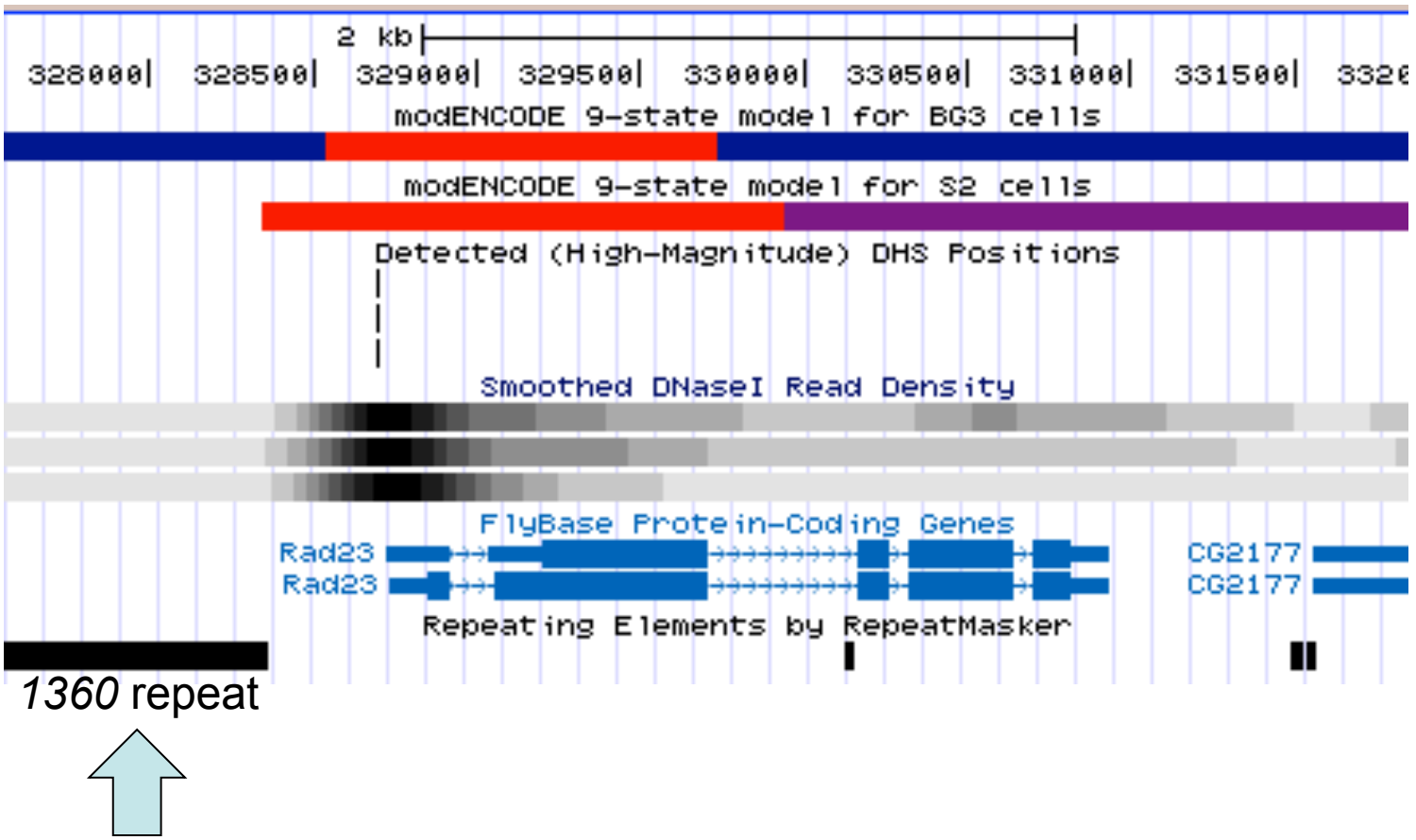
Mapping chromosomal proteins & histone modifications by ChIP-chip: chromosome arm 3L shows a distinct shift between heterochromatin and euchromatin.



An expanded view of the fourth chromosome reveals TSS (state 1, red) and Pc (state 6, grey) domains interspersed within heterochromatin (states 7 & 8, blue).



Most 4th chromosome genes lie in heterochromatic space (blue), but active genes achieve state 1 (red) at the TSS



Future: try to determine what feature drives 4th chromosome gene expression that is absent from euchromatic genes (hsp70).

Heterochromatin formation on the dot chromosome…

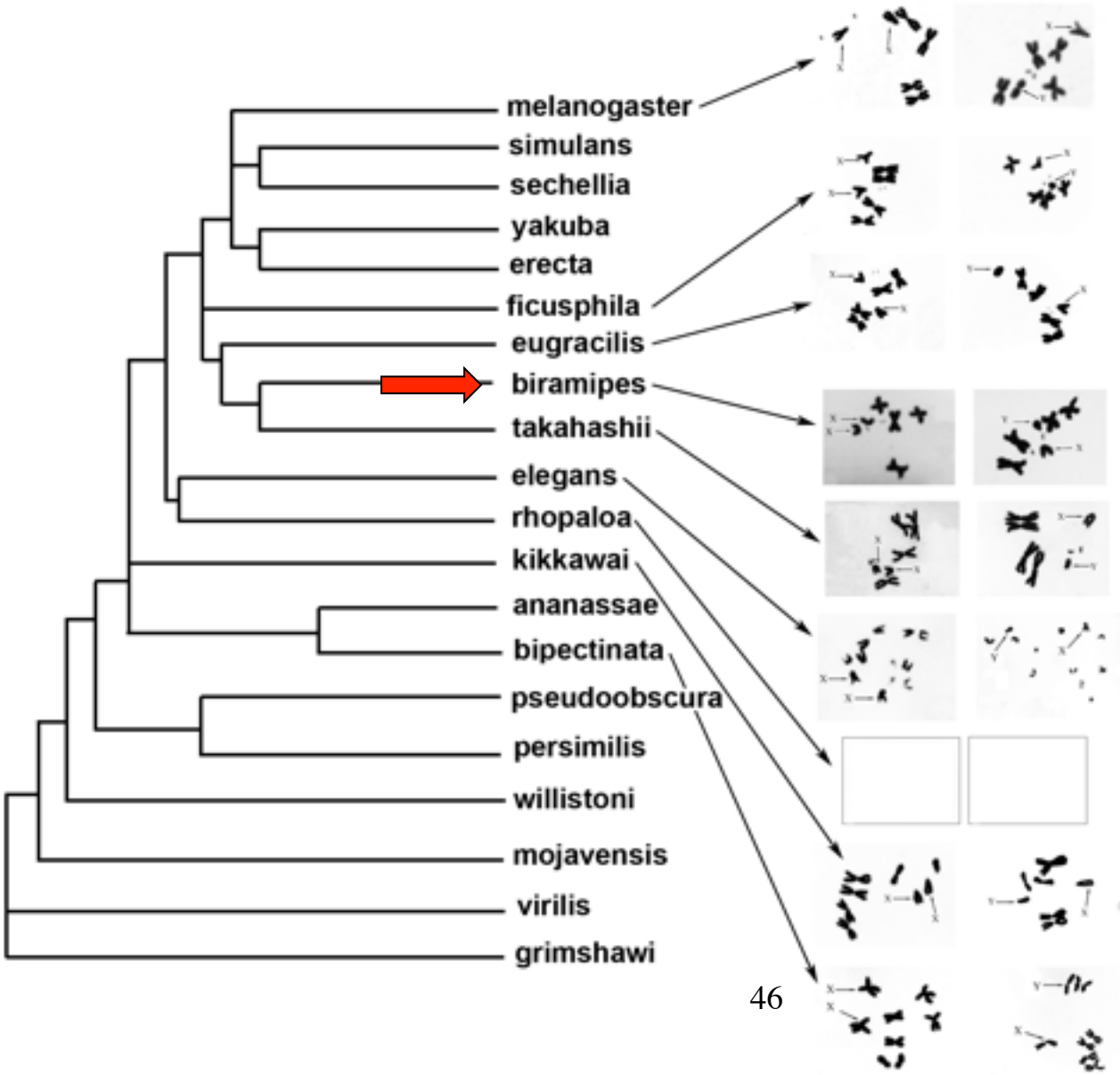
1. Heterochromatin formation changes chromatin at the nucleosome level, eliminating HS sites at the TSS of euchromatic genes; silencing is dependent on HP1a

2. Fourth chromosome genes are larger, have more introns, and less codon bias than euchromatic genes

   Fourth chromosome genes show high levels of HP1a and H3K9 methylation over the body of the gene, but maintain access at the TSS.

   Next steps: what makes fourth chromosome genes robust? Lets look for fourth chromosome motifs!

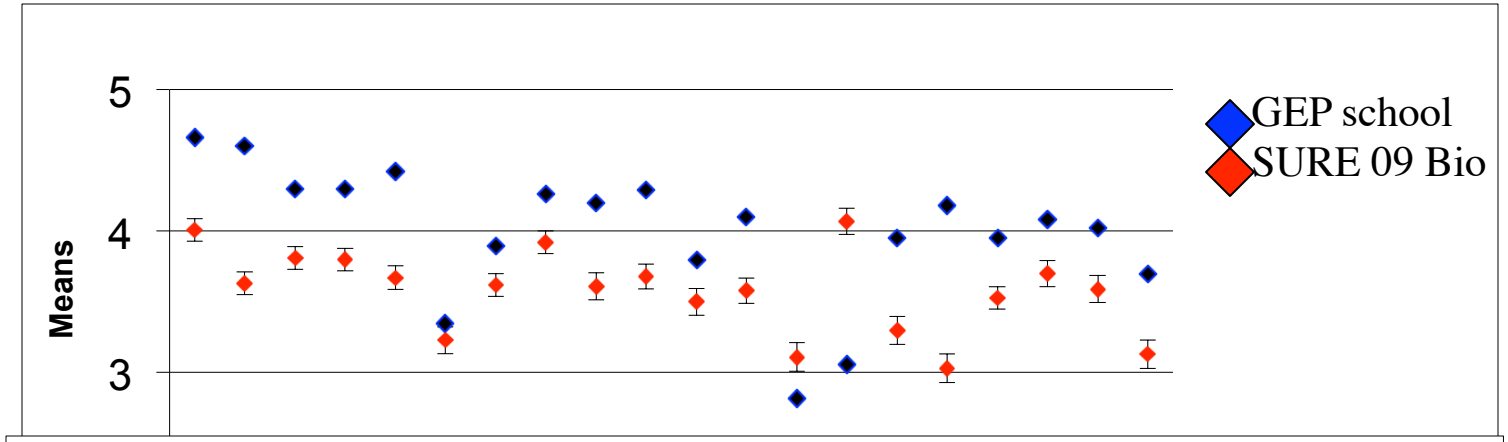
Eight new Drosophila genomes chosen at an evolutionary distance to facilitate motif finding

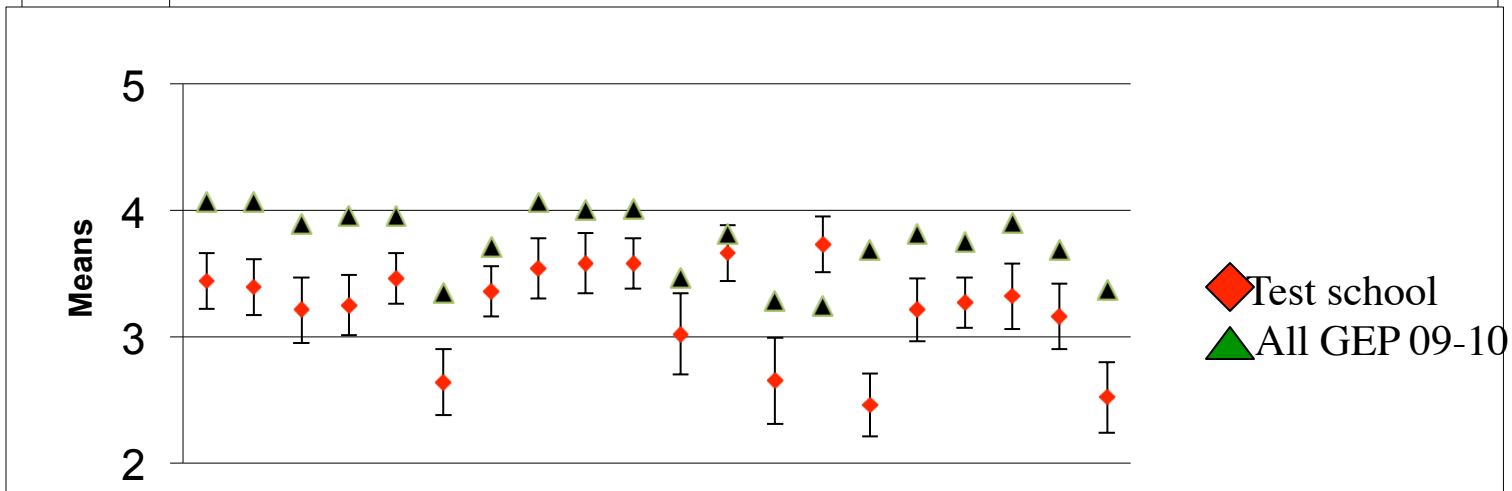


Expanded dot chromosomes?

Questions?

Results using finishing/annotation in semester lab course -112 hr
Quiz scores: precourse 12.8, postcourse 15.6.





Results using annotation in genetics lab – 10 hr
Quiz scores: precourse 4.0, postcourse 9.7

Bioinformatics is a cost-effective research area for undergraduates

- GCAT - Chip, SEEK, SynBio: Malcolm Campbell, Davidson College
  http://www.bio.davidson.edu/projects/gcat/gcat.html

- Teaching Genomics Consortium: Susan Singer, Carleton, Lois Banta, Williams
  http://serc.carleton.edu/genomics/ (on-line curriculum)
  Aiptasia & Chamaecrista Genomics Explorers: Susan Singer, Carleton and Jodi Schwarz, Vassar
  http://serc.carleton.edu/exploring _genomics/aiptasia

- Dynamic Gene, CSHL Dolan Center: Dave Miklos
  http://dynamicgene.dnalc.org/ (annotating rice)
  iPLANT: DNA Subway, with CSHL Dolan Center
  http:// (annotating Arabidopsis)

- Human Microbiome Project: Anne Rosenwald, Georgetown, Jena Canfield, Simmons, JC Ventor Institute

Bioinformatics is a cost-effective research area for undergraduates

JGI programs led by Cheryl Kerfeld, joint with ASM and Brad Goodner, Hiram College

- Undergraduate Research in Microbial Genome Analysis
  http://www.jgi.doe.gov/education/genomeannotation.html
- Tools for undergraduates to annotate microbial genomes as part of the GEBA
  http://www.jgi.doe.gov/programs/GEBA/
- Undergraduate Research in Microbial Functional Genomics
  http://www.jgi.doe.gov/education/functional.html

Microbial work lends itself to vertical integration

Other projects: bar coding for environmental studies
regulatory and metabolic pathways
protein folding (Foldit game)