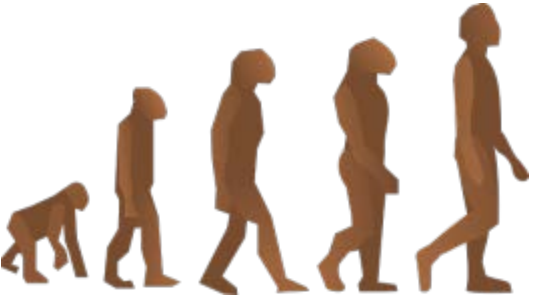
Signatures of past natural selection in the genome

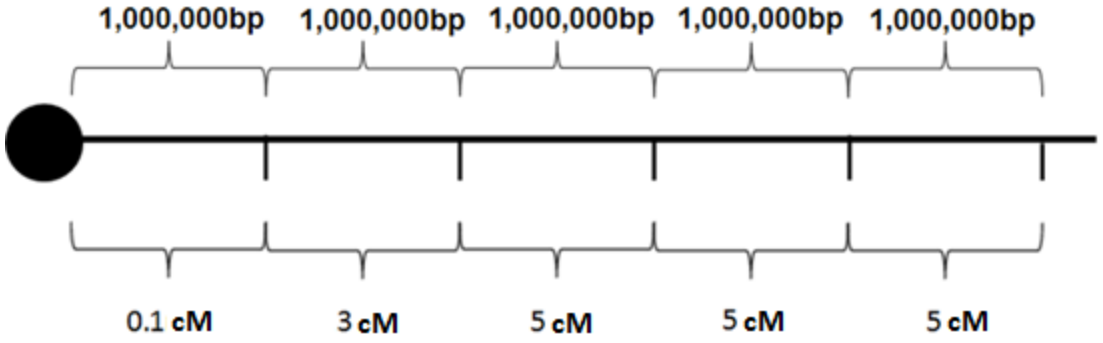


Recombination can also help us detect natural selection

- Recombination rates not constant across genome
  - We already discussed “crossover hotspots”
  - Some regions have more hotspots than others
- Often low near centromeres

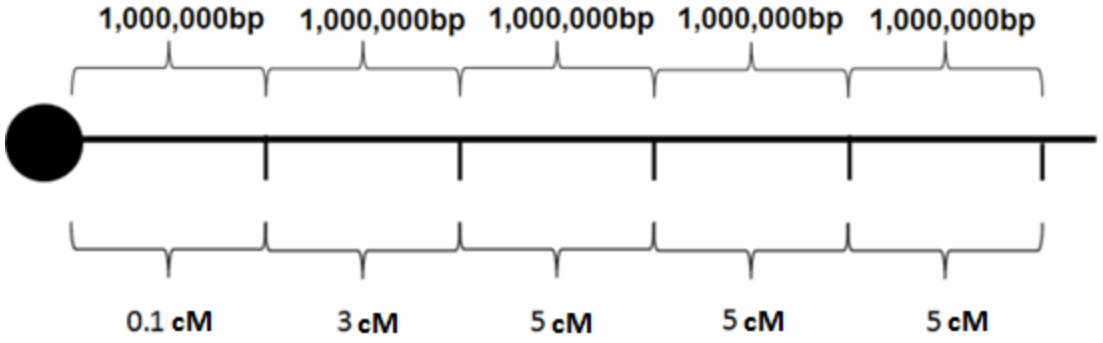
Recombination can also help us detect natural selection

- Recombination rates not constant across genome
  - We already discussed “crossover hotspots”
  - Some regions have more hotspots than others
- Often low near centromeres



Recombination can also help us detect natural selection

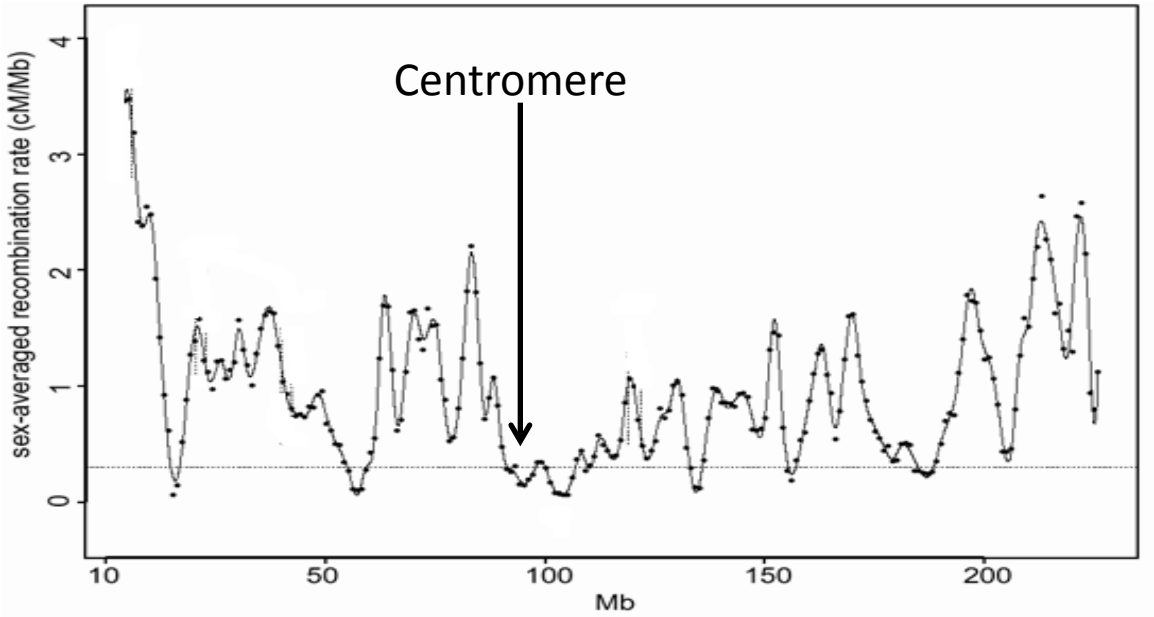
- Recombination rates not constant across genome
  - We already discussed "crossover hotspots"
  - Some regions have more hotspots than others
- Often low near centromeres



Recombination *rate* (cM/Mb) lower near centromere

Recombination rates also vary more generally

- Figure of recombination rates along human chromosome 3



New adaptive alleles arising by mutation periodically

- Sometimes in regions where recombination is rare or absent (very low cM/Mb; no hotspots)

```
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
      *               *                 *   *             *
```

New adaptive alleles arising by mutation periodically

- Sometimes in regions where recombination is rare or absent (very low cM/Mb; no hotspots)

```
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
*       *               *                   *   *               *
```

New adaptive alleles arising by mutation periodically

- Sometimes in regions where recombination is rare or absent (very low cM/Mb; no hotspots)

```
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
ACATGCTAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
*       *               *                   *   *               *
```

New adaptive alleles arising by mutation periodically

- Sometimes in regions where recombination is rare or absent (very low cM/Mb; no hotspots)

```
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
*     *            *               *  *             *
```

New adaptive alleles arising by mutation periodically

- Sometimes in regions where recombination is rare or absent (very low cM/Mb; no hotspots)

```
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
*
```

New adaptive alleles arising by mutation periodically

- Sometimes in regions where recombination is common (high cM/Mb; many hotspots)

```
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
      *               *                 *   *             *
```

New adaptive alleles arising by mutation periodically

- Sometimes in regions where recombination is common (high cM/Mb; many hotspots)

```
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
*       *               *                   *   *               *
```

New adaptive alleles arising by mutation periodically

- Sometimes in regions where recombination is common (high cM/Mb; many hotspots)

```
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
*      *               *                   *   *               *
```

New adaptive alleles arising by mutation periodically

- Sometimes in regions where recombination is common (high cM/Mb; many hotspots)

```
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
GCATGCTAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
GCATGCTAGTCGATTGCGGTATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
GCATGCTAGTCGATTGCGGTATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGATTTC
ACATGCGAGTCGATTGCGGTATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
ACATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
ACATGCGAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
*       *               *                   *   *               *
```

New adaptive alleles arising by mutation periodically

- Sometimes in regions where recombination is common (high cM/Mb; many hotspots)

```
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
GCATGCTAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
GCATGCTAGTCGATTGCGGTATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
GCATGCTAGTCGATTGCGGTATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATGCCCCGGTAGCTGCTGGGTTTC
GCATGCTAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
GCATGCTAGTCGATTGCGGTATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGATTTC
GCATGCTAGTCGATTGCGGTATTTTCGGAGCTGATACCTCGGTAGCTGCTGGGTTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATACCCCGGTAGCTGCTGGGTTTC
GCATGCTAGTCGATTGCGGCATTTTCGGAGCTGATGCCTCGGTAGCTGCTGGGTTTC
*                    *                   *   *               *
```

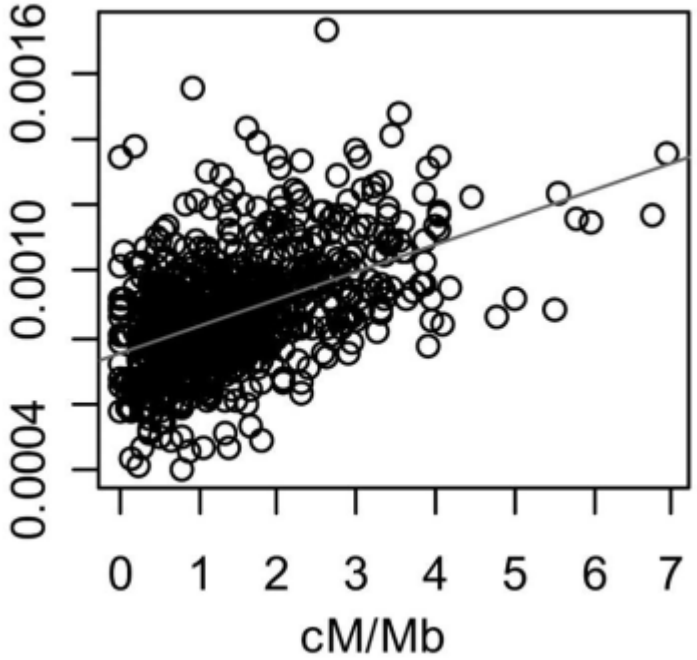
Prediction:

- If you survey regions around the genome, what association should there be between recombination rate and pi?
  - No relationship?
  - Positive (high cM/Mb -> high pi)?
  - Negative (high cM/Mb -> low pi)?

Association in humans

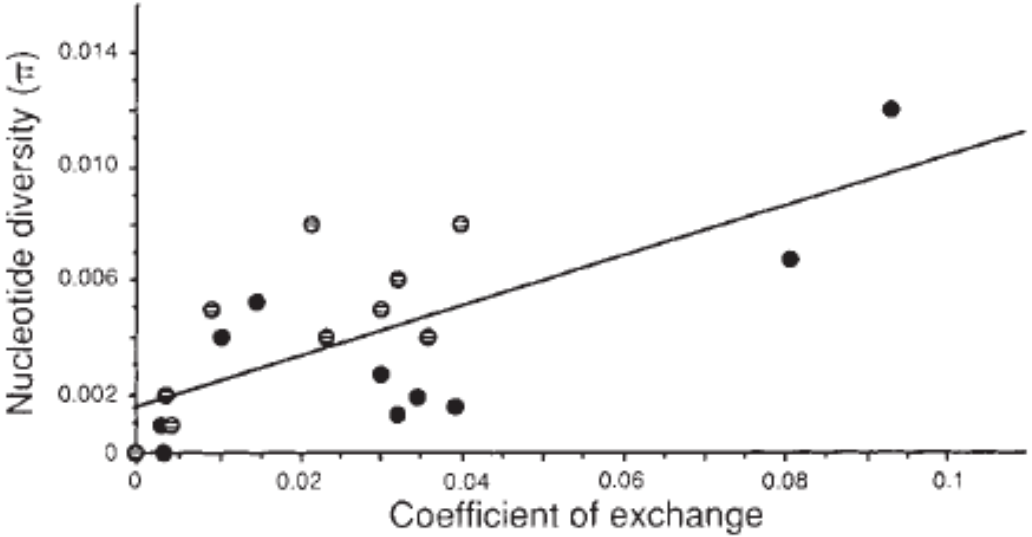
- Matches expectation: high pi in regions of high recombination

- Hellman et al, 2005, *Genome Research*



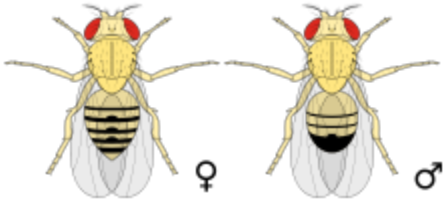
Association in Drosophila

- Matches expectation: high pi in regions of high recombination



Begun & Aquadro, 1992, *Nature*

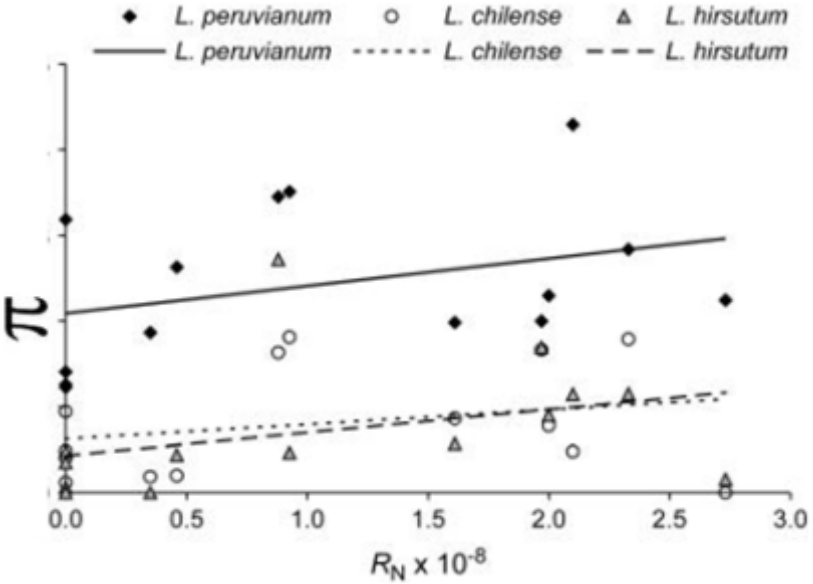
Paper cited >500 times!



Association in tomato species

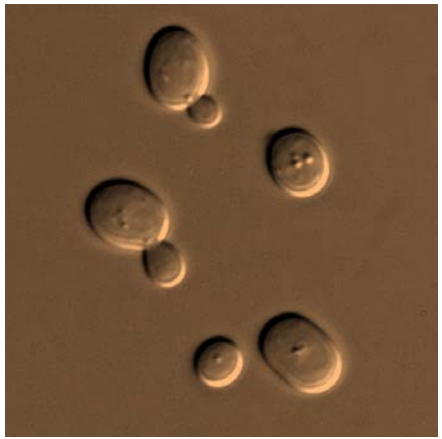
- Matches expectation: high pi in regions of high recombination (but weakly statistically significant)

- Roselius et al, 2005, *Genetics*



Association in yeast

- FAILS to match expectation- no significant positive or negative relationship
  - Perhaps associated with more frequent asexual reproduction?



What does this mean???

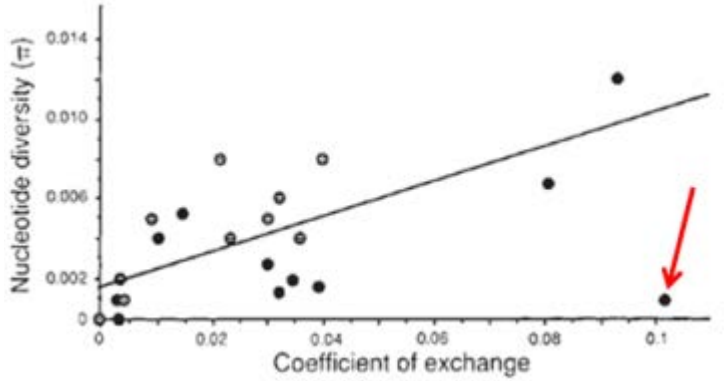
- We see this positive association between recombination rate and pi in many species
- This relationship is predicted from the action of natural selection

What does this mean???

- We see this positive association between recombination rate and pi in many species
- This relationship is predicted from the action of natural selection
- SELECTION'S IMPACT ON GENOMES OF MOST SPECIES IS **HUGE**! SELECTION IS **STRONG**!

Can we leverage this?

- We see that low recombination regions often have low variation (π)
- What if you see very low π in a small region of high recombination?
  - What might you infer happened?



Can we leverage this?

- We see that low recombination regions often have low variation (π)
- What if you see very low π in a small region of high recombination?
  - What might you infer happened?

- **A recent selective sweep!**
  - Can use to find genes that experienced recent selection!

Background selection!

An alternative explanation for why pi and recombination rate are correlated…

…and a continuing, heated debate!

Image Credits, Unit 14-3

- Recombination rate, © 2002 Reprinted by permission from Macmillan Publishers Ltd: Nature, Kong, A., D. F. Gudbjartsson, et al. (2002). "A high-resolution recombination map of the human genome." Nat Genet 31(3): 241-247, all rights reserved
- Crowd, © 2006 Ben Schumin, CC by 2.5, en.wikipedia.org
- Human recombination rates, © 2005 Ines Hellmann, all rights reserved, http://genome.cshlp.org/content/15/9/1222
- Nucleotide diversity, © 1992 Reprinted by permission from Macmillan Publishers Ltd: Nature, David J. Begun & Charles F. Aquadro (1992). "Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster." Nature 356: 519-520, all rights reserved
- Tomatoes, © 2011 Pluma, CC by-SA 3.0, en.wikipedia.org
- Tomato recombination rate, © 2005 Republished with permission from the Genetics Society of America, Kerstin Roselius, Wolfgang Stephan and Thomas Städler, "The Relationship of Nucleotide Polymorphism, Recombination Rate and Selection in Wild Tomato Species", Genetics October 2005 vol. 171 no. 2 753-763, all rights reserved