Recombination, selective sweeps, and hitchhiking
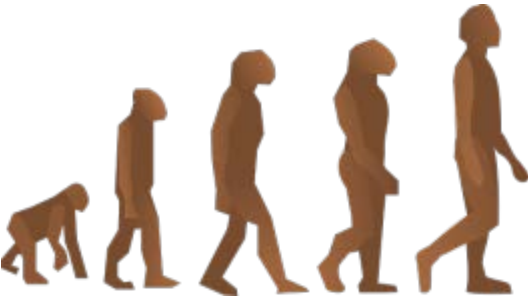
¡VIVA LA EVOLUCIÓN!

… but how does recombination affect molecular evolution?

- We've shown recombination
  - Can combine good mutations
  - Helps selection get rid of bad mutations from a population

- But what about how it affects variation in **neutral** sequences (have no effect on fitness)
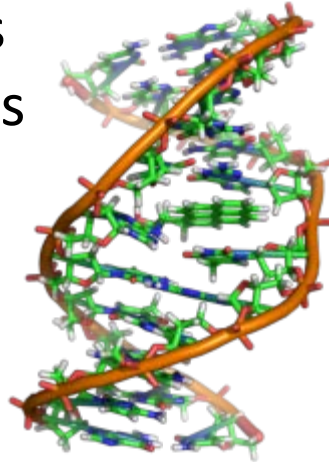
First, how do we "quantify" molecular variation?

- CAGA**G**ATCGATCG**A**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**T**AATTA
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATTA

One means: π ("pi")

- π = average number of pairwise mismatches $(2pq)$
- AG**C** TAT ATT CA**T** CAC GG**G** TT
- AG**T** TAT ATT CA**C** CAC GG**A** TT
- AG**C** TAT ATT CA**T** CAC GG**G** TT

One means: π ("pi")

- π = average number of pairwise mismatches (2pq)

- AG**C** TAT ATT CA**T** CAC GG**G** TT ⎤
                                          ⎬ **3**
- AG**T** TAT ATT CA**C** CAC GG**A** TT ⎦

- AG**C** TAT ATT CA**T** CAC GG**G** TT

One means: π ("pi")

- π = average number of pairwise mismatches (2pq)

- AG**C** TAT ATT CA**T** CAC GG**G** TT

- AG**T** TAT ATT CA**C** CAC GG**A** TT

- AG**C** TAT ATT CA**T** CAC GG**G** TT

3

3

One means: π ("pi")

- π = average number of pairwise mismatches (2pq)

- AG**C** TAT ATT CA**T** CAC GG**G** TT

- AG**T** TAT ATT CA**C** CAC GG**A** TT

- AG**C** TAT ATT CA**T** CAC GG**G** TT

3

3

0

One means: π ("pi")

- π = average number of pairwise mismatches (2pq)

- AG**C** TAT ATT CA**T** CAC GG**G** TT

- AG**T** TAT ATT CA**C** CAC GG**A** TT

- AG**C** TAT ATT CA**T** CAC GG**G** TT

3

3

0

- π = (3 + 3 + 0) / number of sequences (3) = **2**
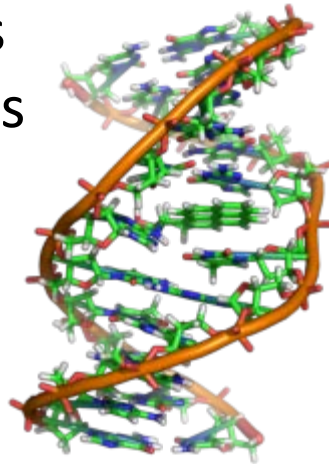
- π *per site* = **2** / # bases (20) = **0.1** (10%)

One means: π ("pi")

- π = average number of pairwise mismatches (2pq)

- AG**C** TAT ATT CA**T** CAC GG**G** TT
- AG**T** TAT ATT CA**C** CAC GG**A** TT   **3**
- AG**C** TAT ATT CA**T** CAC GG**G** TT   **3**   **0**

- π *per site* = **2** / # bases (20) = **0.1** (10%)

- π *per site* is greater when there are more bases differing among individuals

… but how does recombination affect molecular evolution?

- We've shown recombination
  - Can combine good mutations
  - Helps selection get rid of bad mutations from a population

- But what about how it affects variation in **neutral** sequences (have no effect on fitness)

Let's first assume
no recombination

- CAGA**G**ATCGATCG**A**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**G**ATCGATCG**T**TCGCCCCTTTCG**T**AATTA
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**G**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATTA

1   2   3   4

Let's first assume
no recombination

       1   2   3   4

- CAGA**G**ATCGATCG**A**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**G**ATCGATCG**T**TCGCCCCTTTCG**T**AATTA
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**G**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**

Good mutation

Let's first assume
no recombination

       1   2   3   4

- CAGAGATCGATCGATCGCCCCTTTCGCAATTA
- CAGAAATCGATCGTTCGCCCCTTTCGCAATTA
- CAGAAATCGATCGATCGCCCCTTTCGTAATTA
- CAGAGATCGATCGTTCGCCCCTTTCGTAATTA
- CAGAAATCGATCGATCGCCCCTTTCGCAATTG
- CAGAAATCGATCGATCGCCCCTTTCGCAATTG
- CAGAAATCGATCGATCGCCCCTTTCGCAATTG
- CAGAAATCGATCGATCGCCCCTTTCGCAATTG
- CAGAAATCGATCGATCGCCCCTTTCGCAATTG

Good
mutation

Let's first assume
no recombination

        1   2   3   4

- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**

Hitchhiking with a sweep

- "Selective sweep" is spread of an advantageous allele throughout population by natural selection and associated loss of variation near it
  - New mutation at base 4 generated a "selective sweep"
- "Hitchhiking" is spread of other nearby alleles along with the advantageous one because of linkage (lack of recombination)
  - SNPs 1-3 (A-A-C) all "hitchhiked" with base 4
  - **ALL** old variation at those SNPs was lost!

… but what if there WAS lots of **recombination** among the sites???

- CAGA**G**ATCGATCG**A**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**G**ATCGATCG**T**TCGCCCCTTTCG**T**AATTA
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATTA
- CAGA**G**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**

1 2 3 4

Good mutation

… but what if there WAS lots of **recombination** among the sites???

- CAGA**G**ATCGATCG**A**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**G**ATCGATCG**T**TCGCCCCTTTCG**T**AATTA
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATTA
- CAGA**G**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**

1   2   3   4

Good mutation

Get some recombination:

- One heterozygous individual experiences crossing over before making gamete:

CAGAGATCGATCGTTCGCCCCTTTCGTAATTA

CAGAAATCGATCGATCGCCCCTTTCGCAATTG

Get some recombination:

- One heterozygous individual experiences crossing over before making gamete:

CAGAGATCGATCGTTCGCCCCTTTCGTAATTA

CAGAAATCGATCGATCGCCCCTTTCGCAATTG

Get some recombination:

- One heterozygous individual experiences crossing over before making gamete:

CAGA**G**ATCGATCG**T**TCGCCCCTTTCG**T**AATTA

CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**

CAGA**G**ATCGATCG**T**TCGCCCCTTTCG**C**AATT**G**

… but what if there WAS lots of **recombination** among the sites???

- CAGA**G**ATCGATCG**A**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**T**AATTA
- CAGA**G**ATCGATCG**T**TCGCCCCTTTCG**T**AATTA
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**C**AATTA
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATTA
- CAGA**G**ATCGATCG**T**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**

1   2   3   4

Good mutation

… but what if there WAS lots of **recombination** among the sites???

- CAGA**G**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G** ^{4}
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**T**AATT**G**
- CAGA**G**ATCGATCG**T**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**T**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**
- CAGA**G**ATCGATCG**T**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**
- CAGA**A**ATCGATCG**A**TCGCCCCTTTCG**C**AATT**G**

Good mutation

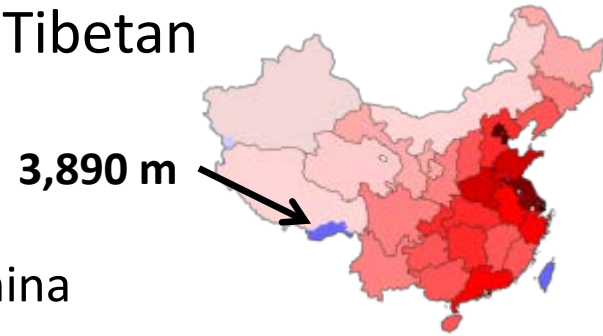# Recap- what happened with recombination?

- Mutation at site 4 still spread

- Recombination allowed SNPs 1 and 2 to maintain variation (2 abundant alleles)

  – Other alleles recombined onto chromosomes bearing the sweeping allele at site 4 (π not reduced in left)

Recap- what happened with recombination?

- Mutation at site 4 still spread

- Recombination allowed SNPs 1 and 2 to maintain variation (2 abundant alleles)
  - Other alleles recombined onto chromosomes bearing the sweeping allele at site 4 (π not reduced in left)

- At SNP 3, **MOST** of the population got one allele
  - **Most** chromosomes have same allele at SNP 3 after sweep
  - One SNP 3 allele became very common by hitchhiking
    - **But other allele was not completely lost**
   - **π greatly reduced in right half (unlike left)**
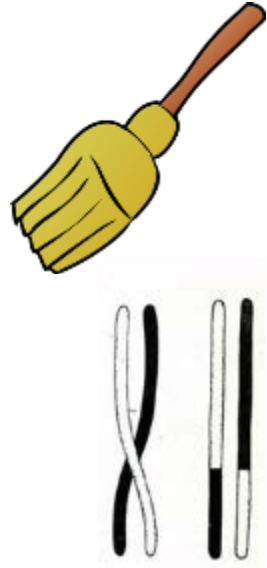
Example: variation at SNP near *EPAS1*

- *EPAS1* affects tolerance to low oxygen

- Evidence for recent sweep & hitchhiking in Tibetan (high-altitude) populations

- "A" SNP frequency in Europe: 0.46
  - Mutation for tolerance to low O_2 near SNP in China
    - Arose on chromosome with "A" allele at SNP

3,890 m

Example: variation at SNP near *EPAS1*

- *EPAS1* affects tolerance to low oxygen

- Evidence for recent sweep & hitchhiking in Tibetan (high-altitude) populations

3,890 m

- "A" SNP frequency in Europe: 0.46
  - Mutation for tolerance to low O_2 near SNP in China
    - Arose on chromosome with "A" allele at SNP

- "A" SNP frequency in Tibet: 0.89
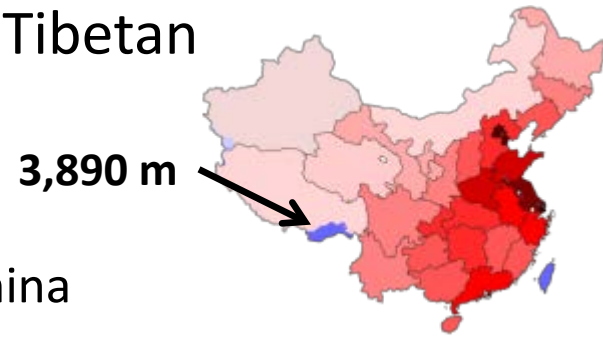
Three Principles:

- Adaptive alleles can "sweep"

- SNP/ marker alleles near adaptive allele (with zero or low recombination) hitchhike
  - π typically reduced

- SNP/ marker alleles far from adaptive allele (with high recombination) don't hitchhike
  - π not typically reduced

Example from a lab exercise

- Started with white-eyed population
- Introduced advantageous red-eye allele
  - Will "Near" or "Far" marker allele hitchhike?
  - Do we expect π to be reduced at "Near" region?
  - Do we expect π to be reduced at "Far" region?

Red / White
eye color gene

22 million bases of DNA

"Near"   "Far"