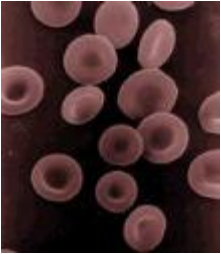




# Differences between populations: Origins and quantifying





# Recap: Calculating genotype/ allele frequencies, and testing HW

- MN blood type in Navajo
    - MM: 305 0.845      $p(M) = 0.845 + \frac{1}{2} (0.144) = 0.917$
    - MN: 52     0.144      $q(N) = 0.011 + \frac{1}{2} (0.144) = 0.083$
    - NN: 4     0.011
- TOTAL: **361**

HW Predicted:

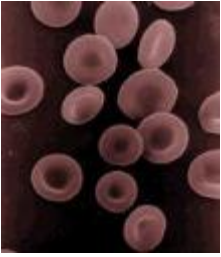
$$MM \quad p^2 = 0.841$$

$$MN \quad 2pq = 0.152$$

$$NN \quad q^2 = 0.007$$

HW





# Recap: Wahlund effect in mixed population

- MN blood type for Aborigine + Navajo
    - MM: 327 0.300  $p(M) = 0.423$
    - MN: 268 0.246  $q(N) = 0.577$
    - NN: 496 0.454
- TOTAL: **1091**

HW Predicted:

$$MM \quad p^2 = 0.179$$

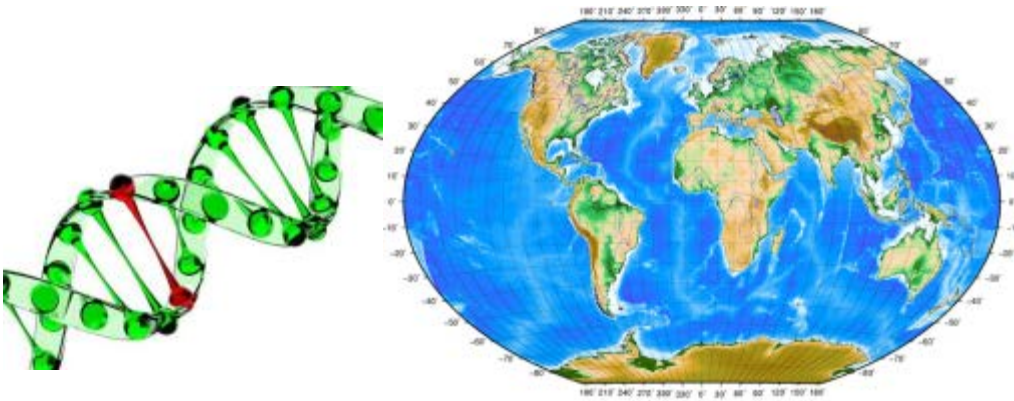
e to HW  $MN \quad 2pq = 0.488$

$$NN \quad q^2 = 0.333$$



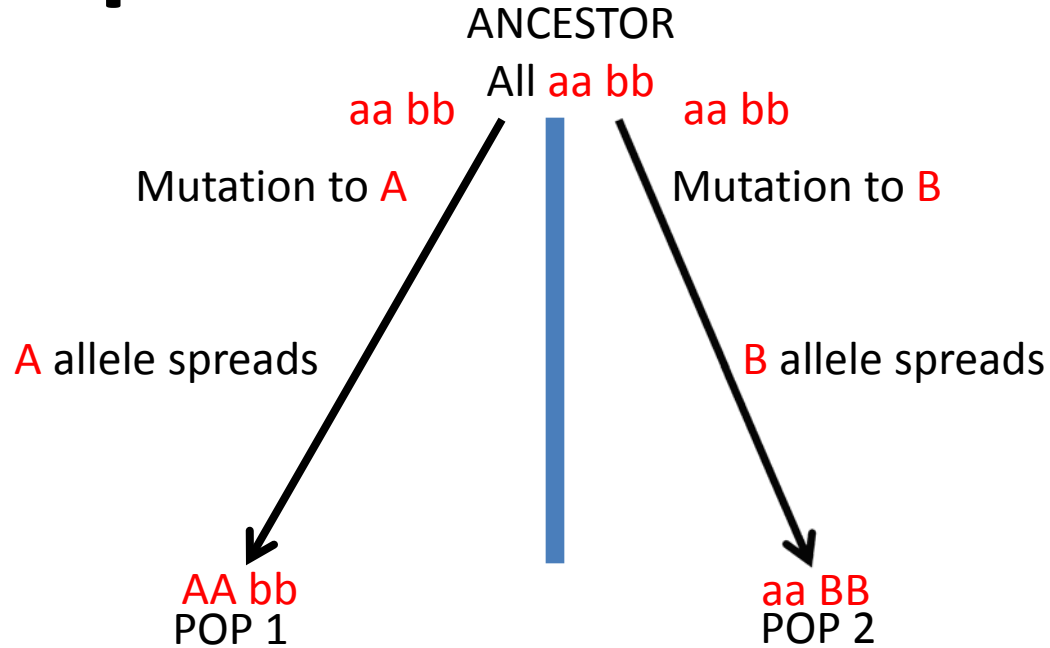
# Populations differ...

- **May have different allele & genotype frequencies**
- May have alleles at some genes not found in other populations if
  - Very recent new mutation, or
  - Complete isolation



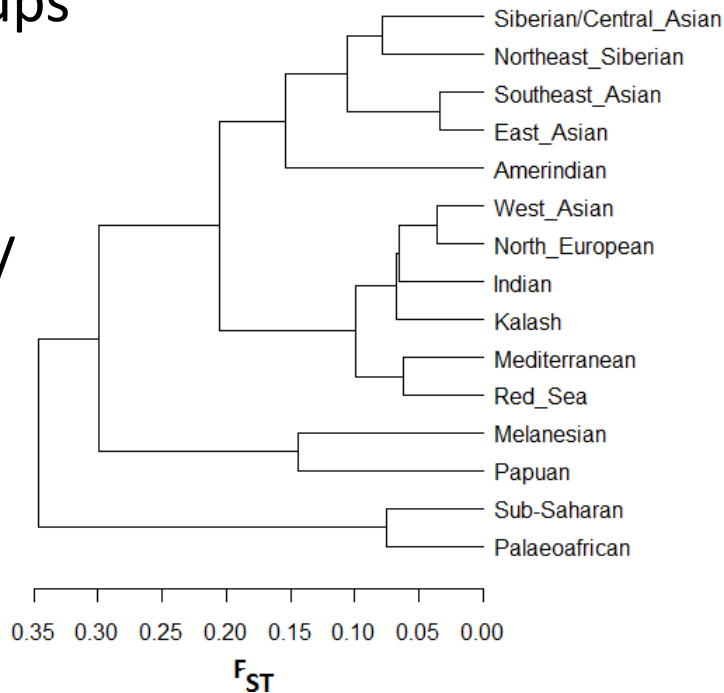
# Difference(s) arose via mutation, then spread

- Ancestors **aa bb**
- Population splits into two groups
  - In #1, new mutation to **A**
  - In #2, new mutation to **B**
- New mutations spread



# Groups within species are different yet related

- May have alleles not found in other groups
- May have very different genotype/allele frequencies



# How to quantify?

- **Simplest to quantify**- all individuals differ
  - “Fixed” difference
    - Population 1: all AA      Population 2: all aa



# How to quantify?



- **Simplest to quantify**- all individuals differ
  - “Fixed” difference
    - Population 1: all AA      Population 2: all aa
- Happens, but not very common within a species, and generally not true among modern human ethnic groups



# How to quantify?



- **More common-** frequency differences of alleles & genotypes
  - Population 1  $p(A)=0.7$
  - Population 2  $p(A)=0.5$
- Measure differences between **POPULATIONS** not **INDIVIDUALS**
  - Can't be applied to individual
- How “quantify”?

# Deviation from HW allows you to quantify allele freq differences!

- Assume two populations at HW
  - If sample *each by itself*, see HW
  - If sample *both together*, see deviation from HW
    - Wahlund effect
- ***How big the deviation is from HW*** when sampling both together quantifies difference in allele frequencies



# Measure we'll use: $F_{ST}$

- $F_{ST}$  ranges from 0 to 1
  - 0 : no allele frequency differences
  - $0 < F_{ST} < 1$  : allele frequencies differ somewhat
  - 1 : “fixed” difference between populations
- $$F_{ST} = \frac{\text{HW predicted } 2pq - \% \text{ observed hetz}}{\text{HW predicted } 2pq}$$



# $F_{ST}$ Example 1



- **POPULATION 1**

– AA	100
– Aa	0
– aa	0

**POPULATION 2**

AA	0
Aa	0
aa	100

TOTAL- AA: 100, Aa: 0, aa: 100



# $F_{ST}$ Example 1



- **POPULATION 1**

– AA	100
– Aa	0
– aa	0

**POPULATION 2**

AA	0
Aa	0
aa	100

TOTAL- AA: 100, Aa: 0, aa: 100 N=200

$$AA: 100/200 = 0.5$$

$$p(A) = 0.5 ; q(a) = 0.5$$

$$Aa: 0/200 = 0$$

$$aa: 100/200 = 0.5$$



# $F_{ST}$ Example 1



- POPULATION 1

– AA	100
– Aa	0
– aa	0

- POPULATION 2

AA	0
Aa	0
aa	100

TOTAL- AA: 100, Aa: 0, aa: 100 N=200

$$AA: 100/200 = 0.5$$

$$p(A) = 0.5 ; q(a) = 0.5$$

$$Aa: 0/200 = 0$$

$$HW \ 2pq = \mathbf{0.50}$$

$$aa: 100/200 = 0.5$$



# $F_{ST}$ Example 1



- POPULATION 1

– AA	100
– Aa	0
– aa	0

POPULATION 2

AA	0
Aa	0
aa	100

TOTAL- AA: 100, Aa: 0, aa: 100 N=200

$$AA: 100/200 = 0.5$$

$$p(A) = 0.5 ; q(a) = 0.5$$

$$Aa: 0/200 = \mathbf{0}$$

$$HW \ 2pq = \mathbf{0.50}$$

$$aa: 100/200 = 0.5$$



# $F_{ST}$ Example 1



- POPULATION 1

– AA	100
– Aa	0
– aa	0

- POPULATION 2

AA	0
Aa	0
aa	100

TOTAL- AA: 100, Aa: 0, aa: 100 N=200

$$AA: 100/200 = 0.5$$

$$p(A) = 0.5 ; q(a) = 0.5$$

$$Aa: 0/200 = 0$$

$$HW \ 2pq = 0.50$$

$$aa: 100/200 = 0.5 \quad (\text{predicted-obs})/\text{predicted}$$

$$F_{ST} = (0.50-0)/(0.50) = 1.00$$



# $F_{ST}$ Example 2



- **POPULATION 1**

– AA	250
– Aa	500
– aa	250

**POPULATION 2**

AA	490
Aa	420
aa	90

TOTAL- AA: 740, Aa: 920, aa: 340

# $F_{ST}$ Example 2



- **POPULATION 1**

– AA	250
– Aa	500
– aa	250

**POPULATION 2**

AA	490
Aa	420
aa	90

TOTAL- AA: 740, Aa: 920, aa: 340 N=2000

$$AA: 740/2000 = 0.37$$

$$Aa: 920/2000 = 0.46$$

$$aa: 340/2000 = 0.17$$

$$p(A) = 0.6; q(a) = 0.4$$

# $F_{ST}$ Example 2



- **POPULATION 1**

– AA	250
– Aa	500
– aa	250

**POPULATION 2**

AA	490
Aa	420
aa	90

TOTAL- AA: 740, Aa: 920, aa: 340 N=2000

$$AA: 740/2000 = 0.37$$

$$Aa: 920/2000 = 0.46$$

$$aa: 340/2000 = 0.17$$

$$p(A) = 0.6; q(a) = 0.4$$

$$HW \ 2pq = \mathbf{0.48}$$

# $F_{ST}$ Example 2



- **POPULATION 1**

– AA	250
– Aa	500
– aa	250

**POPULATION 2**

AA	490
Aa	420
aa	90

TOTAL- AA: 740, Aa: 920, aa: 340 N=2000

AA:  $740/2000 = 0.37$

Aa:  $920/2000 = \mathbf{0.46}$

aa:  $340/2000 = 0.17$

$p(A) = 0.6$ ;  $q(a) = 0.4$

HW  $2pq = \mathbf{0.48}$

# $F_{ST}$ Example 2



- **POPULATION 1**

– AA	250
– Aa	500
– aa	250

**POPULATION 2**

AA	490
Aa	420
aa	90

TOTAL- AA: 740, Aa: 920, aa: 340 N=2000

AA:  $740/2000 = 0.37$

Aa:  $920/2000 = \mathbf{0.46}$

aa:  $340/2000 = 0.17$

$p(A) = 0.6$ ;  $q(a) = 0.4$

HW  $2pq = \mathbf{0.48}$

$$F_{ST} = (0.48 - 0.46) / (0.48) = \mathbf{0.042}$$

FROM EARLIER LECTURE...

# Real data: Mixed population

- MN blood type

- MM: 327

- MN: 268

- NN: 496

## OBSERVED

0.300

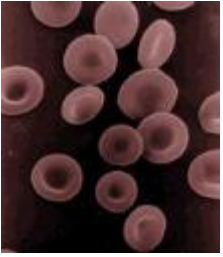
0.246

0.454

$p(M) = 0.423$

$q(N) = 0.577$

$F_{ST} =$



FROM EARLIER LECTURE...

# Real data: Mixed population

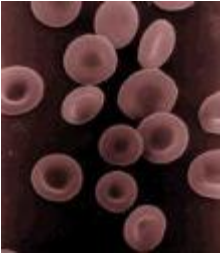
- MN blood type
  - MM: 327
  - MN: 268
  - NN: 496
- |           | OBSERVED |                |
|-----------|----------|----------------|
| – MM: 327 | 0.300    | $p(M) = 0.423$ |
| – MN: 268 | 0.246    | $q(N) = 0.577$ |
| – NN: 496 | 0.454    |                |

$$F_{ST} = \frac{\text{HW predicted } 2pq - \% \text{ observed hetz}}{\text{HW predicted } 2pq}$$





FROM EARLIER LECTURE...

# Real data: Mixed population



• MN blood type	OBSERVED		EXPECTED
– MM: 327	0.300	$p(M) = 0.423$	0.179
– MN: 268	0.246	$q(N) = 0.577$	0.488
– NN: 496	0.454		0.333


$$F_{ST} = \frac{\text{HW predicted } 2pq - \% \text{ observed hetz}}{\text{HW predicted } 2pq}$$



# Recap...

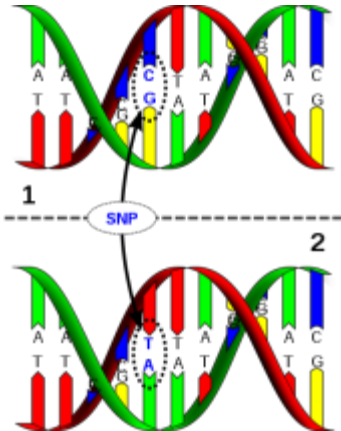
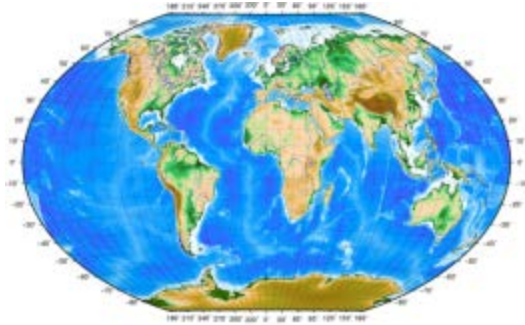
- $F_{ST}$  larger comparing populations that are more different in allele frequencies
  - Aborigine & Navajo VERY different in allele frequencies
- If allele frequencies were identical,  $F_{ST}$  would be 0
- If fixed different,  $F_{ST}$  would be 1



# $F_{ST}$ measures among human populations

Data from 1,110,338 SNPs, 2010 study

- African Americans – Europeans:  $F_{ST} = 0.11$
- African Americans – Chinese:  $F_{ST} = 0.15$
- Europeans – Chinese:  $F_{ST} = 0.11$



$F_{ST}$  among European populations is  $<0.01$

# BIG table of $F_{ST}$ measures



	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Red Sea														
2 East Asian	0.142													
3 Kalash	0.099	0.13												
4 Northeast Siberian	0.175	0.083	0.158											
5 North European	0.068	0.122	0.068	0.146										
6 Palaeoafrican	0.217	0.262	0.248	0.301	0.233									
7 Sub-Saharan	0.158	0.202	0.189	0.24	0.176	0.075								
8 Mediterranean	0.062	0.143	0.092	0.176	0.057	0.238	0.178							
9 Papuan	0.239	0.207	0.234	0.259	0.224	0.346	0.285	0.242						
10 Southeast Asian	0.144	0.034	0.133	0.106	0.125	0.263	0.203	0.146	0.212					
11 West Asian	0.062	0.122	0.063	0.15	0.036	0.223	0.166	0.062	0.222	0.125				
12 Indian	0.085	0.081	0.067	0.119	0.065	0.215	0.157	0.084	0.182	0.085	0.06			
13 Melanesian	0.219	0.173	0.213	0.227	0.203	0.329	0.268	0.221	0.144	0.174	0.201	0.162		
14 Amerindian	0.204	0.136	0.185	0.152	0.167	0.333	0.271	0.205	0.299	0.154	0.175	0.149	0.27	
15 Siberian/Central Asian	0.153	0.059	0.138	0.078	0.126	0.277	0.216	0.151	0.232	0.081	0.129	0.097	0.201	0.144

# What is $F_{ST}$ , in words?

- $F_{ST}$  is the % heterozygous of randomly chosen alleles within populations (observed) relative to that expected in the entire species ( $2pq$ )
  - Measures ***difference in allele frequencies***
    - If identical allele frequencies,  $F_{ST} = 0$
    - If fixed for different alleles,  $F_{ST} = 1$
- Why don't we see higher  $F_{ST}$  among human populations???



# Some $F_{ST}$ assumptions violated in humans

- Supposed to be applied to genes experiencing little/ no natural selection
- Susceptible to differences (and historic changes) in population size among groups
- ... but biggest reason  $F_{ST}$  values aren't larger...



# Image Credits, Unit 6-1

- African children, © 2003 Thomas Schoch, CC by-SA 3.0, en.wikipedia.org
- Aborigines, © 2007 Mombas, CC by-SA 3.0, en.wikipedia.org
- World map, © 2005 David Monniaux, CC by-SA 3.0, en.wikipedia.org
- Green DNA w/ red mutation, © appler, all rights reserved, www.photoxpress.com
- Cladogram, © 2010 Dienekes, all rights reserved, <http://dienekes.blogspot.com/>
- Lightbulb, © 2010 Jacob Henri 6, CC by-SA 3.0, en.wikipedia.org
- DNA SNP, © 2007 David Hall, CC by-SA 3.0, en.wikipedia.org
- Global diversity cartoon, © Dawn Hudson, all rights reserved, www.photoxpress.com
- FST table, © 2010 Dienekes, all rights reserved, <http://dienekes.blogspot.com/>
- Crowd, © 2006 SchuminWeb, CC by-SA 2.5, en.wikipedia.org