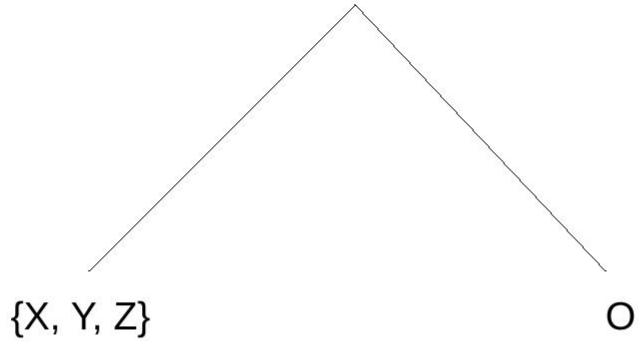
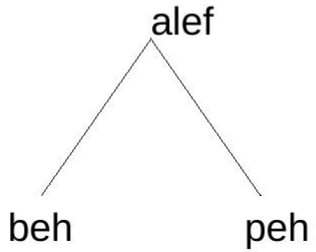
F4 test for investigating admixture events

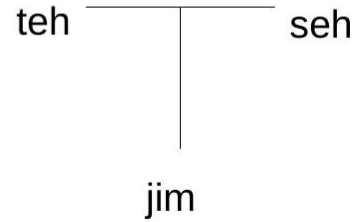
Hassan Shafiey
July 23th, 2018

[From lesson 6] Interpretation of the f4 is not straightforward, so we fix the position of one population as the outermost population in the tree(outgroup population). Now the investigation of other three is easier.

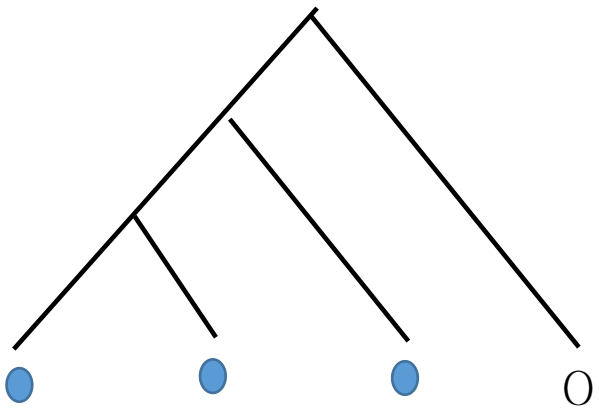


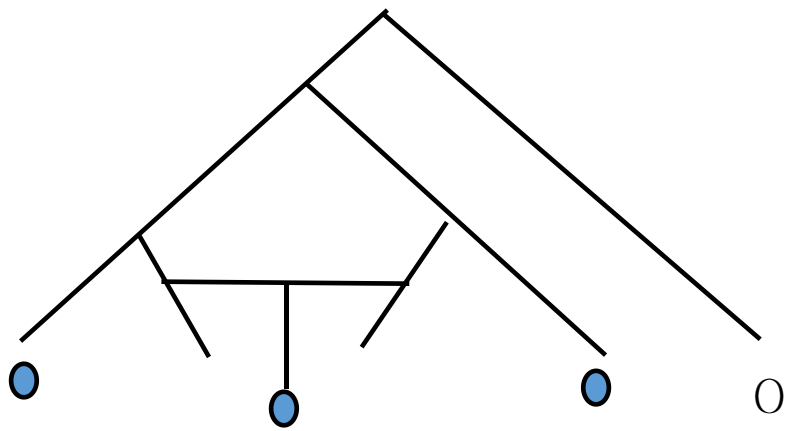
Two kinds of genetic event:

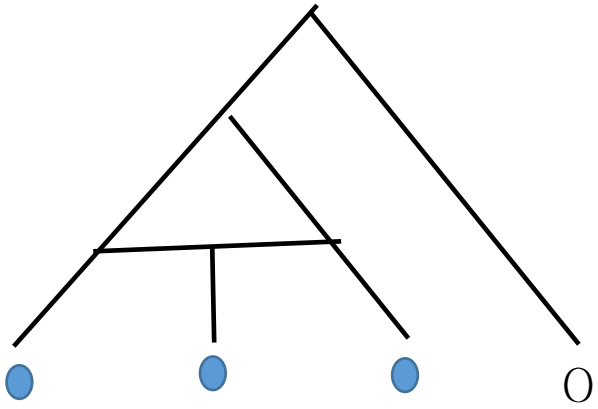




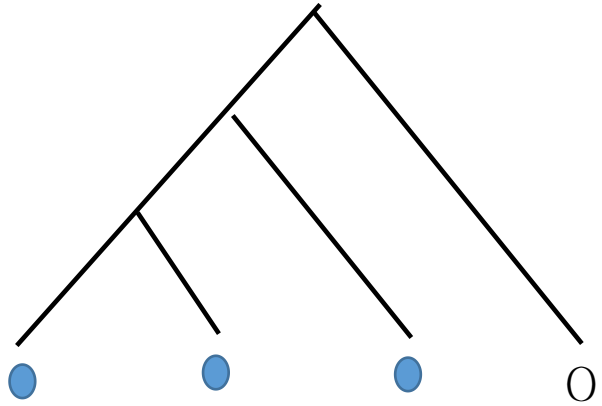
All possible configurations



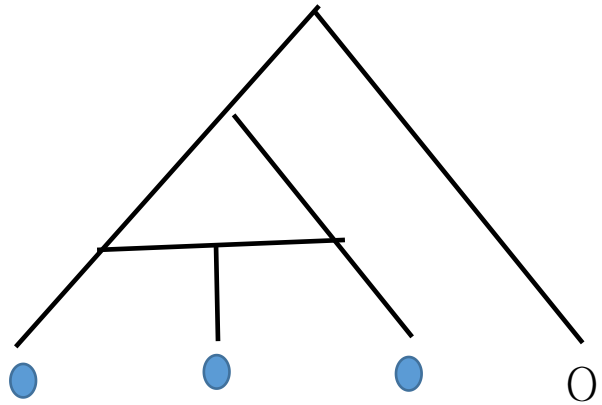




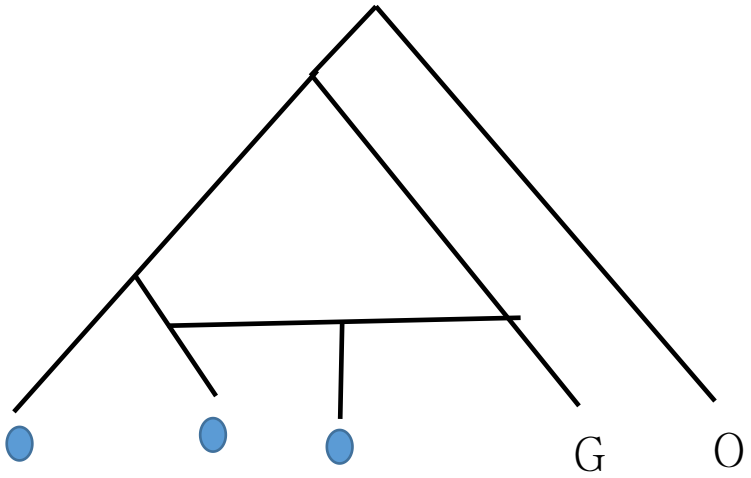
All possible configurations



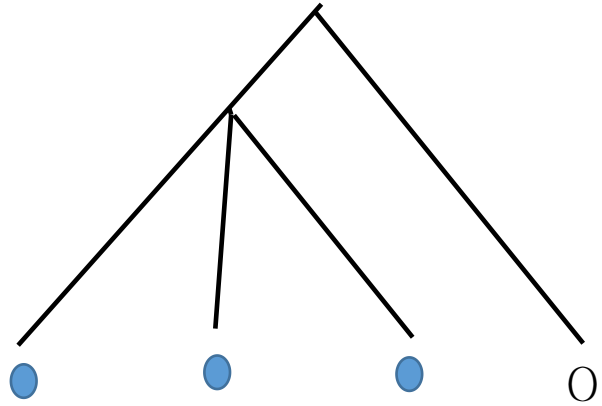
Cladal relationship



Admixture event



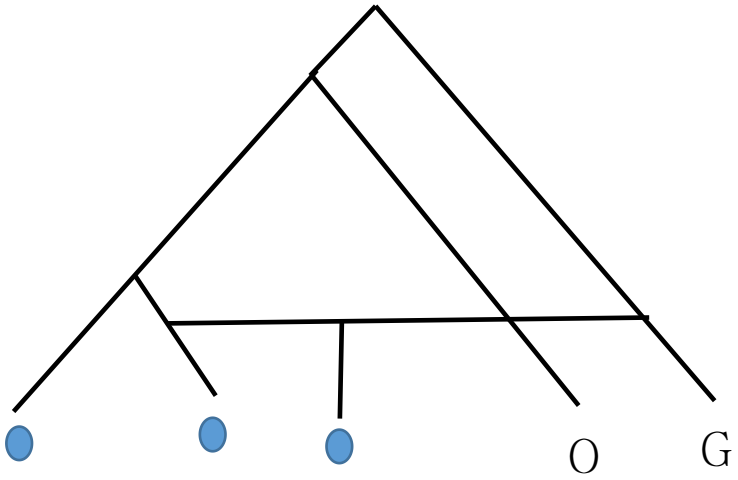
Presence of a ghost population



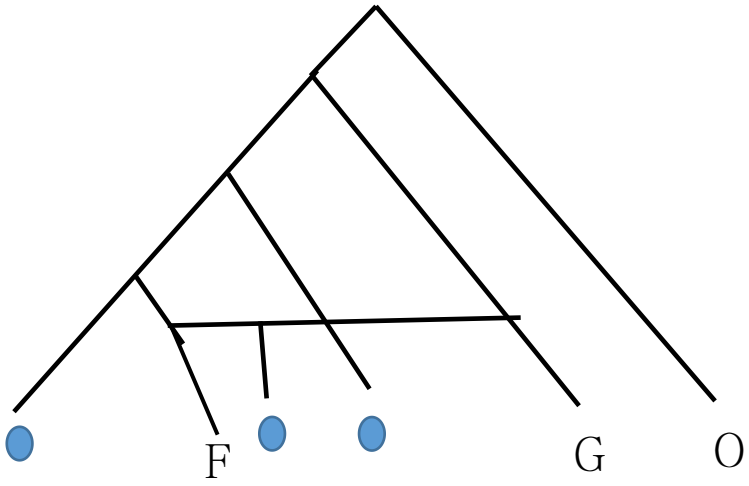
Just for completeness of our list

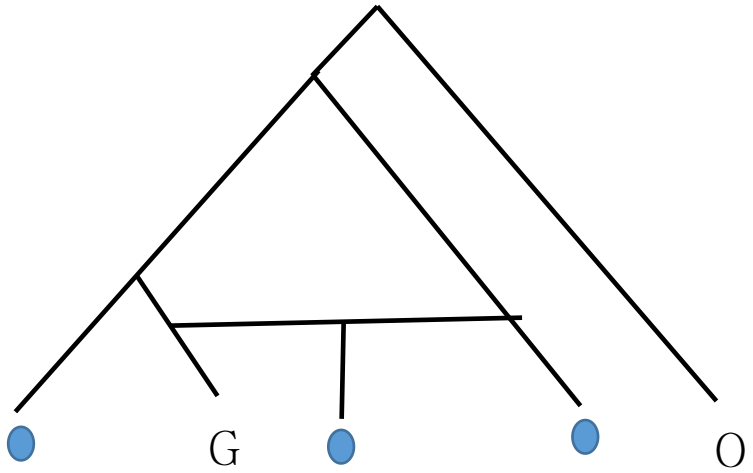
Quiz

what about these topologies: should we include them in our list?



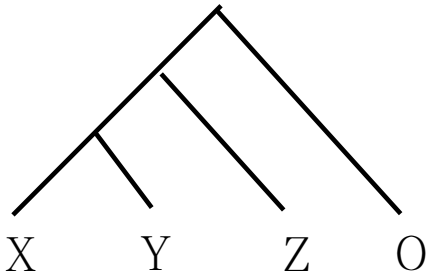
ghost population is outside of all pops



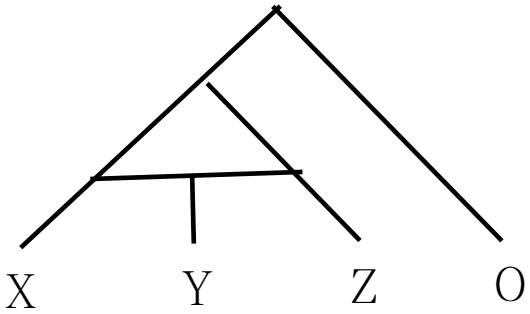


ghost population is inside the three pops

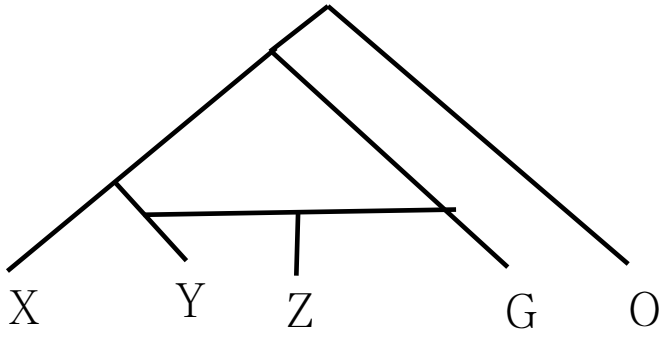
[From lesson 6] Three unique D values for each set of four pops



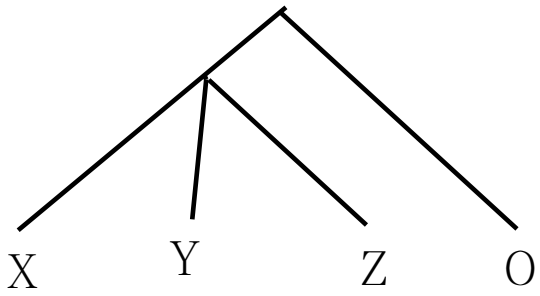
1
D1 (X, Z; Y, O) > 0
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) ~ 0



2
D1 (X, Z; Y, O) = any value
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) < 0



3
D1 (X, Z; Y, O) = any value
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) < 0



4
D1 (X, Z; Y, O) ~ 0
D2 (Y, Z; X, O) ~ 0
D3 (X, Y; Z, O) ~ 0

Example 1: Exploring {AFR, nAFR, Altai}

Quiz:

1. Which one is a correct (and the best) choice for an outgroup?

a. Mbuti (an African population)

b. Bonobo(an African ape)

c. Australian Mouse

d. Denisovan(a Asian sister group for Neanderthal)

e. Homo erectus (One million year old sample)

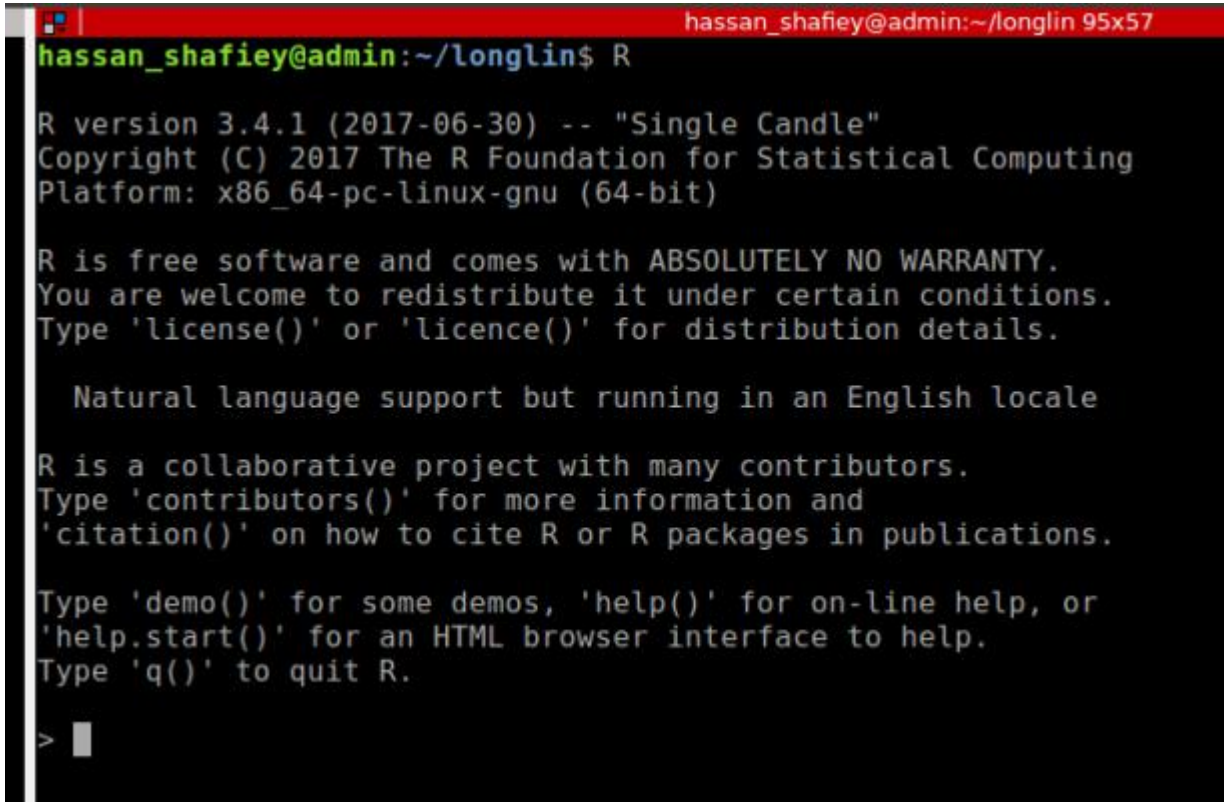
2. What will happen if you decide to put a plant in the place of the outgroup?

Example1: Exploring east Asians relative to Europeans

EAS = c("Han", "She", "Uygur", "Yi")

English (as European) and Chimp (as outgroup)

On the server type R to enter R statistical software



```
> EAS = c("Han", "She", "Uygur", "Yi")
> a = t(combn(EAS, 2)); a
     [,1]    [,2]
[1,] "Han"   "She"
[2,] "Han"   "Uygur"
[3,] "Han"   "Yi"
[4,] "She"   "Uygur"
[5,] "She"   "Yi"
[6,] "Uygur" "Yi"
>
>
> b = cbind(a[, 1], "English", a[, 2], "Chimp"); b
     [,1]    [,2]      [,3]    [,4]
[1,] "Han"   "English" "She"   "Chimp"
[2,] "Han"   "English" "Uygur" "Chimp"
[3,] "Han"   "English" "Yi"    "Chimp"
[4,] "She"   "English" "Uygur" "Chimp"
[5,] "She"   "English" "Yi"    "Chimp"
[6,] "Uygur" "English" "Yi"    "Chimp"
>
>
> b -> D1.ppf
> b[, c(3, 2, 1, 4)] -> D2.ppf
> b[, c(1, 3, 2, 4)] -> D3.ppf
>
> D1.ppf
     [,1]    [,2]      [,3]    [,4]
[1,] "Han"   "English" "She"   "Chimp"
[2,] "Han"   "English" "Uygur" "Chimp"
[3,] "Han"   "English" "Yi"    "Chimp"
[4,] "She"   "English" "Uygur" "Chimp"
[5,] "She"   "English" "Yi"    "Chimp"
[6,] "Uygur" "English" "Yi"    "Chimp"
>
> D2.ppf
     [,1]    [,2]      [,3]    [,4]
[1,] "She"   "English" "Han"   "Chimp"
[2,] "Uygur" "English" "Han"   "Chimp"
[3,] "Yi"    "English" "Han"   "Chimp"
[4,] "Uygur" "English" "She"   "Chimp"
[5,] "Yi"    "English" "She"   "Chimp"
[6,] "Yi"    "English" "Uygur" "Chimp"
>
> D3.ppf
     [,1]    [,2]    [,3]      [,4]
[1,] "Han"   "She"   "English" "Chimp"
[2,] "Han"   "Uygur" "English" "Chimp"
[3,] "Han"   "Yi"    "English" "Chimp"
[4,] "She"   "Uygur" "English" "Chimp"
[5,] "She"   "Yi"    "English" "Chimp"
[6,] "Uygur" "Yi"    "English" "Chimp"
>
> write.table(D1.ppf, "D1.ppf", quote = F, row.names = F, col.names =F )
> write.table(D2.ppf, "D2.ppf", quote = F, row.names = F, col.names =F )
> write.table(D3.ppf, "D3.ppf", quote = F, row.names = F, col.names =F )
> q()
Save workspace image? [y/n/c]: n
hassan_shafiey@admin:~/longlin$
```

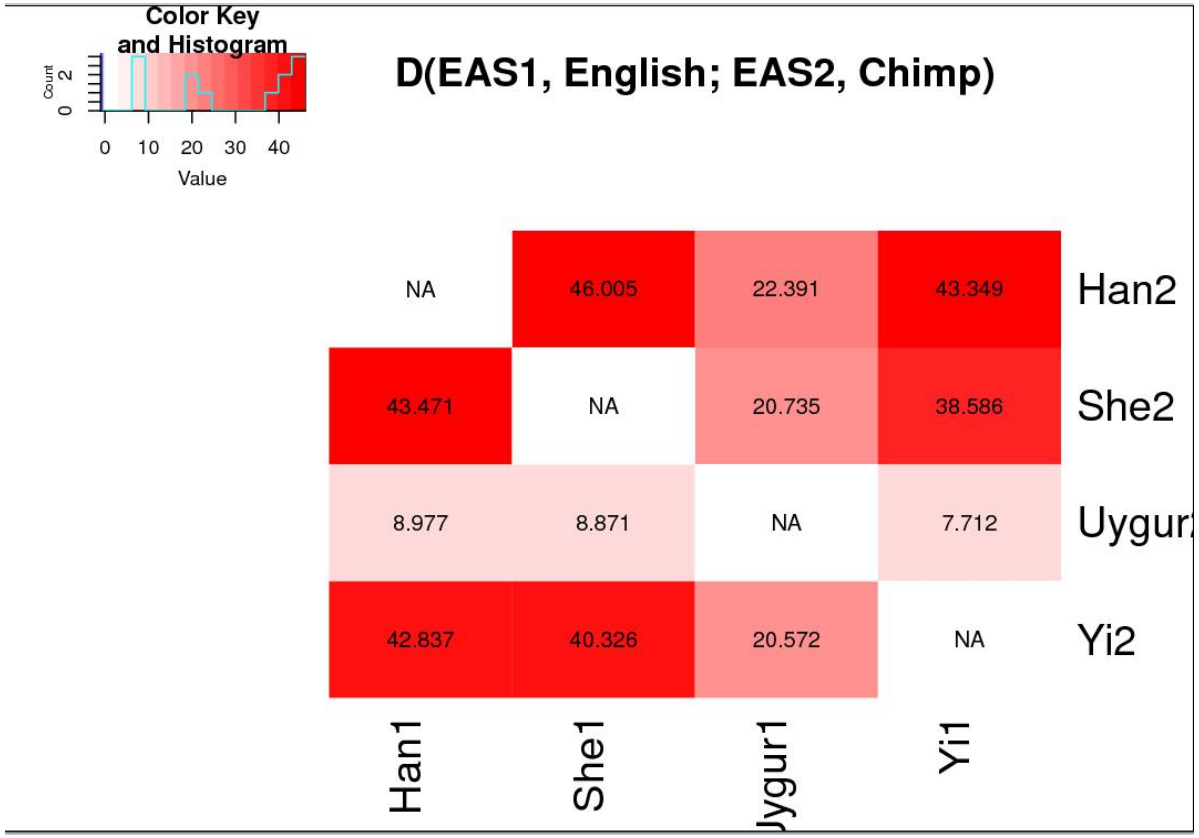
```
hassan_shafiey@admin:~/longlin$ cat D1.ppf
Han English She Chimp
Han English Uygur Chimp
Han English Yi Chimp
She English Uygur Chimp
She English Yi Chimp
Uygur English Yi Chimp
hassan_shafiey@admin:~/longlin$ cat D2.ppf
She English Han Chimp
Uygur English Han Chimp
Yi English Han Chimp
Uygur English She Chimp
Yi English She Chimp
Yi English Uygur Chimp
hassan_shafiey@admin:~/longlin$ cat D3.ppf
Han She English Chimp
Han Uygur English Chimp
Han Yi English Chimp
She Uygur English Chimp
She Yi English Chimp
Uygur Yi English Chimp
hassan_shafiey@admin:~/longlin$
hassan_shafiey@admin:~/longlin$ emacs general.par
hassan_shafiey@admin:~/longlin$ qpDstat -p general.par > D1.log&
[1] 25500
hassan_shafiey@admin:~/longlin$ emacs general.par
hassan_shafiey@admin:~/longlin$ qpDstat -p general.par > D2.log&
[2] 25571
hassan_shafiey@admin:~/longlin$ emacs general.par
hassan_shafiey@admin:~/longlin$ qpDstat -p general.par > D3.log&
[3] 26344
hassan_shafiey@admin:~/longlin$ jobs
[1]   Running                 qpDstat -p general.par > D1.log &
[2]-  Running                 qpDstat -p general.par > D2.log &
[3]+  Running                 qpDstat -p general.par > D3.log &
hassan_shafiey@admin:~/longlin$
```

qpDstat –p Lesson8_general.par > D1.log &

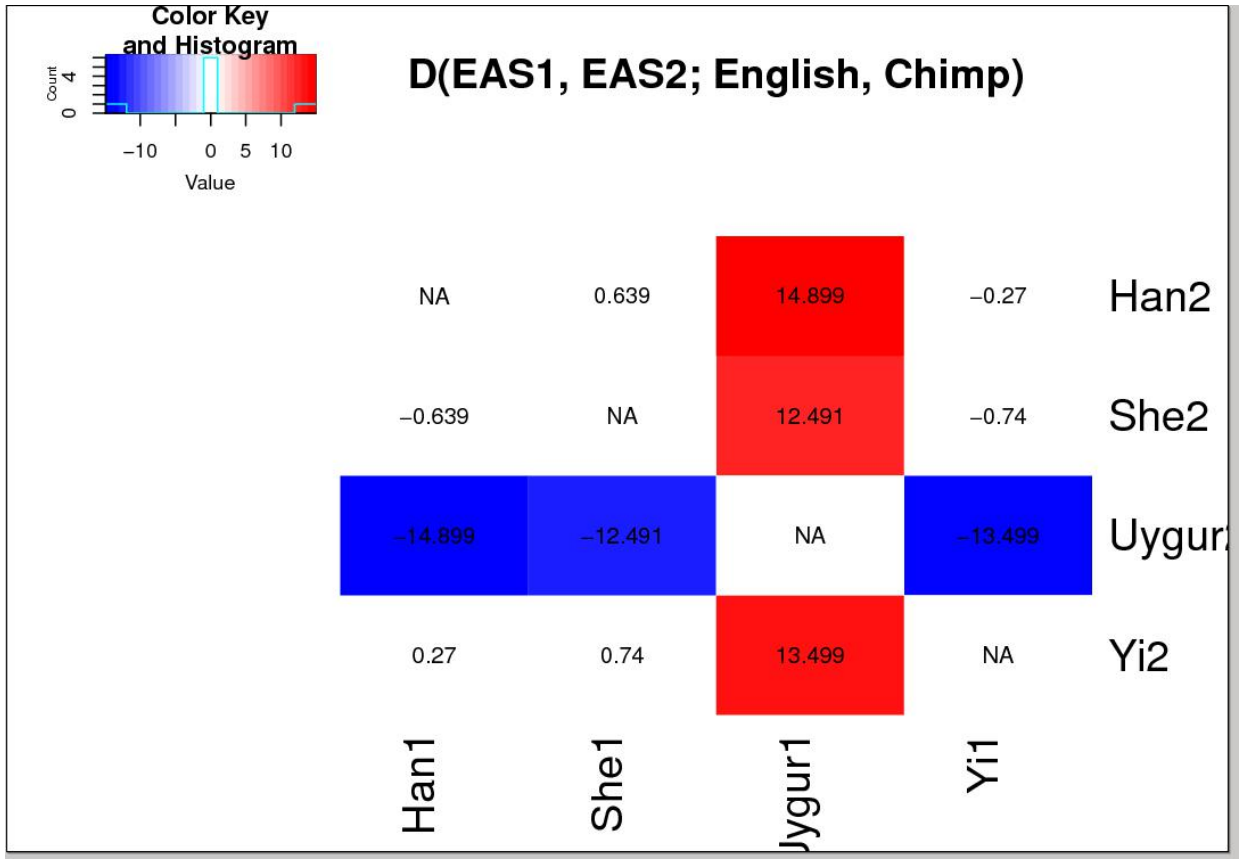
In [1]:
```
%%writefile /public/adna/student/2018class/yang_mel/Lesson8_general.par

genotypename:    /public/adna/student/data/data.eigen.geno
snpname:         /public/adna/student/data/data.eigen.snp
indivname:       /public/adna/student/data/data.eigen.ind
popfilename:     D1.ppf
printsd:         YES
```

```
Writing /public/adna/student/2018class/yang_mel/Lesson8_general.par
```

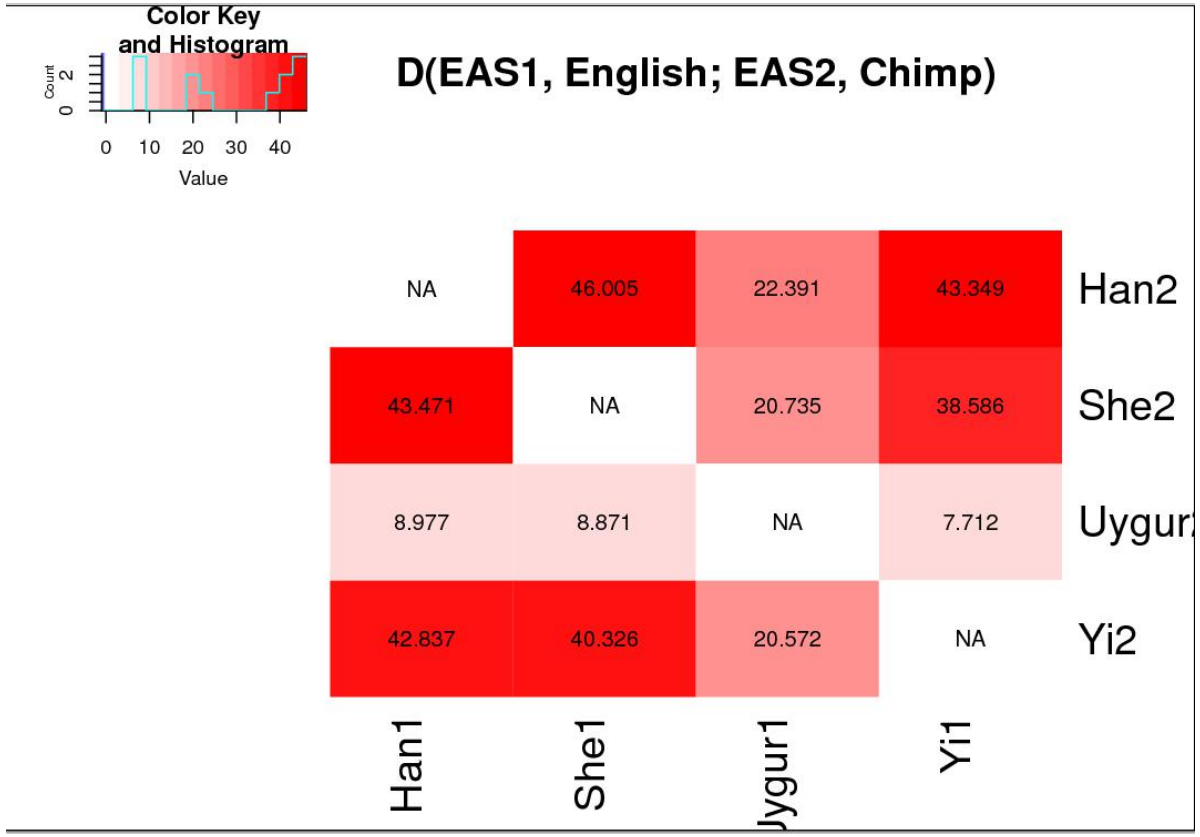
```
#### D1
result:      Han    English       She      Chimp      0.1642     43.471 120216  86298 1988272
result:      Han    English     Uygur      Chimp      0.0340      8.977 105980  99003 1988320
result:      Han    English        Yi      Chimp      0.1567     42.837 119178  86886 1988248
result:      She    English     Uygur      Chimp      0.0369      8.871 106072  98524 1988142
result:      She    English        Yi      Chimp      0.1589     40.326 119165  86483 1988082
result:    Uygur    English        Yi      Chimp      0.0788     20.572 105585  90168 1988121

#### D2
result:      She    English       Han      Chimp      0.1660     46.005 120216  85989 1988272
result:    Uygur    English       Han      Chimp      0.0808     22.391 105980  90127 1988320
result:       Yi    English       Han      Chimp      0.1560     43.349 119178  87018 1988248
result:    Uygur    English       She      Chimp      0.0822     20.735 106072  89960 1988142
result:       Yi    English       She      Chimp      0.1564     38.586 119165  86926 1988082
result:       Yi    English     Uygur      Chimp      0.0313      7.712 105585  99172 1988121

#### D3
result:      Han        She   English      Chimp     -0.0018     -0.639  85989  86298 1988272
result:      Han      Uygur   English      Chimp     -0.0469    -14.899  90127  99003 1988320
result:      Han         Yi   English      Chimp      0.0008      0.270  87018  86886 1988248
result:      She      Uygur   English      Chimp     -0.0454    -12.491  89960  98524 1988142
result:      She         Yi   English      Chimp      0.0026      0.740  86926  86483 1988082
result:    Uygur         Yi   English      Chimp      0.0476     13.499  99172  90168 1988121
```



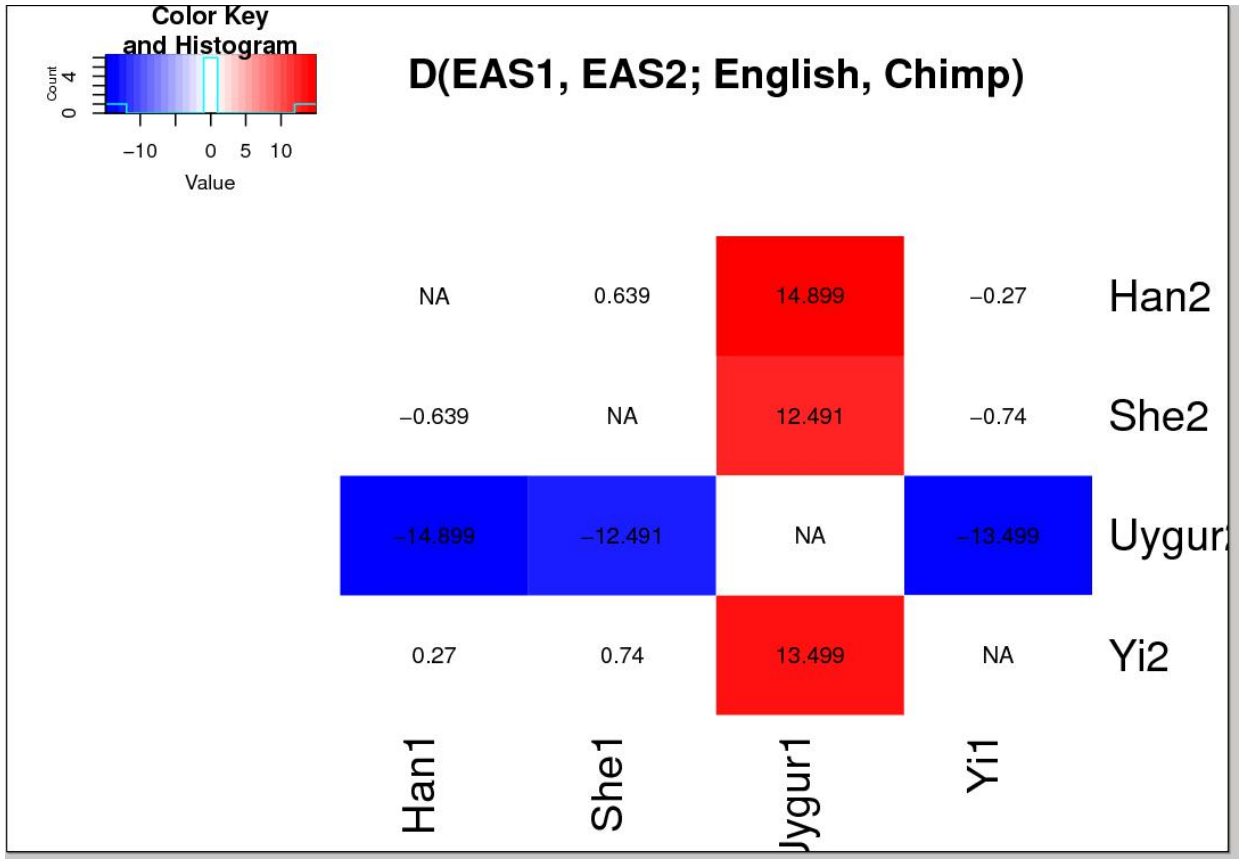
D1 and D2 in one table



D3



D1 and D2 in one table



D3

conclusions: 1. Han, She and Yi form a clade relative to English.
2. Uygur is admixed form of English and east Asians (!?)

Exercise 1: Try to repeat these tests using Mbuti as outgroup.

Example 2: The story of native Americans

It was believed that Native Americans are sister group to East Asians so we expect EUR to have a symmetric connection with both of them; but

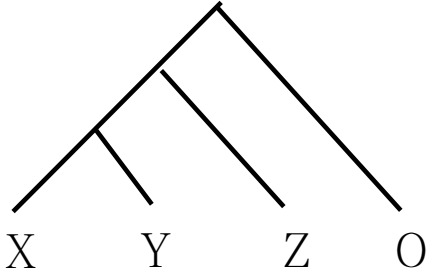
		pop1	pop2	pop3	pop4	D	Z	BABA	ABBA	#SNP
2	#result:	English	Han	Karitiana	Mbuti	-0.0919	-23.515	95713	115074	2054811
3	#result:	French	Han	Karitiana	Mbuti	-0.0858	-23.775	96119	114167	2054961
4	#result:	English	Han	Surui	Mbuti	-0.0924	-22.116	95611	115081	2054503
5	#result:	French	Han	Surui	Mbuti	-0.0909	-23.517	95474	114569	2054635
6										
7	#result:	English	Karitiana	Han	Mbuti	-0.1131	-30.680	91683	115074	2054811
8	#result:	English	Surui	Han	Mbuti	-0.1166	-29.245	91051	115081	2054503
9	#result:	French	Karitiana	Han	Mbuti	-0.1100	-31.438	91532	114167	2054961
10	#result:	French	Surui	Han	Mbuti	-0.1131	-30.114	91291	114569	2054635
11										
12	#result:	Han	Karitiana	English	Mbuti	-0.0215	-6.843	91683	95713	2054811
13	#result:	Han	Surui	English	Mbuti	-0.0244	-6.563	91051	95611	2054503
14	#result:	Han	Karitiana	French	Mbuti	-0.0244	-7.872	91532	96119	2054961
15	#result:	Han	Surui	French	Mbuti	-0.0224	-6.240	91291	95474	2054635

AME = {Karitiana, Surui},
nEUR = {English, French},
EAS = {Han}

Quiz: Try to find three D values for the set {AME, nEUR, Han} using Mbuti as outgroup.

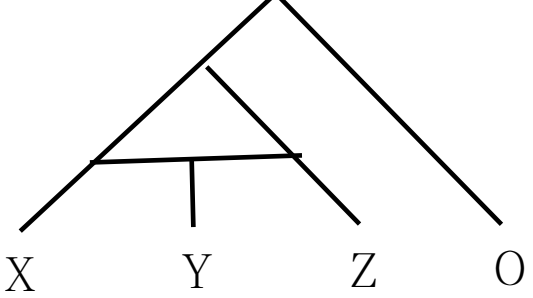
D(EAS, nEUR ; AME, Mbuti)
D(AME, nEUR ; EAS, Mbuti)
D(EAS, AME ; nEUR, Mbuti)

our models:



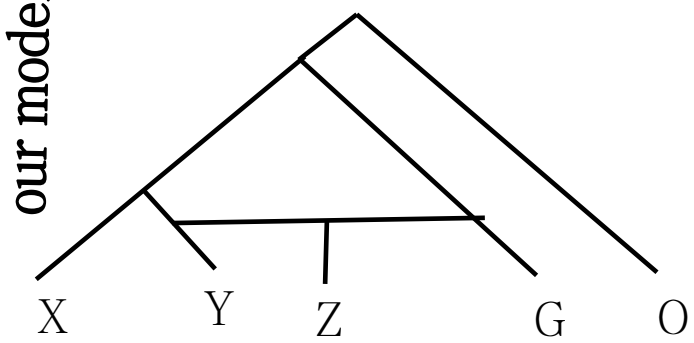
D1 (X, Z; Y, O) > 0
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) ~ 0

1



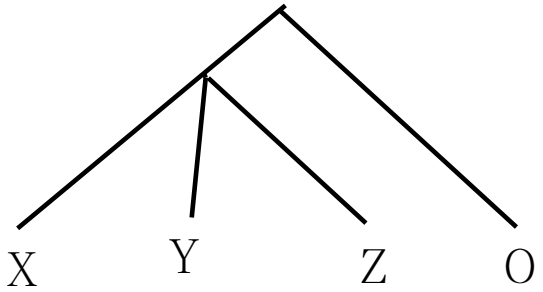
D1 (X, Z; Y, O) = any value
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) < 0

2



D1 (X, Z; Y, O) = any value
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) < 0

3



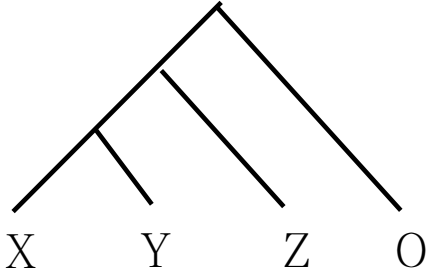
D1 (X, Z; Y, O) ~ 0
D2 (Y, Z; X, O) ~ 0
D3 (X, Y; Z, O) ~ 0

4

our observations:

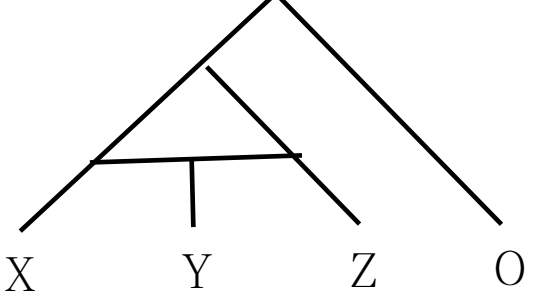
D(EAS, nEUR ; AME, Mbuti) >0, Z ~ 23
D(AME, nEUR ; EAS, Mbuti) >0, Z ~ 30
D(EAS, AME ; nEUR, Mbuti) <0, Z ~ 6

Which model can explain the observations?



D1 (X, Z; Y, O) > 0
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) ~ 0

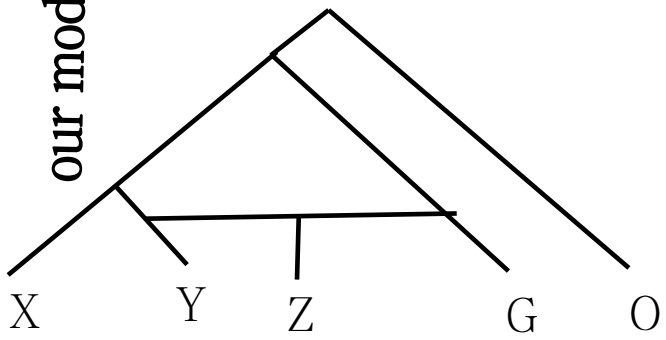
1



D1 (X, Z; Y, O) = any value
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) < 0

2

our models:



D1 (X, Z; Y, O) = any value
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) < 0

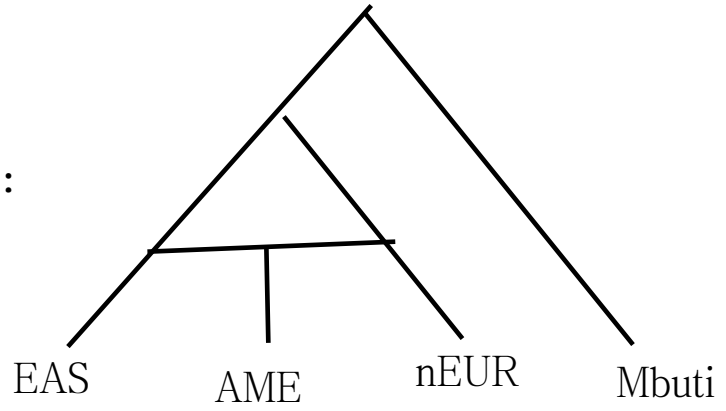
3

our observations:

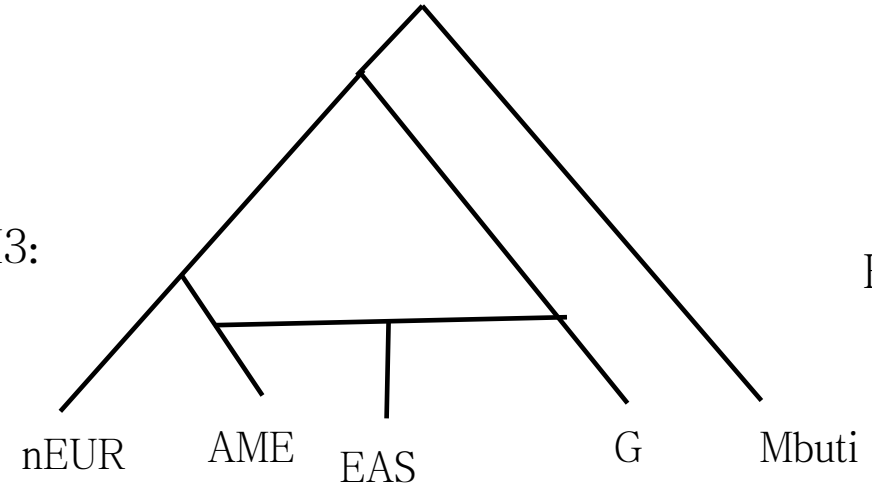
D(EAS, nEUR ; AME, Mbuti) >0, Z ~ 23
D(AME, nEUR ; EAS, Mbuti) >0, Z ~ 30
D(EAS, AME ; nEUR, Mbuti) <0, Z ~ 6

Which model is the true one?

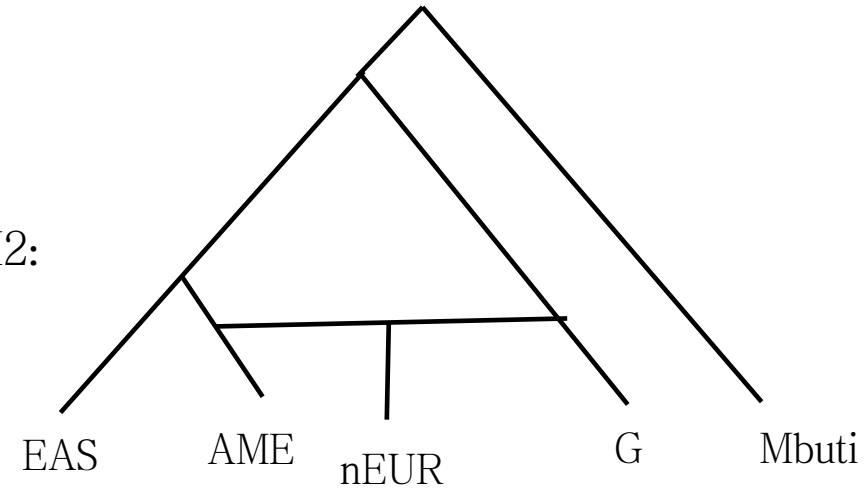
H1:



H3:



H2:



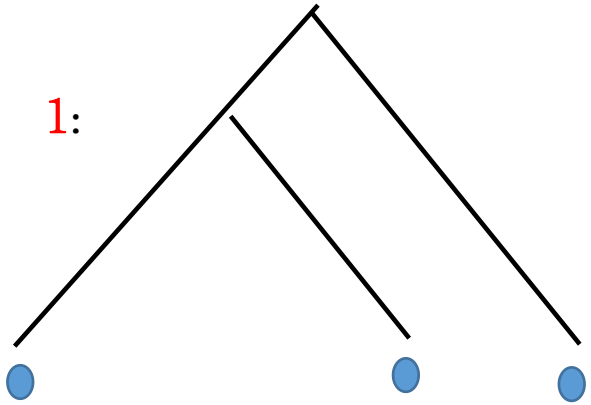
Need a direct measure of admixture; 3 population test

f3 test tells you if target population K is a mixed form of source populations H and G (or their related populations).

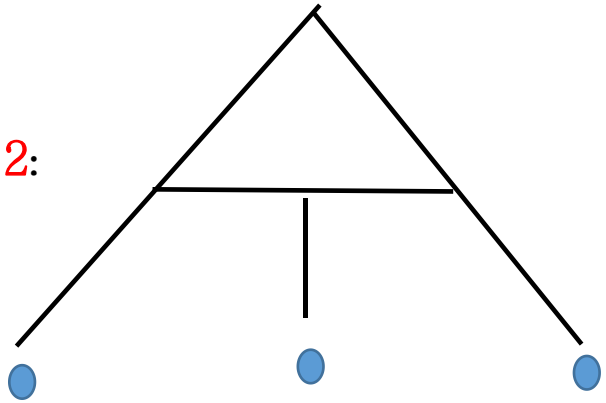
Quiz: What are all possible arrangements of three population?

All possible arrangements of three population:

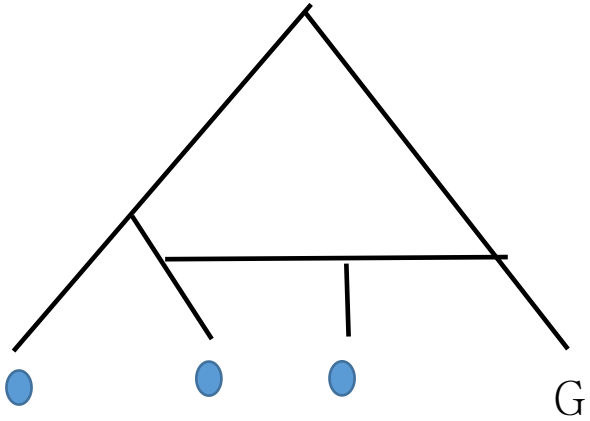
1:



2:

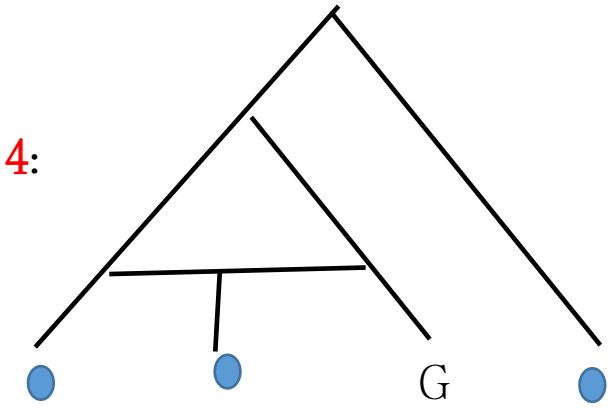


3:



Exercise 2: Convince yourself that all scenarios, except one case of model 2, result in positive value for f3.

4:



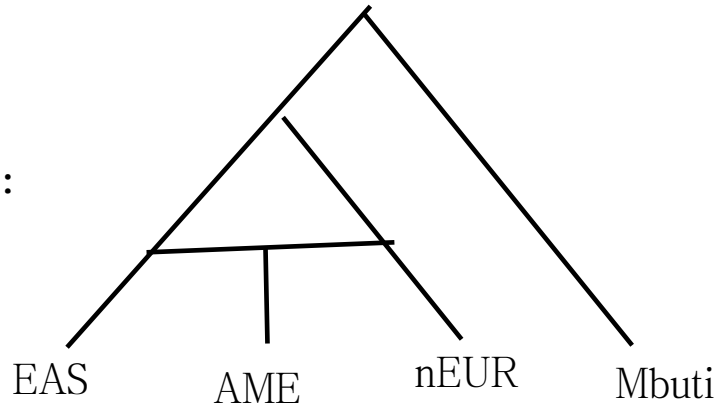
If f3(A, B; C) < 0 then C is a mixed form of ancestries related to A and B.

Quiz: For a 3 population test f3(H, G; K), which one is true?

1. f3 statistic is the overlapping the paths H to K and G to K.
2. f3 = E[(k−h)(k−g)]
3. f3 = cov[(k−h), (k−g)]

Quiz: Using data file "Lesson8_2.log", pick up relevant f3 tests to find an evidence in favor of /against hypothesis H1.

H1:



AME = {Karitiana, Surui},
nEUR = {English, French},
EAS = {Han}

All files relevant to lesson 8 are located in:
/public/adna/student/2018class/shafiey_hassan

Set G = {Sardinian, Greek} in model H1 and check with D and f3 tests.

Quiz: Use file "Lesson8_1.log" and "Lesson8_2.log" to find the following test values.

D(G, AME; nEUR, Mbuti)

D(nEUR, AME; G, Mbuti)

D(G, nEUR; AME, Mbuti)

f3(G, AME; nEUR)

Set G = {Sardinian, Greek} in model H1 and check with D and f3 tests.

Quiz: Use file "Lesson8_1.log" and "Lesson8_2.log" to find the following test values.

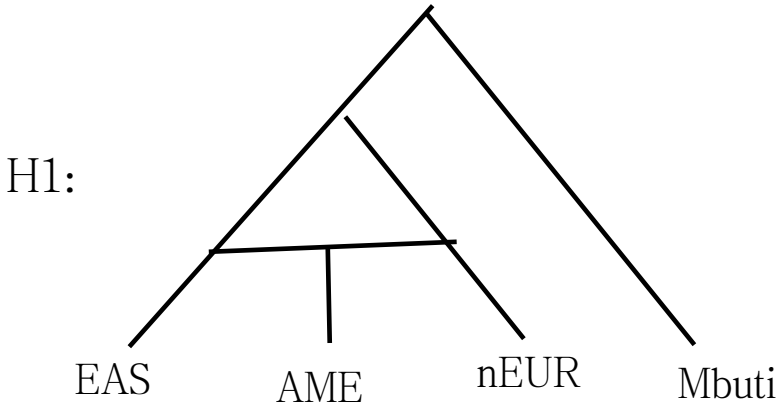
D(G, AME; nEUR, Mbuti) > 0

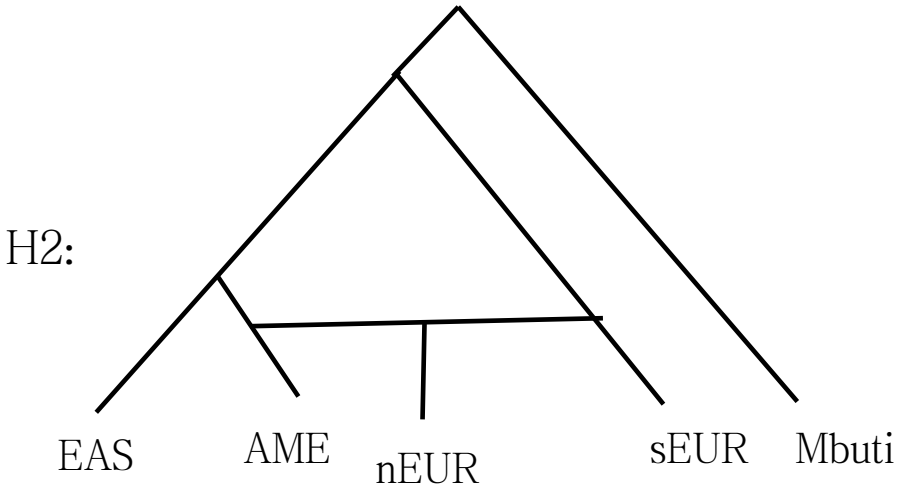
D(nEUR, AME; G, Mbuti) > 0

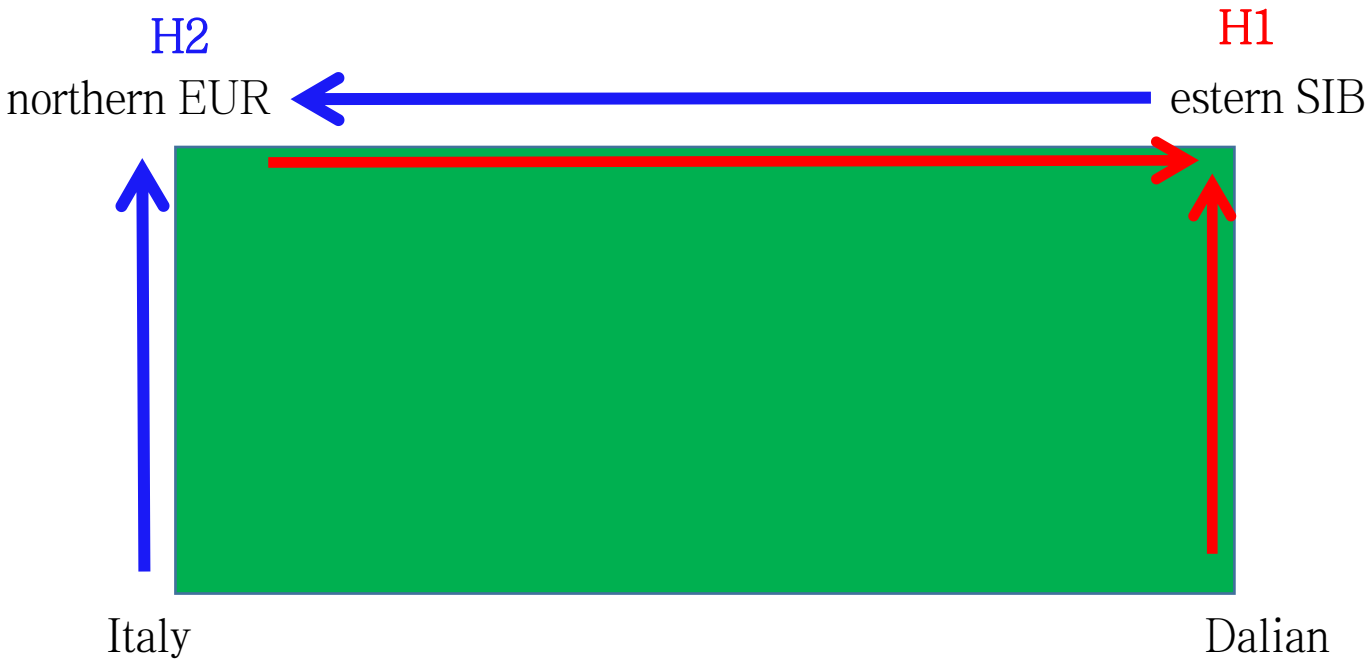
D(G, nEUR; AME, Mbuti) < 0

f3(Sardinian, AME; nEUR) < 0

conclusion: both models are true. How is it possible?

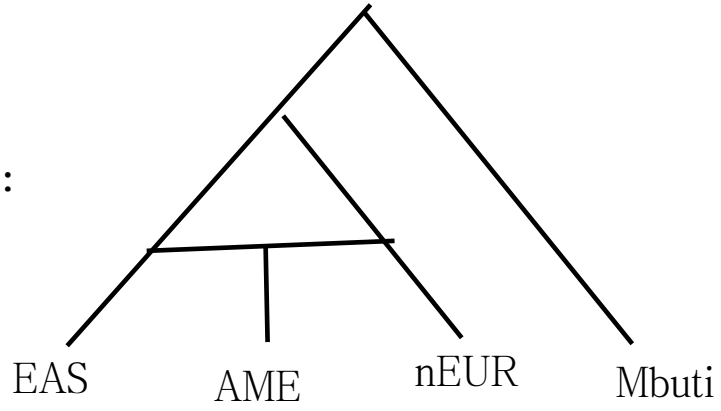




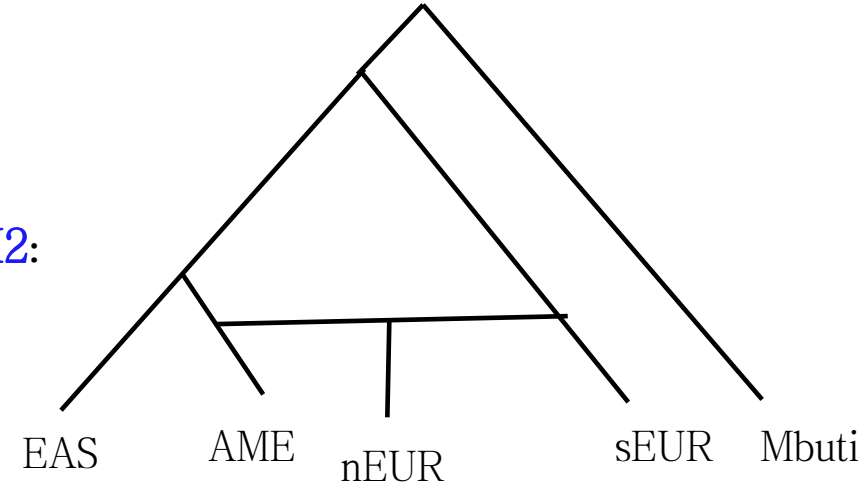


Quiz: Is there a simpler scenario?

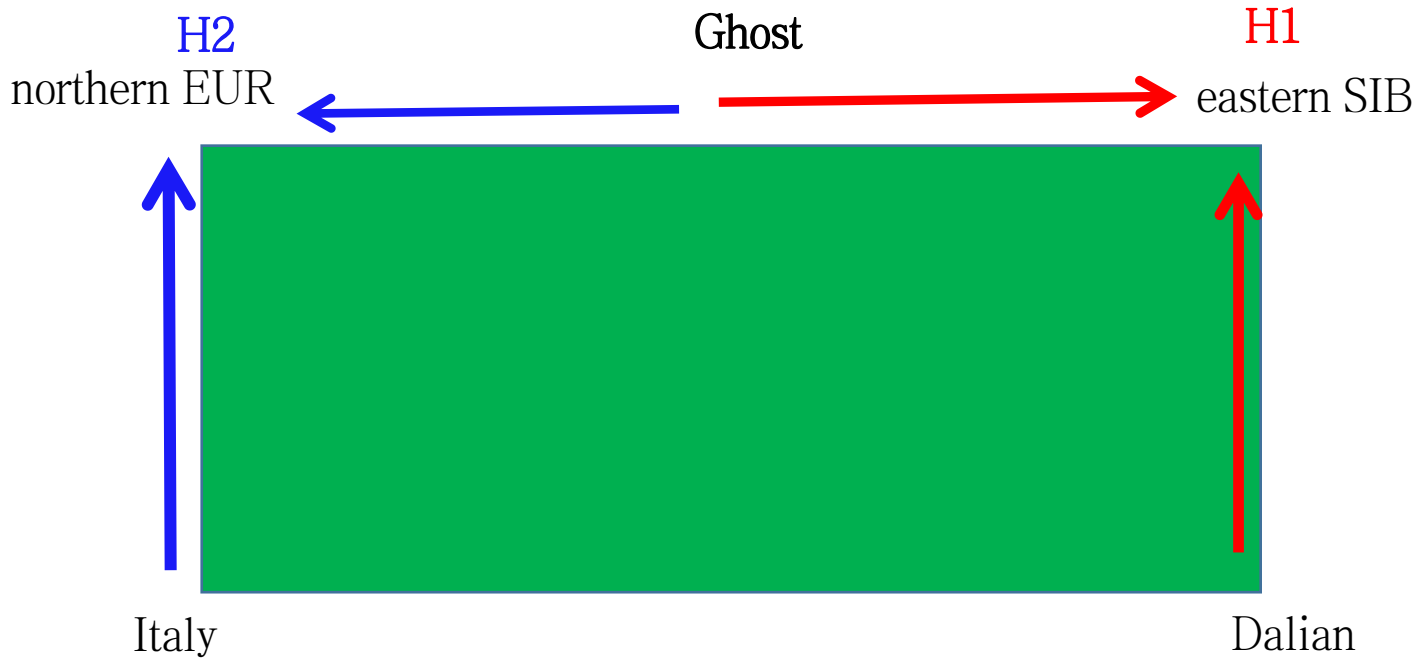
H1:


EAS
AME
nEUR
Mbuti

H2:


EAS
AME
nEUR
sEUR
Mbuti

H2
northern EUR
H1
eastern SIB
Italy
Dalian


H2
northern EUR
Ghost
H1
eastern SIB
Italy
Dalian

Finding ghost

We are looking for a population which

1. lived in central Siberia at least 15 000 years ago.
2. contibuted to the ancestor of native Americans.
3. contributed to the ancestor of northern Europeans.

Exercise 3: One the following populations is the ghost. Run whatever tests you need to figure out which one is the ghost we are looking for.

Malta1, Kostenki14, Tianyuan, UstIshim

Exercise 4: We saw that while f3 test shows admixture in nEUR, it doesn't show admixture in AME? Chat with each other and with Melinda, Hongru, Albert, or myself to figure out the reason.

Example 3: The story of Europeans

Classical archaeological scenario: Modern humans migrated to Eurasia about 60 000 years ago and then splitted into EAS and EUR. So we expect to have a cladal relationship between pdEUR and WHG(Western hunter gatherers).

present−day Europeans

Time (ky BC)	Group	West	Central	East
0.9	Iron Age		Hungary IA (1)	
1.1	Late Bronze Age		Halberstadt (1)	
2.2–1.6	Early Bronze Age		Unetice (8) Hungary BA (2)	
2.5–2.2	Late Neolithic		Alberstedt (1) BenzigerodeHeimburg (3) Bell Beaker (6) Karsdorf (1) Corded Ware (4)	
3.3–2.7	Late Copper Age (steppe)		Hungary CA (1) Sweden NHG (3)	Yamnaya (9)
4–3	Mid Neolithic	La Mina (4)	Iceman (1) Esperstedt (1) Baalberge (3) Sweden MN (1)	
6–5.5	Early Neolithic	Els Trocs (5)	Stuttgart (1) LBK (12) LBKT (1) Hungary EN (8) Starcevo (1)	
6–4.6	Holocene hunter-gatherer	La Brana1 (1)	Loschbour (1) Hungary HG (1) Sweden MHG (1) Motala (7)	Karelia (1) Samara (1)
43–22	Pleistocene hunter-gatherer			MA1 (1) Kostenki14 (1) Ust Ishim (1)

Root

Han

Quiz: What would you expect to get for the following tests if this hypothesis is true?

D(EUR, EAS; WHG, Mbuti)

D(WHG, EAS; EUR, Mbuti)

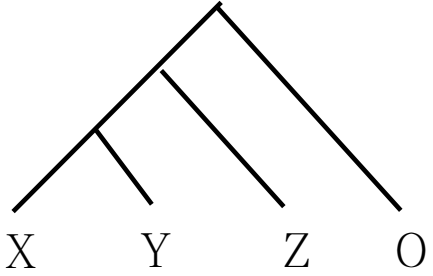
D(EUR, WHG; EAS, Mbuti)

Use the stock of D statistics in /public/adna/student/results/ to find them.

where EUR = {Englsih, French}, EAS = {Han}, WHG = {Loschbour, LaBrana}

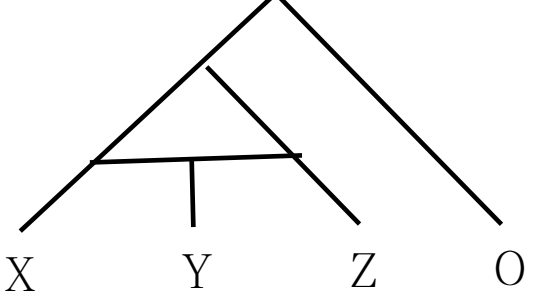
```
20 nrows, ncols: 15 2056289
21 result:    English         Han  Loschbour      Mbuti      0.0959     20.404 114504  94463 2036972
22 result:     French         Han  Loschbour      Mbuti      0.0936     22.839 113745  94274 2037082
23 result:    English         Han   LaBrana1      Mbuti      0.0792     18.571 100766  85968 1824208
24 result:     French         Han   LaBrana1      Mbuti      0.0827     21.389 100679  85291 1824307
25 result:  Loschbour         Han    English      Mbuti      0.1105     24.364 114504  91720 2036972
26 result:  Loschbour         Han     French      Mbuti      0.1042     25.500 113745  92280 2037082
27 result:   LaBrana1         Han    English      Mbuti      0.0895     20.525 100766  84216 1824208
28 result:   LaBrana1         Han     French      Mbuti      0.0890     23.270 100679  84216 1824307
29 result:    English  Loschbour          Han      Mbuti     -0.0147     -3.380  91720  94463 2036972
30 result:     French  Loschbour          Han      Mbuti     -0.0107     -2.649  92280  94274 2037082
31 result:    English   LaBrana1          Han      Mbuti     -0.0103     -2.745  84216  85968 1824208
32 result:     French   LaBrana1          Han      Mbuti     -0.0063     -1.760  84216  85291 1824307
33 ## end of run
```

D(EUR, EAS ; WHG, Mbuti) >0, Z ~20
D(WHG, EAS; EUR, Mbuti) >0, Z ~ 20
D(EUR, WHG ; EAS, Mbuti) <0, Z ~ 3



D1 (X, Z; Y, O) > 0
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) ~ 0

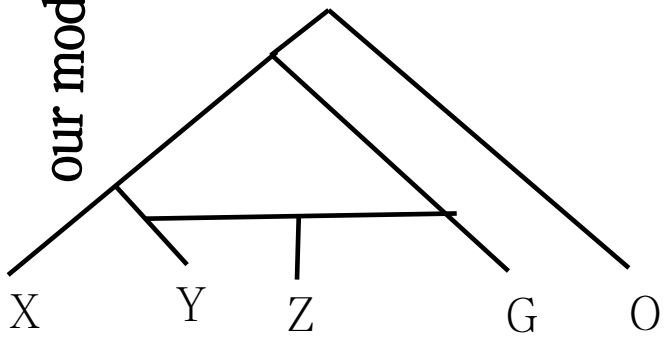
1



D1 (X, Z; Y, O) = any value
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) < 0

2

our models:



D1 (X, Z; Y, O) = any value
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) < 0

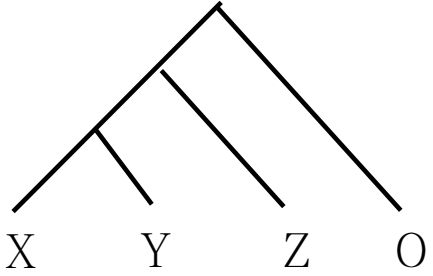
3

our observations:

D(EUR, EAS ; WHG, Mbuti) >0, Z ~20
D(WHG, EAS; EUR, Mbuti) >0, Z ~ 20
D(EUR, WHG ; EAS, Mbuti) <0, Z ~ 3

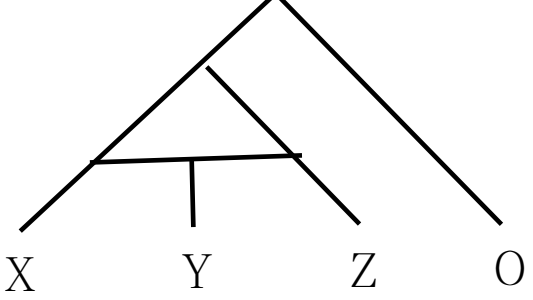
The data doesn't support the cladal scenario.

Quiz: What are the possible scenarios that can explain the observation?



D1 (X, Z; Y, O) > 0
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) ~ 0

1



D1 (X, Z; Y, O) = any value
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) < 0

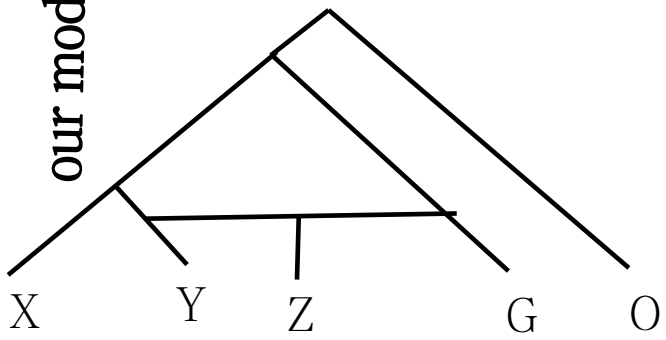
2

our observations:

D(EUR, EAS ; WHG, Mbuti) >0, Z ~20
D(WHG, EAS; EUR, Mbuti) >0, Z ~ 20
D(EUR, WHG ; EAS, Mbuti) <0, Z ~ 3

Which model is the true one?

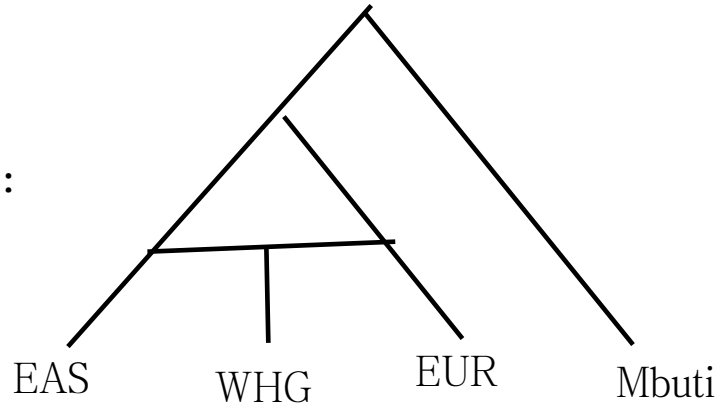
our models:



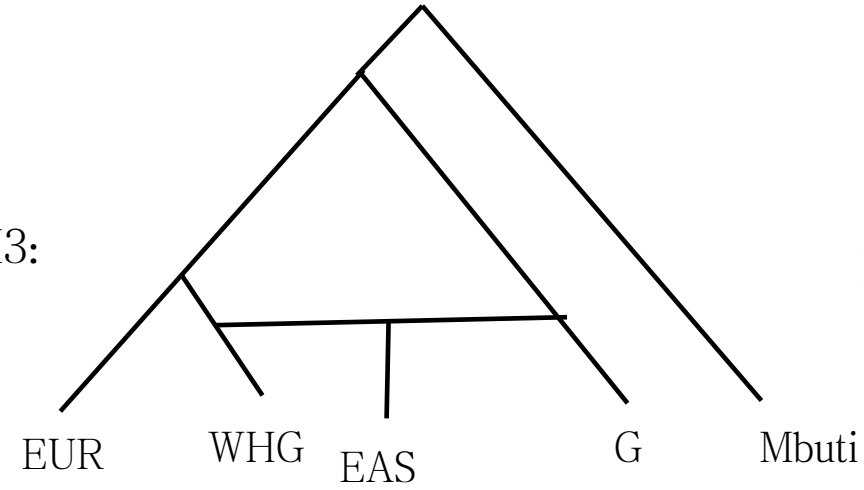
D1 (X, Z; Y, O) = any value
D2 (Y, Z; X, O) > 0
D3 (X, Y; Z, O) < 0

3

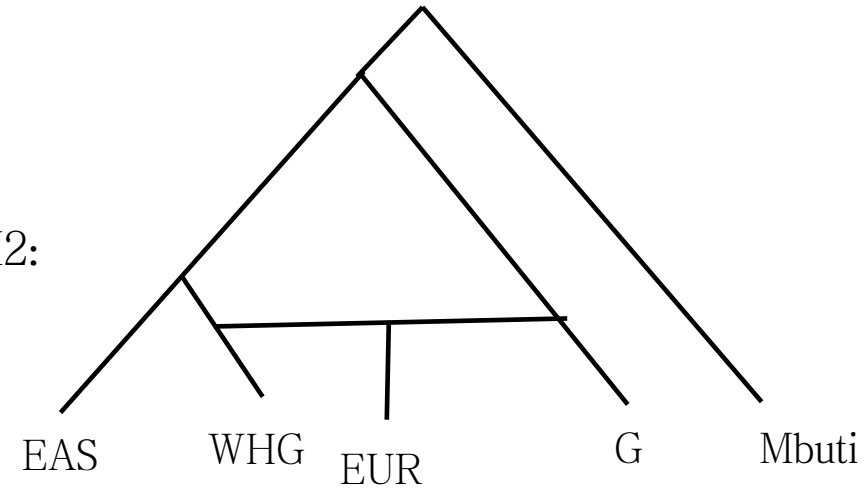
H1:



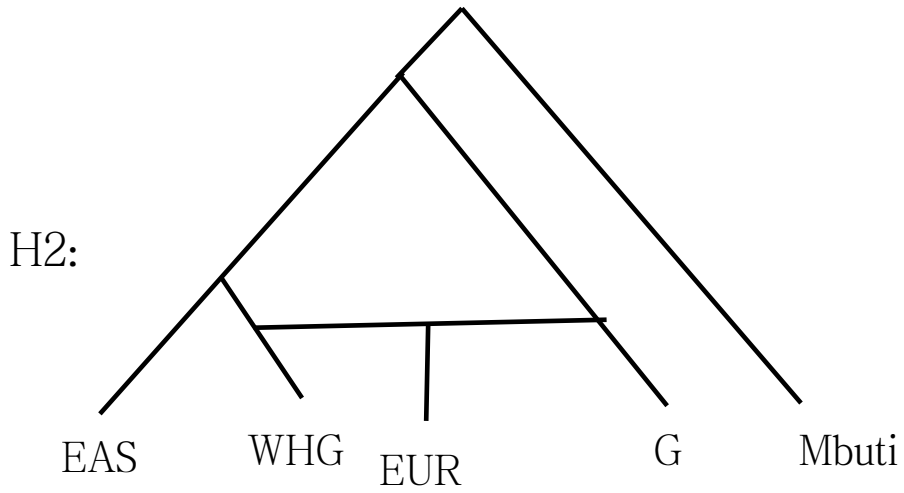
H3:



H2:



Based on previous knowledge we can reject H1 and H3 so we left with H2:



Quiz: Use table S11.1 from page 85 of Lazaridis2014 paper supplementray information file to find most plausible sources for present day Europeans.

Quiz: Use table S11.1 from page 85 of Lazaridis2014 paper supplementray information file to find most plausible sources for present day Europeans.

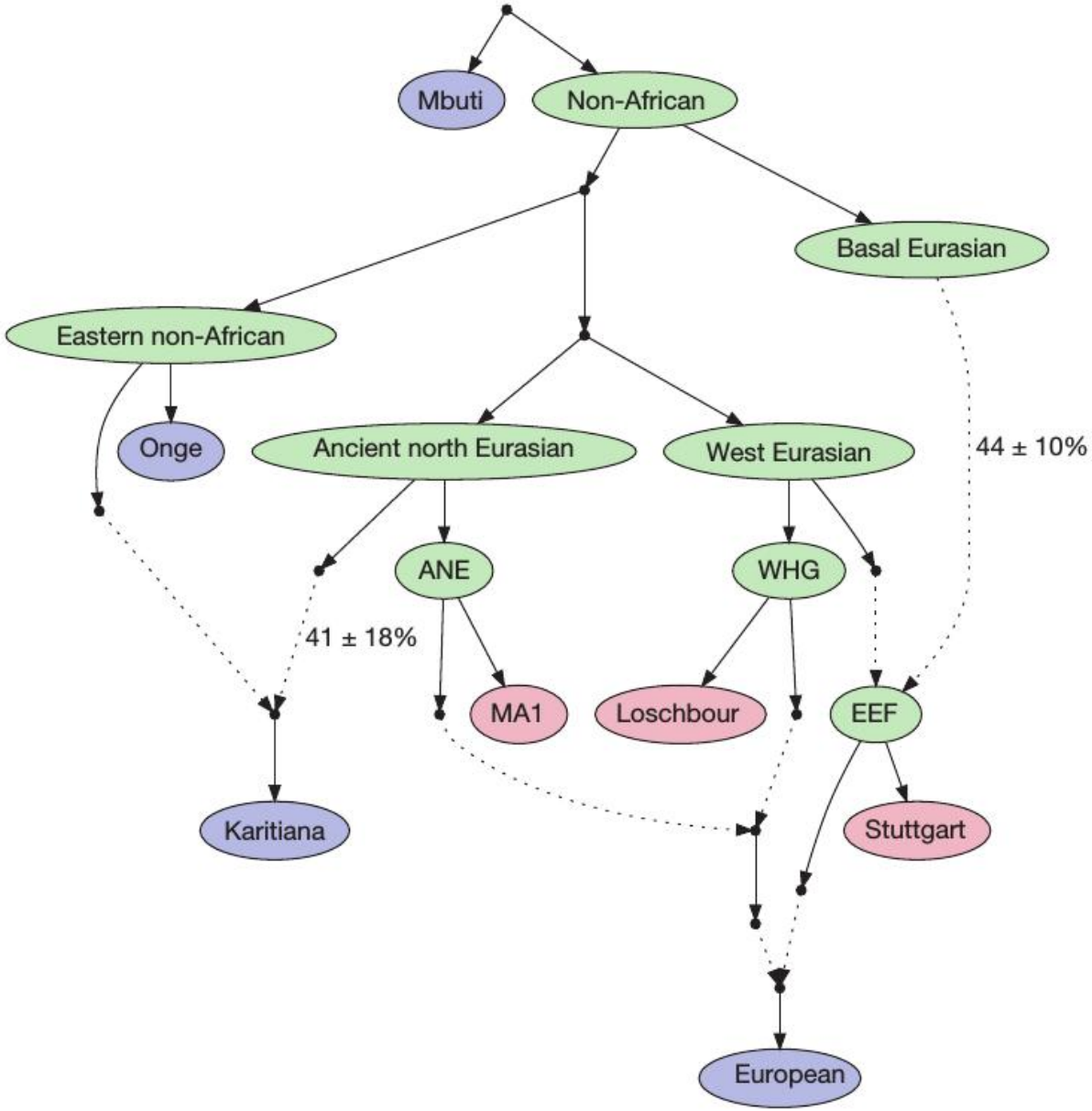
1. present–day Europeans can be best modeled as mixture of either {Stuttgart and ANE} or {WHG and Near Easterns}.

2. [From archaeology] Migrants from Anatolia nd Levant introduced agriculture to Europe.

Stuttgart is an admixed population of near Easterns and WHG.

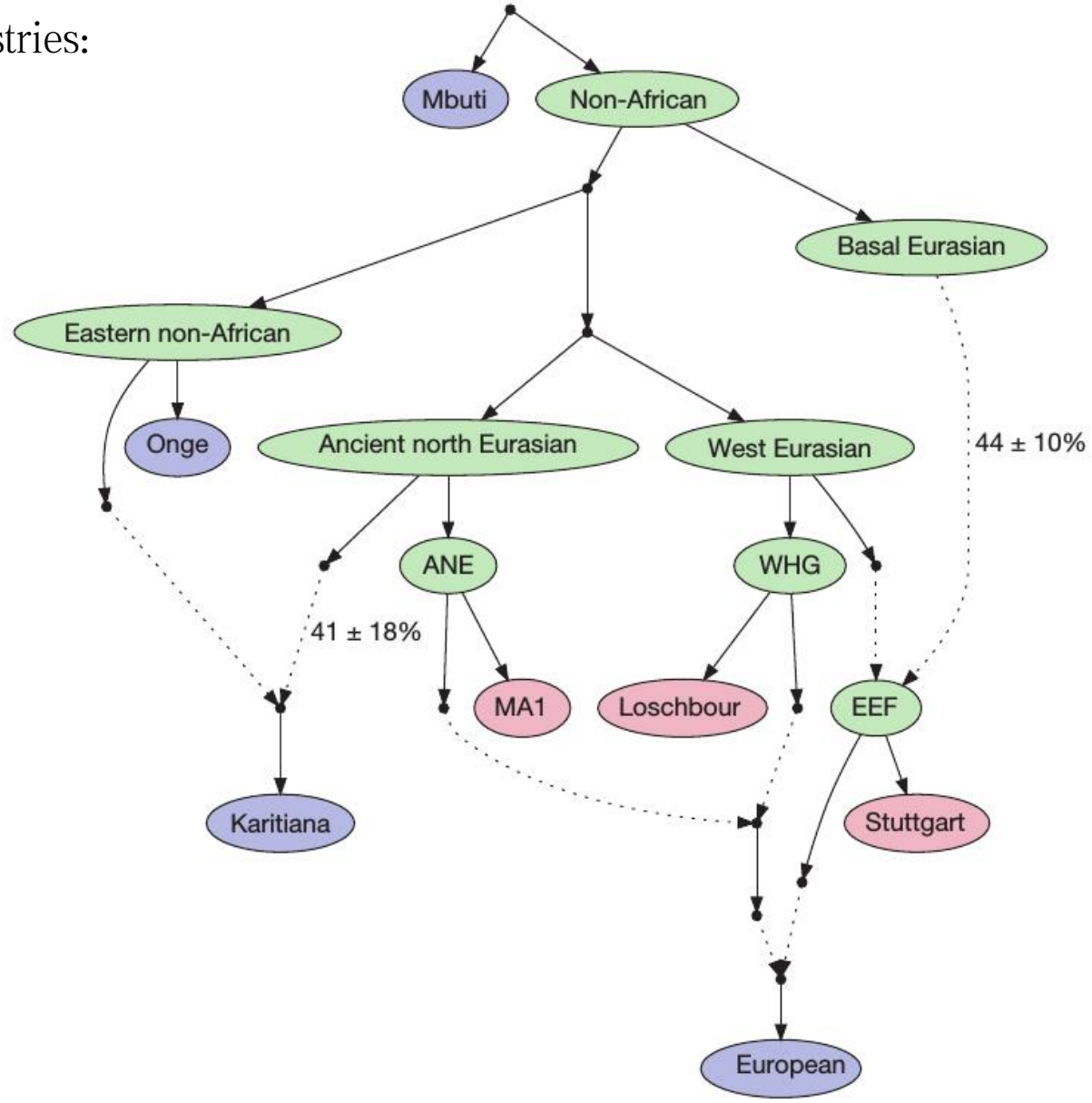
Lazaridis2014: Farmers from near East introduced Basal Eurasian ancestry to Europe and Stuttgart is one representative that carries such ancestry.

Quiz: Present–day Europeans carry three ancestries:
1. Basal Eurasian ancestry.
2. ??
3. ??



Quiz: [Lazaridis2014]:Present–day Europeans carry three ancestries:
1. Basal Eurasian ancestry.
2. ANE
3. WHG

How did ANE come to Europe?

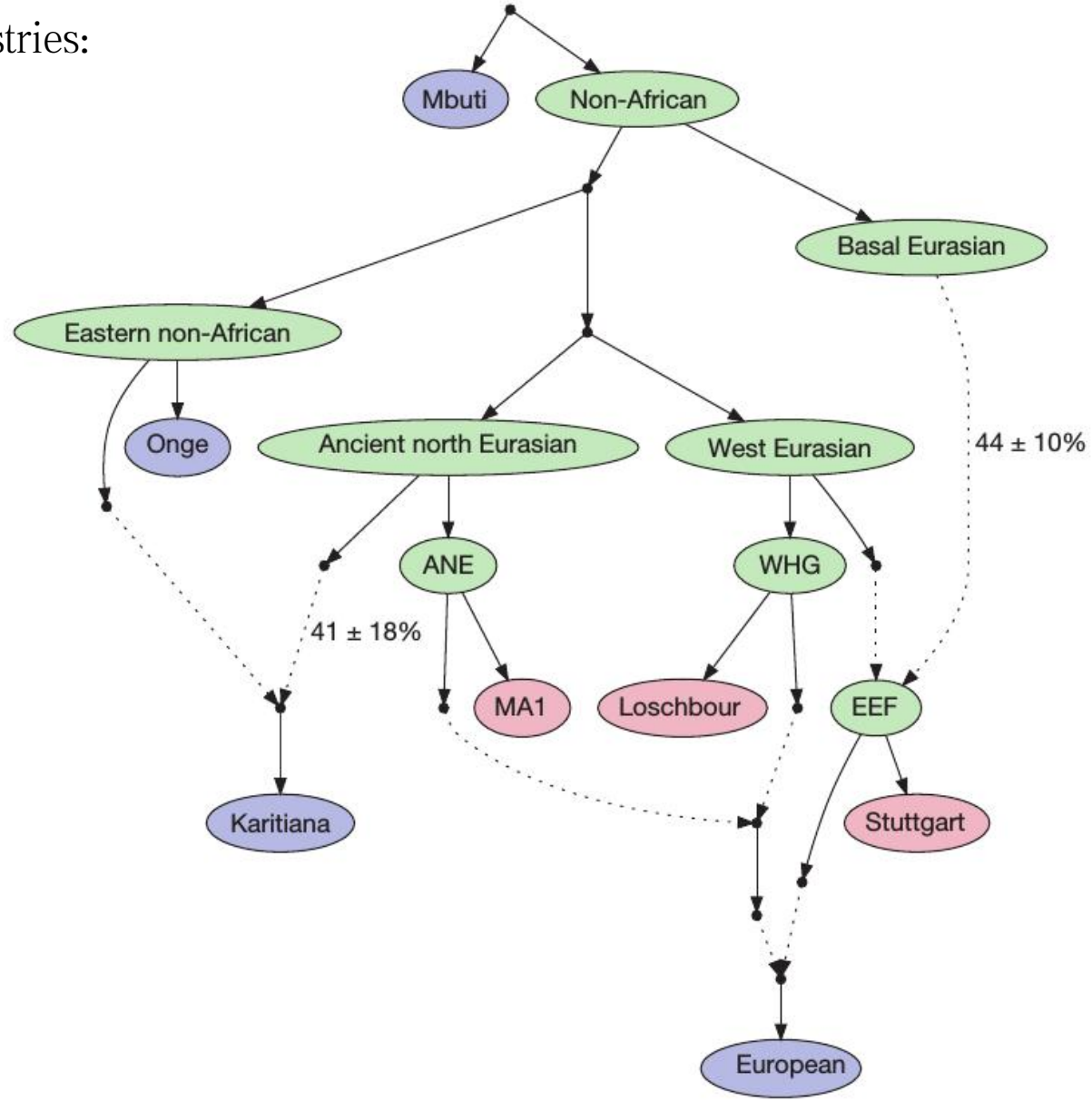


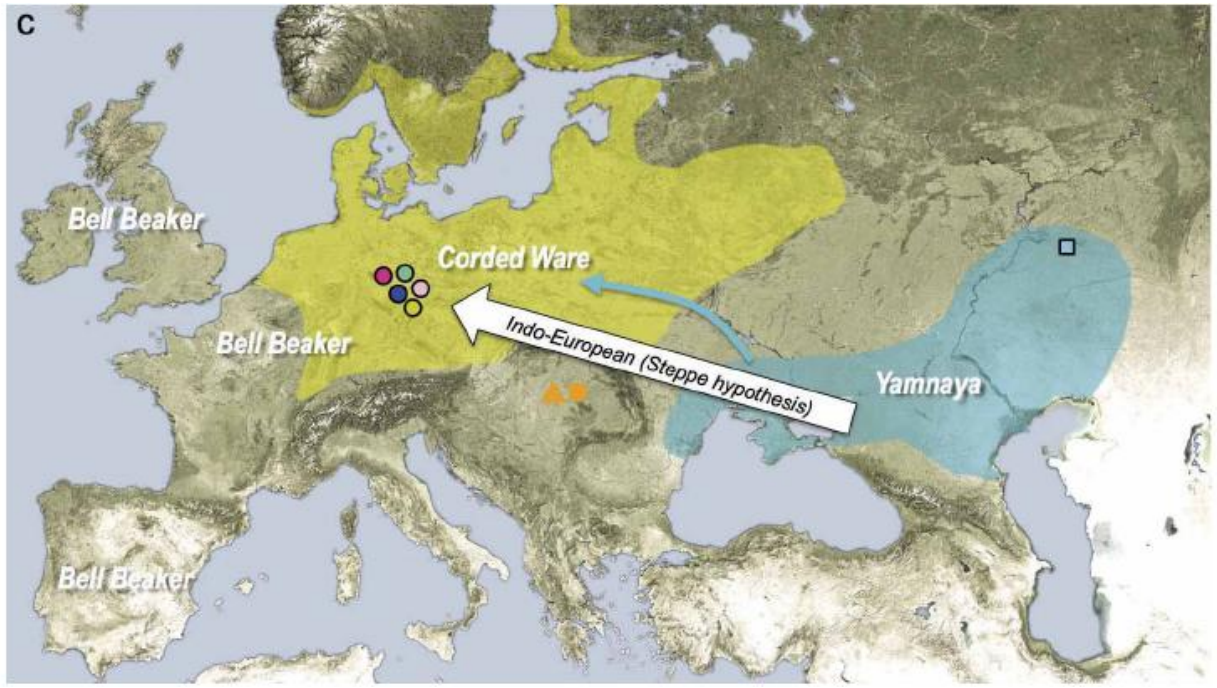
Quiz: [Lazaridis2014]:Present-day Europeans carry three ancestries:
1. Basal Eurasian ancestry.
2. ANE
3. WHG

How did ANE come to Europe [Haak2015]?

ANE → Yamnaya → Corded Ware → Europeans

What is the other message of the Haak2015 paper?





Massive migration from Steppe to Europe can explain the origin of Indo–European languages in Europe. Leo S. Klejn critiques this idea.

Exercise 5: Read the pdf file of his discussion (available on the server) with Haak et.al. and summarize their arguments. Who is right? who argues more scientifically?

Exercise 6: Some of Native Americans have got european ancestry from two different ways. Do you know them? Can you explore these two admixture events using D statistics?

Exercise 7: There are two deep ancestries in Eurasia: East Asian and European. Life in the East of Asia is much more colourful than Europe. Do you think this difference has a genetic basis? Do a literature survey and feel free to contact people in Svante's, David's, and Eske's lab for investigating this.

See you on Wednesday at 9:00 for checking homework. I'll put the slides on the server after the class so you can have exercises.

Thanks for your patience; and those roses are for you all.