Ancient Admixture in Human History

Nick Patterson et al. (2012)

Woojin Jeong

HMDA LAB.

July 25, 2024

Table of Contents

Background

Background

Please refer to materials

Introduction

Abstract

- Admixture between populations is a fundamental process that shapes genetic variation and disease risk.
- Authors present a suite of methods for
  - learning about population mixtures that support formal tests whether mixture occured.
  - inferring proportions and dates of mixture.
- Development of New SNP arrays that was specifically designed for population genetics analysis.

Literature Review

- Local ancestry-based methods
  - Deconvolve ancestry at each locus in the genome and provide individual-level information about ancestry
  - Detect recent admixture events well, but have relatively bad performance for detecting older events.
  - LAMP(2008), HAPMIX(2009), PCADMIX(2010)
- Global ancestry-based methods
  - Powerful tools for detecting population substructure
  - Do not provide any formal tests for admixture
  - PCA(2006), STRUCTURE(2000), ADMIXTURE(2009)
- In this article, by fitting phylogenetic tree-based models,
  - Authors describe a suite of methods that formally test for a history of population mixture.
  - Allow researchers to build models of population mixture that fit genetic data

Methods

- In this Article, authors describe 5 methods...
  - Three-population test
  - D-statistics
  - F_4-ratio estimation
  - Admixture graph fitting
  - Rolloff
- All these methods are implemented in software package, ADMIXTOOLS

Methods

Brief description

The first 4 methods (Three-population test, D-statistics, F_4-ratio estimation, Admixture graph fitting) are based on studying patterns of allele frequency.

- Three-population test is a formal test of admixture even if the events occured hundreds of generations ago.
- D-statistics(Four-population test) is also a formal test for admixture, and also provide informations about the directionality of gene flow
- F_4-ratio estimation allows inference of the mixing proportions of an admixture event
- Admixture graph fitting allows one to build a model of population relationships for an arbitrarily large number of populations simultaneously and assess this model fits allele frequency correlation patterns among populations.

Brief description

The first 4 methods are all based on studying patterns of allele frequency correlations using f-statistics and D-statistics that that we define in what follows.

The expected values of these statistics are functions not just of the demographic history relating the populations, but also of the way that the analyzed polymorphisms were discovered (the so-called ascertainment process)

But the tests of our interests for a history of admixture is based on particular statistics which is has expectation of zero in the absence of admixture, which is robust to almost all ascertainment process. Authors showed this robustness by both simulation and application to real data.

Brief description

The fifth method that we introduce in this study, rolloff, is an approach for estimating the date of admixture which models the decay of admixture linkage disequilibrium in the target population.

Previous method(Pool and Nielsen, 2009) assumes local ancestry inference is perpect, which is not realistic.

Rolloff does not require accurate reconstruction of the breakpoints across the chromosomes or data from good surrogates for the ancestors, making it possible to interrogate older dates.
Simulations that we report in what follows show that rolloff can produce unbiased and quite accurate estimates for dates up to 500 generations in the past.

Basic Assumptions / Notations

- Basic Assumptions
  - We consider biallelic markers only.
  - Ignore recurrent or back mutations.
- Notations
  - $f_2, f_3, f_4..$: statistics
  - $F_2, F_3, F_4..$: parameters for assumed phylogeny
  - Drift : frequency change of an allele along a graph edge

Three population test

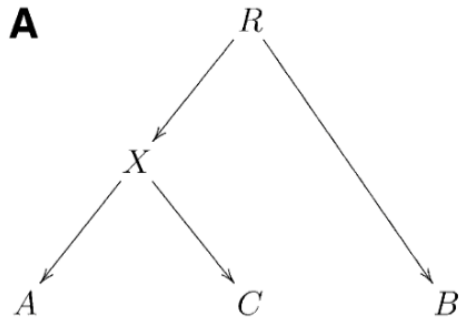


Figure: 1-A

Let a',b',c' be allele frequencies in the population A,B,C at a single polymorphism.

Define Fs,

$$F_2(A,B) = E[(a'-b')^2] : Branch\ Length$$

$$F_3(C;A,B) = E[(c'-a')(c'-b')]$$

$$F_4(A,B;C,D) = E[(a'-b')(c'-d')]$$

Here, Expectation is calculated over each Polymorphism.

Unbiased estimators of F

Theorem (1)

Let a, b, c be sample allele frequencies in the population A,B,C at a single polymorphism.
Here, n_A, n'_A *be the sample counts of variant and reference alleles.*
s_A *is the total number of alleles observed in population A. i.e,* $s_A = n_A + n'_A$.
$\hat{h_A} = \frac{n_A n'_A}{s_A(s_A-1)}$
Then, UE of F are as follow.

- $\hat{F}_2(A,B) = (a-b)^2 - \hat{h_A}/s_A - \hat{h_B}/s_B$
- $\hat{F}_3(C;A,B) = (c-a)(c-b) - \hat{h_C}/s_C$
- $\hat{F}_4(A,B;C,D) = (a-b)(c-d)$

Unbiased estimators of F-statistics

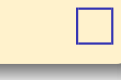
Proof of Theorem (1).

- $\hat{F}_2(A,B) = (a-b)^2 - \hat{h_A}/s_A - \hat{h_B}/s_B$ is UE of $(a'-b')^2$

It is trivial that $s_A * a = n_A \sim Bin(s_A, a')$.

Therefore, $E((a-b)^2 - \hat{h_A}/s_A - \hat{h_B}/s_B)$
$= E((\frac{n_A}{s_A} - \frac{n_B}{s_B})^2 - \frac{n_A(s_A - n_A)}{s_A^2(S_A - 1)} - \frac{n_B(s_B - n_B)}{s_B^2(S_B - 1)})$

Using $E(n_A) = s_A * a'$, $Var(n_A) = s_A * a' * (1-a')$ and a, b are independent,
We can show $\hat{F}_2(A,B)$ is UE of $(a'-b')^2$



f - statistics

We define our f-statistics using former unbiased estimators:

- $f_2(A,B) = \frac{1}{n}\sum_i[(a_i - b_i)^2 - \hat{h_{A_i}}/s_{A_i} - \hat{h_{B_i}}/s_{B_i}]$
- $f_3(C;A,B) = \frac{1}{n}\sum_i[(c_i - a_i)(c_i - b_i) - \hat{h_{C_i}}/s_{C_i}]$
- $f_4(A,B;C,D) = \frac{1}{n}\sum_i(a_i - b_i)(c_i - d_i)$

Unbiased estimators of previous slide were averaged over many markers(i) to form our f-statistics.

Understanding phylogenetic tree

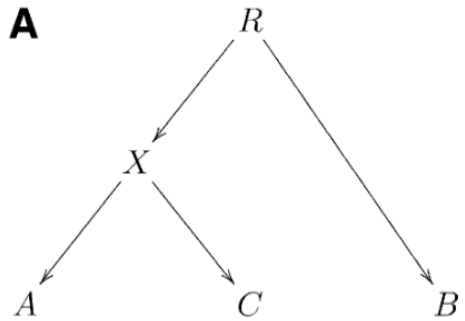


Figure: 1-A

Some assumptions are needed to understand the phylogenetic tree. (Ascertained in an outgroup)

- Each drift has mean of zero i.e. $E[a'|x'] = x', E[x' - a'] = 0$
- Each drift is independent. i.e. $E[(x' - a')(x' - c')] = E[(x' - a')]E[(x' - c')] = 0$

Intuitive way to understand F-statistics

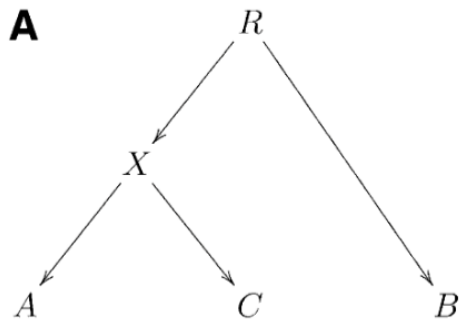


Figure: 1-A

Using the facts that each drift is independent, we can interpret F as follows:

- $F_2(A, C)$: Overlap between the genetic drift paths A -> C, A -> C
- $F_3(C; A, B)$ Overlap between the genetic drift paths C -> A, C -> B
- $F_4(A, B; C, D)$ Overlap between the genetic drift paths A -> B, C -> D

Three - population test

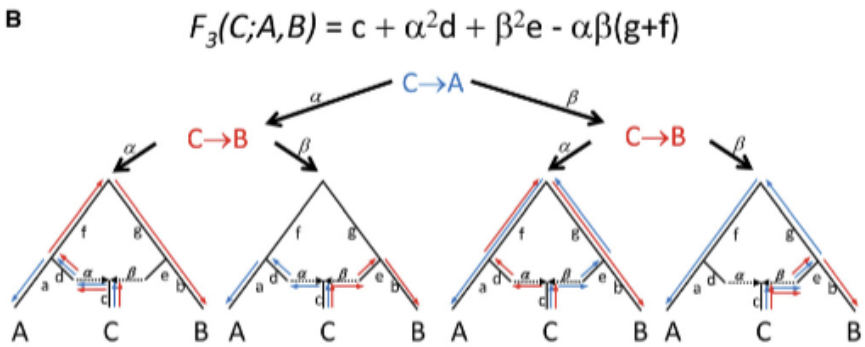


Figure: 2

$F_3(C; A, B) = E[(c' - a')(c' - b')]$ can be negative only if population C has ancestry from population related to both A and B.
Therefore, observation of a significantly negative value of $f_3(C; A, B)$ is evidence of complex phylogeny in C.

- Note that history of admixture does not always result in a negative $f_3(C; A, B)$ statistic.
- Complex history for A or B cannot produce negative $F_3(C; A, B)$

Additivity of F_2 along a tree branch

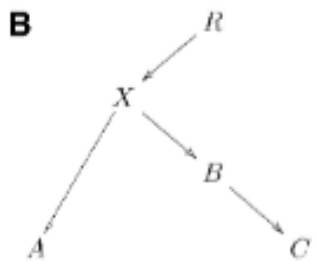


Figure: 1-B

Consider a marker polymorphic at the root. Then,

- Drift on a given edge is a random variable with mean 0 i.e. $E[b'|a'] = a'$
- Drift on two distinct edges of a tree are orthogonal, i.e. $E[(a' - b')(b' - c')] = 0$

It is due to we ascertain in an outgroup.
Also, we assume neutrality and that we can ignore recurrent or backmutations.

Additivity of F_2 along a tree branch

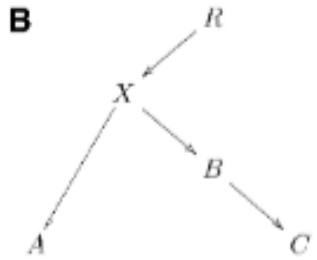


Figure: 1-B

Theorem (Additivity of F_2)

$$F_2(A, C) = F_2(A, B) + F_2(B, C)$$

Proof.

$E[(a' - c')^2] = E[(a' - b' + b' - c')^2] =$
$E[(a' - b')^2] + E[(b' - c')^2] + 2E[(a' - b')(b' - c')]$ □

Expected values of our f-statistics

From Reich et al 2009, supplementary S2.

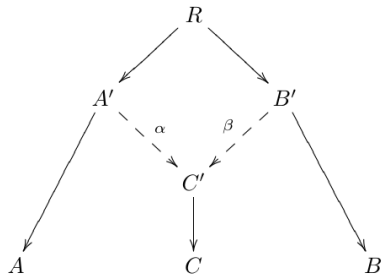


Figure: From Reich et al. 2009

Note that

$$F_3(C; A, B) = F_3(C'; A, B) + F_3(C, C')$$

Label alles at a marker 0,1. Then picking chromosomes from population C independently we can write

$$F_3(C'; A, B) = E(c' - a)(c'' - b)$$

where a,b,c',c" are alleles in populations A,B,C'. However, c' originated from A' with probability α, and so on.

Expected values of our f-statistics

Thus,

$$F_3(C'; A, B) = E(c' - a)(c'' - b)$$
$$= \alpha^2 E(a' - a)(a'' - b) + \beta^2 E(b' - a)(b'' - b)$$
$$+\alpha\beta E(a' - a)(b' - b) + \alpha\beta E(b' - a)(a' - b)$$

where a', a" are independently picked from A' and b', b" from B'. The first 3 terms vanish, Further

$$E(b' - a)(a' - b) = -E((a' - b')^2)$$

, and we obtain

$$F_3(C; A, B) = F_2(C, C') - \alpha\beta F_2(A', B')$$

NOTE that the history of A and B doesn't effect the validity of the test.

The outgroup case

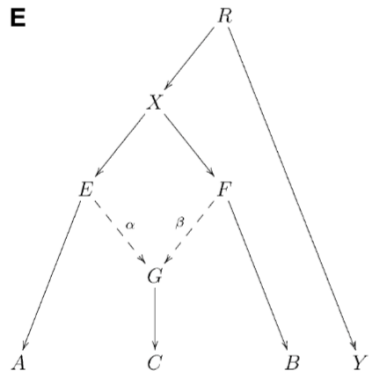


Figure: 1E

Here, Y is an outgroup.

$$F_3(C; A, Y) = F_2(C, G) + \beta^2 F_2(F, X) - \alpha\beta F_2(E, X)$$

The F_3 value can be negative in this case, even if C is not admixed by A and Y. Note that outgroup Y doesn't effect $F_3(C; A, Y)$, i.e. we obtain same F_3 value for any outgroup Y.

Estimates of mixing proportions

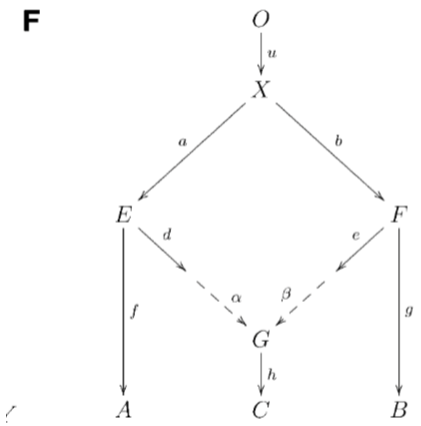


Figure: 1F

For this procedure(Actually in much of work in this article), we need an outgroup for population A,B,C.
We can set an outgroup by letting chimpanzee as a second outgroup, and comparing F_3 statistics, as in previous slide.

Estimates of mixing proportions

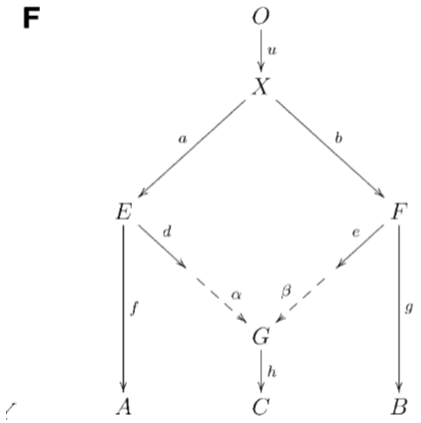


Figure: 1F

Now, we can obtain estimates of

- $Z_0 = u = F_3(O; A, B)$
- $Z_1 = u + \alpha a = F_3(O; A, C)$
- $Z_2 = u + \beta b = F_3(O; B, C)$
- $Z_3 = u + a + f = F_2(O, A)$
- $Z_4 = u + b + g = F_2(O, B)$
- $Z_5 = u + h + \alpha^2(a + d) + \beta^2(b + e) = F_2(O, C)$
- And $F = h - \alpha\beta(a + b) = F_3(C; A, B)$

Estimates of mixing proportions

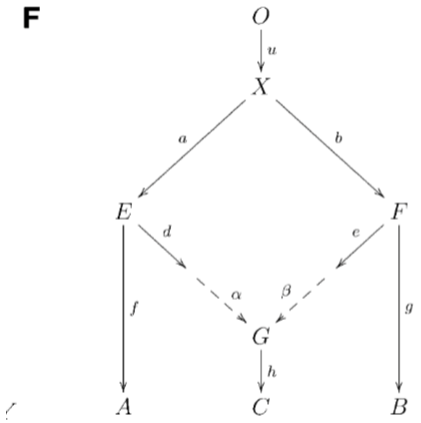


Figure: 1F

Set $Y_i = Z_i - Z_0, i = 0..5$, which eliminates u.
Therefore, any population O which is true outgroup should give similar estimates for Y_i.
We have three inequalities:

- $\alpha \geq Y_1/Y_3$
- $\beta \geq Y_2/Y_4$
- $\alpha\beta(a+b) \leq -F$

Estimates of mixing proportions

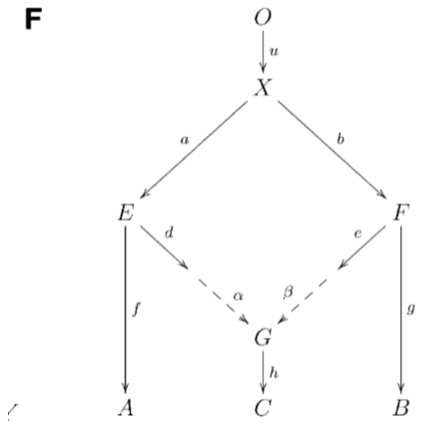


Figure: 1F

We can rewrite these as :

- $Y_1/Y_3 \leq \alpha \leq 1 - Y_2/Y_4$
- $\alpha(Y_2 - Y_1) \geq -F - Y_1$

Now, we have lower bound α_L, upper bound α_U of α.

Estimates of mixing proportions

- Obviously, these bounds are invariant to choices of the outgroup O.
- But choices for source populations A,B may make a substantial difference.
- If we observe that the lower bound exceeds the upper, even when the Z-score for admixture for C is significant, We interpret this as suggesting that our simple model(1F) is wrong.

Z-score?

Normalization of f-statistics

- Magnitude of f_3-statistics strongly depends on the distribution of derived allele frequencies of SNPs examined
- Sign of f_3-statistics is not dependent on the magnitudes of allele frequencies.

We normalize f-statistics using an estimate for each SNP of the heterozygosity of the target population C.

$$T_i = (c' - a')(c' - b')$$

$$B_i = 2c'(1 - c')$$

where B_i is heterozygosity of the target population C.
Now we normalize our f_3 statistic,

$$f_3^* = \frac{\sum_i \hat{T}_i}{\sum_i \hat{B}_i}$$

D-statitics

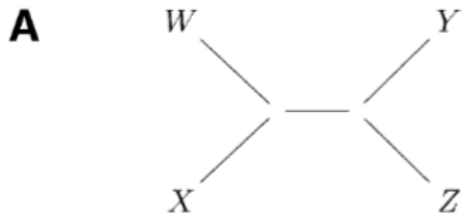


Figure: 3A

D-statistics are used to test whether two populations form a clade.
Let W, X, Y, Z be four populations, with a phylogeny that corresponds to the unrooted tree of Figure 3A.
For SNP i, suppose variant population allele frequencies are w', x', y', z', respectively.

We define

- "BABA" event : the W and Y alleles agree, and the X and Z alleles agree, while the W and X alleles are distinct.
- Define "ABBA" event in similar sense with "BABA" event.

D-statitics

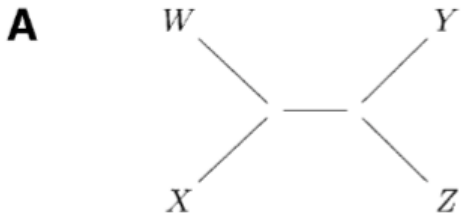


Figure: 3A

Let Num_i and Den_i be the numerator and denominator of the statistic D:

$$Num_i = P(BABA) - P(ABBA) = (w' - x')(y' - z')$$

$$Den_i = P(BABA) + P(ABBA) = (w' + x' - 2w'x')(y' + z' - 2y'z')$$

It's known that replacing w',x'... by w,x.. yields UE of Num_i and Den_i.
Thus, we can define:

$$\hat{Num}_i = (w - x)(y - z)$$

$$\hat{Den}_i = (w + x - 2wx)(y + z - 2yz)$$

Then finally define D-statistic:

Interpretation of D-statitics

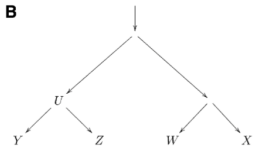


Figure: 3B

Theorem

If we ascertain in an outgroup, then if (W,X) and (Y,Z) are clades in population tree, then $E[Num_i] = 0$.

Proof.

We ascertained in an outgroup, which implies each drift is orthogonal.
Let common ancestor of Y,Z be A, common ancestor of W,X be B.
Then,

$$E(Num_i) = E(w' - x')(y' - z') = E(w' - b' + b' - x')(y' - a' + a' - z') = 0$$

due to the orthogonality of each drift. □

Interpretation of D-statitics, by Durand et al(2011).

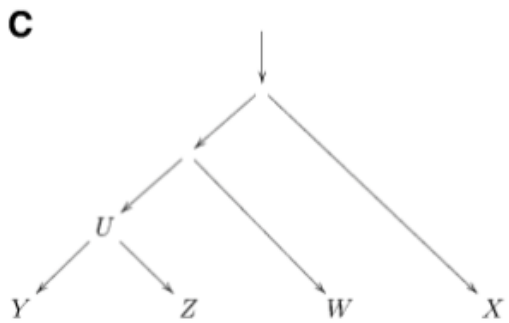


Figure: 3C

- The null hypothesis that we want to test is a demographic scenario in which Y and Z decend from a common ancestral population that diverged from the ancestors of W at an earlier time, without any gene flow between W and Y or Z after they split.
- The alternative hypothesis is that W exchanged genes with Y or Z after these two populations diverged.

Interpretation of D-statitics, by Durand et al(2011).

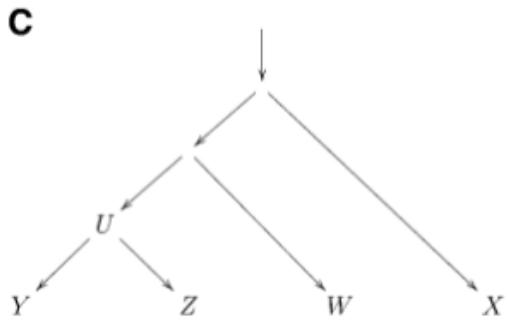


Figure: 3C

- Under the null hypothesis, then $P(BABA) = P(ABBA)$ must hold, which leads D of 0.
- What if D is not 0?
  - If D $>$ 0, It may imply that gene-flow between W to Y is greater than that of W to Z.
  - If D $<$ 0, It may imply that gene-flow between W to Z is greater than that of W to Y.

Interpretation of D-statitics, by Durand et al(2011).

We can compute a standard error for D using weighted block jackknife.
The number of standard errors that this quantity is from zero forms a **Z-Score**, which is approximately normally distributed and thus yields a formal test for whether (W,X) forms a clade.

Weight block jackknife?

- Nearby SNPs are not independent(Linkage Disequilibrium)
- Ignoring LD usually inflates Z-scores.
- Idea is to divide genome into blocks. And Delete block each trial and compute an estimate of D.
- Estimate is weighted, usually the number of SNPs used in block.

F_4 estimation

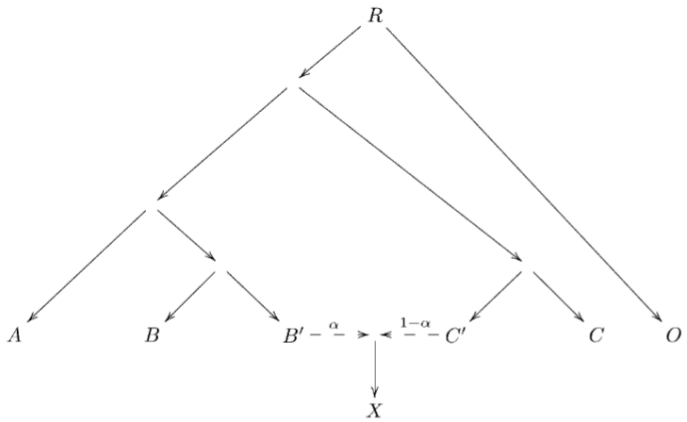


Figure 4 A phylogeny explaining f_4-ratio estimation.

Figure: 4

F_4 estimation is a method for estimating ancestry proportions in admixed population, under assumptions that we have a **correct** historical model.

We have genetic data from A, B, X, C, O.

$$F_4(A, O; X, C) = \alpha F_4(A, O; B', C) = \alpha F_4(A, O; B, C)$$

$$\therefore \hat{\alpha} = \frac{f_4(A, O; X, C)}{f_4(A, O; B, C)}$$

Simulations to test accuracy of f and D statistics

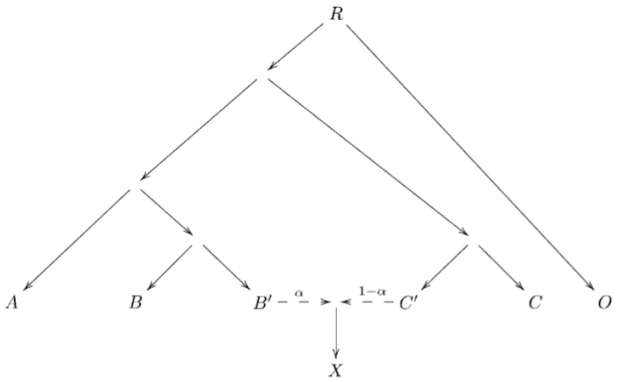


Figure 4 A phylogeny explaining f_4-ratio estimation.

Figure: 4

Table 1 Behavior of *f*- and *D*-statistics for a simulated scenarios of admixture

Scenario	$F_{st}(C, B)$	$F_{st}(O, B)$	$D(A, B; C, O)$	$D(A, X; C, O)$	$f_3(B; A, C)$	$f_3(X; A, C)$	f_4-ratio
Baseline	0.10	0.14	0.00	−0.08	0.002	−0.005	0.47
Vary sample size							
n = 2 from each population	0.10	0.14	0.00	−0.08	0.002	−0.005	0.47
Vary SNP ascertainment							
Use all sites (full sequencing data)	0.10	0.13	0.00	−0.11	0.001	−0.002	0.47
Polymorphic in a single *B* individual	0.10	0.16	−0.01	−0.06	0.003	−0.006	0.47
Polymorphic in a single *C* individual	0.10	0.16	0.00	−0.13	0.003	−0.007	0.46
Polymorphic in a single *X* individual	0.11	0.16	0.00	−0.11	0.003	−0.007	0.49
Polymorphic in two individuals: *B* and *O*	0.10	0.16	−0.01	−0.08	0.002	−0.005	0.46
Vary demography							
N_A = 2,000 (*vs.* 50,000) pop *A* bottleneck	0.10	0.14	0.00	−0.08	0.002	−0.005	0.48
N_B = 2,000 (*vs.* 12,000) pop *B* bottleneck	0.14	0.17	0.00	−0.08	0.011	−0.004	0.48
N_C = 1,000 (*vs.* 25,000) pop *C* bottleneck	0.16	0.14	0.00	−0.08	0.002	−0.005	0.46
N_X = 500 (*vs.* 10,000) pop *X* bottleneck	0.10	0.14	0.00	−0.08	0.002	0.004	0.47
$N_{ABB'}$ = 3,000 (*vs.* 7,000) *ABB′* bottleneck	0.14	0.17	0.00	−0.09	0.002	−0.007	0.47

Figure: table 1

Admixture graph fitting

We first remark that given n populations $P_1, P_2, ...P_n$, then

- The f-statistics span a linear space V_F of dimension $\binom{n}{2}$
- All f-statistics can be found as linear sums of statistics $f_2(P_i, P_j)1 \leq i < j$
- Fix a population (say P_1), Then all f-statistics can be found as linear sums of statistics $f_3(P_1; P_i, P_j), f_2(P_1, P_i)1 < i < j$.

These statements are true(WHY?), both for the theoretical F-values, and for our f-statistics, at least when we have no missing data, so that for all populations our f-statistics are computed on the same set of markers.

Admixture graph fitting

- The f-statistics span a linear space V_F of dimension $\binom{n}{2}$
  - Each F_2 has independent values, without constraint.
- All f-statistics can be found as linear sums of statistics $f_2(P_i, P_j) 1 \leq i < j$
  - $(a-b)^2 = ((c-a)-(c-b))^2 = (c-a)^2 + (c-b)^2 - 2(c-a)(c-b)$, which implies $2f_3(C; A, B) = f_2(C, A) + f_2(C, B) - f_2(A, B)$
  - Next, writing d - b = c - b -(c - d),

$$f_4(C, A; D, B) = f_3(C; A, B) - f_3(C; A, D)$$

    holds.
- Fix a population (say P_1), Then all f-statistics can be found as linear sums of statistics $f_3(P_1; P_i, P_j), f_2(P_1, P_i) 1 < i < j$.
  - $f_2(A, B) = f_2(C, A) + f_2(C, B) - 2f_3(C; A, B)$

Admixture graph fitting(qpGraph)

Procedure for admixture graph fitting is as follows;

1. f-statistics on the basis. Call the resulting $\binom{n}{2}$ long vector f.
2. An estimated error covariance Q of f using the weighted block jackknife.
3. Now, given a graph topology, as well as graph parameters, we can calculated g, the expected value of f.
4. Use score function

$$S_2(g) = -\frac{1}{2}\sum_i \frac{(g_i - f_i)^2}{(Q_{ii} + \lambda)}$$

- A natural score function is

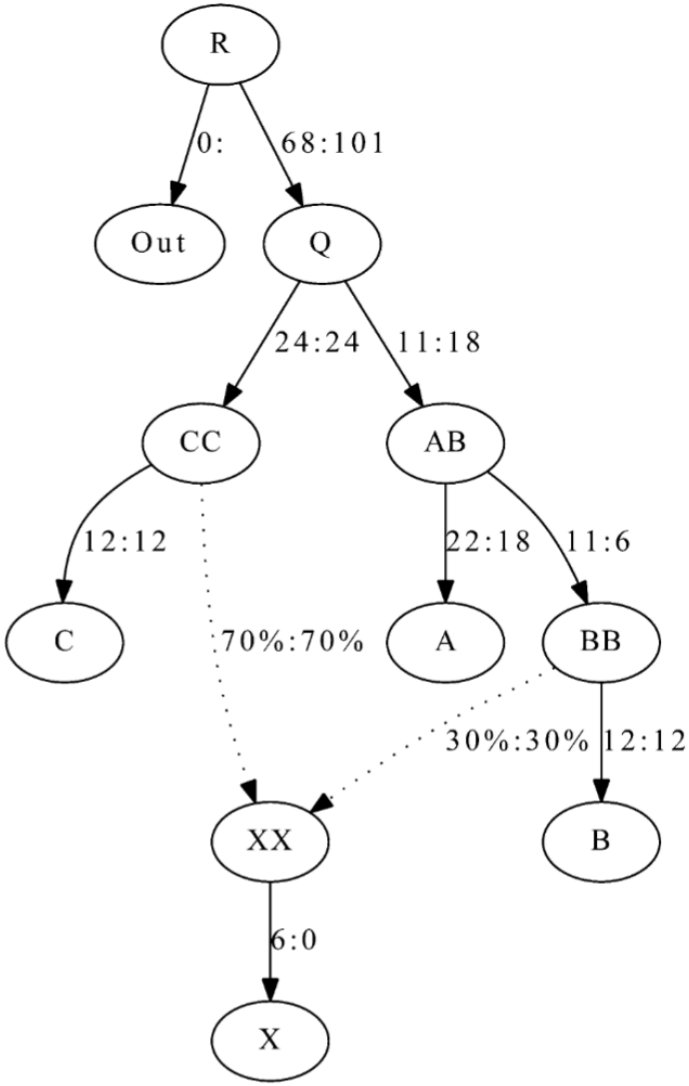
$$S_1(g) = -\frac{1}{2}(g - f)' Q^{-1} (g - f)$$

, but When n is large, our estimate of Q is unstable.

Interpretation and limitation of qpGraph

1. A major use of qpGraph is to show that a hypothesized phylogeny must be incorrect. This generalizes our D- statistic test, which is testing a simple tree on four populations.
2. After fitting parameters, study of which f-statistics fit poorly can lead to insights as to how the model must be wrong
3. Overfitting can be a problem, especially if we hypothesize many admixing events, but only have data for a few populations

Interpretation and limitation of qpGraph



Rolloff

Rolloff studies the decay of admixture linkage disequilibrium with distance to infer the date of admixture.
Rolloff method uses the fact :

Theorem

Consider two alleles on a chromosome in an admixed individual at loci that are a distance d apart.
Then n generations after admixture, with probability e^{-nd} *the two alleles belonged, at the admixing time, to a single chromosome.*

The procedure for rolloff is as follows:

- Set a weight function w at each SNP that is positive when the variant allele has a higher frequency in population A than in B and negative in the reverse situation.
- we compute an LD-based score z(s1 , s2) which is positive if the two variant alleles are in linkage disequilibrium; that is, they appear on the same chromosome more often than would be expected assuming independence

Rolloff

Note that, we can use the difference of empirical frequencies of variant allele as weight function w
And, we can get LD-based score z(s1, s2) as follows :

1. let v_1, v_2 be the vectors of genetype counts of the variant allele.
2. calculate the Pearson correlation ρ between v_1, v_2
3. we use Fisher's z-transformation, i.e

$$z = \frac{\sqrt{m-3}}{2} log(\frac{1+\rho}{1-\rho})$$

where m is the number of samples in which neither s_1 or s_2 has missing data.

Rolloff

The procedure for rolloff is continued:

- Form a correlation between our z-scores and the weight function. Explicitly, for a bin-width x, define the 'bin' $S(d)$, $d = x, 2x, 3x..$by the set of SNP pairs(s_1, s_2), where

$$S(d) = \{(s_1, s_2) | d - x < u_2 - u_1 \leq d\}$$

  where u_i is the genetic position of SNP s_i.
- Then we define $A(d)$ to be the correlation coefficient

$$A(d) = \frac{\sum_{s_1, s_2 \in S(d)} w(s_1) w(s_2) z(s_1, s_2)}{[\sum_{s_1, s_2 \in S(d)} (w(s_1) w(s_2))^2 \sum_{s_1, s_2 \in S(d)} (z(s_1, s_2))^2]^{\frac{1}{2}}}$$

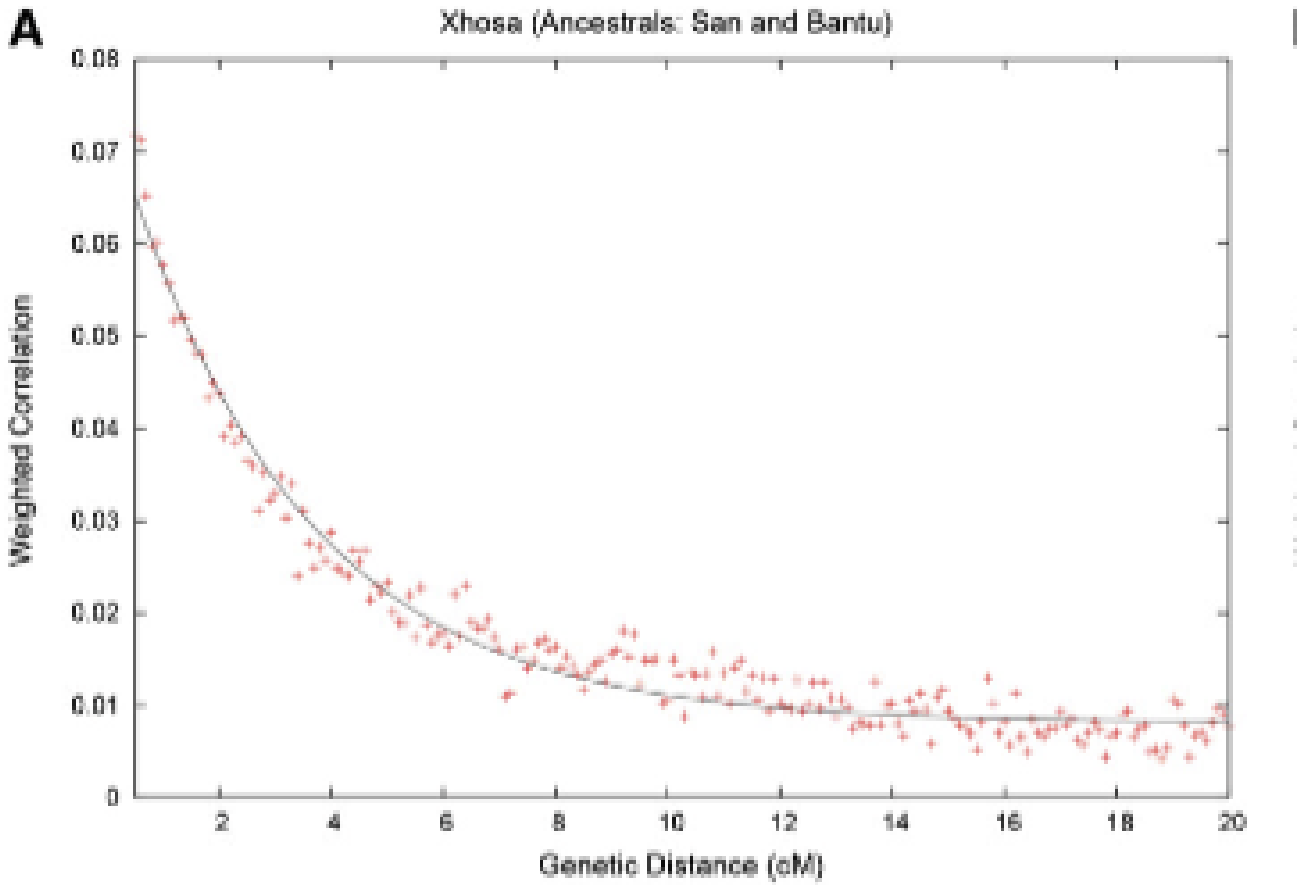
- We fit

$$A(d) \approx A_0 e^{-nd}$$

  then finally we can get the estimate of n.

Rolloff

Interpretation of $A(d)$?



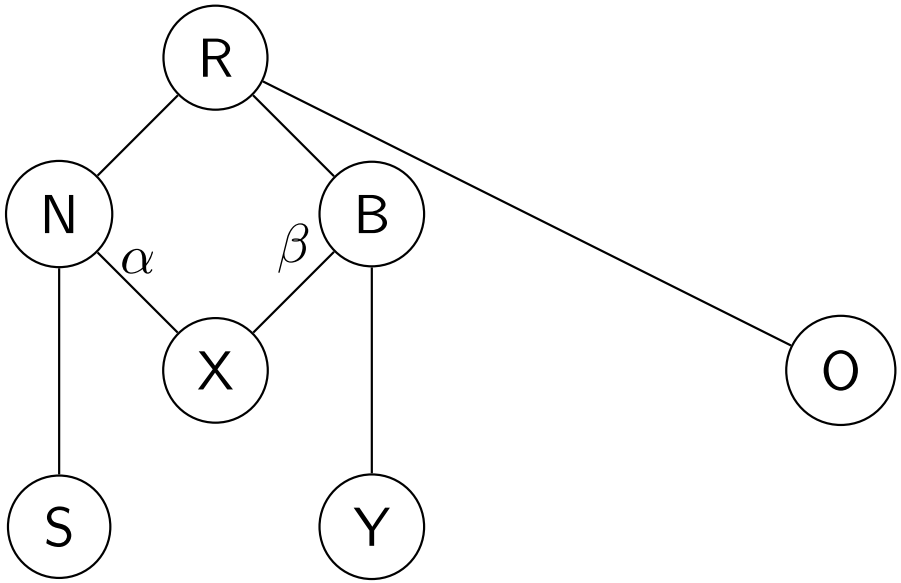
Xhosa: 740 ± 30 years ago

Figure: 7A

Results and Discussion

South African Xhosa

Assumed phylogeny is as follows:
(R : root, N : Nguni, B : Bushman, X : Xhosa, S : San, Y : Yoruba, O : outgroup(Han))



- By running three-population test with San and Yoruba, authors got $f_3(X; S, Y) = -0.009$, Z-score = -33.5
- By using Han as outgroup, got bound of α, which is $0.19 \leq \alpha \leq 0.55$
- By running rolloff using San and Yoruba as reference populations, got estimate of admixing event date of 25.3 += 1.1 generations.

Admixture of Uygur

The Uygur are known to be historically admixed.

- By trying f anlaysis on Uygur using French and Han as source population, authors got

$$0.452 \leq \alpha \leq 0.525$$

  where α is West Eurasian admixture proportion.
- We can discover that
  - gene-flow does not involve an African population. Because

$$D(Yoruba, Han; French, Russian)$$

$$\approx D(any\ sub-saharan\ African, Han; French, Russian)$$

  - gene-flow between the relatives of the Han and Russians. Because

$$D(Yoruba, Han; French, Russian)$$

$= 0.192$, $Z = 26.3$

- By rolloff, estimated admixture date is 790 $+=$ 60 years before present.

Northern European gene flow into Spain

- By running three-population test with Ireland and Sardinia, author discovered that there was geneflow from northern european to spanish.
- By running rolloff, got estimate of admixing event date of 3600 $+=$ 400 years before present.

An example of the outgroup case.

$$f_3(Greece; Albania, Yoruba) = -0.0047, Z = -5.8$$

$$f_3(Greece; Albania, Papuan) = -0.0033, Z = -3.5$$

We can see Yoruba as an Outgroup.

Admixture events detected in Human Genome Diversity Panel populations

Source1	Source2	Target	f_3	Z-score	α_L	α_U	Z_{San}	Z_{Han}	$\alpha_L > \alpha_R$
Daur	Dai	Han	−0.0026	−13.20	0.664	0.26	−7.89	−6.31	*
Han	Orcadian	Tu	−0.0076	−20.64	0.875	0.926	−12.38	−8.98	

Figure: table5

- if $\alpha_L > \alpha_U$, it suggests that our three-population phylogeny is not feasible.
- Tu has complex history. The fact that $f_3(Tu; Han, Orcadian) < 0$ with significant Z-score **does not** mean Tu is admixed from the two given source populations. Instead one should interpret this line as meaning that an East Asian population related genetically to a population ancestral to the Japanese has admixed with a West Eurasian population.

Thank you!