Genetic relatedness analysis: classical and new approaches

Genetic relatedness analysis: modern data and new challenges

by

Bruce S. Weir, Amy D. Anderson (Dept. of Biostatistics, Univ. of Washington),
Amanda B. Hepler (Dept of Statistical Science, Univ. College London)

Nature Reviews (Genetics) 2006, (7), 771-780

Presented and commented by T. Möls: JC, 26th January 2007

Problems

A. Probabilistic problems

- What numerical measures can be used to characterise of how related are two individuals?
- What measures the degree of inbreeding of an individual?

B. Statistical problems

- Given genotypes (SNPs, microsatellites, etc) of two individuals, how related they are?
- Given genotypes (SNPs, microsatellites, etc) of two individuals, in which relationship (brothers, parent-son etc.) they are?

A. It is relatively easy to calculate the probability of a certain type similarity between the genotypes of individuals who are relatives and have a known pedigree.

B. Much more difficult can do the reverse: infer probabilities of various possible relationships between the two individuals given their genotypes.

The simultaneous analysis of more markers (micro-satellite loci, SNPs etc) increases the reliability of relationship inference and allows more detailed statements to be made.

Probabilistic problems

In relatedness analysis, two definition of allele (or gene, SNP, microsatellite etc) identities play the key role:

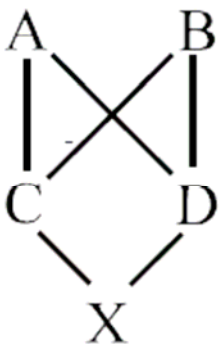
1. Two subunits of DNA are called identical-by-descent (IBD, päritolult identsed) if they are chemical copies of the same subunit of the ancestral DNA (or one of them is copy of the other).
2. Two subunits of DNA are called identical-by-state (IBS, olekult identsed) if they are chemically equal.

It is not possible experimentally distinguish between these identities!

Indiviidi inbriidingukoefitsient mõõdab indiviidi vanemate omavahelist sugulust.

Inbriidingukoefitsient (inbreeding coefficient) on tõenäosus, et alleelid suvaliselt valitud lookuses on päritolult identsed.

Definitsioon ei sõltu lookuse valikust ega alleeli tõenäosusest populatsioonis.

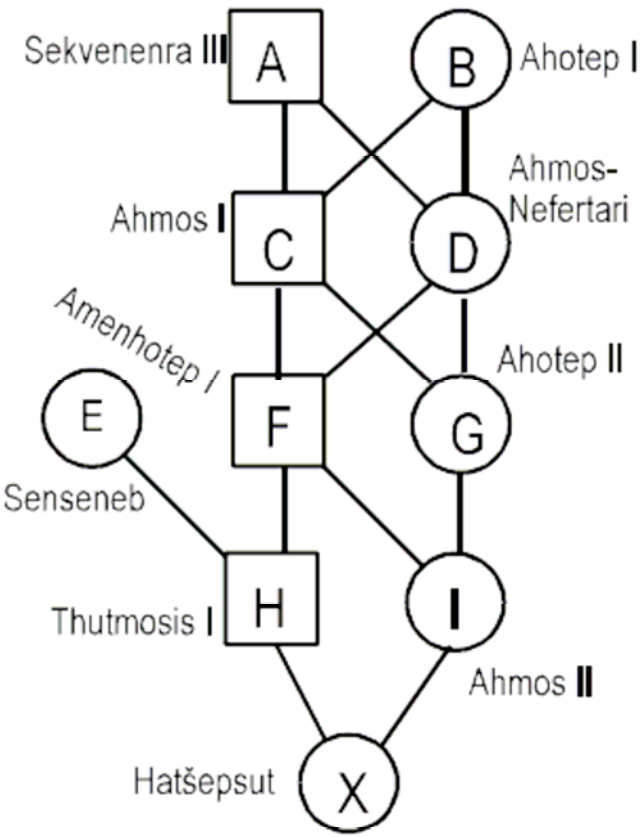


Teoreem (S. Wright, 1921). Indiviidi X IK avaldub valemiga

$$F_X = \sum_{i=1}^{k} \left(\frac{1}{2}\right)^{n_i+1} (1 + F_i), \qquad (10)$$

kus k on trajektooride arv indiviidi X ühe vanema juurest teise juurde läbi ühise eellase, n_i on lõikude arv i-ndal trajektooril ja F_i on trajektoorile i vastava ühise eellase IK.

Joonisel näidatud X inbriidingukoefitsiendi saab leida valemi (10) järgi nii: k=2 (C**A**D ja C**B**D), $n_1=n_2=2$, $F_1=F_2=0$, Seega $F_X=0.25$.



Näide 9. Juuresoleval joonisel on osa Egiptuse vaaraode 18. dünastia (1552 e. m. a.) sugupuust. Leiame Hatšepsuthi IK. Ühised eellased on A, B, C, D ja F, vanemate sugulussidemed on järgmised:

Tee	n_i	F_i
H **F** I	2	¼
HF **C** GI	4	0
HF **D** GI	4	0
HFC **A** DGI	6	0
HFD **A** CGI	6	0
HFC **B** DGI	6	0
HFD **B** CGI	6	0

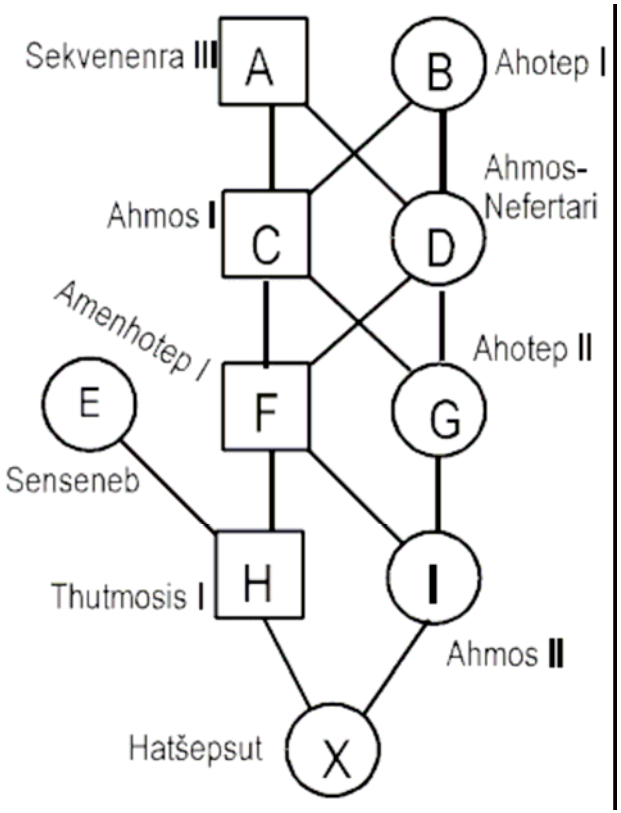
Hatšepsuthi IK eeldusel $F_A=F_B=0$ on seega $F_X = 0.25$ [kontrollimisprogramm: inbrmod2.sas].

```
*Programm inbrmod2;
proc IML;                          *See on SAS moodul;
    *Funktsioon leiab lapse genotyybi vanemate v1 ja v2 genot.
alusel (rekombinatsioon);
start off(v1,v2);                  *=offspring;
     o=j(2,1,0);                   *(2×1) mtrx def.;
     o[1]=v1[(ranuni(0)>0.5)+1];
     o[2]=v2[(ranuni(0)>0.5)+1];
return(o); finish off;
```

```
*Egiptuse vaarao Hatsepsuthi inbriidingukoef. modelleerimine
(teoreetiline väärtus on  0.25);

Start ring(kordi); ikk=0;
do i = 1  to  kordi;   n=0;  k=0;
do i = 1  to  100;

a=j(2,1,0); b=j(2,1,0); e=j(2,1,0); *def. mtrx (2×1), elem=0;

a[1]=1:a[2]=2:b[1]=3:b[2]=4:e[1]=5:e[2]=6:*vanem. alleelid;
c= off(a,b): d= off(a,b): f= off(c,d); g= off(c,d);
h= off(e,f); ii=off(f,g); x= off(h,ii);
n=n+1;
k=k+(abs(x[1]-x[2])<0.01); end;
ikk=ikk+k/n; end;

ikk=ikk/kordi:  print  ikk;  Finish;
```

run ring(200); =>

0.248, 0.252, 0.253
0.244, 0,251, 0.250
0.248, 0.252, 0.253
0.246 /0.250/



It is rather easy to present the pedigree in the simulation procedure:

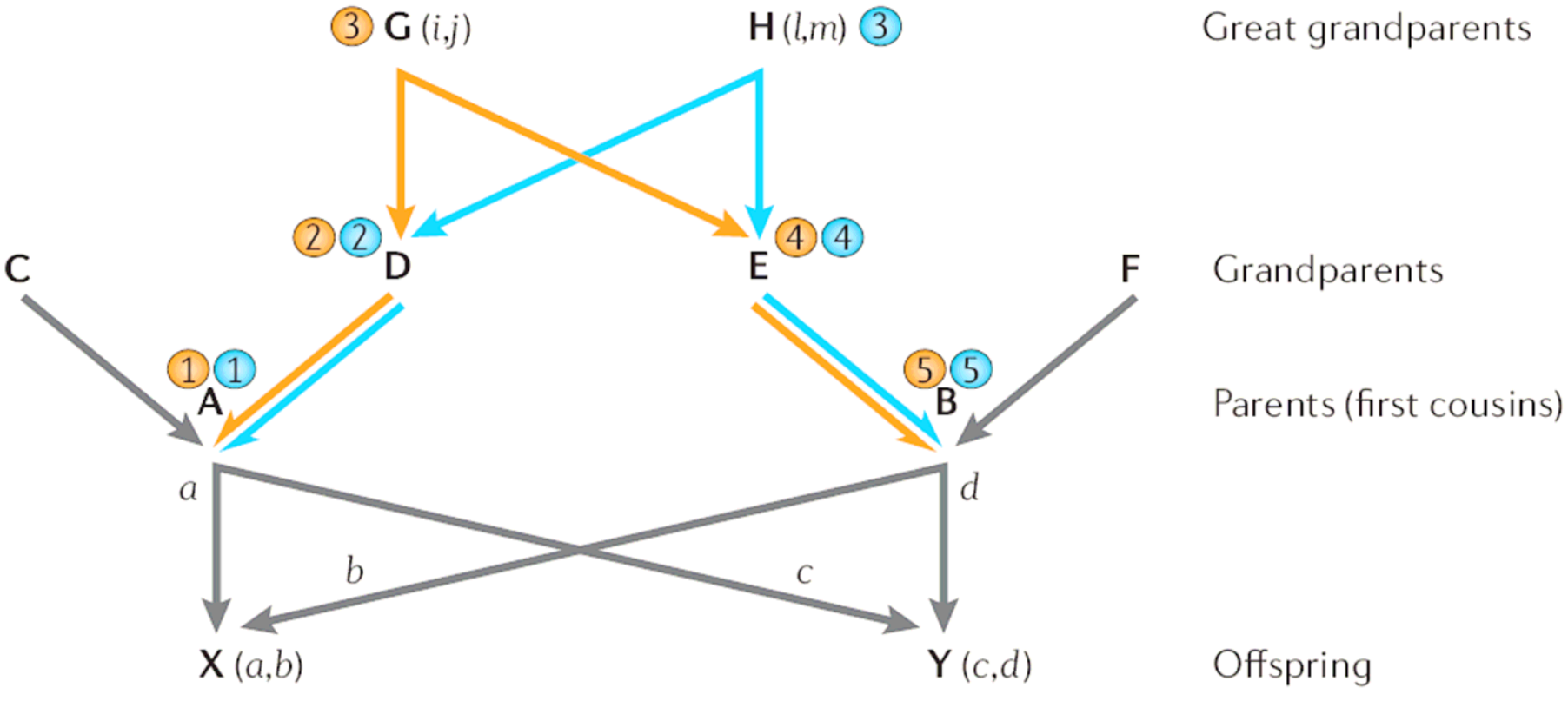
```
c= off(a,b): d= off(a,b): f= off(c,d); g= off(c,d);
h= off(e,f); ii=off(f,g); x= off(h,ii);
```

Kahe indiviidi suguluskoefitsient mõõdab nende omavahelist sugulust.

Kahe indiviidi suguluskoefitsient (coancestry coefficient) on tõenäosus, et valides suvalise lookuse ja valides sellest lookusest kummalgi indiviidil ühe alleeli, on need päritolult identsed.

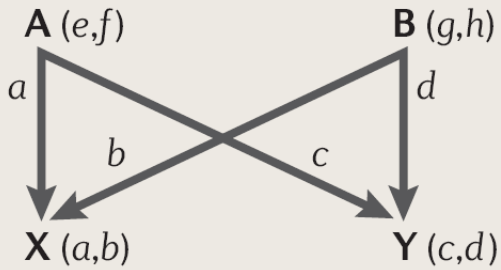
Definitsioon ei sõltu lookuse valikust ega alleeli tõenäosusest populatsioonis.

Calculating the coancestry coefficient



Box 1 | **Measures of relatedness**

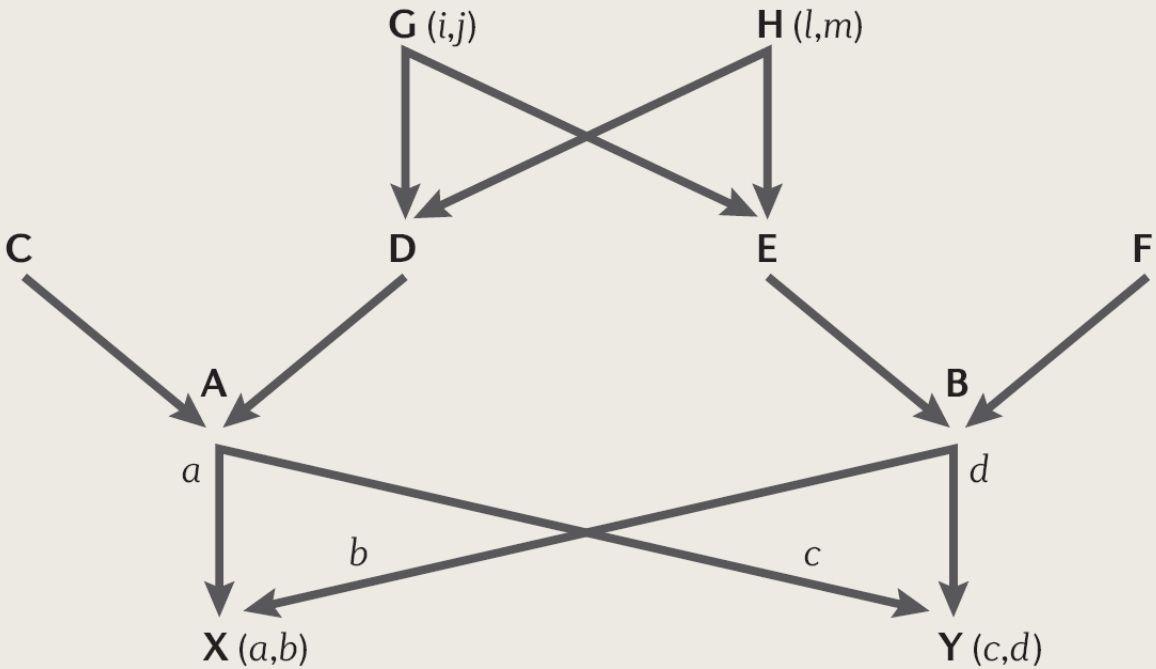
a **Full-siblings**



Parents

Offspring

b **Siblings descended from first cousins**



Great grandparents

Grandparents

Parents (first cousins)

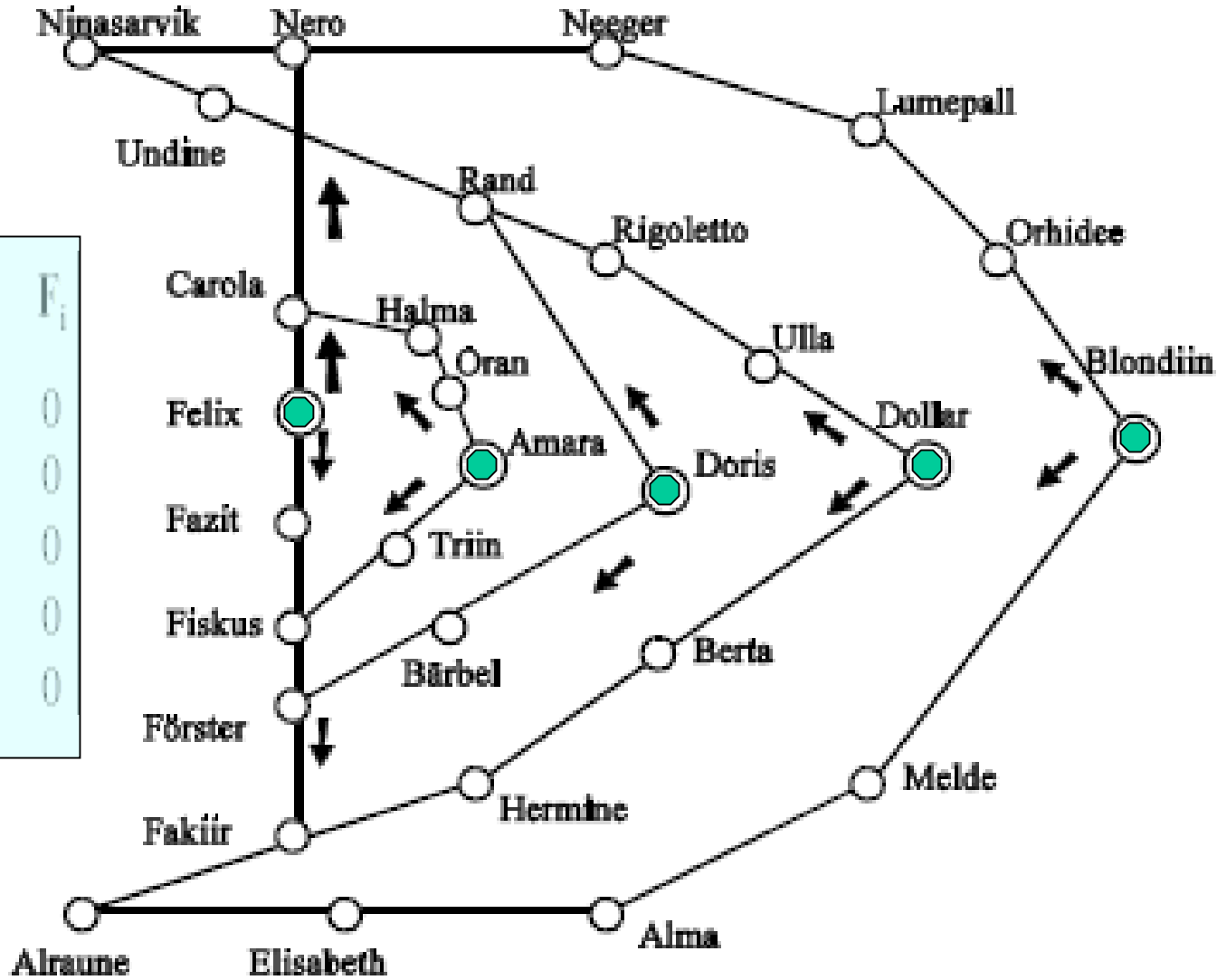
Offspring

Selgub, et kahe indiviidi suguluskoefitsient on võrdne nende järglase oodatava inbriidingukoefitsiendiga. Seega saab suguluskoefitsiendi arvutamiseks kasutada Wrighti valemit (10).

Vaatame näidet.

Felix, Amara, Doris, Dollar ja Blondiin. Leiame F_X, eeldusel, et tihiste eellaste IK on nullid. Arvutused on alljärgnevas tabelis.

Ühine eellane	n_i	F_i
Blondiin	9	0
Dollar	9	0
Doris	7	0
Amara	10	0
Felix	8	0



Ühine eellane	n_i	F_i	F_X liidetav
Blondiin	9	0	$(½)^{10} = 0.0009766$
Dollar	9	0	$(½)^{10} = 0.0009766$
Doris	7	0	$(½)^{8} = 0.0039063$
Amara	10	0	$(½)^{11} = 0.0004883$
Felix	8	0	$(½)^{9} = 0.0019531$
Inbriidingukoefitsient			$F_X = 0.008301$

Siin X on Ninasarviku ja Alraune järglane.

F_X on sama, mis Ninasarviku ja Alraune suguluskoefitsient.

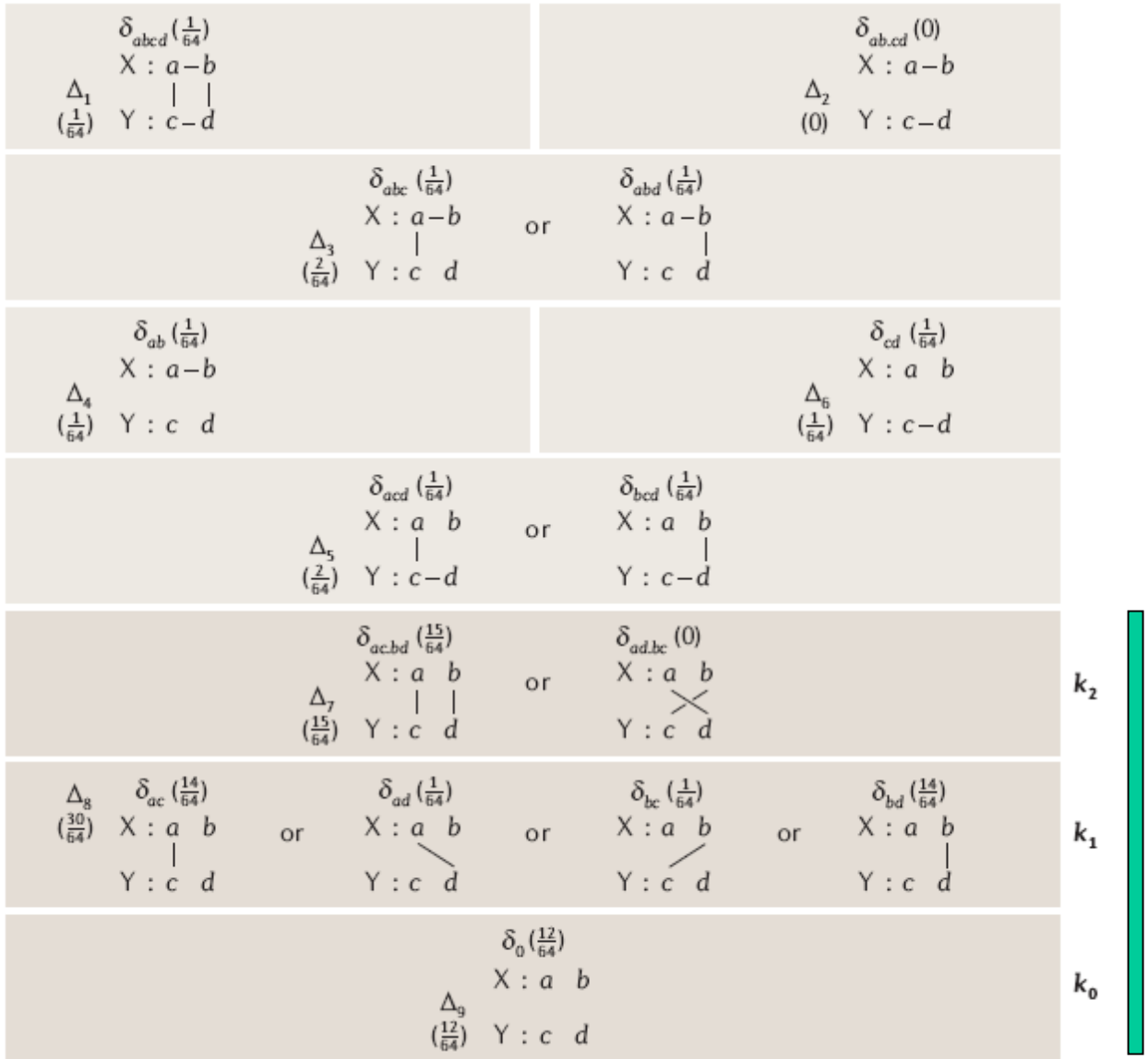
Statistical problems

Comparing the genotypes of two individuals gives information about their relatedness.

Two typical problems:

1) Differentiate between possible types of relatedness.

2) Estimate the coancestry coefficient.



For non-inbred individuals

Table 1 | **Identity-by-descent probabilities for common, non-inbred relatives**

Relationship	k_2	k_1	k_0	$\theta = k_1/4 + k_2/2$
Identical twins	1	0	0	1/2
Full-siblings	1/4	1/2	1/4	1/4
Parent–child	0	1	0	1/4
Double first cousins	1/16	3/8	9/16	1/8
Half-siblings*	0	1/2	1/2	1/8
First cousins	0	1/4	3/4	1/16
Unrelated	0	0	1	0

The coancestry coefficient carries less inormation than the three probabilities.

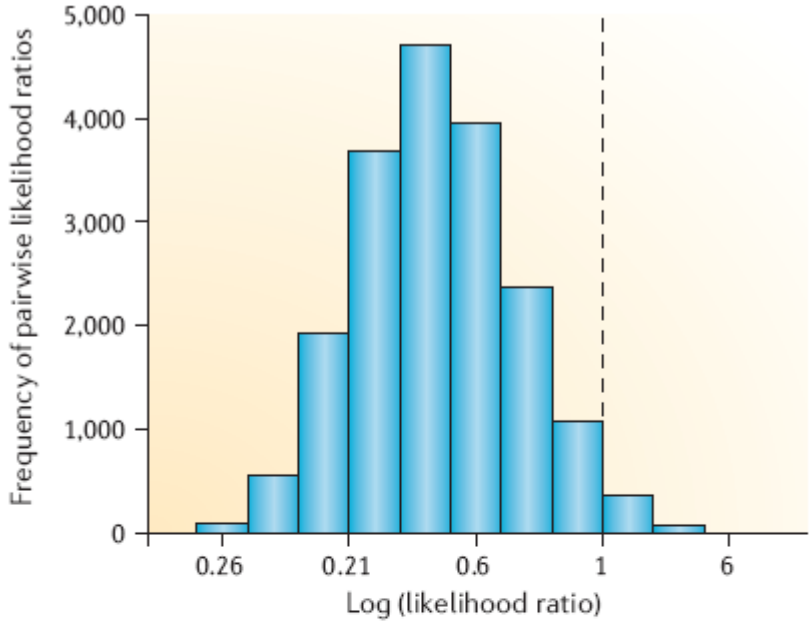


Figure 2 | **Likelihood ratios for putative full-siblings.** Owing to the probabilistic nature of relationship inference, distinguishing between relationships can be difficult, and evidence for a relationship might be found when none is justified. This point is exemplified by a study of 195 Caucasians[35] who were drawn randomly from a population and are presumed to be unrelated. All pairs of individuals in the sample were typed for all of 13 microsatellite loci (the CODIS forensic set[35]). For each pair of individuals, the likelihood ratio for the hypotheses of full-siblings versus unrelated was calculated for each locus and the results were multiplied over loci. A histogram of the 18,195 log-likelihood ratio values is shown: just under 3% of the values have a likelihood ratio greater than 1 (those to the right of the vertical dashed line), which would favour the hypothesis of sibship.

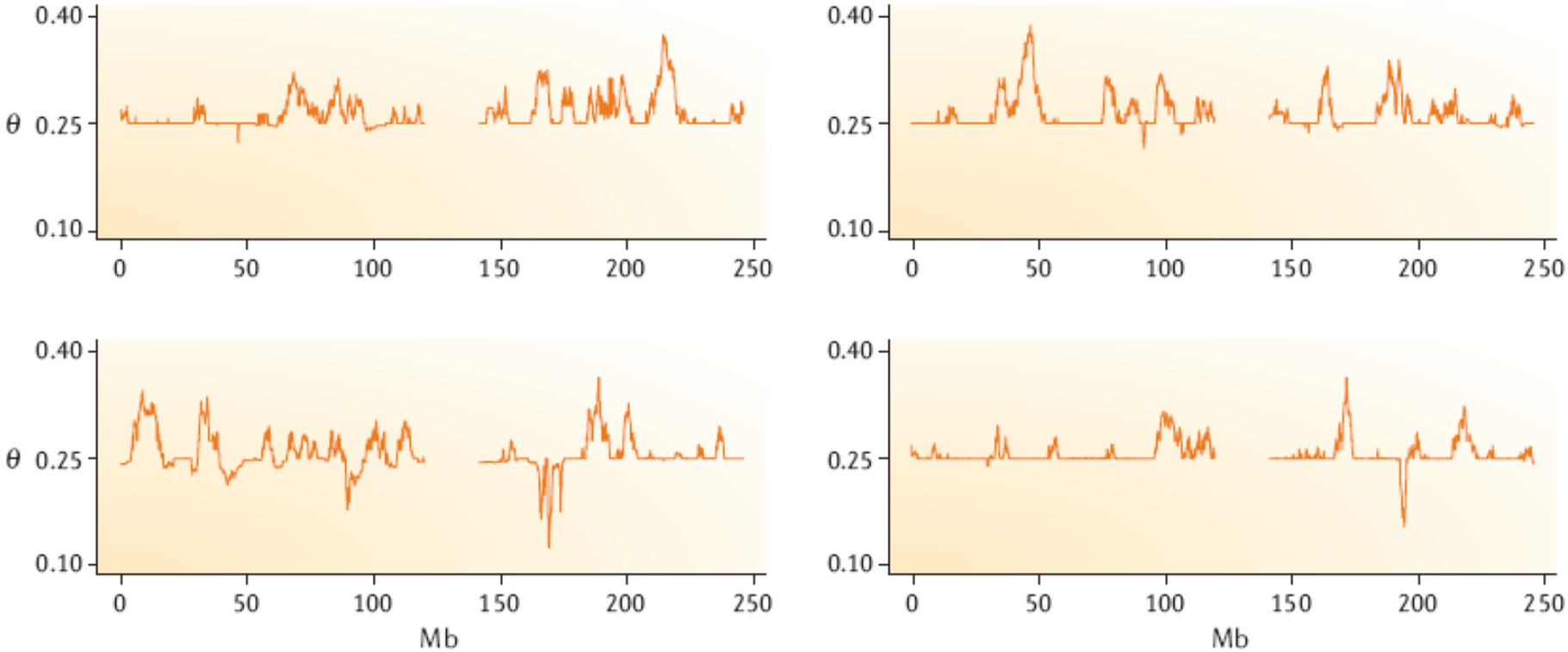
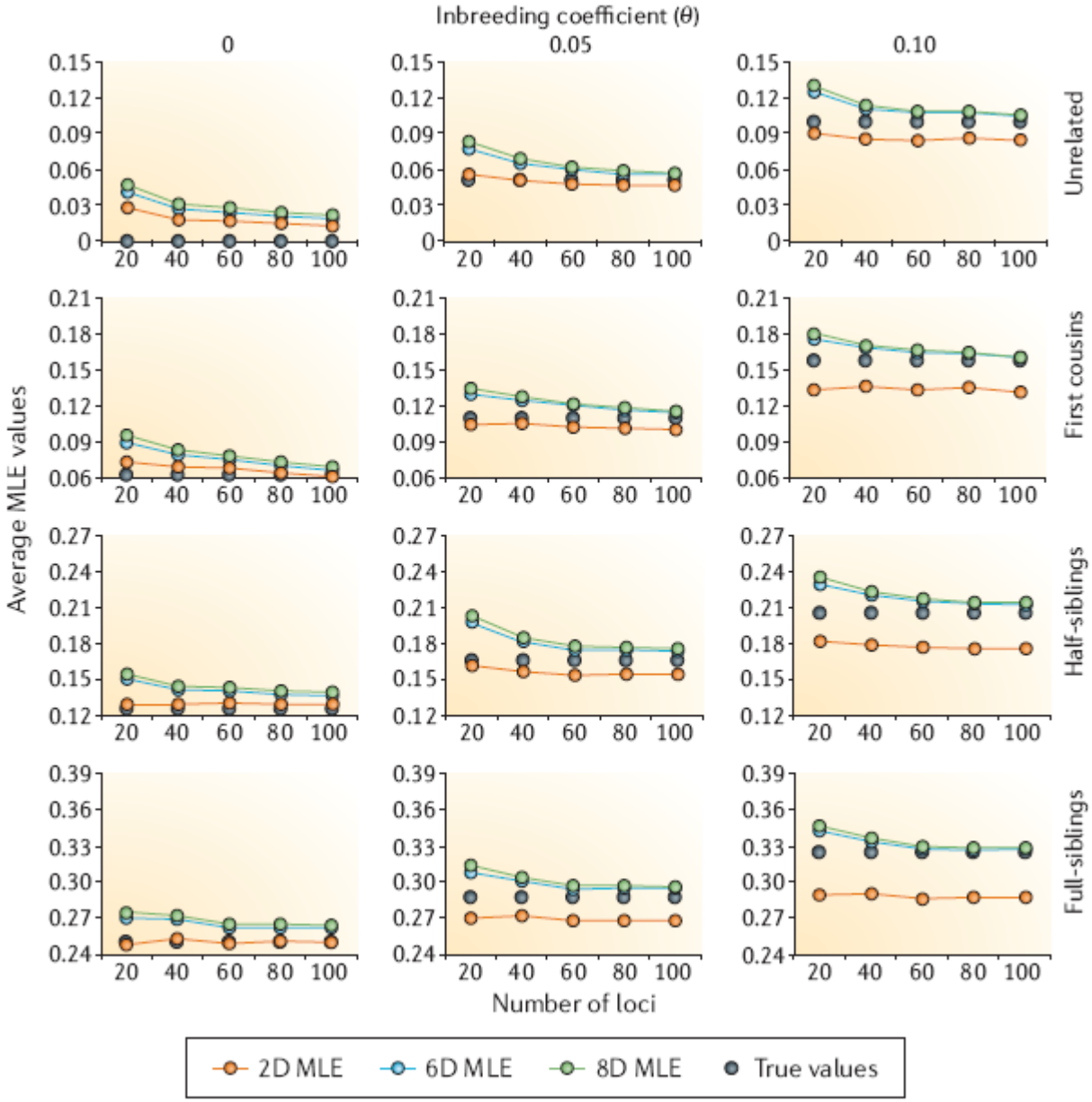
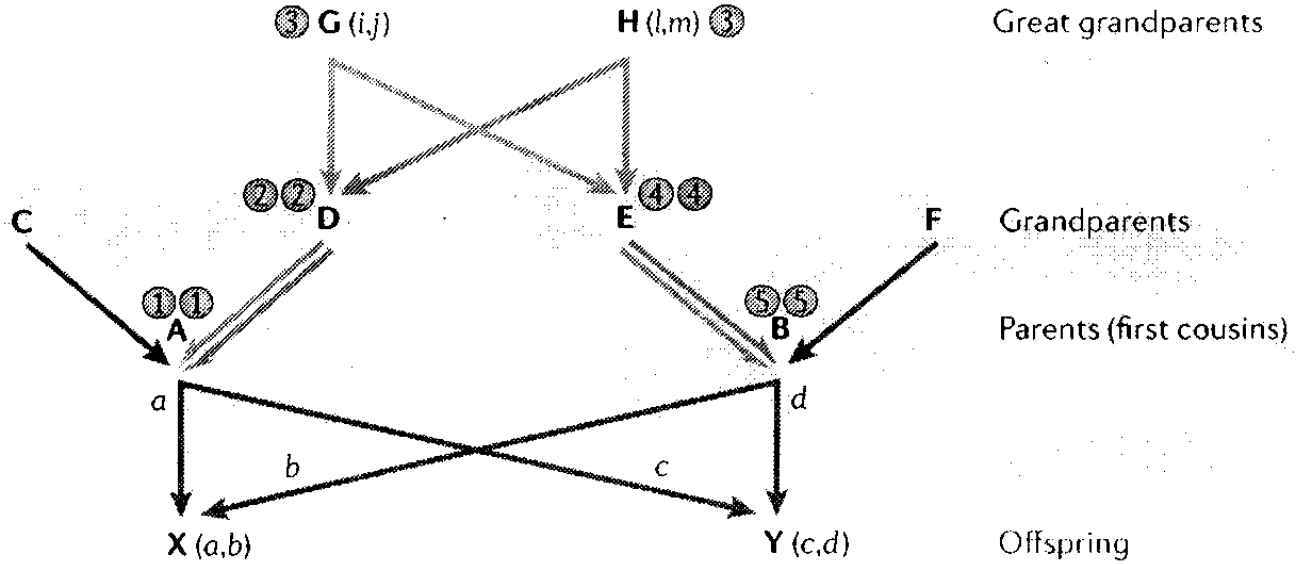


Figure 4 | **Variation in estimated coancestries along a chromosome.** Estimated coancestry coefficients (θ) for four separate parent–child pairs, using successive sets of 100 SNP markers on human chromosome 1. For each pair, the expected coancestry is 0.25, but the estimated actual coancestry varies along the chromosome. The patterns of variation for the four parent–child pairs shown are different, indicating that the variation is not the result of some evolutionary process such as natural selection, but probably represents sampling variation and different genealogies for different chromosomal regions. Because the detailed genealogies of two individuals can vary along the genome, the concept of 'actual' as opposed to 'expected' identity-by-descent becomes important. The variation in actual identity–by-descent is a consequence of evolutionary history and cannot be eliminated, whereas the variation in estimates of identity-by-descent parameters can be reduced (although not eliminated) by using more loci. The smoothing of estimates over loci comes at the expense of masking any variation that might be of real interest.


Inbreeding coefficient (θ)
0
0.05
0.10
Unrelated
First cousins
Half-siblings
Full-siblings
Average MLE values
Number of loci
2D MLE
6D MLE
8D MLE
True values

Box 2 | **Calculating the coancestry coefficient**

Calculating the coancestry coefficient



To calculate the inbreeding coefficients in pedigrees, an individual's genealogy is traced back on both the maternal and paternal sides until an ancestor that is common to both lineages is found. The number *n* of individuals in the pathway that link the parents to the common ancestor, including the parents themselves, is used as a power of 0.5, and the 0.5^n terms are added over all pathways and common ancestors.

For first cousins, such as A and B, $n = 5$; as there is one path (shown in colour) to each of the two grandparents G and H whom they have in common, the inbreeding coefficient of their child X is $2(0.5)^5 = 1/16$. If a common ancestor (for example, G) is himself inbred, then his contribution to the inbreeding coefficient (F) of the descendant is $(1 + F_G)(0.5)^n$.

For the two siblings, X and Y, whose parents, A and B, are first cousins, there are four common ancestors: A and B and the two great-grandparents G and H. There are three people in the paths XAY and XBY, and seven people in each of the paths XADGEBY, XBEGDAY, XADHEBY and XBEHDAY (the paths are not shown). The coancestry coefficient for X and Y is therefore $2(0.5)^3 + 4(0.5)^7 = 9/32$.