GHEP 2024
Advanced theoretical challenges

GHEP kinship exercise 2024: Advanced level

Prepared by: Magnus Dehli Vigeland

General instructions

This is a multiple-choice test consisting of 20 questions. For each question exactly one alternative is correct. You are free to use whatever software you like, but keep in mind that some programs have built-in conventions (e.g., rounding) that may affect the output. If your answer does not precisely match any of the options, choose the closest one.

Files needed to complete the test

- *cousins-data.txt* / *cousins-ibd.txt*. Data for Part II.
- *siblings-data.txt* / *siblings-ibd.txt*. Data for Part III.
- *db.txt*. Allele frequencies for 23 STR markers. (If your software requires database size, use N = 1000.)

Assumptions throughout

- No linkage between markers, no linkage disequilibrium, no deviations from HW equilibrium.
- No drop-outs, drop-ins, silent alleles or mutations.
- Pedigree founders are non-inbred and unrelated to each other.
- The total genetic length of the autosome (chromosomes 1–22) is 3391 cM.

Some definitions

Homologous alleles are *identical by descent* (IBD) if they have the same origin within a given pedigree. The *IBD coefficients* $(\kappa_0, \kappa_1, \kappa_2)$ of non-inbred individuals A and B, are the probabilities of sharing respectively 0, 1 and 2 alleles IBD at a random autosomal locus. They are related to the kinship coefficient φ by the formula $\varphi = \kappa_1/4 + \kappa_2/2$.

The *IBD triangle* (Figure 1) is a convenient tool for visualising IBD coefficients. Note that, since $\kappa_0 + \kappa_1 + \kappa_2 = 1$, any two of them suffice to deduce the third; the choice of κ_0 and κ_2 is simply my personal preference. The online tool *QuickPed* may be useful for calculating IBD coefficients and plotting them in the IBD triangle.

Traditional coefficients like κ and φ measure the **expected** IBD sharing based on the pedigree. In contrast, the *realised* (or *genomic*) relatedness between A and B refers to the **actual** IBD segments they share as a result of recombination (Figure 2). We denote by (k_0, k_1, k_2) the actual proportions of the autosome, in terms of genetic length, where they share 0, 1 and 2 alleles IBD, respectively. The *realised kinship coefficient* is given by $\varphi_R = k_1/4 + k_2/2$.

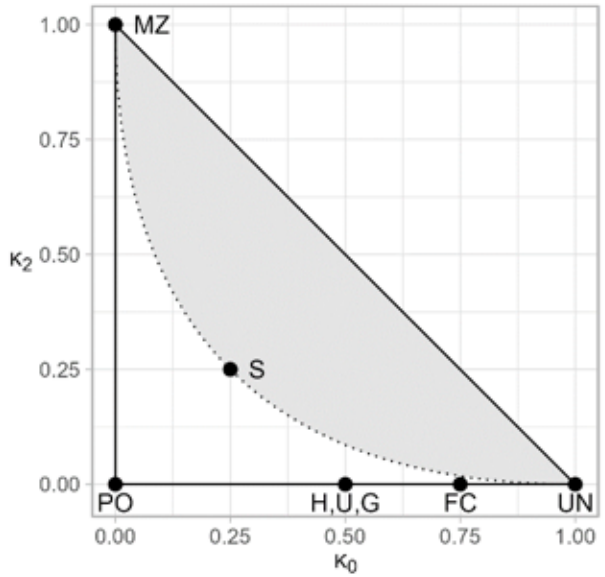


Figure 1. The IBD triangle. FC=first cousins; G=grandparent-grandchild; H=half sibs; MZ=monozygous twins; PO=parent-offspring; S=full sibs; U=uncle-nephew; UN=unrelated

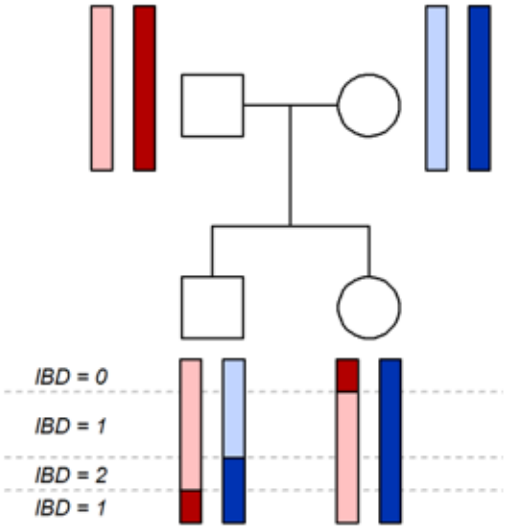


Figure 2. An example of the realised IBD sharing between siblings. The chromosome is divided in segments with IBD status 0, 1 or 2

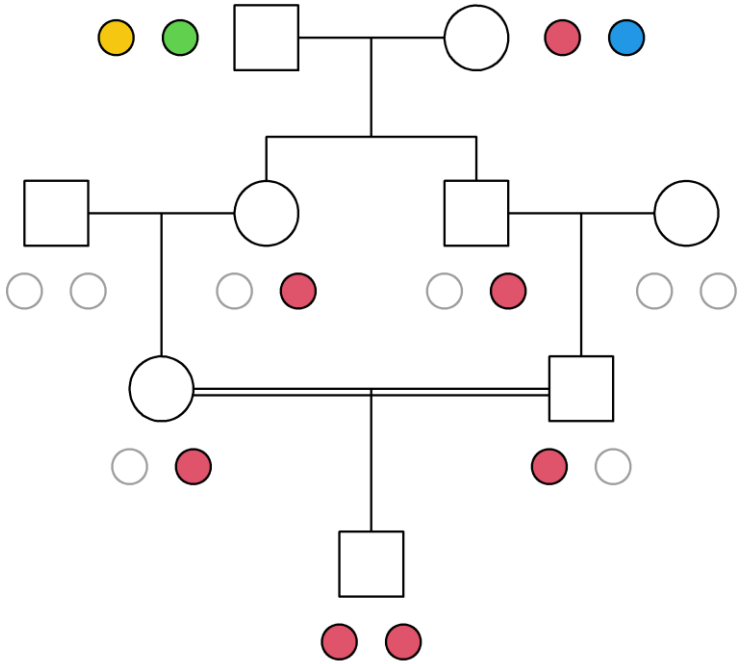
Main topics

- Identity by descent (IBD)
- Realised relatedness

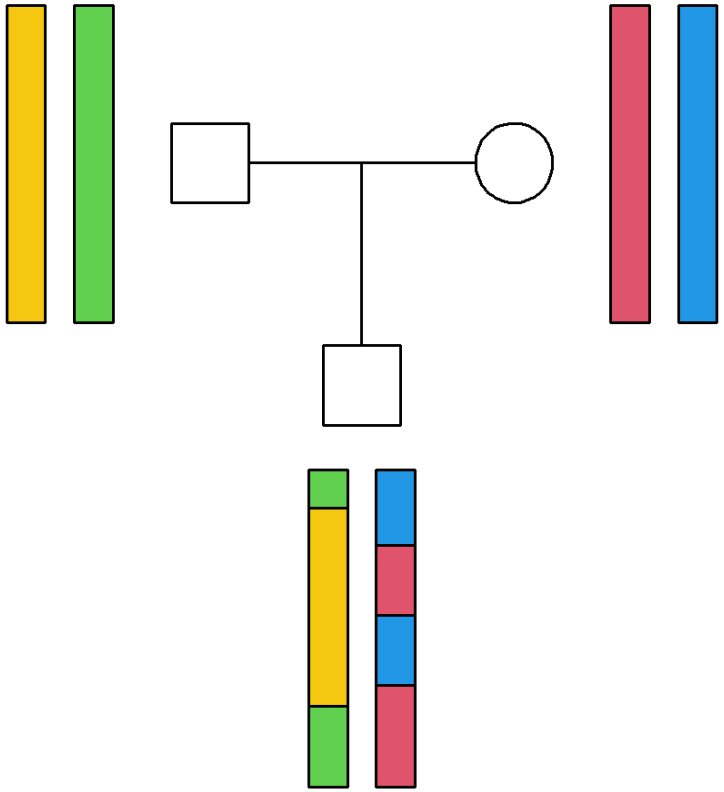
Identity by descent

- IBD = Identity by descent

  = identical alleles with a common origin **in the given pedigree**



Recombination

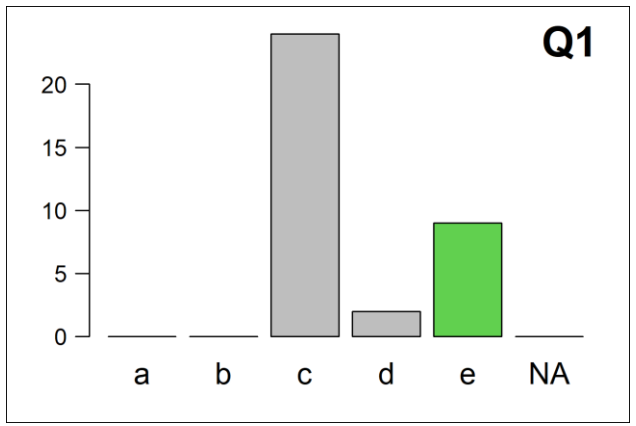


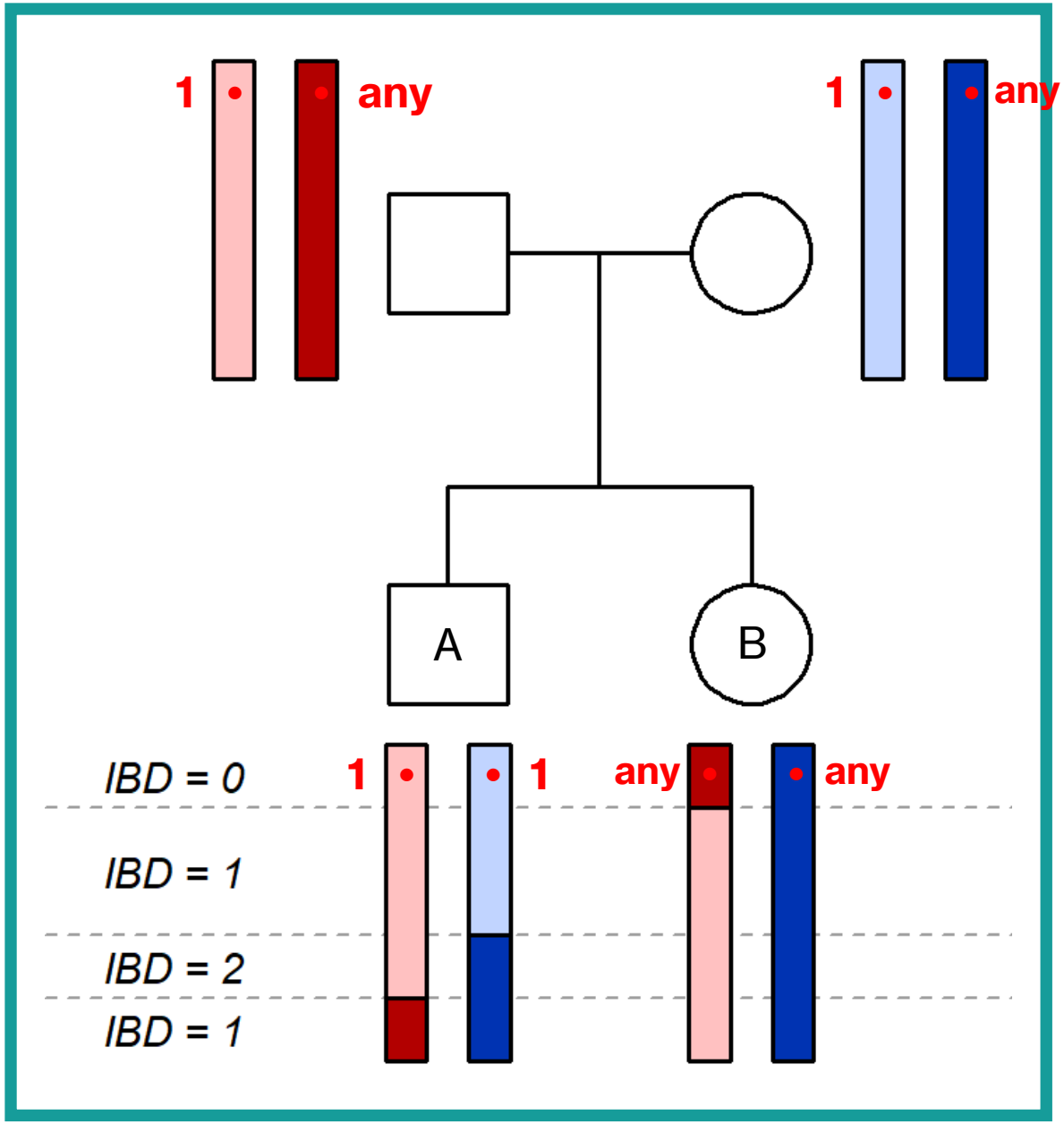
- Genetic distance between two loci:
  = average # crossovers/ meiosis
- Units:
  - 1 Morgan (M) = 1 crossover per meiosis
  - 1 centiMorgan (cM) = 0.01 M
- The human genome: Ca 30 Morgan

Part I: Warm-up

We consider a situation where two individuals, A and B, are typed with a tri-allelic marker. The alleles are labelled 1,2,3, and the allele frequencies are p_1, p_2, p_3, respectively.

1. Suppose the marker lies in a region where A and B have IBD status 0. If A has genotype 1/1, the genotype of B is
   a) 2/3
   b) 2/2 or 3/3
   c) 2/2, 2/3 or 3/3
   d) anything except 1/1
   e) anything





2. Suppose the marker lies in a region with IBD = 1.

If A has genotype 1/2, the genotype of B is

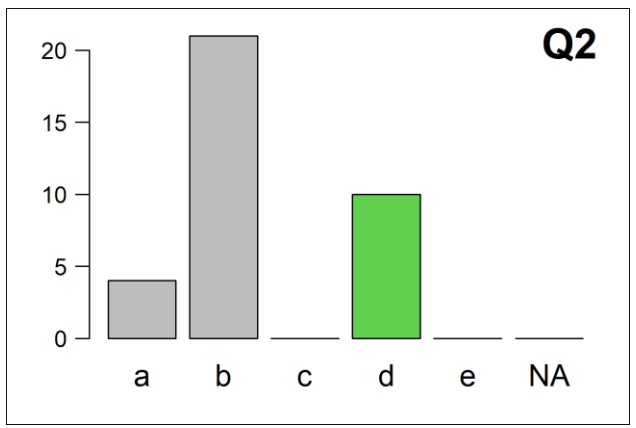
a) 1/3 or 2/3

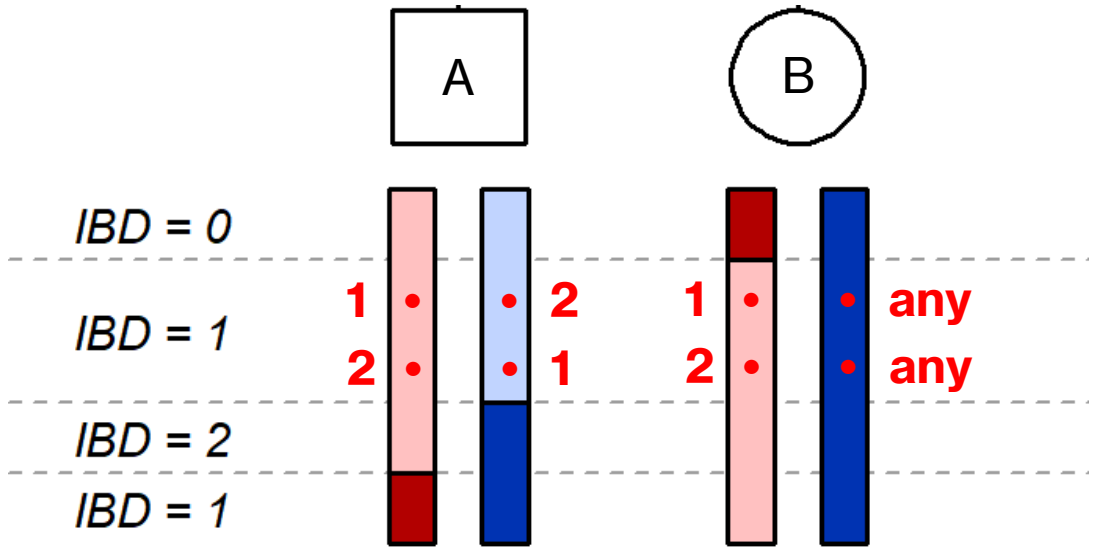
b) 1/1, 2/2, 1/3 or 2/3

c) anything except 1/2

d) anything except 3/3

e) anything





3. Given that the marker is in a region with IBD = 2, the probability that A and B are homozygous for the same allele, is

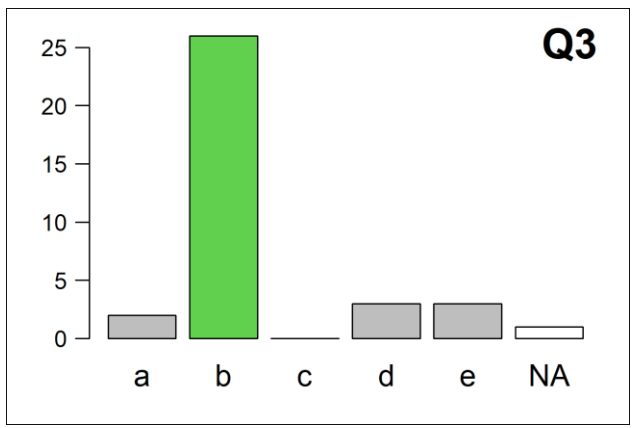
a) 0

b) $p_1^2 + p_2^2 + p_3^2$

c) $p_1(1-p_1) + p_2(1-p_2) + p_3(1-p_3)$

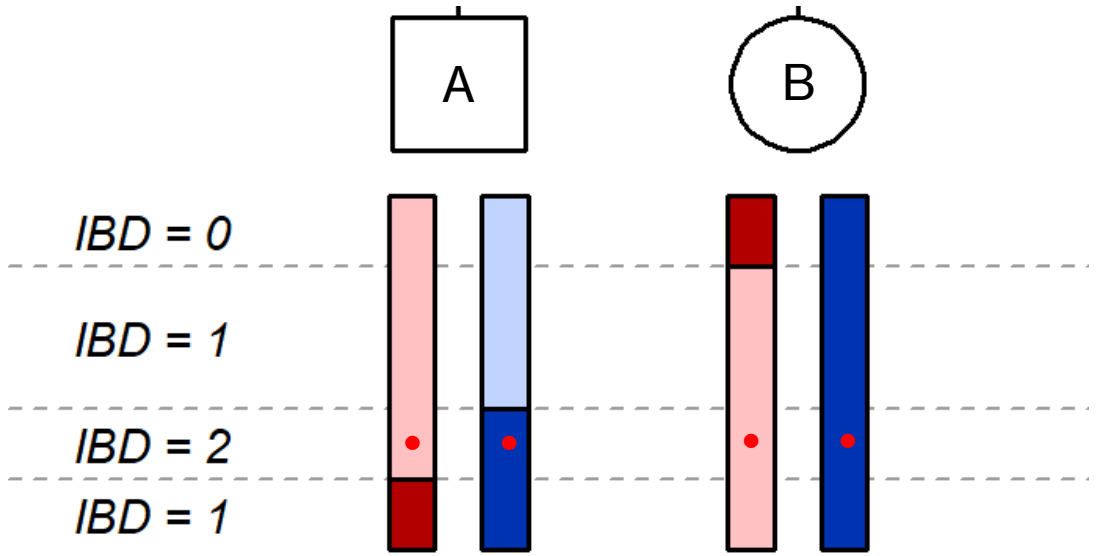
d) $p_1(1-p_1)^2 + p_2(1-p_2)^2 + p_3(1-p_3)^2$

e) 1



- IBD = 2:
  A and B always has the same genotype
- Hardy-Weinberg:

$$P(A = 1/1) = p_1^2$$
$$P(A = 2/2) = p_2^2$$
$$P(A = 3/3) = p_3^2$$



4. Given that the marker is in a region with IBD = 1, the probability of a *full match* (i.e., A and B have the same genotype) is

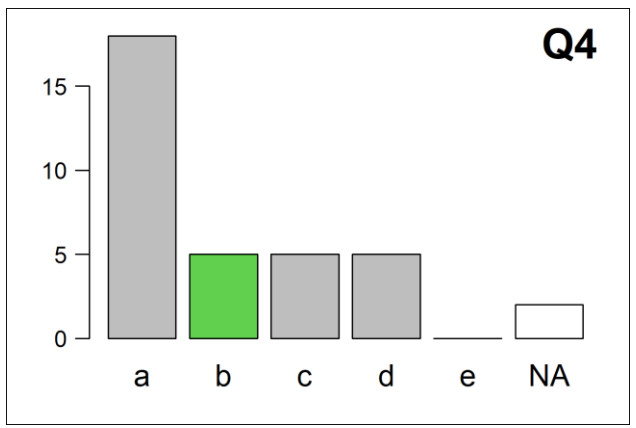
a) 0

b) $p_1^2 + p_2^2 + p_3^2$

c) $p_1(1-p_1) + p_2(1-p_2) + p_3(1-p_3)$

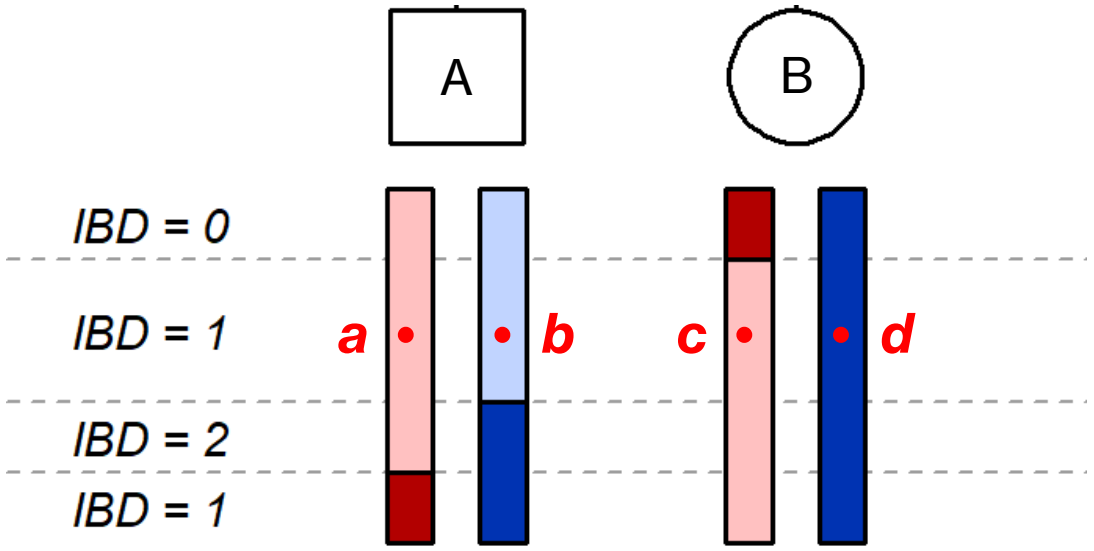
d) $p_1(1-p_1)^2 + p_2(1-p_2)^2 + p_3(1-p_3)^2$

e) 1



- Full match means $\boldsymbol{a/b = c/d}$
- IBD = 1 implies that $\boldsymbol{a = c}$
- What about $\boldsymbol{b = d}$?

$$P(\boldsymbol{b = d} = 1) = p_1 \cdot p_1$$
$$P(\boldsymbol{b = d} = 2) = p_2 \cdot p_2$$
$$P(\boldsymbol{b = d} = 3) = p_3 \cdot p_3$$



5. Given that the marker is in a region with IBD = 1, the probability of a *partial match* (i.e., exactly one shared allele) is

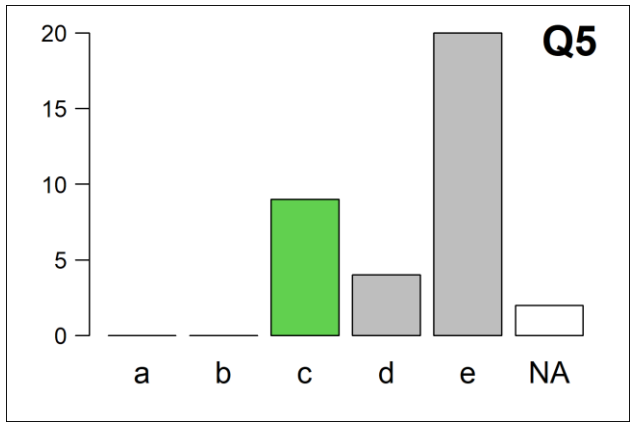
a) 0

b) $p_1^2 + p_2^2 + p_3^2$

c) $p_1(1-p_1) + p_2(1-p_2) + p_3(1-p_3)$

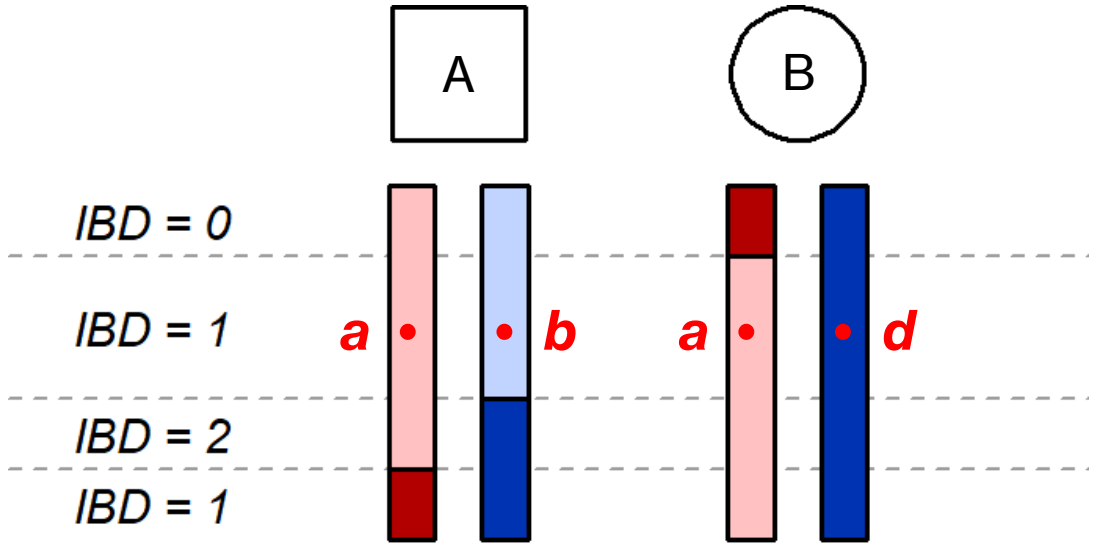
d) $p_1(1-p_1)^2 + p_2(1-p_2)^2 + p_3(1-p_3)^2$

e) 1



Partial match if $\boldsymbol{b \neq d}$

$$P(\boldsymbol{b} = 1, \boldsymbol{d} \neq 1) = p_1(1-p_1)$$
$$P(\boldsymbol{b} = 2, \boldsymbol{d} \neq 2) = p_2(1-p_2)$$
$$P(\boldsymbol{b} = 3, \boldsymbol{d} \neq 3) = p_3(1-p_3)$$

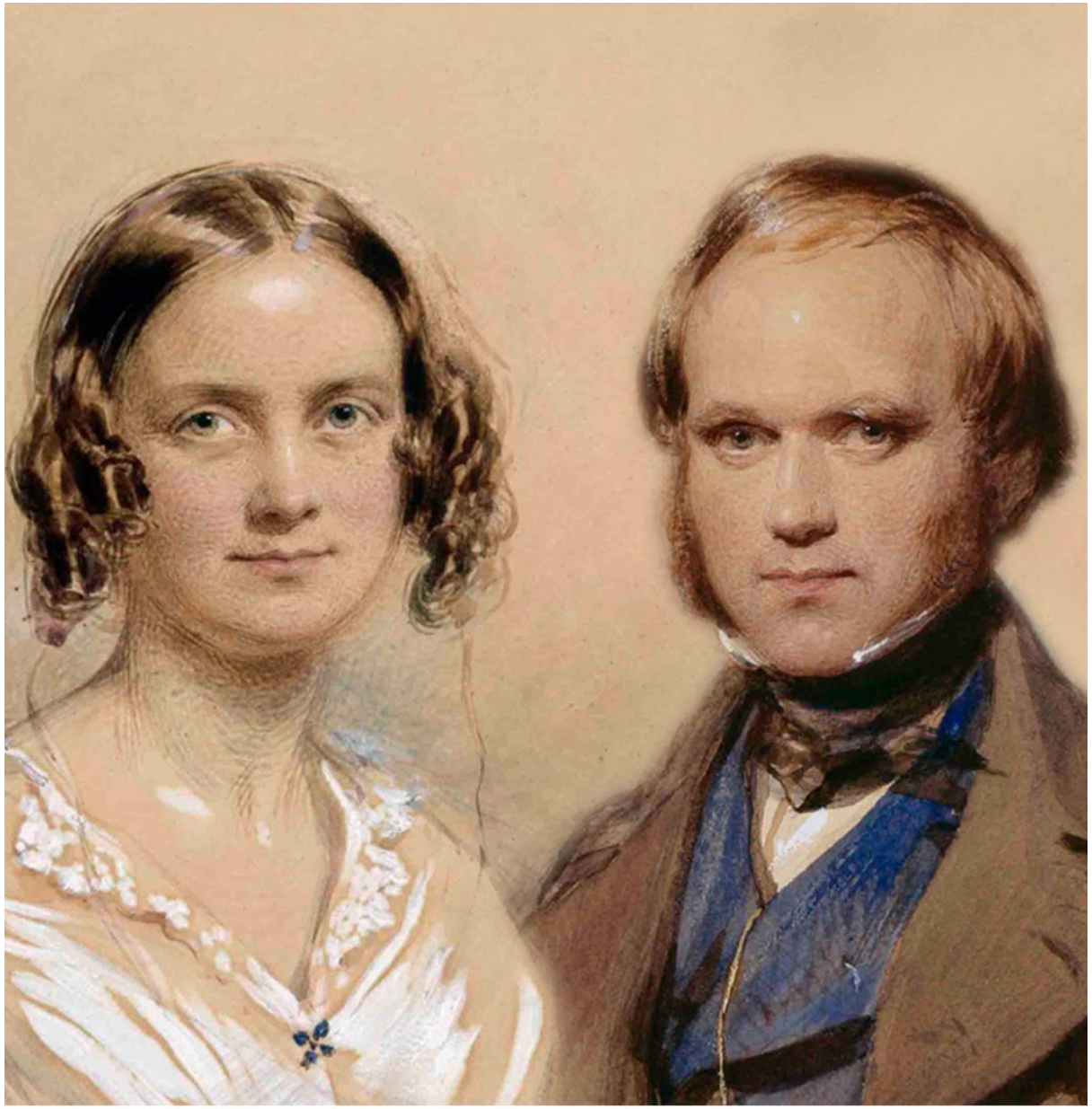


Part II: A case of cousins

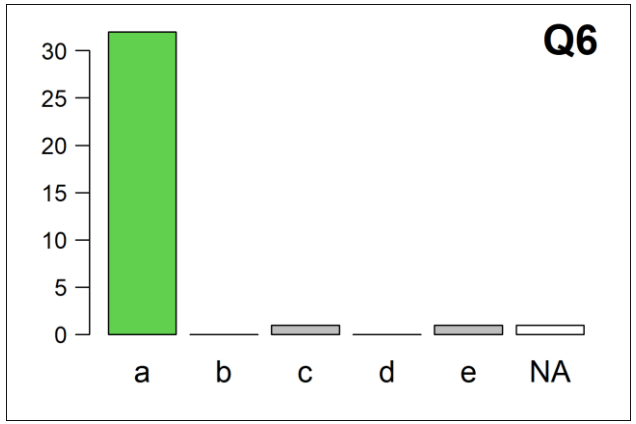
Emma and Carlos are about to get married, but suspect that they are related to each other. They consult a geneticist, who types them with 15 standard STR markers. The resulting genotypes are given in the file *cousins-data.txt*, along with the physical location of each marker. Allele frequencies can be found in *db.txt*.

Note: Recall that linkage is to be ignored in LR calculations. The locations are only used in Exercise 8.

Marker	Chr	Mb	Emma	Carlos
D1S1656	1	230.905	14/17.3	11/12
D2S441	2	68.239	12/14	11/11
D2S1338	2	218.879	20/24	17/17
D3S1358	3	45.582	14/16	15/18
FGA	4	155.509	21/26	22/23
SE33	6	88.987	11.2/20.2	15/21
D7S820	7	83.789	8/13	10/11
TH01	11	2.192	7/9.3	7/9.3
vWA	12	6.093	17/18	16/16
D13S317	13	82.692	11/11	8/11
PentaE	15	97.374	7/12	10/14
D16S539	16	86.386	11/12	9/12
D18S51	18	60.949	16/18	13/16
D19S433	19	30.416	12/14	14/15.2
D22S1045	22	37.536	15/18	12/16

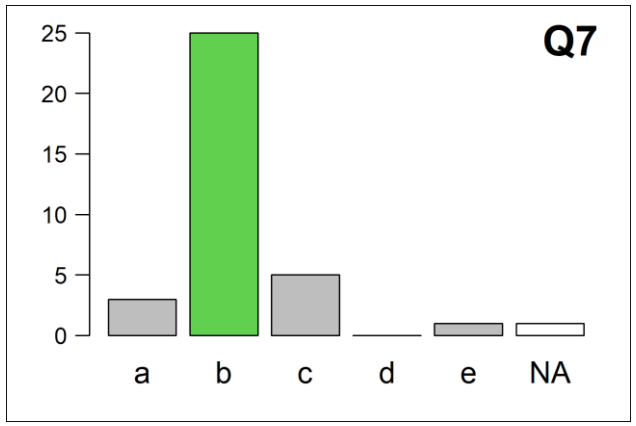


Emma and Charles Darwin (1840)



6. Use all 15 markers to compute the LR comparing the hypothesis that Emma and Carlos are first cousins, to the unrelated alternative. The total LR, rounded to two decimals, is

a) 0.10 ←
b) 0.75
c) 1.00
d) 3.14
e) 13.14

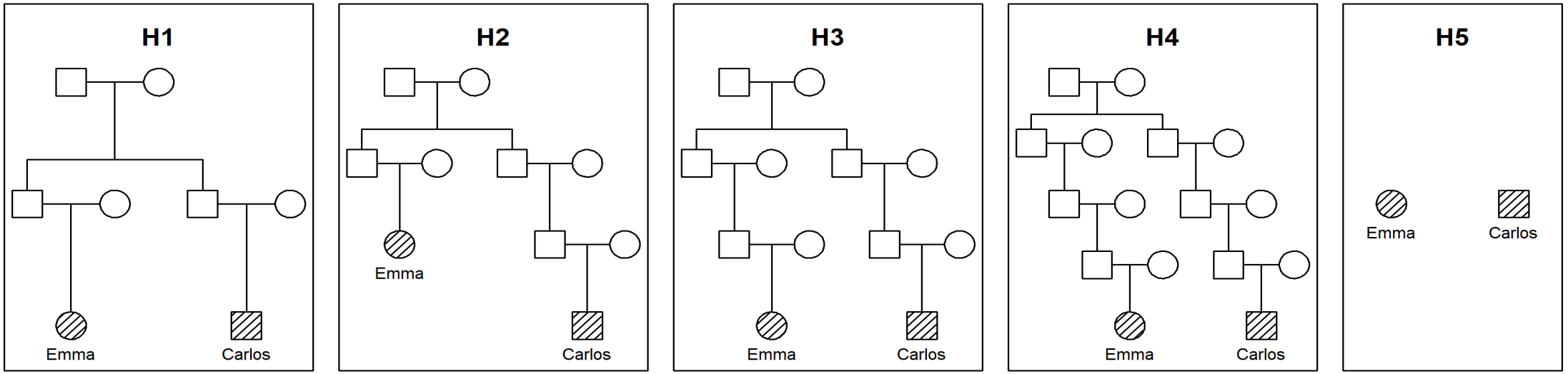


7. The LR comparing the most likely hypothesis with the second most likely, is approximately

a) 1.00
b) 1.12 ←
c) 1.45
d) 3.61
e) 10.18

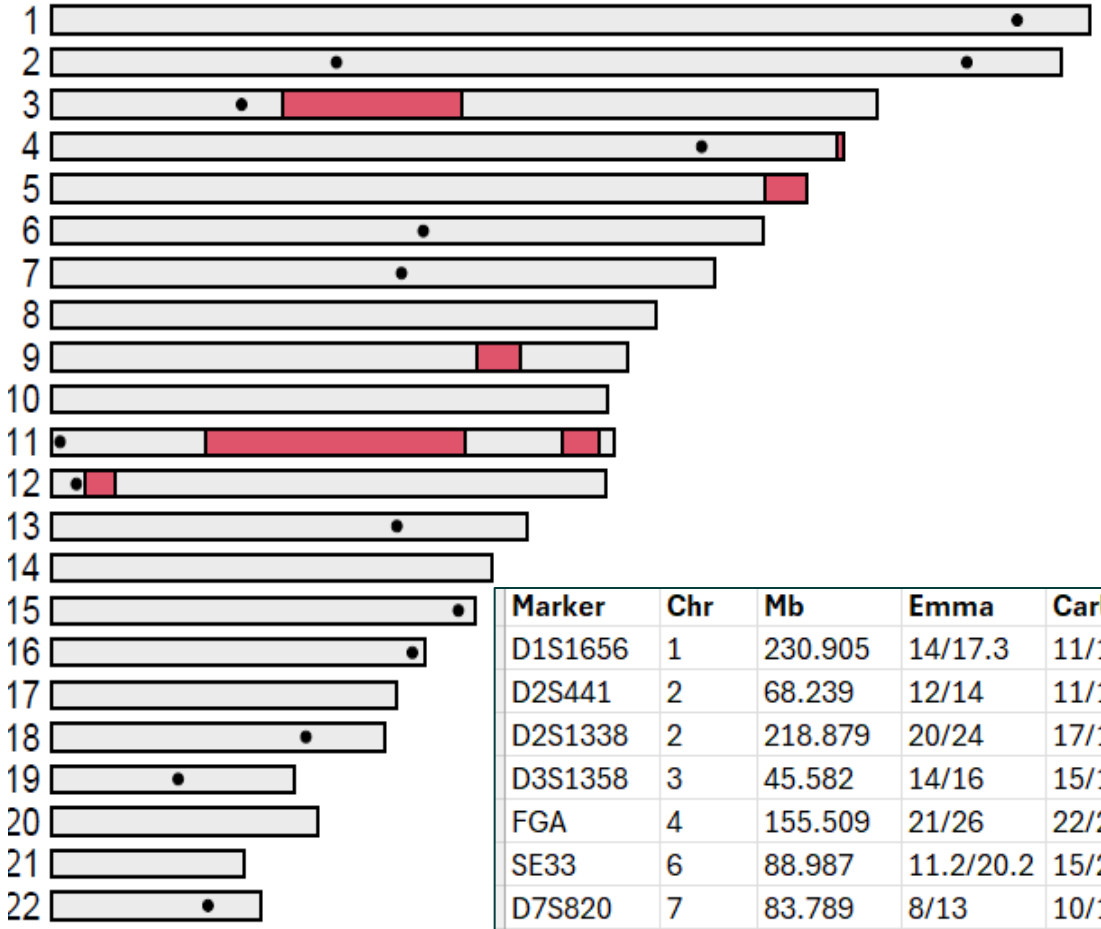
$H5 : H4 = 1/0.8889 \approx 1.12$

```
##      H1:H5     H2:H5     H3:H5     H4:H5
## 0.0982023 0.3544135 0.6121209 0.8889345
```



Shared IBD segments

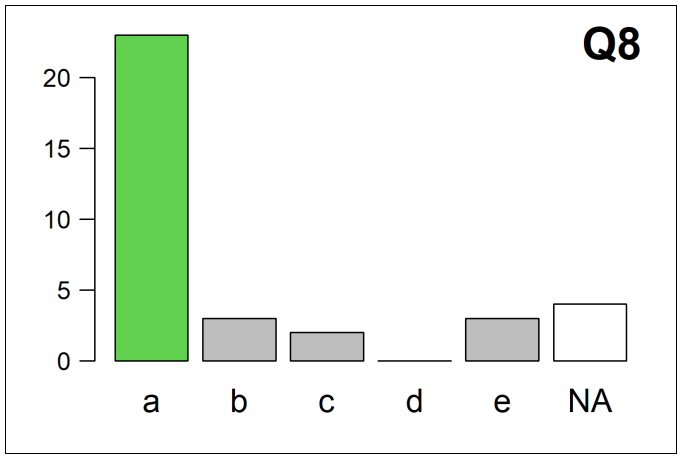
Chr	startMB	endMB	startCM	endCM
3	55.28	98.00	71.71	108.40
4	187.98	189.44	199.80	202.89
5	170.68	180.75	174.66	197.08
9	101.98	112.07	101.00	112.06
11	37.09	98.83	55.18	99.72
11	122.33	131.01	127.17	143.25
12	8.23	15.26	20.31	32.13



Marker	Chr	Mb	Emma	Carlos
D1S1656	1	230.905	14/17.3	11/12
D2S441	2	68.239	12/14	11/11
D2S1338	2	218.879	20/24	17/17
D3S1358	3	45.582	14/16	15/18
FGA	4	155.509	21/26	22/23
SE33	6	88.987	11.2/20.2	15/21
D7S820	7	83.789	8/13	10/11
TH01	11	2.192	7/9.3	7/9.3
vWA	12	6.093	17/18	16/16
D13S317	13	82.692	11/11	8/11
PentaE	15	97.374	7/12	10/14
D16S539	16	86.386	11/12	9/12
D18S51	18	60.949	16/18	13/16
D19S433	19	30.416	12/14	14/15.2
D22S1045	22	37.536	15/18	12/16

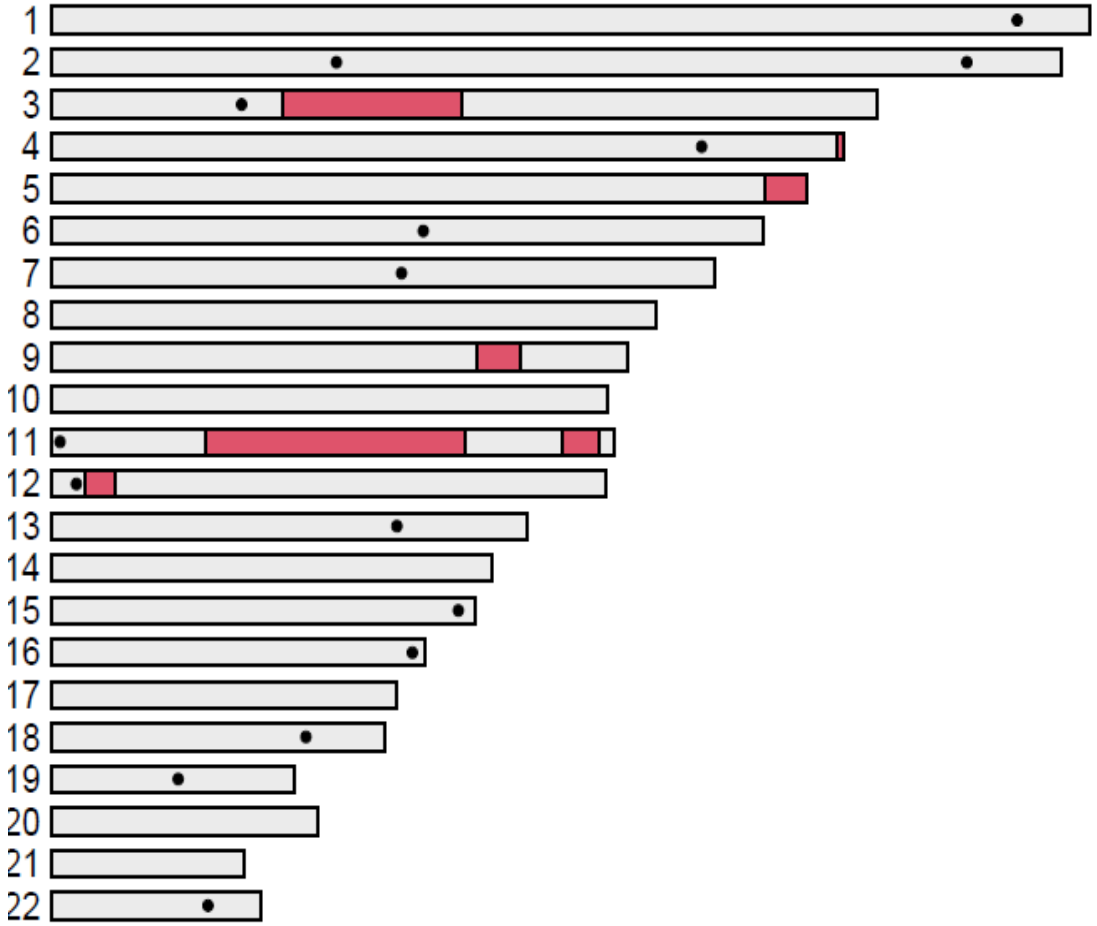
8. Of the 15 STR markers, the number that lie in an IBD region is

a) 0
b) 1
c) 2
d) 3
e) 4



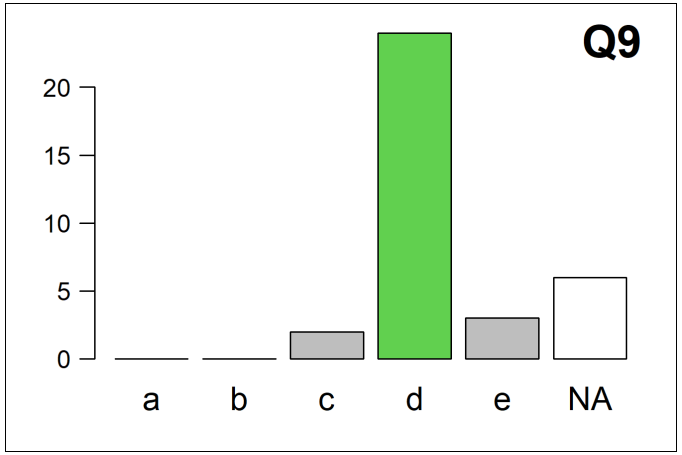
Shared IBD segments

Chr	startMB	endMB	startCM	endCM
3	55.28	98.00	71.71	108.40
4	187.98	189.44	199.80	202.89
5	170.68	180.75	174.66	197.08
9	101.98	112.07	101.00	112.06
11	37.09	98.83	55.18	99.72
11	122.33	131.01	127.17	143.25
12	8.23	15.26	20.31	32.13



9. The observed proportion k_1 of the autosome with IBD status 1, is approximately

a) 1.1%
b) 1.8%
c) 2.7%
d) 4.3%
e) 6.3%



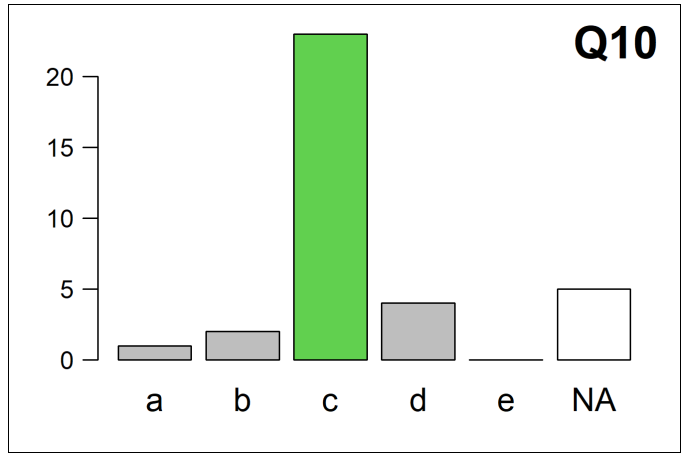
Total length = 145.7 cM

$$k_1 = \frac{145.7}{3391} \approx 0.043$$

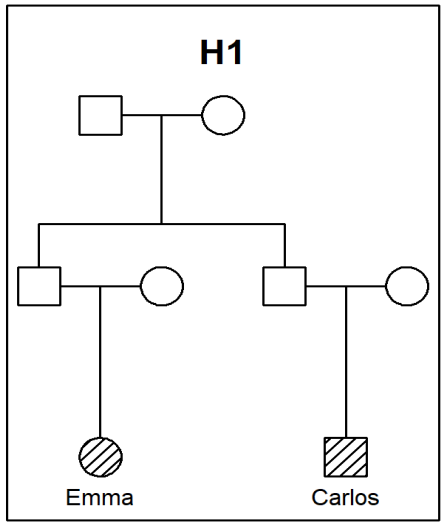
10. Of the following relationships, the one whose κ_1 is closest to the observed k_1 for Emma and Carlos, is

a) first cousins
b) first cousins once removed
c) second cousins
d) third cousins
e) unrelated

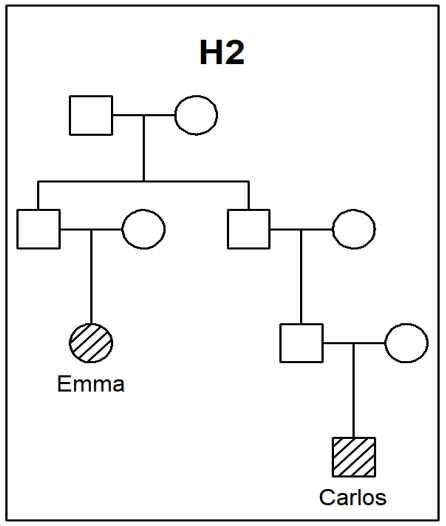
$$k_1 = \frac{145.7}{3391} \approx 0.043$$



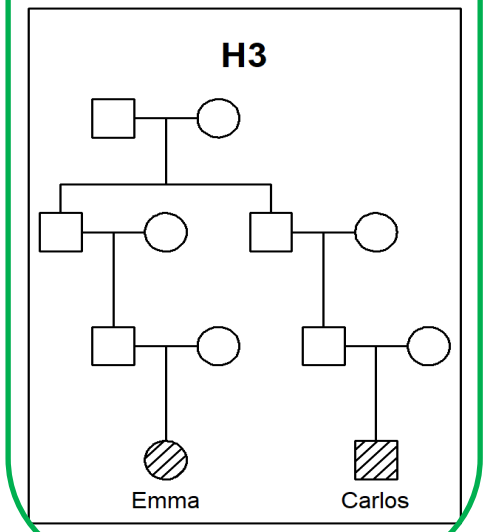
$\kappa_1 = \frac{1}{4} = 0.25$



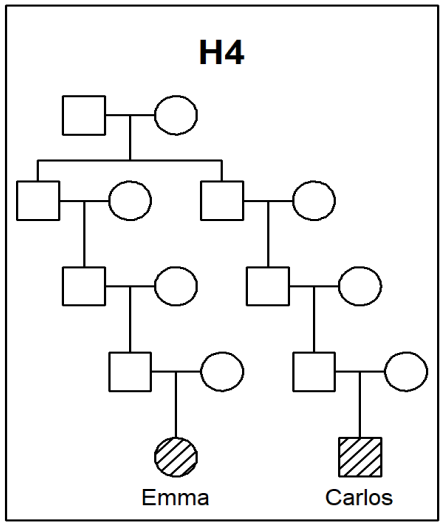
$\kappa_1 = \frac{1}{8} = 0.125$



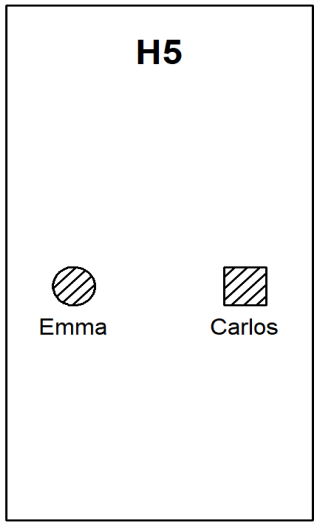
$\kappa_1 = \frac{1}{16} = 0.0625$



$\kappa_1 = \frac{1}{64} = 0.0156$



$\kappa_1 = 0$



Kappas can be computed with QuickPed: https://magnusdv.shinyapps.io/quickped/

11. According to the DNA painter tool, *https://dnapainter.com/tools/sharedcmv2*, the IBD sharing between Emma and Carlos is *not compatible* with (i.e., outside the reported range of)

a) first cousins

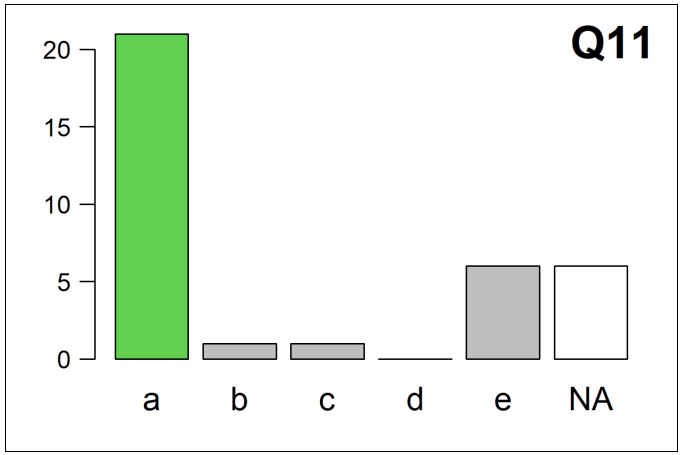
b) first cousins once removed

c) second cousins

d) third cousins

e) several of the above

Total observed IBD = 145.7 cM



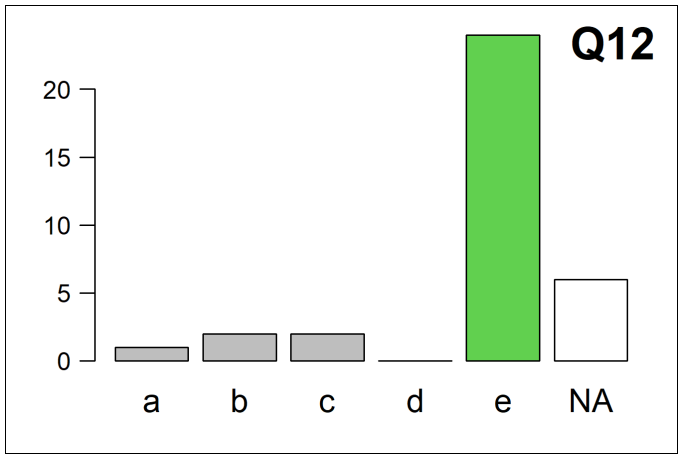
Grandparent 1754 984 – 2462					**Great-Aunt / Uncle** 850 330 – 1467	**1C2R** 221 33 – 471
Half Aunt / Uncle 871 492 – 1315	**Parent** 3485 2376 – 3720			**Aunt / Uncle** 1741 1201 – 2282	**1C1R** 433 102 – 980	**2C1R** 122 14 – 353
Half 1C 449 156 – 979	**Half Sibling** 1759 1160 – 2436	**Sibling** 2613 1613 – 3488	**SELF**	**1C** 866 396 – 1397	**2C** 229 41 – 592	**3C** 73 0 – 234

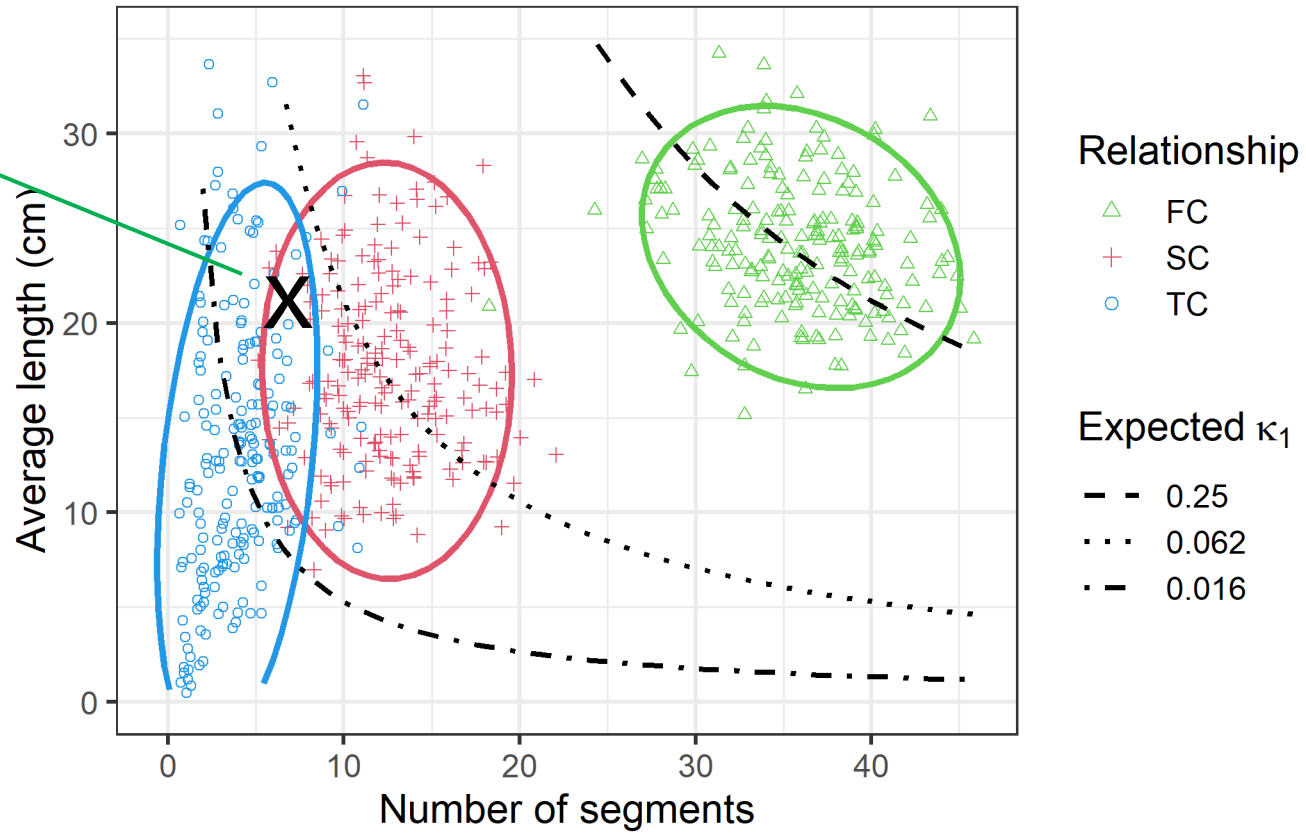
12. Figure 4 shows the distribution of IBD segments in 200 simulations of first, second and third cousins. Based on this plot, the observed data is only compatible (in the sense of being inside the 95% data ellipse) with

a) first cousins (FC)
b) second cousins (SC)
c) third cousins (TC)
d) first and second cousins
e) second and third cousins

Emma & Carlos

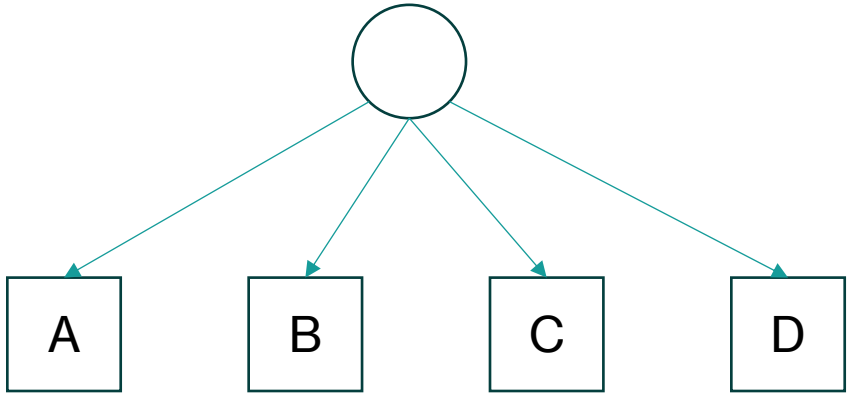
- Number of segments: 7
- Average length: 147.5/7 = 20.8





Part III: A case of sibship

This case involves 4 male individuals, labelled A, B, C and D. It is believed that all four have the same mother, but the paternities are unclear. Genotypes for A, B, C, D at 23 forensic markers can be found in *sibship-data.txt*, with allele frequencies in *db.txt* as before.



	D1S1656	TPOX	D2S441	D2S1338	D3S1358	FGA	D5S818	CSF1PO	SE33	D7S820	D8S1179	D10S1248	TH01	vWA	D1
A	15/16	8/11	10/11	17/19	15/18	22/23	11/13	10/10	19.2/21	10/11	12/13	12/14	9/9.3	16/18	17
B	15/17.3	8/11	10/11	17/19	15/17	20/23	11/13	10/11	19.2/30.2	10/11	10/12	13/14	6/6	16/18	17
C	15/16	8/11	11.3/12	23/25	14/18	22/23	11/11	10/11	19.2/21	10/11	10/12	13/14	6/9	16/18	17
D	16/17.3	8/8	10/11	17/23	14/18	22/23	11/11	11/12	19.2/21	10/10	10/12	13/14	6/9	16/18	17

13. The LR comparing A and B being full siblings versus half siblings, is approximately

a) 0.68 ←

b) 1.00

c) 431.47

d) 133506.3

e) None of the above

14. The LR comparing C and D being full siblings versus half siblings, is approximately

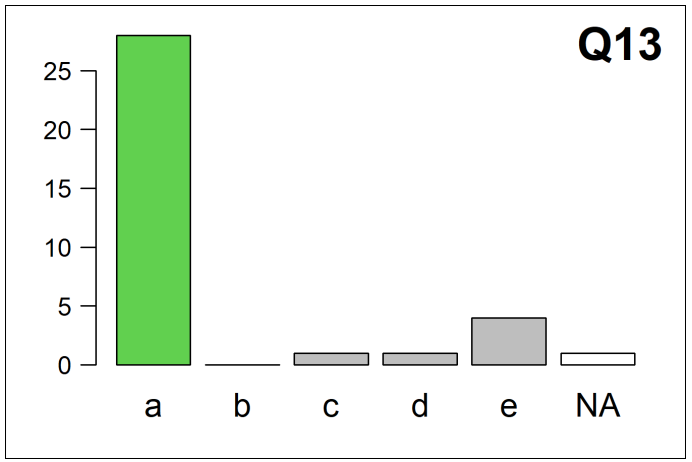
a) 0.68

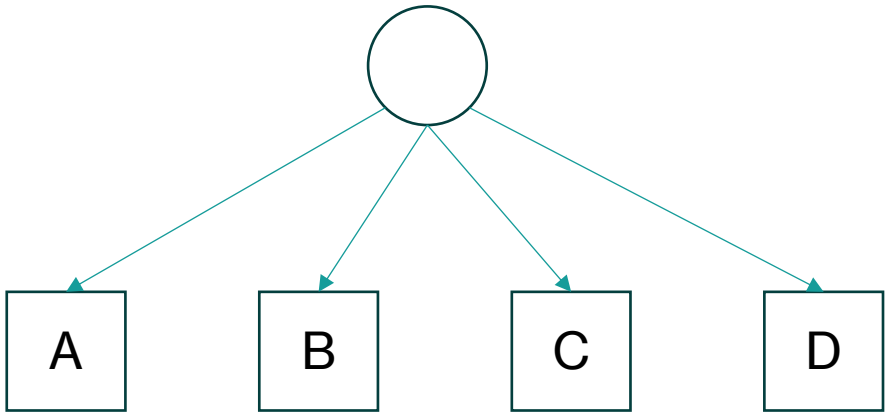
b) 1.00

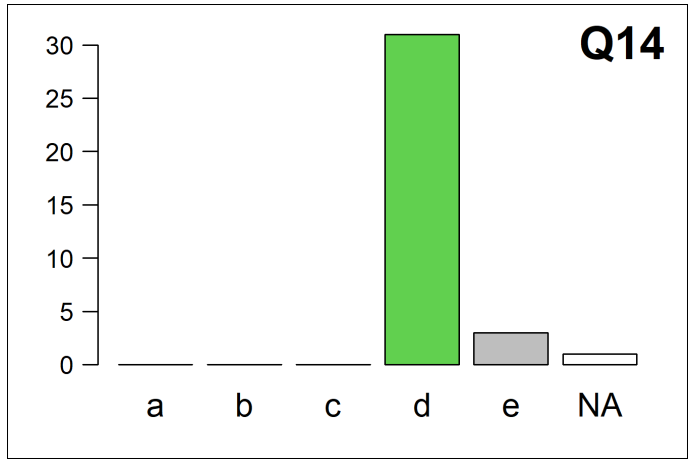
c) 431.47

d) 133506.3 ←

e) None of the above







We now make the following assumptions:

i) all four have the same mother

ii) C and D are full siblings

iii) each pair among A,B,C,D is either half or full siblings, with no further relationships or inbreeding in the pedigree

15. The number of possible hypotheses (pedigrees) connecting A,B,C,D is

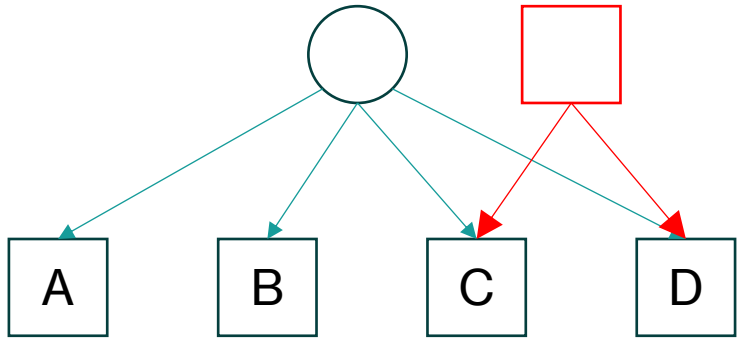
a) 5

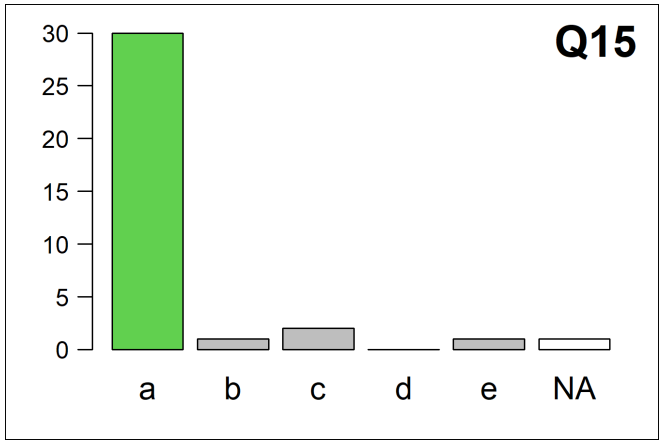
b) 6

c) 7

d) 8

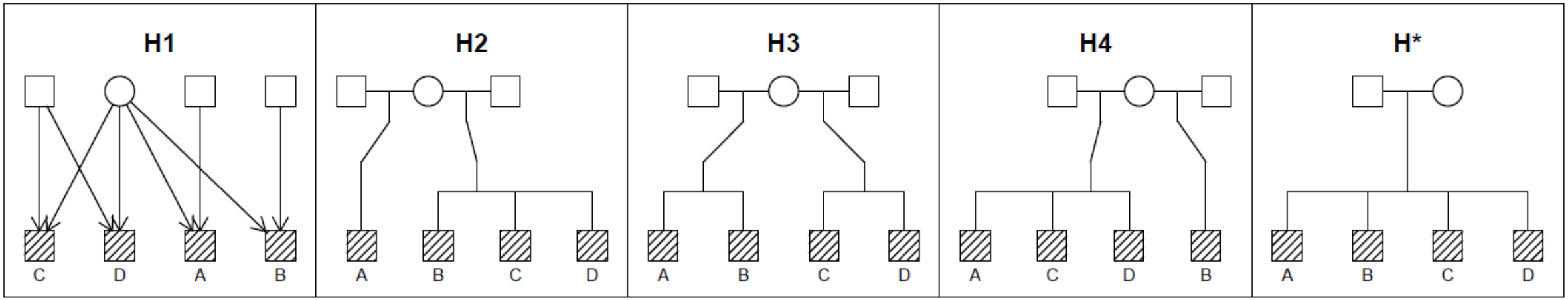
e) 9





Father of C and D can be:

- father of A and B
- father of A, not B
- father of B, not A
- not father of A or B (half sibs)
- not father of A or B (full sibs)



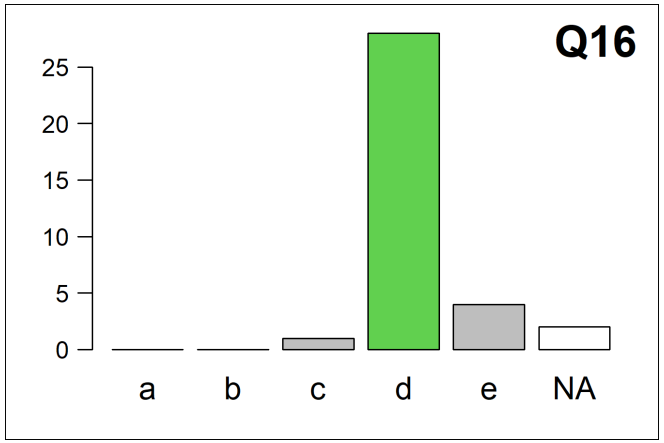
16. Let H^* denote the hypothesis that all four are full siblings. Assuming a flat prior on set of hypotheses from the previous question, the posterior probability of H^* given the marker data, is approximately

a) 0.00
b) 0.01
c) 0.73
d) 0.99
e) 1.00

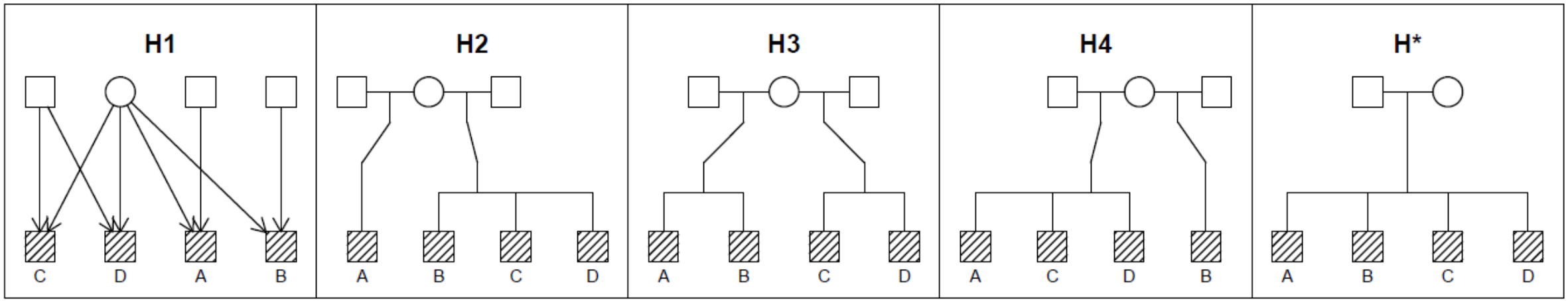
$$P(H^*|E) = \frac{P(E|H^*)P(H^*)}{P(E|H_1)P(H_1) + \ ... + P(E|H^*)P(H^*)}$$

$$= \frac{P(E|H^*)}{P(E|H_1) + \ ... + P(E|H^*)}$$

$$\approx \frac{1}{0.014 + 1} \approx 0.986$$



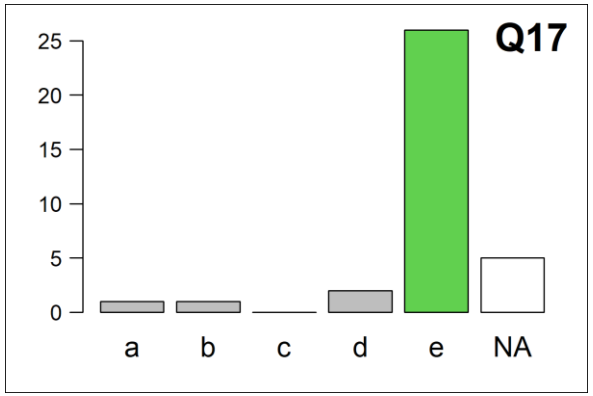

```
##         H1:H*         H2:H*         H3:H*         H4:H*
## 7.119501e-09 6.343537e-07 4.569093e-09 1.400989e-02
```



17. In terms of genetic length, the proportion k_1 of the autosome with IBD status 1, is

a) 0.25
b) 0.30
c) 0.45
d) 0.50
e) 0.60 ←

Solution method
- Segment lengths: **endCM** - **startCM**
- Add lengths of all with IBD = 1
- Divide by 3391 cM → 0.596



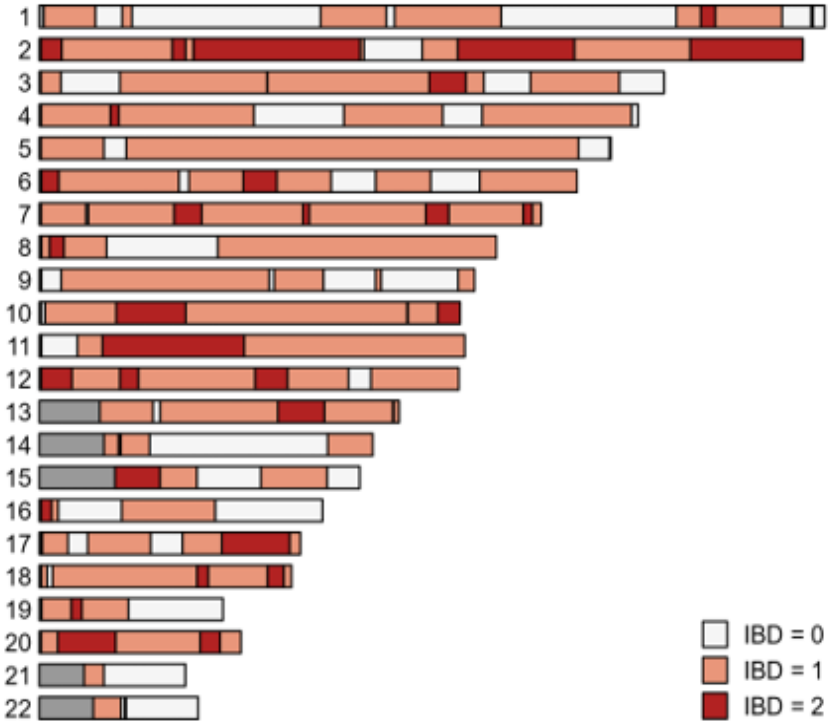


Figure 5. Segments of identity by descent between alleged siblings A and B.

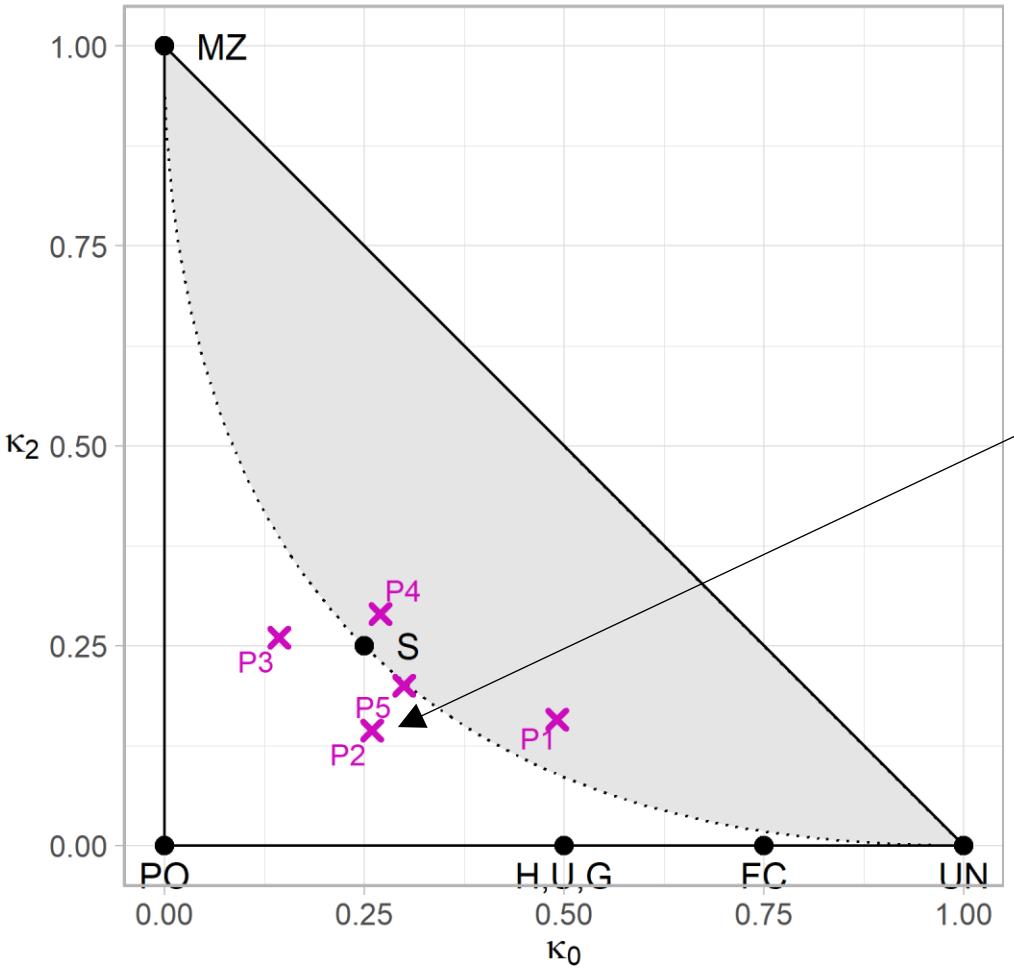
sibs-ibd.txt

IBD segments shared by A and B

Chr	startMB	endMB	startCM	endCM	IBD
1	1.43	17.79	0	34	1
1	26.2	29.32	48.39	50.16	1
1	88.99	109.79	113.15	130.42	1
1	112.39	146.14	134.7	142.82	1
1	201.55	209.27	195.64	208.73	1
1	209.27	213.85	208.73	213.88	2
1	213.85	234.96	213.88	238.86	1
1	244.2	244.81	259.75	261.92	1
2	0.52	7.11	0	14.5	2
2	7.11	42.06	14.5	61.93	1
2	42.06	46.26	61.93	67.9	2
2	46.26	48.78	67.9	71	1
2	48.78	101.48	71	109.73	2
2	101.48	102.69	109.73	110.65	1
2	121.17	132.39	127.82	137.78	1
2	132.39	169.29	137.78	167.39	2
2	169.29	205.94	167.39	197.18	1
2	205.94	241.54	197.18	251.73	2
3	0.52	6.86	0	17.97	1

18. In the IBD triangle in Figure 6, the realised relationship between A and B corresponds to the point

a) P1
b) P2
c) P3
d) P4
e) P5



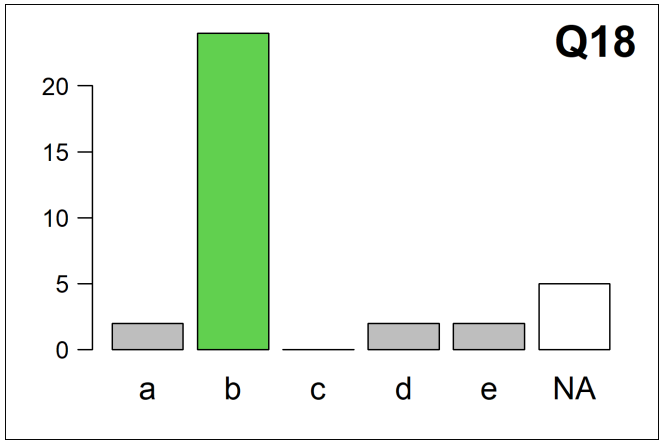
As in previous exercise:

$$k_2 = \frac{\text{total length with } IBD = 2}{3391} = \mathbf{0.14}$$

$$k_0 = 1 - k_1 - k_2 = \mathbf{0.26}$$

Hence the corresponding point is

$$(k_0, k_2) = (0.14, 026)$$



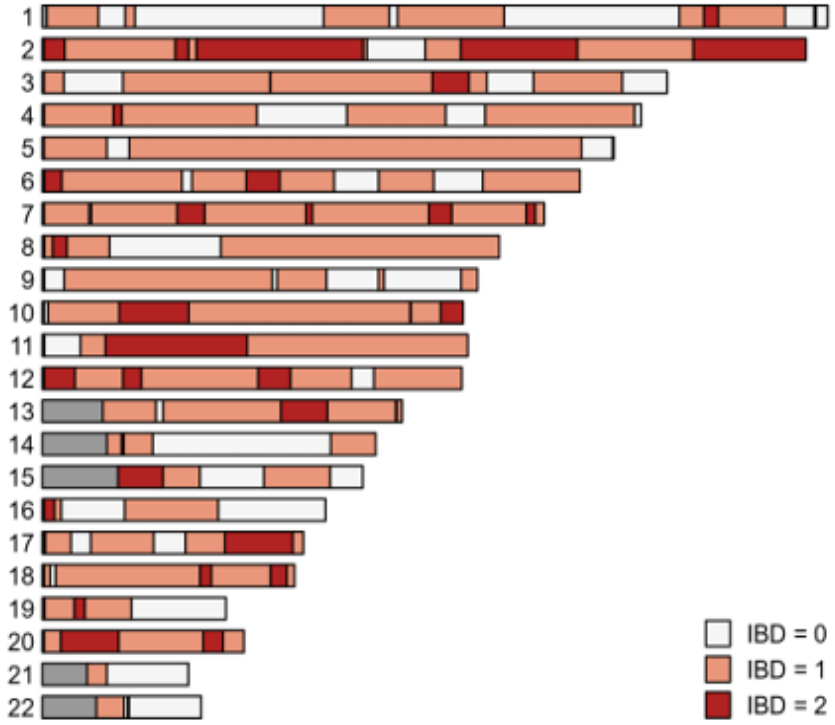


Figure 5. Segments of identity by descent between alleged siblings A and B.

19. Based on the IBD segments, the realised kinship coefficient φ_R between A and B is approximately

a) 0.14

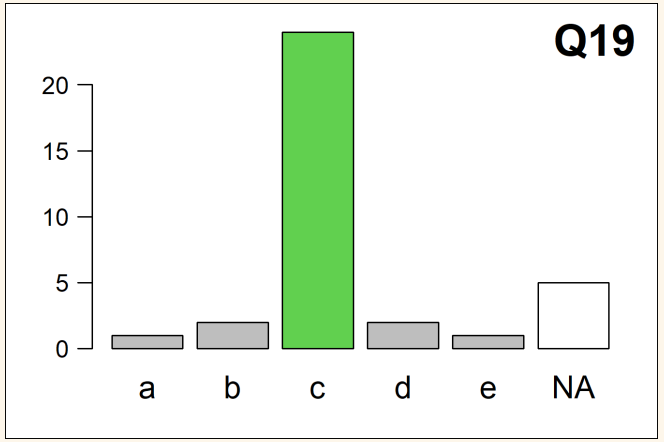
b) 0.19

c) 0.22

d) 0.25

e) 0.26

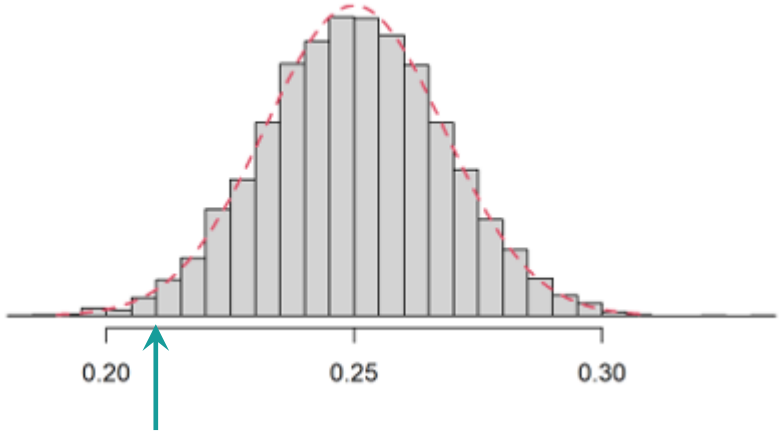
$$\varphi_R = \frac{k_1}{4} + \frac{k_2}{2} = \frac{0.60}{4} + \frac{0.14}{2} = 0.22$$



The histogram in Figure 7 shows the realised kinship coefficient in 5000 simulated pairs of full siblings, closely approximated by a normal distribution with mean $\mu = 0.25$ and standard deviation $\sigma = 0.018$ (dashed red curve).

20. Compared with the normal approximation for full siblings, the observed φ_R falls at the

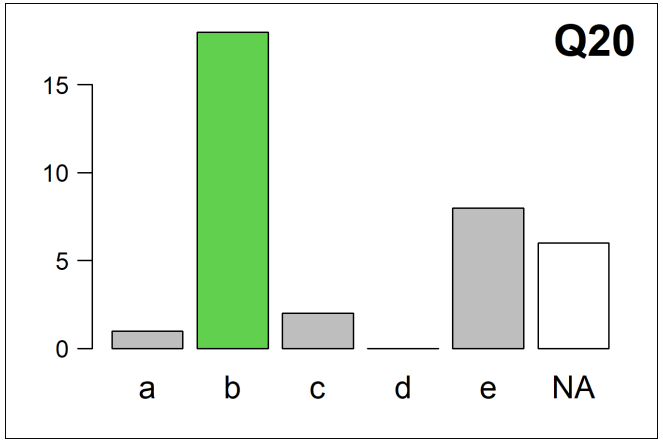
a) 0th percentile
b) 5th percentile
c) 10th percentile
d) 15th percentile
e) 20th percentile



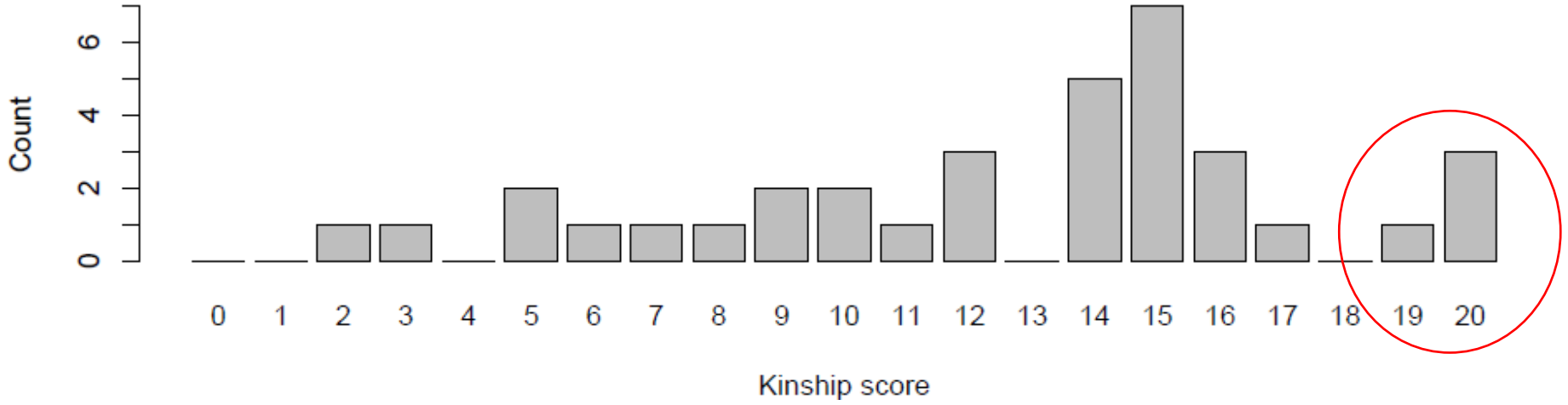
$\varphi_R = 0.22$

Calculation in R

```
> pnorm(phi_R, mean = 0.25, sd = 0.018)
# 0.05367422
```



Statistics



4 labs with perfect or nearly perfect score