

GHEP 2024 desafio teórico de parentesco: nível avançado

Preparado por: Magnus Dehli Vigeland

Instruções gerais

Este exercício é de escolha múltipla com 20 perguntas. Para cada pergunta, existe apenas uma alternativa correta. Pode utilizar qualquer software que prefira, mas tenha em consideração que alguns programas têm convenções integradas (por exemplo, arredondamento) que podem afetar o resultado obtido. Se a sua resposta não coincide totalmente com nenhuma das alternativas, deverá escolher a opção que mais se aproxima.

Ficheiros necessários para realizar o exercício

- *cousins-data.txt* / *cousins-ibd.txt*. Dados para a Parte II.
- *siblings-data.txt* / *siblings-ibd.txt*. Dados para a Parte III.
- *db.txt*. Frequências alélicas para 23 marcadores STR. (Se o seu software requer o tamanho da base de dados, utilize $N = 1000$).

Simplificações assumidas ao longo do exercício

- Ausência de *linkage* entre os marcadores, ausência de *linkage disequilibrium*, ausência de desvios ao equilíbrio de Hardy-Weinberg.
- Ausência de *drop-out*, *drop-in*, alelos silenciosos ou mutações.
- Os fundadores dos *pedigrees* não são consanguíneos e não são parentes entre si.
- O comprimento genético total do genoma autossômico (cromossomas 1-22) é de 3391 cM.

Algumas definições

Alelos homólogos são idênticos por descendência (IBD) se tiverem a mesma origem dentro de um determinado *pedigree*. Os coeficientes IBD ($\kappa_0, \kappa_1, \kappa_2$) de indivíduos A e B não consanguíneos, correspondem às probabilidades de partilharem, respetivamente, 0, 1 e 2 alelos IBD num locus autossômico aleatório. Estão relacionados com o coeficiente de parentesco φ pela fórmula $\varphi = \kappa_1/4 + \kappa_2/2$.

O triângulo IBD (Figura 1) é uma forma útil para visualizar coeficientes IBD. Note-se que, como $\kappa_0 + \kappa_1 + \kappa_2 = 1$, basta conhecer dois dos coeficientes para deduzir o terceiro; a escolha de κ_0 e κ_2 é, simplesmente, a minha preferida. A ferramenta online [QuickPed](#) pode ser útil para calcular coeficientes IBD e representá-los no triângulo IBD.

Coefficientes tradicionais como κ e φ medem a partilha IBD **esperada** com base no *pedigree*. Em contraste, o **parentesco real** entre A e B refere-se aos segmentos de IBD que eles realmente compartilham como resultado da recombinação (Figura 2). Denominamos por (k_0, k_1, k_2) as proporções reais do genoma autossômico, em termos de comprimento genético, onde eles compartilham 0, 1 e 2 alelos IBD, respectivamente. O *coeficiente de parentesco real* é dado por $\varphi_R = k_1/4 + k_2/2$.

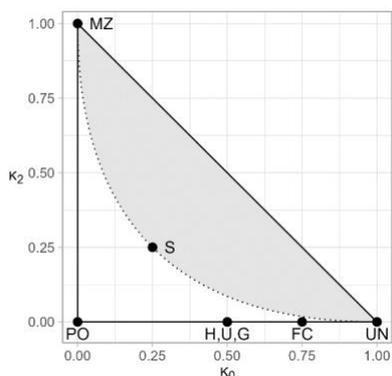


Figura 1. Triângulo de IBD. FC=primos em primeiro grau; G=avô-neto; H=meios-irmãos; MZ=gêmeos monozigóticos; PO=progenitor-filho; S=irmãos inteiros; U=tio-sobrinho; UN=não aparentados

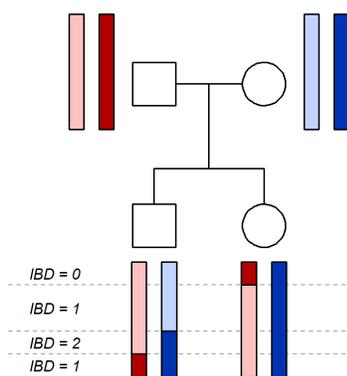


Figura 2. Um exemplo de partilha IBD real entre irmãos. O cromossoma divide-se em segmentos com estado IBD 0, 1 ou 2

Questões

Parte I: O aquecimento

Considere a situação em que dois indivíduos, A e B, são genotipados para um marcador tri-alelético. Os alelos são designados como 1, 2, 3, e as frequências alélicas são p_1, p_2, p_3 , respetivamente.

- Suponha que o marcador se localiza numa região onde A e B apresentam estado IBD de 0. Se A tem o genótipo 1/1, o genótipo de B é
 - 2/3
 - 2/2 ou 3/3
 - 2/2, 2/3 ou 3/3
 - qualquer genótipo exceto 1/1
 - qualquer genótipo
- Suponha que o marcador se localiza numa região com IBD = 1. Se A tem o genótipo 1/2, o genótipo de B é
 - 1/3 ou 2/3
 - 1/1, 2/2, 1/3 ou 2/3
 - qualquer genótipo exceto 1/2
 - qualquer genótipo exceto 3/3
 - qualquer genótipo

Nas próximas três perguntas, assume-se que os genótipos de ambos os indivíduos são desconhecidos.

- Sabendo que o marcador se encontra numa região com IBD = 2, a probabilidade de A e B serem homocigóticos para o mesmo alelo é
 - 0
 - $p_1^2 + p_2^2 + p_3^2$
 - $p_1(1 - p_1) + p_2(1 - p_2) + p_3(1 - p_3)$
 - $p_1(1 - p_1)^2 + p_2(1 - p_2)^2 + p_3(1 - p_3)^2$
 - 1
 - Sabendo que o marcador se encontra numa região com IBD = 1, a probabilidade de serem completamente coincidentes (*full match*, ou seja, A e B têm o mesmo genótipo) é
 - 0
 - $p_1^2 + p_2^2 + p_3^2$
 - $p_1(1 - p_1) + p_2(1 - p_2) + p_3(1 - p_3)$
 - $p_1(1 - p_1)^2 + p_2(1 - p_2)^2 + p_3(1 - p_3)^2$
 - 1
 - Sabendo que o marcador se encontra numa região com IBD = 1, a probabilidade de serem parcialmente coincidentes (*partial match*, ou seja, partilham um só alelo) é
 - 0
 - $p_1^2 + p_2^2 + p_3^2$
 - $p_1(1 - p_1) + p_2(1 - p_2) + p_3(1 - p_3)$
 - $p_1(1 - p_1)^2 + p_2(1 - p_2)^2 + p_3(1 - p_3)^2$
 - 1
-

Parte II: Um caso de primos

Emma e Carlos estão prestes a casar-se, mas suspeitam que sejam aparentados. Consultam um geneticista que os genotipa para 15 marcadores STR convencionais. Os genótipos resultantes encontram-se no ficheiro *cousins-data.txt*, o qual também contém a localização física de cada marcador. As frequências alélicas encontram-se no ficheiro *db.txt*.

Nota: lembre-se que deve ignorar linkage nos cálculos de LR. As localizações são usadas apenas no Exercício 8.

6. Utilize os resultados dos 15 marcadores para calcular o LR que compara a hipótese de Emma e Carlos serem primos em primeiro grau (*first cousins*) com a hipótese alternativa de não serem aparentados. O LR total, arredondado a duas decimais, é
- 0,10
 - 0,75
 - 1,00
 - 3,14
 - 13,14

Numa abordagem mais exaustiva, o geneticista compara as 5 hipóteses seguintes:

- H1: primos em primeiro grau (*first cousins*)
- H2: meios primos em primeiro grau (*first cousins once removed*)
- H3: primos em segundo grau (*second cousins*)
- H4: primos em terceiro grau (*third cousins*)
- H5: não aparentados (*unrelated*)

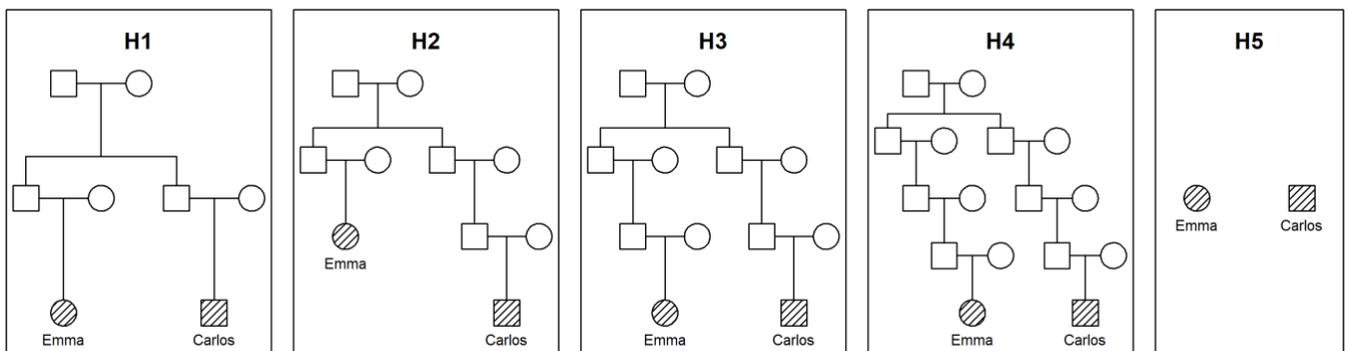


Figura 3. As hipóteses para a Questão 7

7. O LR que compara a hipótese mais provável com a segunda mais provável é, aproximadamente
- 1,00
 - 1,12
 - 1,45
 - 3,61
 - 10,18

Não estando completamente satisfeitos com a análise de STRs, Emma e Carlos recorrem à sequenciação completa do genoma, e o geneticista usa estes dados para identificar regiões de IBD compartilhadas entre eles. O resultado encontra-se no ficheiro *cousins-ibd.txt*, reproduzido na íntegra a seguir. Em todos os segmentos, Emma e Carlos compartilham um haplótipo, ou seja, $IBD = 1$

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

8. Dos 15 marcadores STR, o número que se encontra numa região IBD é
- 0
 - 1
 - 2
 - 3
 - 4
9. A proporção observada k_1 de autossomas com IBD=1 é, aproximadamente
- 1,1%
 - 1,8%
 - 2,7%
 - 4,3%
 - 6,3%
10. Das relações seguintes, aquela cujo κ_1 está mais próximo do k_1 observado para Emma e Carlos, é
- primos em primeiro grau (*first cousins*, ver Figura 3, H1)
 - meios primos em primeiro grau (*first cousins once removed*, ver Figura 3, H2)
 - primos em segundo grau (*second cousins*, ver Figura 3, H3)
 - primos em terceiro grau (*third cousins*, ver Figura 3, H4)
 - não aparentados (*unrelated*, ver Figura 3, H5)
11. De acordo com a ferramenta *DNA painter*, <https://dnainter.com/tools/sharedcmv2>, a partilha de IBD entre Emma e Carlos não é compatível com (ou seja, está fora do intervalo reportado para)
- primos em primeiro grau (*first cousins*, ver Figura 3, H1)
 - meios primos em primeiro grau (*first cousins once removed*, ver Figura 3, H2)
 - primos em segundo grau (*second cousins*, ver Figura 3, H3)
 - primos em terceiro grau (*third cousins*, ver Figura 3, H4)
 - vários dos anteriores
12. A Figura 4 mostra a distribuição de segmentos de IBD em 200 simulações para primos em primeiro, segundo e terceiro graus. Com base neste gráfico, os dados observados são apenas compatíveis (no sentido de estarem dentro da elipse de 95% de dados) com
- primos em primeiro grau (*first cousins*, FC)
 - primos em segundo grau (*second cousins*, SC)
 - primos em terceiro grau (*third cousins*, TC)
 - primos em primeiro e em segundo grau (*first and second cousins*)
 - primos em segundo e em terceiro grau (*second and third cousins*)

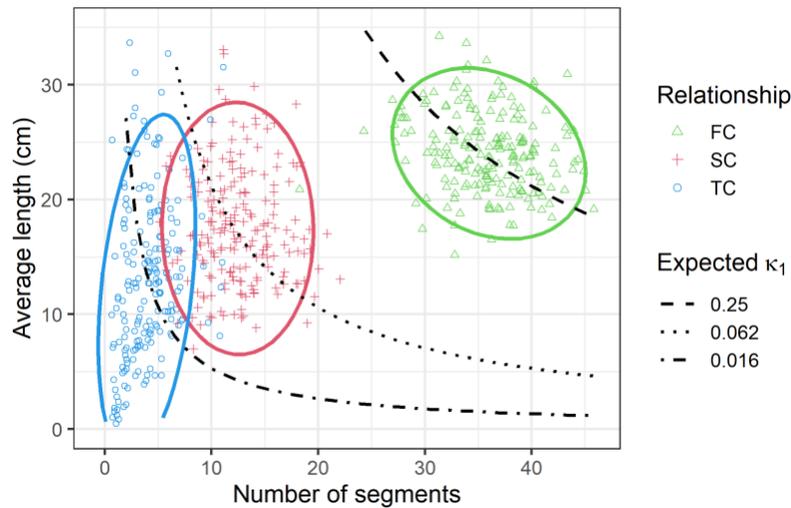


Figura 4. Distribuição IBD para primos em primeiro, segundo e terceiro graus.

Parte III: Um caso de irmandade

Este caso envolve 4 indivíduos masculinos, designados por A, B, C e D. Acredita-se que os quatro tenham a mesma mãe, mas as paternidades são pouco claras. Os genótipos de A, B, C, e D em 23 marcadores forenses encontram-se no ficheiro *sibship-data.txt*, e as frequências alélicas em *db.txt* como anteriormente.

13. O LR que compara A e B como irmãos inteiros (germanos) versus meios-irmãos é, aproximadamente,
 - a) 0,68
 - b) 1,00
 - c) 431,47
 - d) 133506,3
 - e) Nenhuma das anteriores

14. O LR que compara C e D como irmãos inteiros versus meios-irmãos é, aproximadamente,
 - a) 0,68
 - b) 1,00
 - c) 431,47
 - d) 133506,3
 - e) Nenhuma das anteriores

Consideremos agora as seguintes suposições:

- i) os quatro têm a mesma mãe
 - ii) C e D são irmão inteiros
 - iii) todos os pares formados entre A,B,C,D correspondem a meios-irmãos ou irmãos inteiros, sem qualquer outra relação ou endogamia no *pedigree*
15. O número de hipóteses possíveis (*pedigrees*) que relacionam A,B,C,D é
 - a) 5
 - b) 6
 - c) 7
 - d) 8
 - e) 9

16. Consideremos que H^* corresponde à hipótese de que os quatro são irmãos inteiros. Assumindo um *a priori* equiprovável (*flat prior*) no conjunto de hipóteses da questão anterior, a probabilidade *a posteriori* de H^* , dados os resultados genotípicos, é aproximadamente

- a) 0,00
- b) 0,01
- c) 0,73
- d) 0,99
- e) 1,00

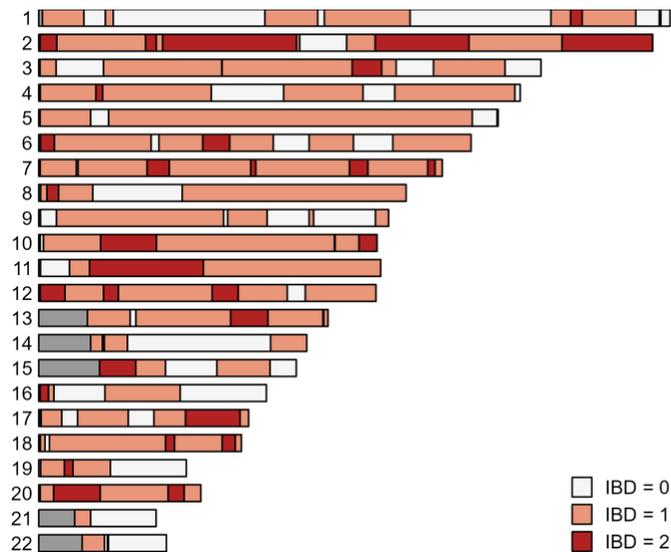


Figura 5. Segmentos de identidade por descendência entre os supostos irmãos A e B.

Para melhor investigar a relação entre A e B, é realizada a sequenciação completa do genoma, e os seus segmentos IBD são determinados conforme se mostra na Figura 5. As coordenadas encontram-se no ficheiro *sibs-ibd.txt*.

17. Em termos de comprimento genético, a proporção k_1 autossómica com IBD=1 é

- a) 0,25
- b) 0,30
- c) 0,45
- d) 0,50
- e) 0,60

18. No triângulo IBD da Figura 6, o parentesco real entre A e B corresponde ao ponto

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

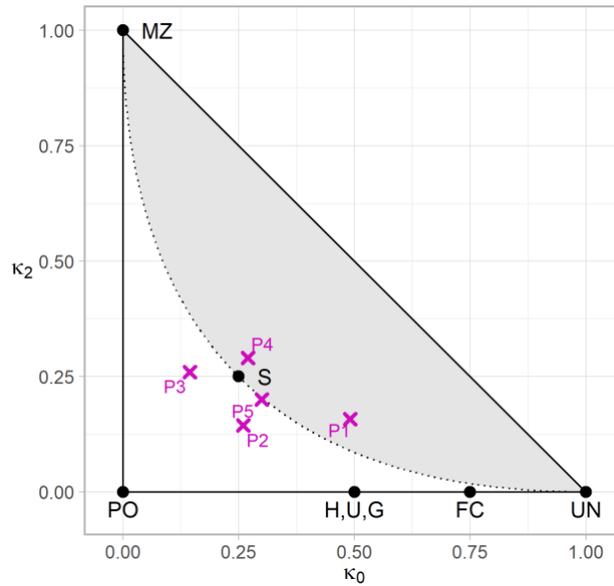


Figura 6. Triângulo IBD com as alternativas para o parentesco real entre A e B.

19. Com base nos segmentos IBD, o coeficiente de parentesco real φ_R entre A e B é, aproximadamente,
- 0,14
 - 0,19
 - 0,22
 - 0,25
 - 0,26

O histograma na Figura 7 mostra o coeficiente de parentesco real em 5000 pares simulados de irmãos inteiros, aproximado por uma distribuição normal com média $\mu = 0,25$ e desvio padrão $\sigma = 0,018$ (curva tracejada a vermelho).

20. Comparado com a aproximação normal para irmãos inteiros, o valor observado φ_R cai no
- 0º percentil
 - 5º percentil
 - 10º percentil
 - 15º percentil
 - 20º percentil

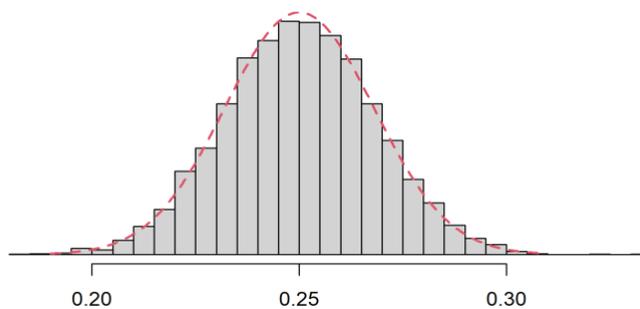


Figura 7. Simulações do parentesco real entre irmãos inteiros

Responsável pelo desenvolvimento do exercício:

Magnus Dehli Vigeland

magnusdv@gmail.com

Tradução:

Cíntia Alves (cintialves70@gmail.com)

Comité Executivo do GHEP-ISFG

info@ghép-isfg.org