

Exercício de parentesco GHEP 2025: Nível avançado

Desenvolvido 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 desejar, mas lembre-se de que alguns programas possuem 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 aproxime.

Suposições: Ao longo do exercício, assumimos a ausência de desvios do equilíbrio de Hardy-Weinberg, desequilíbrio de ligamento (*linkage disequilibrium*), drop-outs, drop-ins, alelos silenciosos ou mutações, a menos que explicitamente indicado.

Introdução

O tema do desafio deste ano é *linkage* (ligamento) e efeito de marcadores ligados em testes de parentesco. Por conveniência, relembramos aqui alguns conceitos básicos; para mais detalhes, foram incluídas algumas referências no final.

A taxa de recombinação θ entre dois marcadores é a fração, a longo prazo, de gametas (produtos meióticos) cujos alelos, nos dois marcadores, têm diferentes origens parentais. Por exemplo, a Figura 1 mostra uma mãe com genótipos A/a e B/b em dois marcadores. Para simplificar, assumimos que a fase gamética é conhecida, conforme indicado pela linha vertical. Neste caso, há quatro haplótipos possíveis, conforme mostrado na parte inferior da Figura 1, junto com as correspondentes probabilidades de transmissão. Os dois haplótipos externos são não-recombinantes; os internos são recombinantes.

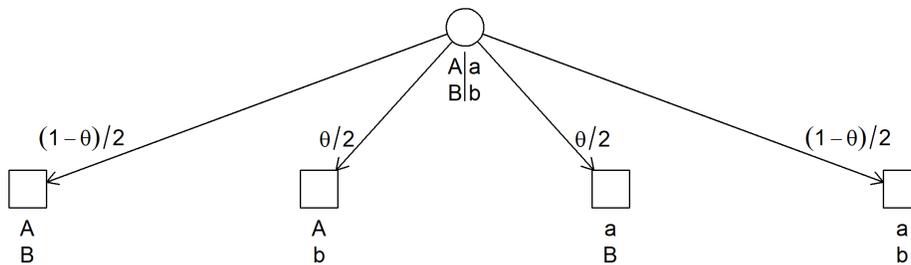


Figure 1. Ilustração das probabilidades de recombinação e de transmissão de haplótipos em termos de θ .

Para dois loci independentes, por ex., localizados em cromossomas diferentes, os quatro haplótipos representados na Figura 1 são igualmente prováveis, e $\theta = 0,5$. Tais loci são ditos não-ligados (*unlinked*). Pelo contrário, loci com $\theta < 0,5$ encontram-se ligados (*linked*).

A distância genética (*genetic map distance*) entre dois loci no mesmo cromossoma é definida como o número esperado de *crossovers* entre eles por meiose. A unidade básica é 1 Morgan, correspondendo a 1 *crossover* esperado por meiose. Uma unidade mais comum é 1 centiMorgan (cM) = 0,01 Morgan.

Existem várias aproximações para converter taxas de recombinação em distância genética, e vice-versa. A mais simples é a função de mapeamento de Haldane (*Haldane's map function*), que relaciona a taxa de recombinação θ entre dois loci, e sua distância genética d medida em M:

$$\theta = \frac{1}{2}(1 - e^{-2d}).$$

Software: Abaixo está uma lista (não exaustiva) de programas que contemplam marcadores ligados e que podem ser úteis para este desafio (Clique nos links para mais informações). Algumas instruções para o KLINK estão incluídas no final.

- **KLINK:** Aplicação desenvolvida recentemente com base no pedsuite, compatível com o software Familias.

- *FamLink2*: Software autónomo bem estabelecido para Windows. Aceita arquivos do Famílias como entrada.
- *pedsuite*: Conjunto de pacotes R para análise de pedigrees, incluindo uma ampla gama de funções para testes de parentesco.
- *MERLIN*: Software popular para análise de *linkage*, embora não tenha sido desenvolvido especificamente para testes de parentesco.

Parte I: Um caso delicado de herança

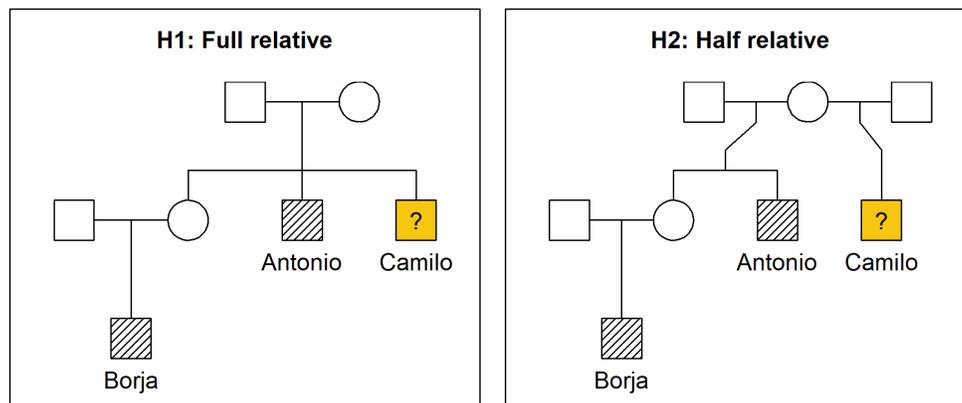
O detetive Vargas já tinha visto sua cota de disputas familiares, mas era facilmente reconhecível que esta levaria provavelmente a desafios ou complicações. Antonio e o seu sobrinho Borja, ambos com olhos azuis penetrantes, inclinaram-se sobre a sua mesa, argumentando agressivamente que Camilo, sentado calmamente ao lado deles, não tinha direito à herança. O pai de Antonio tinha sido um homem rico, e havia agora milhões em jogo.

"Ele não é meu irmão inteiro", disse Antonio. "Sempre suspeitei que a minha mãe tinha um caso com o jardineiro. Ou com aquele malabarista desprezível que ela costumava convidar para as festas."

"Basta olhar para os olhos castanhos dele!", acrescentou Borja. "E ele nem consegue enrolar a língua como nós!"

Ambos começaram a enrolar a língua freneticamente. Vargas fechou os olhos e recostou-se na cadeira. Truques mendelianos de festa não resolveriam o assunto — distinguir irmãos inteiros de meios-irmãos exigia uma abordagem mais científica. Os testes de parentesco habituais seriam suficientes? Com um suspiro, pegou no telefone e ligou para o melhor geneticista forense da cidade...

Após o telefonema do detetive Vargas, você esboça os pedigrees que representam as hipóteses alternativas deste caso:



Os seguintes arquivos estão disponíveis para este exercício:

- *vargasA.ped* e *vargasA.fam*: Dataset A no formato standard ped e no formato Famílias
- *vargasCombined.ped* e *vargasCombined.fam*: Datasets A e B combinados
- *db35.txt*: Base de dados de frequências para todos os marcadores usados neste caso
- *map35.txt*: Mapa com as posições cromossômicas e cM de todos os marcadores

Para um dado conjunto de dados envolvendo marcadores em *linkage*, introduzimos a seguinte notação:

- LR^l representa o LR considerando *linkage*, ou seja, onde o *linkage* é levada em consideração.
- LR^u representa o LR quando é ignorado o *linkage*, ou seja, obtido pela simples multiplicação de todos os LRs obtidos para cada marcador.

Todos os LRs referem-se às hipóteses H1 e H2 mostradas acima, com H1 como numerador e H2 como denominador.

Os três homens são inicialmente tipados com um kit padrão incluindo 19 STRs autossômicos. Este é o dataset A (Table 1).

1. O LR^u total para o dataset A é aproximadamente:
 - a) 0,71
 - b) 2,01
 - c) 35,78
 - d) 109,23
 - e) 213,07

Table 1: Dataset A

Marker	Antonio	Borja	Camilo
D1S1656	16.3/20.3	16.3/16.3	12/20.3
TPOX	11/11	8/11	8/11
D2S1338	17/23	17/23	17/19
D3S1358	16/16	14/15	16/16
FGA	19/23	20/21	19/20
CSF1PO	10/12	9/12	8/12
D6S474	17/18	14/16	15/16
D7S820	8/12	8/8	10/12
D8S1179	12/13	12/14	12/13
D10S1248	13/16	13/16	14/16
TH01	9/9.3	7/9	9/9.3
vWA	17/18	17/17	17/18
D13S317	12/12	8/12	12/12
PentaE	5/12	5/12	12/20
D16S539	9/11	9/11	9/11
D18S51	19/20	13/16	14/19
D19S433	13/14	14/15	13/14
PentaD	11/12	9/11	11/12
D22S1045	15/16	16/16	11/15

Table 2: Dataset B

Marker	Antonio	Borja	Camilo
D3S3045	11/15	11/14	9/15
D5S2800	17/17	14/17	14/17
D7S3048	21/24	21/23	21/24
D8S1132	19/22	20/22	19/22
D9S1122	12/13	12/12	10/13
D10S1435	11/12	12/13	11/12
D11S2368	19/20	18/19	19/20
D12S391	20/24	17/24	17/20
D13S325	18/19	18/19	18/19
D14S1434	10/11	10/10	10/11
D15S659	17/18	14/18	17/18
D17S1301	11/11	11/12	11/12
D18S1364	13/18	14/14	13/16
D19S253	7/12	12/13	7/12
D20S482	11/14	14/15	12/16
D21S11	28/31.2	28/28	28/30

2. Dois marcadores do dataset A estão localizados no mesmo cromossoma. De acordo com a função de mapeamento de Haldane, a taxa de recombinação entre eles, arredondada para duas decimais, é:
 - a) 0,41
 - b) 0,42
 - c) 0,45
 - d) 0,48
 - e) 0,49
3. Seja FC a razão de mudança do quociente LR^l/LR^u para os dois marcadores da questão anterior. Então:
 - a) $FC < 1$
 - b) $FC = 1$
 - c) $1 < FC \leq 1,0001$
 - d) $1,0001 < FC \leq 1,001$
 - e) $FC > 1,001$

Você decide tipar os três indivíduos com outro kit, adicionando 16 novos marcadores STR. Este é o dataset B (Table 2). A seguir, é dito que dois marcadores formam um *par ligado* se estiverem localizados no mesmo cromossoma.

4. O número de *pares ligados* no dataset total (incluindo o par do Exercício 2) é:
 - a) 11
 - b) 12
 - c) 13
 - d) 14
 - e) 15
5. O LR^l total para os 35 marcadores, após arredondamento, é:
 - a) 213
 - b) 1 031
 - c) 7 624
 - d) 15 401
 - e) 23 125

6. O valor total da razão de mudança LR^{ℓ}/LR^u para os 35 marcadores é aproximadamente:
- 0,55
 - 1,01
 - 2,02
 - 3,03
 - 4,04
7. O número de *pares ligados* para os quais LR^{ℓ} é superior a LR^u , é:
- 4
 - 6
 - 8
 - 10
 - 12
8. O número de *pares ligados* para os quais LR^{ℓ} e LR^u apoiam *diferentes* hipóteses, é:
- 0
 - 1
 - 2
 - 3
 - 4

Uma medida adequada de quanto o *linkage* afeta a LR é o valor absoluto do logaritmo de base 2 da razão de mudança, ou seja,

$$W = |\log_2(LR^{\ell}/LR^u)|.$$

9. Encontre o valor de W para cada um dos *pares ligados*. Então,
- o par mais próximo tem o maior W , e o par mais distante tem o menor W .
 - o par mais próximo tem o maior W , mas o par mais distante não tem o menor W .
 - o par mais próximo não tem o maior W , mas o par mais distante tem o menor W .
 - o par mais distante tem o maior W .
 - Nenhuma das anteriores.

No seu relatório ao detetive Vargas, formula uma conclusão com base no conjunto de dados completo. De acordo com o protocolo do seu laboratório, $LR > 10\,000$ é considerado como forte suporte para H1 e $LR < 0.0001$ como forte suporte para H2; limites de 1 000 and 0,001 fornecem suporte moderado. Valores entre estes limites são considerados inconclusivos.

Se LR^{ℓ} e LR^u levarem a conclusões diferentes, terá que decidir como lidar com isso.

10. Conclui-se que os dados:
- apoiam fortemente um parentesco completo (como em H1).
 - apoiam moderadamente um parentesco completo.
 - apoiam fortemente um meio parentesco (como em H2).
 - apoiam moderadamente um meio parentesco.
 - fornecem evidências inconclusivas.

Parte II: Linkage Lab

Os próximos exercícios são baseados na aplicação Shiny **Linkage Lab**, disponível online em <https://magnusdv.shinyapps.io/linkagelab/>. Esta aplicação foi desenhada para explorar o efeito do *linkage* em casos simples de parentesco.

Cada caso compara dois relacionamentos possíveis entre os indivíduos A e B, genotipados para dois marcadores, M1 e M2. Ambos os marcadores têm quatro alelos, denominados por 1, 2, 3 e 4, cujas frequências alélicas e taxas de mutação podem ser definidas pelo usuário (assume-se que estes parâmetros são os mesmos para ambos os marcadores). A distância entre os marcadores pode ser dada em centiMorgans ou como uma taxa de recombinação. A aplicação calcula a razão de verossimilhança (LR) para os dois relacionamentos e apresenta os resultados num gráfico, onde se pode variar, livremente, um dos parâmetros.

Todas as perguntas abaixo dizem respeito a comparações *Sibs* : *half-sibs* (Irmãos : meios-irmãos). Inicialmente, considere todas as frequências alélicas como sendo de 0,25 e uma taxa de mutação de 0.

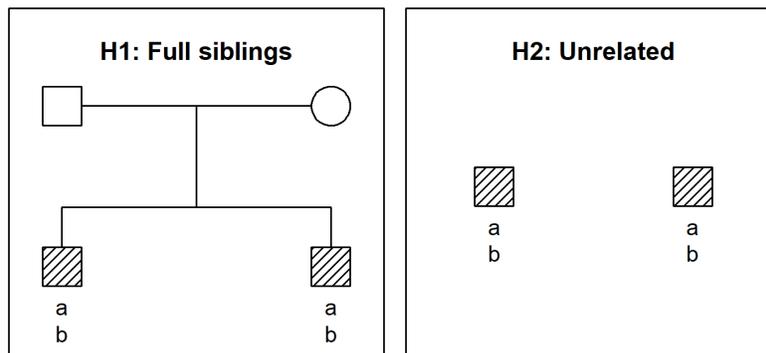
11. Considere que A tem o genótipo 1/2 para ambos os marcadores, enquanto B é 3/4 para ambos os marcadores. Então, *não* é verdade que:
 - a) o LR aumenta à medida que a ligação se torna mais forte
 - b) o LR, quando a ligação é completa, é exatamente o dobro do que sem ligação.
 - c) para ambas as hipóteses, a probabilidade é maior quando a taxa de recombinação é 0.
 - d) os dados apoiam a hipótese de irmãos inteiros se a taxa de recombinação é 0.
 - e) os dados apoiam a hipótese de meios-irmãos se a taxa de recombinação é 0,5.
12. Agora altere o genótipo de B para 1/2 no M1. Então, *não* é verdade que:
 - a) o LR aumenta à medida que a ligação se torna mais fraca.
 - b) o LR, na ausência de *linkage*, é exatamente o dobro do que quando a ligação é completa.
 - c) para ambas as hipóteses, a probabilidade é maior quando a taxa de recombinação é 0,5.
 - d) os dados apoiam a hipótese de meios-irmãos se a taxa de recombinação é 0.
 - e) os dados apoiam a hipótese de irmãos inteiros se a taxa de recombinação é 0,5.
13. Agora altere o genótipo de B para 1/3 no M2 (mantenha M1 como no exercício anterior). Então, *não* é verdade que:
 - a) o LR é inferior a 1 se os marcadores estiverem completamente ligados.
 - b) o LR é maior se a taxa de recombinação for 0,25.
 - c) a probabilidade de Ped 1 depende do *linkage*, mas a probabilidade de Ped 2 não.
 - d) os dados apoiam a hipótese de irmãos inteiros se os marcadores não estiverem ligados.
 - e) o LR é igual a 1 para marcadores separados por x cM, para um determinado valor de x entre 0 e 10.

Finalmente, considere que A e B são ambos 1/1 para M1, e 2/2 e 3/3, respectivamente, para M2. Defina a variável "*PLot variable*" como sendo "Frequency of '1' allele" (frequência do alelo 1), a distância dos marcadores como 50 cM, e a taxa de mutação como 0,005. Mantenha as frequências dos alelos '2' e '3' em 0,25, como anteriormente.

14. A frequência de '1' para a qual $LR = 1$, é aproximadamente:
 - a) 0,1
 - b) 0,2
 - c) 0,3
 - d) 0,4
 - e) 0,5

Parte III: *Linkage* em X

Os restantes exercícios exploram o efeito do *linkage* entre SNPs do cromossoma X, em testes de irmandade. Suponha que dois rapazes são ambos hemizigóticos para o alelo a no marcador X1, e alelo b no marcador X2, como mostrado na figura abaixo.



Sejam $p = P(a)$ e $q = P(b)$ as frequências alélicas de a e b . Em todos os exercícios que se seguem, assumo que p e q são ambos próximos a 0. Seja θ a taxa de recombinação entre os marcadores.

15. A probabilidade de uma pessoa ao acaso, do sexo feminino, ter o alelo a no *locus* X1, é aproximadamente
- $p^2 - p$
 - $p/2$
 - p
 - $2p$
 - p^2
16. A probabilidade da hipótese H_1 , ignorando um fator constante¹, aproxima-se melhor por
- $pq\theta(1 - \theta)$
 - $pq(1 - \theta)^2$
 - $p^2q^2(1 - \theta)^2$
 - $pq(\theta^2 + (1 - \theta)^2)$
 - $p^2q^2(\theta^2 + (1 - \theta)^2)$
17. O LR que compara H_1 com H_2 , ignorando um fator constante, aproxima-se melhor por
- $(\theta^2 + (1 - \theta)^2)/(pq)$
 - $1/(pq)$
 - $pq/(\theta^2 + (1 - \theta)^2)$
 - $pq/(\theta(1 - \theta))$
 - pq
18. Numa situação destas,
- é impossível a ligação completa a menos que tenha ocorrido uma mutação.
 - a ligação completa duplica, aproximadamente, o LR em comparação com marcadores não ligados.
 - o LR é maior se $\theta = 0,25$.
 - o LR é maior se os marcadores não estiverem ligados.
 - o LR é, aproximadamente, independente do *linkage*.

Considere agora a mesma situação, mas onde o rapaz mais à esquerda tem o alelo comum A no marcador X1.

19. Numa situação destas,
- é impossível a ligação completa a menos que tenha ocorrido uma mutação.
 - a ligação completa duplica, aproximadamente, o LR em comparação com marcadores não ligados.
 - o LR é maior se $\theta = 0,25$.
 - o LR é maior se os marcadores não estiverem ligados.
 - o LR é, aproximadamente, independente do *linkage*.

Finalmente, considere a mesma situação, mas onde os dois rapazes têm o alelo comum A no marcador X1.

20. Numa situação destas,
- é impossível a ligação completa a menos que tenha ocorrido uma mutação.
 - a ligação completa duplica, aproximadamente, o LR em comparação com marcadores não ligados.
 - o LR é maior se $\theta = 0,25$.
 - o LR é maior se os marcadores não estiverem ligados.
 - o LR é, aproximadamente, independente do *linkage*.

BOA SORTE!

Magnus

¹ Por exemplo, se a verdade for $3pq\theta$, então $pq\theta$ é considerado correto. (O objetivo desta formulação é evitar soluções triviais verificando diretamente as alternativas.)

Uso do KLINK

Aqui estão algumas dicas para a aplicação KLINK para testes de parentesco com pares de marcadores ligados:

- A aplicação online está disponível em <https://magnusdv.shinyapps.io/klink/>.
- Carregue o arquivo .fam com o conjunto de dados que deseja analisar.
- Defina *Marker map* como "Custom" e carregue o ficheiro de mapa.
- Defina *Map function* como "Haldane".
- Talvez queira ajustar a configuração *Ignore Linkage above (cM)*.
- Pulse *Calculate LR*. Note que os resultados incluem LRs assumindo a presença e a ausência de linkage.
- Aumente *Decimals* caso necessite uma maior precisão.

Leitura adicional

Os livros a seguir fornecem mais detalhes sobre *linkage* em geral e sobre marcadores ligados, em testes de parentesco.

- Thompson, *Statistical Inference from Genetic Data on Pedigrees*. Institute of Mathematical Statistics, 2000.
- Egeland, Kling & Mostad, *Relationship Inference with Families and R*. Academic Press, 2016.

Responsável pelo desenvolvimento do exercício:

Magnus Dehli Vigeland

magnusdv@gmail.com

Tradução:

Cíntia Alves (cintialves70@gmail.com)

Comité Executivo do GHEP-ISFG

info@ghep-isfg.org