

# Interação entre genes: a epistasia



**José Osório Azevedo Neto<sup>1</sup>, Andréa Cristina Peripato<sup>2</sup>, Sergio Russo Matioli<sup>3</sup>**

<sup>1</sup>Área de Informática, *Campus* de Pirituba, Instituto Federal de São Paulo, SP

<sup>2</sup>Departamento de Genética e Evolução, Universidade Federal de São Carlos, SP

<sup>3</sup>Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, SP

Autor para correspondência - [srmatioli@ib.usp.br](mailto:srmatioli@ib.usp.br)

**Palavras-chave:** interação interlôcus, extensões das leis de Mendel

O termo “epistasia” refere-se à interação entre genótipos de genes (lócus) diferentes (interação interlócus), em que a expressão de um lócus depende do genótipo de outro lócus de forma não independente. Esse conceito pode ser aplicado a dois domínios: um de natureza fisiológica/bioquímica, voltado para a maneira como os produtos gênicos interagem em processos biológicos. O segundo é de caráter quantitativo e mais matemático, referindo-se aos efeitos do fenômeno sobre um fenótipo quantitativo. Ambas as definições são valiosas e válidas em seus respectivos domínios, oferecendo perspectivas diferentes sobre o conceito de epistasia. Neste artigo abordaremos as diferentes definições de epistasia e sua importância, incluindo exemplos. O objetivo do artigo é oferecer subsídios para a discussão do fenômeno no ensino superior.

Antes de detalharmos o conceito de epistasia propriamente dito, é importante que revisemos o que existe por trás dos conceitos mendelianos de dominância e recessividade. A expressão de um genótipo específico em uma característica pode ser entendida pela interação entre os alelos em um mesmo lócus. Nessa interação intralócus, um organismo tem dois alelos diferentes para uma característica e, no contraste entre genótipos homocigóticos e heterocigótico, é possível identificar a relação de dominância e recessividade. A característica será dominante quando observada nos genótipos homocigóticos e heterocigóticos e recessiva quando observada somente em um genótipo homocigótico. A base molecular desse fenômeno deve-se à bioquímica dos produtos dos alelos, que muitas vezes são proteínas, e às suas interações. Abaixo, listamos algumas das causas moleculares da dominância e recessividade. Um aprofundamento sobre o assunto pode ser encontrado no artigo “Dominante ou Recessivo?” publicado na *Genética na Escola* (v.9, n.2 2012).

**1. Atividade Enzimática:** Muitos genes codificam enzimas. Um alelo dominante pode produzir uma enzima funcional, enquanto um alelo recessivo pode produzir uma enzima não funcional. Se a quanti-

dade da enzima produzida na presença de apenas um alelo é suficiente para um fenótipo normal (haplossuficiência), então esse alelo é dominante. Isso frequentemente resulta em dominância completa.

**2. Função e Estrutura da Proteína:** Em casos em que a estrutura da proteína é crítica para sua função, um alelo dominante pode produzir uma proteína normal, enquanto um alelo recessivo produz uma proteína não funcional, caso semelhante ao da atividade enzimática.

**3. Efeitos Dominantes-Negativos:** Algumas mutações exercem um efeito dominante produzindo uma proteína que não apenas falha em funcionar adequadamente, mas também interfere na função da proteína produzida pelo alelo normal. Esse tipo de mutação pode levar a um fenótipo mesmo quando apenas uma cópia (alelo) carrega a mutação em um lócus.

**4. Elementos Regulatórios:** Nem todas as mutações que levam a mudanças fenotípicas ocorrem na região codificante dos genes. Algumas ocorrem em regiões reguladoras que controlam quando, onde e quanto de uma proteína é produzida. Fenótipos dominantes ou recessivos podem surgir a partir de mutações nessas áreas.

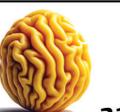
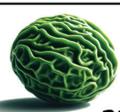
Na interação intralocus, nem todos os alelos mostram dominância completa. Na dominância incompleta, um indivíduo heterozigótico exibirá um fenótipo intermediário entre os dois fenótipos dos indivíduos homozigóticos. Um exemplo disso é a **hipercolesterolemia hereditária**, em que indivíduos HH são homozigóticos para a capacidade de produzir receptores **LDL** nas células, e, assim, possuem fenótipo normal, enquanto os indivíduos hh são incapazes de produzir receptores LDL, por isso possuem a forma severa da hipercolesterolemia. Os indivíduos Hh possuem o fenótipo intermediário, com a doença manifestando-se também de maneira intermediária. Na codominância, ambos os alelos são expressos e ambos os fenótipos são exibidos simultaneamente (por exemplo, o tipo sanguíneo AB em humanos).

O conceito de epistasia vem sendo estabelecido desde os primórdios da Genética moderna, no começo do século XX. A primeira menção ao fenômeno de interação entre alelos de locus diferentes tinha natureza fisiológica. A palavra em inglês *epistasis* foi cunhada por Bateson em 1909, para descrever a influência de um locus em outro na cor da pelagem de coelhos. O termo “epistasia” é derivado do grego *epi-*, que significa “sobre” ou “além”, e *stasis*, que significa “parada” ou “estagnação”.

A interação epistática é diferente da interação simples entre alelos de genes diferentes. No cruzamento entre diíbridos (duplos heterozigóticos), no caso da segregação independente, esperam-se as proporções dos fenótipos 9:3:3:1. As combinações entre formatos e cores da ervilha, no trabalho clássico de Mendel, exemplificam esse tipo de interação (Figura 1).

**Hipercolesterolemia hereditária** - condição genética caracterizada por níveis elevados e persistentes de colesterol no sangue, especialmente das lipoproteínas de baixa densidade (LDL), conhecido como “colesterol ruim”.

**LDL** – *Low Density Lipoprotein* (lipoproteína de baixa densidade) - uma classe de partículas responsável pelo transporte de colesterol no sangue. Seu nível elevado permite o acúmulo de placas nas artérias, aumentando o risco de doenças cardiovasculares.

	AB	Ab	aB	ab
AB	 <b>AABB</b>	 <b>AABb</b>	 <b>AaBB</b>	 <b>AaBb</b>
Ab	 <b>AABb</b>	 <b>AAbb</b>	 <b>AaBb</b>	 <b>Aabb</b>
aB	 <b>AaBB</b>	 <b>AaBb</b>	 <b>aaBB</b>	 <b>aaBb</b>
ab	 <b>AaBb</b>	 <b>Aabb</b>	 <b>aaBb</b>	 <b>aabb</b>

**Figura 1.** Resultados do experimento de Mendel que mostram a independência da segregação dos caracteres “forma” e “cor” das ervilhas de *Pisum sativum*, representados no quadro de Punnett com as proporções 9:3:3:1 dos fenótipos amarelo/liso, amarelo/rugoso, verde/liso e verde/rugoso, respectivamente. O locus A refere-se à forma da ervilha, com alelos A e a; o locus B refere-se à coloração das ervilhas, com alelos B e b. As letras A e B foram utilizadas por Mendel em seu artigo original.

Entretanto, no caso do cruzamento entre coelhos diíbridos descrito pelo geneticista norte-americano William Bateson, foi observada uma proporção 9:3:4 (Figura 2).

As interações epistáticas podem ocorrer em vários contextos e, por envolver mais de um

locus gênico que resulta em um produto final, a mutação em um desses alelos pode produzir um fenótipo alterado. Temos dois casos que permitem ter um maior entendimento da interconexão entre os produtos gênicos associados a um fenótipo, a epistasia recessiva e a epistasia dominante.

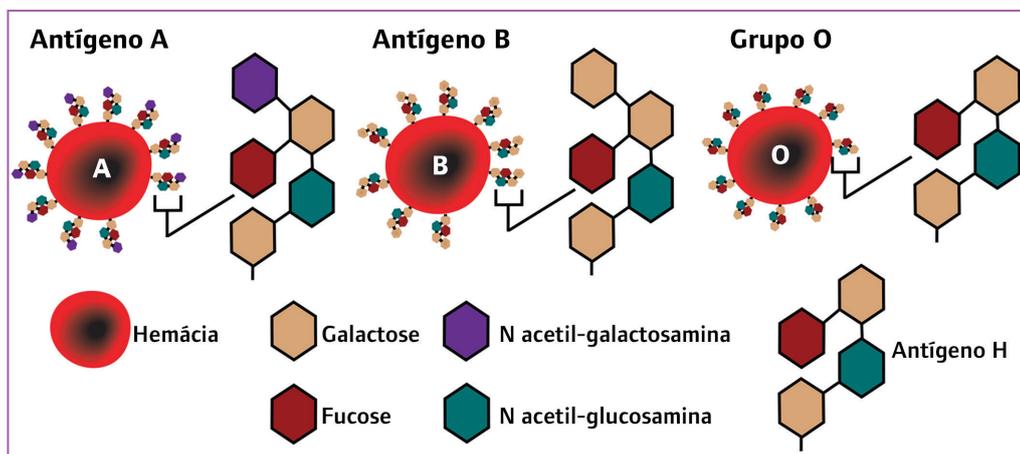
	CG	Cg	cG	cg
CG	 CCGG	 CCGg	 CcGG	 CcGg
Cg	 CCGg	 CCgg	 CcGg	 Ccgg
cG	 CcGG	 CcGg	 ccGG	 ccGg
cg	 CcGg	 Ccgg	 ccGg	 ccgg

**Figura 2.** Resultados do experimento relatado por Bateson que mostram o fenômeno da epistasia na coloração de pelagem de coelhos com as proporções 9:3:4 dos fenótipos cinza, preto e albino, respectivamente. A determinação da coloração vai depender dos genótipos nos locos C (alelos C e c) e G (alelos G e g). A presença do genótipo “cc” implica fenótipo branco independentemente dos alelos do locus G, enquanto o genótipo “gg” acentua a cor na presença do alelo “C” do locus C.

## 1. Epistasia recessiva: fenótipo Bombaim

Nesse tipo de interação, o homocigoto recessivo mascara a expressão fenotípica dos alelos codominantes ou dominantes. O fenótipo Bombaim (descoberto em Bombaim, Índia) é um caso raro em que a pessoa parecer o tipo sanguíneo O, cujo genótipo seria  $ii$ , mas, se ela se unir a uma pessoa tipo O ( $ii$ ), poderá gerar uma criança com o tipo A ( $I^A i$ ), ou tipo B ( $I^B i$ ). O sistema ABO é bem conhecido e parece resultar de herança monogênica, mas o fenótipo Bombaim mostra que pode ser um pouco mais complicado. Os alelos  $I^A$ ,  $I^B$  e  $i$  estão situados no cromosomo

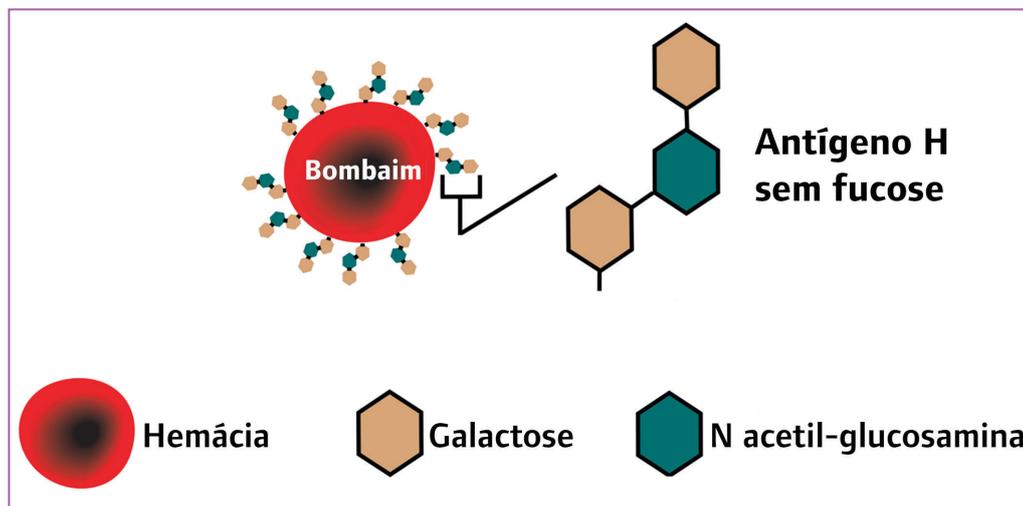
9 e seus produtos primários são enzimas (glicosiltransferases) que adicionam carboidratos específicos a um substrato de base, chamado de antígeno H, localizado na superfície das hemácias (Figura 3). As especificidades do tipo sanguíneo são determinadas pelos aminoácidos próximos da ligação das enzimas glicosiltransferases. Se o indivíduo for do grupo A, o açúcar N-acetilgalactosamina será adicionado. Se o indivíduo for do grupo B, a D-galactose será adicionada, como mostrado na Figura 3. Indivíduos do grupo O não produzem transferases.



**Figura 3.** Bases moleculares dos grupos sanguíneos do sistema ABO. O açúcar N-galactosamina adicionado ao antígeno H confere o caráter A, as hemácias com essa molécula reagem com o anticorpo anti-A. O açúcar galactose adicionado ao antígeno H confere o caráter B, as hemácias com essa molécula reagem com o anticorpo anti-B. O sangue dos indivíduos que somente apresentam o antígeno H não reagem com os anticorpos anti-A ou anti-B.

O substrato de base, em que ocorre o acoplamento das transferases, chamado de antígeno H, é o produto do gene *fucosiltransferase* localizado no cromossomo 19. Pessoas com genótipo *HH* ou *Hh*, apresentam o substrato H na superfície de suas hemácias. Já os *hh* não produzirão o antígeno H, impossibilitando a associação das glicosiltransferases e, assim, impedindo o reconhecimento como tipo sanguíneo A, AB ou B (Figura 4). Pessoas com genótipo *hh* são as que apresentam o fenótipo Bombaim, que, por não apresentarem reação ao anticorpo anti-A ou anti-B, podem ser classificadas

como tipo O. Mas, geneticamente elas terão os alelos  $I^A$ ,  $I^B$ , ou  $i$ , dependendo do seu genótipo. Se a sua progênie tiver pelo menos um alelo  $H$ , os descendentes poderão apresentar o tipo sanguíneo de acordo com seu genótipo no sistema ABO. Esse é o caso de epistasia recessiva, em que o genótipo *hh* mascara o efeito dominante ( $I^A i$ ,  $I^B i$ ) ou codominante ( $I^A I^B$ ) dos alelos no outro locus. Na Figura 3 estão representados os antígenos presentes em cada um dos fenótipos de grupos sanguíneos que são decorrentes dos tipos de carboidratos presentes em cada um deles.



**Figura 4.** Indivíduos do fenótipo Bombaim são homocigóticos para o alelo que confere a deficiência da enzima Fucosiltransferase (genótipo *hh*). Nas suas hemácias o antígeno H não contém a fucose, que faz com que não haja a ligação dos açúcares N acetil galactosamina (que confere o caráter A do grupo sanguíneo ABO), ou da segunda molécula da galactose (que confere o caráter B). Assim, como o sangue dos indivíduos *hh* não reage tanto com o anticorpo anti-A ou com o anticorpo anti-B, eles são confundidos com indivíduos do grupo O, mesmo que o genótipo deles contenha os alelos  $I^A$  ou  $I^B$ .

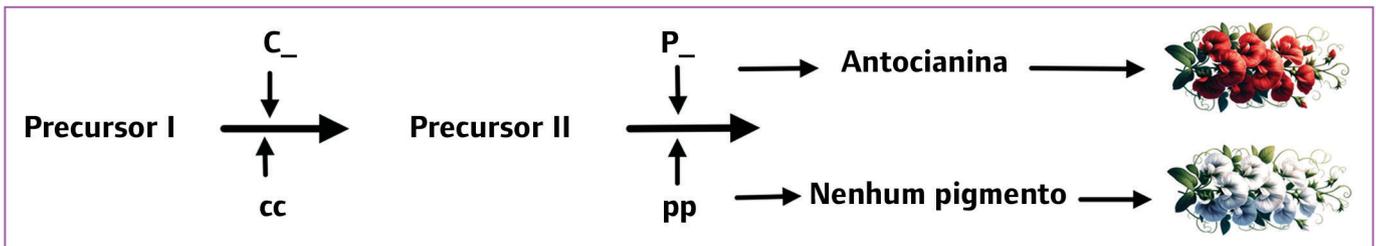
## 2. Epistasia dominante: cor das pétalas de flores de ervilha

Na epistasia dominante, é necessária a presença de pelo menos um alelo dominante em cada um dos dois loci para que a flor seja vermelha; se houver genótipo homocigoto recessivo em qualquer um dos dois loci, a flor será branca. Na Figura 5 está o exemplo, sob a forma do quadro de Punnett, com os resultados do cruzamento entre plantas híbridas também descrito por Bateson e se refere à herança da cor das pétalas das flores de ervilha.

Estudos posteriores mostraram que os genes envolvidos na coloração das pétalas das flores codificavam enzimas que participam da síntese da antocianina (pigmento responsável pela coloração de frutas, flores e folhas) em uma via metabólica em série. Assim, qualquer deficiência em uma dessas duas enzimas inviabiliza a produção do pigmento pela interrupção do fluxo de metabólitos nessa via (Figura 6).

	CP	Cp	cP	cp
CP	 CCPP	 CCPp	 CcPP	 CcPp
Cp	 CCPp	 CCpp	 CcPp	 Ccpp
cP	 CcPP	 CcPp	 ccPP	 ccPp
cp	 CcPp	 Ccpp	 ccPp	 ccpp

**Figura 5.** Resultados dos cruzamentos de ervilhas relatados por Bateson que mostram o fenômeno da epistasia na coloração das pétalas das flores, com as proporções 9:7 dos fenótipos vermelho e branco, respectivamente. A presença de alelos dominantes simultaneamente nos dois loci ( $C\_P\_$ ) implica em fenótipo vermelho, enquanto o genótipo duplo recessivo em qualquer um dos loci resulta na cor branca.



**Figura 6.** Representação da via metabólica para coloração das pétalas das flores de ervilha. Os genótipos  $C\_$  e  $P\_$  indicam que tanto  $CC/Cc$  ou  $PP/Pp$  apresentam a mesma relação com o fenótipo, pois ter apenas um dos alelos dominantes já é o suficiente para a conversão na via. Os genótipos  $cc$  e  $pp$  representam homocigotos recessivos.

## O conceito quantitativo de epistasia

Ronald A. Fisher, um dos pioneiros da genética e estatística modernas, introduziu o conceito de *epistacy* (que atualmente chamamos de "epistasia") em seu trabalho sobre a genética da variação contínua. Fisher usou o termo para descrever a interação entre dois ou mais loci genéticos que afetam a expressão fenotípica de uma característica.

No contexto da genética quantitativa, a epistasia de Fisher refere-se à interação não aditiva entre alelos de genes diferentes, de modo que a expressão conjunta de alelos em dois ou mais genes tem um efeito que não é a soma simples dos efeitos individuais de cada

alelo. Em outras palavras, a presença de um alelo em um gene pode modificar a expressão de um alelo em um segundo gene.

Para entendermos a implicação da epistasia na Genética quantitativa, em primeiro lugar precisamos caracterizar o efeito aditivo. Suponhamos dois loci diferentes, **A** e **B**, nos quais segregam dois alelos em cada um: **A**, **a**, **B** e **b**, respectivamente. Suponhamos, também, que estes genes influenciam um caráter quantitativo, ou seja, um caráter que possa ser medido e expresso em termos numéricos, como peso, altura, quantidade de um pigmento, teor de um metabólito etc. Suponhamos ainda que, para um caráter quantitativo, a presença de um alelo do tipo "maiúsculo" (**A** ou **B**) adicione uma determinada quantidade. A Tabela 1 apresenta os valores numéricos hipotéticos na situação na qual o genótipo **aabb** tem

o menor valor da característica, que é 20 (por exemplo, a altura de um arbusto), enquanto os demais genótipos têm o seu valor acrescido de 10 para cada alelo maiúsculo que apresentam (**A** ou **B**, indistintamente). Assim, quando consideramos os dois loci,

na ausência do alelo maiúsculo o arbusto terá 20 cm. A cada alelo “maiúsculo” que a planta tiver em um dos loci ela terá 10 cm a mais que o tamanho base e, no caso, as plantas **AABB** terão o potencial de atingir 60 cm de altura.

	<b>AA</b>	<b>Aa</b>	<b>aa</b>
<b>BB</b>	60	50	40
<b>Bb</b>	50	40	30
<b>bb</b>	40	30	20

**Tabela 1.**

Efeito aditivo de dois genes em um genótipo, no qual o valor base é 20, para o fenótipo correspondente ao genótipo **aabb** e cada alelo “maiúsculo” (**A** ou **B**) adiciona um valor de 10 ao fenótipo.

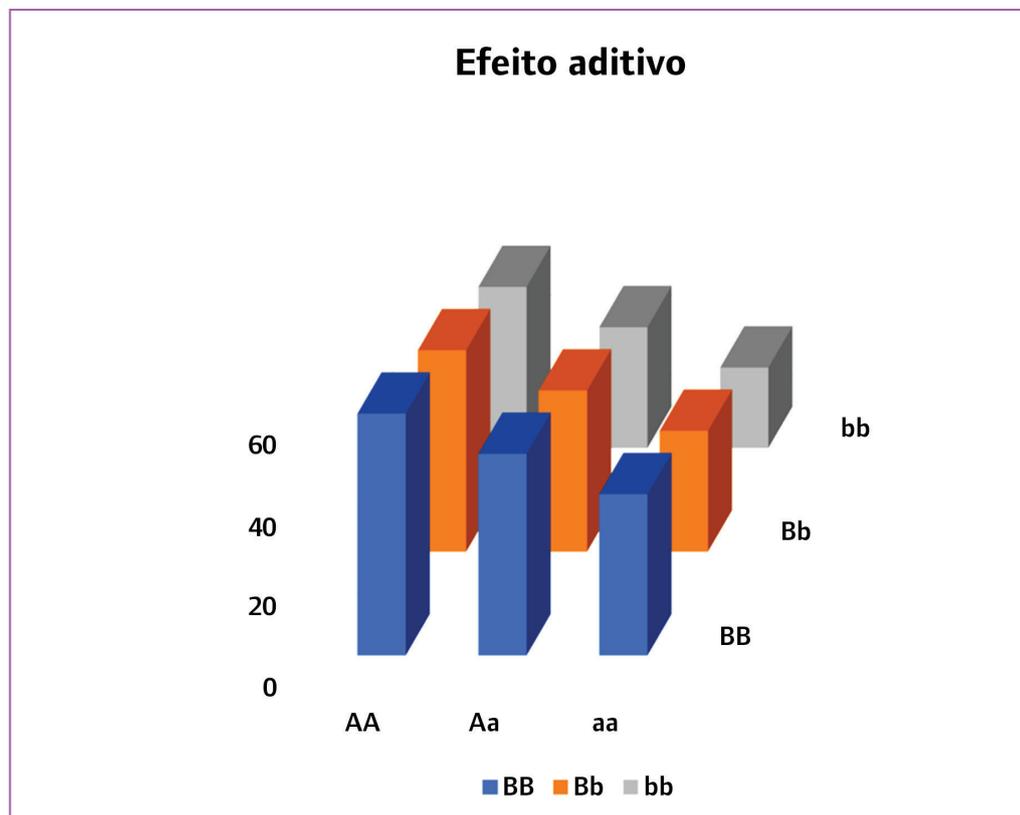
Podemos representar os valores da Tabela 1 graficamente, tal como representado na Figura 7.

No contexto de **efeito aditivo**, podemos pensar que a contribuição de um único gameta, independentemente de ser feminino ou masculino, já permite inferir a contribuição do alelo na característica da progênie. Se o espermatozoide ou o ovócito tiver um **A** e/ou um **B**, o potencial do fenótipo já está

no gameta. Os **efeitos não aditivos** não funcionam dessa mesma maneira e dependem da fecundação, pois são influenciados pelas interações intralocus e interlocus, que somente serão estabelecidas após a união dos gametas. Se for interação intralocus, somente será possível verificar os efeitos não aditivos de dominância conhecendo os dois alelos no mesmo locus, e na interação interlocus, a epistasia, é necessário conhecer os alelos do outro locus para identificar tal efeito.

**Efeitos não aditivos** - são efeitos genéticos que não podem ser explicados pela simples soma dos efeitos dos alelos individuais. Incluem interações entre alelos, como dominância e epistasia.

**Efeitos aditivos** - refere-se à contribuição cumulativa de alelos para uma característica fenotípica. A soma dos efeitos de cada alelo afeta diretamente o fenótipo observado.



**Figura 7.**

Efeito aditivo de dois genes em um fenótipo, no qual o valor base é 20, para o fenótipo correspondente ao genótipo **aabb** e cada alelo “maiúsculo” (**A** ou **B**) adiciona um valor de 10 ao fenótipo.

Fisher considerou a epistasia como uma complicação na análise da variação genética, já que ela introduz uma fonte adicional de variação que não é facilmente separável dos efeitos aditivos dos genes individuais. Em sua abordagem, ele desenvolveu métodos estatísticos para discriminar os efeitos aditivos, dominantes e epistáticos na variação fenotípica populacional.

Uma vez caracterizado o efeito aditivo fica mais intuitivo entender o efeito não aditivo

da epistasia. As possibilidades de epistasia são muito vastas e não há consenso na nomenclatura de seus tipos. Dentre os tipos “clássicos” verificamos a Tabela 2 e Figura 8, que representam uma situação de epistasia dominante na herança quantitativa considerando-se dois locus, A e B. Se houver pelo menos um alelo “maiúsculo” em cada um dos locus (A\_B\_), o fenótipo corresponderá ao valor 60, mas esse valor será 30 se houver um duplo recessivo (aa\_\_ ou \_\_bb).

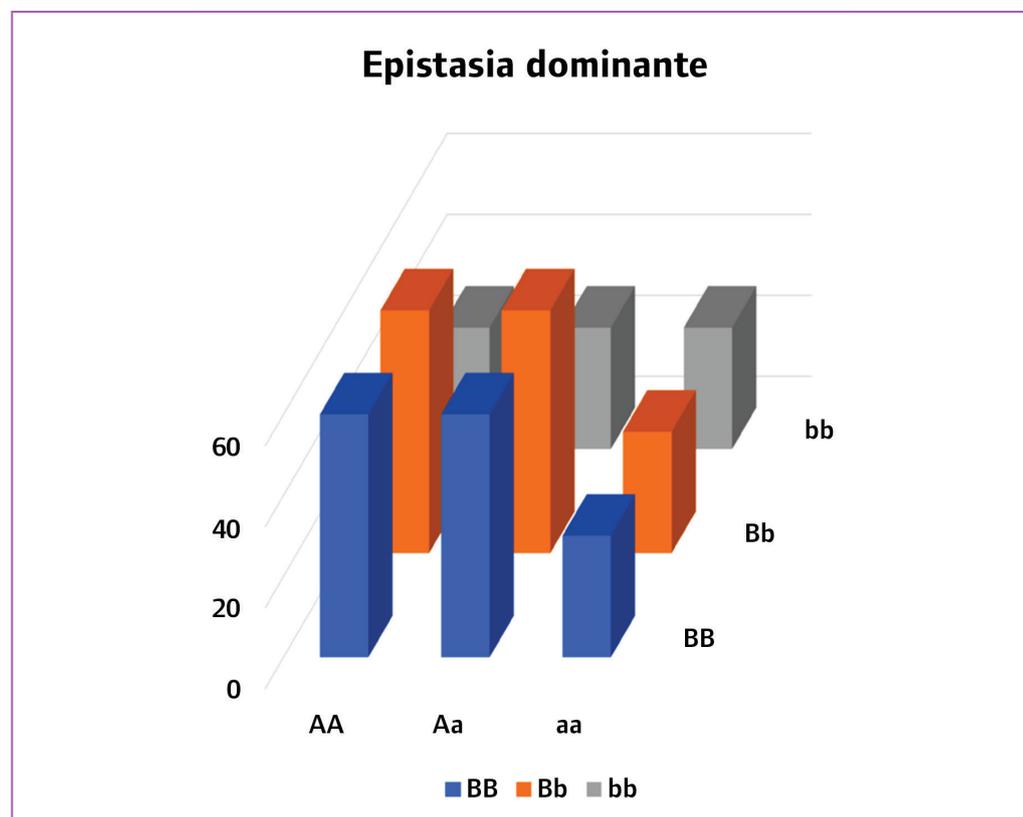
	AA	Aa	aa
BB	60	60	30
Bb	60	60	30
bb	30	30	30

**Tabela 2.**

Epistasia dominante, no qual o valor é 30 para o fenótipo correspondente aos genótipos que contêm um duplo recessivo em qualquer um dos genes (aa\_\_ ou \_\_bb) e os genótipos que contêm pelo menos um alelo dominante em cada um dos locus (A\_B\_) apresentam o valor 60 para o fenótipo considerado.

A representação gráfica dos valores da Tabela 2 está apresentada na Figura 8.

Outro exemplo é a epistasia codominante, que está representada na Tabela 3 e Figura 9.



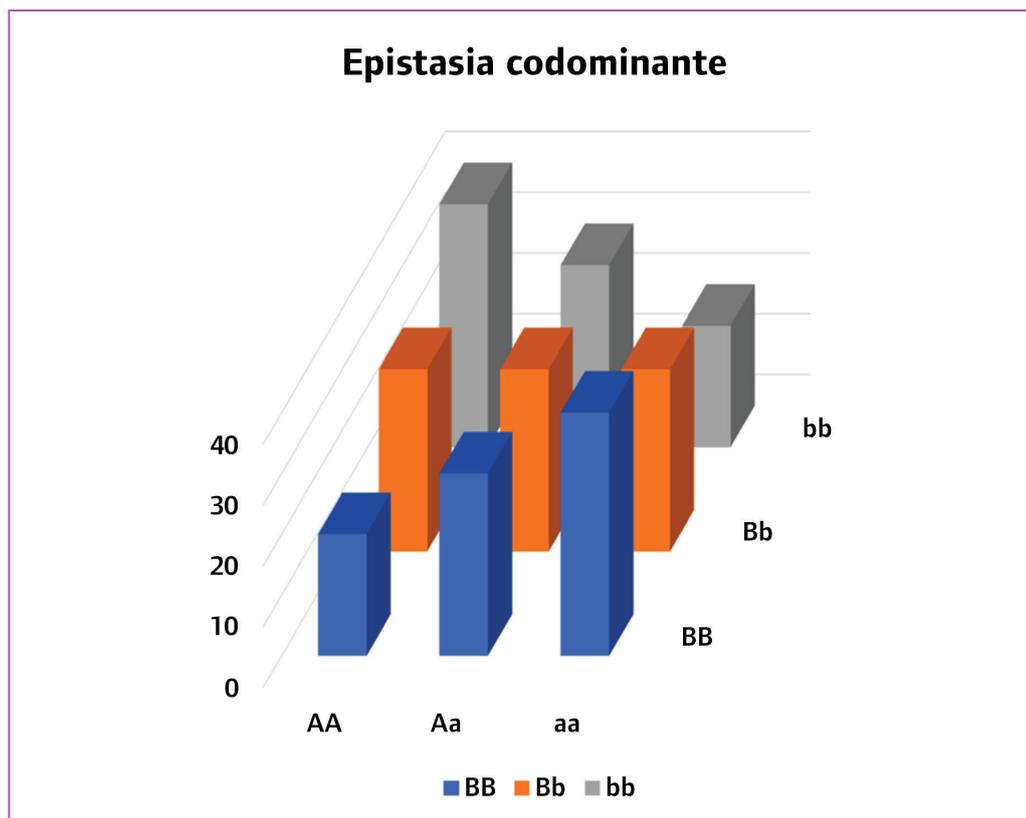
**Figura 8.**

Epistasia dominante em um fenótipo quantitativo, no qual o valor do fenótipo é 60 para os genótipos que contenham pelo menos um alelo maiúsculo (A\_B\_) e valor 30 para os genótipos que sejam duplos minúsculos em pelo menos um dos genes (aa\_\_ ou \_\_bb).

	AA	Aa	aa
BB	20	30	40
Bb	30	30	30
bb	40	30	20

**Tabela 3.**

Epistasia codominante, na qual o valor é 30 para o fenótipo correspondente aos genótipos que contêm um heterozigótico em qualquer um dos genes (**Aa\_\_** ou **\_\_Bb**). Os genótipos homozigóticos **AAbb** ou **bbBB** terão o valor 40 e os genótipos homozigóticos **aaBB** ou **AAbb** terão o valor 20.



**Figura 9.**

**Epistasia codominante quantitativa.** O valor é 30 para o fenótipo correspondente aos genótipos que contêm um heterozigótico em qualquer um dos genes (**Aa\_\_** ou **\_\_Bb**), os genótipos homozigóticos **AAbb** ou **bbBB** terão o valor 40 e os genótipos homozigóticos **aaBB** ou **AAbb** terão o valor 20.

## Consequências da epistasia

A importância da epistasia na evolução e na genética tem sido um tópico de debate contínuo. Enquanto Fisher acreditava que a epistasia era relativamente insignificante para a adaptação e evolução, outros geneticistas e evolucionistas argumentaram que a epistasia pode desempenhar um papel crucial na determinação da trajetória adaptativa de populações.

Durante o processo de adaptação de uma espécie a um ambiente, pode haver situações

em que continuamente modificam-se genes que têm a propriedade de alterar efeitos de outros genes. Isso geraria os chamados “**complexos gênicos coadaptados**”. A existência de tais complexos pode ter implicações importantes no isolamento reprodutivo entre espécies, dado que os híbridos que se formam podem ter desequilíbrios entre genes de complexos diferentes, causando inviabilidade ou esterilidade. Por outro lado, os complexos gênicos coadaptados quando se encontram em indivíduos híbridos a partir do encontro de populações isoladas pode contribuir com o fenômeno da **heterose**, situação na qual o fenótipo do heterozigoto é superior (em valor) àqueles dos homozigotos.

### Complexos gênicos coadaptados

- refere-se a um grupo de genes que interagem entre si e coevoluem devido à seleção natural. A coadaptação ocorre quando a combinação específica de alelos em vários loci diferentes é vantajosa.

**Heterose** - também conhecida como vigor híbrido, é a superioridade do fenótipo de um organismo híbrido em comparação aos seus pais homozigóticos.

Nesse contexto, *epistacy* ou epistasia, como concebido por Ronald A. Fisher, refere-se à interação genética na qual a expressão combinada de alelos em múltiplos loci é diferente da soma de seus efeitos individuais. Fisher desenvolveu métodos para analisar essas interações no contexto da genética quantitativa, proporcionando uma base para estudos subsequentes sobre a importância da epistasia na genética e evolução.

Outra consequência do fenômeno da epistasia é tornar muito complexa a tarefa de entendermos as relações entre genótipos e fenótipos, mesmo que existam dados referentes a muitos genótipos, distribuídos por todo o genoma e seus respectivos fenótipos. Por exemplo, em um estudo típico de associação, os indivíduos correspondentes a grupos de fenótipos são separados e então se verifica se existem genótipos que são significativamente mais ou menos frequentes entre os grupos. Como não sabemos, em princípio, se o efeito de um determinado gene é influenciado por outros, é necessário verificar o efeito da combinação de genótipos de genes diferentes. Isso implica o aumento da dimensionalidade do estudo e resulta naquilo que é conhecido como a “maldição da dimensionalidade”, que surge em vários campos, incluindo estatística, análise de dados e aprendizado de máquina.

Essa problemática também é verificada quando se busca identificar as variações genéticas associadas a doenças ou características quantitativas. O objetivo desses estudos é identificar polimorfismos de nucleotídeo único (SNPs - *Single Nucleotide Polymorphisms* - Quadro 1) ou regiões do genoma que estejam correlacionadas com a característica de interesse e, devido a efeitos não aditivos, como os epistáticos, a “maldição da dimensionalidade” apresenta desafios únicos. A abordagem que utiliza genomas de muitos indivíduos não aparentados na busca por variações genéticas associadas a doenças é conhecida como estudos de GWAS (*Genome Wide Association Studies* - Estudos de Associação Genômica Ampla - Quadro 2). Outra abordagem, que utiliza cruzamentos planejados para verificar o papel dos genes nas características, é conhecida como estu-

dos de QTL (*Quantitative Trait Loci*, loci de características quantitativas - Quadro 3). Em ambas as abordagens os desafios são semelhantes. Os estudos de QTLs são empregados em animais domesticados e em plantas cultivadas, enquanto em seres humanos somente são realizados GWAS. Não se emprega estudos de QTLs em humanos por motivos éticos, dado que se baseiam em cruzamentos programados.

A “maldição da dimensionalidade” refere-se ao aumento exponencial no volume associado à adição de dimensões extras a um espaço matemático. No contexto de GWAS/QTL, à medida que consideramos mais e mais SNPs (dimensões) para detectar interações epistáticas, o número de possíveis combinações de SNPs que precisam ser testadas para interações cresce exponencialmente. Por exemplo, em um único locus com 2 alelos (*A* e *a*) são possíveis 3 genótipos (*AA*, *Aa* e *aa*), para dois loci que contribuem para um caráter são possíveis 9 genótipos, para 3 loci são 27, para 4 são 81.

Os desafios na busca por efeitos epistáticos são a explosão combinatória, problemas de natureza estatística e a esparsidade dos dados, ou seja, na situação em que as amostras de muitas classes possíveis são muito pequenas ou até inexistentes. Se existirem um milhão de SNPs, testar todas as possíveis interações em pares requer a avaliação de quase 500 bilhões de combinações! Isso é computacionalmente intenso e demorado. Testar todas as possíveis combinações de trios é computacionalmente impraticável com a tecnologia atual. Com relação às limitações de natureza estatística, à medida que o número de testes aumenta, a probabilidade de falsos positivos (**erros do Tipo I**) também aumenta. Para contabilizar isso, são aplicadas correções rigorosas de múltiplos testes, que, por sua vez, reduzem o poder estatístico de detectar as interações que realmente ocorrem. Além disso, com o aumento do número de dimensões (SNPs), os dados disponíveis tornam-se esparsos. Isso significa que para muitas possíveis combinações de SNPs, pode haver poucos ou nenhum indivíduo com aquela combinação particular, tornando difícil se tirar conclusões significativas.

**Erros do Tipo I** - em testes estatísticos, um erro do Tipo I ocorre quando se rejeita incorretamente uma hipótese nula verdadeira. Também conhecido como falso positivo.

Devido à maldição da dimensionalidade, muitos GWAS estão focados principalmente na detecção de efeitos principais (efeitos de SNP individuais) em vez de interações, embora interações epistáticas possam desempenhar um papel significativo na **arquitetura genética** de características com herança multifatorial. Métodos e algoritmos especializados foram desenvolvidos para buscar interações epistáticas de maneira mais computacionalmente eficiente, mas o desafio permanece significativo. Enquanto os GWAS têm sido bem-sucedidos na identificação de muitas associações genéticas com doenças e características, a maldição da dimensionalidade representa um desafio significativo na busca por efeitos epistáticos.

No cenário atual não existe a possibilidade de se negligenciar a contribuição da epistasia em características complexas, principalmente envolvendo doenças, para conseguir expli-

car todos os fatores genéticos a elas associados. Superar esse desafio requer abordagens computacionais inovadoras, compreensão mais profunda da arquitetura genética e conjuntos de dados maiores para aumentar o poder estatístico.

Em conclusão, por conta do fenômeno de interação entre partes diferentes do genoma, estamos ainda muito distantes do “cálculo sagrado” da Genética, que seria a possibilidade de prever, com grande precisão, todas as características de um organismo simplesmente conhecendo a sequência de nucleotídeos de seu genoma. Podemos ousar afirmar que essa será uma tarefa até mesmo impossível dada a natureza combinatória das possibilidades de interação que existem. Tal conclusão, ao invés de ser frustrante, pode ser edificadora, pois nos permite assumir que cada indivíduo é uma combinação única e, portanto, preciosa.

**Arquitetura genética** - número total de regiões gênicas envolvidas em um fenótipo, seus efeitos diretos (aditivos), interações intralocus, interlocus e destas com o ambiente. A arquitetura genética é uma propriedade de uma população em uma dada geração, visto que tanto a composição genética quanto o ambiente podem mudar ao longo do tempo.

O genoma humano possui 23 pares de cromossomos e cada um deles é uma molécula longa de DNA, constituída por unidades fundamentais, os nucleotídeos, que por sua vez são constituídos por um açúcar (uma pentose), um grupo fosfato e uma base nitrogenada (adenina, timina, guanina ou citosina). Na molécula de DNA as bases estão emparelhadas duas a duas e geralmente os cientistas se referem ao número de pares de bases para descrever o tamanho do genoma de uma espécie. Em uma célula haploide (por exemplo, um gameta) em humanos, teremos em seu núcleo 3,2 bilhões de pares de bases e, se considerarmos uma célula somática, considerando as duas cópias (materna e paterna) de cada cromossomo, teremos o dobro disso, ou seja 6,4 bilhões de pares de base. Entre cada duas pessoas na população, 99,6% desses pares são iguais, em média. Isso significa que quase 13 milhões de pares de bases são diferentes entre duas pessoas. Essas diferenças no genótipo podem se refletir em diferenças nos fenótipos das pessoas, inclusive propensão a doenças. As diferenças nas bases estão localizadas em certos pontos do genoma chamados polimorfismos de nucleotídeo único, abreviadamente SNPs, da sigla em inglês *Single Nucleotide Polymorphisms* (pronuncia-se “isnîps”). Nessas posições há um alelo principal, que é a base mais frequente nessa posição para uma certa população, e um alelo alternativo, que é o alelo menos frequente (é raro o SNP com mais de dois alelos em uma certa população). Uma característica importante do SNP é a frequência do alelo alternativo, ou MAF, do inglês *Minor Allele Frequency*. Quanto maior for a frequência, mais importante é o SNP para estudos populacionais e de doenças de grande prevalência. Convém destacar que, no caso dos SNPs, está se usando o termo “alelo” para indicar cada uma das bases nitrogenadas alternativas, por exemplo A ou G, numa dada posição, ainda que muitos SNPs estejam localizados fora dos genes e essa variação não se enquadre na definição convencional de alelo.

**Quadro 1.**  
O que são os Polimorfismos de Nucleotídeo Único ou SNPs.

**Quadro 2.**

O que são os estudos de Associação de Genoma Inteiro ou GWAS (*Genome Wide Association Studies*).

A intenção dos estudos de GWAS é procurar em todo o genoma quais pontos estão contribuindo para a variação de algum fenótipo em particular, tal como a propensão a uma determinada doença. Esses estudos, ao contrário do que o nome indica, não utilizam o genoma inteiro. Eles analisam somente os SNPs que estão espalhados ao longo do genoma, pois sequenciar genomas completos ainda é um processo caro. Tais estudos procuram diferenças entre genótipos de pessoas que apresentam um fenótipo de interesse (grupo “casos”) e aquelas que não o apresentam (grupo “controle”). Esses estudos utilizam milhares de casos e controles para comparação.

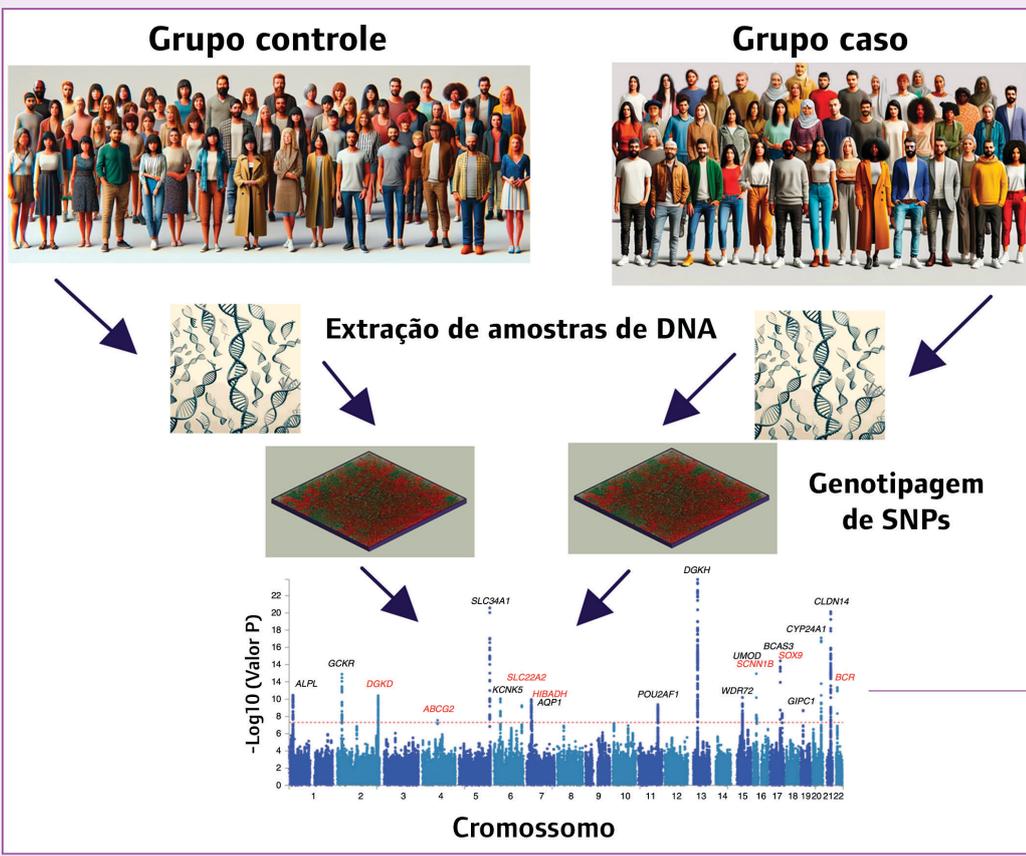
Para apoiar os GWAS, empresas de biotecnologia fabricam plataformas de genotipagem, que permitem saber qual alelo está presente em cada um dos SNP amostrados. Essas plataformas podem conter sondas que avaliam centenas de milhares ou até milhões deles simultaneamente. Na prática, essas plataformas de genotipagem são arranjos bidimensionais de fragmentos de DNA ligados a uma lâmina de vidro que correspondem a sondas capazes de hibridar com cada uma das duas sequências de bases possíveis para cada SNP. Por causa do arranjo bidimensional desses pontos em uma lâmina, por analogia com os “chips” eletrônicos, essas plataformas também são chamadas de “chips de DNA”. Os autores do GWAS aplicam, em cada um desses “chips”, o material biológico de cada indivíduo, caso ou controle, que frequentemente é DNA extraído da saliva ou de células do interior da bochecha. Essas plataformas retornam, então, à empresa, que as submete a um tratamento químico e computacional, apresentando, posteriormente, o resultado aos pesquisadores, que é basicamente o conjunto de dados correspondentes aos alelos presentes em cada SNP de cada caso e cada controle. Em posse desses dados, é aplicado um tratamento estatístico que verifica a probabilidade de que a proporção de alelos seja a mesma entre casos e controles para cada SNP, ou seja, informa se essa proporção difere entre os grupos.

Os resultados dos GWAS são tipicamente mostrados sob a forma de um **plote de Manhattan**, em que a posição cromossômica de cada SNP é relacionada com o valor negativo do logaritmo da sua probabilidade de ser semelhante entre casos e controles. Por exemplo, se a probabilidade de que as proporções sejam iguais for alta, tal como 80%, o valor de  $-\text{Log}_{10}(0,80)$  é aproximadamente 0,10. Entretanto, se o valor for baixíssimo, tal como 0,000000001, ou um em um bilhão, o valor do seu  $-\text{Log}_{10}$  é 9. Considera-se, então, que os SNPs que estão próximos a genes que seguramente influenciam o fenótipo são aqueles que têm uma probabilidade muito baixa de estarem por acaso em quantidades diferentes nos casos e controles.

Em resumo, um estudo de GWAS indica, por meio dos SNPs neles localizados, quais as regiões genômicas com alta probabilidade de conter variações ou genes que influenciam o fenótipo em estudo.

**Plote Manhattan** - é a maneira gráfica usual de se mostrar as posições onde se encontram os SNPs que estão associados a um fenótipo a partir de um GWAS. No eixo das abscissas encontram-se as posições dos cromossomos humanos e, nas ordenadas, os valores do logaritmo negativo do valor da probabilidade de que as proporções de alelos sejam iguais entre casos e controles. É chamado assim porque se parece com a silhueta da ilha de Manhattan, em Nova Iorque (EUA), onde edifícios muito altos destacam-se em um horizonte plano.

**Plote de Manhattan.** Sarah A. Howles, Akira Wiberg, Michelle Goldsworthy, Asha L. Bayliss, Anna K. Gluck, Michael Ng, Emily Grout, Chizu Tanikawa, Yoichiro Kamatani, Chikashi Terao, Atsushi Takahashi, Michiaki Kubo, Koichi Matsuda, Rajesh V. Thakker, Benjamin W. Turney & Dominic Furniss, CC BY 4.0 <<https://creativecommons.org/licenses/by/4.0/>>, via Wikimedia Commons

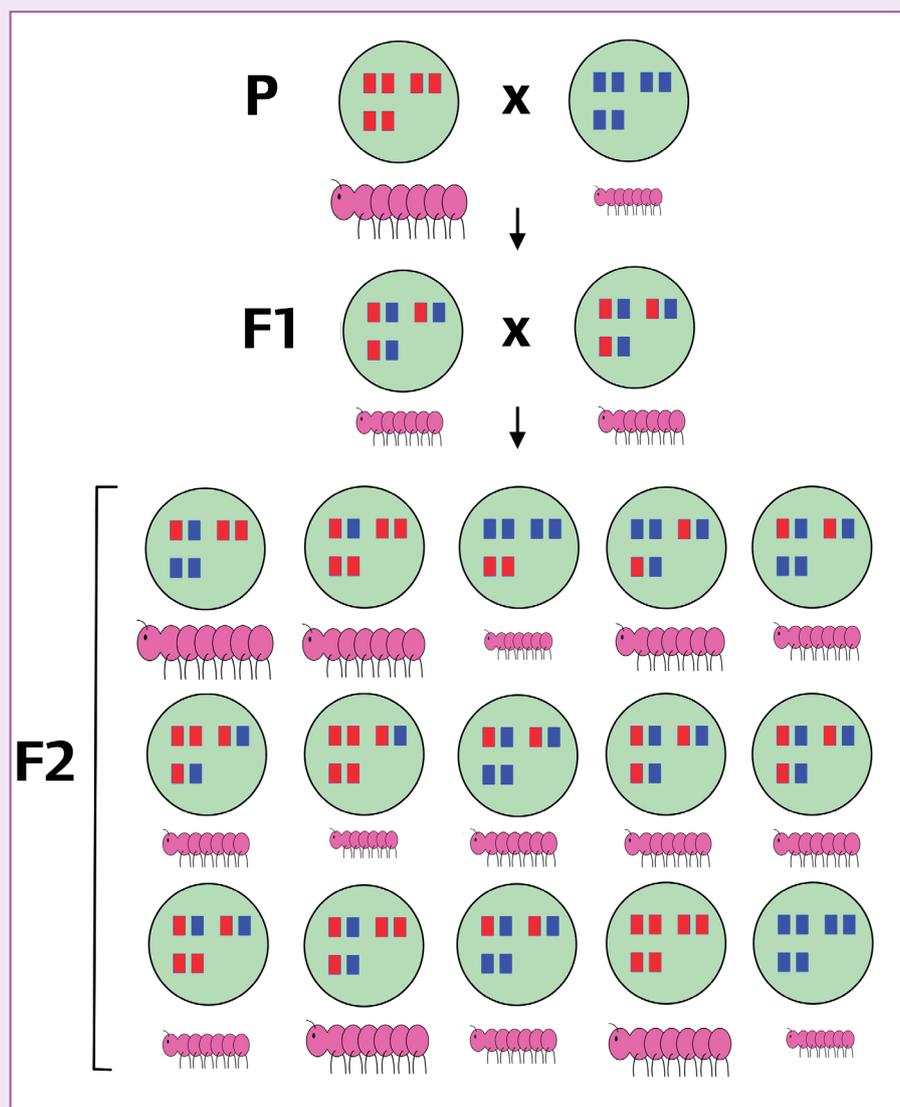


**Quadro 3.**

O que são os estudos de Locus de Caracteres Quantitativos ou QTL (*Quantitative Trait Loci*)?

Estudos de QTL são elaborados para identificar regiões do genoma que contribuem para características quantitativas em linhagens de animais ou plantas que podem ser cruzados entre si. Esses estudos são importantes na compreensão da contribuição genética para variações em características complexas, ou seja, se as diferenças fenotípicas encontradas na população de estudo são devidas a poucos lócus de grande efeito ou de muitos lócus, cada um com pouco efeito no fenótipo. A abordagem em questão é frequentemente aplicada em setores como agricultura e melhoramento animal, oferecendo aos pesquisadores a capacidade de identificar regiões no genoma associadas a características desejáveis, como resistência a doenças ou maior produtividade.

Nos estudos de QTL são realizados cruzamentos planejados, envolvendo a combinação de duas linhagens de organismos (sejam plantas, animais etc.) que apresentam diferenças marcantes na característica quantitativa de interesse. Utilizando o mapeamento genético, que emprega marcadores moleculares como SNPs e microssatélites (regiões pequenas do genoma que consistem em repetições curtas de sequências de bases como, por exemplo, ACACACAC, que são muito variáveis em número de repetições), os descendentes dos cruzamentos são analisados para identificar variações fenotípicas associadas à variação genética herdada. A análise não apenas possibilita a busca por genes com efeitos individuais na característica, mas também permite a investigação da associação entre duas regiões genéticas que explicam a variação observada na característica. Dessa forma, é possível explorar QTLs epistáticos, compreendendo interações complexas entre diferentes lócus gênicos. A diferença fundamental do estudo de QTL é o planejamento dos cruzamentos que permitirão o embaralhamento entre os genomas dos indivíduos com características e genótipos contrastantes, conforme exemplo abaixo. A genotipagem e passos seguintes são similares ao GWAS.



Esquema de cruzamentos para um estudo de QTLs, no qual dois animais parentais hipotéticos apresentam fenótipos contrastantes (tamanho grande e pequeno). Os retângulos coloridos representam regiões do genoma, em que cores iguais significam homocigose para aquele gene e cores diferentes representam heterocigose. Assim, os parentais são contrastantes não somente no fenótipo, mas também no genótipo, em que cada um apresenta combinação homocigótica diferente do outro. A geração F1 terá a contribuição do genoma dos parentais contrastantes e todos os descendentes serão heterocigóticos, assim, na relação genótipo/fenótipo, todos terão o mesmo genótipo/fenótipo. A partir da geração F2, devido à recombinação gênica, os filhotes poderão ter variação no genótipo que estará associada à variação no fenótipo. A partir dessa variação será possível a busca por QTLs individuais ou epistáticos.

## Atividades práticas publicadas na revista *Genética na Escola* sobre o conceito de epistasia

CAPELLI, L. P.; SILVEIRA, R. V. M. O “X” da questão: a diversidade gerada pela segregação cromossômica independente e a cor da pelagem de gatos. *Genética na Escola* v. 4, n. 1, p. 17-24, 2009.

HAHN, E. C.; SCHIENGOLD, M. Coloração da pelagem canina: integrando conceitos básicos de genética clássica. *Genética na Escola* v. 12, n. 1, p. 44-57, 2017.

HEPP, D.; CORSO, J.; LISBOA, C. P. Os genes de coloração MC1R e ASIP como modelos para o

ensino de genética – dominância e interação gênica. *Genética na Escola* v. 12, n. 2, p. 176-181, 2017.

LEITE, L. M.; FERRO, A. R.; SAMPAIO, L. F.; CAPARROZ, R. Dominó gênico: interagindo para compreender a interação gênica. *Genética na Escola* v. 9, n. 1, p. 30-37, 2014.

MIYAKI, C. Y.; MORI, L.; ARIAS, M. C. Dos genes aos fenótipos. *Genética na escola* v. 2, n. 1, p. 10-13, 2006.



### Agradecimentos

Os autores agradecem aos professores Alan R. Templeton e James M. Cheverud pelos ensinamentos que receberam, direta ou indiretamente, sobre o fascinante fenômeno que é a epistasia, às Dras. Regina Célia Mingroni Netto e Eliana Maria Beluzzo Dessen pelas sugestões e à Srta. Lara B. Matioli pela ajuda com as figuras. Alguns trechos foram editados e adequados a partir de consultas ao ChatGPT (<https://openai.com/blog/chatgpt>). Algumas figuras também foram editadas a partir de figuras geradas pelo modelo de linguagem Dal-E, da plataforma ChatGPT.

### Para saber mais

BATESON, W. *Mendel's Principles of Heredity*. Cambridge University Press, Cambridge, UK/London/New York, 1909. (facsimile em <https://www.biodiversitylibrary.org/bibliography/46238>)

CHEVERUD, J. M.; ROUTMAN, E. J. Epistasis and its contribution to genetic variance components. *Genetics*, v.139, n. 3, p. 1455–1461, 1995. (<https://doi.org/10.1093/genetics/139.3.1455>)

FISHER, R. A. The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, v.52, p. 399-433, 1918. (<https://doi.org/10.1017/0080456800012163>)

MINGRONI-NETTO, R. C. Dominante ou recessivo? *Genética na Escola* v. 7, n. 2, p. 28-33, 2012.