

CAPÍTULO 3

Pais desconhecidos em pedigrees de animais domésticos: abordagem e metodologias aplicadas

Daniele Alves, Saymon Rocha, Thais Machado, Jeriel Dias, Heloísa Furlani, Kettlen Neves, Sirlene Lazaro, Hinayah Rojas de Oliveira, Thaise Pinto de Melo

<https://doi.org/10.69570/mp.978-65-84548-39-8.c3>

Resumo

Informações perdidas no pedigree podem reduzir as acurácias de avaliações genéticas em animais domésticos. Algumas metodologias que incluem informação genômica têm auxiliado a melhorar as acurácias dos valores genéticos nestas situações. Métodos como metafundadores (MFs), grupos de pais desconhecidos (UPGs) e pedigrees truncados têm sido combinados à informação genômica, com resultados positivos em algumas espécies de animais domésticos como solução a problemática dos pais desconhecidos. Este capítulo tem como objetivo apresentar as bases teóricas e exemplos de aplicações práticas utilizando as metodologias dos UPGs, metafundadores e pedigrees truncados, incluindo um tutorial passo a passo de como realizar estas metodologias nos programas da família BLUPF90: <https://github.com/danielealves-sys/TutorialUpgMfTrunc>.

Palavras chave: avaliação genética, BLUPF90, grupos de pais desconhecidos, metafundadores, pedigree truncado.

1. Introdução

A avaliação genética tem sido primordial nos programas de melhoramento genético, evoluindo constantemente com o acesso a novas tecnologias. Inicialmente, as avaliações genéticas eram baseadas somente em dados fenotípicos e de pedigree, o que apesar de eficazes, apresentavam algumas limitações, principalmente quando confrontadas com dados genealógicos incompletos ou erros de coleta de dados (Pimentel *et al.*, 2024).

A inconsistência de informações genealógicas, como em pedigrees incompletos, pode comprometer a construção das matrizes **A** e **H**, e, conseqüentemente, a precisão dos valores genéticos dos animais (Masuda *et al.*, 2022). Atualmente, diferentes estratégias têm sido utilizadas para contornar essa limitação, como o uso de metafundadores (MFs). Segundo Christensen (2012), o conceito de MFs propõe a inclusão de uma matriz de covariância entre fundadores, possibilitando que eles compartilhem a ancestralidade parcial, o que difere da suposição tradicional onde os fundadores não são relacionados.

Além disso, outra solução amplamente utilizada é uso de *Unknown Parent Groups* (UPGs) proposta por Quaas (1988), a qual modela efeitos genéticos médios para animais com registros parentais desconhecidos, utilizando como base o ano de nascimento, sexo ou origem geográfica. O uso de pedigrees truncados também pode ser utilizado como uma estratégia para lidar com a problemática de pedigrees perdidos, o qual consiste em remover animais de gerações mais antigas.

Nesse sentido, Lourenco *et al.* (2014) afirmam que ao utilizar fenótipos dos últimos dez anos ou das últimas três gerações, podemos reduzir o viés de seleção acumulado ao longo do tempo nos dados históricos, tornando as predições genéticas para os animais candidatos a seleção mais acuradas e menos influenciadas por informações ancestrais com erros ou ausentes (Bussiman *et al.*, 2023). Entender a teoria, as aplicações e os desafios associados ao uso de pedigree incompleto é um importante passo para melhorar os modelos de avaliação genética em animais domésticos.

2. Grupos de pais desconhecidos (UPGs)

Os UPGs podem ser definidos como "*grupos de pais fantasmas*" ou uns ancestrais representativos. Quando o pai ou a mãe de um animal é desconhecido, ao em vez de manter a informação perdida assumindo que ele vem de uma base genética antiga com valor genético zero, este indivíduo é incluído em um grupo. Esse grupo representa a média genética esperado para aquele grupo de pais e/ou mães desconhecidos.

Modelos que ignoram os UPGs tendem a subestimar severamente o progresso genético real da população, comprometendo o monitoramento do programa e as projeções econômicas (Meyer, 2021). A metodologia não apenas aumenta a acurácia das previsões, mas também reduz drasticamente o viés sistemático, especialmente para animais jovens (Bradford *et al.*, 2019).

A principal consequência da inclusão das UPGs é a correção da tendência genética nas estimativas. O efeito dos UPGs sobre os valores genéticos é estatisticamente e quantitativamente relevante (Misztal *et al.*, 2013), sobretudo em populações com animais jovens com pedigrees incompletos. Segundo Masuda *et al.* (2022), a ausência de pedigree pode introduzir vies nas previsões genômicas pelo ssGBLUP. A inclusão de UPG corrige parte dessas distorções, ajustando diferenças genéticas médias entre animais sem ascendência informada e reduzindo tendências genéticas. No entanto, os autores destacam que, em bases de dados pequenas ou incompletas, o modelo pode subestimar os efeitos dos UPG.

Na prática, o pedigree passa a apresentar uma codificação para o UPG que fica no lugar do Pai desconhecido. Esse UPG se torna um ponto de partida na rede de parentesco, conectando os animais com pais desconhecidos através de uma média estimada para os UPGs. A Figura 1 ilustra tais relações.

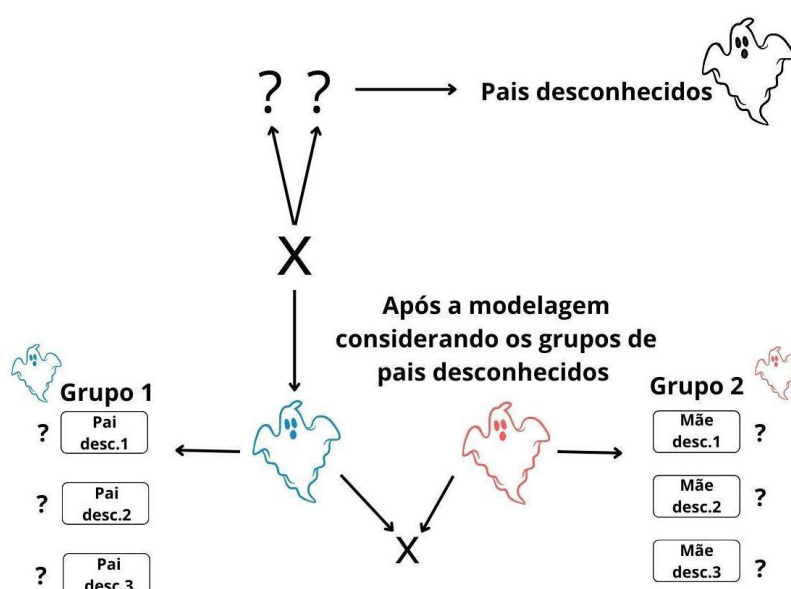


Figura 1. Representação esquemática dos UPGs, que substituem os pais ausentes no pedigree e conectam os animais por meio de médias genéticas estimadas para cada grupo.

A construção dos UPG é uma das primeiras etapas da análise e deve levar em conta possíveis diferenças entre os indivíduos da população. O objetivo nesta etapa é construir UPGs mais homogêneos possíveis. Ao avaliar uma população multirracial de bovinos de corte, por exemplo, os UPGs podem ser construídos combinando as informações do sexo (no caso de características avaliadas em ambos os sexos), ano/geração de nascimento e composição racial, como segue:

a) UPG1: MACHO_BRAFORD_2000-2010: Representa o valor genético médio de todos os pais desconhecidos da raça Braford, nascidos entre 2000 e 2010.

b) UPG2: FEMEA_NELORE_2011-2015: Representa o valor genético médio de todas as mães desconhecidas da raça Nelore, nascidas entre 2011 e 2015.

A inclusão de UPGs modifica o modelo animal clássico, conforme descrito por Legarra *et al.* (2014). Assim, o modelo ajustado se torna:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{ZQg} + \mathbf{Za}^* + \mathbf{e}$$

Neste modelo, $\mathbf{y}$ é o vetor de observações, $\boldsymbol{\beta}$ é o vetor de efeitos fixos, $\mathbf{X}$ é a matriz de incidência dos efeitos fixos que relaciona as observações no vetor $\mathbf{y}$ aos seus respectivos níveis de efeitos fixos no vetor $\boldsymbol{\beta}$, $\mathbf{Z}$ é a matriz de incidência dos efeitos aleatórios, relaciona as observações no vetor $\mathbf{y}$ aos animais correspondentes no vetor $\mathbf{a}$; e $\mathbf{e}$ é o vetor de erros ou resíduos aleatórios. Assume-se que $\mathbf{e}$ segue uma distribuição normal com média zero. O valor genético dos animais ($\mathbf{a}_i$) é decomposto em efeito médio do grupo ao qual seu pai desconhecido pertence ($\mathbf{g}_j$) e o desvio do indivíduo em relação a essa média ($\mathbf{a}_i^*$).

Para solucionar as equações de modelo misto, a inversa da matriz de parentesco é modificada para acomodar os UPGs, tratando-os como "animais fantasmas" sem variância de segregação Mendeliana, conforme a transformação de Quaas-Pollak (Quaas, 1988). A inversa da matriz de covariâncias para os efeitos genéticos e de grupo ($\mathbf{A}^{*-1}$) é então construída da seguinte forma:

$$\mathbf{A}^{*-1} = \begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{Q} \\ -\mathbf{Q}'\mathbf{A}^{-1} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} \end{bmatrix}$$

Nesta matriz, $\mathbf{A}^{*-1}$ é a inversa da matriz de parentesco por pedigree de todos os animais, e $\mathbf{Q}$ é a matriz de incidência que aloca a fração de cada ancestral desconhecido de um indivíduo ao seu respectivo UPG. Esta estrutura permite que o modelo ajuste corretamente o valor genético do animal, considerando a média do seu grupo de origem.

A inclusão da informação genômica nas avaliações genéticas têm o potencial de melhorar as estimativas dos efeitos dos UPGs. Ao incluir a informação genômica junto aos UPGs no método Single-Step GBLUP (ssGBLUP) por exemplo, a matriz de parentesco baseada no pedigree ($\mathbf{A}$) é combinada com a matriz de parentesco genômico ($\mathbf{G}$), que é calculada a partir dos efeitos marcadores (SNPs) (Vanraden, 2008). O resultado é uma matriz unificada, a matriz $\mathbf{H}$, cuja inversa ($\mathbf{H}^{-1}$) substitui a $\mathbf{A}^{-1}$ nas equações do modelo misto. A forma geral da $\mathbf{H}^{-1}$ é dada por (Aguilar *et al.*, 2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

Onde $\mathbf{A}_{22}$ é um subconjunto da matriz $\mathbf{A}$, considerando apenas os animais genotipados. Ao considerar os UPGs nas análises, a informação dos mesmos precisa ser incluída em ambas as matrizes, do pedigree tradicional ($\mathbf{A}^{-1}$) e do pedigree apenas dos animais genotipados ($\mathbf{A}_{22}^{-1}$), para garantir a compatibilidade entre as bases genética e genômica. A abordagem mais comum e robusta, conhecida como *altered QP model* (Masuda *et al.*, 2022) inclui o UPG na matriz $\mathbf{A}$ e traz a seguinte modificação para a $\mathbf{H}^{-1}$:

$$\mathbf{H}_{UPG}^{-1} = \mathbf{A}_{UPG}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22,UPG}^{-1} \end{bmatrix}$$

Nesta equação, $\mathbf{A}_{22}^{-1}$ e $\mathbf{A}_{22,UPG}^{-1}$ são as inversas das matrizes de parentesco ajustadas para os UPGs. Alternativamente, outra abordagem (ssGBLUP/H_UPG) consiste em incluir os UPGs em todos os elementos da matriz $\mathbf{H}$. A sua inversa ($\mathbf{H}^{-1}$) é construída da seguinte forma, conforme detalhado por Masuda *et al.* (2022):

$$\mathbf{H}_{QP}^{-1} = \mathbf{A}_{UPG}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} & -(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2 \\ \mathbf{0} & -\mathbf{Q}_2'(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1}) & \mathbf{Q}_2'(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2 \end{bmatrix}$$

Onde: $\mathbf{A}_{22}^{-1}$ é a inversa da matriz de parentesco por pedigree já ajustada para os UPGs; $\mathbf{G}^{-1}$ e $\mathbf{A}_{22}^{-1}$ são as inversas das matrizes de parentesco genômico e por pedigree (apenas para animais genotipados), respectivamente; $\mathbf{Q}_2$ é a submatriz de alocação que relaciona os animais genotipados aos seus respectivos UPGs. A principal diferença desta matriz para a anterior é a inclusão dos termos $-(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2$ e $\mathbf{Q}_2'(\mathbf{G}^{-1} - \mathbf{A}_{22}^{-1})\mathbf{Q}_2$, que explicitamente modelam a covariância entre os valores genéticos dos animais genotipados e os efeitos dos UPGs.

3. Uso de metafundadores em avaliação genética com pais desconhecidos

Tradicionalmente, a utilização dos UPGs tem sido uma das principais estratégias adotadas para representar ancestrais desconhecidos em pedigrees de animais domésticos. Entretanto, sua principal limitação reside na suposição de independência entre os grupos formados, desconsiderando possíveis relações de parentesco. Visando superar essa restrição, podemos aplicar a metodologia dos metafundadores (MF), que expande a teoria dos UPG ao incorporar informações de relacionamento e endogamia entre populações fundadoras, configurando-se como uma alternativa promissora para aprimorar a compatibilidade entre as matrizes de parentesco ($\mathbf{A}$) e matriz genômica ($\mathbf{G}$) (Masuda *et al.*, 2022).

A Figura 2 ilustra, de forma esquemática, o conceito de metafundador e como os mesmos são considerados nos pedigrees. No diagrama, o “Metafundador” representa uma população ancestral hipotética, conectada aos indivíduos da população base (animais 2 a 6) por linhas pontilhadas que caracterizam um pseudo-pedigree, isto é, uma relação genética estimada a partir de frequências alélicas, sem registro genealógico real. A partir dessa base, o fluxo de descendência segue pelas gerações seguintes (animais 7 a 13), evidenciando como a inclusão do metafundador permite modelar a variabilidade genética original e corrigir a falta de informações completas do pedigree. Essa representação facilita compreender como os MFs incorporam a estrutura da população fundadora na matriz de parentesco, melhorando as acurácias de predição dos valores genéticos estimados.

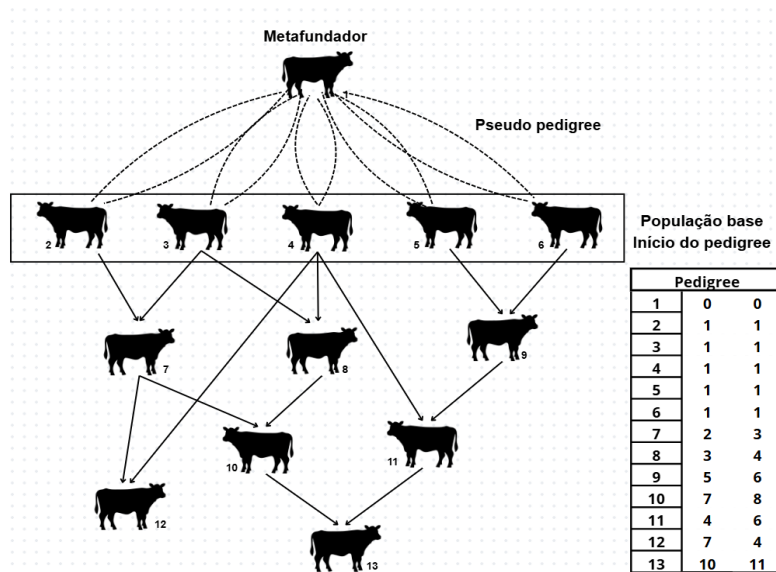


Figura 2. Representação esquemática de uma população de bovinos com metafundador. Adaptado de Legarra *et al.* (2015).

Os MF podem ser definidos como pseudoindivíduos que representam os ancestrais base de uma população. Em contraste com os UPGs tradicionais, que são considerados efeitos fixos no modelo estatístico, os MFs apresentam vantagens, pois permitem a modelagem da endogamia e das relações genéticas entre grupos de base, sendo tratados como efeitos aleatórios. Além disso, promovem ajustes na matriz de parentesco **A**, de modo a torná-la compatível com a matriz genômica **G**, considerando frequências alélicas baseadas em 0,5 (Aguilar *et al.*, 2010; Christensen, 2012).

No ssGBLUP, a matriz de relacionamento **H** integra pedigree e genômica, sendo escrita na forma clássica (sem metafundador) como:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

Em que **A** é a matriz de parentesco do pedigree, **A₂₂** o bloco entre genotipados e **G** a matriz genômica. Quando se utiliza metafundadores, a base ancestral passa a considerar parentesco entre fundadores e **A** é substituída por **A***, incorporando a matriz de covariância **Γ** (Gama), conforme descrito em Legarra *et al.* (2015) e Kluska *et al.* (2021) e :

$$\mathbf{\Gamma} = \mathbf{A}^{*-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{*-1} \end{bmatrix}$$

Segundo Masuda *et al.* (2022), a versão sem MF assume fundadores não relacionados, podendo gerar incompatibilidade entre **A** e **G** em pedigrees incompletos, enquanto a inclusão de MFs ajusta essa base, reduz viés e melhora a consistência das estimativas genômica. O elemento central dessa abordagem é a matriz **Γ**, que descreve as relações dentro e entre MFs. Essa matriz apresenta interpretações específicas:

- $\Gamma_{ii} < 1$: indica uma população base com elevada variabilidade genética;
- $\Gamma_{ii} > 1$: caracteriza uma população base endogâmica;
- $\Gamma_{ij} > 0$: reflete sobreposição genética entre populações fundadoras;
- $\Gamma_{ij} < 0$: sugere divergência entre populações fundadoras.

A estimação de **Γ** é realizada a partir de dados genotípicos, geralmente por meio de métodos de mínimos quadrados generalizados (*Generalized Least Squares* – GLS). Para garantir acurácias adequadas com o uso da metodologia, é necessário um número adequado de animais genotipados que apresentem conexão com os metafundadores.

Uso de MFs tem sido aplicado em diferentes espécies, especialmente em estudos com populações multirraciais de bovinos de corte (Junqueira *et al.*, 2020; Tonussi *et al.*, 2021) e suínos (Aldridge *et al.*, 2020; Fu *et al.*, 2021). Neste sentido, a utilização de MFs tem trazido maior acurácia de seleção em cenários multirraciais e envolvendo cruzamentos. No entanto, apesar desses avanços, a aplicação prática ainda enfrenta desafios relevantes. Em especial, a estimação da matriz **Γ**, que representa as covariâncias entre os MFs, pode resultar em valores pouco plausíveis do ponto de vista biológico, sobretudo em populações mais complexas (Kudinov *et al.*, 2022). Além disso, a obtenção de estimativas robustas requer quantidade adequada de dados genômicos, já que a baixa conectividade entre indivíduos pode comprometer os resultados.

Outro aspecto importante refere-se ao escalonamento de variâncias, visto que, em alguns contextos, não é recomendado estimar diretamente os componentes de variância com o uso de MFs (Legarra *et al.*, 2015). O conceito

de MFs representa um avanço teórico e aplicado nas avaliações genômicas, ao possibilitar maior integração entre informações de pedigree e genômica, favorecer o tratamento de pedigrees incompletos e proporcionar incremento na acurácia das predições genéticas.

4. Pedigrees truncados em avaliações genética de animais domésticos

Em avaliações genômicas, nem sempre é necessário utilizar todas as informações históricas disponíveis do animal ou população analisada. Nesse contexto, o truncamento de pedigrees se apresenta como uma forma de limitar a quantidade de informações trabalhadas, mantendo apenas as mais significativas, visto que, pedigrees incompletos ou populações cuja estrutura é incerta podem comprometer as avaliações genéticas (Meyer, 2021). Assim, o truncamento de pedigrees consiste na remoção de informações antigas, desnecessárias ou incompletas do pedigree utilizado na avaliação (Bussiman *et al.*, 2023), reduzindo vieses e otimizando os custos computacionais sem comprometer a acurácia das predições genéticas. Essa manobra é útil principalmente quando não há conhecimento sobre a genealogia dos progenitores dos animais fundadores da população. Nestes casos, utilizar o truncamento a partir de gerações mais novas pode ser mais vantajoso do que utilizar gerações mais antigas com informações inconsistentes.

Estudos em diferentes espécies mostram que o truncamento pode contribuir no aumento dos ganhos genéticos, reduzindo vieses nas predições dos valores genéticos. Cesarani *et al.* (2021), avaliando bovinos de leite, concluíram que, a combinação do truncamento do pedigree com informações genômicas de UPG trouxe vantagens em termos de predição. Em suínos, Pocrnic *et al.* (2017) mostrou que reduzir a profundidade do pedigree para cerca de 2-5 gerações não afetou a acurácia das predições genômicas para animais genotipados. Já em búfalos, o truncamento combinado a dados genômicos pode melhorar a acurácia e reduzir vieses (Melo *et al.*, 2024). Além disso, o truncamento pode atuar na redução de viés associados a mudanças na definição de características ao longo do tempo, causadas por alterações de manejo dos animais, ambiente ou metodologias de mensuração (Tsuruta *et al.*, 2005; Howard *et al.*, 2018).

Em bubalinos da raça Murrah, Melo *et al.* (2024), observam que o número de gerações utilizadas depende da estrutura da população e da característica avaliada. Já em bovinos de leite e suínos, Lourenco *et al.* (2014). afirma que a utilização de cerca de duas gerações foi suficiente para manter, e em alguns casos até melhorar, a acurácia das predições. Tal característica é particularmente interessante em espécies com intervalos de geração mais curtos, em que a informação das gerações mais antigas têm menor relevância (Mäntysaari *et al.*, 2020).

Outro aspecto a ser levado em consideração nas avaliações genéticas é o elevado processamento computacional requerido, visto que, o uso de grandes volumes de dados aumenta consideravelmente o custo das análises genômicas, especialmente em populações intensamente genotipadas. Nessa perspectiva, Bussiman *et al.* (2023) destaca que o truncamento reduz o número de equações a serem resolvidas, o que diminui os custos computacionais. Tal prática foi aplicada com sucesso em estudos com bovinos de leite, suínos e ovinos de leite (Lourenco *et al.*, 2014; Howard *et al.*, 2018; Hollifield *et al.*, 2021; Macedo *et al.*, 2022). Contudo, a decisão sobre quantas gerações manter deve considerar fatores como herdabilidade da característica, número de registros e intervalo de gerações, contribuindo para eficiência e confiabilidade dos programas genéticos, qualidade das informações, custo computacional e acurácia das predições genéticas.

5. O Papel da genômica na resolução de pedigrees incompletos

Historicamente, os programas de melhoramento genético dependiam quase que exclusivamente de informações fenotípicas e de registros genealógicos para estimar o mérito genético dos animais domésticos. Durante décadas, essa abordagem foi a base do melhoramento genético, possibilitando avanços importantes na produção de animais geneticamente superiores com as características desejadas (Çelik, 2024). No entanto, a principal limitação dessa prática estava diretamente relacionada à qualidade e à correta coleta das informações de pedigree. Em muitos casos, a falta de registros parentais ou erros na coleta de dados compromete a confiabilidade das estimativas genéticas, levando a interpretações equivocadas da contribuição genética dos indivíduos

(Pimentel *et al.*, 2024). Nesse sentido, Wolak e Reid (2017) destacam que, em populações base, é difícil encontrar pedigrees completos, resultando em conexões genealógicas incertas. Apesar de levantar tal discussão para populações de animais silvestres/selvagens, tal premissa se aplica também a populações de animais domésticos. De forma semelhante, Speed e Balding (2015) apontam que a falta de informações completas nos registros implica em coeficientes de parentesco imprecisos, prejudicando cálculos de endogamia e dificultando a distinção entre efeitos genéticos e ambientais. Essa fragilidade afeta não apenas a seleção dos melhores animais, mas também a previsão do ganho genético esperado, que passa a ser subestimado em função da ausência de vínculos genealógicos confiáveis (Gómez *et al.*, 2021).

Com os avanços da biologia molecular, novas ferramentas começaram a ser incorporadas aos programas de melhoramento genético. O uso de marcadores moleculares clássicos, como microssatélites e RFLPs, que permitem identificar associações entre regiões específicas do genoma e características quantitativas de interesse. Esses marcadores possibilitaram o mapeamento de QTLs (*Quantitative Trait Loci*) especialmente para características quantitativas, facilitando a identificação de genes responsáveis por variações genéticas de interesse (Meuwissen *et al.*, 2001). A seleção genômica, por outro lado, significativamente impactou no aumento da acurácia das predições genéticas. Koning (2016) descreve essa abordagem com base na utilização de dados genotípicos e fenotípicos da população de referência para construir equações de previsão do mérito genético dos indivíduos candidatos à seleção. Dessa forma, foi possível avaliar animais jovens sem depender de informações de progênie ou apenas de registros fenotípicos individuais. Ibtisham *et al.* (2017) reforçam que esse benefício da seleção genômica foi fundamental, pois permitiu reduzir o intervalo de gerações e aumentar a taxa de ganho genético. Além disso, a avaliação pode ser aplicada inclusive em embriões, antecipando decisões de seleção que antes dependiam de longos períodos de observação. Especialmente no caso de pedigrees incompletos, Tonussi *et al.* (2017) ressaltam que, com informações genômicas, é possível identificar relações genéticas mesmo entre animais cujo pedigree é desconhecido ou

incerto, otimizando estratégias de acasalamento e aumentando a rentabilidade do rebanho.

Entre os modelos estatísticos que mais se destacam na incorporação de informações genômicas está o GBLUP (*Genomic Best Linear Unbiased Prediction*), que substitui a matriz de parentesco tradicional, derivada do pedigree, pela matriz genômica construída a partir de das informações dos marcadores moleculares (Boichard *et al.*, 2016). Posteriormente, Misztal *et al.* (2009) desenvolveram o método ssGBLUP (*Single-Step GBLUP*), que combina em um procedimento de único passo a informação de todos os animais disponíveis, sejam eles genotipados ou não. Para calcular o parentesco entre os animais usando esse modelo, é necessário a utilização da matriz $\mathbf{H}^{-1}$, que integra informações das matrizes $\mathbf{A}^{-1}$ e a inversa da matriz de parentesco genômico ($\mathbf{G}^{-1}$) (Legarra *et al.*, 2009; Aguilar *et al.*, 2010; VanRaden, 2020). Enquanto, para estimar os valores genéticos genômicos (GEBVs), é necessário substituir a inversa da matriz $\mathbf{A}^{-1}$ pela inversa da matriz $\mathbf{H}^{-1}$. Apesar de sua ampla utilização e de sua capacidade de reduzir vieses em comparação a outros métodos, o ssGBLUP ainda depende, em certa medida, da consistência dos pedigrees. Quando há muitas informações ausentes, a formação das matrizes $\mathbf{A}$ e $\mathbf{G}$ pode ser afetada, resultando em GEBVs tendenciosos (Masuda *et al.*, 2022).

No contexto da era genômica, o uso de haplótipos também tem se mostrado uma ferramenta interessante para superar as limitações impostas por impostas por pedigrees incompletos no melhoramento genético animal. Os haplótipos são blocos de variações no DNA que são herdados em conjunto, o que proporciona uma visão mais detalhada e precisa das relações dos indivíduos (Legarra *et al.*, 2009). Essa metodologia é valiosa quando pensamos em pedigrees incompletos, situação comum em muitas populações comerciais e de base (Junqueira *et al.*, 2020). A partir da comparação dos haplótipos é possível identificar segmentos genômicos idênticos por descendência (IBD), revelando vínculos genéticos que não foram detectados pelo pedigree tradicional (Périte *et al.*, 2016), conforme mostra a Figura 3.

A integração de informações de haplótipos contribui significativamente para corrigir falhas nas informações parentais, reduzindo o viés nas estimativas genéticas e aumentando a acurácia da predição de mérito genético, o que se

traduz em programas de seleção mais eficientes e confiáveis (Romeiro *et al.*, 2013; Garrick, 2011). Isso permite que programas de seleção superem as barreiras tradicionais impostas por pedigrees incompletos, buscando sempre manter um progresso contínuo nos programas de melhoramento (Hayes *et al.*, 2007).

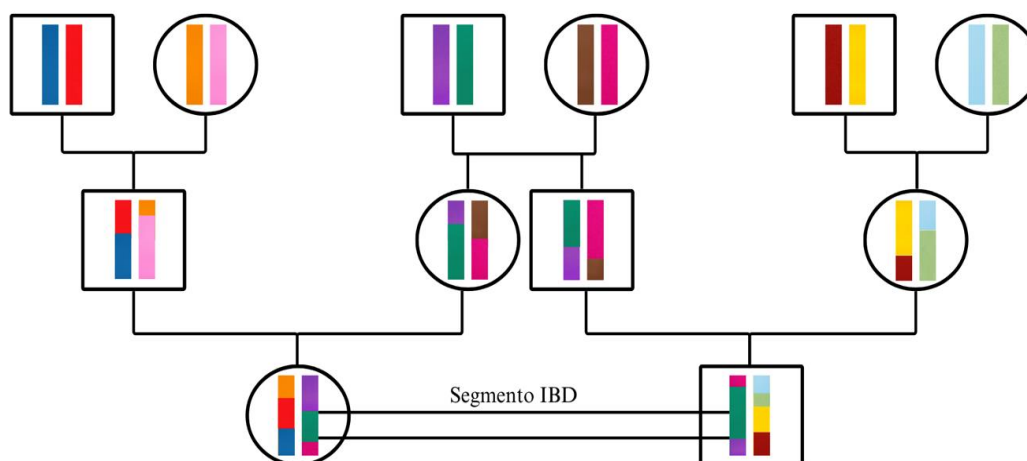


Figura 3. Representação da transmissão de haplótipos em uma genealogia, destacando um segmento IBD herdado de um ancestral comum.

6. Tutoria GitHub:

Link de acesso ao tutorial para o passo a passo para realizar análises BLUP, ssGBLUP, com inclusão de UPG, MFs e sistema de truncamento em avaliações genéticas: <https://github.com/danielealves-sys/TutorialUpgMfTrunc>

7. Considerações finais

As metodologias apresentadas neste capítulo são potenciais e efetivas para solucionar a problemática dos pais desconhecidos em pedigrees de animais domésticos. Cada abordagem oferece uma forma distinta de lidar com essas incertezas, contribuindo para melhorar de maneira significativa as estimativas dos valores genéticos em avaliações tradicional e genômica. Assim a escolha entre utilizar UPGs, MFs ou pedigree truncado deve ser cuidadosamente orientada pelas características específicas da população em estudo.

8. Referências bibliográficas

- AGUILAR, Ignacio *et al.* Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. **J Dairy Sci** 93:743–752. 2010. <https://doi.org/10.3168/JDS.2009-2730>.
- ALDRIDGE, Michael *et al.* Variance estimates are similar using pedigree or genomic relationships with or without the use of metafounders or the algorithm for proven and young animals 1. **J Anim Sci** 98. 2020.<https://doi.org/10.1093/JAS/SKAA019>.
- BOICHARD, Didier *et al.* Genomic selection in domestic animals: Principles, applications and perspectives. **C R Biol** 339:274–277. 2016. <https://doi.org/10.1016/j.crv.2016.04.007>.
- BRADFORD, Heather *et al.* Modeling missing pedigree in single-step genomic BLUP. **J Dairy Sci** 102:2336–2346. 2019. <https://doi.org/10.3168/jds.2018-15434>.
- BUSSIMAN, Fernando *et al.* Boundaries for genotype, phenotype, and pedigree truncation in genomic evaluations in pigs. **J Anim Sci** 101. 2023. <https://doi.org/10.1093/jas/skad273>.
- CESARANI, Alberto *et al.* Genomic predictions for yield traits in US Holsteins with unknown parent groups. **J Dairy Sci** 104:5843–5853. 2021. <https://doi.org/10.3168/jds.2020-19789>.
- CHRISTENSEN, Ole. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. 2012. <https://doi.org/10.1186/1297-9686-44-37>.
- ÇELİK, Şeyda. Bibliometric analysis of genomic selection in breeding of animal from 1993 to 2024: global trends and advancements. **Front Genet** 15. 2024. <https://doi.org/10.3389/fgene.2024.1402140>
- FU, Chuanke *et al.* Single-step genomic evaluation with metafounders for feed conversion ratio and average daily gain in Danish Landrace and Yorkshire pigs. **Genetics Selection Evolution** 53.2021. <https://doi.org/10.1186/s12711-021-00670-x>.
- GARRICK, J. Dorian. The nature, scope and impact of genomic prediction in beef cattle in the United States. **Genetics Selection Evolution** 43. 2011. <https://doi.org/10.1186/1297-9686-43-17>.
- GÓMEZ, Mayra *et al.* Accounting for Genetic Differences Among Unknown Parents in *Bubalus bubalis*: A Case Study From the Italian Mediterranean Buffalo. **Front Genet** 12. 2021. <https://doi.org/10.3389/fgene.2021.625335>.
- HAYES, John *et al.* Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. **Genet Res** 89:215–220.2007. <https://doi.org/10.1017/S0016672307008865>.

- HOLLIFIELD, Mary *et al.* Determining the stability of accuracy of genomic estimated breeding values in future generations in commercial pig populations. **J Anim Sci** 99.2021. <https://doi.org/10.1093/jas/skab085>.
- HOWARD, Jeremy *et al.* The impact of truncating data on the predictive ability for single-step genomic best linear unbiased prediction. **Journal of Animal Breeding and Genetics** 135:251–262.2018. <https://doi.org/10.1111/jbg.12334>.
- IBTISHAM, Fahar *et al.* Genomic selection and its application in animal breeding. **Thai Journal of Veterinary Medicine** 47:301–310. 2017. <https://doi.org/10.56808/2985-1130.2838>.
- JUNQUEIRA, Vinícius *et al.* Applying the Metafounders Approach for Genomic Evaluation in a Multibreed Beef Cattle Population. **Front Genet** 11. 2020. <https://doi.org/10.3389/fgene.2020.556399>.
- KLUSKA, Sabrina *et al.* Metafounders May Reduce Bias in Composite Cattle Genomic Predictions. **Front Genet** 12. 2021. <https://doi.org/10.3389/fgene.2021.678587>.
- KONING, Dirk-Jan *et al.* On genomic selection. **Genetics** 203:5–7.2016. <https://doi.org/10.1534/genetics.116.189795>.
- KUDINOV, Andrei *et al.* Single-step genomic BLUP with many metafounders. **Front Genet** 13. 2022. <https://doi.org/10.3389/fgene.2022.1012205>.
- LEGARRA, Andrés; AGUILAR, Ignacio and MISZTAL, Inacy. A relationship matrix including full pedigree and genomic information. **J Dairy Sci** 92:4656–4663.2009. <https://doi.org/10.3168/jds.2009-2061>.
- LEGARRA, Andrés *et al.* Ancestral relationships using metafounders: Finite ancestral populations and across population relationships. **Genetics** 200:455–468.2015. <https://doi.org/10.1534/genetics.115.177014>.
- LOURENCO, Daniela *et al.* Are evaluations on young genotyped animals benefiting from the past generations? **J Dairy Sci** 97:3930–3942.2014. <https://doi.org/10.3168/JDS.2013-7769>.
- MACEDO, Fernando *et al.* Removing data and using metafounders alleviates biases for all traits in Lacaune dairy sheep predictions. **J Dairy Sci** 105:2439–2452. 2022. <https://doi.org/10.3168/jds.2021-20860>.
- MÄNTYSAARI, Esa; KOIVULA, Minna and STRANDÉN, Ismo. Symposium review: Single-step genomic evaluations in dairy cattle. **J Dairy Sci** 103:5314–5326.2020. <https://doi.org/10.3168/jds.2019-17754>.
- MASUDA, Yutaka, P.M. *et al.* Invited review: Unknown-parent groups and metafounders in single-step genomic BLUP. **J Dairy Sci** 105:923–939. 2022. <https://doi.org/10.3168/jds.2021-20293>.

- MELO, Thaise *et al.* Unknown parent groups and truncated pedigree in single-step genomic evaluations of Murrah buffaloes. **J Dairy Sci** 107:8130–8140. 2024. <https://doi.org/10.3168/jds.2023-24608>.
- MEUWISSEN, T.H.E., B.J. HAYES, and M.E. GODDARD. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. 2001. **Genetics Society of America**. <https://doi.org/10.1093/genetics/157.4.1819>
- MEYER, Karin. Impact of missing pedigrees in single-step genomic evaluation. **Anim Prod Sci** 61:1760–1773. 2021. <https://doi.org/10.1071/AN21045>.
- MISZTAL, Inacy; LEGARRA, Andrés and AGUILAR, Ignacio. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. **J Dairy Sci** 92:4648–4655. 2009. <https://doi.org/10.3168/JDS.2009-2064>.
- MISZTAL, Inacy *et al.* Unknown-parent group in single-step genomic evaluation. *Journal of Animal Breeding and Genetics*, Berlin, v. 130, n. 4, p. 252-258, 1 Aug.213. <https://doi.org/10.1111/jbg.12025>.
- MISZTAL, Inacy *et al.* Current status of genomic evaluation. **Journal of Animal Science**, v.98(4), p. 1-14. 2020. Available at <https://doi.org/10.1093/jas/skaa101>.
- PÉRTILE, Simone *et al.* Seleção e associação genômica ampla para o melhoramento genético animal com uso do método ssGBLUP. **Pesqui Agropecu Bras** 51:1729–1736. 2016. <https://doi.org/10.1590/S0100-204X2016001000004>.
- PIMENTEL, E.C.G. *et al.* How pedigree errors affect genetic evaluations and validation statistics. **J Dairy Sci** 107:3716–3723. 2024. <https://doi.org/10.3168/jds.2023-24070>
- POCRNIC, Ivan *et al.* Technical note: Impact of pedigree depth on convergence of single-step genomic BLUP in a purebred swine population. **J Anim Sci** 95:3391. 2017. <https://doi.org/10.2527/jas2017.1581>.
- QUAAS, Richard. Additive Genetic Model with Groups and Relationships. **J Dairy Sci** 71:1338–1345. 1988. [https://doi.org/10.3168/jds.S0022-0302\(88\)79691-5](https://doi.org/10.3168/jds.S0022-0302(88)79691-5).
- ROMEIRO, Gilberto., *et al.* Genômica Aplicada ao Melhoramento Genético de Gado de Corte. 2013. <https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/980608/1/MelhoramentoGeneticoCapitulo17.pdf> (acessado 16 de setembro de 2025).
- SARMENTO, José *et al.* Desmistificando a avaliação genética animal. **Revista Científica de Produção Animal**. V25 (26–35). 1415-563x. 2023.
- SPEED, Doug and Balding, David. Relatedness in the post-genomic era: Is it still useful? **Nat Rev Genet** 16:33–44. 2015. <https://doi.org/10.1038/nrg3821>.

- TONUSSI, Rafael *et al.* Application of single step genomic BLUP under different uncertain paternity scenarios using simulated data. **PLoS One** 12. 2017. <https://doi.org/10.1371/journal.pone.0181752>.
- TSURUTA, Shogo; MISZTAL, Ignacy and LAWLOR, Thomas. Changing definition of productive life in US holsteins: Effect on genetic correlations. **J Dairy Sci** 88:1156–1165. 2005. [https://doi.org/10.3168/jds.S0022-0302\(05\)72782-X](https://doi.org/10.3168/jds.S0022-0302(05)72782-X).
- VANRADEN, Paul. Efficient methods to compute genomic predictions. **J Dairy Sci** 91:4414–4423. 2008. <https://doi.org/10.3168/jds.2007-0980>.
- VANRADEN, Paul. Symposium review: How to implement genomic selection. **J Dairy Sci** 103:5291–5301. 2020. <https://doi.org/10.3168/JDS.2019-17684>.
- WOLAK, Matthew and REID, Jane. Accounting for genetic differences among unknown parents in microevolutionary studies: how to include genetic groups in quantitative genetic animal models. **Journal of Animal Ecology** 86:7–20. 2017. <https://doi.org/10.1111/1365-2656.12597>.

Agradecimentos

Agradecemos o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (Capes).

Autores

Daniele Alves¹, Saymon Rocha¹, Thais Machado¹, Jeriel Dias¹, Heloísa Furlani¹, Kettlen Neves¹, Sirlene Lazaro², Hinayah Rojas de Oliveira³, Thaise Pinto de Melo¹

¹ Departamento de Zootecnia, Universidade Federal de Santa Maria, Santa Maria, Brasil; ² Departamento de Biociência Animal, Universidade de Guelph, Canadá; ³ Departamento Zootecnia, Universidade de Purdue, West Laffayette, Estados Unidos.