

BRUNNA DE FIGUEIREDO DUARTE

**ANÁLISE DE ASSOCIAÇÃO GENÔMICA AMPLA PARA CARACTERES
MORFOLÓGICOS, AGRONÔMICOS E DE BROTAÇÃO EM PAINEL DE BATATA
TETRAPLOIDE**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

Orientador: Guilherme da Silva Pereira

**VIÇOSA - MINAS GERAIS
2024**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

D812a
2024

Duarte, Brunna de Figueiredo, 1995-
Análise de associação genômica ampla para caracteres
morfológicos, agronômicos e de brotação em painel de batata
tetraploide / Brunna de Figueiredo Duarte. – Viçosa, MG, 2024.
1 dissertação eletrônica (48 f.): il. (algumas color.).

Orientador: Guilherme da Silva Pereira.

Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Agronomia, 2024.

Referências bibliográficas: f. 41-48.

DOI: <https://doi.org/10.47328/ufvbbt.2024.534>

Modo de acesso: World Wide Web.

1. Batata - Melhoramento genético. 2. Batata -
Conservação. 3. Marcadores genéticos. I. Pereira, Guilherme da
Silva, 1986-. II. Universidade Federal de Viçosa. Departamento
de Agronomia. Programa de Pós-Graduação em Genética e
Melhoramento. III. Título.

CDD 22. ed. 635.212

Bibliotecário(a) responsável: Euzébio Luiz Pinto CRB-6/3317


BRUNNA DE FIGUEIREDO DUARTE

**ANÁLISE DE ASSOCIAÇÃO GENÔMICA AMPLA PARA CARACTERES
MORFOLÓGICOS, AGRONÔMICOS E DE BROTAÇÃO EM PAINEL DE BATATA
TETRAPLOIDE**


Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Magister Scientiae*.

APROVADA: 29 de maio de 2024

Assentimento:

Documento assinado digitalmente
 **BRUNNA DE FIGUEIREDO DUARTE**
Data: 27/08/2024 17:50:08-0300
Verifique em <https://validar.iti.gov.br>

Brunna de Figueiredo Duarte
Autora

Documento assinado digitalmente
 **GUILHERME DA SILVA PEREIRA**
Data: 27/08/2024 09:13:45-0300
Verifique em <https://validar.iti.gov.br>

Guilherme da Silva Pereira
Orientador

AGRADECIMENTOS

A Deus.

À minha mãe, Maria Alice de Figueiredo, por todo apoio, suporte e amor incondicional.

A meu pai, Vicente Soares Duarte, por todo suporte e incentivo no amor pela ciência.

Aos meus irmãos, Matheus, Izadora, Davi e Lucas por se fazerem presentes independente da distância.

Ao meu namorado, Nathan, que mesmo distante não mediu esforços para ser sempre meu companheiro e incentivador.

À Universidade Federal de Viçosa e ao Programa de Genética e Melhoramento por nos proporcionar um ensino e experiências de qualidade.

Ao meu orientador, Dr. Guilherme da Silva Pereira, por todos os ensinamentos e confiança.

A EMBRAPA Clima Temperado, em especial a pesquisadora Dra. Caroline Marques Castro, pela disponibilidade dos dados, auxílio e confiança.

Aos meus amigos que se tornaram a minha família, Ana Paula, Ana Luiza, Jéssica, Vivian e Thiago. Agradeço por sempre me apoiarem e estarem comigo nos momentos em que mais precisei.

Aos meus colegas de laboratório, Carla, João Roberto, Iara e Ana Paula por todo auxílio.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pela concessão da bolsa de mestrado.

Aos secretários do PPGGM, Marco Túlio e Odilon, por estarem sempre disponíveis a nos ajudar.

A todos que de alguma forma me auxiliaram diretamente e indiretamente na minha jornada pessoal e profissional. Manifesto aqui os meus sinceros agradecimentos.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, e Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) – Código de Financiamento APQ-02529-22.

RESUMO

DUARTE, Brunna de Figueiredo, M.Sc, Universidade Federal de Viçosa, maio de 2024. **Análise de associação genômica ampla para caracteres morfológicos, agronômicos e de brotação em painel de batata tetraploide.** Orientador: Guilherme da Silva Pereira

A batata tetraploide (*Solanum tuberosum*) é a sexta espécie de maior importância econômica do mundo, devido a sua versatilidade de consumo e pelo seu equilíbrio adequado de conteúdo de proteínas e carboidratos. Porém, ao ser armazenada por períodos prolongados e/ou com má refrigeração, pode ocorrer a brotação nos tubérculos, o que impede o consumo humano. Em razão da natureza autotetraploide e por ser altamente heterozigótica, a seleção de genótipos superiores para os caracteres de interesse se dá após avaliações sucessivas em campo de progênies F₁, o que pode levar cerca de dez anos. Seleção assistida por marcadores pode, portanto, acelerar esse processo. Assim, o objetivo deste estudo foi identificar polimorfismos de nucleotídeo único (SNPs) associados a caracteres morfológicos, agronômicos e relacionados à brotação em painel de batata. O painel, pertencente à Embrapa Clima Temperado, consistiu em 137 acessos, genotipados para 8.303 SNPs e avaliados para quatorze caracteres em casa-de-vegetação e câmara fria ao longo de dois anos, 2020 e 2021. Os experimentos foram conduzidos em delineamento em blocos aumentos de Federer com 11 e 9 blocos para os respectivos anos. Ao realizar análise de dosagem alélica via pacote R *fitTetra* v.1.0 e filtragem dos SNPs com base em frequência alélica mínima inferior a 5% e de dados perdidos superior 20%, restaram 4.879 SNPs. A análise de estrutura populacional e do parentesco entre os acessos foram realizadas, respectivamente, via análise de componentes principais e pela matriz de relacionamento genômico de VanRaden por meio dos pacotes R *pcaMethods* v.1.86.0 e *AGHmatrix* v.2.0.4. O desequilíbrio de ligação foi estimado pelo pacote R *ldsep* v.2.1.5. Para os estudos de GWAS, médias fenotípicas ajustadas foram obtidas, e usadas para identificar as associações utilizando o pacote R *GWASpoly* v.2.1.3 para diversos modelos e formas de parametrização. Nas análises de estrutura populacional (Q) e relacionamento genômico (K), foram encontrados agrupamentos entre acessos com origens semelhantes. Nas análises de desequilíbrio de ligação, observou-se que o decaimento ocorreu por volta de 1 Mb. O modelo de GWAS que se destacou foi o Q + K, por meio do qual foi possível encontrar associações significativas para oito caracteres. Foram eles: Formato (Form), Largura (LarTub) e Comprimento (CompTub) do tubérculo, Cor da polpa (CorPol), Peso seco da parte aérea (PSPA), Peso fresco do tubérculo (PFT), Dias para iniciar a brotação (DIB) e Comprimento do broto apical (CompBrAp). Ao estudar as regiões com associações significativas, QTLs observados previamente para Form e CorPol foram identificados. Além disso, dois, cinco, três e um genes candidatos foram observados nas regiões para CompTub, CompBrAp, DIB e PFT, respectivamente. Os resultados encontrados neste presente estudo são de grande importância para o conhecimento genético dos caracteres estudados, além de proporcionar a criação de ferramentas moleculares para seleção de genótipos superiores e com brotação tardia.

Palavras-chave: *Solanum tuberosum*. GWAS. SNP. Desequilíbrio de ligação

ABSTRACT

DUARTE, Brunna de Figueiredo, M.Sc, Universidade Federal de Viçosa, May, 2024. **Genome-wide association analysis for morphological, agronomic and sprouting related traits in a tetraploid potato panel.** Adviser: Guilherme da Silva Pereira

Tetraploid potato (*Solanum tuberosum*) is the sixth most economically important species in the world due to its versatility in consumption and its balanced protein and carbohydrate content. However, when stored for prolonged periods or under poor refrigeration conditions, sprouting may occur in potato tubers, making them unsuitable for human consumption. Due to its autotetraploid nature and high heterozygosity, the detection of superior phenotypes for target traits occurs through field evaluation of F₁ progenies. Thus, the objective of this study was to identify single nucleotide polymorphisms (SNPs) associated with morphological, agronomic and sprouting related traits in a potato panel, estimate genetic parameters for these traits, study linkage disequilibrium decay, analyze population structure, and genomic relationships within tetraploid potato accessions. A panel of 137 accessions from Embrapa Temperate Agriculture genotyped with 8,303 SNPs was used. These genotypes were also evaluated for fourteen traits in greenhouse and cold chamber conditions in 2020 and 2021, with experiments conducted using Federer's augmented block design with 11 and 9 blocks for the respective years. Allele dosage analysis was performed using the R package *fitTetra* v.1.0, and SNP filtering based on a minor allele frequency less than 5% and missing data up to 20% resulted in 4,879 SNPs for the studies. Population structure analysis and relatedness among accessions were conducted via principal component analysis and the VanRaden genomic relationship matrix using the R packages *pcaMethods* v.1.86.0 and *AGHmatrix* v.2.0.4, respectively. Linkage disequilibrium was computed using the R package *ldsep* v.2.1.5. For GWAS studies, adjusted phenotypic means were estimated, and associations were identified using the R package *GWASpoly* v.2.1.3 with various models and parameterizations. Population and relatedness analysis revealed groupings among accessions with similar origins. Linkage disequilibrium analysis has decayed around 1 Mb. For GWAS analyses, the Q + K model with eight parameterizations stood out, revealing significant associations for eight traits: Shape (Form), Width (LarTub) and Length (CompTub) of the tuber, Flesh Color (CorPol), Dry weight of the aerial part (PSPA), Fresh weight of the tuber (PFT), Days to start sprouting (DIB) and Length of the apical bud (CompBrAp). When studying the association regions for their respective traits, previously identified QTLs were found for Form and CorPol, as well as two, five, three, and one candidate genes for CompTub, CompBrAp, DIB, and PFT, respectively. The results of this study are of great importance for understanding the genetic basis of the studied traits and serve as a basis for molecular tool for selecting superior genotypes with delayed sprouting.

Keywords: *Solanum tuberosum*. GWAS. SNP. Linkage disequilibrium.

SUMÁRIO

1. INTRODUÇÃO	9
2. REVISÃO DE LITERATURA.....	12
2.1. DOSAGEM ALÉLICA DE SNPs.....	16
2.2. ESTUDOS POPULACIONAIS E RELACIONAMENTO GENÔMICO.....	17
2.3. DESEQUILÍBRIO DE LIGAÇÃO E ESTUDOS DE ASSOCIAÇÃO.....	18
3. MATERIAIS E MÉTODOS	20
3.1. MATERIAL VEGETAL.....	20
3.2. DADOS FENOTÍPICOS	20
3.2.1. DADOS AVALIADOS EM CASA DE VEGETAÇÃO	20
3.2.2. DADOS AVALIADOS EM CÂMARA FRIA.....	22
3.3. ANÁLISES FENOTÍPICAS.....	22
3.3.1. ESTIMAÇÃO DE HERDABILIDADES	22
3.3.2. ESTIMAÇÃO DE MÉDIAS AJUSTADAS.....	23
3.4. DADOS GENOTÍPICOS E CHAMADA DE DOSES.....	23
3.5. ANÁLISES GENOTÍPICAS	24
3.6. ASSOCIAÇÃO GENÔMICA AMPLA (GWAS).....	24
4. RESULTADOS E DISCUSSÃO	27
4.1. ANÁLISES FENOTÍPICAS.....	27
4.2. ANÁLISES GENOTÍPICAS	31
4.3. ASSOCIAÇÃO GENÔMICA AMPLA (GWAS).....	33
5. CONCLUSÃO	40
6. REFERÊNCIAS.....	41

LISTA DE TABELAS

Tabela 1 – Modos de avaliação dos quatorze caracteres de batata avaliados em painel de diversidade da Embrapa.....	21
Tabela 2 – A estatística do teste de razão da verossimilhança (LRT) e seu respectivo P -valor ($**P < 0,01$, $***P < 0,001$) para efeitos de genótipos e estimativas de herdabilidade, H^2 , para os quatorze caracteres em dois anos de avaliação	28
Tabela 3 - Médias ajustadas gerais, mínimas e máximas para quatorze caracteres em dois anos de avaliação	29
Tabela 4 – Limiares $-\log_{10}(P)$ obtidos pelo método ‘M.eff’ para os oito modelos do pacote <i>GWASpoly</i>	34
Tabela 5 – Caracteres com picos de associação e seus respectivos marcadores mais significativos.....	39

LISTA DE FIGURAS

Figura 1 – Esquema geral de melhoramento clássico de batata.	15
Figura 2 – Gráfico com formas de parametrização para tetraploides dos marcadores SNP. (A) Modelo aditivo, (B) Efeito diploidizado aditivo, (C) Efeito 1-dom-alt, (D) Efeito 1-dom-ref, (E) Efeito 2-dom-alt e (F) Efeito 2-dom-ref.....	26
Figura 3 – Boxplot da distribuição dos dados brutos dos caracteres fenotípicos em dois anos de avaliação	27
Figura 4 – Correlação entre as médias ajustadas de quatorze caracteres em dois anos de avaliação, 2020 (topo) e 2021 (base).....	30
Figura 5 – Análise dos componentes principais (PC) das 240 amostras a partir de dosagem alélica de 4.879 SNPs	31
Figura 6 – Dendrograma obtido a partir da matriz de relacionamento genômico realizado entre 240 genótipos de batata genotipados com 4.789 SNPs	32
Figura 7 – Decaimento do desequilíbrio de ligação em painel de 240 acessos de batata tetraploide genotipado com 4.789 SNPs.....	33
Figura 8 – Gráficos de Manhattan e de QQ dos caracteres morfológicos.....	37
Figura 9 – Gráficos de Manhattan e de QQ dos caracteres relacionados a produção	38
Figura 10 – Gráficos de Manhattan e de QQ dos caracteres relacionados a brotação	38

1. INTRODUÇÃO

A batata cultivada (*Solanum tuberosum* L.) é a sexta cultura com maior produção em escala mundial, produzindo por volta de 374 milhões de toneladas no ano de 2022 (FAOSTAT, 2022), devido ao seu alto valor nutricional e versatilidade culinária (Douches & Jastrzebski, 1993). Todavia, devido à demanda crescente do mercado, há a necessidade do tubérculo ser armazenado por períodos prolongados, tornando a brotação natural dos tubérculos armazenados um desafio (Boivin et al., 2020).

Após a colheita, os tubérculos entram em um estado de dormência, durante o qual a brotação é suprimida. No entanto, o período de dormência muitas vezes não coincide com a necessidade de armazenamento, fazendo-se necessário controlar ou prolongar os dias para iniciar a brotação para evitar a perda nutricional da batata e a formação de compostos tóxicos, como a solanina (Boivin et al., 2020; Zhang et al., 2018). Assim, embora haja a demanda por variedades com brotação tardia, o desenvolvimento dessas variedades por meio de métodos de melhoramento tradicionais é demorado. Nesse esquema, há apenas coleta de informações fenotípicas em campo, sem o auxílio de tecnologias de marcadores moleculares. Apesar de ser uma forma ainda muito utilizada nos programas de melhoramento, a conexão entre métodos tradicionais e moleculares trouxe, dentre outros benefícios, a diminuição no tempo de obtenção de cultivares ao permitir seguir com indivíduos que, potencialmente, já possuem a característica de interesse (Andersson et al., 2017; Bhatia & Bajwa, 2022; Igwe, 2021). Além disso, o melhoramento conjunto de caracteres morfológicos e agrônômicos devem ser realizadas simultaneamente aos de brotação devido às restritas preferências do consumidor e do produtor (Mudege et al., 2021).

A complexidade do melhoramento genético da cultura da batata, caracterizada como autotetraploide e altamente heterozigótica, contribui para um processo moroso na obtenção de novas cultivares, em função da necessidade de um maior número de indivíduos e de cruzamentos para obtenção de cultivares com as características de interesse. (Wang et al., 2022). Para superar as limitações das abordagens convencionais, como exemplo a necessidade de elevado número de indivíduos e influência ambiental, técnicas de biotecnologia e ferramentas de bioinformática se tornam fundamentais (Barrell et al., 2013). Dentre estas técnicas, destacam-se os marcadores de DNA, sobretudo aqueles baseados em polimorfismos de nucleotídeo único (do inglês, *single nucleotide polymorphisms* – SNPs). Em batata, intensidade dos alelos dos SNPs podem ser obtidos por meio do sistema *Illumina Infinium*

(Felcher et al., 2012), e modelados para dosagem com auxílios de pacotes como *fitTetra* v.1.0 (Voorrips et al., 2011). Em conjunto, essas abordagens permitem acessar a dosagem alélica de cada loco marcador, aumentando sua informatividade. Os SNPs têm sido empregados na cultura da batata para diversos fins, incluindo estudos de diversidade genética e estrutura populacional (Deperi et al., 2018), estudos de relacionamento genômico (Villanueva et al., 2021), seleção genômica (Sverrisdóttir et al., 2017), mapeamento de locos de herança quantitativa (do inglês, *quantitative trait loci* – QTL) (da Silva Pereira et al., 2021; Odilbekov et al., 2020), estudos de associação genômica ampla (do inglês, *genome-wide association studies* – GWAS) (Yuan et al., 2020), entre outras aplicações.

Ao utilizar GWAS, é comum realizar análises da estrutura populacional e do relacionamento genômico de indivíduos. A compreensão do relacionamento entre indivíduos e de endogamia é imprescindível, pois, em maiores magnitudes, pode levar à redução da diversidade genética dentro da população (Villanueva et al., 2021), sem a qual seria impossível identificar variação em caracteres de interesse. A análise de estrutura populacional em um painel de indivíduos permite avaliar a estrutura genética presente na população estudada (Deperi et al., 2018), sendo a Análise de Componentes Principais (PCA) um dos métodos mais utilizados para este propósito. Descrita por Pearson (1901) e aplicada por Hotelling (1933, 1936), essa análise converte as variáveis pertinentes (neste caso, os SNPs) em um conjunto menor de variáveis que explicam a variabilidade dos dados, permitindo a redução de dimensionalidade (Wang et al., 2020).

Já o relacionamento genético entre os indivíduos pode ser conhecido por meio do coeficiente de endogamia, que calcula a probabilidade de dois alelos serem idênticos por descendência, e os coeficientes de relacionamento, que quantificam a relação genética entre indivíduos com base na ancestralidade compartilhada (Bernstein et al., 2018; Villanueva et al., 2021). Essas medidas de relacionamento podem ser obtidas diretamente via identidade por descendência (do inglês, *identity-by-descent* – IBD), referindo-se a alelos idênticos que foram herdados do mesmo ancestral comum, ou indiretamente via identidade por estado (do inglês, *identity-by-state* – IBS), em que os alelos são idênticos sem necessariamente compartilharem origens ancestrais recentes. Para GWAS, o IBS é o mais comumente utilizado devido à limitação sobre o conhecimento da estrutura familiar ancestral das amostras, sendo, ainda assim, útil para auxiliar na identificação de regiões do genoma associadas a características de interesse (Berdugo-Cely et al., 2023).

Ao se conduzir GWAS em espécies poliploides, como a batata, o pacote do R *GWASpoly*, desenvolvido por Rosyara et al. (2016), tem sido amplamente utilizado para nas análises usando o modelo linear $Q + K$ (Yu et al., 2006), a fim de controlar associações espúrias, ou seja, uma associação com efeitos tendenciosos e sinais falsos devido a estruturação ou forte relacionamento (Naeem et al., 2021; Yuan et al., 2020). Neste modelo, Q representa a estrutura populacional e K é a matriz de relacionamento genômico. Assim, um estudo de utilizando o modelo $Q + K$ irá, potencialmente, controlar os efeitos de estruturação de população e do relacionamento entre indivíduos que possuem uma origem ou ancestralidade em comum (Rosyara et al., 2016).

Além disso, outro ponto importante a ser avaliado em GWAS é o decaimento do desequilíbrio de ligação (DL). Esta análise permite quantificar o grau de associação não-aleatória, ou preferencial, entre alelos de diferentes locos em uma população. Caso o conjunto de SNPs não seja denso o suficiente, permitindo uma cobertura ampla do genoma, não será possível identificar marcadores associados a caracteres de interesse (Hill & Robertson, 1968; Vos et al., 2017). Recentemente, Gerard (2021) propôs um método para estimar o DL especificamente em espécies poliploides, por meio do qual é possível estimar o DL em autopoliploides mesmo considerando a incerteza genotípica, sob a hipótese de equilíbrio de Hardy-Weinberg, utilizando probabilidades das dosagens do genótipo.

Portanto, o objetivo deste estudo foi identificar marcadores SNPs associados a caracteres morfológicos, agronômicos e relacionados à brotação em batata tetraploide. Durante o estudo, buscou-se também estimar parâmetros genéticos para caracteres agronômicos e relacionados à brotação de batata tetraploide, estudar o decaimento do desequilíbrio de ligação entre pares de marcadores SNP em painel de batata tetraploide, e analisar a estrutura genética populacional e o relacionamento genômico realizado de acessos de batata tetraploide.

2. REVISÃO DE LITERATURA

O gênero *Solanum*, ao qual as batatas cultivadas e selvagens pertencem, possui mais de 1 mil espécies reconhecidas e é composto por outras plantas economicamente importantes como, berinjela e tomate (Kaunda & Zhang, 2019). As espécies selvagens de batata possuem distribuição nas Américas, desde o sudoeste dos Estados Unidos até o centro da Argentina e do Chile (Hijmans & Spooner, 2001). Elas possuem dois centros de diversidade, um na América do Norte e Central, com centro no México, e a outra é na América do Sul, com centro nos Andes, se prolongando da Venezuela ao Chile. Dentre as espécies cultivadas, há diversas ploidias variando de diploide ($2n = 2x = 24$), a pentaploide ($2n = 5x = 60$), e, como exemplo, se pode citar as espécies *S. ajanhuiri* e *S. stenotomum* como diploides, *S. tuberosum* como tetraploide, e *S. curtilobum* como pentaploide (Spooner et al., 2007).

A espécie *S. tuberosum* L. é a mais utilizada comercialmente e sua origem ocorreu na América do Sul (Adu et al., 2019), de um progenitor silvestre do complexo *S. brevicaulle*, localizado no sul do Peru (Machida-Hirano, 2015). *S. tuberosum* constitui uma das espécies de batata mais importantes para o mundo devido a seu alto valor nutricional (Douches & Jastrzebski, 1993), produzindo por volta de 374 milhões de toneladas no ano de 2022 (FAOSTAT, 2022). Desse total, 54,3% estão concentrados no continente asiático, tendo, como maiores produtores, China, Índia e Rússia, enquanto 26,2% da produção permanece na Europa, com destaque para Ucrânia, e 11,9% no continente americano, tendo os Estados Unidos como principal produtor. Os restantes 7,6% da produção são derivados da África e Oceania (FAOSTAT, 2022). O Brasil está na 19ª posição na produção global de batata, produzindo por volta de 3 milhões de toneladas em 2022. Para o consumo de batata fresca, o Brasil é autossuficiente e não há necessidade de importação. No entanto, quando se trata do mercado de alimentos congelados, o Brasil depende de importação já que a demanda é maior do que a capacidade de produção local (Horodyska et al., 2021; Silva et al., 2022).

Apesar de ser tratada como anual, para fins comerciais, a planta da batata é uma planta perene. No geral, as batateiras são advindas de tubérculo-semente ou batata-semente, que nada mais é do que o tubérculo plantado do qual surge a planta. Este conceito é diferente de uma semente verdadeira, que é nomeada semente botânica e é utilizada apenas para fins de melhoramento genético (Fortes & Pereira, 2003). A planta apresenta caule flexível, e é dividido entre aéreo e subterrâneo. Suas flores possuem o formato de cimeira e são hermafroditas, podendo ser branca, rosada ou arroxeadas. Geralmente, a flor possui incapacidade de ser

fecundada pelo seu próprio pólen, o que impede a autofecundação. Outra estrutura, a qual se dá o nome de batata, são os tubérculos. Eles consistem em caules subterrâneos que se adaptaram para ser o sistema de reserva de amido da planta, sendo também utilizados como meio de propagação. Na película da batata, pode-se observar os olhos, que são gemas dormentes de onde sairão as hastes e os estolões quando é quebrada a dormência e inicia-se brotação (Fortes & Pereira, 2003).

Do ponto de vista genético, *S. tuberosum* é uma espécie autotetraploide ($2n = 4x = 48$), possuindo quatro cópias homólogas de cada um dos seus 12 cromossomos. A origem da ploidia mais elevada da batata, em comparação com outras espécies cultivadas, que são diploides (milho, soja, arroz, feijão, outras), é sugerida pela habilidade da batata diploide em produzir grãos de pólen $2n$ de forma natural (Rodríguez, 2010). A existência de quatro cromossomos homólogos em autotetraploides pode resultar em um pareamento cromossômico entre dois cromossomos (bivalentes) ou entre mais de dois cromossomos homólogos (multivalentes) de forma aleatória, caracterizando uma herança tetrassômica (Bradshaw, 2007). Adicionalmente, distorções sistemáticas na segregação, podem acontecer a partir de um fenômeno conhecido como dupla redução, adicionando dificuldade ao estudo genético de espécies autotetraploides (Amadeu et al., 2016).

O cultivo da batata ocorre por meio do seu tubérculo por propagação vegetativa, sendo utilizada para fins de produção agrícola e manutenção de germoplasma. Pelo seu grande consumo, a batata precisa ser normalmente armazenada por vários meses, e isso pode ser problemático em razão do fenômeno natural de brotação. Após a colheita, os tubérculos ficam com suas gemas (olhos) dormentes, não ocorrendo brotação. Normalmente, esse período de dormência é menor que o de armazenamento, fazendo com que a brotação precise ser controlada e/ou que tenha o tempo de ocorrência aumentado para não afetar o valor nutricional da batata (Boivin et al., 2020). Um grande problema de tubérculos com brotos é a alta concentração de glicoalcalóides, como a solanina. Essa substância é tóxica para os humanos quando consumida em grande quantidade (Zhang et al., 2018).

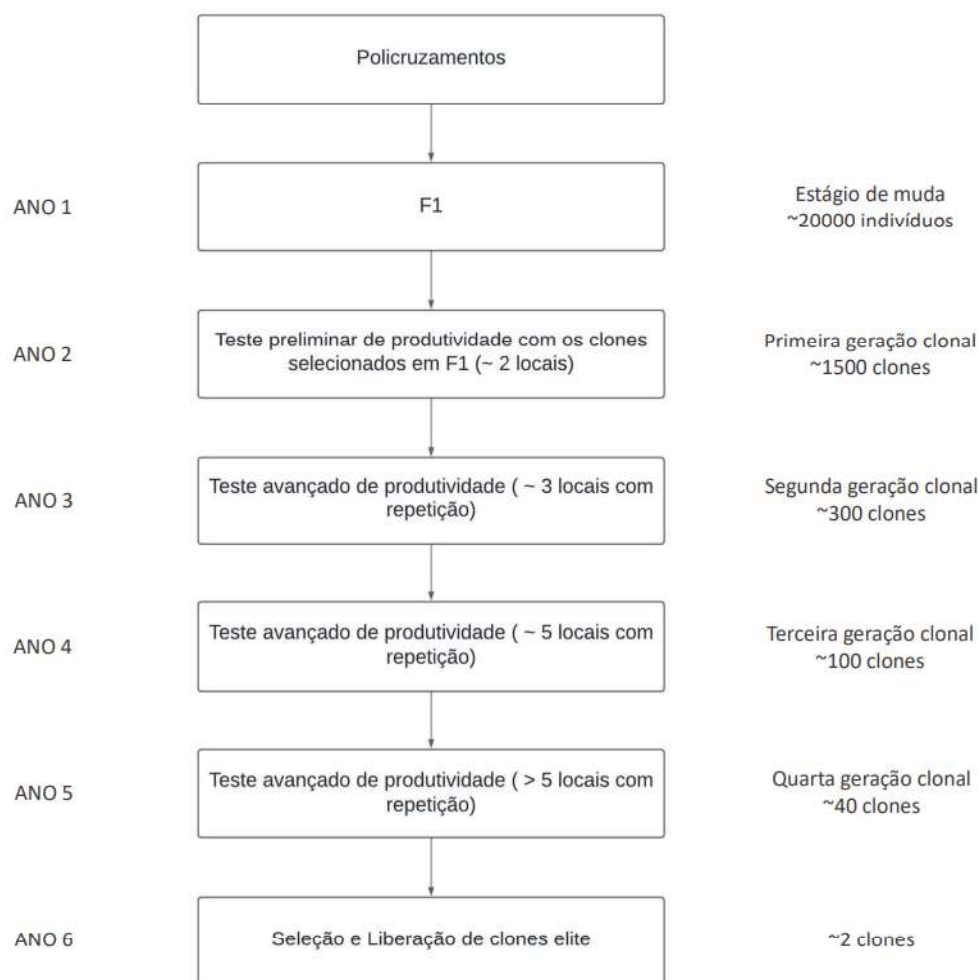
Além de caracteres relacionados à brotação, outros tipos de caracteres como os morfológicos e relacionados a produção são importantes de serem avaliados visando os segmentos de mercado da batata, sendo alguns deles: mesa, *chips* e pré-frita. Como exemplo, para o mercado de batata *chips* e de mesa, o formato do tubérculo esperado é redondo e ovalado enquanto que para o mercado de pré-frita o mais adequado são cultivares com formato alongados (Pandey et al., 2022). O peso seco da parte aérea está correlacionado positivamente

ao conteúdo de amido nos tubérculos, logo, teor de matéria seca é utilizado para monitorar a eficiência do processamento na indústria de fritura, evitando fatias quebradiças e danificadas (Gościnna et al., 2024; Su & Sun, 2017). A cor da polpa também tem influência quando se fala sobre segmentos de mercado. Para o mercado de batata *chips*, há uma preferência da cor da polpa branca, normalmente mais firme. Já para o mercado de mesa, a preferência dos consumidores é pela cor da polpa creme ou amarelada (Nick & Borém, 2017).

O primeiro passo no melhoramento genético da batata (Figura 1) é a seleção dos genitores e o cruzamento entre eles. Esses genitores são altamente heterozigotos, e cada semente botânica resultante desse cruzamento difere bastante entre si em virtude da segregação dos múltiplos locos (Campos & Ortiz, 2020). A seleção F_1 em batatas envolve a avaliação das plântulas originadas da germinação e emergência de sementes verdadeiras, bem como do tubérculo desenvolvido subsequentemente. Portanto, sementes verdadeiras, obtidas por hibridização e normalmente plantadas em casa de vegetação, são cruciais para selecionar características desejáveis no melhoramento da batata (Aydın & Öztürk, 2022; Koyuturk & Yilmaz, 2022). A seleção assistida de marcadores pode facilitar essa seleção ao identificar se os genótipos F_1 apresentam marcas para o perfil de produto desejado ou se é necessário o descarte, sem haver a necessidade de seguir seleção visual ou etapas subsequentes.

Ao realizar a seleção de plântulas F_1 , começam os testes de produtividade e de características de interesse por meio da propagação clonal das batata-sementes em campo. Dentre essas características, podem ser citadas tamanho e formato do tubérculo, profundidade dos olhos, cor e textura da casca, peso de matéria seca, qualidade, além de características relacionadas a doenças. Nos dois primeiros anos (Figura 1), são realizadas avaliação de características morfológicas e, nos quatro anos seguintes, os genótipos são avaliados em mais locais e com repetições, com ênfase em características mais quantitativas, como aquelas relacionadas à produção (Bonierbale et al., 2020; Pandey, 2021). Por fim, ao chegar no último ano com um clone elite, o foco é no desenvolvimento do mercado e a promoção das novas variedades, explorando os interesses de produtor, consumidor e indústria para uma melhor aceitação no mercado (Pandey, 2021).

Figura 1 – Esquema geral de melhoramento clássico de batata.



Fonte: adaptado de Bonierbale et al. (2020).

Assim como outras características de interesse, o desenvolvimento de variedades com uma brotação mais tardia por métodos de melhoramento tradicionais é complexo e demanda certo tempo (Andersson et al., 2017). As razões pelas quais há demora na obtenção de cultivares está relacionada aos fatos de a genética da cultura ser complexa em função da poliploidia, herança tetrassômica e alta heteroziguidade, relatados em tópicos acima, exigindo a amostragem de milhares de sementes botânicas e a propagação clonal em ensaios subsequentes com crescente tamanho de parcelas e locais/anos de avaliação. Técnicas de biotecnologia podem ser utilizadas para contornar algumas dificuldades das técnicas convencionais (Barrell et al., 2013).

Dentre as técnicas de biotecnologia disponíveis, têm-se os chamados marcadores moleculares. Diferentes tipos ou sistemas de marcadores foram desenvolvidos e utilizados para a batata até o momento, como, por exemplo, marcadores do tipo de polimorfismos de tamanho

de fragmento amplificado (do inglês, *amplified fragment length polymorphisms* – AFLP) (del Rio & Bamberg, 2021) e de repetições de sequências simples (do inglês, *simple sequence repeats* – SSR) (Ivanova-Pozdejeva et al., 2022). Mais recentemente, o foco dos estudos com esta cultura tem sido nos SNPs, por serem marcadores de alta densidade e rendimento, além de mais informativos quando comparados aos marcadores AFLP e SSR, por exemplo.

2.1. DOSAGEM ALÉLICA DE SNPs

A genotipagem em batata tetraploide é desafiadora porque seus locos marcadores podem ter até quatro alelos diferentes cada. Por consequência, para análise genética, é necessário um sistema de genotipagem que possa diferenciar tais alelos (haplótipos), ou, ao menos, quantificar o número de cópias de cada alelo a partir de um sistema de análise bialélico. Esse número de cópias dos alelos de marcadores bialélicos (A e a), como os SNPs, também chamada de zigosidade, pode ser categorizada como nulliplex ($aaaa$), simplex ($Aaaa$), duplex ($AAaa$), triplex ($AAAa$) e quadruplex ($AAAA$) em tetraploides (Uitdewilligen et al., 2013). Até recentemente, a identificação das dosagens alélicas não era claramente realizada, devido às limitações apresentadas por sistemas de marcadores baseados em reação em cadeia da polimerase (PCR) e géis eletroforéticos, exclusivamente. Atualmente, isso é possível pelo desenvolvimento de plataformas de genotipagem de SNPs que possibilitam inferência de dosagem a partir de sinais proporcionais à quantidade de cada alelo nos indivíduos e pacotes como *fitTetra v.1.0*. (Voorrips et al., 2011).

Um exemplo de plataforma para genotipagem para poliploides, por meio de SNPs, é a Illumina Infinium. Ela se baseia na tecnologia BeadArray da Illumina, que nada mais é que um *chip* com esferas de sílica em forma de matriz. Em cada grânulo há a sequência específica de captura do seu polimorfismo correspondente, e ao ser processado, o DNA se hibridiza com as esferas e emitem sinais fluorescentes. Como os SNPs são bialélicos, esse sistema descreve a intensidade de fluorescência de duas sondas, uma correspondendo a cada alelo, em cada loco, como coordenadas cartesianas (X, Y) . Logo, é possível identificar a dosagem alélica pois espera-se que as intensidades dos sinais, $R = Y/(X + Y)$, sejam proporcionais às respectivas dosagens (Felcher et al., 2012; Mollinari & Serang, 2015). Simplificadamente, no caso do *fitPoly*, um modelo de mistura de normais é ajustado, com cada distribuição normal correspondendo a uma distribuição de valores R originadas das cinco possíveis classes (0, 1, ..., 4), e, assim, as classificações de doses são obtidas.

2.2. ESTUDOS POPULACIONAIS E RELACIONAMENTO GENÔMICO

Após dosagem e filtragem para qualidade, os dados de SNPs podem ser utilizados para avaliar a estrutura populacional em um painel de indivíduos. O objetivo é caracterizar a estrutura genética presente na população disponível, que poderá conter indivíduos oriundos de diferentes origens e regiões geográficas, contemplando, assim, histórias evolutivas específicas (Deperi et al., 2018). Um dos métodos amplamente utilizados para isso é a análise de componentes principais (PCA). Tal análise, descrita inicialmente por Pearson (1901) e posteriormente aplicada por Hotelling (1933, 1936), converte as variáveis pertinentes (no caso, os SNPs) em um mesmo número de variáveis, chamados de componentes principais, de tal modo que maior variabilidade do conjunto de dados possa ser explicada pelos primeiros componentes (Wang et al., 2020). Este procedimento possibilita a redução da dimensionalidade multivariada com pouca perda sobre a explicação da variação inicial existente.

A PCA aplicada em dados bialélicos, como SNPs, encontra autovalores e autovetores da matriz de covariância das frequências dos alelos. Com base nos componentes principais, é possível visualizar as amostras e suas distâncias em um gráfico de dispersão ao projetar os genótipos no espaço ampliado pelos eixos dos componentes principais, sendo possível observar evidência de identidade causada por origem ou ancestralidade comum. Isto é uma propriedade importante porque essas distâncias entre os grupos refletem as distâncias genéticas da população (Elhaik, 2022).

No contexto dos estudos de associação genômica ampla (GWAS), além da estrutura populacional, é fundamental investigar o relacionamento genético entre os indivíduos. Esse relacionamento pode ser expresso por meio de uma matriz de relacionamento genômico realizado em um painel de indivíduos com base nos dados de SNPs, considerando a covariância de um indivíduo com ele mesmo – coeficiente de endogamia, diferença em relação a 1 na diagonal principal – e a covariância entre os pares de indivíduos – coeficientes de relacionamento, fora da diagonal principal. O estudo desse relacionamento ou endogamia é fundamental, pois, em maior magnitude, pode causar a redução da diversidade genética dentro da população (Villanueva et al., 2021).

Uma abordagem baseada em SNP para caracterizar o relacionamento genômico entre indivíduos foi proposta por VanRaden (2008) e adaptada para autoploiploides por Ashraf et al. (2016) e Endelman et al. (2018). Ela se dá pela fórmula:

$$G = \frac{WW^T}{\sum_k 4p_kq_k}$$

em que W é a matriz que especifica as dosagens alélicas dos indivíduos, centrada na média, p_k é a frequência do primeiro alelo no loco k e q_k é a frequência do segundo alelo no loco k . A diagonal de WW^T mostra o número de locos homocigóticos para cada indivíduo, e seu valor menos 1 representa o coeficiente de endogamia genômica, que será maior se o indivíduo for mais endogâmico.

2.3. DESEQUILÍBRIO DE LIGAÇÃO E ESTUDOS DE ASSOCIAÇÃO

O desequilíbrio de ligação (DL) determina o grau de associação não aleatória, ou preferencial, entre alelos de diferentes locos numa população. Há diversas medidas que podem ser utilizadas para estimar o DL (Flint-Garcia et al., 2003). Dentre elas, o coeficiente de correlação de Pearson ao quadrado (r^2) é o mais comumente utilizado (Hill & Robertson, 1968). No entanto, tal medida foi proposta no contexto de espécies diploides, sendo necessário utilizar abordagem alternativa para espécies poliploides, como a batata. Recentemente, Gerard (2021) propôs um método para estimar o DL especificamente em espécies poliploides, que permite assumir o erro em conhecer os genótipos (dosagem alélica), e para poliploides essa incerteza é muito maior do que em diploides.

Vários fatores evolutivos afetam o DL (Flint-Garcia et al., 2003), dentre eles: mutação, recombinação, fluxo gênico e seleção. Enquanto mutação, fluxo gênico e seleção podem elevar o DL entre locos, a recombinação pode reduzir este parâmetro ao longo das gerações. Com sua propagação vegetativa, a batata possui relativamente poucos eventos históricos de recombinação, resultando, de modo geral, um decaimento moderado de DL (Tagliotti et al., 2021). O sucesso de estudos genéticos, como GWAS, pode se dar pela extensão de DL presente entre locos. Isso porque sua extensão ao longo do genoma determinará se o número de SNPs genotipados será suficiente para detectar associações entre genótipo e fenótipo (Vos et al., 2017). Se o DL decair rapidamente, maior quantidade de SNPs será necessária para detectar tais associações (D'hoop et al., 2010).

O DL é a base para identificar marcadores associados a caracteres de interesse por meio da metodologia de GWAS. Nesse tipo de estudo, um painel de germoplasma tem examinados – e eventualmente controlados – a estrutura de população e o relacionamento genômico estimados a partir de marcadores. Nas populações utilizadas para GWAS, indivíduos de uma mesma origem, ou aparentados, tendem a compartilhar alelos em comum no genoma, formando

subpopulações de indivíduos dentro da população (estrutura). Por isso, se a caráter estiver com alta frequência na subpopulação, os SNPs mais comuns dentro dela podem apresentar associação espúria com a caráter estudada (Tibbs Cortes et al., 2021). Logo, painéis cujas amostras apresentam baixa estrutura populacional ou parentesco diminuem o risco de se obter associações espúrias entre genótipo e fenótipo. Se um certo painel de GWAS englobar indivíduos de diferentes germoplasmas, que tiveram suas respectivas histórias evolutivas, e se realizar o controle de estrutura e parentesco, o estudo auxiliará no controle de associações não-espúrias entre genótipo e fenótipo (Rosyara et al., 2016).

Em espécies autoploiploides, o pacote do R *GWASpoly* desenvolvido por Rosyara et al. (2016) é aproveitado para conduzir o GWAS usando o modelo linear misto $Q + K$ em SNPs para controlar associações espúrias (Naeem et al., 2021; Yuan et al., 2020), em que Q representa a estrutura populacional e K é a matriz de relacionamento genômico. Trata-se de uma abordagem de modelo misto que auxilia no controle de associações espúrias de marcador-caráter, com modelos apropriados para espécies poliploides.

3. MATERIAIS E MÉTODOS

3.1. MATERIAL VEGETAL

O painel de indivíduos *genotipados* contém 240 genótipos oriundos de diferentes bancos de germoplasma dos países Brasil (159), Quênia (79) e EUA (2), e foram utilizados nos estudos de estrutura de população e relacionamento genômico.

Já o painel de indivíduos *fenotipados* contém 351 genótipos, e foram utilizados na estimação de herdabilidades e obtenção de médias ajustadas. Destes, três genótipos foram utilizados como testemunha, sendo eles as linhagens Asterix, BRS IPR Bel e Markies.

Comparando os indivíduos dos dois painéis, um total de 137 continham informações genóticas e fenóticas, efetivamente usadas em GWAS. Ambos os painéis foram obtidos, avaliados (genotipagem e fenotipagem), e disponibilizados pela Embrapa Clima Temperado.

3.2. DADOS FENOTÍPICOS

3.2.1. DADOS AVALIADOS EM CASA DE VEGETAÇÃO

Nos anos de 2020 e 2021, os genótipos foram avaliados em casa de vegetação na Embrapa Clima Temperado, Pelota – RS. Em 2020, o plantio foi realizado no inverno e a colheita na primavera, nos meses agosto e novembro, respectivamente. Em 2021, o plantio e colheita foram realizadas no outono, entre os meses março e junho. Os experimentos foram conduzidos em delineamento de blocos aumentados de Federer, com testemunhas comuns, tendo 11 e 9 blocos incompletos nos respectivos anos. Para os dois anos, cada bloco foi formado por quatro linhas de 3,6 m de comprimento. A distância entre os blocos foi de 60 cm, mas apenas entre os blocos sete e oito a distância foi de 1,68 m entre eles.

Em 2020, os onze blocos foram constituídos por 27 tratamentos regulares (genótipos) e 3 tratamentos comuns (testemunhas), possuindo um total de 30 parcelas, sendo cada parcela constituída de 2 vasos de 24 cm de diâmetro. Multiplicando-se os 30 genótipos pelos 11 blocos, houve um total de 330 genótipos plantados. No ano seguinte, 2021, os nove blocos também foram constituídos por 27 tratamentos regulares (genótipos) e 3 tratamentos comuns (testemunhas). Houve um total de 270 genótipos plantados devido haver 30 genótipos e 9 blocos. No conjunto dos dois anos, foi possível observar um total de 351 genótipos avaliados, sendo 105 genótipos avaliados em ambos os anos.

A avaliação fenotípica foi realizada para seis caracteres morfológicos (Asp, CorPel, CorPol, LarTub, CompTub e Form) e cinco caracteres de produção (NH, AF, PSPA, NTT, PFT), descritos com mais detalhes na Tabela 1.

Tabela 1 – Modos de avaliação dos quatorze caracteres de batata avaliados em painel de diversidade da Embrapa

Caracteres	Sigla	Modo de avaliação
Caracteres morfológicos		
Aspereza da película do tubérculo	Asp	Escala de 1 a 5, sendo: (1) liso; (2) intermediária entre liso e médio; (3) médio; (4) intermediária entre médio e Russet; (5) Russet
Cor da película	CorPel	Escala de 1 a 12, sendo: (1) amarela; (2) amarela clara; (3) intermediária entre amarela claro e média; (4) amarela médio; (5) intermediária entre amarela média e intensa; (6) amarela intensa; (7) vermelha; (8) vermelha clara; (9) intermediária entre vermelho claro e média; (10) vermelha média; (11) intermediária entre vermelha média e intensa; (12) vermelha intensa
Cor da polpa do tubérculo	CorPol	Escala de 1 a 5, sendo: (1) branca; (2) creme; (3) amarela clara; (4) amarela média; (5) amarela escura
Comprimento do tubérculo	CompTub	Milímetros (mm)
Largura do tubérculo	LarTub	Milímetros (mm)
Formato do tubérculo	Form	Escala de 1 a 5, sendo: (1) redondo; (2) intermediário entre redondo e ovalado; (3) ovalado; (4) intermediário entre ovalado e alongado; (5) alongado
Caracteres relacionados a produção		
Número de hastes	NH	Número de hastes por planta
Área foliar	AF	Escala de 1 a 5, sendo: (1) quando a área coberta pela planta ocupava até 25 % do vaso; (2) entre 25 e 50%; (3) entre 50 e 75%; (4) entre 75 e 99% e (5) quando a área coberta era de 100%
Peso seco da parte aérea	PSPA	Gramas (g)
Número total de tubérculos pertencentes ao total da amostra	NTT	Quantidade de tubérculos
Peso fresco de tubérculos	PFT	Gramas (g)
Caracteres relacionados a brotação		
Dias para iniciar a brotação após entrada na câmara fria	DIB	Escala de 1 a 5, sendo: (1) até 30 dias; (2) 31 a 60 dias; (3) 61 a 90 dias; (4) 91 a 120 dias; (5) acima de 120 dias
Comprimento do broto apical	CompBrAp	Milímetros (mm)
Número de brotos laterais	NBrLat	Milímetros (mm)

Fonte: a autora.

3.2.2. DADOS AVALIADOS EM CÂMARA FRIA

Para avaliação dos três caracteres de brotação (DIB, CompBrAp, NBrLat), também descritos na Tabela 1, foi retirado, em cada ano, um tubérculo por vaso de planta e eles foram dispostos em caixas divididas por quadrados internos formados pela interseção de 5 linhas e 8 colunas. Cada genótipo era representado pelos seus 2 vasos de planta em duas linhas, ou seja, um tubérculo por linha, mantendo a linha número 3 vazia. Logo, cada caixa tinha até 16 genótipos caso todas as plantas produzissem tubérculos, com duas caixas correspondendo a um bloco do experimento. Após todo esse processo, os tubérculos foram armazenados e avaliados por 120 dias em câmara fria com temperatura de armazenamento de 7,5 °C.

3.3. ANÁLISES FENOTÍPICAS

Todas as análises foram realizadas no software estatístico R (R Core Team, 2022), utilizando o pacote *ASReml-R* (Gilmour et al., 2015).

3.3.1. ESTIMAÇÃO DE HERDABILIDADES

Para cálculo da herdabilidade, os dados fenotípicos (Tabela 1) foram analisados por meio do seguinte modelo aleatório, para cada ano separadamente:

$$y_{ij} = \mu + G_i + B_j + \varepsilon_{ij}$$

em que y_{ij} é o valor fenotípico (mensuração do caráter) do genótipo i ($i = 1, 2, \dots, 189$ ou 271 , a depender do ano), no bloco j ($j = 1, 2, \dots, 9$ ou 11 , a depender do ano); μ é a média geral (intercepto); G_i é o efeito aleatório de genótipo, com $G_i \sim N(0, \sigma_g^2)$; B_j é o efeito aleatório de bloco, com $B_j \sim N(0, \sigma_b^2)$; e ε_{ij} é o erro residual aleatório, com $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$.

Os componentes de variância foram estimados via máxima verossimilhança restrita (do inglês, *Restricted Maximum Likelihood* – REML; PATTERSON & THOMPSON, 1971). Foi verificado se o efeito de genótipos possuía significância pelo teste da razão da verossimilhança (do inglês, *Likelihood Ratio Test* – LRT) com probabilidade de 5%. A herdabilidade no sentido amplo (H^2) foi estimada, para cada caráter, pela seguinte equação:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_\varepsilon^2}$$

em que σ_g^2 é a variância genotípica, σ_ε^2 é a variância residual.

3.3.2. ESTIMAÇÃO DE MÉDIAS AJUSTADAS

Para realização de GWAS, as médias ajustadas foram obtidas a partir de modelo misto para cada ano, separadamente:

$$y_{ij} = \mu + G_i + B_j + \varepsilon_{ij}$$

em que y_{ij} é o valor fenotípico (mensuração do caráter) do genótipo i ($i = 1, 2, \dots, 189$ ou 271 , a depender do ano), no bloco j ($j = 1, 2, \dots, 9$ ou 11 , a depender do ano); μ é a média geral (intercepto); G_i é o efeito fixo de genótipo; B_j é o efeito aleatório de bloco, com $B_j \sim N(0, \sigma_b^2)$; ε_{ij} é o erro residual aleatório, com $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$.

As correlações genéticas entre os caracteres foram calculadas entre as médias ajustadas por meio do coeficiente de correlação de Pearson.

3.4. DADOS GENOTÍPICOS E CHAMADA DE DOSES

Um total de 240 genótipos de batata tetraploide foram genotipados no presente estudo, pertencentes a um painel com acessos oriundos de bancos de germoplasmas vindos dos programas de melhoramento do Brasil, Quênia e Estados Unidos. Esses acessos foram genotipados para 8.303 marcadores moleculares do tipo SNP, com base na plataforma Infinium 8303 Potato da Illumina (SolCAP 8K) (Felcher et al., 2012). A sequência da cultivar Atlantic foi utilizada para o desenvolvimento do SolCAP 8k (Naeem et al., 2021), por isso, ela é utilizada como controle de genotipagem neste painel. As posições dos SNPs foram obtidas por utilizando a versão 4.0.3 do genoma de *S. tuberosum* (da Silva Pereira et al., 2021).

A partir dos valores de intensidade obtidos na genotipagem, utilizou-se um modelo de mistura, implementado no pacote do R *fitTetra v.1.0* (Voorrips et al., 2011) para realizar a chamada genotípica de amostras autotetraploides com marcadores SNPs. O algoritmo é baseado na razão dos sinais dos alelos, e ajusta os modelos de misturas das cinco distribuições normais relativas a essas razões. Assim, foi possível categorizar os marcadores nas cinco classes genotípicas possíveis de dosagens dos alelos (*aaaa*, *Aaaa*, *AAaa*, *AAAA*). SNPs com intensidades de sinal que não seguiram o ajuste esperado entre o modelo de misturas e as razões geram dados não-confiáveis, sendo, logo, descartados do conjunto final dos dados. Os SNPs que possuíam frequência de alelos raros (do inglês, *minor allele frequency* - MAF) inferior a 5%, e acima de 20% de dados perdidos foram removidos do conjunto total de marcadores.

3.5. ANÁLISES GENOTÍPICAS

Antes de realizar as análises de GWAS, os dados de SNPs foram utilizados para estudar a estrutura populacional e o relacionamento genômico realizado entre os 240 indivíduos do painel de batata tetraploide. A estrutura populacional foi investigada pela análise de componentes principais (PCA), via pacote do R *pcaMethods* v.1.94.0, utilizando o algoritmo NIPALS (Algoritmo de Mínimos Quadrados Parciais Iterativos Não Lineares) (Stacklies et al., 2007). O relacionamento genômico realizado foi obtido pelo método proposto por (VanRaden, 2008), utilizando-se o pacote do R *AGHmatrix* v.2.0.4 (Amadeu et al., 2016). Para fazer essas análises, foram utilizados os dados filtrados, com base no tópico 3.4.

Para analisar a queda do DL com a distância genética, foi utilizado o método proposto por Gerard (2021) implementado no pacote *ldsep* v.2.1.5, especificamente desenvolvido para espécies poliploides. As posições dos SNPs foram baseados no genoma de referência v. 4.0.3 conforme descrito em da Silva Pereira et al. (2021), a partir do qual verificou-se a distribuição de SNPs nos 12 cromossomos. Os SNPs que não se encaixaram em nenhum cromossomo foram colocados no cromossomo 0, os quais, para fins de estudos de desequilíbrio de ligação, foram removidos da análise.

3.6. ASSOCIAÇÃO GENÔMICA AMPLA (GWAS)

No presente estudo, a análise de GWAS foi realizada com base na abordagem de modelos mistos proposta por Yu et al. (2006), estendida por Rosyara et al. (2016) para o contexto de poliploides, utilizando-se o pacote do R *GWASpoly* v.2.11, com as médias ajustadas de ambos os anos modela conjuntamente conforme o seguinte modelo:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{S}\boldsymbol{\alpha} + \mathbf{Q}\mathbf{v} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

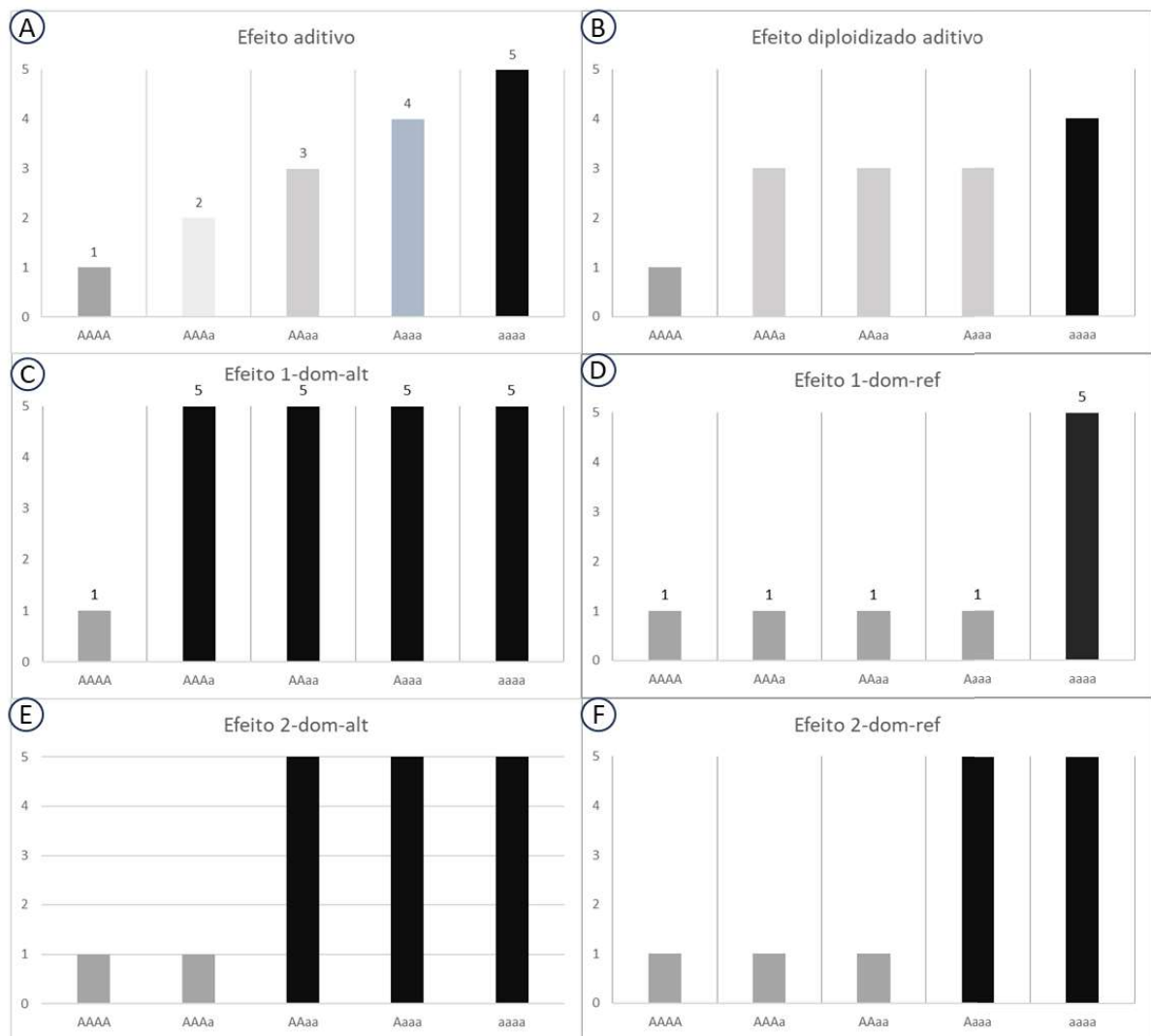
em que \mathbf{y} é o vetor correspondente às médias ajustadas originadas do modelo misto (genótipo como fixo); $\boldsymbol{\beta}$ é o vetor de efeitos fixos, tendo \mathbf{X} como matriz delineamento de anos, além do intercepto; $\boldsymbol{\alpha}$ é o vetor dos efeitos fixos do i -ésimo SNP e \mathbf{S} a matriz de marcadores parametrizada de acordo com os vários modelos (detalhados a seguir); \mathbf{v} é o vetor dos efeitos fixos populacionais tendo \mathbf{Q} como matriz de incidência para covariáveis subpopulacionais originadas de PCA (seis primeiros componentes), relacionando \mathbf{y} a \mathbf{v} ; \mathbf{u} é o vetor de efeitos aleatórios com covariância entre indivíduos proporcional a uma matriz de parentesco \mathbf{K} , com $\mathbf{u} \sim N(\mathbf{0}, \sigma_g^2 \mathbf{K})$; $\boldsymbol{\varepsilon}$ é o vetor de efeitos residuais aleatórios, com $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I})$; \mathbf{X} e \mathbf{Z} são

matrizes de incidência de 1s e 0s relacionando \mathbf{y} a $\boldsymbol{\beta}$ e \mathbf{u} , respectivamente; e σ_g^2 e σ_ε^2 são as variâncias genética e do resíduo, respectivamente.

Foram investigados quatro modelos estatísticos, sendo eles: i) Modelo Naïve, sem controle de estrutura populacional (Q) e parentesco (K); ii) Modelo K , controlando apenas o parentesco; iii) Modelo Q , controlando apenas os efeitos da estrutura populacional; e iv) Modelo $Q + K$, controlando tanto o parentesco quanto a estrutura populacional. O diagnóstico de tais modelos foi realizado por gráficos de quantil-quantil (Q-Q), considerando os P -valores, na escala logarítmica, $-\log_{10}(P)$, observados *versus* esperados para saber qual modelo melhor se ajustou aos dados. Para controlar os falsos positivos (erro tipo I), foi utilizada a correção de múltiplos testes (M.eff) proposto por Moskvina & Schmidt (2008), consistindo numa modificação da correção de Bonferroni que usa um número efetivo de marcadores e considera o DL entre os marcadores para estabelecimento do limiar.

Como mencionado, cada modelo considerou oito formas de parametrização dos marcadores: aditivo (Figura 2A), aditivo diploidizado (diplo-aditivo) (Figura 2B), simplex dominante alternativo (1-dom-alt) (Figura 2C), simplex dominante referência (1-dom-ref) (Figura 2D), duplex dominante alternativo (2-dom-alt) (Figura 2E), duplex dominante referência (2-dom-ref) (Figura 2F), geral diploidizado (diplo-geral) e geral. No modelo aditivo, o efeito do SNP é proporcional à dosagem do alelo menor. Nos modelos simplex dominante, todos os três heterozigotos (AAa , Aaa , $Aaaa$) são correspondentes a um dos homozigotos. Nos modelos duplex dominantes, o estado duplex ($AAaa$) possui o mesmo efeito que todos os outros estados: simplex ($Aaaa$) e nulliplex ($aaaa$), ou triplex ($AAAA$) e quadriplex ($AAAA$). Nos modelos diploidizados, todas as três classes heterozigóticas possuem o mesmo efeito, transformando os dados com dosagem tetraploides ($AAAA$, $AAAa$, $AAaa$, $Aaaa$, $aaaa$) em dosagem diploide (AA , Aa , aa), e os seus modelos de ação gênica englobam os efeitos aditivos. O modelo geral possibilita que qualquer classe de dosagem do SNP seja testada em contraste às demais classes (Ferrão et al., 2018).

Figura 2 – Gráfico com formas de parametrização para tetraploides dos marcadores SNP. (A) Modelo aditivo, (B) Efeito diploidizado aditivo, (C) Efeito 1-dom-alt, (D) Efeito 1-dom-ref, (E) Efeito 2-dom-alt e (F) Efeito 2-dom-ref



Fonte: adaptado de Rosyara et al. (2016)

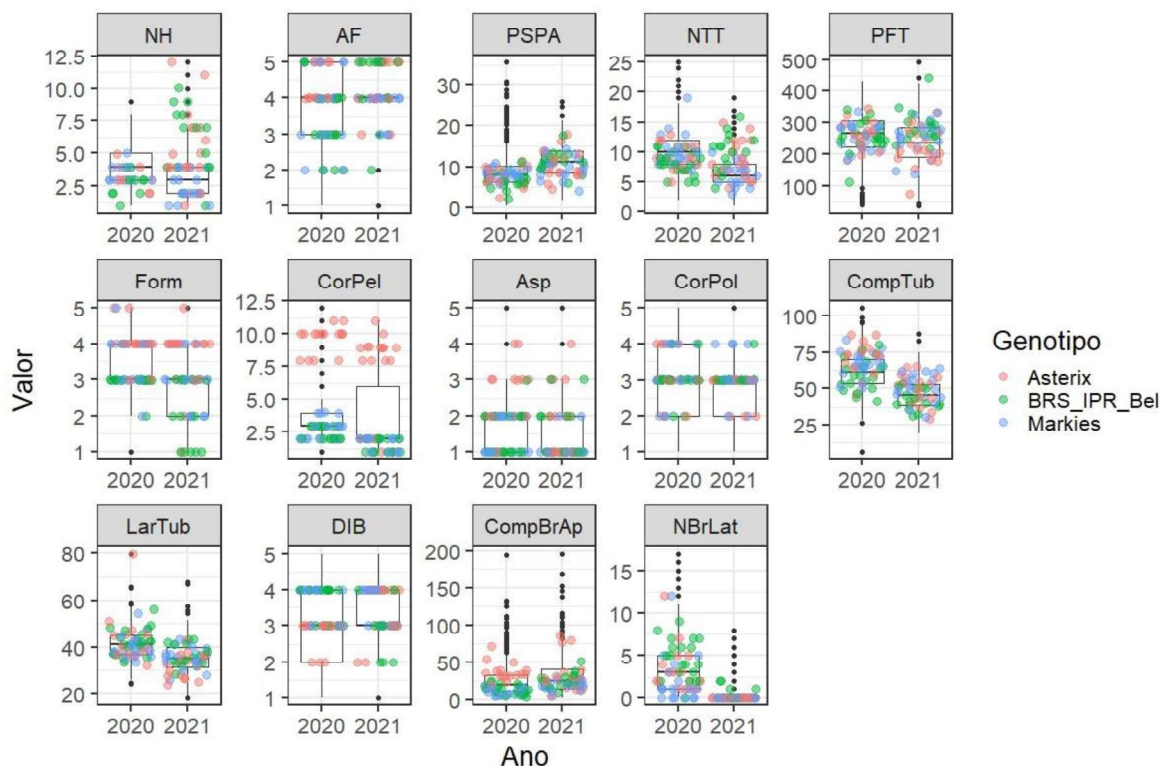
4. RESULTADOS E DISCUSSÃO

4.1. ANÁLISES FENOTÍPICAS

A análise de gráficos *boxplots* (Figura 3) dos dados brutos exibiu a existência de poucos *outliers*, com os caracteres apresentando uma boa distribuição e conformidade com os valores esperados para os caracteres avaliados.

Ao analisar os modelos considerando os efeitos de genótipo como aleatório, foi possível observar significância desses efeitos em todos os caracteres e nos dois anos por meio do teste da razão da verossimilhança (Tabela 2). Os valores de herdabilidade variaram de 0,27 a 0,96 para caracteres morfológicos, e 0,33 a 0,66 para caracteres de produção. Estudos similares feitos por Bisognin et al. (2012), Bradshaw et al. (2008), Fekadu et al. (2013) e Haynes et al. (1996) reportaram herdabilidades aproximadas às encontradas neste estudo (Tabela 2) para os caracteres PSPA, Form, NH e CorPol, respectivamente. Para caracteres relacionados à brotação, observou-se valores de 0,69, 0,69, 0,33 para DIB, CompBrAp, NBrLat, respectivamente. Estudos realizados por Bisognin et al. (2018) revelaram um valor de 0,79 para o caráter DIB.

Figura 3 – Boxplot da distribuição dos dados brutos dos caracteres fenotípicos em dois anos de avaliação



Fonte: a autora.

Tabela 2 – A estatística do teste de razão da verossimilhança (LRT) e seu respectivo *P*-valor (***P* < 0,01, ****P* < 0,001) para efeitos de genótipos e estimativas de herdabilidade, *H*², para os quatorze caracteres em dois anos de avaliação

Classe	Caracteres	Ano 1 (2020)			Ano 2 (2021)		
		LRT	<i>P</i> -valor	<i>H</i> ²	LRT	<i>P</i> -valor	<i>H</i> ²
Morfológicos	Asp	266,42	< 2,2e-16 ***	0,75	67,51	< 2,2e-16 ***	0,52
	CorPel	883,62	< 2,2e-16 ***	0,96	438,69	< 2,2e-16 ***	0,92
	CorPol	313,03	< 2,2e-16 ***	0,81	252,25	< 2,2e-16 ***	0,84
	Form	426,00	< 2,2e-16 ***	0,86	123,45	< 2,2e-16 ***	0,61
	CompTub	85,73	< 2,2e-16 ***	0,49	18,46	8,6e-06 ***	0,27
	LarTub	42,19	4,1e-11 ***	0,37	18,74	7,5e-06 ***	0,27
Produção	AF	92,64	< 2,2e-16 ***	0,51	36,89	6,2e-10 ***	0,37
	PSPA	163,53	< 2,2e-16 ***	0,66	68,58	< 2,2e-16 ***	0,53
	NTT	52,86	1,8e-13 ***	0,39	47,21	3,1e-12 ***	0,33
	NH	8,43	0,001848 **	0,56	60,44	3,7e-15 ***	0,33
	PFT	55,83	3,9e-14 ***	0,43	49,51	9,8e-13 ***	0,44
Brotação	DIB	210,76	< 2,2e-16 ***	0,69	56,49	2,8e-14 ***	0,50
	CompBrAp	194,64	< 2,2e-16 ***	0,69	158,14	< 2,2e-16 ***	0,73
	NBrLat	28,52	4,6e-08 ***	0,30	28,82	3,9e-08 ***	0,36

Fonte: a autora.

Ao tratar o genótipo como efeito fixo, foi possível estimar suas médias ajustadas (Tabela 3). Para cada caráter nos anos 2020 e 2021, notaram-se as respectivas variações de valores fenotípicos mínimos e máximos (entre parênteses): Asp (nota 0,40 e 0,63), CorPel (nota 0,55 e 0,93), CorPol (nota 0,40 e 0,36), CompTub (comprimento de 9,42 mm e 8,65 mm), LarTub (largura de 5,08 mm e 5,78 mm), Form (nota 0,33 e 0,56), NH (1,11 e 1,53 hastes por planta), AF (nota 0,68 e 0,62), PSPA (2,49 e 2,70 g), NTT (2,74 e 2,25 tubérculos), PFT (48,89 e 54,32 g), DIB (nota 0,55 e 0,60), CompBrAp (comprimento de 12,32 mm e 15,21 mm) e NBrLat (número de 2,76 e 0,87 brotos laterais por tubérculo).

As correlações entre as médias ajustadas para o ano 1 (Figura 4A) foram positivas significativas para os caracteres: Form e CompTub (0,63), NTT e NH (0,46), CompTub e LarTub (0,36), AF e NTT (0,34), e PFT e CompTub (0,34). Avaliando os valores de correlações negativas, os seguintes caracteres deram significativas: DIB e CompBrAp (-0,55), NTT e CompTub (-0,32), NTT e LarTub (-0,25) e NH e LarTub (-0,22). Para o ano 2 (Figura 4B), foi possível observar correlações positivas significativas para os seguintes caracteres: Form e CompTub (0,71), NTT e NH (0,54), PSPA e PFT (0,54), NH e AF (0,49), CompTub e LarTub (0,47), AF e NTT (0,39) e PFT e CompTub (0,30). Ao analisar os valores das correlações negativas, foi possível observar esse padrão nos seguintes caracteres: DIB e CompBrAp (-0,61), DIB e NBrLat (-0,27) e CorPol e CompBrAp (-0,25). Entre os dois anos, as correlações positivas mais significativas se repetiram entre eles, com exceção entre PSPA e PFT que foi significativa apenas no ano 2.

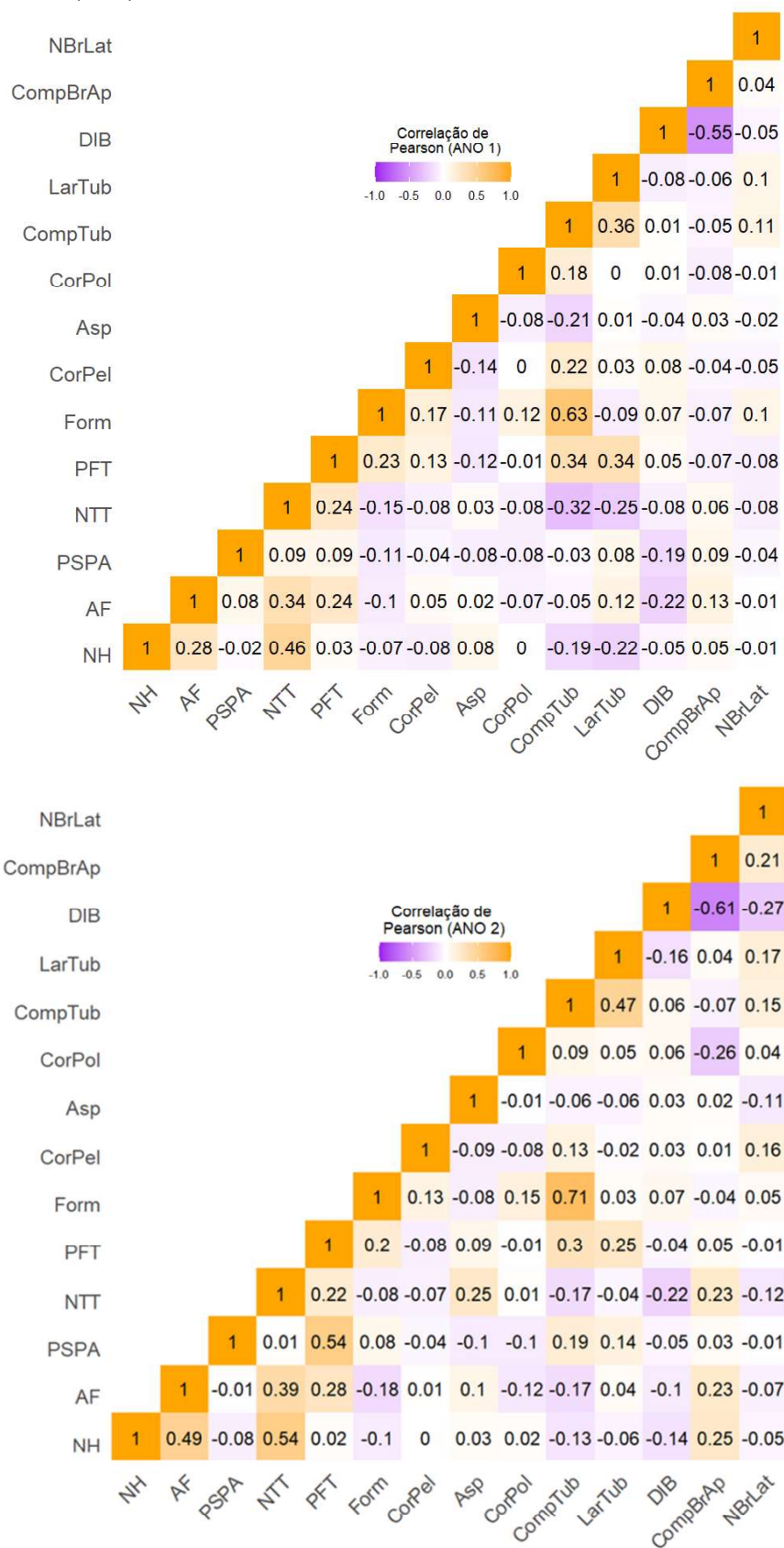
Tabela 3 - Médias ajustadas gerais, mínimas e máximas para quatorze caracteres em dois anos de avaliação

Classe	Caracteres	Ano 1 (2020)		Ano 2 (2021)	
		Média geral	Médias min – max	Média geral	Médias min – max
Morfológicos	Asp	1,93	0,64 – 5,01	1,61	0,53 – 5,38
	CorPel	4,38	0,67 – 12,56	3,71	0,07 – 12,72
	CorPol	2,90	0,57 – 4,86	2,75	0,85 – 5,07
	CompTub	61,37	30,28 – 99,25	46,11	24,37 – 77,06
	LarTub	41,31	24,45 – 56,77	35,94	24,78 – 53,62
	Form	3,20	0,49 – 5,18	2,59	0,46 – 5,45
Produção	NH	3,99	0,99 – 9,00	3,05	0,17 – 7,96
	AF	3,84	1,08 – 5,68	3,92	1,00 – 5,00
	PSPA	8,89	1,10 – 35,63	11,57	3,55 – 20,86
	NTT	10,25	5,49 – 21,5	6,52	2,99 – 15,69
	PFT	261,03	65,21 – 395,33	231,92	45,13 – 450,84
Brotação	DIB	2,89	0,99 – 5,00	3,29	0,99 – 5,00
	CompBrAp	26,09	3,09 – 159,73	33,56	2,66 – 182,37
	NBrLat	3,60	0,00 – 16,11	0,54	0,00 – 6,50

Fonte: a autora.

As correlações negativas tenderam também a serem significativas entre os anos, com destaque a correlação entre DIB e CompBrAp ($-0,55$ e $-0,61$) que possuiu um valor de correlação negativa maior das demais. Zhang et al. (2022) relataram uma correlação positiva de $0,33$ entre as características Largura e Comprimento do Tubérculo (LarTub e CompTub) utilizando 108 acessos. Este valor foi semelhante ao encontrado neste estudo no ano 1 ($0,36$) e relativamente inferior ao encontrado no ano 2 ($0,47$). No estudo realizado por Bisognin et al. (2012), encontrou-se uma correlação entre Form e PFT, apesar deste estudo não ter encontrado uma alta significância entre esses caracteres, foi encontrado correlação entre PFT e CompTub o que pode justificar esta correlação. A correlação positiva entre NTT e NH ($0,46$ e $0,54$ nos respectivos anos) também foi descrita por Silva et al. (2020), em função de uma tendência de as plantas com maior número de hastes tenderem a apresentar um maior número de tubérculos. No entanto, os autores também citaram que NTT está correlacionado negativamente com CompTub e LarTub, ou seja, quanto mais tubérculos por planta menor é o tamanho deles, refletindo no CompTub e LarTub.

Figura 4 – Correlação entre as médias ajustadas de quatorze caracteres em dois anos de avaliação, 2020 (topo) e 2021 (base)



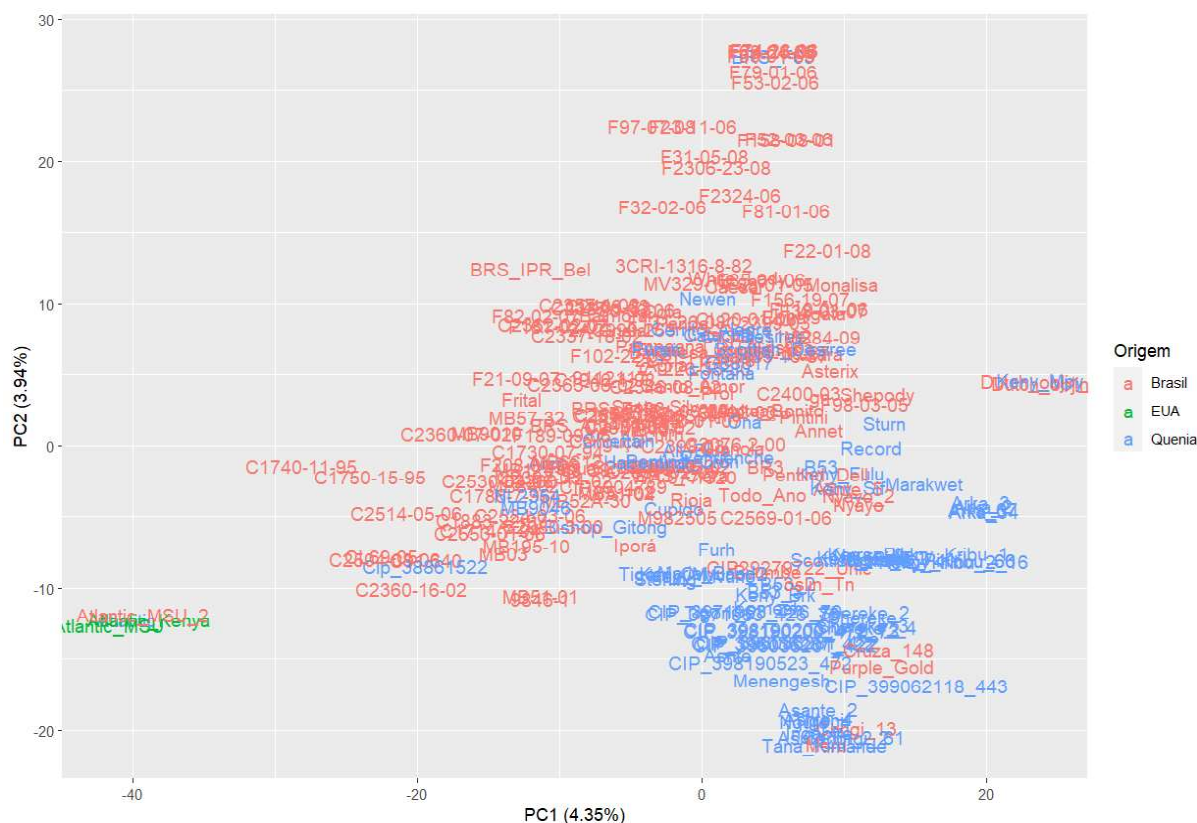
Fonte: a autora.

4.2. ANÁLISES GENOTÍPICAS

Por meio do pacote *RfitTetra* v.1.0, foi possível caracterizar os 240 acessos genotipados em uma das cinco possíveis classes de dosagens alélicas para cada SNP. Após a classificação dos marcadores, dos 8.303 SNPs restaram 6.092 SNP em função da remoção de marcadores cujas categorização foi dada como não confiável. Com os dados ajustados à suas devidas dosagens e ao fazer a filtragem de dados perdidos a 20% e MAF a 5%, o total de SNPs para realizar as demais análises foi para 4.879 SNPs.

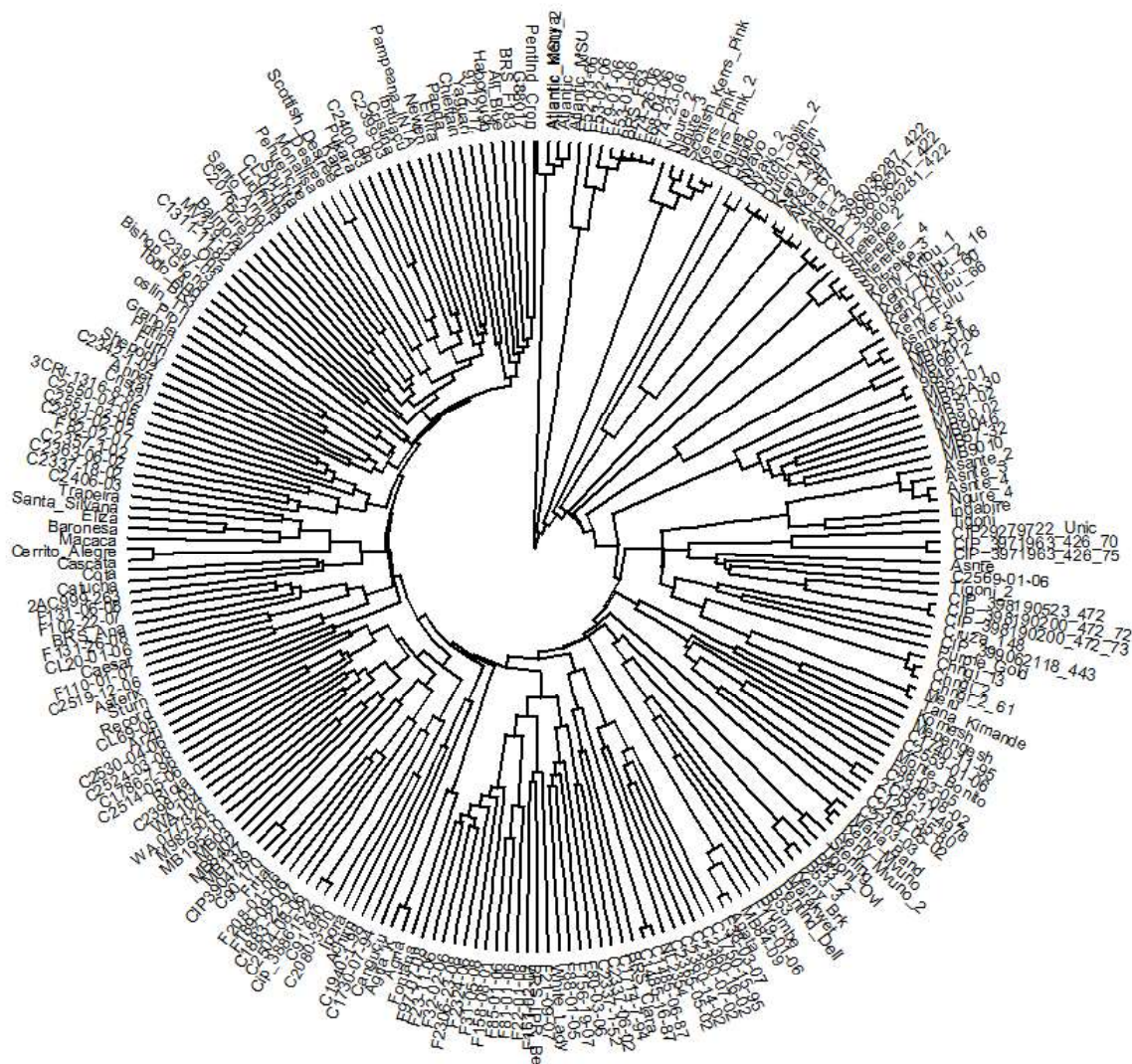
Na análise gráfica dos dois primeiros componentes principais (Figura 5), observou-se um padrão de agrupamento entre genótipos de origem comum a partir de dois países, Brasil e Quênia, com alguns genótipos sobrepostos entre as duas origens, possivelmente devido a intercâmbios de germoplasma. Já os genótipos mais afastados correspondem a amostras de Atlantic com origens diferentes (EUA, Brasil e Quênia), demonstrando que se trata, de fato, da mesma cultivar.

Figura 5 – Análise dos componentes principais (PC) das 240 amostras a partir de dosagem alélica de 4.879 SNPs



Fonte: a autora.

Figura 6 – Dendrograma obtido a partir da matriz de relacionamento genômico realizado entre 240 genótipos de batata genotipados com 4.789 SNPs



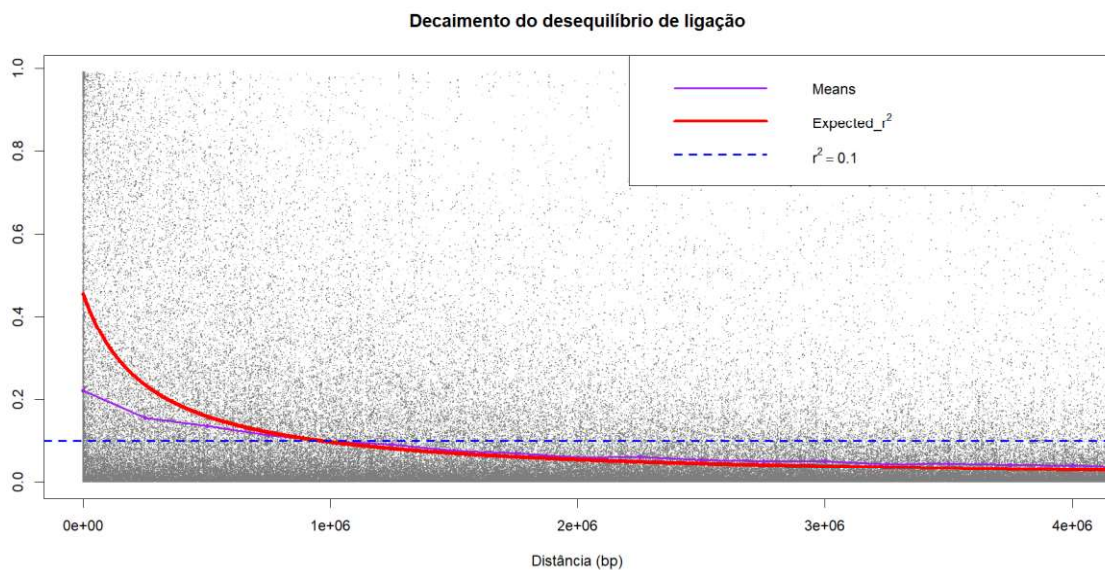
Fonte: a autora.

Analisando o dendrograma obtido a partir da matriz de relacionamento genômico realizado de VanRaden (Figura 6), foi possível observar resultados semelhantes aos encontrados no estudo de PCA. Ou seja, os genótipos com origem semelhante ficaram agrupados mais proximamente, enquanto aqueles de origens distintas ficaram mais afastados. A maior divergência foi dos acessos Atlantic, agrupados entre si, em relação aos demais acessos.

Com o objetivo de avaliar a extensão média dos blocos de desequilíbrio de ligação, médias de r^2 entre pares de SNPs dentro de janelas de 250.000 bp foram utilizadas para ajustar um modelo não-linear de decaimento de DL (Figura 7). A distância que nos fornece valores $r^2 \leq 0.1$ deu-se por volta de 1.000.000 bp, ou seja, 1 Mb. Os valores de DL podem variar por

alguns motivos, dentre os quais destacam-se o tipo de população e de marcadores moleculares, métodos de genotipagem e limites de significância selecionados para seleção do DL. No caso de painéis de batata tetraploide, por possuir poucos eventos históricos de recombinação, alguns autores, como D'hoop et al. (2010) e Simko et al. (2006), obtiveram como conclusão um decaimento por volta de 5 e 10 cM. De acordo com Vos et al. (2017), a distância de 2 e 4 Mb é, em média, equivalente a distância de 5 e 10 cM. Zia et al. (2020) observaram um decaimento por volta de 1,22 Mb utilizando um *array* SolCAP 12K (Ellis et al., 2018), similar ao valor encontrado neste estudo. A queda do desequilíbrio deste estudo foi mais curta do que encontrada em outros estudos, provavelmente devido à menor densidade de SNPs e à composição da população, majoritariamente formada a partir de duas origens.

Figura 7 – Decaimento do desequilíbrio de ligação em painel de 240 acessos de batata tetraploide genotipado com 4.789 SNPs.



Fonte: a autora.

4.3. ASSOCIAÇÃO GENÔMICA AMPLA (GWAS)

Para realizar as análises de GWAS, foram utilizadas médias ajustadas de 137 indivíduos genotipados com 4.789 SNPs. Para controlar a estrutura de parentesco, considerou-se a matriz relacionamento genômico, conforme implementado pela função *set.K()* do pacote GWASpoly baseado nos marcadores de todos os cromossomos exceto daquele cujo SNP está sendo testado (*leave-one-chromossome-out* - LOCO). Além da MAF, também foi utilizado a frequência

genotípica máxima para remover os SNPs que estão abaixo de um limite de frequência ($1 - 5/N$).

Ao se inspecionar gráficos QQ para os modelos Naive, Q+K, Q e K para os quatorze caracteres agrônômicas e de brotação, o modelo Q+K foi o que aparentou melhor ajuste (Figura 10, Figura 8 e Figura 9). No geral, os valores de $-\log_{10}(P)$ observados foram semelhantes aos valores esperados dos modelos, à exceção das regiões de aparente associação. Os limiares de associação, obtido via método de M.eff a 5% de significância, variaram de 4.61 (modelo 1-dom-alt) a 4.82 (modelo aditivo) (Tabela 4). De acordo com o modelo Q+K utilizado, observaram-se 2, 2, 1 e 1 picos de associação para os caracteres morfológicos Form, CorPol, LarTub e CompTub, respectivamente (Figura 7), 4 e 1 para os caracteres de produção PSPA e PFT, respectivamente (Figura 8), e 4 e 1 para os caracteres de brotação DIB e CompBrAp, respectivamente (Figura 9) para os diversos modelos de ação gênica.

Tabela 4 – Limiares $-\log_{10}(P)$ obtidos pelo método ‘M.eff’ para os oito modelos do pacote *GWASpoly*

Modelos	Aditivo	Diplo-aditivo	1-dom-alt	1-dom-ref	2-dom-alt	2-dom-ref	Diplo-geral	Geral
$-\log_{10}(P)$	4.82	4.80	4.61	4.50	4.65	4.62	4.80	4.82

Fonte: a autora.

No presente estudo, foi encontrado pico de associação na posição 48,8 Mb do cromossomo 10 para forma de tubérculo (Form) para o modelo aditivo, diplo-aditivo e 1-dom-ref, além de picos de associação no cromossomo 0 (marcas sem posicionamento no demais cromossomos) na posição 22,1 Mb para os modelos aditivo, diplo-aditivo e 1-dom-ref (Figura 8, Tabela 5). Rosyara et al. (2016), ao realizarem estudo de associação genômica, observaram picos de associação significativas para forma de tubérculo no cromossomo 10 e na posição 48,8 Mb (genoma v.4.03) para o modelo aditivo. Fan et al. (2022) mapearam cinco genes candidatos na região *Ro* do cromossomo 10, putativamente relacionados a Form, empregando um mapa de ligação genética de alta densidade, originado de uma população segregante diploide. A região *Ro* é conhecido como o principal loco que controla formato do tubérculo, sendo o alelo *Ro* dominante e responsável pelo formato redondo, e o alelo *ro* recessivo responsável pelo formato longo (Chen et al., 2018).

Para o caráter CorPol, foi encontrada uma região de maior efeito no cromossomo 3 (42,6 Mb), responsável pelo aumento do valor deste caráter para os modelos diplo-aditivo e 1-dom-alt. Também no cromossomo 3, duas outras regiões significativas, uma a 11,2 Mb e outra a 49,4 Mb, contribuem para a diminuição da coloração da polpa para o modelo aditivo, diplo-aditivo,

1-dom-alt, 2-dom-alt e 1-dom-ref (Figura 8, Tabela 5). Em um estudo feito por Uitdewilligen et al. (2013) utilizando 84 cultivares genotipadas com 129.156 variantes, observou-se um QTL no cromossomo 3 na região de aproximadamente 68,3 Mb (genoma v.3) para cor da polpa, cujas 27 variantes significativas encontravam-se localizadas dentro e perto do gene *CHY2* (*β-caroteno hidroxilase*), que codifica para uma enzima parte da via de carotenoides.. O estudo realizado por Pandey et al. (2022) encontrou dois QTLs para CorPol, sendo eles no cromossomo 3 e nas posições de 48,5 Mb e 43,9 Mb (genoma v.4.03). Ao analisar a região entre 49,4 Mb e 43,9 Mb do cromossomo 3 no genoma v.4.03, região encontrada com associação neste estudo e no estudo de Pandey et al. (2022), respectivamente, foi possível observar o gene *CHY2* (*PGSC0003DMT400026363*) na posição 43,9 Mb e um possível gene relacionado a enzima *CHY2* (*PGSC0003DMT400024575*), localizado por volta de 44 Mb no cromossomo 3, se aproximando mais às regiões encontradas neste estudo.

Na região entre 36,9 Mb e 37,9 Mb do cromossomo 2, em que houve evidência de associação para CompTub, há um gene, nomeados como *PGSC0003DMT400043117*, responsável pela síntese da enzima Sucrose synthase (SuSy). Uma revisão feita por Stein & Granot (2019) sobre uma visão geral da síntese de sacarose em diversas espécies vegetais mostra que a enzima SuSy desempenha uma função fundamental no metabolismo da sacarose, principalmente nos tubérculos e raízes. Quando a atividade de SuSy é reduzida, as plantas também terão um crescimento reduzido, além de redução da síntese de amido. Já quando há superexpressão da SuSy, ocorre um aumento no crescimento da planta. Como foi observado correlação positiva entre PFT e CompTub, o gene proposto acima pode ser considerado como candidato para estudos de CompTub.

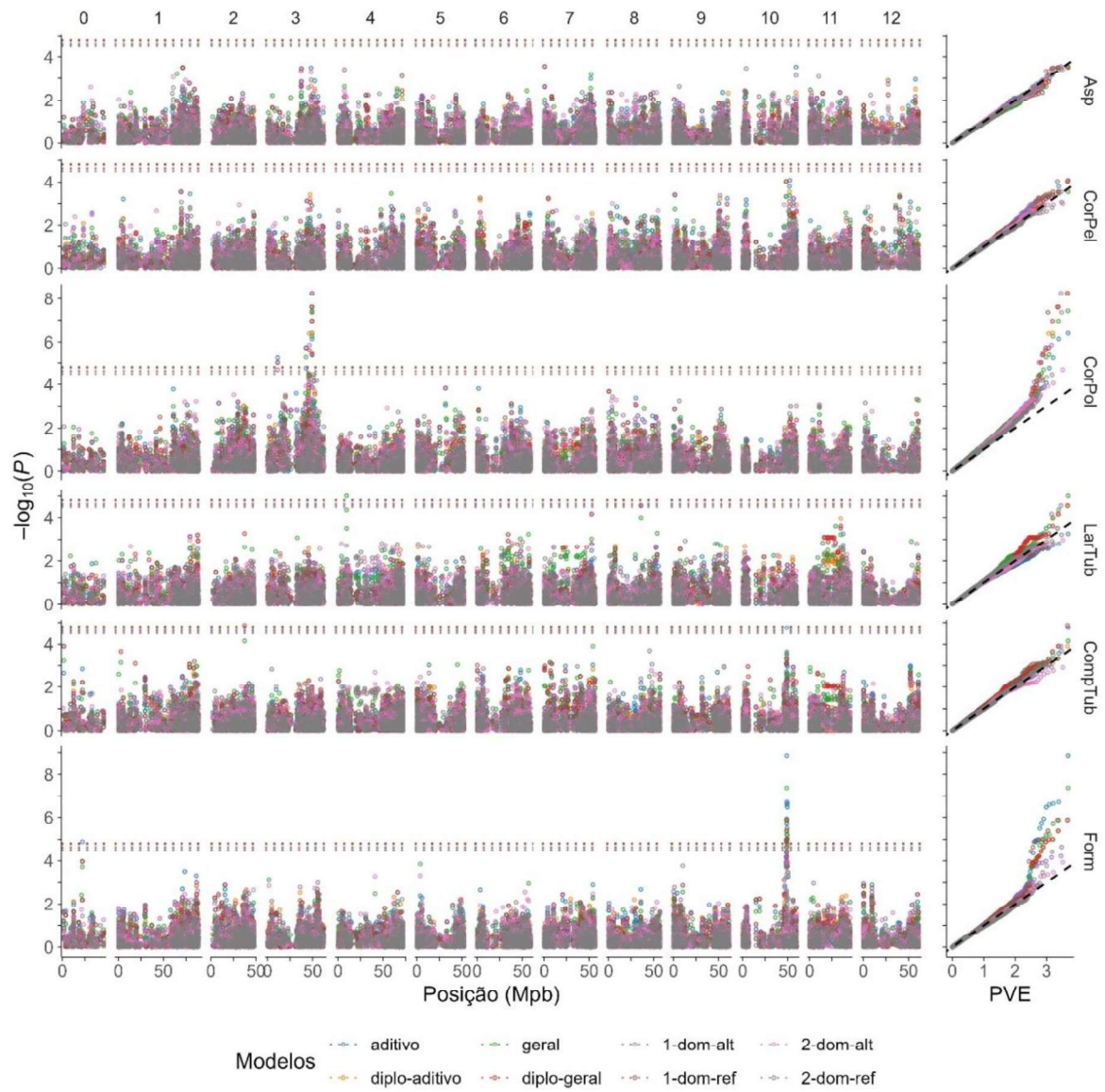
Bisognin et al. (2018) realizaram uma análise de QTL para dormência e brotação de tubérculos, em uma população segregante diploide (MSX902) constituída de 131 indivíduos genotipados com 8.303 SNPs da plataforma Infinium. Nos cromossomos 2, 3 e 7 foram detectados QTLs associados à liberação de dormência e da dominância apical. Como citado pelos mesmos autores, genes relacionados a resposta ao ácido abscísico (ABA), genes *TUBBY* e mudança hormonal das giberelinas (Giberelina 2-oxidase 2 e Giberelina 3-beta-dioxigenase) estão associados a dormência e liberação da dormência apical, influenciando, assim, o comprimento do broto apical. Apesar de os nossos estudos não encontrarem associações para DIB nestes cromossomos, foram encontradas associações para o CompBrAp no cromossomo 2 (27,5 Mb) e 7 (52,1 Mb), próximas às regiões descritas no estudo de Bisognin et al., (2018), nos cromossomos 2 (27,1 Mb) e 7 (47,4 Mb). Um gene próximo a posição 20 Mb no

cromossomo 2, nomeado como *PGSC0003DMT400024385*, está relacionado à síntese de giberelina. No cromossomo 7, foram encontrados os genes *PGSC0003DMT400071045*, *PGSC0003DMT400071043*, *PGSC0003DMT400071094* seguidos um do outro e que também estão relacionados na síntese de giberelina, e o gene *PGSC0003DMT400018209* envolvido na sinalização de ABA.

Na região de associação ao caráter DIB no cromossomo 12 (4,9 Mb), observou-se um gene nomeado de *PGSC0003DMT400000852* com definição de Aldeído desidrogenase, que transforma o acetaldeído em ácido acético, e outro gene nomeado de *PGSC0003DMT400002750*, uma proteína de ligação responsiva à presença de etileno. Em estudos feitos por Claassens (2002), ao expor tubérculos de batata ao etileno, acetaldeído, e ácido acético, após 5 dias houve 90, 11 e 9% de brotação dos tubérculos, respectivamente. Indicando que os produtos acetaldeído e ácido acético inibem a brotação, ao contrário do etileno, que induz a brotação nos tubérculos. Outro gene candidato é o *PGSC0003DMT400000858* responsável pela ativação da proteína GTPase para uma proteína contendo gene do fator de ribosilação de ADP (ARF). Liu et al. (2012), ao realizarem o estudo de identificação de genes associados à liberação de dormência de tubérculos de batata, observaram que a variação de expressão do gene *ARF1* no tubérculo, tornando-o um gene putativo na regulação da dormência e brotação na batata. Uma das explicações é que o transporte de fito-hormônios realizados por *ARF1* e a mudança de amido em sacarose são necessários para tecidos que possuem um rápido crescimento, como os brotos (Liu et al., 2012).

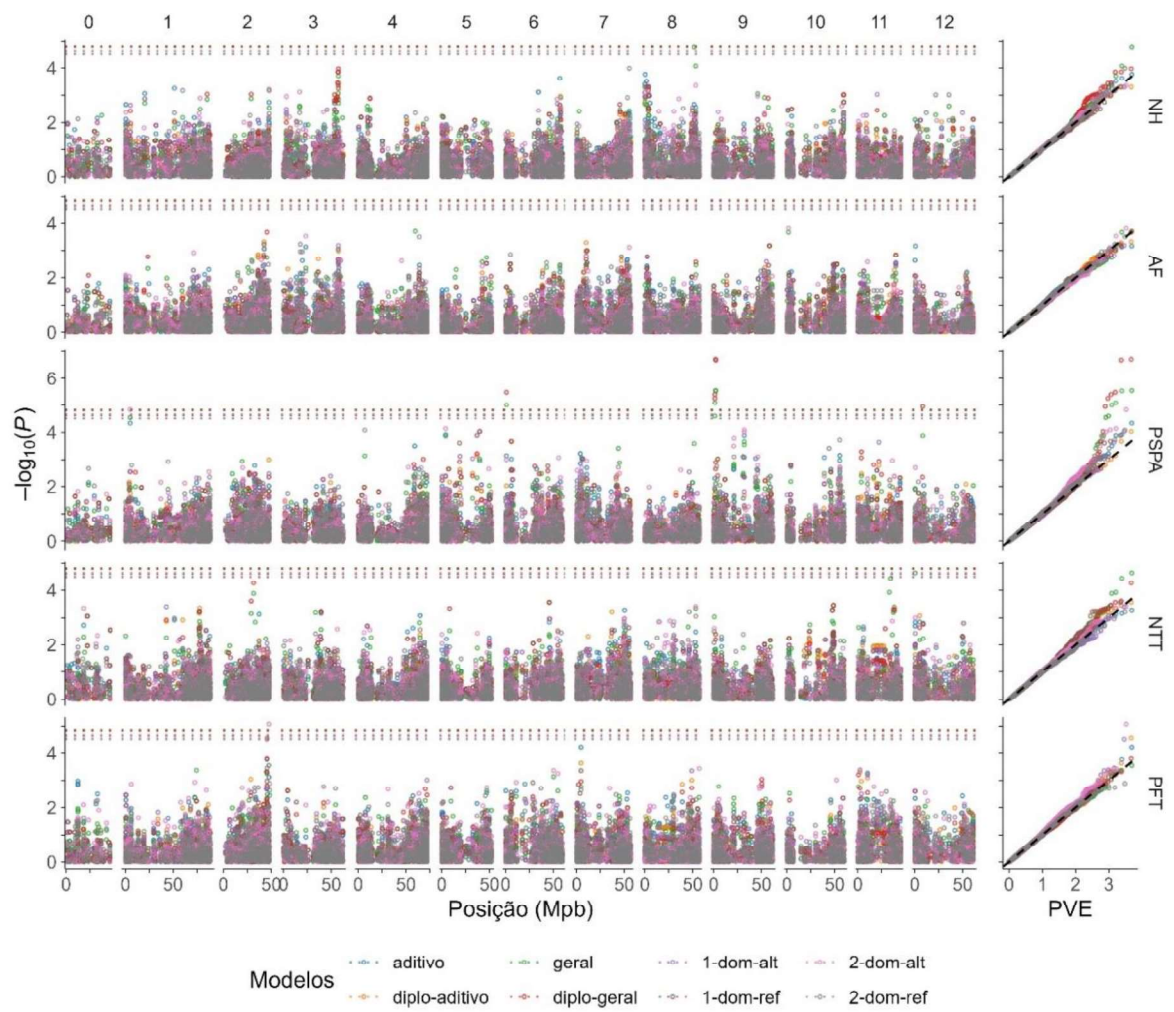
Um gene candidato para a características PFT é *PGSC0003DMT400052115*, próximo à região 47,6 Mb no cromossomo 2, constitui-se de uma subunidade regulatória da proteína fosfatase 2A. García et al. (2023) estudaram o funcionamento da proteína para diversas características agrônomicas e parâmetros fisiológicos. Eles puderam concluir que folhas de plantas batata com super expressão da subunidade catalítica da proteína fosfatase 2A exibem altos níveis de carboidratos e o transporte de sacarose com uma regulação negativa. Isto sugere um transporte diminuído de sacarose das folhas para os tubérculos, contribuindo para uma redução do peso dos tubérculos nas plantas.

Figura 8 – Gráficos de Manhattan e de QQ dos caracteres morfológicos



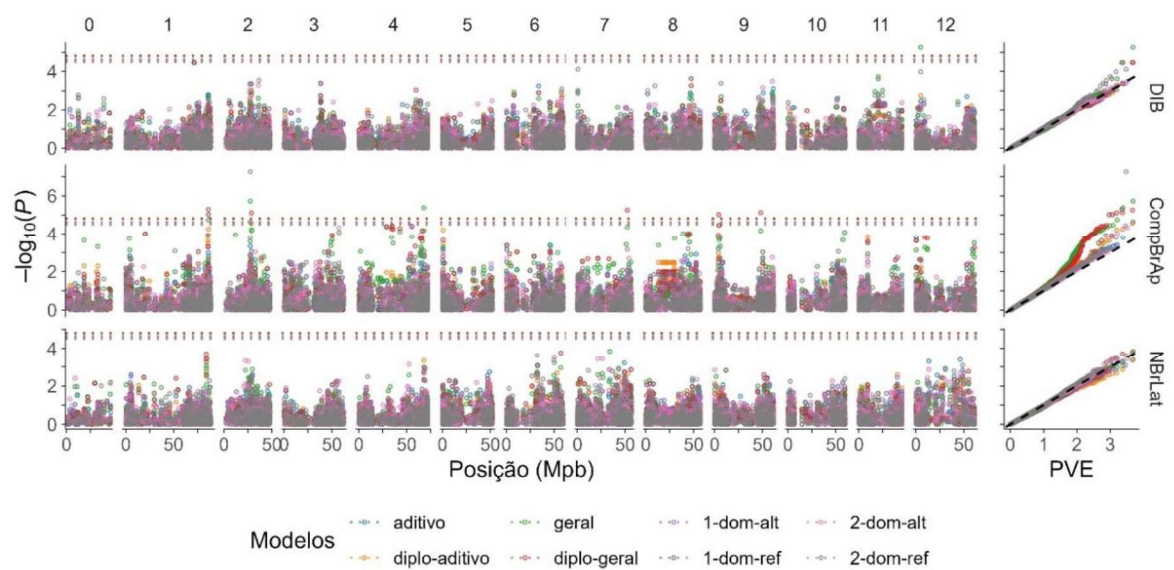
Fonte: a autora.

Figura 9 – Gráficos de Manhattan e de QQ dos caracteres relacionados a produção



Fonte: a autora.

Figura 10 – Gráficos de Manhattan e de QQ dos caracteres relacionados a brotação



Fonte: a autora.

Tabela 5 – Caracteres com picos de associação e seus respectivos marcadores mais significativos

Caráter	Modelo gênico	Marcador	Crom.	Posição	Score	Efeito
Caracteres morfológicos						
Form	Aditivo	solcap_snp_c2_56344	0	22129989	4,89	0,48
	Aditivo	solcap_snp_c1_8020	10	48863048	8,86	0,52
	Diplo-Aditivo	solcap_snp_c2_25471	10	48808404	5,88	0,67
	1-dom-ref	solcap_snp_c2_25471	10	48808404	5,88	0,67
	2-dom-ref	solcap_snp_c1_8020	10	48863048	4,65	0,72
	Geral	solcap_snp_c1_8020	10	48863048	7,38	
	Diplo-Geral	solcap_snp_c2_25471	10	48808404	5,88	
CorPol	Aditivo	solcap_snp_c2_5736	3	11287783	5,25	-0,42
	Diplo-Aditivo	solcap_snp_c1_9141	3	42570907	5,73	1,19
	Diplo-Aditivo	solcap_snp_c2_5736	3	11287783	5,04	-0,69
	1-dom-alt	solcap_snp_c1_9141	3	42570907	5,73	1,19
	1-dom-alt	solcap_snp_c2_5736	3	11287783	5,04	-0,69
	1-dom-ref	solcap_snp_c2_20227	3	49231930	7,61	-0,94
	2-dom-alt	solcap_snp_c2_5736	3	11287783	4,7	-0,97
	Geral	solcap_snp_c2_20282	3	49393157	7,42	
	Diplo-Geral	solcap_snp_c2_20282	3	49393157	8,25	
	LarTub	Geral	solcap_snp_c2_26790	4	9520486	5,01
CompTub	Diplo-Geral	solcap_snp_c2_42172	2	37439083	4,87	
Caracteres relacionados a produção						
PFT	2-dom-alt	solcap_snp_c2_38552	2	47627484	5,08	47,69
PSPA	2-dom-alt	solcap_snp_c2_55010	1	4952171	4,82	8,00
	Geral	solcap_snp_c2_30612	6	833354	5	
	Geral	solcap_snp_c1_3597	9	2081425	5,54	
	Diplo-Geral	solcap_snp_c2_30612	6	833354	5,48	
	Diplo-Geral	solcap_snp_c1_3597	9	2081425	6,7	
	Diplo-Geral	solcap_snp_c2_34762	12	8063612	4,94	
	Caracteres relacionados a brotação					
CompBrAp	1-dom-ref	solcap_snp_c2_14703	1	86354960	5,3	18,85
	2-dom-ref	solcap_snp_c2_41980	2	27557527	7,27	39,75
	Geral	solcap_snp_c2_41980	2	27557527	5,72	
	Geral	solcap_snp_c2_34865	4	67983299	5,38	
	Diplo-Geral	solcap_snp_c2_14703	1	86354960	5,02	
	Diplo-Geral	solcap_snp_c2_19692	2	28746087	5,11	
	Diplo-Geral	solcap_snp_c2_42760	7	52191287	5,26	
DIB	Geral	solcap_snp_c2_49209	12	4907175	5,27	

Fonte: a autora.

5. CONCLUSÃO

Por meio deste trabalho, ao realizar os estudos de população e de parentesco observou-se uma nítida estruturação entre acessos de origens distintas, o que a nível de GWAS é importante para controlar seu viés utilizando o modelo $Q + K$. Ao analisar a queda do desequilíbrio de ligação por volta de 1 Mb foi possível concluir que, apesar do número reduzido de SNP, os marcadores cobriram uma boa parte do genoma. Com isso, após verificar pontos importantes de influência na análise de GWAS, foi encontrar associações significativas entre SNP e caracteres para o modelo $Q + K$ seguindo as parametrizações aditivo, diplo-aditivo, 1-dom-alt, 1-dom-ref, 2-dom-alt, 2-dom-ref, diplo-geral e geral. Dentre elas, os QTLs encontrados em Form e CorPol validaram o presente estudo com base em diversas literaturas. Sabendo disso, encontraram-se nas regiões de associação dois genes candidatos para CompTub, cinco genes candidatos para CompBrAp, três genes candidatos para DIB e um gene candidato para PFT. Correlações importantes entre CompBrAp e DIB (-0,55) e PFT e CompTub (0,33) foram observadas e nos ajudam a entender alguns resultados de co-localização de QTLs aqui encontrados. Novos estudos de validação dos genes propostos são necessários para entender como funciona o controle genético desses caracteres e a forma em que os genes controlam a expressão desses caracteres.

Estes resultados são de grande importância pois, com eles, será possível implementar novas estratégias nos programas de melhoramento genético de batata tetraploide. Tendo como foco final, diminuir no tempo e custos no processo de melhoramento, o que é de extrema importância devido ser uma cultura rústica e sem muitos financiamentos para estudo. Ao se conseguir selecionar genótipos mais tardios para brotação em etapas iniciais do processo, será possível estender o tempo de prateleira e diminuir os riscos envolvidos no consumo de tubérculos que apresentam brotação.

6. REFERÊNCIAS

- Adu, G. B., Badu-Apraku, B., Akromah, R., Garcia-Oliveira, A. L., Frederick Justice, A., & Melaku, G. (2019). Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. *PLOS ONE*, *14*(4), e0214810. <https://doi.org/10.1371/journal.pone.0214810>
- Amadeu, R. R., Cellon, C., Olmstead, J. W., Garcia, A. A. F., Resende, M. F. R., & Muñoz, P. R. (2016). AGHmatrix: R Package to Construct Relationship Matrices for Autotetraploid and Diploid Species: A Blueberry Example. *The Plant Genome*, *9*(3). <https://doi.org/10.3835/plantgenome2016.01.0009>
- Andersson, M., Turesson, H., Nicolia, A., Fält, A.-S., Samuelsson, M., & Hofvander, P. (2017). Efficient targeted multiallelic mutagenesis in tetraploid potato (*Solanum tuberosum*) by transient CRISPR-Cas9 expression in protoplasts. *Plant Cell Reports*, *36*(1), 117–128. <https://doi.org/10.1007/s00299-016-2062-3>
- Ashraf, B. H., Byrne, S., Fé, D., Czaban, A., Asp, T., Pedersen, M. G., Lenk, I., Roulund, N., Didion, T., Jensen, C. S., Jensen, J., & Janss, L. L. (2016). Estimating genomic heritabilities at the level of family-pool samples of perennial ryegrass using genotyping-by-sequencing. *Theoretical and Applied Genetics*, *129*(1), 45–52. <https://doi.org/10.1007/s00122-015-2607-9>
- Aydin, M. A., & Öztürk, G. (2022). Klon 6/7 X 101 Patates (*Solanum tuberosum* L.) Melezinin İlk Generasyonunda Verim Komponentlerine Dayalı Klon Seleksiyonu. *ANADOLU Ege Tarımsal Araştırma Enstitüsü Dergisi*, *32*(1), 40–49. <https://doi.org/10.18615/anadolu.1129968>
- Barrell, P. J., Meiyalaghan, S., Jacobs, J. M. E., & Conner, A. J. (2013). Applications of biotechnology and genomics in potato improvement. *Plant Biotechnology Journal*, *11*(8), 907–920. <https://doi.org/10.1111/pbi.12099>
- Berdugo-Cely, J. A., Céron-Lasso, M. del S., & Yockteng, R. (2023). Phenotypic and molecular analyses in diploid and tetraploid genotypes of *Solanum tuberosum* L. reveal promising genotypes and candidate genes associated with phenolic compounds, ascorbic acid contents, and antioxidant activity. *Frontiers in Plant Science*, *13*. <https://doi.org/10.3389/fpls.2022.1007104>
- Bernstein, R., Plate, M., Hoppe, A., & Bienefeld, K. (2018). *Computing inbreeding coefficients and the inverse numerator relationship matrix in large populations of honey bees. Journal of Animal Breeding and Genetics*. <https://onlinelibrary.wiley.com/doi/10.1111/jbg.12347>
- Bhatia, D., & Bajwa, G. S. (2022). Molecular Marker Techniques and Recent Advancements. *Em Genotyping by Sequencing for Crop Improvement* (p. 1–21). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119745686.ch1>
- Bisognin, D. A., Manrique-Carpintero, N. C., & Douches, D. S. (2018). QTL Analysis of Tuber Dormancy and Sprouting in Potato. *American Journal of Potato Research*, *95*(4), 374–382. <https://doi.org/10.1007/s12230-018-9638-0>

- Bisognin, D. A., Rigão, M. H., Lopes, S. J., & Storck, L. (2012). Heritability and correlation among potato tuber traits. *Crop Breeding and Applied Biotechnology*, *12*, 215–219. <https://doi.org/10.1590/S1984-70332012000300009>
- Boivin, M., Bourdeau, N., Barnabé, S., & Desgagné-Penix, I. (2020). Sprout Suppressive Molecules Effective on Potato (*Solanum tuberosum*) Tubers during Storage: A Review. *American Journal of Potato Research*, *97*(5), 451–463. <https://doi.org/10.1007/s12230-020-09794-0>
- Bonierbale, M. W., Amoros, W. R., Salas, E., & de Jong, W. (2020). Potato Breeding. Em H. Campos & O. Ortiz (Orgs.), *The Potato Crop: Its Agricultural, Nutritional and Social Contribution to Humankind* (p. 163–217). Springer International Publishing. https://doi.org/10.1007/978-3-030-28683-5_6
- Bradshaw, J. E. (2007). The Canon of Potato Science: 4. Tetrasomic Inheritance. *Potato Research*, *50*(3), 219–222. <https://doi.org/10.1007/s11540-008-9041-1>
- Bradshaw, J. E., Hackett, C. A., Pande, B., Waugh, R., & Bryan, G. J. (2008). QTL mapping of yield, agronomic and quality traits in tetraploid potato (*Solanum tuberosum* subsp. *Tuberosum*). *Theoretical and Applied Genetics*, *116*(2), 193–211. <https://doi.org/10.1007/s00122-007-0659-1>
- Campos, H., & Ortiz, O. (Orgs.). (2020). *The Potato Crop: Its Agricultural, Nutritional and Social Contribution to Humankind*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-28683-5>
- Chen, N., Zhu, W., Xu, J., Duan, S., Bian, C., Hu, J., Wang, W., Li, G., & Jin, L. (2018). Molecular marker development and primary physical map construction for the tuber shape Ro gene locus in diploid potato (*Solanum tuberosum* L.). *Molecular Breeding*, *39*(1), 6. <https://doi.org/10.1007/s11032-018-0913-z>
- Claassens, M. M. J. (2002). *Carbohydrate metabolism during potato tuber dormancy and sprouting*.
- da Silva Pereira, G., Mollinari, M., Schumann, M. J., Clough, M. E., Zeng, Z.-B., & Yenko, G. C. (2021). The recombination landscape and multiple QTL mapping in a *Solanum tuberosum* cv. ‘Atlantic’-derived F1 population. *Heredity*, *126*(5), Artigo 5. <https://doi.org/10.1038/s41437-021-00416-x>
- del Rio, A., & Bamberg, J. (2021). An AFLP Marker Core Subset for the Cultivated Potato Species *Solanum phureja* (*Solanum tuberosum* L. subsp. *Andigenum*). *American Journal of Potato Research*, *98*(5), 374–383. <https://doi.org/10.1007/s12230-021-09849-w>
- Deperi, Sofia. I., Tagliotti, Martín. E., Bedogni, M. C., Manrique-Carpintero, N. C., Coombs, J., Zhang, R., Douches, D., & Huarte, Marcelo. A. (2018). Discriminant analysis of principal components and pedigree assessment of genetic diversity and population structure in a tetraploid potato panel using SNPs. *PLOS ONE*, *13*(3), e0194398. <https://doi.org/10.1371/journal.pone.0194398>

- D'hoop, B. B., Paulo, M. J., Kowitwanich, K., Sengers, M., Visser, R. G. F., van Eck, H. J., & van Eeuwijk, F. A. (2010). Population structure and linkage disequilibrium unravelled in tetraploid potato. *Theoretical and Applied Genetics*, *121*(6), 1151–1170. <https://doi.org/10.1007/s00122-010-1379-5>
- Douches, D. S., & Jastrzebski, K. (1993). 44 - Potato: *Solanum tuberosum* L. Em G. Kalloo & B. O. Bergh (Orgs.), *Genetic Improvement of Vegetable Crops* (p. 605–644). Pergamon. <https://doi.org/10.1016/B978-0-08-040826-2.50048-5>
- Elhaik, E. (2022). Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports*, *12*(1), 14683. <https://doi.org/10.1038/s41598-022-14395-4>
- Ellis, D., Chavez, O., Coombs, J., Soto, J., Gomez, R., Douches, D., Panta, A., Silvestre, R., & L., N. (2018). *Genetic identity in genebanks: Application of the SolCAP 12K SNP array in fingerprinting and diversity analysis in the global in trust potato collection*. *61*(7). <https://doi.org/10.1139/gen-2017-0201>
- Endelman, J. B., Carley, C. A. S., Bethke, P. C., Coombs, J. J., Clough, M. E., da Silva, W. L., De Jong, W. S., Douches, D. S., Frederick, C. M., Haynes, K. G., Holm, D. G., Miller, J. C., Muñoz, P. R., Navarro, F. M., Novy, R. G., Palta, J. P., Porter, G. A., Rak, K. T., Sathuvalli, V. R., ... Yencho, G. C. (2018). Genetic Variance Partitioning and Genome-Wide Prediction with Allele Dosage Information in Autotetraploid Potato. *Genetics*, *209*(1), 77–87. <https://doi.org/10.1534/genetics.118.300685>
- Fan, G., Wang, Q., Xu, J., Chen, N., Zhu, W., Duan, S., Yang, X., De Jong, W. S., Guo, Y., Jin, L., & Li, G. (2022). Fine Mapping and Candidate Gene Prediction of Tuber Shape Controlling Ro Locus Based on Integrating Genetic and Transcriptomic Analyses in Potato. *International Journal of Molecular Sciences*, *23*(3), Artigo 3. <https://doi.org/10.3390/ijms23031470>
- FAOSTAT. (2022). *FAOSTAT*. <https://www.fao.org/faostat/en/#data/QCL/visualize>
- Fekadu, A., Petros, Y., & Zelleke, H. (2013). *Genetic variability and association between agronomic characters in some potato (Solanum tuberosum L.) genotypes in SNNPRS, Ethiopia*. *5*(8), 523–528. <https://doi.org/10.5897/IJBC2013.0548>
- Felcher, K. J., Coombs, J. J., Massa, A. N., Hansey, C. N., Hamilton, J. P., Veilleux, R. E., Buell, C. R., & Douches, D. S. (2012). Integration of Two Diploid Potato Linkage Maps with the Potato Genome Sequence. *PLoS ONE*, *7*(4), e36347. <https://doi.org/10.1371/journal.pone.0036347>
- Ferrão, L. F. V., Benevenuto, J., Oliveira, I. de B., Cellon, C., Olmstead, J., Kirst, M., Resende, M. F. R., & Munoz, P. (2018). Insights Into the Genetic Basis of Blueberry Fruit-Related Traits Using Diploid and Polyploid Models in a GWAS Context. *Frontiers in Ecology and Evolution*, *6*. <https://www.frontiersin.org/articles/10.3389/fevo.2018.00107>
- Flint-Garcia, S., Thornsberry, J., & Buckler, E. (2003). Structure of Linkage Disequilibrium in Plants. *Annual review of plant biology*, *54*, 357–374. <https://doi.org/10.1146/annurev.arplant.54.031902.134907>

- Fortes, G. R. de L., & Pereira, J. E. S. (2003). *Classificação e descrição botânica*. <http://www.alice.cnptia.embrapa.br/handle/doc/502639>
- García, M. N. M., Baroli, I., Cortelezzi, J. I., Zubillaga, M., & Capiati, D. A. (2023). Genetic manipulation of protein phosphatase 2A affects multiple agronomic traits and physiological parameters in potato (*Solanum tuberosum*). *Functional Plant Biology*, *50*(12), 1117–1129. <https://doi.org/10.1071/FP23163>
- Gerard, D. (2021). Pairwise linkage disequilibrium estimation for polyploids. *Molecular Ecology Resources*, *21*(4), 1230–1242. <https://doi.org/10.1111/1755-0998.13349>
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J., & Thompson, R. (2015). *ASReml User Guide Release 4.1 Structural Specification*. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. www.vsnl.co.uk
- Gościnną, K., Retmańska, K., Wszelaczyńska, E., & Pobereźny, J. (2024). Influence of Edible Potato Production Technologies with the Use of Soil Conditioner on the Nutritional Value of Tubers. *Agronomy*, *14*(3), Artigo 3. <https://doi.org/10.3390/agronomy14030549>
- Haynes, K. G., Siczka, J. B., Henninger, M. R., & Fleck, D. L. (1996). Clone × Environment Interactions for Yellow-flesh Intensity in Tetraploid Potatoes. *Journal of the American Society for Horticultural Science*, *121*(2), 175–177. <https://doi.org/10.21273/JASHS.121.2.175>
- Hijmans, R. J., & Spooner, D. M. (2001). Geographic distribution of wild potato species. *American Journal of Botany*, *88*(11), 2101–2112. <https://doi.org/10.2307/3558435>
- Hill, W. G., & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, *38*(6), 226–231. <https://doi.org/10.1007/BF01245622>
- Horodyska, J., Pigat, S., Wonik, J., Bompola, F., Cai, D., Rehm, C. D., & Gonzalez, T. D. (2021). Impact of sociodemographic factors on the consumption of tubers in Brazil. *Nutrition Journal*, *20*(1), 54. <https://doi.org/10.1186/s12937-021-00709-1>
- Hotelling, H. (1933). *Analysis of a complex of statistical variables into principal components*.
- Hotelling, H. (1936). *Simplifield calculation of principal components*. 498–520.
- Igwe, D. O. (2021). Molecular Markers: Potential Facilitators in Plant Breeding and Germplasm Conservation. Em O. O. Babalola (Org.), *Food Security and Safety: African Perspectives* (p. 611–646). Springer International Publishing. https://doi.org/10.1007/978-3-030-50672-8_31
- Ivanova-Pozdejeva, A., Kivistik, A., Kübarsepp, L., Tähtjärv, T., Tsahkna, A., Droz, E., & Laanemets, K. (2022). Fingerprinting of Potato Genotypes from Estonian Genebank Collection Using SSR Markers. *Potato Research*, *65*(1), 153–170. <https://doi.org/10.1007/s11540-021-09514-z>

- Kaunda, J. S., & Zhang, Y.-J. (2019). The Genus Solanum: An Ethnopharmacological, Phytochemical and Biological Properties Review. *Natural Products and Bioprospecting*, 9(2), 77–137. <https://doi.org/10.1007/s13659-019-0201-6>
- Koyuturk, Ö., & Yilmaz, G. (2022). EARLY GENERATION SELECTION IN DIFFERENT LOCATIONS IN POTATO BREEDING PROGRAM. *Eurasian Journal of Science Engineering and Technology*, 3(1), 1–13. <https://doi.org/10.55696/ejset.1031168>
- Liu, B., Zhang, N., Wen, Y., Si, H., & Wang, D. (2012). Identification of differentially expressed genes in potato associated with tuber dormancy release. *Molecular Biology Reports*, 39(12), 11277–11287. <https://doi.org/10.1007/s11033-012-2037-6>
- Machida-Hirano, R. (2015). Diversity of potato genetic resources. *Breeding Science*, 65(1), 26–40. <https://doi.org/10.1270/jsbbs.65.26>
- Mollinari, M., & Serang, O. (2015). Quantitative SNP Genotyping of Polyploids with MassARRAY and Other Platforms. Em J. Batley (Org.), *Plant Genotyping: Methods and Protocols* (p. 215–241). Springer. https://doi.org/10.1007/978-1-4939-1966-6_17
- Moskvina, V., & Schmidt, K. M. (2008). On multiple-testing correction in genome-wide association studies. *Genetic Epidemiology*, 32(6), 567–573. <https://doi.org/10.1002/gepi.20331>
- Mudege, N. N., Mayanja, S., Nyaga, J., Nakitto, M., Tinyiro, S. E., Magala, D. B., Achora, J. C., Kisakye, S., Bamwirire, D., Mendes, T., & Muzhingi, T. (2021). Prioritising quality traits for gender-responsive breeding for boiled potato in Uganda. *International Journal of Food Science & Technology*, 56(3), 1362–1375. <https://doi.org/10.1111/ijfs.14840>
- Naeem, M., Demirel, U., Yousaf, M. F., Caliskan, S., & Caliskan, M. E. (2021). Overview on domestication, breeding, genetic gain and improvement of tuber quality traits of potato using fast forwarding technique (GWAS): A review. *Plant Breeding*, 140(4), 519–542. <https://doi.org/10.1111/pbr.12927>
- Nick, C., & Borém, A. (2017). *Batata: Do plantio à colheita*. Ed. UFV.
- Odilbekov, F., Selga, C., Ortiz, R., Chawade, A., & Liljeroth, E. (2020). QTL Mapping for Resistance to Early Blight in a Tetraploid Potato Population. *Agronomy*, 10(5), Artigo 5. <https://doi.org/10.3390/agronomy10050728>
- Pandey, J. (2021). *MOLECULAR CHARACTERIZATION, GENOME-WIDE ASSOCIATION STUDIES AND GENOMIC SELECTION OF ADVANCED POTATO CLONES FROM THE TEXAS A&M POTATO BREEDING PROGRAM* (<https://oaktrust.library.tamu.edu/bitstream/handle/1969.1/196252/PANDEY-DISSERTATION-2021.pdf?sequence=1&isAllowed=y>).
- Pandey, J., Scheuring, D. C., & Vales, M. I. (2022). Genomic regions associated with tuber traits in tetraploid potatoes and identification of superior clones for breeding purposes. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.952263>

- PATTERSON, H. D., & THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3), 545–554. <https://doi.org/10.1093/biomet/58.3.545>
- Pearson, K. (1901). Principal components analysis. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 6(2), 559.
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing* [Software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rodríguez, L. E. (2010). Origen y evolución de la papa cultivada. Una revisión. *Agronomía Colombiana*, 28(1), 9–17.
- Rosyara, U. R., De Jong, W. S., Douches, D. S., & Endelman, J. B. (2016). Software for Genome-Wide Association Studies in Autopolyploids and Its Application to Potato. *The Plant Genome*, 9(2). <https://doi.org/10.3835/plantgenome2015.08.0073>
- Silva, G. O. da, Azevedo, F. Q., Melo, J. W. de, Pereira, G. E., Patiño-Torres, A. J., Carvalho, A. D. de, Ragassi, C. F., & Pereira, A. da S. (2022). Growth, fresh mass accumulation and distribution in new Brazilian potato cultivars. *Horticultura Brasileira*, 40, 208–213. <https://doi.org/10.1590/s0102-0536-20220210>
- Silva, G. O. da, Azevedo, F. Q., Ragassi, C. F., Carvalho, A. D. F. de, Pereira, G. E., & Pereira, A. da S. (2020). Growth analysis of potato genotypes. *Revista Ceres*, 67, 207–215. <https://doi.org/10.1590/0034-737X202067030006>
- Simko, I., Haynes, K. G., & Jones, R. W. (2006). Assessment of Linkage Disequilibrium in Potato Genome With Single Nucleotide Polymorphism Markers. *Genetics*, 173(4), 2237–2245. <https://doi.org/10.1534/genetics.106.060905>
- Spooner, D. M., Núñez, J., Trujillo, G., del Rosario Herrera, M., Guzmán, F., & Ghislain, M. (2007). Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification. *Proceedings of the National Academy of Sciences*, 104(49), 19398–19403. <https://doi.org/10.1073/pnas.0709796104>
- Stacklies, W., Redestig, H., Scholz, M., Walther, D., & Selbig, J. (2007). pcaMethods a bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, 23(9), 1164–1167. <https://doi.org/10.1093/bioinformatics/btm069>
- Stein, O., & Granot, D. (2019). An Overview of Sucrose Synthases in Plants. *Frontiers in Plant Science*, 10. <https://doi.org/10.3389/fpls.2019.00095>
- Su, W.-H., & Sun, D.-W. (2017). Chemical imaging for measuring the time series variations of tuber dry matter and starch concentration. *Computers and Electronics in Agriculture*, 140, 361–373. <https://doi.org/10.1016/j.compag.2017.06.013>
- Sverrisdóttir, E., Byrne, S., Sundmark, E. H. R., Johnsen, H. Ø., Kirk, H. G., Asp, T., Janss, L., & Nielsen, K. L. (2017). Genomic prediction of starch content and chipping quality in

- tetraploid potato using genotyping-by-sequencing. *Theoretical and Applied Genetics*, 130(10), 2091–2108. <https://doi.org/10.1007/s00122-017-2944-y>
- Tagliotti, M. E., Deperi, S. I., Bedogni, M. C., & Huarte, M. A. (2021). Genome-wide association analysis of agronomical and physiological traits linked to drought tolerance in a diverse potatoes (*Solanum tuberosum*) panel. *Plant Breeding*, 140(4), 654–664. <https://doi.org/10.1111/pbr.12938>
- Tibbs Cortes, L., Zhang, Z., & Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *The Plant Genome*, 14(1), e20077. <https://doi.org/10.1002/tpg2.20077>
- Uitdewilligen, J. G. A. M. L., Wolters, A.-M. A., D'hoop, B. B., Borm, T. J. A., Visser, R. G. F., & van Eck, H. J. (2013). A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. *PLoS ONE*, 8(5), e62355. <https://doi.org/10.1371/journal.pone.0062355>
- VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, 91(11), 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Villanueva, B., Fernández, A., Saura, M., Caballero, A., Fernández, J., Morales-González, E., Toro, M. A., & Pong-Wong, R. (2021). The value of genomic relationship matrices to estimate levels of inbreeding. *Genetics Selection Evolution*, 53(1), 42. <https://doi.org/10.1186/s12711-021-00635-0>
- Voorrips, R. E., Gort, G., & Vosman, B. (2011). Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinformatics*, 12(1), 172. <https://doi.org/10.1186/1471-2105-12-172>
- Vos, P. G., Paulo, M. J., Voorrips, R. E., Visser, R. G. F., van Eck, H. J., & van Eeuwijk, F. A. (2017). Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theoretical and Applied Genetics*, 130(1), 123–135. <https://doi.org/10.1007/s00122-016-2798-8>
- Wang, F., Xia, Z., Zou, M., Zhao, L., Jiang, S., Zhou, Y., Zhang, C., Ma, Y., Bao, Y., Sun, H., Wang, W., & Wang, J. (2022). The autotetraploid potato genome provides insights into highly heterozygous species. *Plant Biotechnology Journal*, 20(10), 1996–2005. <https://doi.org/10.1111/pbi.13883>
- Wang, Guo, T., Wang, Y., Xing, Y., Wang, Y., & He, X. (2020). Exploring the optimization of water and fertilizer management practices for potato production in the sandy loam soils of Northwest China based on PCA. *Agricultural Water Management*, 237, 106180. <https://doi.org/10.1016/j.agwat.2020.106180>
- Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38(2), 203–208. <https://doi.org/10.1038/ng1702>

- Yuan, J., Bizimungu, B., De Koeyer, D., Rosyara, U., Wen, Z., & Lagüe, M. (2020). Genome-Wide Association Study of Resistance to Potato Common Scab. *Potato Research*, 63(2), 253–266. <https://doi.org/10.1007/s11540-019-09437-w>
- Zhang, F., Qu, L., Gu, Y., Xu, Z.-H., & Xue, H.-W. (2022). Resequencing and genome-wide association studies of autotetraploid potato. *Molecular Horticulture*, 2, 6. <https://doi.org/10.1186/s43897-022-00027-y>
- Zhang, Li, M., Zhang, G., Wu, L., Cai, D., & Wu, Z. (2018). Inhibiting Sprouting and Decreasing α -Solanine Amount of Stored Potatoes Using Hydrophobic Nanosilica. *ACS Sustainable Chemistry & Engineering*, 6(8), 10517–10525. <https://doi.org/10.1021/acssuschemeng.8b01860>
- Zia, M. A. B., Demirel, U., Nadeem, M. A., & Çaliskan, M. E. (2020). Genome-wide association study identifies various loci underlying agronomic and morphological traits in diversified potato panel. *Physiology and Molecular Biology of Plants*, 26(5), 1003–1020. <https://doi.org/10.1007/s12298-020-00785-3>