



**UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E
MELHORAMENTO DE PLANTAS**

**Desenvolvimento de marcadores
ambientômicos para o melhoramento de
arroz de terras altas (*Oryza sativa* L.) em
território brasileiro.**

MARCO ANTÔNIO MARCELINO BAHIA

Orientador(a):
Prof. Dr. Rafael Tassinari Resende



UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE AGRONOMIA

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO (TECA) PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TESES

E DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a [Lei 9.610/98](#), o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo das Teses e Dissertações disponibilizado na BDTD/UFG é de responsabilidade exclusiva do autor. Ao encaminhar o produto final, o autor(a) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do material bibliográfico

☒ Dissertação ☐ Tese ☐ Outro*: _____

*No caso de mestrado/doutorado profissional, indique o formato do Trabalho de Conclusão de Curso, permitido no documento de área, correspondente ao programa de pós-graduação, orientado pela legislação vigente da CAPES.

Exemplos: Estudo de caso ou Revisão sistemática ou outros formatos.

2. Nome completo do autor

Marco Antônio Marcelino Bahia

3. Título do trabalho

Desenvolvimento de marcadores ambientômicos para o melhoramento de arroz de terras altas (*Oryza sativa* L.) em território brasileiro.

4. Informações de acesso ao documento (este campo deve ser preenchido pelo orientador)

Concorda com a liberação total do documento ☒ SIM ☐ NÃO¹

[1] Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante:

a) consulta ao(à) autor(a) e ao(à) orientador(a);

b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo da tese ou dissertação.

O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;

- Publicação da dissertação/tese em livro.

Obs. Este termo deverá ser assinado no SEI pelo orientador e pelo autor.



Documento assinado eletronicamente por **Marco Antônio Marcelino Bahia, Discente**, em 25/07/2023, às 16:34, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rafael Tassinari Resende, Professor do Magistério Superior**, em 25/07/2023, às 21:11, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3860389** e o código CRC **866518A0**.

Referência: Processo nº 23070.036911/2023-41

SEI nº 3860389

MARCO ANTÔNIO MARCELINO BAHIA

Desenvolvimento de marcadores ambientais para o melhoramento de arroz de terras altas (*Oryza sativa* L.) em território brasileiro.

Dissertação apresentada ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas, da Universidade Federal de Goiás, como requisito parcial para a obtenção do título de Mestre em Genética e Melhoramento de Plantas.

Área de Concentração: Genética e Melhoramento de Plantas

Orientador:

Prof. Dr. Rafael Tassinari Resende

Coorientadores:

Prof. Dr.^a Sônia Milagres Teixeira

Prof. Dr. Carlos Roberto Sette Júnior

Goiânia GO – Brasil

2023

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Bahia, Marco Antônio Marcelino
Desenvolvimento de marcadores ambientômicos para o
melhoramento de arroz de terras altas (*Oryza sativa* L.) em território
brasileiro. [manuscrito] / Marco Antônio Marcelino Bahia. - 2023.
39 f.: il.

Orientador: Prof. Dr. Rafael Tassinari Resende; co-orientadora
Dra. Sônia Milagres Teixeira; co-orientador Dr. Carlos Roberto Sette
Júnior.

Dissertação (Mestrado) - Universidade Federal de Goiás, Escola
de Agronomia (EA), Programa de Pós-graduação em Genética e
Melhoramento de Plantas, Goiânia, 2023.

Bibliografia. Apêndice.

Inclui mapas, gráfico, tabelas.

1. Ambientômica. 2. Melhoramento Genético. 3. Bioinformática. 4.
Rizicultura. 5. Monte Carlo. I. Resende, Rafael Tassinari, orient. II.
Título.



UNIVERSIDADE FEDERAL DE GOIÁS

ESCOLA DE AGRONOMIA

ATA DE DEFESA DE DISSERTAÇÃO

Ata nº 112 da sessão de Defesa de Dissertação de Marco Antônio Marcelino Bahia, que confere o título de Mestre em Genética e Melhoramento de Plantas, na área de concentração em Genética e Melhoramento de Plantas.

Aos trinta dias do mês de junho de dois mil e vinte e três, a partir das quatorze horas, na Sala de Reuniões do PPGGMP, realizou-se a sessão pública de Defesa de Dissertação intitulada: "Desenvolvimento de marcadores ambientômicos para o melhoramento de arroz de terras altas (*Oryza sativa* L.) em território brasileiro". Os trabalhos foram instalados pelo Orientador, Doutor Rafael Tassinari Resende (PPGGMP/UFG), com a participação dos demais membros da Banca Examinadora: Doutora Patrícia Guimarães Santos Melo (PPGGMP/UFG), membro titular interna e Doutora Úrsula Ramos Zaidan (EA/UFG), membro titular externa. Durante a arguição os membros da banca não fizeram sugestão de alteração do título do trabalho. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Dissertação, tendo sido o candidato APROVADO pelos seus membros. Proclamados os resultados pelo Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **Patrícia Guimarães Santos Melo, Professora do Magistério Superior**, em 30/06/2023, às 17:16, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rafael Tassinari Resende, Professor do Magistério Superior**, em 03/07/2023, às 16:45, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ursula Ramos Zaidan, Professor do Magistério Superior-Substituto**, em 03/07/2023, às 19:27, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3860388** e o código CRC **1A87078E**.

AGRADECIMENTOS

Primeiramente agradeço do fundo do meu coração àquela que divide comigo o sal e em fortuitas ocasiões divide também o açúcar. Obrigado Deborah Mendes Vieira.

Agradeço ao meu orientador Rafael Tassinari Resende pelos ensinamentos, paciência e amizade. Um obrigado sincero pelos cafezinhos também.

Agradeço aos meus pais Sandra e Betinho, minhas irmãs Lorena e Luciana, meus sogros Deuselina e Ivan, e meus cunhados Ricardo e Jean. Ter uma família que te apoia e te incentiva é essencial no processo de formação continuada.

Na esperança de que um dia os mais jovens se interessem por ciência, cito aqui meus sobrinhos Maria Olivia e Davi, que são uma fonte imensa de alegria e sorrisos para o titio.

Agradeço aos meus coorientadores Sônia Milagres Teixeira e Carlos Roberto Sette Júnior, pelo voto de confiança e acompanhamento durante a jornada.

Agradeço aos meus colegas do Laboratório de Melhoramento de Precisão: Augusto, Adriano, Bruno, César, Lucas e Marília. Espero poder ajudá-los cada dia mais.

Agradeço também aos amigos que se mantiveram desde a graduação e continuam sempre a me incentivar e apoiar: Matheus, Thiago, Christian, Lucas, João Victor, Arthur, Quezia, Karol, Hallefy, Daniel, Raquel, Francieudes, Jorge e Hery.

Agradeço a todos professores do PPGGMP pelos ensinamentos e cuidados. Agradeço aos técnicos e demais servidores do setor de melhoramento por facilitarem em muito nosso dia-a-dia. Agradeço aos colegas discentes do PPGGMP pelo companheirismo e ajuda em momentos de necessidade.

Agradeço aos meus meninos Guilherme Augusto e Maria Clara pela ajuda e companhia. Agradeço também ao Bruno, Victor Hugo, Thiago, Pedro e Elisa pela amizade.

Agradeço à Embrapa Arroz e Feijão e em especial ao Dr. Flávio Breseghello pela parceria nesse trabalho e por ceder o banco de dados para nossas análises.

Agradeço ao professor Dr. Gustavo Eduardo Marcatti por compartilhar de forma gratuita na internet material de qualidade sobre programação e geoprocessamento e por vez ou outra me ajudar com problemas e soluções para meus códigos.

Agradeço à CAPES e a FAPEG pelo financiamento do projeto.

SUMÁRIO

RESUMO.....	5
ABSTRACT	6
1 INTRODUÇÃO	7
2 REFERENCIAL TEÓRICO	9
2.1 Ambientômica: Histórico e visão geral	9
2.2 Obtenção de dados e tecnologias de manuseio.....	10
2.3 Aplicações da ambientômica no melhoramento de plantas.....	11
3 MATERIAL E MÉTODOS	13
3.1 Obtenção dos dados fenotípicos do arroz	13
3.2 Extração e manipulação de dados.....	16
3.3 Estudos de associação ambientômica ampla (EWAS)	18
3.4 Repositórios de dados climáticos	18
4 RESULTADOS	21
4.1 Mapeamento Brasileiro para melhoramento de Arroz de Terras Altas	21
4.2 Importâncias das covariáveis ambientais na construção dos índices	22
5 DISCUSSÃO	26
5.1 Explorando a variabilidade genética do arroz de terras altas no território brasileiro	26
5.2 Derivação de multimarcadores ambientômicos e a importância das covariáveis ambientais.....	27
5.3 Leitura dos marcadores ambientômicos e sua influência na produtividade do arroz de terras altas	28
6 CONCLUSÃO.....	32
7 REFERÊNCIAS	33
8 APÊNDICE	38

RESUMO

BAHIA, M.A.M. **Desenvolvimento de marcadores ambientômicos para o melhoramento de arroz de terras altas (*Oryza sativa* L.) em território brasileiro.** Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 2023^{1, 2}.

O arroz (*Oryza sativa* L.) é um dos alimentos base da dieta brasileira, sendo, portanto, estrategicamente imprescindível seu cultivo e a independência produtiva visando a garantia alimentar da população. Dentro da rizicultura, a escolha do genótipo adequado para o plantio é o fator que impacta com maior força o resultado do empreendimento. Buscando embasar essa tomada de decisão, a ambientômica vem sendo aplicada com o objetivo de selecionar genótipos com maior potencial produtivo para certas áreas de interesse. O objetivo desse trabalho foi gerar e analisar a participação dos marcadores ambientômicos para os dados de produção total de arroz de terras altas em território brasileiro. Os dados experimentais foram fornecidos pela Embrapa Arroz e Feijão, sendo que, foram avaliados 2119 genótipos de arroz em 187 municípios ou localidades do país, no período compreendido entre 1982 a 2018. Para a geração dos marcadores ambientômicos foram utilizados dados das plataformas *SoilGrids*, *WorldClim* e NASA POWER; ao todo, foram coletadas 393 covariáveis ambientais. A geração dos marcadores ambientômicos foi realizada utilizando o método de Monte Carlo, para 10.000 iterações e considerando sempre a presença dos 187 municípios onde estão alocados os experimentos da EMBRAPA. Utilizando o pacote Random Forest e os métodos *IncMSE* e *IncNodePurity* foram avaliados o grau de importância de cada covariável para o modelo aplicado em toda a extensão do território brasileiro, chegando ao resultado que aponta que o coeficiente de variação para a precipitação sazonal foi a covariável mais importante para ambos os modelos.

Palavras-chaves: Ambientômica; melhoramento genético; bioinformática; rizicultura; Monte Carlo.

¹ Orientador Prof. Dr. Rafael Tassinari Resende

² Coorientadores: Profs. Dra. Sônia Milagres Teixeira, Dr. Carlos Roberto Sette Júnior

ABSTRACT

BAHIA, M.A.M. **Development of enviromic markers for upland rice (*Oryza sativa* L.) breeding in brazilian territory.** Dissertation (Master in Genetics and Plant Breeding) – School of Agronomy, Federal University of Goiás, Goiânia, 2023^{1, 2}.

Rice (*Oryza sativa* L.) is one of the staple foods in the Brazilian diet, and therefore, its cultivation and productive independence are strategically essential for ensuring the population's food security. Within rice farming, selecting the appropriate genotype for planting is the factor that most strongly impacts the outcome of the endeavor. In order to support this decision-making process, enviromics has been applied with the objective of selecting genotypes with higher productive potential for specific areas of interest. The aim of this study was to generate and analyze the contribution of enviromic markers to the total upland rice production data in Brazilian territory. The experimental data were provided by Embrapa Rice and Beans and involved the evaluation of 2,119 rice genotypes in 187 municipalities or localities across the country, spanning the period from 1982 to 2018. For the generation of enviromic markers, data from the SoilGrids, WorldClim, and NASA POWER platforms were used, resulting in a total of 393 environmental covariates collected. The generation of enviromic markers was performed using the Monte Carlo method, with 10,000 iterations and always considering the presence of the 187 municipalities where the Embrapa experiments were conducted. The Random Forest package and the *IncMSE* and *IncNodePurity* methods were used to evaluate the importance of each covariate for the model applied throughout the Brazilian territory. The results showed that the coefficient of variation for seasonal precipitation was the most important covariate for both models.

Keywords: Enviromics; plant breeding; bioinformatics; rice cultivation; Monte Carlo.

¹ Advisor Prof. Dr. Rafael Tassinari Resende

² Co-advisors: Profs. Dra. Sônia Milagres Teixeira, Dr. Carlos Roberto Sette Júnior

1. INTRODUÇÃO

A produção de arroz no Brasil desempenha um papel significativo no mercado global, já que o país se destaca como um dos principais *players* mundiais desse cereal. No Brasil, a produção de arroz é realizada em dois modelos distintos: o cultivo irrigado e o de terras altas, sendo o sistema irrigado o mais predominante (Reed et al., 2006; de Carvalho et al., 2020). O estudo da interação entre genótipos de arroz e o ambiente tem se mostrado inovador no campo do melhoramento vegetal, incorporando os conceitos da ambientômica (Lima et al., 2017). A análise de multivariáveis ambientais permite mensurar o impacto dessa interação, permitindo entender como os genótipos do arroz se relaciona com o ambiente, influenciando a performance fenotípica. Essas abordagens podem ser aplicadas para aumentar o ganho de seleção e coletar informações sobre o crescimento, desempenho e composição genotípica do arroz, tanto em situações controladas como em campo (Lima et al., 2017).

Modelos de interação genótipos por ambientes (G x A) como AMMI, GGE biplot e análise fatorial têm sido utilizados em estudos da integração genótipos por ambientes do arroz (Haider et al., 2017; Devi et al., 2020; Hashim et al., 2021). No entanto, o avanço tecnológico do geoprocessamento tem se mostrado uma ferramenta promissora no melhoramento genético dessa cultura (Costa-Neto et al., 2020). Por meio da coleta e análise de informações ambientais e espaciais, como dados de satélite, informações climáticas e características do solo, é possível mapear as condições favoráveis e desfavoráveis para o crescimento e desenvolvimento do arroz (Tang et al., 2009; Hashim et al., 2021). Esses dados serão essenciais para a tomada de decisões sobre a escolha de genótipos e ambientes adequados para o plantio, auxiliando na seleção de variedades adaptadas a cada região específica. O uso de técnicas avançadas de geoprocessamento oferece uma abordagem mais precisa e eficiente para o melhoramento genético do arroz, permitindo um aumento na eficácia do processo seletivo e a maximização da produtividade e resistência das variedades desenvolvidas.

O uso das técnicas de ambientômica na agricultura e no melhoramento de plantas é recente, trazida das ciências médicas para os estudos de interação Genótipos por Ambientes de plantas e animais em 2019 (*preprint* de Resende et al., 2019, publicado em (Resende et al., 2020)). O uso da ambientômica, visa coletar a informação não somente do indivíduo ou do material genético, mas mapear de forma clara as informações acerca do ambiente e correlacioná-las com as informações de genótipo e/ou fenótipo gerando múltiplos marcadores ambientômicos (McOmish et al., 2014; Resende et al., 2022). Para a coleta das informações

ambientais com o objetivo de embasar a modelagem da região, pode-se usar as mais diversas estratégias e tecnologias disponíveis, tais como imagens de satélites, informações de estações meteorológicas, informações climáticas de temperatura, precipitação, direção e força do vento, umidade do ar, informações para a caracterização físico-química do solo, entre muitas outras (Xu, 2016; Marcatti et al., 2017; Resende et al., 2020).

Os trabalhos recentes de zoneamento indicaram o potencial de produção em regiões não tradicionais, porém aptas do ponto de vista técnico (Steinmetz & Silva, 2017; Januário et al., 2018; Heinemann et al., 2021). A expansão da fronteira da rizicultura no Brasil é de interesse amplo da comunidade do agronegócio tendo em vista a alta demanda no mercado interno e a possibilidade de exportação do grão (Silva & Assad, 2001), bem como recomendações assertivas de materiais genéticos específicos para cada região. O melhoramento genético do arroz de terras altas no Brasil teve início na década de 1970, com o Programa de Melhoramento do Arroz (PMA), coordenado pela Embrapa e outras instituições de pesquisa agrícola, e ainda há espaço para uma expansão ainda maior frente à produção de arroz irrigado.

No contexto do desenvolvimento de novas variedades de arroz que se adaptem às condições climáticas e de solo do país, o qual apresenta uma grande diversidade geográfica e climática (Breseghello et al., 2021), busca-se melhorar a concordância entre genótipos e ambiente. As perspectivas para o avanço do melhoramento do arroz no Brasil são bastante positivas. A pesquisa em melhoramento genético do arroz tem sido uma área importante de investimento em muitas instituições de pesquisa, como a Embrapa, universidades e outras instituições públicas e privadas.

Para o melhoramento genético do arroz no Brasil, são esperados avanços significativos, tais como o desenvolvimento de variedades que demonstrem maior resistência a doenças, pragas e adversidades climáticas, o incremento na produtividade e o aprimoramento da qualidade dos grãos (Coelho et al., 2017; Heinemann et al., 2019; Breseghello et al., 2021). Como externalidade positiva, o avanço do melhoramento genético para o arroz no Brasil poderá também acarretar em desenvolvimento e treinamento de profissionais no uso de técnicas avançadas de melhoramento genético de precisão no âmbito do geoprocessamento.

O melhoramento genético do arroz tem um grande potencial para contribuir para a agricultura brasileira e a segurança alimentar da população. Esse trabalho tem o objetivo de apresentar o processo de criação e avaliação de marcadores ambientômicos para arroz de terras altas em toda a extensão do território brasileiro, visando embasar a tomada de decisão sobre escolhas de genótipos e ambientes adequados para o plantio.

2. REFERENCIAL TEÓRICO

2.1. Ambientômica: Histórico e Visão Geral

Os conceitos relacionados à *enviromics* têm sido discutidos no campo científico há algum tempo, especialmente nas ciências médicas (Cuartas Arias, 2011; McOmish et al., 2014). No campo do melhoramento vegetal, o estudo das variáveis ambientais locais é considerado inovador, permitindo a mensuração do efeito da interação entre o genótipo e o ambiente (Resende et al., 2020).

Em comparação às outras ‘-ômicas’, a genômica se concentra no estudo da organização da informação genética dentro do genoma e como essa organização influencia a expressão fenotípica. Por outro lado, a fenômica coleta informações sobre o fenótipo, visando melhorar a seleção ao avaliar o crescimento, desempenho e composição do genótipo, tanto em ambientes controlados quanto em campo. A ambientômica, por sua vez, busca coletar informações não apenas sobre o indivíduo ou organismo, mas também mapear de forma clara as informações ambientais e correlacioná-las com informações genotípicas e/ou fenotípicas. Alguns autores tendem a dissociar a ambientômica da genética, considerando-a apenas como o estudo geral do ambiente (Crossa et al., 2021).

Para coletar informações ambientais a fim de embasar a modelagem de uma região, diversas estratégias e tecnologias podem ser utilizadas, incluindo imagens de satélite, dados de estações meteorológicas, informações climáticas como temperatura, precipitação, direção e intensidade do vento, umidade do ar, bem como informações físico-químicas do solo, entre outros (Marcatti et al., 2017; Resende et al., 2022). Ao trabalhar com a genômica, é usual se trabalhar com informações de conjuntos de genes, uma vez que a identificação de um gene específico altamente influente na interação com o fenótipo é rara na genética quantitativa. Da mesma forma, na ambientômica é comum lidar com um grande volume de informações, pois não se pode atribuir a um único fator ambiental a responsabilidade por uma expressão fenotípica específica (Giani et al., 2020; Meuwissen et al., 2020).

Com um grande volume de informações, no entanto, pode-se sobrepor diferentes fatores ambientais e compreender como esse conjunto de informações está relacionado às características fenotípicas de interesse, por meio de modelagem do ambiente. É importante ressaltar que, quanto maior o volume de informações disponíveis, maior é a probabilidade de o modelo explicar adequadamente as características fenotípicas.

Em termos estruturais, as informações do ambientoma (*envirome*) são baseadas na coleta e organização de pixels, que são as unidades básicas dos mapas de dados. Cada pixel contém uma informação individual, que pode ser discreta ou contínua, dependendo da variável em questão. Cada núcleo de informação é chamado de ambiótipo (*envirotipo*). A sobreposição de várias camadas de pixels ou camadas nos fornece um modelo composto por informações de diversas fontes, como dados experimentais, informações de fazendas ou dados de parcelas de uma empresa multinacional. Ao sobrepor essas camadas de informações ambientais, constrói-se um ambientoma, por analogia ao genoma (Resende et al., 2020).

Assim como o genoma engloba todos os marcadores mensuráveis de DNA que um indivíduo ou população pode ter, o ambientoma engloba todas as informações ambientais que podem ser coletadas em uma determinada área de estudo. É importante lembrar que, por se tratar de variáveis ambientais, é necessário respeitar os limites do que é economicamente viável, uma vez que dificilmente serão alocados recursos para coletar dados que não sejam essenciais para o método do estudo.

Quanto à análise dos dados no contexto da ambientômica, não existe um método único, sendo que qualquer abordagem e modelagem confiáveis podem ser utilizadas. De acordo com Van Eeuwijk et al. (2016), os modelos mais comumente utilizados para o estudo da interação genótipo x ambiente são os modelos lineares mistos e os modelos bayesianos. Estudos recentes em universidades brasileiras também têm utilizado modelos de crescimento de culturas combinados com estatística bayesiana, indicando a ausência de uma técnica predominante.

2.2. Obtenção de dados e tecnologias de manuseio

No campo da ambientômica, a coleta de dados precisa e abrangente é uma etapa crucial. A utilização de sensores e drones remotos é essencial nesse contexto, e possibilita uma coleta eficiente e econômica dos dados, afim de analisar as respostas das plantas ao ambiente circundante. O sensoriamento remoto é uma tecnologia poderosa para coletar dados em larga escala. Ao se utilizar drones e satélites com sensores especializados para coletar imagens de parâmetros ambientais, como temperatura, umidade do solo, índices de vegetação e outros, os pesquisadores podem estudar o comportamento dos caracteres climáticos durante um certo período de tempo e a resposta fenotípica das culturas aos diferentes ambientes (Mosleh et al., 2015; Carrasco et al., 2022; Wei et al., 2023).

O uso de drones e sensores tem mudado a forma como se coleta informações sobre as plantas. Os agricultores podem monitorar cada planta em tempo real com sensores portáteis de alta precisão. Compreender a abrangência das relações entre as plantas e o ambiente, permite a implementação de métodos de manejo mais precisos e eficazes com o objetivo de maximizar o potencial produtivo das plantas. A análise avançada de dados coletados é agora possível graças à facilidade de armazenamento e manipulação de dados, somadas à integração de tecnologias de internet das coisas e inteligência artificial. Essa integração de tecnologias e a presença de profissionais bem capacitados vem auxiliando na tomada de decisões inteligentes sobre proceder no manejo e melhoramento de plantas (Dorok et al., 2015; Morales et al., 2020; Zhao et al., 2021; Muenchen, 2023).

Essas tecnologias de manuseio sofisticadas não apenas fornecem informações detalhadas sobre as condições ambientais, mas também ajudam a criar cultivares mais fortes e produtivas. Ao obter uma melhor compreensão de como as plantas reagem aos estressores e desafios ambientais, os agricultores e pesquisadores podem escolher características que serão adequadas para a criação de variedades adaptadas a diferentes ambientes (Wang et al., 2021). Com a disponibilidade dessas ferramentas, a ambientômica auxilia os pesquisadores a avançar no campo do melhoramento de plantas, promovendo uma agricultura mais eficiente, sustentável e adaptada aos desafios do futuro.

2.3. Aplicações da ambientômica no melhoramento de plantas

Nos últimos anos, a ambientômica, tem sido apresentada como uma nova técnica para o cultivo e melhoramento de plantas. Essa abordagem interdisciplinar combina dados de várias fontes biológicas e ambientais para entender e otimizar as reações das plantas às condições ambientais variáveis (Resende et al., 2020; Costa-Neto et al., 2021; Costa-Neto & Fritsche-Neto, 2021). A ambientômica integra informações genéticas, de fenótipo e ambientais para fornecer uma visão abrangente e detalhada dos processos que ocorrem nas plantas em diferentes sítios. Isso oferece oportunidades inéditas para o desenvolvimento de variedades de plantas mais resistentes e produtivas (Resende et al., 2018, 2022).

A ambientômica também é importante para otimizar as estratégias de cultivo, especialmente quando se trata de mudanças climáticas. O desempenho das plantas em diferentes condições climáticas pode ser previsto usando dados ambientais e genômicos, o que permite aos agricultores ajustar suas técnicas agrícolas para minimizar os efeitos

negativos nas colheitas. Isso resulta em uma produção agrícola mais sustentável, o que garante a segurança alimentar em um mundo onde as condições estão se tornando cada vez mais instáveis (Morais et al., 2008; Brar et al., 2017; Heinemann et al., 2019).

A ambientômica também pode ser usada para acelerar o processo de melhoramento de plantas por meio da seleção assistida por marcadores. Os melhoristas podem selecionar genótipos promissores com antecedência identificando marcadores genéticos associados a características desejáveis. Isso reduz significativamente o tempo necessário para desenvolver novas variedades. A produtividade e a diversidade das culturas agrícolas aumentam positivamente com essa abordagem de melhoramento acelerado (Glaubitz et al., 2014; Wang et al., 2020; Reyes et al., 2022).

A ambientômica abre caminhos de melhorar as plantas e adaptar a agricultura às necessidades de um mundo que está mudando rapidamente, auxiliando a se obter uma agricultura mais resiliente, sustentável e produtiva e na garantia da segurança alimentar global no futuro através da compreensão aprofundada das respostas das plantas aos estresses ambientais e da seleção de genótipos que terão uma resposta excelente ao ambiente de interesse.

3. MATERIAL E MÉTODOS

3.1. Obtenção dos dados fenotípicos de arroz

Os dados utilizados para a caracterização fenotípica da cultura de arroz de terras altas no Brasil foram cedidos pela Embrapa Arroz e Feijão, Breseghello et al. (2021) apresentam com maior detalhamento os procedimentos de instalação dos experimentos e a obtenção dos caracteres fenotípicos. Para este estudo, foram utilizadas informações relacionadas ao genótipo, dados de localização e data de plantio de cada experimento e dados brutos relativos à produtividade (Tabela 1).

Embora a análise da Embrapa também abranja os dados de arroz irrigado, o estudo se ateve apenas aos dados de arroz de terras altas. Os primeiros resultados datam do ano 1982 e os dados apresentados chegam ao ano de 2018, totalizando 37 anos de análise e acompanhamento do projeto. Os dados apresentam experimentos em 187 municípios ou localidades distribuídos em 21 estados e Distrito Federal, contemplando todas as 5 regiões brasileiras.

Ao todo, foram considerados 2119 genótipos de arroz de terras altas e 1593 experimentos. A distribuição dos municípios e genótipos analisados por região pode ser observada na Figura 1. Entre os genótipos destaca-se a predominância do BRS Primavera, utilizado em 803 experimentos e em 143 localidades (representando 50,4% e 76,4% respectivamente). Pode-se observar também a presença expressiva tanto do BRS Bonança, utilizado em 508 experimentos e encontrado em 117 localidades (31,8% e 62,5%) e do Caiapó, registrado em 462 ocasiões e 106 locais (29% e 56%).

Tanto os números de municípios, genótipos e de experimentos se concentram na Região Centro-Oeste, devido principalmente à presença da Embrapa Arroz e Feijão no município de Santo Antônio de Goiás, Goiás. Existe uma alta taxa de correspondência, acima de 91%, entre os materiais genéticos presentes em todas as outras regiões com os materiais testados no Centro-Oeste, e é interessante observar também, que há uma taxa de correspondência de apenas 54% entre os genótipos testados no Sudeste com os testados no Norte. A região Sul apresenta os menores números de cidades e experimentos o que se traduz em uma alta taxa de correspondência com as outras regiões (Tabela 1).

Os dados fornecidos pela Embrapa não contam com as coordenadas geográficas exatas dos locais do experimento. A título de aproximação, foram utilizadas as coordenadas geográficas do centroide do município, dado obtido nos repositórios gratuitos do IBGE (2010).

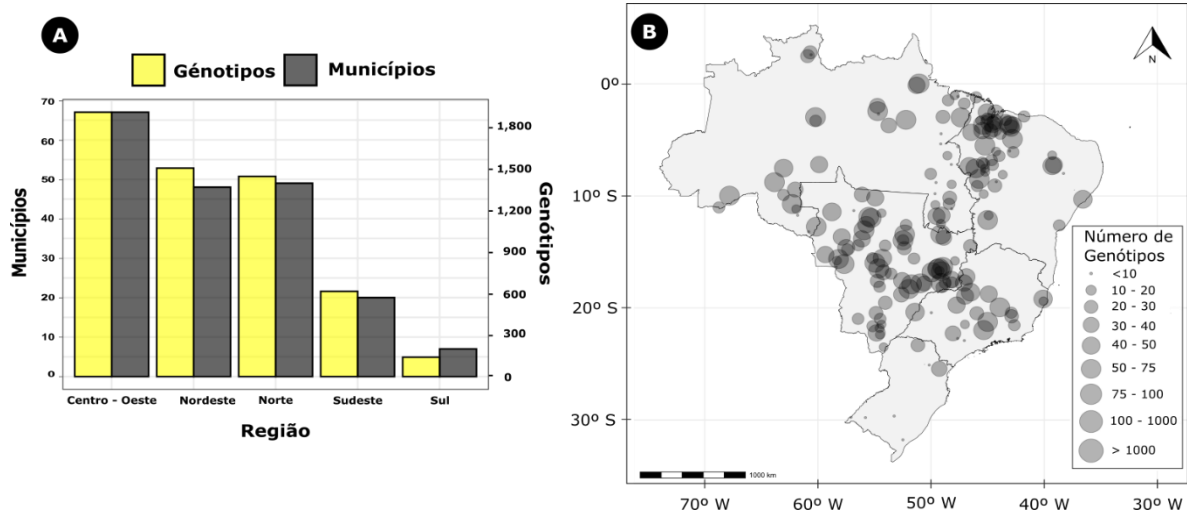


Figura 1. (A) Quantidade de genótipos e municípios analisados por região. Colunas amarelas se referem à quantidade de genótipos (eixo à direita), colunas pretas são referentes ao número de municípios onde há experimentos da Embrapa. (B) Distribuição e densidade de genótipos nos sítios de experimentação considerados nesse projeto.

Tabela 1. Número de genótipos, municípios e experimentos; produtividade e coincidência entre genótipos por região.

Região	Número de Genótipos	Número de Municípios	Número de Experimentos	Produtividade (Kg. ha ⁻¹)	Número de Genótipos Coincidentes				
					Centro - Oeste	Nordeste	Norte	Sudeste	Sul
Centro – Oeste	1909	65	848	3036,23 ± 1495,22	1909 (1,00)	1374 (0,72)	1333 (0,70)	592 (0,31)	134 (0,07)
Nordeste	1505	48	228	3217,85 ± 1384,23	1374 (0,91)	1505 (1,00)	1162 (0,77)	454 (0,3)	130 (0,09)
Norte	1447	47	403	3198,59 ± 1333,91	1333 (0,92)	1162 (0,80)	1447 (1,00)	334 (0,23)	127 (0,09)
Sudeste	616	20	92	2946,02 ± 1347,51	592 (0,96)	454 (0,74)	334 (0,54)	616 (1,00)	130 (0,21)
Sul	137	7	17	3921,37 ± 1462,09	134 (0,98)	130 (0,95)	127 (0,93)	130 (0,95)	137 (1,00)

3.2. Extração e manipulação de dados

Uma vez com os dados de produtividade e informações ambientais necessárias, foi preciso agrupá-los em todo o território brasileiro para montar o ambientoma. Optou-se, portanto, por uma resolução de estudo de 10 km por aresta de pixel, equilibrando a necessidade de detalhes dos dados com a capacidade presente de processamento computacional. Utilizando a ferramenta "Pontos regulares" no software QGIS, foi criada uma malha de pontos espaçados regularmente a cada 10 km, formando um gradiente de pixels de 100 km² em todo o território brasileiro.

Concatenando a base de dados da “Embrapa Arroz e Feijão” com as coordenadas obtidas através dos dados do IBGE, usando o nome do município como dado de amarração, criou-se uma base de dados indicando as 187 localidades onde houve coleta de dados fenotípicos para o projeto (Figura 1B). Utilizando o software R foi realizada a extração dos dados ambientais para os dois arquivos vetoriais criados. Primeiramente foi feita para a base de 187 pontos e em sequência para nossa malha de 87155 pontos.

O primeiro arquivo produzido será o arquivo de treinamento contendo informações referentes ao genótipo, fenótipo e ambiótipo. O segundo arquivo contém apenas informações ambientais. Aplicando modelos matemáticos (Equações 1, 2) foi possível prever o potencial existente para cada localidade (Figura 3A), baseado na produtividade média para localidades semelhantes, segundo o apresentado por Resende et al. (2020).

Foram adotados dois modelos para estimar e prever a produtividade local por pixel, devido ao delineamento dos experimentos conduzidos pela Embrapa. Enquanto em algumas localidades havia a presença de diversos ensaios (*trials*), em outros havia apenas um, inviabilizando o uso dos BLUPs. Para corrigir isso, foi o modelo BLUE, tratando esse próprio ensaio como o intercepto do modelo.

$$y = Xb + Wr + e, \quad [1]$$

$$y = Wb + Zg + Wr + e, \quad [2]$$

Em que, para ambos modelos: y é variável observada; X é uma matriz de incidência sobre os efeitos fixos; Z e W são matrizes de incidência sobre os efeitos aleatórios; ‘ b ’ são os efeitos fixos (sendo localidades e genótipos no modelo da Eq. [1], e apenas localidades na Eq.[2]); ‘ g ’ é o vetor de efeitos genéticos, assumido como de natureza

aleatória e estrutura de variância e covariância $g|A, \sigma_g^2 \sim N(0, A\sigma_g^2)$, sendo A uma matriz de parentesco para genótipos não aparentados, ou simplesmente uma matriz identidade de ordem 2119 (genótipos); ‘ r ’ são os efeitos aleatórios relativos à variação de tipo de delineamento, presença de blocos e/ou repetições (delineamento látice e ano de implantação), com estrutura de variância e covariância $r|\sigma_r^2 \sim N(0, I\sigma_r^2)$; e ‘ ε ’ é o vetor de erros aleatórios, com distribuição de variância e covariância $e|\sigma_e^2 \sim N(0, I\sigma_e^2)$.

Como um parâmetro indicativo de ganhos de seleção local, foi calculada a herdabilidade no sentido amplo, ou um coeficiente de variação genética, para cada um dos municípios avaliados, conforme o modelo da Equação 3, utilizando as variâncias genéticas, variâncias de efeitos aleatórios relativos ao delineamento e ao ano de implantação e variâncias dos erros aleatórios (Tabela 2). A partir dos dados da herdabilidade calculada e utilizando uma rede neural através do pacote [nnet](#), os resultados da herdabilidade foram extrapolados para todos os pixels do território brasileiro (Ripley & Venables, 2002). Em seu processo de treinamento foram feitas 20.000 interações para a definição dos parâmetros da rede e posterior validação com nossos pontos de controle, sendo esses, os pixels onde estão instalados os experimentos.

$$h_{CALC}^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_r^2 + \sigma_e^2} \quad [3]$$

Em que, σ_g^2 é variância genética, σ_r^2 é variância dos efeitos aleatórios relativos ao delineamento e ao ano de implantação; e σ_e^2 é variância dos erros aleatórios.

O modelo utilizado para a projeção dos dados de h_{CALC}^2 foi uma rede neural testada para o conjunto de dados $\{h_{CALC}^2, h_{PROJ}^2\}$ nos 187 pontos de experimentação utilizando dados normalizados de longitude, latitude e altitude. Os parâmetros da rede foram definidos pela maior correlação entre h_{CALC}^2 x h_{PROJ}^2 e são compostos pelos argumentos:

- **Size:** Número de nós em cada camada da rede neural. Para este estudo foi utilizada uma única camada com 6 nós;
- **Rang:** Especifica os valores de limite inferior e superior dos pesos da rede neural. Os limites para essa rede foram de -0.1 e 0.1, respectivamente.
- **Decay:** Valor que especifica a taxa de decaimento para a regularização da rede. Utilizou-se em nosso estudo um *decay* de 5e-4; e
- **Maxit:** Máximo de iterações permitidas no ajuste da rede. Para este estudo foram consideradas 1000 como número máximo de iterações.

3.3. Estudos de associação ambientômica ampla (EWAS)

A modelagem para esse trabalho envolve a criação de índices ambientais que serão usados para compor uma matriz de índices a partir da técnica de *Monte Carlo*, tendo como base as 393 covariáveis ambientais extraídas dos 3 repositórios de dados que serão descritos na próxima seção desse trabalho. Essa matriz de marcadores é composta por 87155 pixels (linhas) por 10.000 simulações de ambientômas (colunas).

Para a avaliação do estudo de associação ambientômica ampla (EWAS – *Enviromics-Wide Association Study*) deste trabalho, foi preciso definir o grau de importância de cada covariável dentro da construção de cada um dos 10.000 ambientômas (Costa - Neto et al., 2023). Para tal, foram utilizadas duas medidas de avaliação através do pacote [Random Forest](#) (Breiman, 2001), sendo elas: *IncMSE* e *IncNodePurity*. Enquanto *IncMSE* é o aumento médio na soma do erro quadrático do modelo, *IncNodePurity* é a medida do aumento na homogeneidade nos conjuntos dos dados, visando maximizar a pureza dos subconjuntos (Soil Survey Staff, 2023). Para uma classificação final a covariável mais relevante para o índice ambiental, recebeu dez pontos e cada covariável abaixo dela recebeu uma redução de 50% em sua pontuação, para cada ambientôma, sendo a pontuação final a soma da pontuação obtida em cada uma das 10.000 interações.

3.4. Repositórios de dados climáticos

Buscando realizar a caracterização do ambiente foram consultados três repositórios de dados geoclimáticos: *WorldClim*, Nasa POWER e *SoilGrids*. Todos repositórios oferecem gratuitamente seus serviços e são complementares quanto ao tipo e periodicidade dos dados disponibilizados.

O *WorldClim* é um repositório de dados que fornece informações acerca de variáveis bioclimáticas globais para uso em pesquisas voltadas para as mais diversas áreas. Os dados disponíveis nesse repositório abrangem variáveis como índices médios de precipitação, temperatura, oscilação de temperatura, umidade, radiação solar, velocidade do vento etc. Para a realização desse trabalho, foi utilizada a versão 2.1 da plataforma (Cerasoli et al., 2022)

Coletou-se, também, dados provenientes do repositório Nasa POWER (Prediction Of Worldwide Energy Resource), que é uma plataforma gratuita mantida e abastecida pela agência estadunidense. Similarmente ao *WorldClim* ela apresenta dados

climáticos históricos de temperatura umidade, vento, médias de precipitação e radiação solar. (Monteiro et al., 2018; Quansah et al., 2022).

Visando considerar as diversas diferenças quanto à composição do solo dentro do território brasileiro, recorreu-se ao repositório *SoilGrids 2.0* que oferece informações valiosas nesse quesito, incluindo dados sobre composição, textura, pH, capacidade de troca catiônica, ponto de murcha e capacidade de campo. Como diferencial, o *SoilGrids 2.0* detalha as propriedades químicas do solo em diferentes profundidades e em quatro marcos da distribuição padrão dos valores; permitindo, assim, uma visão mais completa que embasa pesquisas de diferentes culturas agronômicas e florestais (Poggio et al., 2021; Dandabathula et al., 2022).

Tabela 2. Informações adicionais sobre as 10 localidades com maior herdabilidade calculada.

Localidade	Estado	Região	σ^2g	σ^2r	σ^2e	h^2c	h^2p	BLUEm	GYm
Londrina	PR	Sul	1334795,80	0,00	312992,70	0,810	0,684	4653,5	4177,1
Ponta Grossa	PR	Sul	373920,88	152,97	94773,48	0,798	0,522	2197,3	1538,0
Jacundá	PA	Norte	383106,02	20684,58	117070,80	0,736	0,321	4098,5	3840,2
Diamantino	MT	Centro - Oeste	94234,63	3281,23	41553,64	0,678	0,352	1560,7	1111,1
Fátima do Sul	MS	Centro - Oeste	102447,06	11422,39	45509,10	0,643	0,325	2989,8	2724,2
Rio Quente	GO	Centro - Oeste	426104,20	3696,54	234871,20	0,641	0,551	2717,3	2822,1
Capão do Leão	RS	Sul	861945,68	105706,60	401930,10	0,629	0,632	4090,6	3809,4
Redenção	PA	Norte	215422,87	0,00	127470,00	0,628	0,326	2118,6	2320,3
Alto Parnaíba	MA	Nordeste	561372,18	2076,25	362303,00	0,606	0,357	3503,0	3295,3
Formoso	GO	Centro - Oeste	654435,84	65494,76	398202,40	0,590	0,459	5459,5	5663,2

Sendo, σ^2g = variância genética; σ^2r = variância dos efeitos aleatórios relativos ao delineamento e ao ano de implantação; σ^2e = variância dos erros aleatórios; h^2c = herdabilidade calculada a partir da Equação 3; h^2p = herdabilidade projetada a partir da rede neural; **BLUEm** = Média da produtividade calculada a partir das Equações 1 e 2 e **GYm** = Média da produtividade obtida no banco de dados da Embrapa.

4. RESULTADOS

4.1. Mapeamento Brasileiro para melhoramento de Arroz de Terras Altas

Na Tabela 2, são apresentados os dados das dez localidades com maior herdabilidade calculada, sendo quatro na região Centro-Oeste, três na região Sul, dois na região Norte e um na região Nordeste. É importante ressaltar que a herdabilidade calculada fornece resultados marginais que podem dar uma falsa sensação de uma herdabilidade acima da média. Por outro lado, a herdabilidade projetada leva em consideração a dependência espacial e a vizinhança, proporcionando uma representação mais precisa dos dados. Não foi encontrada nenhuma forte correlação entre a herdabilidade calculada e os parâmetros: número de genótipos, número de *trials*, número de repetições, número de blocos e número de anos em que se realizou experimento em cada localidade.

Os dados da herdabilidade projetada são mostrados na Figura 2A, onde é possível observar uma faixa de valores elevados que se estende do norte do Rio Grande do Sul até o sudeste de Goiás. Também são visíveis áreas com valores mais baixos na região costeira do Nordeste e no interior dos estados do Maranhão e Piauí. A distribuição desses dados é apresentada na Figura 2B, com uma média de herdabilidade projetada de 0,345 e um valor máximo de 0,745. Observa-se um pico de herdabilidade em torno de 0,25, indicando que a herdabilidade da maioria dos pixels da área se concentra em torno desse valor. O modelo gerado pela aplicação da rede neural projetou a herdabilidade para os pontos de experimento (Figura 2C) com uma correlação aproximada de 60%.

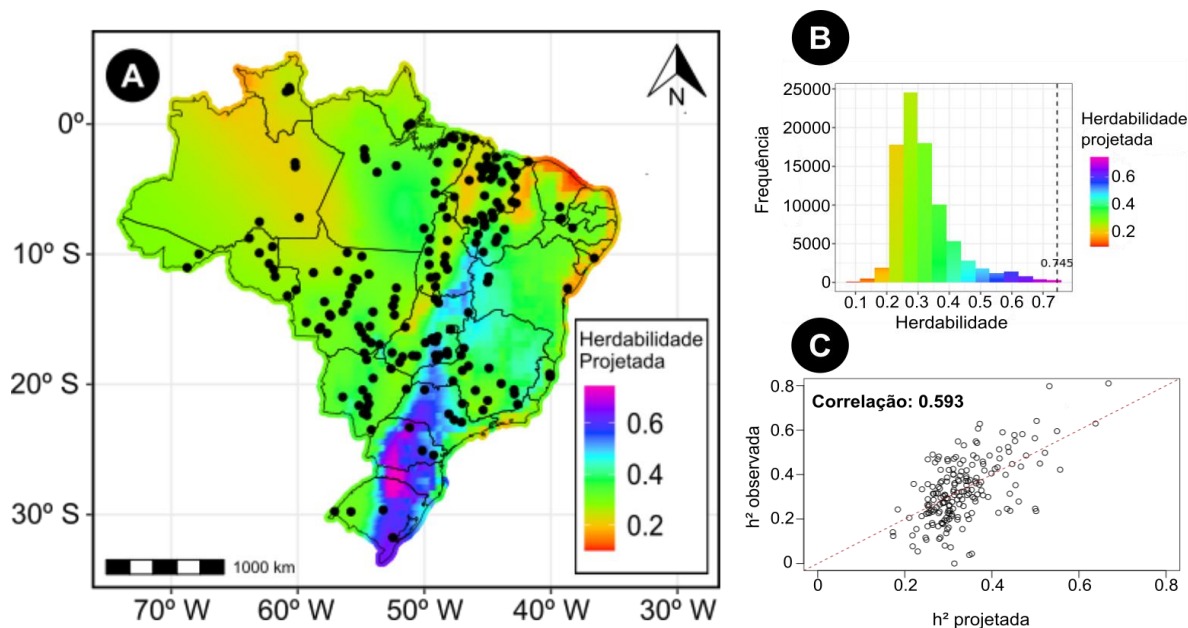


Figura 2. (A) Mapa de herdabilidade projetada para cada pixel do território brasileiro, pontos indicam a localização dos experimentos da Embrapa, (B) histograma apresentando os dados da projeção e (C) correlação entre as herdabilidades observadas e projetadas nas 187 localidades de experimentos.

4.2. Importâncias das covariáveis ambientais na construção dos índices

Pode-se observar as covariáveis mais importantes para o método *IncMSE*, seu repositório de origem, quantas vezes elas figuraram em cada posição do top-5 e sua pontuação na Tabela 3:

Tabela 3. Vinte covariáveis melhor ranqueadas pelo método *IncMSE*.

Covariável ¹	Base ²	TOP 1 ³	TOP 2	TOP 3	TOP 4	TOP 5	Pontuação
bio_15	WC	8435	1245	180	54	27	91116,07
cfvo_30-60cm_Q0p95	SG	1355	5210	1340	573	403	44036,25
ocd_5-15cm_Q0p05	SG	91	971	1779	1388	1027	12979,41
bio_12	WC	21	411	1012	1031	880	7070,09
nitrogen_100-200cm_mean	SG	25	411	912	881	842	6605,41
ocd_15-30cm_Q0p05	SG	6	358	819	881	726	5836,25
cfvo_30-60cm_mean	SG	16	261	871	786	618	5313,73
nitrogen_30-60cm_mean	SG	7	173	449	537	523	3354,39

silt_100-200cm_Q0p05	SG	9	146	390	513	525	3036,72
T2M_MAX_2010a2019OCT	NP	2	119	363	566	587	2945,43
T2MDEW_2010a2019JUN	NP	10	172	362	432	460	2941,09
soc_30-60cm_mean	SG	9	161	251	258	267	2169,39
PRECTOTCORR_10a19APR	NP	5	89	178	226	279	1554,35
cec_100-200cm_mean	SG	2	31	103	150	212	907,62
cfvo_5-15cm_Q0p95	SG	1	19	146	151	155	867,62
cec_100-200cm_Q0p5	SG	0	28	63	141	219	748,65
RH2M_2010a2019OCT	NP	0	28	91	110	148	694,76
bio_7	WC	0	15	49	112	131	540,35
bio_18	WC	0	11	34	80	123	426,95
PRECTOTCORR_10a19MAR	NP	0	18	42	65	75	379,61

¹ A descrição quanto ao rótulo das covariáveis se encontra no Apêndice A.

² Informação quanto ao repositório de origem, sendo: NP = Nasa POWER, SG = *SoilGrids*; WC = *WorldClim*.

³ Número de vezes em que cada covariável ocupou cada posição dentro do top-5.

As covariáveis melhores ranqueadas foram o coeficiente de variação para a precipitação sazonal, o percentual de fragmentos grossos na camada de 30 – 60 cm do solo e a densidade do carbono orgânico na camada de 5 – 15 cm. Juntas, essas covariáveis foram as mais importantes do índice ambiental em 98,81% dos casos, sendo que a covariável que aborda a precipitação foi considerada a mais importante para o ambiente em 84,35% dos casos.

O *IncMSE* é utilizado para medir a importância relativa de cada covariável na predição de um modelo, sendo que as covariáveis com maiores incrementos relativos são tidas como mais importantes, indicando que sua presença é essencial para a precisão do modelo. É interessante notar também que variáveis de todos repositórios estão representadas nesse ranking e quando ampliadas para as 20 covariáveis com maior pontuação, pode-se observar uma distribuição mais equilibrada entre as fontes.

As covariáveis mais importantes para o método *IncNodePurity*, seu repositório de origem, quantas vez elas figuraram em cada posição do top-5 e sua pontuação se encontram na Tabela 4:

Tabela 4. Vinte covariáveis melhor ranqueadas pelo método *IncNodePurity*.

Covariável ¹	Base ²	TOP 1 ³	TOP 2	TOP 3	TOP 4	TOP 5	Pontuação
bio_15	WC	9425	467	67	29	4	96792,17
phh2o_30-60cm_Q0p5	SG	367	4801	2052	1040	608	34665,41
phh2o_60-100cm_Q0p5	SG	124	2222	2395	1454	1010	21132,13
bio_7	WC	16	887	1975	2277	1759	14001,37
bio_12	WC	7	450	878	1102	1223	7310,41
phh2o_15-30cm_Q0p5	SG	12	489	845	1008	956	7055,02
ocd_5-15cm_Q0p05	SG	7	165	528	840	1240	4740,41
T2M_MAX_2010a2019OCT	NP	0	101	303	570	815	3149,51
nitrogen_30-60cm_Q0p95	SG	35	201	271	376	432	3109,31
T2MDEW_2010a2019AUG	NP	2	74	210	391	524	2172,93
T2M_2010a2019AUG	NP	1	59	152	248	408	1666,70
cfvo_30-60cm_Q0p95	SG	3	31	126	261	363	1379,18
silt_100-200cm_Q0p05	SG	0	6	52	117	233	717,64
ALLSKY_DWN_10a19OCT	NP	0	23	73	130	156	679,30
ALLSKY_DWN_10a19NOV	NP	0	8	25	30	62	245,59
PRECTOTCORR_10a19APR	NP	0	5	7	24	39	189,55
nitrogen_30-60cm_mean	SG	0	3	9	25	43	176,62
cfvo_100-200cm_Q0p95	SG	0	4	7	9	9	65,18
WS2M_2010a2019JUL	NP	1	0	3	9	9	55,70
ocd_15-30cm_Q0p05	SG	0	0	1	4	5	41,05

¹ A descrição quanto ao rótulo das covariáveis se encontra no Apêndice A.

² Informação quanto ao repositório de origem, sendo: NP = Nasa POWER, SG = *SoilGrids*; WC = *WorldClim*.

³ Número de vezes em que cada covariável ocupou cada posição dentro do top-5.

Para o segundo método foi possível observar ainda mais claramente uma segregação quanto a importância das covariáveis, o que é esperado quando se utiliza o *IncNodePurity*. As três variáveis mais importantes, coeficiente de variação para a precipitação sazonal e o Ph do solo medido em água, para as profundidades 30-60 cm e 60-100 cm, ocuparam a primeira posição de importância em 99,16%, sendo que a covariável que remete à pluviometria foi a mais importante em 94,25% das interações testadas.

O método *IncNodePurity* mede o incremento de homogeneidade dos nós a partir do índice de Gini e determina a importância relativa de uma covariável para a construção da

árvore de decisão, então é natural que, quando há uma covariável que seja extremamente mais relevante, os resultados sejam mais estratificados. Observa-se também, uma distribuição mais equilibrada quanto aos repositórios de origem quando comparada com o método anterior. Adicionalmente, se pode notar que quatro covariáveis estão presentes nos dois rankings: coeficiente de variação para a precipitação sazonal, precipitação anual, densidade do carbono orgânico na camada de 5 – 15 cm e temperatura máxima medida no mês de outubro.

A produtividade média projetada de cada pixel é apresentada na Figura 3A, e foi calculada a partir dos dados obtidos pelos métodos BLUE e BLUP nas 187 localidades de experimentos da Embrapa. É interessante notar a concentração dos pixels mais produtivos na região Nordeste, em algumas localidades da região Sul e no estado de Minas Gerais. Além disso, na Figura 2B, pode-se analisar o desvio padrão em relação à média projetada, na qual se observa que a região Nordeste apresenta as maiores concentrações de desvio padrão por pixel, ao contrário das outras regiões do Brasil, que em sua maioria possuem desvios pertencentes às duas primeiras classes.

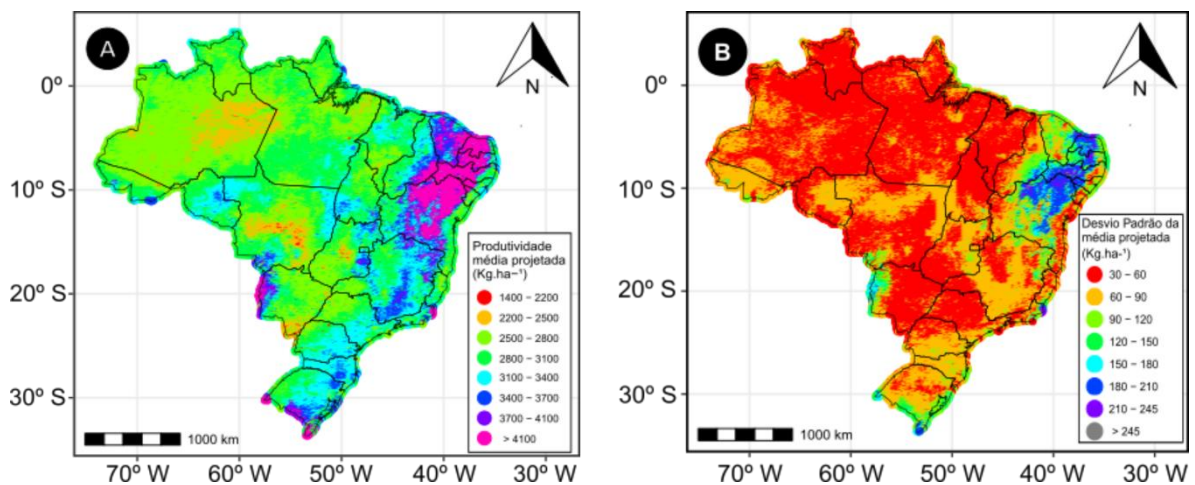


Figura 3. Mapas de (A) Produtividade média projetada obtido a partir do método de Monte Carlo para os dados obtidos de BLUE/BLUP dos materiais genéticos; e (B) desvio padrão da média projetada da parte A desta figura, para todo o território brasileiro.

5. DISCUSSÃO

5.1. Explorando a variabilidade genética do arroz de terras altas no território brasileiro

A utilização de um grande volume de dados fenotípicos de arroz, abrangendo uma ampla área de estudo e em múltiplos anos pode enriquecer significativamente a análise de melhoramento (Arief et al., 2019). Essa abordagem possibilita uma compreensão mais completa dos fenótipos do arroz, permitindo identificar padrões e relações importantes. No entanto, para detectar melhor os padrões da interação genótipos por ambientes ($G \times A$), os dados provenientes de terras altas se apresentaram mais bem distribuídos em uma maior gama de ambientes em comparação com os dados de sistemas irrigados (Breseghello et al., 2021). Isso ocorre porque os fenótipos do arroz de terras altas são mais suscetíveis às variações ambientais do que os de terras irrigadas.

Com posse de uma estrutura de *big data*, ao considerar a interação $G \times A$, é mais interessante tratar dados genéticos desbalanceados em vários locais do que ter uma distribuição balanceada em poucos locais (Resende et al., 2020). Essa abordagem permite capturar a variabilidade genética e ambiental em diferentes contextos, fornecendo informações valiosas sobre a adaptação e o desempenho dos genótipos em ambientes diversos.

A integração de informações geográficas também desempenha um papel fundamental nesse contexto. Tal como será discutido no presente trabalho, por meio da espacialização dos resultados, é possível obter uma melhor compreensão dos padrões espaciais envolvidos aos dados fenotípicos e genéticos do arroz. Isso não apenas auxilia na visualização dos padrões geográficos, mas também facilita a construção de estratégias específicas para cada região, levando em consideração as particularidades e necessidades locais do arroz de terras altas (van Eeuwijk et al., 2019; Costa-Neto et al., 2020).

São apresentados os dados de herdabilidade projetada em cada pixel do território brasileiro (Figura 2A). É possível identificar uma faixa de localidades que exibem altas herdabilidades, estendendo-se de Goiás até o Rio Grande do Sul, com os valores mais expressivos concentrados na região Sul do país. Na maior parte do território, a herdabilidade variou entre 0,2 e 0,4, o que ressalta a importância de uma análise aprofundada do comportamento dos genótipos de interesse e sua estabilidade local em relação a características ambientais específicas (van Frank et al., 2020).

5.2. Derivação de multimarcadores ambientômicos e a importância das covariáveis ambientais

O uso das técnicas de ambientômica na agricultura e no melhoramento de plantas vem se consolidando desde o final da década passada. O principal foco é em agilizar as etapas iniciais do melhoramento utilizando informações fenotípicas de plantios já consolidados para realizar uma melhor seleção genotípica a partir da caracterização do ambientoma local (Binkley et al., 2017; Grattapaglia et al., 2018; Rubilar et al., 2020; Crossa et al., 2021).

Os novos estudos utilizando a ambientômica tendem a aperfeiçoar o conhecimento já desenvolvido até o momento. Os parâmetros utilizados por pesquisadores do zoneamento climático e da agricultura de precisão continuam válidos e estão sendo validados pela ambientômica. Como apresentado anteriormente, tanto o método *IncMSE* quanto o *IncNodePurity* retornaram que a covariável mais importante para a modelagem do ambientoma foi o coeficiente de variação para a precipitação sazonal, contando ainda com a precipitação anual (bio_12) figurando entre as 5 covariáveis ambientais mais importantes, desenhando assim uma clara relação entre as taxas de precipitação e suas variáveis relacionadas com a produtividade da cultura (Silvando et al., 1998; Silva & Assad, 2001; Lorençoni et al., 2010; Januário et al., 2018; Santana et al., 2022).

O percentual de fragmentos grossos na camada 30-60 cm do solo e a densidade de carbono orgânico na camada de 5-15 cm foram outras covariáveis apontadas como sendo de extrema importância para a montagem do ambientoma pelo método *IncMSE*. O percentual de fragmentos grossos pode reduzir a capacidade de drenagem do solo e criar uma camada compactada que impede o crescimento e penetração das raízes (Yadav et al., 2019). A densidade de carbono orgânico, por sua vez, desempenha uma função importante na disponibilização de água e outros nutrientes para a planta além de contribuir para a manutenção da atividade microbiana (Samal et al., 2017; Yadav et al., 2019).

O pH do solo medido em água aparece duas vezes como sendo uma das cinco variáveis mais importantes para a montagem do ambientoma quando analisado o método *IncNodePurity*, e isso pode ser facilmente explicado ao considerar a disponibilidade de macro e micronutrientes em gradientes de pH (Shareef et al., 2019). Solos mais ácidos tendem a aumentar a concentração de Al, Fe e Mn atrapalhando o crescimento e desenvolvimento da planta, enquanto pH's mais elevados tendem a atrapalhar o

desenvolvimento radicular e diminuir a área foliar dos indivíduos (Neina, 2019; Shareef et al., 2019; Medina-Litardo et al., 2022).

Outras covariáveis aparecem com frequência dentre as vinte mais importantes para os modelos, como os percentuais de nitrogênio e silte, temperatura máxima, amplitude térmica e a irradiação total. A ambientômica apresenta uma visão holística a respeito da quantidade de covariáveis analisáveis, o grau de precisão dos modelos de predição e os graus de detalhamento a níveis genéticos e territoriais (Binkley et al., 2020).

O uso da ambientômica no melhoramento genético de plantas abrirá caminhos para a instalação de novos sítios produtivos em locais que antes poderiam ter sido considerados inadequados a uma cultura. Cada ambiente deve ser analisado de forma única e um laudo de aptidão para o plantio de interesse deve levar em consideração a vasta gama de materiais genéticos disponíveis e suas diferentes necessidades (Mosleh et al., 2015; Resende et al., 2018).

5.3. Leitura dos marcadores ambientômicos e sua influência na produtividade do arroz de terras altas

Todas as regiões brasileiras foram representadas com experimentos contendo genótipos de arroz de terras altas, e cada marcador ambientômico mostra um perfil diferente de potencial produtivo da área, com resultados por pixel. No entanto, nem sempre dois locais semelhantes ambientalmente apresentam padrões de interação genótipo-ambiente semelhantes. Isso significa que um ambiente pode ser adequado para um conjunto de genótipos e inadequado para outros devido à complexidade dos fatores ambientais e sua correlação. Além disso, nem sempre locais próximos apresentam interação GxA semelhante, o que pode tornar a seleção de genótipos mais desafiadora (Costa-Neto et al., 2021).

Através da análise da Figura 3A, nota-se que a área central do Nordeste, caracterizada pelo clima tropical semiárido, destaca-se como uma região promissora em termos de produtividade média. Isso sugere que, mesmo com a alternância dos genótipos testados no experimento de Monte Carlo, a região continua apresentando altos níveis de produtividade. No entanto, essa mesma região está associada a um desvio padrão significativo em relação à média produtiva, indicando que, apesar da tendência de manter uma alta produtividade média ao alternar os genótipos testados, ainda há uma variação considerável na escolha dos genótipos, afetando os níveis de produtividade (Li et al., 2013; Ferreira, 2019; Moeinizade et al., 2021).

No contexto da região Nordeste, a distribuição dos experimentos está fortemente concentrada no estado do Maranhão (Figura 2A). Novos experimentos devem ser realizados em regiões que apresentam alta produtividade média projetada e alto desvio padrão. Uma alta variância genética é desejável para a identificação, desenvolvimento e teste de genótipos específicos. Portanto, uma localidade que apresenta alto potencial produtivo associado a uma alta variância é um local adequado para a expansão de plantios e testes.

Já no sul do estado de Minas Gerais e na divisa de Santa Catarina com o Rio Grande do Sul, observa-se regiões com produtividade acima da média, associadas com um baixo desvio padrão. Para essas localidades existe a tendência de ter bons resultados produtivos com baixa variação independentemente do genótipo plantado e das condições ambientais. Para o desenvolvimento de linhagens essas regiões são valiosas, uma vez que a incerteza do efeito ambiental não vai ser um fator de imprevisibilidade dos resultados (Alves et al., 2021).

Os locais com menor produtividade, conforme demonstrado na Figura 3A, estão concentrados nos estados de Mato Grosso, Mato Grosso do Sul, Goiás e Amazonas, em áreas próximas a experimentos. Essa situação pode ser atribuída à testagem de genótipos que não são adequados para essas localidades, resultando em uma diminuição geral na produtividade média do local. É importante ressaltar que uma baixa produtividade projetada em alguma dessas áreas não significa que ela seja inadequada para o cultivo de arroz, pode-se apenas presumir que alguns dos genótipos testados não apresentaram seu melhor desempenho nas condições ambientais específicas desses locais. Para melhorar a produtividade desses locais, além de uma recomendação genotípica mais assertiva, pode-se também pensar em adequação de práticas culturais como otimização das datas de plantio, implementação de técnicas adequadas de manejo da água, adoção de estratégias adequadas de controle de plantas daninhas e pragas, e otimização das taxas de aplicação de fertilizantes (Saito et al., 2018; Kusbiantoro et al., 2020).

Trabalhos de zoneamento climático para a cultura da *Oryza sativa* usaram parâmetros como temperatura média, precipitação acumulada, riscos de intempéries, etc, para apresentarem suas previsões. Porém, os parâmetros eram na maioria das vezes escolhidos de maneira empírica e em número reduzido. Ferramentas tecnológicas como o desenvolvimento de pacotes para análise de dados e repositórios de informações climáticas contribuem para a criação e aplicação de modelos robustos que englobem diversos fatores climáticos, bioclimáticos, ambientais, geográficos e pedológicos (da Silva et al., 1995; Silva & Assad, 2001; Costa-Neto & Fritsche-Neto, 2021).

Os trabalhos de zoneamento tendem a se ater na análise única ou prioritária do balanço hídrico e do risco associado à cultura quando deslocada das áreas historicamente consolidadas para o plantio. A ambientômica permite realizar um estudo mais profundo voltado não somente para os caracteres ambientais de cada macro região de um estado, mas para cada pequena unidade de terra disponível para plantio, isso fica evidente ao se observar classificações díspares em uma região diminuta, como nos estados da região Centro – Oeste e em alguns estados da região Norte. A participação das agências de pesquisa e consultoria a níveis municipais podem ajudar a calibrar ainda mais o modelo, fornecendo dados locais mais precisos das condições climáticas e de solo (Silvando et al., 1998; Steinmetz & Silvando, 2017).

Em uma revisão sobre o futuro do arroz de terras altas no país, Santana et al. (2022) citam a possibilidade de expansão da cultura para localidades sob agricultura intensiva. O uso da ambientômica auxiliaria na escolha do binômio genótipo-ambiente, sendo um fator importante na tomada de decisão para a determinação em realizar ou não a troca de cultura (ou cultivar) a partir de parâmetros econômicos de produtividade e gastos associados. É interessante observar que a plataforma NASA POWER oferece ferramentas e dados preditivos que podem ser incorporados ao modelo a depender da necessidade da análise.

Em um cenário de mudanças climáticas e de um custo elevado relacionado à abertura de novos sítios de plantio, deve-se prezar pela rápida tomada de decisão e melhoria da produtividade local. A gestão de recursos genéticos embasada pela ambientômica vem em auxílio dos produtores e melhoristas que poderão definir melhores combinações de localidades e genótipos, explorando o potencial máximo dos sítios já existentes (Marcatti et al., 2017; Resende et al., 2018; Binkley et al., 2020).

Os índices ambientais associados a cada um dos pontos da malha brasileira obtidos nesse estudo são extremamente úteis não só para o cultivo de arroz de terras altas, mas como para outras culturas anuais ou perenes. Marcatti et al. (2017), em estudo sobre recomendação de genótipos de eucalipto, apresentam de forma clara os usos e benefícios dos índices ambientais para plantios florestais equiâneos. Utilizando também índices ambientais, Curin et al. (2020) analisaram condições ótimas para a maximização da produtividade do milho na Argentina.

É importante ressaltar que os marcadores ambientômicos são ferramentas de baixo custo associado, muitos deles estando disponíveis online e de forma gratuita, o que permite não só a coleta e armazenamento de dados brutos, mas sua manipulação facilitada

(Resende et al., 2020; Costa-Neto et al., 2021). Ainda mais, os marcadores ambientômicos não são de uso exclusivo para uma determinada espécie, o que pode colaborar na diversificação e crescimento da produção da agricultura brasileira, auxiliando principalmente pequenos produtores que trabalham com uma diversidade maior de plantios

Integrando a ambientômica ao melhoramento de plantas é possível ainda trabalhar na predição da produção e na determinação de materiais genéticos específicos que maximizarão a produtividade local. A criação dos marcadores ambientais é apenas um dos passos dentro da ambientômica.

6. CONCLUSÃO

Considerando a amplitude do território brasileiro e a complexidade dos requerimentos para a montagem de um ambiente é importante dar prioridade à realização de experimentos com diversos tipos de materiais genéticos em diferentes localidades e a obtenção da maior quantidade e diversidade de dados que possibilitem a caracterização do ambiente de forma mais completa.

Para os dois métodos de avaliação utilizados nesse estudo a covariável mais importante para a montagem de um ambiente foi o coeficiente de variação da precipitação sazonal. Em todos os rankings de importância foi observada presença de covariáveis dos três repositórios analisados, reforçando a necessidade de sempre manter um banco de dados robusto e atualizado para os caracteres ambientais.

A porção central da Região Nordeste tem potencial para receber áreas de testagem de genótipos. O sul de Minas Gerais e a região da divisa entre Santa Catarina e Rio Grande do Sul podem ser usadas para o desenvolvimento de linhagens. Os valores de herdabilidade projetada para a grande maioria do território brasileiro embasam o investimento em testagem e experimentação a fim de conhecer a estabilidade local de cada genótipo.

7. REFERÊNCIAS

- ALVES, F. C. et al. Impact of the complexity of genotype by environment and dominance modeling on the predictive accuracy of maize hybrids in multi-environment prediction models. *Euphytica*, v. 217, n. 37, 1 mar. 2021.
- ARIEF, V. N. et al. Utilization of multiyear plant breeding data to better predict genotype performance. *Crop Science*, v. 59, n. 2, p. 480–490, 1 mar. 2019.
- BINKLEY, D. et al. Variation in whole-rotation yield among Eucalyptus genotypes in response to water and heat stresses: The TECHS project. *Forest Ecology and Management*, v. 462, n. January, p. 117953, 2020.
- BRAR, D. S.; SINGH, K.; KHUSH, G. S. *Frontiers in Rice Breeding. Em: The Future Rice Strategy for India*. [s.l.] Elsevier, 2017. p. 137–160.
- BREIMAN, L. Random Forests. *Machine Learning*, v. 45, p. 5–32, 2001.
- BRESEGHELLO, F. et al. Crop Science Building the Embrapa rice breeding dataset for efficient data reuse. 2021.
- CARRASCO, L. et al. Historical mapping of rice fields in Japan using phenology and temporally aggregated Landsat images in Google Earth Engine. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 191, p. 277–289, 1 set. 2022.
- CERASOLI, F.; D’ALESSANDRO, P.; BIONDI, M. Worldclim 2.1 versus Worldclim 1.4: Climatic niche and grid resolution affect between-version mismatches in Habitat Suitability Models predictions across Europe. *Ecology and Evolution*, v. 12, n. 2, 1 fev. 2022.
- COELHO, G. R. C. et al. Genetic diversity of high performance cultivars of upland and irrigated Brazilian rice. *Genetics and Molecular Research*, v. 16, n. 3, 21 set. 2017.
- COSTA-NETO, G. et al. A novel GIS-based tool to reveal spatial trends in reaction norm: upland rice case study. *Euphytica*, v. 216, n. 37, 1 mar. 2020.
- COSTA-NETO, G. et al. EnvRtype: A software to interplay enviromics and quantitative genomics in agriculture. *G3: Genes, Genomes, Genetics*, v. 11, n. 4, 2021.
- COSTA-NETO, G. et al. Envirome-wide associations enhance multi-year genome-based prediction of historical wheat breeding data. *G3 Genes|Genomes|Genetics*, v. 13, n. 2, 9 fev. 2023.
- COSTA-NETO, G.; FRITSCHÉ-NETO, R. Enviromics: bridging different sources of data, building one framework. *Crop Breeding and Applied Biotechnology*, v. 21, n. e393521S12, p. 1–14, 2021.
- CROSSA, J. et al. The Modern Plant Breeding Triangle: Optimizing the Use of Genomics, Phenomics, and Enviromics Data. *Frontiers in Plant Science*, v. 12, n. April, p. 1–6, 2021.
- CUARTAS ARIAS, M. Functional enviromics and personality disorders. *International Journal of Psychological Research*, v. 4, n. 2, p. 5–8, 2011.

- CURIN, F. et al. Water and radiation use efficiencies in maize: Breeding effects on single-cross Argentine hybrids released between 1980 and 2012. *Field Crops Research*, v. 246, 1 fev. 2020.
- DA SILVA, S. C. et al. Zoneamento agroclimático para o arroz de sequeiro no estado de Goiás. *Embrapa SPI*, v. 43, p. 80, 1995.
- DANDABATHULA, G. et al. Validation of SoilGrids 2.0 in an Arid Region of India using In Situ Measurements. *European Journal of Environment and Earth Sciences*, v. 3, n. 6, p. 49–58, 22 nov. 2022.
- DE CARVALHO, D. F. et al. Evolution and current scenario of irrigated area in Brazil: Systematic data analysis. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 24, n. 8, p. 505–511, 1 ago. 2020.
- DEVI, K. R. et al. AMMI Biplot Analysis for Genotype X Environment Interaction and Stability for Yield in Hybrid Rice (*Oryza sativa* L.) Under Different Production Seasons. *Current Journal of Applied Science and Technology*, v. 39, n. 48, p. 169–175, 31 dez. 2020.
- DOROK, S. et al. Flexible Analysis of Plant Genomes in a Database Management System. *EDTB*, p. 509–512, 2015.
- FERREIRA, G. M. L. Atlas geográfico: espaço mundial. 5. ed. São Paulo: Editora Moderna, 2019.
- GIANI, A. M. et al. Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal* Elsevier B.V., , 1 jan. 2020.
- GLAUBITZ, J. C. et al. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. 2014.
- HAIDER, Z. et al. Comparison of GGE biplot and AMMI analysis of multi-environment trial (MET) data to assess adaptability and stability of rice genotypes. *African Journal of Agricultural Research*, v. 12, n. 51, p. 3542–3548, 21 dez. 2017.
- HASHIM, N. et al. Integrating multivariate and univariate statistical models to investigate genotype–environment interaction of advanced fragrant rice genotypes under rainfed condition. *Sustainability (Switzerland)*, v. 13, n. 8, 2 abr. 2021.
- HEINEMANN, A. B. et al. Upland rice breeding led to increased drought sensitivity in Brazil. *Field Crops Research*, v. 231, p. 57–67, 2019.
- HEINEMANN, A. B. et al. RISCO CLIMÁTICO E PERÍODO DE SEMEADURA PARA O ARROZ IRRIGADO NO TOCANTINS. *AGRI-ENVIRONMENTAL SCIENCES*, v. 7, n. 2, p. 13, 14 jun. 2021.
- JANUÁRIO, I. R.; MENEZES, R. N.; BONFIM, O. T. Zoneamento agroclimático para o plantio do arroz na mesorregião central do Estado do Maranhão. *Ciência e Natura*, v. 40, p. 13, 27 mar. 2018.
- KUSBIANTORO, D. et al. Development and Improvement of Upland Rice Productivity through Dry Land Utilization. *Scitepress*, 31 jul. 2020.

- LI, R. et al. A Monte Carlo Permutation Test for Random Mating Using Genome Sequences. *PLoS ONE*, v. 8, n. 8, 5 ago. 2013.
- LIMA, N. et al. Plasticidade Fenotípica. *Revista de Ciência Elementar*, v. 5, n. 2, 2017.
- LORENÇONI, R.; NETO, D. D.; HEINEMANN, A. B. Calibration and evaluation of the ORYZA-APSIM crop model for upland rice in Brazil. *Revista Ciencia Agronomica*, v. 41, n. 4, p. 605–613, 2010.
- MARCATTI, G. E. et al. GIS-based approach applied to optimizing recommendations of Eucalyptus genotypes. *Forest Ecology and Management*, v. 392, p. 144–153, 2017.
- MCOMISH, C. E.; BURROWS, E. L.; HANNAN, A. J. Identifying novel interventional strategies for psychiatric disorders: integrating genomics, “enviromics” and gene-environment interactions in valid preclinical models. *British journal of pharmacology*, v. 171, n. 20, p. 4719–4728, 2014.
- MEDINA-LITARDO, R. C. et al. Salinity of soil and irrigation water on rice productivity in the canton of San Jacinto de Yaguachi, Ecuador. *Agrociencia*, v. 56, n. 7, 2022.
- MEUWISSEN, T. H. E. et al. Management of Genetic Diversity in the Era of Genomics. *Frontiers in Genetics*, v. 11, 13 ago. 2020.
- MOEINIZADE, S. et al. A look-ahead Monte Carlo simulation method for improving parental selection in trait introgression. *Scientific Reports*, v. 11, n. 1, 1 dez. 2021.
- MONTEIRO, L. A.; SENTELHAS, P. C.; PEDRA, G. U. Assessment of NASA/POWER satellite-based weather system for Brazilian conditions and its impact on sugarcane yield simulation. *International Journal of Climatology*, v. 38, n. 3, p. 1571–1581, 1 mar. 2018.
- MORAIS, L. K. DE et al. Phenotypic adaptability and stability of genotype of irrigation Rice for the state of São Paulo. *Revistas Biociências*, v. 14, n. 1, 2008.
- MORALES, N. et al. ImageBreed: Open-access plant breeding web–database for image-based phenotyping. *Plant Phenome Journal*, v. 3, n. 1, 2020.
- MOSLEH, M. K.; HASSAN, Q. K.; CHOWDHURY, E. H. Application of remote sensors in mapping rice area and forecasting its production: A review. *Sensors (Switzerland)*, v. 15, n. 1, p. 769–791, 2015.
- MUENCHEN, R. Update to Data Science Software Popularity. Disponível em: <<https://r4stats.com/2023/06/07/update-to-data-science-software-popularity/>>. Acesso em: 19 jun. 2023.
- NEINA, D. The Role of Soil pH in Plant Nutrition and Soil Remediation. *Applied and Environmental Soil Science* Hindawi Limited, , 2019.
- POGGIO, L. et al. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL*, v. 7, n. 1, p. 217–240, 14 jun. 2021.
- QUANSAH, A. D. et al. Assessment of solar radiation resource from the NASA-POWER reanalysis products for tropical climates in Ghana towards clean energy application. *Sci Rep*, v. 12, p. 10684, 2022.

REED, P. L.; STORR, C. L.; ANTHONY, J. C. Drug dependence enviromics: Job strain in the work environment and risk of becoming drug-dependent. *American Journal of Epidemiology*, v. 163, n. 5, p. 404–411, 2006.

RESENDE, R. T. et al. Environmental uniformity, site quality and tree competition interact to determine stand productivity of clonal Eucalyptus. *Forest Ecology and Management*, v. 410, n. September 2017, p. 76–83, 2018.

RESENDE, R. T. et al. Enviromics in breeding: applications and perspectives on envirotypic-assisted selection. *Theoretical and Applied Genetics*, v. 134, n. 1, p. 95–112, 2020.

RESENDE, R. T. et al. Editorial: Enviromics in Plant Breeding. *Frontiers in Plant Science*, v. 13, p. 1939, 30 jun. 2022.

REYES, V. P. et al. Utilization of Genotyping-by-Sequencing (GBS) for Rice Pre-Breeding and Improvement: A Review. *Life* 2022, Vol. 12, Page 1752, v. 12, n. 11, p. 1752, 1 nov. 2022.

RIPLEY, B. D.; VENABLES, W. N. *Modern Applied Statistics with S*. fourth ed. New York: Springer, 2002.

SAITO, K. et al. Progress in varietal improvement for increasing upland rice productivity in the tropics. *Plant Production Science* Taylor and Francis Ltd., , 3 jul. 2018.

SAMAL, S. K. et al. Evaluation of long-term conservation agriculture and crop intensification in rice-wheat rotation of Indo-Gangetic Plains of South Asia: Carbon dynamics and productivity. *European Journal of Agronomy*, v. 90, p. 198–208, 1 out. 2017.

SANTANA, C. A. M.; SOUZA, G. DA S. E; GOMES, E. G. O futuro do arroz de terras altas no Brasil: cultivo de oportunidade. *Revista de Política Agrícola*, v. 31, n. 2, p. 51, 2022.

SHAREEF, R. S.; MAMAT, A. S.; AL-SHAHEEN, M. R. The Effect of Soil PH, High-Calcium Compost and Cadmium on Some of Growth Characters in Corn (*Zea mays*l.). *ARC Journal of Pharmaceutical Sciences*, v. 5, n. 4, p. 16–27, 2019.

SILVA, S. C. DA; ASSAD, E. D. Zoneamento de riscos climáticos para o arroz de sequeiro nos estados de Goiás, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Tocantins e Bahia. *Revista Brasileira de Meteorologia*, v. 9, n. 3, p. 536–543, 2001.

SILVANDO, C. DA S.; BRITES, R. S.; ASSAD, E. D. IDENTIFICAÇÃO DE RISCO CLIMÁTICO PARA A CULTURA DE ARROZ DE SEQUEIRO NO ESTADO DE GOIÁS. *Pesquisa Agropecuária Brasileira*, v. 33, n. 7, p. 1005–1011, 1998.

SOIL SURVEY STAFF. Tree-based Models. Em: *Statistics for Soil Survey - Part 2*. [s.l.: s.n.].

STEINMETZ, S.; SILVA, S. C. Início dos estudos sobre Zoneamento Agrícola de Risco Climático (ZARC) no Brasil. Documentos Santo Antônio de Goiás, GOEmbrapa Arroz e Feijão, , 2017.

STEINMETZ, S.; SILVANDO, C. DA S. Início dos Estudos sobre Zoneamento Agrícola de Risco Climático (ZARC) no BrasilEm. Santo Antônio de Goiás: [s.n.]. Disponível em: <www.embrapa.br/fale-conosco/sac>. Acesso em: 26 mar. 2023.

TANG, L. et al. RiceGrow: A rice growth and productivity model. *NJAS-Wageningen Journal of Life Sciences*, v. 57, p. 83–92, 2009.

VAN EEUWIJK, F. A. et al. Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Science*, v. 282, p. 23–39, 1 maio 2019.

VAN EEUWIJK, F. A.; BUSTOS-KORTS, D. V.; MALOSETTI, M. What should students in plant breeding know about the statistical aspects of genotype $\times$ Environment interactions? *Crop Science*, v. 56, n. 5, p. 2119–2140, 2016.

VAN FRANK, G. et al. Genetic diversity and stability of performance of wheat population varieties developed by participatory breeding. *Sustainability (Switzerland)*, v. 12, n. 1, p. 1–21, 1 jan. 2020.

WANG, B. et al. MeMYB26, a drought-responsive transcription factor in cassava (*Manihot esculenta* Crantz). *Crop Breeding and Applied Biotechnology*, v. 21, n. 1, p. 1–9, 2021.

WANG, N. et al. Applications of genotyping-by-sequencing (GBS) in maize genetics and breeding. *Scientific Reports*, v. 10, n. 16308, 2020.

WEI, M. et al. Investigating the Potential of Crop Discrimination in Early Growing Stage of Change Analysis in Remote Sensing Crop Profiles. *Remote Sensing 2023*, Vol. 15, Page 853, v. 15, n. 3, p. 853, 3 fev. 2023.

XU, Y. Envirotyping for deciphering environmental impacts on crop plants. *Theoretical and Applied Genetics*, v. 129, n. 4, p. 653–673, 2016.

YADAV, G. S. et al. Conservation tillage and nutrient management effects on productivity and soil carbon sequestration under double cropping of rice in north eastern region of India. *Ecological Indicators*, v. 105, p. 303–315, 1 out. 2019.

ZHAO, Y. et al. A high-performance database management system for managing and analyzing large-scale snp data in plant genotyping and breeding applications. *Agriculture*, v. 11, n. 11, 1 nov. 2021.

Apêndice A. Descrição das covariáveis que figuraram entre as vinte mais importantes para os métodos IncMSE e IncNodePurity.

Covariável	Base	Descrição
ALLSKY_DWN_10a19NOV	NP ¹	All Sky Surface Shortwave Downward Irradiance in November
ALLSKY_DWN_10a19OCT	NP	All Sky Surface Shortwave Downward Irradiance in October
bio_12	WC ²	Annual Precipitation
bio_15	WC	Precipitation Seasonality (CoV)
bio_18	WC	Precipitation of Warmest Quarter
bio_7	WC	Temperature Annual Range
cec_100-200cm_mean	SG ³	Mean of Cation exchange capacity on 100-200 depth
cec_100-200cm_Q0p5	SG	Cation exchange capacity on 100-200 depth (50%)
cfvo_100-200cm_Q0p95	SG	Coarse fragments on 100-200 cm depth (95%)
cfvo_30-60cm_mean	SG	Mean of coarse fragments on 30-60 cm depth
cfvo_30-60cm_Q0p95	SG	Coarse fragments on 30-60 cm depth (95%)
cfvo_5-15cm_Q0p95	SG	Coarse fragments on 5-15 cm depth (95%)
nitrogen_100-200cm_mean	SG	Mean of total nitrogen on 100-200 cm depth
nitrogen_30-60cm_mean	SG	mean of total nitrogen on 30-60 cm depth
nitrogen_30-60cm_Q0p95	SG	Total nitrogen on 30-60 cm depth (95%)
ocd_15-30cm_Q0p05	SG	Organic Carbon Density on 15-30 depth (5%)
ocd_5-15cm_Q0p05	SG	Organic Carbon Density on 5 -15 depth (5%)
phh2o_15-30cm_Q0p5	SG	Soil pH on water for the 15-30 depth (50%)
phh2o_30-60cm_Q0p5	SG	Soil pH on water for the 30-60 depth (50%)
phh2o_60-100cm_Q0p5	SG	Soil pH on water for the 60-100 depth (50%)
PRECTOTCORR_10a19APR	NP	Precipitation Corrected in April
PRECTOTCORR_10a19MAR	NP	Precipitation Corrected in March
RH2M_2010a2019OCT	NP	Relative Humidity at 2 Meters in October
silt_100-200cm_Q0p05	SG	Silt percentage on 100-200 cm depth (5%)
soc_30-60cm_mean	SG	Mean of Organic Carbon Concentration on 30-60 cm depth
T2M_2010a2019AUG	NP	Average temperature at 2 Meters in August
T2M_MAX_2010a2019OCT	NP	Maximum Temperature at 2 Meters in October

T2MDEW_2010a2019AUG	NP	Dew point temperature at 2 meters in August
T2MDEW_2010a2019JUN	NP	Dew point temperature at 2 meters in June
WS2M_2010a2019JUL	NP	Wind Speed at 2 Meters in July

*Dados extraídos dos repositórios:

¹<https://power.larc.nasa.gov/#resources>

²<https://www.worldclim.org/data/bioclim.html>

³<https://data.isric.org/geonetwork/srv/por/catalog.search#/home>