UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO - UNIRIO

CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE – CCBS

PROGRAMA DE PÓS-GRADUAÇÃO EM ALIMENTOS E NUTRIÇÃO – PPGAN

Carolina Thomaz dos Santos D'Almeida

Protein-digestibility and sorghum phenolic compounds from flowering to maturation: processing, bioaccessibility and interactions by foodomics strategies

Rio de Janeiro

August, 2024

Carolina Thomaz dos Santos D'Almeida

Protein-digestibility and sorghum phenolic compounds from flowering to maturation: processing, bioaccessibility and interactions by foodomics strategies

PhD defense presented to the jury of the Graduate Program in Food and Nutrition, Federal University of the State of Rio de Janeiro, as partial requirement for the PhD title.

Supervisor: Prof. Dr. Mariana Simões Larraz Ferreira
Co-supervisor: Dr. Hamza Mameri

Rio de Janeiro
August, 2024

Catalogação informatizada pelo(a) autor(a)

D148 D'Almeida, Carolina Thomaz dos Santos
Protein-digestibility and sorghum phenolic compounds from flowering to maturation: processing, bioaccessibility and interactions by foodomics strategies / Carolina Thomaz dos Santos D'Almeida. -- Rio de Janeiro, 2024.
256

Orientadora: Mariana Simões Larraz Ferreira.
Coorientadora: Hamza Mameri.
Tese (Doutorado) - Universidade Federal do Estado do Rio de Janeiro, Programa de Pós-Graduação em Alimentos e Nutrição, 2024.

1. digestibility. 2. metabolomics. 3. phenolic profile. I. Ferreira, Mariana Simões Larraz, orient. II. Mameri, Hamza, coorient. III. Título.

Carolina Thomaz dos Santos D'Almeida

Protein-digestibility and sorghum phenolic compounds from flowering to maturation: processing, bioaccessibility and interactions by foodomics strategies

PhD defense presented to the jury of the Graduate Program in Food and Nutrition, Federal University of the State of Rio de Janeiro, as partial requirement for the PhD title.

Approved on: 22/08/2024

JURY

Documento assinado digitalmente
gov.br MARIANA SIMOES LARRAZ FERREIRA
Data: 24/09/2024 19:44:40-0300
Verifique em https://validar.iti.gov.br

Prof. Dr Mariana Simões Larraz Ferreira

Universidade Federal do Estado do Rio de Janeiro (UNIRIO)

Digitally signed by trust_
Date: 2024.09.23 10:48:54 -05'00'

Dr Monica Patricia Cala Molina

Universidad de los Andes – Colômbia

Documento assinado digitalmente
gov.br FREDERICO AUGUSTO RIBEIRO DE BARROS
Data: 25/09/2024 10:24:10-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Frederico Augusto Ribeiro de Barros

Universidade Federal de Viçosa (UFV)

Documento assinado digitalmente
gov.br TALITA PIMENTA DO NASCIMENTO
Data: 23/09/2024 17:43:29-0300
Verifique em https://validar.iti.gov.br

Dra. Talita Pimenta do Nascimento

Embrapa Agroenergia

Documento assinado digitalmente
gov.br GUILHERME MIRANDA TAVARES
Data: 24/09/2024 12:09:46-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Guilherme Miranda Tavares

Universidade Estadual de Campinas (UNICAMP)

Dedico esse trabalho aos meus pais Carlos e Claudia e ao meu marido Rafael, que não mediram forças para me ver alcançando sonhos.

Agradeço à Universidade Federal do Estado do Rio de Janeiro (UNIRIO) e ao Programa de Pós-Graduação em Alimentos e Nutrição (PPGAN) por todo suporte financeiro e oportunidades concedidas ao longo de toda da minha jornada científica.

À l'Institut Agro Montpellier, pour la période doctorale en alternance.

À Universidade Estadual de Campinas (UNICAMP), pelo período de Mobilidade Acadêmica Nacional.

Ao CNPq pela concessão da bolsa de doutorado (Processo n° 140245/2020-0), que não só possibilitou o financiamento da pesquisa, mas também permitiu exercer a Mobilidade Acadêmica Nacional.

À CAPES pela bolsa concedida para execução do Doutorado Sanduíche.

Aos membros da banca pela avaliação deste trabalho.

Ao Instituto « Amélioration génétique et adaptation des plantes méditerranéennes et tropicales » (AGAP) e à Nancy Terrier pelo fornecimento de amostras e pelo suporte científico durante o desenvolvimento das análises e elaboração dos artigos.

Ao “Maatschap de Milliano-Meijer” e ao Walter de Milliano pelo fornecimento das amostras de sorgos holandeses.

À minha orientadora Profa. Dra. Mariana Simões Larraz Ferreira, que abriu não só as portas do laboratório para mim, mas do mundo. Agradeço a parceria e confiança que foram construídas nesses quase 10 anos juntas. Por sempre buscar o melhor para mim e todo o grupo, tentando novas oportunidades, parcerias, financiamentos e infraestrutura. Por ter me permitido viver experiências nacionais e internacionais. Por toda formação, conhecimentos compartilhados, dicas científicas e trabalhos corrigidos. Por me impulsionar a tentar voos mais altos. Por me mostrar como é ser ética, democrática, assertiva e respeitada no ambiente profissional. Por confiar em mim e no meu trabalho mais do que eu mesma. Por me inspirar a ser pesquisadora e tentar buscar evoluir a cada dia mais. Obrigada por tudo. Sem a sua orientação e amizade, nada disso seria possível.

À mon co-encadrant Dr. Hamza Mameri, pour m'avoir si bien accueilli à Montpellier pendant mon doctorat et pour le formidable partenariat du master au doctorat. Pour m'avoir encouragé à formuler des hypothèses à la fin de chaque manip. Pour les encouragements constants à chaque rencontre qui m'ont poussé à évoluer et à

m'améliorer. Pour toutes les corrections des articles et résumés. Pour avoir rendu possibles de nouveaux partenariats et de nouvelles expériences scientifiques. Ce fut un immense plaisir d'avoir ton encadrement tout au long de cette thèse.

Ao Prof. Dr. Guilherme Tavares, por ter me permitido viver uma excelente experiência durante a minha Mobilidade Acadêmica Nacional. Por me oferecer toda infraestrutura e suporte para a realização de parte dessa tese. Por não ter me deixado desanimar quando os resultados não corresponderam às nossas expectativas. Por ter me apresentado análises que contribuirão para a minha formação. Por ter me encorajado e me estimulado, mesmo quando eu já estava no meu "último respiro" do doutorado. Por todos os cafés, bolos e "e aí, está dando certo?". Foi um prazer fazer parte do LQA nesses meses.

À Profa. Dra. Édira Castello Branco de Andrade Gonçalves pela excelente gestão do PPGAN enquanto estive no doutorado. Por estar sempre presente, buscar financiamento e oportunidades aos discentes e por trabalhar com excelência para garantir a qualidade do Programa.

Ao meu queridíssimo Laboratório de Bioativos. Seria injusto citar nomes, pois TODOS que passaram pelo laboratório nesses 9 anos que estive lá foram fundamentais para a minha formação. Agradeço por todos os conhecimentos compartilhados. Por todos os cafés e bolos. Por todas as risadas e união. Por todos os desabafos. Por todas as viagens rumo a congressos. Por terem divido a bancada e a vida comigo. O LabBio é uma grande família e que bom que faço parte dela.

Um agradecimento especial à minhas amigas metabolômicas: Lhulhu, Mari e Micbs. Sem vocês nada disso seria possível. Obrigada por me ensinarem tanto diariamente e pela amizade. Minha admiração por vocês é imensa.

Ao queridíssimo Laboratório de Química de Alimentos: Álvaro, Ana, Anne, Claudia, Lê, Lívia, Mari, Regiane e Rose. Por terem me recebido tão bem e por terem me incluído na família (mesmo sendo filha bastarda). Eu estava vivendo o final do doutorado, casamento, mudança e distância ao mesmo tempo, mas vocês tornaram tudo mais leve. Agradeço pelos almoços, por me ouvirem sempre, por me ajudarem com as análises, pelas saídas para tomar uma cerveja e por todo o carinho que me trataram. Por terem vindo ao Rio viver um momento único comigo. Vocês são um pedaço do meu coração em Campinas.

À toute l'équipe française qui m'a accueilli à bras ouverts et m'a permis de réaliser une partie de cette thèse. Un grand merci, en particulier, à Dr. Marie-Hélène Morel, Joëlle

Bonicel et Dr. Lynda Abdelbost, pour leur amour et leur aide constants. Votre soutien a été inestimable et je vous en suis éternellement reconnaissante.

Aos meus pais Carlos e Claudia com todo o meu amor. Mesmo sem entender em detalhes o que eu fazia no laboratório, nunca mediram esforços para que eu vivesse meu sonho. Por terem me incentivado aos estudos e me oferecido todo suporte financeiro e emocional para trilhar esse caminho.

Ao meu marido Rafael Braga, que lutou todas as grandes batalhas comigo. Por ter superado a distância internacional e nacional. Por ter sido corajoso o suficiente para planejar um casamento, uma obra e uma mudança comigo na reta final do meu doutorado e à distância. Por todas as caronas até Campinas, idas até a rodoviária e idas ao aeroporto. Por sempre opinar sobre os meus gráficos, figuras e apresentações. Por não ter me deixado desistir em nenhum momento. Por ter tornado todo esse processo mais leve.

Ao Bud, meu companheiro de escrita. Por ter me acompanhado durante a escrita da tese. Por ter me forçado a pausas para lambeijos e carinhos. Por ter entendido quando eu estava atarefada e os passeios precisavam ser mais curtos.

Aos meus familiares e amigos, por todo suporte e por compreenderem a minha ausência em alguns momentos.

A Deus, que não só me dá força para viver diariamente, como também foi o grande responsável por esses agradecimentos enormes. Tenho certeza de que foi Ele quem colocou tantas pessoas queridas no meu caminho.

A todos que, de alguma forma, contribuíram para a conclusão dessa etapa da minha vida.

À ciência e à pesquisa, que mudaram a minha vida.

Obrigada. Merci.

“Happiness is not to be found in knowledge, but in the acquisition of knowledge.”

Edgar Allan Poe

RESUMO

A inclusão do sorgo (*Sorghum bicolor* L.) na dieta humana tem crescido devido aos seus benefícios agronômicos e à saúde, atribuídos aos seus compostos fenólicos (CF). No entanto, as kafirinas, proteínas de reserva do sorgo, formam corpos proteicos e complexos insolúveis com os CF, afetando sua digestibilidade. Apesar dos avanços, ainda há lacunas na compreensão da síntese das kafirinas e CF e na interação entre esses componentes do grão de sorgo. O objetivo desta tese foi fornecer informações detalhadas sobre: i) a acumulação de CF e proteínas durante a maturação dos grãos; ii) as modificações após processamento biológico, físico e térmico, incluindo a digestibilidade das kafirinas e a bioacessibilidade dos CF; e iii) os mecanismos de interação entre proteínas e CF do sorgo. Foram estudados genótipos de sorgo holandês e francês, contrastantes em níveis de tanino e coloração do pericarpo. Grãos foram colhidos em dez estágios de maturação (7 a 40 DAF), da divisão celular, enchimento do grão, maturidade fisiológica à colheita (n=20). Grãos maduros passaram por três processamentos: germinação (n=9), descorticação (n=12) e cozimento (n=24). CF e proteínas foram extraídos sequencialmente e analisados por metabolômica (UHPLC-ESI-QTOF/MS) e cromatografia de exclusão por tamanho (SE-HPLC), respectivamente. Interações proteína-fenólico foram avaliadas por espectroscopia de fluorescência e *docking* molecular, e a estabilidade térmica dos CF por HPLC-DAD-MS/MS. CF e proteínas do sorgo são sintetizados desde os estágios iniciais e continuam ao longo do desenvolvimento. Foram anotados 318 CF em todas as amostras. Nos estágios iniciais (10-17 DAF), os ácidos fenólicos predominam (50% da abundância), enquanto a biossíntese de flavonoides se torna prioritária a partir do estágio pastoso. Nesse ponto, as kafirinas se polimerizam por pontes dissulfeto, reduzindo a digestibilidade. Em grãos maduros, a abundância de CF aumenta nove vezes em relação ao estágio 10 DAF. A descorticação reduziu o teor de CF em 45% e alterou seu perfil. Após a germinação, os principais CF, como daidzina e ácidos cafeico, isoferúlico e p-cumárico, diminuíram. Ambos os processos foram eficazes na despolimerização de procianidinas, presentes mesmo em grãos imaturos e nos genótipos sem tanino. O cozimento não alterou o perfil de CF, mas aumentou seu teor em 24%. O efeito genético teve maior impacto no perfil de CF e de kafirinas que o processamento. Todos os processamentos reduziram os níveis de flavonoides, enquanto apenas a germinação aumentou a bioacessibilidade geral. A digestibilidade proteica *in vitro* diminuiu ao longo da maturação e com o cozimento dos grãos, mas a descorticação mitigou esse efeito. Não houve diferença entre genótipos com e sem taninos, sendo a natureza polimérica e hidrofóbica das kafirinas o principal fator limitante da proteólise e digestibilidade. No modelo proteico estudado, flavonoides se ligaram preferencialmente à região a-hélice por interações hidrofóbicas, enquanto os ácidos fenólicos à região β-barril por ligações de hidrogênio e forças de van der Waals. Os complexos proteína-fenólicos mostraram efeito sinérgico quanto ao potencial antioxidante, mas efeito protetor térmico apenas nos CF isolados. Os resultados da tese apresentam informações cruciais sobre os CF e proteínas do sorgo, que podem aumentar seu consumo e expandir seu cultivo.

Palavras-chave: complexação de proteínas; digestibilidade; metabolômica; perfil fenólico, taninos.

ABSTRACT

Sorghum grains (*Sorghum bicolor* L.) are becoming more common in human diets due their agronomic and potential health benefits associated with their phenolic compounds (PC). However, sorghum storage proteins, kafirins, can form protein bodies and interact with PC, reducing its digestibility. Despite significant advances, substantial gaps remain regarding the proteins and PC, as well as the interactions mechanisms between these two components of the grain. The aim of this thesis was to provide deep information about: i) PC and protein accumulation throughout sorghum grain maturation; ii) their changes after biological, physical and thermal processing, including kafirins digestibility and PC bioaccessibility; and iii) the mechanisms of interaction between proteins and sorghum PC. Dutch and French sorghum genotypes were used in this study, contrasting in pericarp colors and tannin levels. Grains were harvested at ten different stages: from 7 to 40 days after flowering (DAF); from cellular division, grain filling, physiological maturity until mature (n = 20). Mature grains underwent three processing methods: germination (n=9), decortication (n=12), and cooking (n=24). PC and proteins of sorghum grains were sequentially extracted and analyzed using high-resolution untargeted metabolomics (UHPLC-ESI-QTOF/MS) and size-exclusion chromatography (SE-HPLC), respectively. Protein-phenolic interactions were elucidated by fluorescence spectroscopy and molecular docking and PC' thermal stability was tested by HPLC-DAD-MS/MS. PC and sorghum proteins are synthesized from the early stages and continue throughout development. A total of 318 PC were annotated in all samples. In the early stages (10-17 DAF), phenolic acids were predominant (50% of abundance), while flavonoid biosynthesis becomes a priority from the dough stage. At this point, kafirins polymerize via disulfide bonds, reducing digestibility. In mature grains, PC abundance increases ninefold compared to the 10 DAF stage. Dehulling reduced PC content by 45% and changed their profile. After germination, daidzin and caffeic, isoferulic, and p-coumaric acids decreased. Both processes were effective in depolymerizing procyanidins, present even in immature grains and in tannin-free genotypes. Cooking did not impact the PC profile but increased their content by 24%. Genetic effects had a greater impact on the PC and kafirin profiles than processing. All processes reduced flavonoid levels, while only germination resulted in an increase in overall bioaccessibility. *In vitro* protein digestibility decreased during maturation and in cooked grains, but decortication mitigated this effect. There was no difference between tannin and non-tannin genotypes, and the polymeric and hydrophobic nature of kafirins was pointed as the main factor limiting proteolysis and digestibility. In the proposed protein model, flavonoids preferentially bound to the a-helix region through hydrophobic interactions, while phenolic acids bound to the β-barrel region through hydrogen bonds and van der Waals forces. Protein-phenolic complexes showed a synergistic effect on antioxidant potential but a protective thermal effect only in isolated PC. The results of this thesis provide crucial information on sorghum PCs and proteins, which can increase its consumption and expand its cultivation.

Keywords: protein complexation; digestibility; metabolomics; phenolic profile, tannins.

List of figures

List of tables

List of abbreviations

AA – amino acid

ANOVA - Analysis of variance

DAF – Days after flowering

DG – decorticated grain

ESI – Electrospray source ionization

HCA - Hierarchical Cluster Analysis

HPLC – High-Performance Liquid Chromatography

LC-MS – Liquid Chromatography Mass Spectrometer

MS - Mass spectrometry

MS/MS - Tandem mass spectrometry

m/z – mass/charge ratio

OPLS-DA - Orthogonal Partial Least Squares-Discriminant Analysis

PB – protein body

PC - Phenolic compounds

PCA - Principal Component Analysis

QC – Quality control

QTOF - Quadrupole Time-of-Flight

ROS - Reactive oxygen species

SDS-PAGE- Sodium dodecyl-sulfate polyacrylamide gel electrophoresis

SE-HPLC – Size Exclusion-High-Performance Liquid Chromatography

UHPLC-MS/MS - Ultra Performance Liquid Chromatograph coupled to tandem Mass Spectrometry

UHPLC-MS^E - Ultra Performance Liquid Chromatograph coupled to Mass Spectrometry based on multiplex data independent acquisition

VIP - Variance Important Projection

Summary

Introduction

Cereals play a crucial role in human nutrition and global food security by providing significant sources of energy, nutrients, and fiber (Stefoska-Needham *et al.*, 2015). They contribute to more than half (56%) of the calories and 50% of the total proteins consumed worldwide (Van Hung, 2016). The consumption of whole grains is widely recommended due to their association with a reduced risk of chronic diseases, attributed to dietary fiber and bioactive molecules, such as phenolic compounds (Tammi *et al.*, 2024). Among cereals, sorghum has been gaining prominence due to its high phenolic content, which is ten times greater than other cereals (Ragaee *et al.*, 2006).

Sorghum (*Sorghum bicolor* (L.) Moench) has long been a staple food in the semiarid regions of Africa and parts of Asia and Central America (Awika, 2017). In these countries, sorghum grains are submitted to different processing methods, ranging from homemade to industrial techniques, to produce sorghum-based products for human consumption (Li *et al.*, 2022). Recently, sorghum's use in the global food supply has increased due to its high-water use efficiency and lower requirements for inputs, such as fertilizers and pesticides, compared to other major cereal grains (Awika, 2017). Additionally, sorghum is heat-tolerant and can remain dormant during climatic stress (Xiong *et al.*, 2019b). This grain cereal could be an alternative crop for the Cerrado Biome and Northeast regions of Brazil, extending its cultivation until the north of the country (Vilarinho *et al.*, 2007).

Although sorghum has several cultivation and biological advantages over other cereals, its proteins present technological and nutritional challenges. Being a gluten-free cereal, sorghum flour lacks rheological properties such as viscoelasticity, needing the addition of other components (starch, gums, and hydrocolloids) to improve texture (Adiamo *et al.*, 2018). Additionally, the mainly proteins in sorghum, known as kafirins, have low digestibility, due to intrinsic factors, such as covalent bonds that form protein aggregates, and external factors, such as protein-phenolics interaction (Duodu *et al.*, 2003). The content and profile of kafirins and phenolic compounds, as well as their interaction, are influenced by genotype, cultivation and harvest conditions, but also grain processing (Dykes & Rooney, 2006). However, there is limited information on the synthesis of these compounds throughout grain development and their influence on sorghum digestibility.

In this context, the central question of this thesis was *"How to value sorghum in human nutrition, providing strategies to enrich its bioactive compounds and improve its protein quality?"* and the main hypothesis was that sorghum protein digestibility is more influenced by the formation of protein-protein aggregates than by protein-phenolic complexes. Hence, different factors should be evaluated since cultivation local, growing season, sorghum genotypes and races (*i.e.* tannin presence), maturity stage, and processing can influence the sorghum functionality and impact nutritional properties, such as protein digestibility.

In this way, different statements and experimental designs were proposed to answer the main question and the starting point hypothesis:

i) proteins and phenolics are synthesized during sorghum grain maturation, exhibiting variations according to genotypes and affecting the digestibility;
ii) technological processing can enhance the nutritional quality of sorghum and its bioactive potential;
iii) phenolic bioaccessibility and protein digestibility are influenced by genetic factors, agricultural practices, and processing methods;
iv) phenolic compounds and proteins interact to form complexes that can present a thermo-protective effect.

Based on this context, the main objectives of this thesis can be outlined as follows:

(i) To characterize, in a comprehensive way, the sorghum phenolic compounds by untargeted metabolomic approaches, considering grain development stages, different genotypes and location diversity, grain milling fractions, and different processing methods;
(ii) To deep understand how grain development and processing impact sorghum proteins assembly, focusing on their synthesis, complexation, solubility, and digestibility;
(iii) To evaluate factors that impact sorghum value for human health, such as antioxidant potential, phenolic bioaccessibility, and protein digestibility;
(iv) To elucidate the formation of sorghum protein-phenolic complexes by identifying the preferred binding sites and the primary forces involved in a simple model.

This PhD thesis was carried out at the Graduate Program in Food and Nutrition (Programa de Pós-Graduação em Alimentos e Nutrição – PPGAN) within the research lines "Processing, quality, valorization of food, coproducts and residues" and "Functional foods" which aim to address critical issues in food science and technology, promoting sustainable practices and enhancing public health. From these lines of research, innovative processing methods and new functional foods can be developed based on the use of robust and modern analytical tools. This study contemplates 4 of the 17 Sustainable Development Goals instituted by the UN to protect the planet: #2 (zero hunger), #3 (good health and well-being), #11 (Sustainable cities and communities) and #12 (Partnerships for the goals), since this project count with research partnership from Brazil, France, the Netherlands, and Canada to enhance knowledge and promote the better utilization of sorghum grains.

The results obtained during this thesis are structured and presented in five scientific papers, one of them are already published in a peer reviewed scientific journal, one is under submission, and the last three are in preparation for submission. Additionally, the first chapter covers a literature review, while the final chapter offers an overall discussion of the current work, summarizes the main conclusions, and provides perspectives for future research. The manuscript is organized into five major parts encompassing seven chapters (organizational scheme below):

- Part I of this thesis is entitled "**Sorghum: a star crop for combating adverse weather conditions and global food insecurity**". This initial part of the thesis is composed of a unique chapter (**Chapter 1**) which presents a comprehensive literature review, highlighting the agronomic and bioactive significance of sorghum grain. It discusses how current world circumstances offer a unique opportunity to introduce underutilized technologies and crops to tackle the pressing issue of food insecurity and hunger. Additionally, the review delves into the structural and chemical composition of sorghum grain, its primary limitations for human consumption, and explores various processing and analytical techniques that can enhance and ensure its potential as a promising food source.
- Part II is entitled "**From flowering to grain maturity: a comprehensive study on grain development in sorghum**" brings an innovative and temporal approach to follow sorghum grain development. Two chapters are presented in this part: **Chapter 2** focus on sorghum grain filling and protein complexation using classical

biochemical methods, reflecting on its final digestibility; and **Chapter 3** aims to elucidate PC biosynthesis during grain growth, studying the effect of genotypes and growing season on the PC profile by metabolomic tools.

- Part III is entitled "**Effect of traditional processing methods on the protein and bioactive composition of sorghum**". Two chapters are presented in this part: **Chapter 4** shows the impacts of fractionation (grain dehulling) and heat treatment (grain boiling) on the protein and phenolic profile, including the physicochemical characterization, *in vitro* protein digestibility, bioaccessibility and metabolomic profile of phenolic compounds; **Chapter 5** aims to understand the impact of germination and seedling growth on the phenolic profile of sorghum grains genotypes, highlighting the compounds bioaccessibility in each gastrointestinal tract phase (oral, gastric and intestinal).
- Part IV is entitled "**Interactions between protein and sorghum phenolic compounds: an unprecedented study**" and is composed by a single chapter (**Chapter 6**). This part of the thesis explores the mechanisms and forces of interactions between a model protein (β-lg) and phenolic compounds extracted from mature sorghum bran of different genotypes. This chapter also demonstrates the protective effect of this complex against the thermal degradation of phenolic compounds. As the title suggests, this is an unprecedented study for sorghum grains extracts and may significantly contribute to their industrial application.
- Part V is dedicated to a general discussion (**Chapter 7**) encompassing the main results found in this thesis, the concluding remarks, and some proposed prospects for future works.

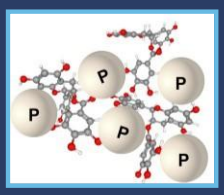
Thesis design

Phenolic compounds

Pericarp (4%)

Endosperm (86%)

Proteins (kafirins)

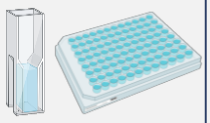
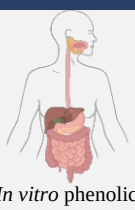
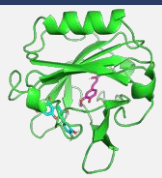


Impact factors on the nutritional variability of sorghum

- Genotypes or cultivars
- Environmental factors
- Maturation stages
- Technological processing

Main question: how to value sorghum in human nutrition, providing strategies to enrich its bioactive compounds and improve its protein quality?

Key methods

Spectrophotometric and spectroscopic analysis	HPLC methods	*In vitro* simulated gastrointestinal digestion	Metabolomic tools and chemometric analyses	Molecular docking
				
Antioxidant methods and fluorescence quenching of protein-phenolic complexes	Amino acid and protein composition	*In vitro* phenolic bioaccessibility and protein digestibility	Tentative identification, relative quantification and phenolic profile of sorghum	Mechanisms and forces of interaction between phenolics and proteins

Thesis organization

PART I	PART II		PART III		PART IV	PART V
Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6	Chapter 7
Literature review	"Digestibility, biochemical changes and kafirin polymerization analysis during sorghum grain development"	"Dynamic metabolomic changes in the phenolic compounds profile and antioxidant activity in developmental sorghum grains"	"Enhancing the nutritional value and bioavailability of sorghum grains bred for northern Europe through processing: A perspective on phenolic and protein digestibility"	"Tracking the changes and bioaccessibility of phenolic compounds of sorghum grains (*Sorghum bicolor* (L.) Moench) upon germination and seedling growth by UHPLC-QTOF-MS/MS"	"β-lactoglobulin and sorghum phenolic compounds molecular binding: interaction mechanism and thermal stability impact"	General discussion, conclusions and perspectives

Target result: innovative processing techniques and/or the consumption of grains at earlier developmental stages may enhance biological quality of sorghum for human consumption and geographical diffusion.

Part I

Sorghum: a star crop for combating adverse weather conditions and global food insecurity

Chapter 1 – Literature review

1. Key challenges in global food security

According to Food and Agriculture Organization (FAO, 2008), "food security" refers to the availability of sufficient, safe, and nutritious food to meet the dietary needs and preferences of all individuals for an active and healthy life. Currently, about 30% of the global population – 2.4 billion people – were moderately or severely food insecure, meaning they did not have access to adequate food (FAO, 2023). Projections show that the world is not on track to achieve the Sustainable Development Goal of zero hunger by 2030 (FAO *et al.*, 2023). This is due to the contemporary world encountering a series of challenges that directly threaten food security, including: (1) climate changes, which can disrupt agricultural production through extreme weather events, altered precipitation patterns, and increased incidence of pests and diseases; (2) disasters and geopolitical conflicts, which leads to fluctuating food prices and reduced access to food for vulnerable populations; (3) productivity stagnation of some agri-food crops, due to the almost total occupation of cultivation areas and maximum yield; and (4) the estimated population growth, increasing food demand.

Climate changes

Climate change is interconnected to agriculture, serving as it is the main cause of both biotic and abiotic stresses that negatively impact agricultural production (Neupane *et al.*, 2022). These environmental variations influence annual precipitation, average temperature, heat waves, concentrations of CO_2 and ozone, pest and microorganism activity, and wind composition, as well as the frequency and intensity of natural disasters such as landslides, floods, and droughts (Neupane *et al.*, 2022). With global temperatures projected to rise by 2 °C at the end of this decade, significant crops including wheat, corn, and soybeans are projected to reduce its average productivity up to 50% (Muluneh, 2021; Neupane *et al.*, 2022). In future scenarios, without ongoing agricultural innovation, over half of the global population could face the risk of undernourishment by 2050 (Dawson *et al.*, 2016).

Disasters and geopolitical conflicts

The COVID-19 pandemic has had a profound impact on lives and livelihoods of global population, resulting in an economic recession that ended three decades of global progress in poverty reduction. In a single year (2019–2020), estimated increase of close to 90 million people faced hunger due to this biological disaster (FAO *et al.*, 2023). Although 2021 marked a partial recovery from the pandemic's damage, its residual effects led to an additional increase of approximately 38 million people facing hunger. (FAO *et al.*, 2023) In 2022, just as the weight of the pandemic was beginning to lift, the outbreak of war in Ukraine, involving two of the world's largest agricultural commodity producers, increasing food prices and reducing access to food for vulnerable populations (Abay *et al.*, 2023). Among commodities, cereals were drastically impacted, as Russia and Ukraine together account for about 34, 27 and 17 percent of the global wheat, barley and maize trade, respectively (Abay *et al.*, 2023).

Stagnation in crops yields

Recent data has suggested that yields for many crucial crops may be stagnating in some regions around the world (Neupane *et al.*, 2022; Ray *et al.*, 2012). Specifically, concerns have been raised regarding potential stagnation or decline in yields of three key global crops (maize, rice, and wheat) which together account for approximately 57% of the world's food energy input (Ray *et al.*, 2012). For example, rice (35%) and wheat (37%) have substantial areas that are now witnessing yield stagnation, while maize (26%) has less area in yield stagnation (Ray *et al.*, 2012). This implies that, at least for wheat and rice, yield stagnation could significantly impact agriculture's capacity to meet the increasing global demand for these essential commodities.

Global population growth

Finally, the substantial increase in global population for the coming years further threatens food security. According to the United Nations, it is estimated that by 2050 the global population will range between 9.4 and 10.1 billion people (UN-DESA, 2022). Over half of this predicted growth is expected to occur in sub-Saharan Africa, a region comprised predominantly of low-income countries. Consequently, these nations will probably face limited resources and access to technology to sustainably produce sufficient food for their expanding populations (Molotoks *et al.*, 2021).

Cultivation of alternative crops

Given the challenges mentioned above, innovations in agriculture are essential, innovations in agriculture are necessary, such as encouraging the production and consumption of non-conventional cereals. Sorghum is a versatile cereal capable of adapting to diverse environments, particularly under conditions of water deficit. This adaptability makes it highly valuable in regions with irregular rainfall and elevated temperatures. Consequently, sorghum cultivation can be expanded to areas prone to water deficit, as the crop demonstrates remarkable potential to grow in hostile environments with minimal inputs and care. With a nutritional composition comparable to traditional cereals, incorporating sorghum grain into human nutrition represents a viable strategy for ensuring global food security (Hossain *et al.*, 2022).

2. Sorghum: evolutionary origins and contemporary importance

Sorghum, a cereal from *Poaceae* family and *Andropogonae* tribe, is characterized by its diploid nature, comprising 10 chromosomes (2n = 20), and a relatively small genome (~ 730 megabase pairs) when compared to other cereals (Paterson *et al.*, 2009). There are four major types of sorghum namely grain, forage, biomass and sweet sorghum. Forage and biomass sorghum are mainly used for livestock and biomass industries, respectively. Sweet sorghum can be used as alternative to table sugar and mainly used in brewery in the production of alcoholic beverages. Grain sorghum holds the highest economic value and can be further subdivided into five morphologically distinct races: bicolor, kafir, caudatum, durra, and guinea (Lazarides *et al.*, 1991). Additionally, there are ten stabilized hybrid races produced through genetic crosses between these five races.

Sorghum cultivation has a long historical trajectory. This grain is native to Africa, with evidence suggesting its domestication occurred approximately 5,000 years ago in Northeast Africa (Ananda *et al.*, 2020). In the Americas, the introduction of sorghum is relatively recent, between the 17th and 18th centuries through the transatlantic slave trade (Venkateswaran *et al.*, 2019). Currently, sorghum is cultivated globally and ranks as the fifth most produced crop, following maize, wheat, rice, and barley. Its cultivation significantly contributes to the global economy, with 43.7 million hectares dedicated to sorghum production, yielding 58 million tons annually, and generating a turnover of 1.6 billion U.S. dollars per year (USDA, 2024b).

From an agronomic perspective, sorghum morphological and genetic characteristics confer resistance to specific abiotic stresses, including drought, salinity, and high temperatures (Hossain *et al.*, 2022). These attributes are particularly significant given the global trend towards increasing temperatures due to climate change, coupled with the stagnation of large crop yields, limited potential for expanding cultivated areas, and population growth (Lotze-Campen, 2011). To sum up, sorghum is an adaptable cereal with the potential to support food security in the face of these challenges.

From a nutritional standpoint, sorghum is a carbohydrate-rich crop and a key staple food for millions of residents in semi-arid tropical regions (George *et al.*, 2022). The nutritional composition of sorghum varies by genotype, but generally, it comprises approximately 75% carbohydrates (including starch and other polysaccharides), 15% protein, 8% fiber, and 3% lipids (Xiong *et al.*, 2019b). Given its high energy input, sorghum is a dietary staple for more than 30 countries. In these regions, sorghum is consumed as whole grain or processed into flour for making breads, porridges, and other foods, either alone or mixed with other flours (Hossain *et al.*, 2022). Due to their high content of resistant starch, sorghum grains are slowly digested, making them suitable for diabetic patients as they result in a lower glycemic response (Taylor & Duodu, 2023). Furthermore, sorghum is a gluten-free cereal, making it a promising and safe food source for individuals with celiac or allergic diseases. The rising demand for gluten-free foods in Western and developing countries underscores sorghum's potential as an excellent food ingredient and nutrient source (Xiong *et al.*, 2019b).

However, the main current interest in incorporating sorghum grains into the global human-diet is due their bioactive compounds content, particularly phenolic compounds (PC). Sorghum grains have been identified as potential sources of PC (Kang *et al.*, 2016), and have been recognized for their potential in preventing and modulating chronic diseases (Arbex *et al.*, 2018; Espitia-Hernández *et al.*, 2022a). The efficacy of these compounds is attributed to their antioxidant properties, which neutralizes free radicals and prevent oxidative stress (Xiong *et al.*, 2019b). Beyond direct antioxidant effects, sorghum's phenolic compounds also appear to induce the activity of endogenous detoxifying enzymes (phase II enzymes). These enzymes convert reactive oxygen and nitrogen species into non-toxic compounds, thereby indirectly enhancing the body's defense against oxidative damage (Awika *et al.*, 2009).

To fully harness the nutritional and bioactive benefits of sorghum, it is crucial to understand its developmental stages, from flowering to final maturity. These stages are vital for optimizing sorghum quality, maximizing its yield and elucidating the complete synthesis of these phenolic compounds in the grain. Detailed knowledge of sorghum's growth enables allows for targeted agronomic practices that can improve both crop quantity and quality. This, in turn, supports an advancement of sorghum as a functional food and ensuring its continued contribution to food security and public health.

3. Sorghum plant and grain development

Sorghum plant undergoes ten differents developmental stages, from planting to maturity, which can be categorized into three groups: G1 - vegetative state (S0-S3), G2 - reproductive state (S4-S6), and G3 - grain filling (S7-S9) (Figure 1) (Gerik *et al.*, 2003).

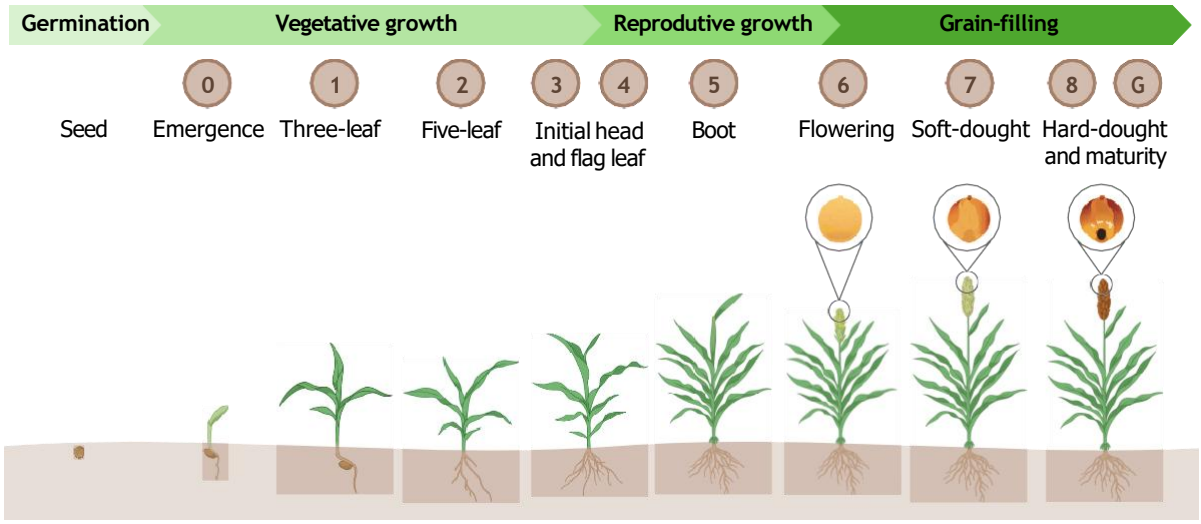


Figure 1. Development stages of sorghum crop (Gerik *et al.*, 2003).

During the G1 vegetative state, the plant develops its structures, such as leaves and tillers, which are crucial for supporting grain formation and growth (Gerik *et al.*, 2003). The duration of this stage is temperature-dependent. The G2 state encompasses from the panicle emergence until flowering (anthesis). It is visually marked by the appearance of protrusions on the plant's growing point (Roozeboom & Prasad, 2019). This stage begins approximately 30 days post-emergence and is critical for final grain yield, as any disturbances during this period can reduce seed formation (Gerik *et al.*, 2003; Roozeboom & Prasad, 2019).

The final G3 grain-filling state is crucial for grain sorghum and initiates after complete emergence of the panicle. The anthers, which will encase the developing grains,

first appear at the top of the panicle and subsequently progress downward as development continues, *i.e.*, grains at the base of the panicle are the last to reach each developmental stage (Roozeboom & Prasad, 2019). Post-flowering, nutrients produced by leaves and roots (*e.g.*, sugars, amino acids, proteins) are rapidly transported to the grain, reflecting on its dry-mass accumulation through its conversion to starch and proteins (Gerik *et al.*, 2003). Sorghum grain-filling can be divided into four phases: milky, soft dough, hard dough and physiological maturity.

Milky stage happens 7 to 10 days after flowering (DAF). During this period, the grain undergoes rapid expansion, with a high water content (above 70%) (Gerik *et al.*, 2003). The grain at this stage is characterized by its softness and the presence of an internal milky fluid. Rapidly, sorghum grains begin to accumulate dry-mass, reaching 50% of its final weight during the **soft-dough stage**. This phase occurs 15 to 25 DAF and the plant is quite susceptible to bird feeding at this time. The **hard-dough stage** occurs when the grain has accumulated approximately 75% of its dry matter and cannot be compressed between the fingers. Finally, at **physiological maturity**, the grain attains its maximum dry weight, visually indicated by the appearance of a black-layer above the point of kernel attachment in the floret near the kernel base. Mature sorghums still have relatively high water contents (30%) and therefore require post-harvest drying to ensure microbiological safety during storage (Gerik *et al.*, 2003).

Besides morphological changes, physicochemical properties are also altered as the grain growth. Ranathunga & Suwannaporn (2022) reported that molecular weight distribution and degree of polymerization of the two main cereal components (starch and protein) are modified according to maturation. Immature grains showed lower gelatinization temperatures and enthalpy, attributed to a reduced proportion of long-chain amylopectin; while their proteins are less polymerized, potentially enhancing their digestibility (Ranathunga & Suwannaporn, 2022). These changes significantly influence the technological and nutritional properties of the grain. Furthermore, immature cereals contain higher levels of fibers, vitamins, and bioactive compounds, mainly phenolic compounds, compared to mature grains (Kim & Kim, 2017; Ranathunga & Suwannaporn, 2022). Although these nutritional and bioactive properties, coupled with the shorter cultivation period required, have aroused the interest of consuming cereals before their final maturation (Alkay *et al.*, 2022; Çetin-Babaoğlu *et al.*, 2020; Kim & Kim, 2017), to

our knowledge there is no detailed study on these changes throughout sorghum grain development.

4. Structure and composition of mature sorghum grains

Grain physiochemistry and structure are important factors of the end-use quality of cereals, essential for their applications in food, feed, and fuel. Mature sorghum grain is typically round, spherical, or elliptical, with a diameter ranging from 4 to 8 mm and a weight between 3 to 80 mg, depending on the genotype (Awika *et al.*, 2005; Waniska & Rooney, 2000). The grain is composed of three main components: the tegument (pericarp, 3-6% of the grain), the germ (embryo, 5-10% of the grain), and the endosperm (reserve tissue, 84-90% of the grain). The seed coat, known as the testa, is located between the pericarp and the endosperm (Earp *et al.*, 2004) (Figure 2).

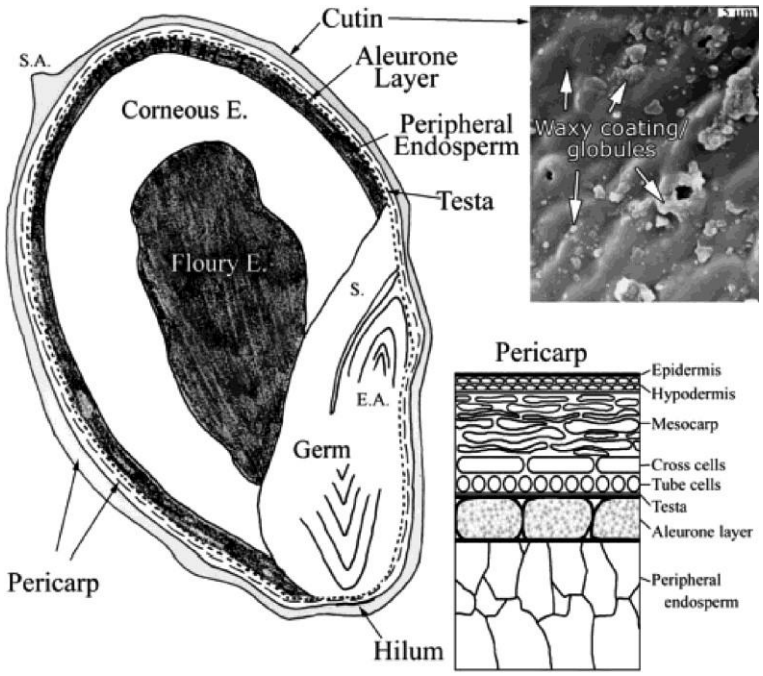


Figure 2. Sorghum kernel structure (Earp *et al.*, 2004).

The testa is a layer of thick-walled cells that may present pigmentation and the presence of condensed tannins, characteristics that can influence the nutritional and sensory qualities of food products derived from the grain. The presence of tannins is associated with increased resistance of the grains to fungi, pathogens, and bird attacks. It is important to note that grain color is not an indicator of tannin presence; grains may be white or light-colored and still contain tannins. The presence of the pigmented testa is regulated by the B1 and B2 genes, which must both be dominant for pigmentation to

occur. Additionally, these genes, together with the R and Y genes, play a crucial role in determining the color of the pericarp, resulting in variations between white, black, yellow, and red sorghum grains (Earp *et al.*, 2004).

The endosperm is the storage tissue and constitutes most of the sorghum grain. It comprises the aleurone layer, as well as the peripheral, corneous, and floury zones. Positioned just beneath the pericarp, the aleurone layer consists of cells with thick walls. This layer contains minerals, hydrolytic enzymes (mainly synthesized upon grain imbibition), and lipids (Waniska & Rooney, 2000). The corneous and floury zones of endosperm are composed of starch granules, a protein matrix, protein bodies, and cell walls rich in glucuronoarabinoxylans. In the peripheral and corneous regions of the endosperm, starch granules and protein bodies are trapped in a protein matrix. The floury endosperm, located in the core of the grain, has a significantly lower abundance of protein bodies compared to the corneous and peripheral regions (Waniska & Rooney, 2000). The ratio of floury to corneous endosperm can influence grain milling yield and digestibility (Khoddami *et al.*, 2023).

The germ is divided into two major parts: the embryonic axis and scutellum. The embryonic axis forms the new plant; while the scutellum is the single cotyledon of the sorghum seed whose cells are responsible for storing nutrients. Germ is firmly adherent to the grain and does not easily separate during the milling process (Waniska & Rooney, 2000).

The pericarp, the outermost part of the grain, is divided into three layers: the epicarp, mesocarp, and endocarp. The epicarp consists of rectangular cells, typically containing pigments, and may be covered by a thin wax layer. The mesocarp is characterized by the presence of small starch granules within its cells, a feature unique to sorghum among cereals. The endocarp, the innermost sublayer of the pericarp, comprises crossed cells and a layer of tubular cells responsible for transporting water into the kernel (Earp *et al.*, 2004).

The pericarp thickness is a crucial factor in grain processing. A thick pericarp combined with a corneous endosperm is necessary to produce a high yield of sorghum flour. A hard and thick pericarp is easier to remove through pounding, while a thin pericarp takes more time to remove, mainly achieved through dehulling (Guindo *et al.*, 2016). However, sorghum varieties with a thick pericarp have certain disadvantages, such

as susceptibility to mold, grain weathering in the field during maturation, and rapid deterioration of grains during storage (Earp *et al.*, 2004; Guindo *et al.*, 2016). Overall, the pericarp is composed of non-starch polysaccharides, carotenoids, and phenolic compounds such as 3-deoxyanthocyanidins, tannins, and phenolic acids (Awika *et al.*, 2005; Earp *et al.*, 2004).

5. Phenolic compounds of sorghum grains

Phenolic compounds, belonging to secondary metabolites, are well-known to naturally bio-synthesize in plants. They are a large group of heterogeneous compounds characterized by their basic structure consisting of a benzene ring with hydroxyl constituents and range from simple phenolic acids to highly polymerized tannins (Li *et al.*, 2021). Studies have highlighted the health benefits promoted by phenolic compounds, mainly due to the high antioxidant capacity against the overproduction of reactive oxygen species (ROS) (Xu *et al.*, 2021; Zhang *et al.*, 2015). The functional properties of phenolic compounds classify them as a bioactive compound, positively influencing human health and reducing the risk of chronic diseases, including inflammation, diabetes, atherosclerosis, and cardiovascular disease (Zhang *et al.*, 2015).

Sorghum is a rich source of phenolic compounds, mainly comprising phenolic acids, flavonoids, and condensed tannins. These compounds are synthesized via the phenylpropanoid pathways, and their quantitative and qualitative profiles are influenced by the genotype and environmental cultivation conditions (Xu *et al.*, 2021). The extraction method is another critical factor in determining the phenolic content of sorghum (Li *et al.*, 2021). Typically, sorghum phenolics are extracted using ethanol, methanol, or acetone solvents. However, the solvent selection must be meticulously considered, as different extractors can present different results from the same study. Additionally, the application of acids (*e.g.*, HCl) and bases (*e.g.*, NaOH) during extraction can facilitate the release of bound phenolic compounds, promoting the hydrolysis of flavonoid glycosides and esterified phenolics, respectively (Shelembe *et al.*, 2014).

In the phenolic profile of sorghum, flavonoids constitute the majority class (D'Almeida *et al.*, 2021b). These compounds are structurally characterized by a basic skeleton consisting of 15 carbon atoms in a C6-C3-C6 configuration. Dominant flavonoids in sorghum include anthocyanins, flavonols, and flavanones (Xiong *et al.*, 2019b). Anthocyanins account for 79% of the flavonoid content and are among the most

important water-soluble pigments in plants. This subclass can exist as glycosides (anthocyanins) or aglycones (anthocyanidins). In sorghum grains, the main anthocyanins identified are 3-deoxyanthocyanidins (*e.g.*, luteolinidin, apigeninidin, 5-methoxyluteolinidin, and 7-methoxy apigeninidin) (Ofosu *et al.*, 2021). Among flavonols in sorghum grains, kaempferol and quercetin are the most investigated, and naringenin is the most well-known compound in flavanones. The flavanol catechin is also well reported in sorghum grains (Li *et al.*, 2021).

Tannins are the most extensively studied phenolic compound in sorghum due to their presence in a condensed form (proanthocyanidins) with high molecular weight and significant polymerization, which are uncommon among major cereals (Wu *et al.*, 2012). These compounds influence the formation of resistant starch in sorghum grains, thereby reducing postprandial blood glucose levels, lowering the caloric load of starchy foods, and enhancing satiety after consumption (Taylor & Duodu, 2023). Tannins can significantly modify the rheological properties of gluten in blended flours, enhancing gluten strength (Girard *et al.*, 2016). However, one of the most recognized properties of tannins is their potent antioxidant activity; they are 15-30 times more effective at scavenging free radicals, making them crucial for promoting health (Hagerman *et al.*, 1998).

Sorghum grain also contains substantial levels of phenolic acids, with hydroxybenzoic and hydroxycinnamic acids being the predominant subclasses (Xiong *et al.*, 2019b). These acids possess free radical scavenging ability due to their polarity and hydroxyl groups. Notable phenolic acids extensively studied in sorghum include caffeic acid, p-coumaric acid, sinapic acid, gallic acid, protocatechuic acid, p-hydroxybenzoic acid, and ferulic acid, with ferulic acid being the most prevalent (Li *et al.*, 2021). Phenolic acids are typically bound to the cell wall (lignin) via covalent bonds, and requires acid or alkaline hydrolysis at high temperatures for their extraction (Wu *et al.*, 2017). These bound phenolic acids are not hydrolyzed by human digestive enzymes, which decreases their bioaccessibility. However, they can be fermented by the colon microbiota (Hole *et al.*, 2012).

6. Phenolic compound bioaccessibility in gastrointestinal tract

To understand the physiological response to phenolic compounds of foods, it is necessary to follow the digestive processes within the human digestive tract. To this end,

static *in vitro* methods have been employed as a suitable, cost-effective, and simple alternative to simulate gastrointestinal digestion and evaluate the bioaccessibility of specific compounds from the food matrix (Rasera *et al.*, 2023). Currently, the "INFOGEST static *in vitro* simulation of gastrointestinal digestion" serves as the standard protocol for digesting food matrices (Brodkorb *et al.*, 2019; Rasera *et al.*, 2023). In this protocol, food samples undergo sequential oral, gastric, and intestinal digestion; the large intestine is not considered as compound absorption mainly occurs in the small intestine (Brodkorb *et al.*, 2019). The use of digestive fluids (salivary, gastric, and intestinal), enzymes, and optimal conditions (ratios of enzyme to substrate, pH, temperature, and duration of each digestive phase) enhances the method's accuracy, closely reflecting the processes occurring within the human organism.

The oral phase of *in vitro* digestion is controversial among researchers. Many argue that this stage does not significantly affect phenolic bioaccessibility, due to the short action time of amylase (Rasera *et al.*, 2023). However, Brodkorb *et al.* (2019) assert that the oral phase should be included in all simulated digestion procedures to ensure dilution consistency and the formation of a swallowable bolus. Furthermore, it is important to note that this stage is influenced by factors such as pH, electrolytes, temperature, darkness, reduced O_2 content, and enzyme activation, which can lead to hydrolysis or degradation reactions. For exemple, anthocyanins can be oxidized by O_2, deprotonated at pH 7, and spontaneously cleaved during this phase of gastrointestinal digestion (Sánchez-Velázquez *et al.*, 2021). Dou *et al.* (2019) also observed an important role of the oral phase in the release of phenolic acids, especially for caffeoylquinic acid, 3-caffeoylquinic acid and catechin.

In the gastric phase, the pH is reduced (pH 2.0) and pepsin is introduced. This phase is characterized by intense phenolic release. Pepsin hydrolyzes peptide chains, liberating phenolic acids, flavonoid-diglycosidic compounds, and monoglycosides (Guo *et al.*, 2017). The acidic pH also facilitates the conversion of flavonoid glucosides to aglycones, thereby reducing flavonoid glucoside bioaccessibility by the end of the gastric phase (Tomé-Sánchez *et al.*, 2021). de Morais *et al.* (2020b) documented the conversion of procyanidins and epicatechin gallate to epicatechin, as well as the release of gallic acid from galloylated catechins. This underscores the role of pH in this stage, where glycosidic bonds are broken, forming compounds that are more stable under acidic conditions.

The intestinal phase is particularly critical for phenolic compounds, as it can lead to their degradation, precipitation, and increased instability. During this phase digestive enzymes and pH can interfere with the hydroxyl groups of phenolics, triggering oxidation and polymerization reactions, and resulting in the formation of high molecular weight phenolic derivatives with low solubility (Rasera *et al.*, 2023). These reactions vary depending on the phenolic class, with anthocyanins being notably affected, leading to significantly reduced bioaccessibility in the intestinal phase (Nagar *et al.*, 2021). Changes in the chemical structure of phenolic compounds induced by intestinal enzymes are not only related to bioaccessibility; these interactions with hydroxyl groups also modify the antioxidant potential and solubility of the phenolics (Neto *et al.*, 2017).

7. Sorghum proteins and its digestibility

Sorghum grain proteins are traditionally classified based on their solubility in: albumins (water soluble), globulins (salt soluble), kafirins (prolamins, alcohol soluble), cross-linked kafirins (alcohol + reducing agents soluble), cross-linked glutelins (alkali soluble), and insoluble protein residues (non-extractable) (Jambunathan *et al.*, 1975a). A simplified classification divides these proteins into kafirins and non-kafirins, with kafirins being the most abundant (77-82%) (Belton *et al.*, 2006).

Kafirins are further classified based on molecular weight, solubility, structure, and amino acid composition into three main types: a-kafirins (66-84%), β-kafirins (8-13%), γ-kafirins (9-21%) and δ-kafirins (1%). γ-Kafirins have the highest molecular weight (28 kDa), followed by a-kafirins (22 and 26 kDa), β-kafirins (18 kDa) and δ-kafirins (13 kDa) (Belton *et al.*, 2006). These proteins are synthesized and translocated into the endoplasmic reticulum lumen, where they form protein bodies. These bodies, 0.4-2.0 μm in diameter, consist of β and γ-kafirins in the shell (peripheral protein bodies) and a-kafirins in the interior. In mature grain endosperm, protein bodies form a matrix with starch granules, contributing to sorghum's hardness, digestibility, and processing quality (Labuschagne, 2018).

The secondary structure of kafirin also has been elucidated using Fourier transform infrared (FTIR) spectroscopy (Shah *et al.*, 2024). In their native state, kafirin polypeptides are highly folded, with a-helices comprising 40-60% of the secondary structure, β-sheets approximately 27%, and disordered structures, including random coils, around 24% . There are limited studies on the secondary structures of individual kafirin

subunits (Shah *et al.*, 2024). FTIR analysis reveals that a-kafirin subunits contain a higher proportion of a-helices compared to other subunits, whereas γ-kafirin exhibits more random coils and β-sheet structures than a-kafirin subunits (Belton *et al.*, 2006).

The a-helical structure in kafirin significantly contributes to its hydrophobicity, as the hydrophilic amino acid side chains are closely packed and buried within the helices, while the hydrophobic aliphatic amino acid side chains point outward . Conversely, the β-sheet secondary structure is more open and unfolded, with hydrophilic amino acid residues more exposed on the exterior of the structure (Da Silva *et al.*, 2011). A comprehensive understanding of primary and secundary kafirins can facilitate the development of new methodologies to optimize the functional properties related to sorghum proteins.

a-kafirins are rich in nonpolar amino acids (proline, leucine, and alanine) and primarily form intramolecular disulfide bonds. In contrast, β and γ-kafirins, which are rich in cysteine, form both intra- and intermolecular disulfide bonds. This extensive polymerization of β and γ-kafirins enhances their resistance to proteases and protects the internal a-kafirins from proteolysis (Labuschagne, 2018). This structural conformation significantly affects the digestibility of sorghum proteins (Wu *et al.*, 2013). Additionally, sorghum prolamins have a higher proportion of cross-linked fractions and greater hydrophobicity, which explains their increased tendency to form protein aggregates compared to prolamins from other cereals (Belton *et al.*, 2006).

The percent digestibility of whole grain sorghum (60%) is lower compared to other cereals, such as wheat (80%) (Duodu *et al.*, 2002). Sorghum proteins are poorly digestible when wet cooked (35%), due to the resistant starch formation after cooling. This starch can complex with kafirins, rendering them less susceptible to enzymatic attack by peptidases (Duodu *et al.*, 2002). The lower digestibility of sorghum grains is reflected in sorghum-based products, where the nutritional availability of proteins is reduced compared to other cereals. This characteristic is a significant limiting factor for the application of sorghum in human nutrition, as it affects the overall protein intake and nutrient absorption from sorghum-based foods.

8. Protein–phenolic compound interactions

Besides the inherent nature of sorghum proteins and their organization within the grain, the interaction of proteins with non-protein components, including phenolic compounds, also contributes to the issue of sorghum digestibility (Belton *et al.*, 2006). Phenolic compounds can interact with dietary proteins through two different mechanisms: non-covalent and covalent interactions.

Non-covalent interactions are reversible and typically involve four types of binding forces: hydrogen bonding, hydrophobic interactions, electrostatic interactions, and van der Waals forces. Phenolic groups are well-known hydrogen donors and can form hydrogen bonds with the carboxyl groups of proteins. These hydrogen bonds also form through interactions between -OH groups of phenolics and amino groups of proteins. Hydrophobic interactions mainly occur due to associations between the non-polar aromatic rings of phenolics and the hydrophobic sites of protein molecules (Zhang *et al.*, 2021). Electrostatic interactions arise from the charged groups on the protein surface and the -OH groups of phenolics, while van der Waals forces involve interactions between atoms influenced by the surrounding solvents (Ozdal *et al.*, 2013).

For covalent interactions, literature suggests five mechanisms steps. Specifically, **(1)** the diphenol portion of a polyphenol can be oxidized to an orthoquinone, either enzymatically, as occurs in plant tissues, or via molecular oxygen. **(2)** This orthoquinone can form a dimer in a side reaction or **(3)** react with the amino or thiol side chains of polypeptides to form covalent bonds with the phenolic ring, regenerating hydroquinone. **(4)** The hydroquinone can be reoxidized and bind to a second polypeptide, resulting in a covalent cross-links between proteins and phenolic compounds. **(5)** Alternatively, two orthoquinones can dimerize to form cross-links (Strauss & Gibson, 2004).

Physicochemical properties of proteins, such as solubility or thermal stability are strongly affected by their interaction with different phenolic compounds (Czubinski & Dwiecki, 2017). However, the most important factor affected by the formation of protein-phenolic compound complexes is the decrease in protein digestibility (Duodu *et al.*, 2003). Tannins are the main phenolic class responsible for reducing sorghum digestibility, mainly through the formation of protein complexes via -OH groups and precipitation due to these large compound size (Duodu et al., 2003). However, it may not be the precipitation itself that reduces digestibility, but rather a potential change in protein conformation and steric effects that inhibit enzyme accessibility (Duodu *et al.*, 2003).

Although not widely reported in the literature, phenolic acids and flavonoids also interact and form complexes with proteins (Duodu *et al.*, 2003).

Although protein-phenolic complexes formation negatively impacts protein digestibility, the thermal denaturation temperatures of these complexes were found to be higher than those of the isolated phenolic compounds (Liu *et al.*, 2017). This finding is technologically significant for the development of cereal-based products, as most industrial processing involves the application of thermal methods. To better elucidate the strength and types of interactions involved in complexes, several studies have employed a model protein (β-lg) and isolates phenolic compounds (Chen *et al.*, 2022a; Dai *et al.*, 2022; Liu *et al.*, 2022; Qie *et al.*, 2021; Zhang *et al.*, 2022). Globally, the β-lg-phenolic complexes are formed through non-covalent bonds, but the interaction strength, reaction type, and thermal protective effect vary according to the structure and phenolic class. While these findings are promising for understanding the interaction between kafirins and phenolics, to our knowledge there are no studies specifically investigating these interactions using proteins and/or phenolic extracts from sorghum grain.

9. Sorghum processing for food products

Sorghum grain processing is considered a prerequisite for the production and/or consumption of sorghum-based food products. Food processing includes mechanical, traditional thermal, biological and non-thermal treatments, the first three being the most applied to sorghum grains (Li *et al.*, 2022). In general, treatments can change the functional and nutritional properties of sorghum grains, enhancing their palatability for human consumption (Rashwan *et al.*, 2021). The primary nutritional modifications resulting from the three main types of processing, with a focus on grain proteins and phenolics, will be briefly discussed below.

Mechanical processing

Mechanical processing methods involve physical techniques to change the structure and properties of sorghum grains. This technique is typically employed as a pre-processing step; due to the hardness of sorghum grains, mechanical methods are applied to prepare the grains for subsequent processing (Li *et al.*, 2022). Among mechanical methods, dehulling is particularly notable for enhancing storage safety and improving the

sensory quality of sorghum (Buitimea-Cantúa *et al.*, 2013). Dehulling is the process of removing the grain pericarp and most of the testa layer. This outer layer contains compounds that contribute to the strong flavor and lipophilic compounds; thus, their removal results in grains with more pleasant odors and flavors, reduced rancidity, and enhanced storage safety (Buitimea-Cantúa *et al.*, 2013). However, decortication also reduces the levels of compounds concentrated in the pericarp and testa layer, leading to paradoxical effects: (1) a lower bioactive and antioxidant potential of sorghum grains due to the loss of a significant portion of phenolics, and (2) the removal of a large portion of phenolics can reduce the formation of protein-phenolic complexes, thereby increasing the digestibility of sorghum proteins (Buitimea-Cantúa *et al.*, 2013; Li *et al.*, 2022).

Thermal processing

Thermal processing methods use heat to modify the chemical and physical characteristics of sorghum grains; and can be divided into wet (boiling in water) and dry cooking (roasting) (Li *et al.*, 2022). Thermal processing can inactivate microorganisms and enzymes in sorghum, thereby ensuring food safety; however, it also adversely affects the nutritional and sensory properties of sorghum products (Rashwan *et al.*, 2021). For exemple, wet cooking treatment of sorghum grains decreases the overall content of bioactive compounds and potential biological activities, but increases the content of free phenolic compounds. This can be attributed to the thermosensitivity of phenolic compounds at high temperatures and leaching losses (Li *et al.*, 2022). Conversely, high temperatures can damage the cellular structure of the grains, releasing bound phenolic compounds (Xiong *et al.*, 2019a).

Wet heat treatment also affects sorghum proteins and their digestibility. Cooling sorghum grains (or products) after cooking promotes the formation of resistant starch, which can form complexes with kafirin proteins, making them less susceptible to enzymatic attack (Duodu *et al.*, 2002). Boiling also activates several endogenous factors such as trypsin inhibition, disulfide and non-disulfide crosslinking, increased kafirin hydrophobicity, and alterations in protein secondary structure (Duodu *et al.*, 2002; Rashwan *et al.*, 2021).

Biological processing

Biological processing methods involve the use of microorganisms or enzymes to transform sorghum grains. Germination is one of the significant biological methods that is widely used to improve the palatability and nutritional characteristic of cereals for human consumption (Rashwan *et al.*, 2021). It brings the degradation of main macronutrients, reduction of anti-nutritional factors, and increment of different bioactive compounds and many bioactivities. During germination, phytates, tannins, and enzyme inhibitors are degraded, leading to improved palatability and enhanced bioavailability of minerals such as iron and zinc (Afify *et al.*, 2011; Nkhata *et al.*, 2018).

In proteins, germination enhances solubility, *in vitro* digestibility, and free amino acid content (Abdelbost *et al.*, 2024). This improvement is due to the activation of proteolytic enzymes, which depolymerize cross-linked kafirins and glutelins. Additionally, the activity of intrinsic amylases, proteases, phytases, and fiber-degrading enzymes increases during germination, thereby positively impacting nutrient digestibility (Abdelbost *et al.*, 2023; Afify *et al.*, 2012). Concerning phenolic compounds, Hithamani & Srinivasan (2014) demonstrated that germination did not significantly alter the total phenol content but did result in a significant reduction in the total flavonoid content.

In summary, sorghum grain processing offers methods to remove/reduce antinutritional factors, value bioactive compounds, and enhance protein quality, thereby improving the quality of sorghum-based products. It is important to note that the effects of differents treatments on sorghum grains are dependent on specific conditions, such as temperature, time, water content, and the genotype used, among other factors. Optimizing these processing techniques is crucial for maximizing the sorghum value. Therefore, it is necessary to apply efficient and sensitive techniques to profile the changes in compounds present in sorghum grain throughout different types of processing, elucidating the main mechanisms involved.

10. Methods for assessing transformations in phenolic compounds and proteins

Given the complexity of the sorghum matrix and the need to evaluate changes after processing and throughout development, analytical tools have been employed to characterize phenolic compounds and proteins in sorghum grain, as well as their interactions (Czubinski & Dwiecki, 2017; Khoddami *et al.*, 2013; Punia *et al.*, 2020).

For the characterization of phenolic compounds in sorghum grain, metabolomic techniques, such as liquid chromatography coupled to mass spectrometry (LC-MS), have been emphasized. Mass spectrometry is an analytical technique used to identify and quantify chemical species in their ionized forms by measuring their mass/charge ratios (*m/z*) in the gas phase. This technique can characterize complex matrices and elucidating structures by comparing chromatograms and multistage mass spectra with standards and data libraries.

Metabolomics has been successfully applied to characterize the vast diversity of specialized metabolites, such as phenolic compounds in food integrating the concept of "Foodomics." It has been recently defined as a new discipline that studies food and nutrition domains through the application of advanced omics technologies in which MS techniques are considered indispensable (Cifuentes, 2009). This concept employs omics tools to understand and map chemical compounds and to characterize contaminants in foods (Herrero *et al.*, 2012). In cereal grains, it has been successfully applied to characterize the evolution of phenolic compounds in different stages of grain development, the phenolic fingerprint in different genotypes and milling fractions, for instance (Lima *et al.*, 2024; Lima *et al.*, 2023; Santos *et al.*, 2019b). In sorghum, this technique has been essential for understanding not only the levels but also the phenolic profile of different sorghum genotypes with contrasting pericarp colors and grown in various regions worldwide (Ofosu *et al.*, 2021; Pontieri *et al.*, 2021; Xiong *et al.*, 2020). Additionally, this technique can elucidate the transformations these metabolites undergo after sorghum processing (D'Almeida *et al.*, 2021b; Montini *et al.*, 2020).

For sorghum proteins analysis, studies range from simpler analyses using the Kjeldahl method to advanced tools such as proteomics, which also utilizes LC-MS to map proteins and elucidate their biological functions. Although these analyses provide valuable information about proteins, in sorghum grains, they have been primarily applied to obtain information about plant growth and interactions with the environment, such as drought tolerance and heat stress (Labuschagne, 2018). Alternatively, size-exclusion high-performance liquid chromatography (SE-HPLC) has been widely used to separate and quantify protein fractions, allowing the analysis of protein aggregation, degradation, and conformational changes (Abdelbost *et al.*, 2024; Gallo *et al.*, 2024).

To understand the protein-phenolic complexes in plant matrices, the fluorescence quenching technique has been widely applied (Ozdal *et al.*, 2013). Fluorescence

spectroscopy is a valuable tool in investigating protein structure, function, and reactivity. Wavelength shifts and changes in the intensity of fluorescence emission of tryptophan residues can be used to monitor local interactions between proteins and other grain components, such as phenolic compounds (de Morais *et al.*, 2020a; Qie *et al.*, 2021). Furthermore, the computational technique of molecular docking has been efficient in predicting interactions between proteins and phenolic compounds in plant matrices (Dai *et al.*, 2022; Liu *et al.*, 2022; Zhang *et al.*, 2022). This method simulates the binding process by positioning the phenolic molecule in the active site or binding region of the protein, elucidating preferential sites and the strength of the interaction at the molecular level.

Despite the efficiency of these mentioned techniques, there is still a lack of information on both phenolic compounds and proteins in sorghum grains. The synthesis and evolution of phenolic compound and protein levels throughout sorghum maturation and during some processing methods remain underexplored. Additionally, the study of the interactions between phenolic compounds and sorghum proteins appears to be unprecedented. Obtaining this information can help in understanding biotic and abiotic factors and in developing processes from harvest onwards, aiming at higher protein quality and retention of bioactive compounds in the grains and processed sorghum-based products. This information can help in selecting genotypes, maturation stages, and processing types with greater bioactive potential and better protein digestibility.

Part II

From flowering to grain maturity: a comprehensive study on grain development in sorghum

Chapter 2 – *"Digestibility, biochemical changes and kafirin polymerization analysis during sorghum grain development"*

Chapter 3 – *"Dynamic metabolomic changes in the phenolic compounds profile and antioxidant activity in developmental sorghum grains*

Chapter II - Digestibility, biochemical changes and kafirin polymerization analysis during sorghum grain development

Carolina Thomaz dos Santos D'Almeida [1,2], Marie-Hélène Morel [3], Joëlle Bonicel [3], Nancy Terrier [4], Mariana Simões Larraz Ferreira [1,2], Hamza Mameri [3*]

[1] Laboratory of Bioactives, Food and Nutrition Graduate Program, Federal University of State of Rio de Janeiro, UNIRIO, Brazil.

[2] Center of Innovation in Mass Spectrometry, Laboratory of Protein Biochemistry, UNIRIO, Rio de Janeiro, 22290-240, Brazil.

[3] UMR 1208 IATE, Univ. Montpellier, INRAE, L'Institut-Agro Montpellier, F-34060 Montpellier, France

[4] AGAP Institute, Univ. Montpellier, INRAE, CIRAD, F-34398 Montpellier, France

This chapter is currently undergoing internal revisions and is expected to be submitted to *Nature Communications* (Impact Factor: 14.7). To enhance the manuscript, proteomic data currently being collected will be added, aiming to provide in-depth insights into the molecular and cellular events during sorghum grain growth that result in low kafirin digestibility. The submission is expected within the next six months and, for confidentiality reasons, it is not available in the online version of the thesis.

Chapter III - Dynamic metabolomic changes in the phenolic compounds profile and antioxidant activity in developmental sorghum grains

Carolina Thomaz dos Santos D'Almeida [1,2], Marie-Hélène Morel [3], Nancy Terrier [4], Hamza Mameri [3], Mariana Simões Larraz Ferreira [1,2], [*]

[1] Laboratory of Bioactives, Food and Nutrition Graduate Program, Federal University of State of Rio de Janeiro, UNIRIO, Brazil.

[2] Center of Innovation in Mass Spectrometry, Laboratory of Protein Biochemistry, UNIRIO, Rio de Janeiro, 22290-240, Brazil.

[3] UMR 1208 IATE, Univ. Montpellier, INRAE, L'Institut-Agro Montpellier, F-34060 Montpellier, France

[4] AGAP Institute, Univ. Montpellier, INRAE, CIRAD, F-34398 Montpellier, France

Highlights

- Folin-Ciocalteu and antioxidant capacity showed a progressive increase towards grain maturity
- Sorghum polyphenols accumulated and become more insoluble by the physiological grain maturity
- Free to bound phenolic compounds ratio showed a progressive reduction along the grain growth
- Phenolic profiles are different among genotypes and during the grain development
- Eight dimer and trimer isomers of procyanidins have been annotated
- Differentially expressed phenolics evidences the variability of genotypes/growing season

This chapter was submitted to the Journal of Agricultural and Food Chemistry (Impact Factor: 5.7).

ABSTRACT

Phenolic compounds (PC) were analyzed by UHPLC-ESI-QTOF-MS^E in two sorghum genotypes, harvested in two growing seasons (GS) at five days after flowering (DAF) to evaluate how genotype/GS influence the PC synthesis and antioxidant capacity during grain growth. Total phenolic contents were strongly correlated with antioxidant capacity ($r > 0.9$, $p < 0.05$). Globally, 97 PC were annotated, including 20 PC found irrespectively of grain developmental stage and genotype/GS. Phenolic profile clearly differs between stages, phenolic acids were the most abundant class in early stages (50%) and flavonoid accumulation becomes predominant in late ones (3/5 of total ion abundance). Dimeric and trimeric procyanidins were identified even in 10DAF grains. Chemometry revealed great PC variability between genotypes (27%) and important biomarkers of GS differentiation (*e.g.*, ferulic acid). This work can input open databases of PC and paves the way to understand biosynthetic pathways of PC in sorghum and future sorghum selection

Keywords: antioxidant compounds, polyphenols, *Sorghum bicolor*, UHPLC-MS^E

1. Introduction

Sorghum, botanically known as *Sorghum bicolor* L. Moench is considered the fifth most important carbohydrate-rich crop in the world. Despite sorghum is considered a staple food for about 500 million people in 30 countries in Africa and Asia, most of the sorghum is used as animal feed in almost all western countries (Khoddami *et al.*, 2023). Due to climate oscillations, the use of sorghum in human nutrition has been required, since this cereal can support drastic agronomic and environmental circumstances, mainly low rainfall, low availability of water for irrigation and salinity, i.e., sorghum would be a possible solution to food production stagnation and can ensure world food security (Hossain *et al.*, 2022). In addition to its agronomic advantages, sorghum grain is a rich source of nutrients (Xiong *et al.*, 2019b), and most importantly, contains a diverse range of health-promoting bioactive phenolic compounds (Cardoso *et al.*, 2017).

Phenolic compounds (PC) are a group of specialized metabolites, naturally biosynthesized by plants to act as defense agents in response to possible stresses caused during their development. These compounds are associated with a diverse human health benefits, such as reducing oxidative stress (antioxidant capacity) and cancer prevention (Xiong *et al.*, 2019b) until the modulation of the intestinal microbiota and the barrier function (Tiozon *et al.*, 2022). In sorghum, PC are concentrated in the grain outermost layers (bran), and they have a diversified profile, with the classes of flavonoids (such as flavonols, flavones, flavanones and tannins) and phenolic acids being the most abundant (Xiong *et al.*, 2019b).

Comprehensive knowledge about these compounds is the prerequisite for its industrial applications and classification, e.g., sorghum has been traditionally classified according to its tannin contents into tannin-rich and -poor sorghum. Tannins are positively related to reduced postprandial blood glucose release (Amoako & Awika, 2016), reducing the caloric value of starchy foods (Girard & Awika, 2018) and high antioxidant capacity (Tian *et al.*, 2012). However, high molecular weight condensed tannins are known to bind with proteins and severely limiting their bioaccessibility and digestibility (Duodu *et al.*, 2003). The profile and levels of tannins and others PC depends on the genotype, pedoclimatic and growth conditions.

The sorghum plant develops in a predictable manner characterized by three distinct growth stages - vegetative growth, panicle initiation and grain filling. The latter stage begins with flowering and continues until dry matter accumulation (physiological

maturity or when grain attains the maximum dry weight). Tannin and others PC synthesis begins at this stage (60-90 days after sorghum crop planting) (Roozeboom *et al.*, 2016). Sorghum grain development progresses from milk to physiological maturity over a 25- to 45-days after flowering (DAF), depending on genotype and environmental conditions (Gerik *et al.*, 2003). Although the variation in macronutrient composition during this process is well established in the literature (Gerik *et al.*, 2003), the PC synthesis during the development process are largely unknown.

Recent advances in the metabolomics field have contributed to a better understanding of plant metabolism, metabolome analyses in crop science can provide valuable information that goes beyond biomarkers identification to a tool for discovering active drivers involved in biological processes (Hamany Djande *et al.*, 2020). In this study, we aimed to investigate the temporal changes in sorghum grain phenolic profile and antioxidant capacity of different genotypes and growing seasons at five developmental stages. These results provide insights into the PC biosynthesis in sorghum (*Sorghum bicolor* L.) during the grain development.

2. Material and methods

2.1. Chemicals and Reagents

The following reference standards, as well as MS-grade acetonitrile and methanol, were purchased from Sigma–Aldrich (St. Louis, MO, USA): vanillic acid, p-coumaric acid, catechin, caffeic acid, ellagic acid, trans-ferulic acid, kaempferol, myricetin, pyrogallol, flavanone, quercetin, gallic acid, epicatechin, 4-hydroxybenzyl alcohol, 4-hydroxyxy benzaldehyde acid, 4-hydroxybenzoic acid, 4-hydroxybenzoic acid, 4-phenylacetic acid, synapinic acid, benzoic acid, quercetin 3 glycoside, 3,4-diOH phenylacetic acid, epigallocatechin, epigatechin gallate, chlorogenic acid, 2,5-dihydroxy benzoic acid, 4-methoxycinnamic acid, 2-hydroxycinnamic acid, 3-hydroxy-4-methoxycinnamic acid, trans cinnamic acid, 3-methoxycinnamic acid, L-(-)-3 phenylacetic acid. Formic acid was purchased from Fluka (Switzerland). MilliQ water was obtained through the Barnstead™ Smart2Pure™ (Thermo Fisher Scientific, USA) purification system. Other unmarked reagents were of analytical grade.

2.2. Samples

Immature sorghum (*Sorghum bicolor L.*, caudatum race) grains from two different genotypes: 1) red pericarp (IS15752, pigmented testa and presence of condensed tannins)

and 2) white pericarp (Macia, without pigmented testa and tannin-free) were cultivated and collected by the unit "Genetic improvement and adaptation of tropical and Mediterranean plants" (UMR-AGAP, CIRAD, INRAE, Montpellier, France). The grains were cultivated at Mauguio and Lavalette domaines (Montpellier, South region, France) in two growing seasons (Wu *et al.*): 2017 (GS1) and 2018 (GS2) (Figure 11A).

The grains were collected at five stages, namely, 10DAF, 17DAF, 25DAF (grain filling stage), 33DAF (dough stage) and 40DAF (harvest maturity stage) (Figure 11A); for each stage, three replicates were collected, and each replicate consisted of 3 panicles of independent randomly selected plants. Panicles were stored at -60°C after harvest. Whole grains were freeze-dried at 12% moisture content, cryogenically ground for 2 min using a ball mill and the resulting powder was maintained at -80 °C until analysis.

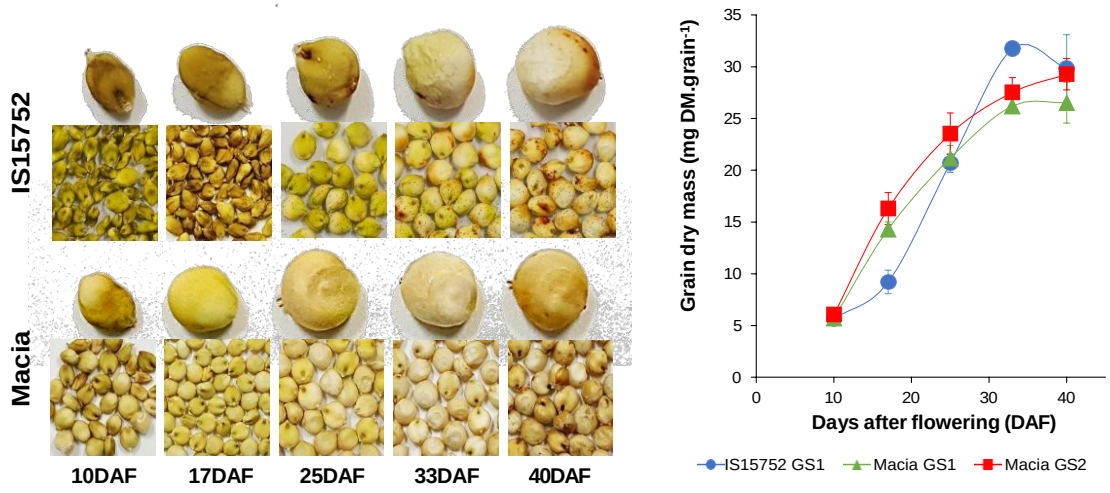


Figure 3. Photograph of each genotype (on the left) and the average weight (on the right) on each day after flowering (DAF) of the sorghum grain. Results are expressed as mean ± standard deviation (n=3).

2.3. Free and bound compounds extraction

To obtain free and conjugated PC from sorghum flour, the extraction technique according to Santos *et al.* (2019a) was performed in triplicate with some modifications. Free phenolic compounds (FPC, soluble) were extracted in 80% ethanol at a ratio 1:20 (w/v) and stirred at room temperature (25°C, 200 rpm) for 10 min. After 10 min of centrifugation at 5,000 *g* at 25°C, the supernatant was removed and stored in Eppendorf tubes (-80 ºC). Extraction was performed twice, and the extracts obtained were pooled. The pellets resulting from FPC extraction were submitted to alkaline hydrolysis with 1:70 (v/v) of 4M NaOH (submerged in an ultrasonic bath - 42 kHz, 90 min, 40 °C). After, the

acid hydrolysis was performed with concentrated HCl (~pH 2), and the samples were centrifuged (2,000 x*g*, 5 min). The supernatant was washed three times with ethyl acetate (7 mL) and centrifuged between each wash step (10,000 x*g*, 5 min, 10 °C), to obtain the bound phenolic compounds (BPC, insoluble). Both extracts (FPC and BPC) were evaporated (SpeedVac Savant, ThermoFisher Scientific, USA) and reconstituted in 500 µL of methanol, acetonitrile and Milli-Q water (2:5:93, v/v/v). The reconstituted extracts were filtered (0.22 µm, hydrophilic PTFE, Analytical) and stored in vials at −80 °C.

2.4. Total reducing capacity

The total reducing capacity (TRC) was determined by Folin-Ciocalteu method, in triplicate, according to Singleton *et al.* (1999), adapted for microplates. Extracts (100 µL) were added to 700 µL of Milli-Q water in test tubes. After homogenization, 50 µL of Folin-Ciocalteu reagent and 150 µL of 20% sodium carbonate were added. The mixture was incubated (30 min, 40°C) and 300 µL of the final solution was transferred to a microplate. The absorbance reading at 750 nm was performed in a FlexStation III microplate reader (Molecular Devices). Solvent blank and standard curve analyze were performed with gallic acid (5 to 130 µg/mL). Results were expressed in mg of gallic acid equivalents (GAE) per 100 g of sample, in dry basis.

2.5. Determination of antioxidant capacity

The antioxidant capacity of samples was determined, in triplicates, by the DPPH (2,2-diphenyl-1-picrylhydrazyl) radical scavenging method and the ferric reducing antioxidant power method (FRAP), adapted to microplates (Sompong *et al.*, 2011). For the DPPH method, a 20 µL aliquot of each extract was combined with 280 µL of the DPPH solution (32 µg/mL) and incubated (30 min, in the dark, 25 °C). For FRAP assays, the reagent was prepared in acetate buffer (0.3 M, pH 3.6), FeCl3.6H2O (20 mM) and TPTZ solution (10 mM) in a 10:1:1 ratio. A 20 µL aliquot of each extract was combined with 15 µL of Milli-Q water and 265 µL of FRAP reagent, gently vortexed and incubated (30 min, 37 °C). Absorbance was measured using a microplate reader (FlexStation III, Molecular Devices, USA) at 715 nm and 595 nm, respectively, and results were expressed as µmol of trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid) equivalents (TE) per 100 g of sample, in dry basis.

2.6. Metabolomic analysis of sorghum grains by UHPLC-MS^{E}

The phenolic profiling was performed injecting 5 µL of each sample into an Ultra High Performance Liquid Chromatography (UHPLC) Acquity system (Waters, USA) coupled with XEVO G2S Q-Tof (Waters, England) equipped with ionization source electrospray. An UHPLC HSS T3 C18 column (100 x 2.1 mm, 1.8 µm particle diameter; Waters) at 30 °C and flow rate of 0.5 mL/min of ultra-pure water containing 0.3% formic acid and 5 mM ammonium formate (mobile phase A) and acetonitrile containing 0.3% formic acid (mobile phase B) was used according to the gradient method: : 0 min - 97% A; 11.80 min - 50% A; 12.38 min - 15% A; 14.11 - 97% A. Data were acquired in triplicate in MS^E negative and centroid mode between *m/z* 50 and 1200, collision energy ramp from 30 to 55 V; cone voltage 30 V; capillary voltage 3.0 kV; desolvation gas (N_2) 1,200 L/h at 600 °C; cone gas 50 L/h; source at 150 °C; and using leucine enkephalin (Leu-Enk, *m/z* 554.2615, [M-H]-) for calibration. A mix containing 33 analytical standards of phenolic compounds (10 ppm) was prepared and injected in triplicate, prior to the injection of the samples, to ensure the reproducibility of the instrument and to confirm phenolic compounds identification. Besides the injection of the chemical standards, a set of quality control (QC) samples was also prepared by pooling equal volumes of each sorghum extracts and were injected after each batch of six runs of sorghum samples to monitor the instrument's stability.

MassLynx v 4.1 software (Waters, USA) was used to acquire MS data and Progenesis QI (Waters, USA) software was applied to data processing. Non-targeted identification was performed according to Metabolomics Initiative Standard as described by Sumner *et al.* (2007) considering a customized database built from PubChem and online database Phenol-Explorer. The metabolites identification (level 1) was based on standards runs parameters, such as: isotope distribution of neutral mass, exact mass, retention time and MS/MS fragments spectra. The following parameters were applied to annotated metabolites (level 2 and 3): exact mass error (<10 ppm); isotopic similarity (>80%); score (>30) and the highest score of fragmentation, generated by the software. Data from the literature and chemical characteristics of the molecules were also used to help the tentative annotation of unknown compounds. In addition, only compounds present in the three technical replicates and showing CV < 30% were considered. The resulting compounds had their normalized relative abundance divided by one hundred and multiplied by the average grain dry weight (mg $DM.grain^{-1}$), in order to calculate the relative phenolic abundance per grain. Metabolic pathways were proposed based on the

observed phenolic changes in this study and their comparison with the KEGG phenylpropanoid biosynthesis pathway (map00940).

2.7. Statistical analysis

Statistical analysis was performed with Tukey's test ($p < 0.05$) and one-way ANOVA, using XLSTAT software (Addinsoft, France). The raw data obtained by UHPLC-MS^E was submitted to the normalization step in Progenesis QI software to allow quantitative comparisons among samples, by abundance obtained from ion mass spectra. Data generated were exported to perform multivariate analysis such as the hierarchical cluster analysis by XLSTAT and the orthogonal partial least-squares discriminate analysis (OPLS-DA) by EZinfo 3.0. The efficiency and reliability of the OPLS-DA models were verified by percent variation of the y variables explained by the model (R2Y) and the predictive performance of the model (Q2), using Metaboanalyst 5.0 web server (https://www.metaboanalyst.ca/). In addition, permutation tests were carried out with 100 random permutations, to validate the OPLS-DA models.

3. Results and discussion

3.1. Evaluation of sorghum grain development

Figure 11 presents images of the sorghum grain and their respective growth curves. Visible morphological changes (Figure 11B) and the sigmoidal growth curve of the average grain weight analyzed across contrasting genotypes and growing seasons (Wu *et al.*) (Figure 11A) clearly delineated the filling stages of the sorghum grain. Both genotypes and GS demonstrated rapid growth during the initial grain development stages (10-25DAF), indicative of rapid grain filling post-flowering. By 33DAF, the grains reach their maximum weight, followed by a stabilization phase at 40DAF.

The sorghum grain formation involves four stages: 1) milk stage, initial stage of grain development that occurs around 10DAF; 2) soft and 3) hard dough stages, stages where the grain reaches about 50% and 75% of its total dry weight, respectively; and 4) physiological maturity, indicating that maturation is complete after full grain filling (100% of its total dry weight) (Gerik *et al.*, 2003). In the present study, all these stages of grain development were covered and would be effective in understanding PC synthesis during sorghum grain maturation: 10DAF = milk stage; 17DAF = soft dough stage; 25DAF = hard dough stage; and 33 and 40DAF = physiological maturity (Figure 11B).

Similarities and differences between genotypes and GS (two crop years) during grain formation were observed. While for Macia genotype there was no variation: Macia GS1 and Macia GS2 are statistically equal in all stages showing similar profiles of grain dry mass accumulation; the IS15752 genotype showed a peculiar behavior in some stages ($p<0.05$) (Figure 11B): at 17DAF it showed lower grain dry mass values than Macia GS1 and GS2 (-36 and -44%, respectively), while at 33DAF these values were higher (+21 and +14%, respectively), suggesting the conversion of sugars and amino acids into starch and protein, respectively, may occur later in IS15752, but its effective filling in later stages forms heavier grains.

3.2. Phenolic compounds and antioxidant capacity

The TRC was determined in both, free (FPC) and bound (BPC) extracts, throughout grain development in the two genotypes and two GS (Figure 12). IS15752 GS1 ranged from 7.38±0.29 to 26.03±3.39 for the total extract (TPC corresponds to the sum of FPC and BPC), and, in general, the values at each stage were three-fold greater than that found in Macia GS1 (3.62±0.29 to 9.13± 0.94) and GS2 (2.30±0.19 to 8.07±1.11). The superiority of IS15752 was found in all antioxidant analysis ($p<0.05$) and can be explained by the presence of tannins in this genotype (D'Almeida *et al.*, 2021b). When comparing the different GS (for Macia genotype), the most immature stage (10DAF) showed TRC values 57% higher in GS1 when compared to GS2 ($p<0.01$); while in the other stages, the values were similar. The hypothesis is that the Macia GS1 genotype suffered some abiotic stress in the initial stage of grain growth, which favored the synthesis of PC to protect plants from oxidative stress (Sharma *et al.*, 2019).

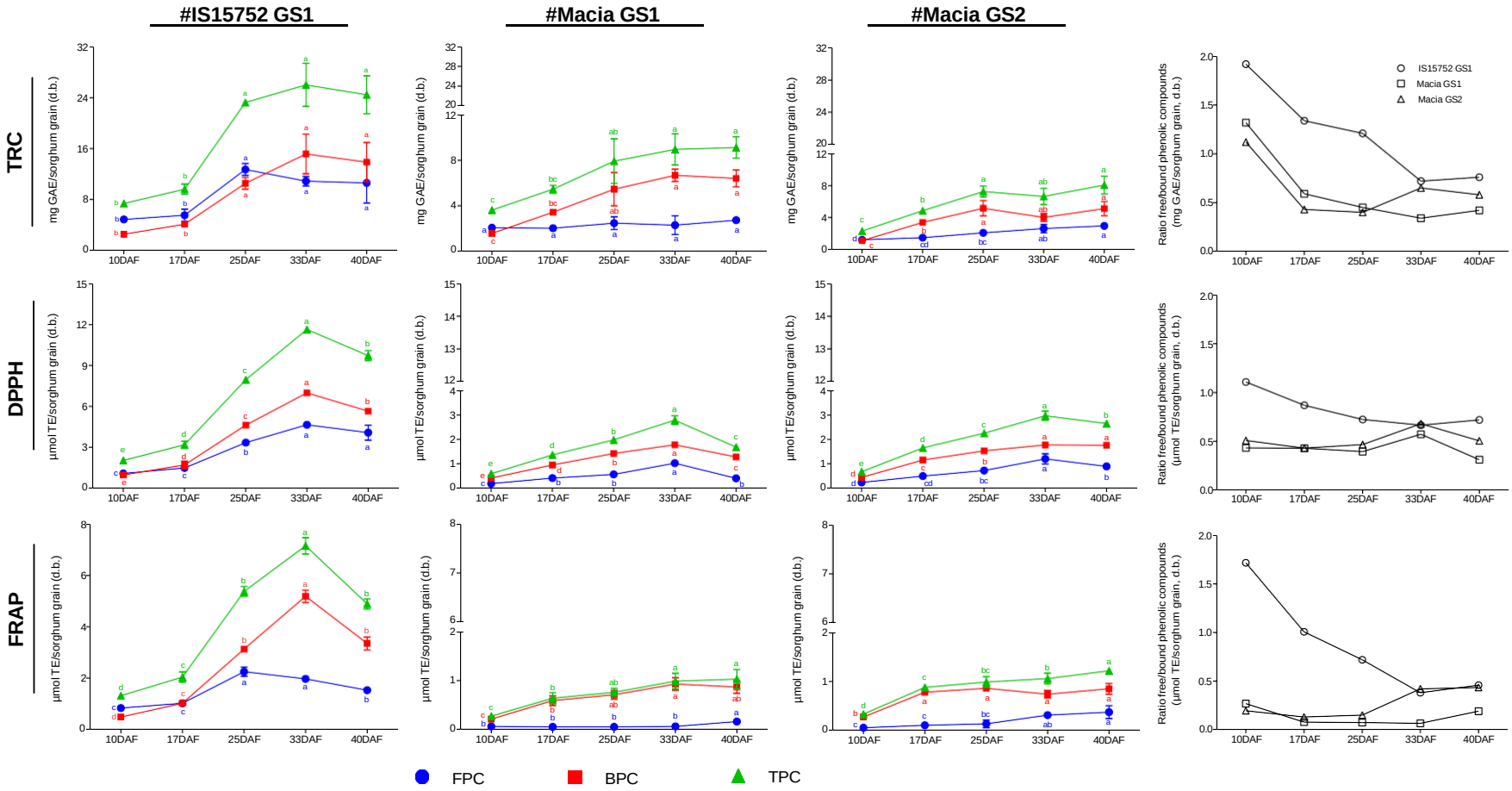


Figure 4. Evaluation of days after flowering (DAF) in total reducing capacity (TRC) and antioxidant capacity (DPPH and FRAP methods) in free (FPC), bound (BPC) and total (TPC) phenolic compounds extracts, in different genotypes and growing seasons (Wu *et al.*) of sorghum grains. The ratio between the FPC and BPC values in each analysis is shown in the last column. Results are expressed as mean ± standard deviation (n=3). Different letters indicate a significant difference between DAF (Tukey, $p<0.05$).

The changes of TRC throughout grain development behaved in a similar way in all samples, showing a progressive increase towards maturation. The initial stages (10DAF and 17DAF) have lower TRC values, followed by a significant increase in 25DAF (142, 46 and 49% in IS15752 GS1, Macia GS1 and Macia GS2, respectively, compared to 10DAF), and constant values in mature stages (33 and 40DAF) (Figure 12). In Macia genotypes, this increase was essentially due to the BPC, mainly in GS1, while for IS15752 genotype the significantly increase occurred until 25DAF for both FBC and BPC, and then stabilized. In contrast to what has been reported in wheat grain (Santos et al., 2019), this result indicates an insolubilization and complexation of phenolic compounds during the sorghum grain development. Indeed, FPC:BPC ratio showed a progressive reduction along the grain development for IS15752 GS1 (from 1.92 to 0.76), and Macia GS1 (from 1.32 to 0.42) and GS2 (from 1.12 to 0.58) samples, indicating that BPC are the main responsible for TPC increase. These results corroborate previously published data with maize kernels (Zhang *et al.*, 2020); however, in this case, FPC is the main responsible for the increase, highlighting a dissimilarity between sorghum and maize.

Looking at each extract (free and bound) separately, FPC extract was predominant in Macia genotype at 10DAF (57% of TPC), but the significant synthesis of BPC in the soft dough stage makes it the majority (BPC averaged 54% of TPC from 25DAF). During *in vivo* digestion, BPC reach the colon and are processed/transformed by microbial activity, presenting potential beneficial effects on human health (Rocchetti *et al.*, 2022). In IS15752 GS1, the BPC start to be produced (synthesized and linked to other components) early in the grain development and show the maximum by 25-33DAF. This result is expected for this genotype, since it is classified as high condensed tannins (proanthocyanidins), and this phenolic class is usually bound to components of the plant matrix (Xiong *et al.*, 2019b). Despite this, it is important to consider that although the interactions between condensed tannins and other matrix components can be broken by the action of acid hydrolysis, the method applied in the present study was not efficient to depolymerize and, consequently, to extract and to quantify these compounds; also the Folin-Ciocalteu method present interferences with other reducing power substances such as ascorbic acid, aromatic amines and sugars.

As expected, the antioxidant capacity measured by DPPH and FRAP methods, showed a strong correlation with TRC (0.9846 and 0.9737, $p < 0.05$, respectively). Through these different methods it is possible to observe that, although the general behavior during sorghum grain development is similar between genotypes and GS, there are variations in their proportions in the extracts (free and bound). In Macia genotypes, irrespective to GS, BPC extract was almost superposed to TPC in FRAP and DPPH results. Since IS15752 GS1 had the same growing conditions as Macia GS1, it is believed that these variations can be associated with the proanthocyanidin-rich composition of the first genotype.

3.3. Identification of phenolic compounds by UHPLC-MS^E

The phenolic profiles of the different sorghum genotypes and GS were followed during different grain development stages by UHPLC-MS^E method, providing the most comprehensive screening in sorghum grains to date. Globally, a total of 97 PC were tentatively identified, among them 11 compounds were fully confirmed by reference standards (Supplementary Table 2, compounds in bold): 7 were present in both extracts (free and bound); 2 were identified only in free extract; and 2 were identified only in bound extract. Contrary to findings in the literature, which indicated that the number of compounds identified in immature cereal grain samples and mature whole sorghum samples was greater in bound extracts compared to free extracts (D'Almeida *et al.*, 2021b; Santos *et al.*, 2019a; Shao *et al.*, 2014), our study found that the majority of PC were present in free extracts (exclusive 47 PC) rather than in bound extracts (exclusive 32 PC). A total of 18 PC were commonly identified in both extracts. The predominance of FPC is also observed in the total relative ion abundance, where the abundance of FPC was 112% higher than BPC. The annotated phenolic compounds belonged mainly to the flavonoid class (54%), followed by phenolic acids (32%), other polyphenols (12%) and lignans (2%). One compound ($[M-H]^-$10.16 *m/z* 191.0343) could not be assigned to a class and was classified as unknown compound.

Some compounds were systematically present independently of the development stage (32 PC) or the genotype and GS (26 PC) (Figure 13A and B, intersection of the Venn diagram). These compounds and their relative abundance by total ion counting are described in Table 1, where the confirmed identifications with phenolic pattern are in bold. Additionally, among these compounds, 20 PC were common across all samples,

regardless of developmental stage or genotype/GS (Table 1, compounds marked with an asterisk).

Table 1. Phenolic conpounds (PC) found in connon between the different developnental stages indicated by days after flowering (DAF, n = 32) or between genotypes/growing season (GS, n = 26) of sorghun grains (per seed).

Name of compound	Molecular formula	*m/z*	RT (min)	Class	IS15752 GS1					Macia GS1					Macia GS2				
					10DAF	17DAF	25DAF	33DAF	40DAF	10DAF	17DAF	25DAF	33DAF	40DAF	10DAF	17DAF	25DAF	33DAF	40DAF
PC in common among all days after flowering (n = 32)																			
4-Hydroxybenzaldehyde*	**$C_7H_6O_2$**	**121.0289**	**6.14**	**OP**	$4.52x10^5$	$6.97x10^5$	$1.67x10^6$	$1.58x10^6$	0	$4.46x10^4$	$8.09x10^4$	$8.80x10^4$	$7.94x10^4$	$6.10x10^4$	$5.35x10^4$	$4.19x10^4$	$8.18x10^4$	0	$2.22x10^5$
Caffeic acid*	**$C_9H_8O_4$**	**179.0343**	**6.24**	**PA**	$9.42x10^4$	$8.05x10^4$	$1.81x10^5$	$4.14x10^5$	$5.29x10^5$	$1.17x10^4$	$3.26x10^4$	0	$4.46x10^5$	$6.08x10^5$	$2.02x10^4$	$4.08x10^4$	0	$5.10x10^5$	$3.63x10^5$
trans-ferulic acid*	**$C_{10}H_{10}O_4$**	**193.0499**	**8.00**	**PA**	$6.70x10^4$	$1.15x10^5$	$1.49x10^5$	$2.28x10^5$	$1.99x10^5$	$9.20x10^4$	$1.92x10^5$	$2.87x10^4$	$4.21x10^5$	$3.43x10^5$	$4.54x10^3$	$3.30x10^5$	$4.53x10^5$	$3.53x10^5$	$3.47x10^5$
p-coumaric acid*	**$C_9H_8O_3$**	**163.0393**	**7.48**	**PA**	$2.51x10^4$	$5.07x10^4$	$1.28x10^5$	$6.56x10^4$	$2.38x10^5$	$6.97x10^3$	$1.01x10^5$	$5.60x10^4$	$2.88x10^5$	$4.68x10^4$	$1.03x10^4$	$1.31x10^5$	$4.48x10^4$	$3.20x10^5$	$4.95x10^5$
Scutellarein*	$C_{15}H_{10}O_6$	285.0392	10.16	F	$8.49x10^2$	$1.21x10^4$	$1.66x10^4$	0	$1.76x10^3$	0	$2.54x10^4$	0	0	$2.96x10^5$	0	$4.65x10^4$	$1.71x10^5$	$4.09x10^5$	$5.54x10^5$
3'-Hydroxymelanettin*	$C_{16}H_{12}O_6$	299.0551	11.03	F	$2.98x10^2$	0	$4.96x10^3$	0	0	$5.83x10^3$	$1.03x10^3$	$2.45x10^3$	0	0	$7.92x10^3$	$3.10x10^4$	$2.19x10^5$	$3.35x10^5$	$6.93x10^5$
Esculetin*	$C_9H_6O_4$	177.0185	6.10	OP	$2.20x10^3$	$2.00x10^3$	$3.73x10^3$	$1.54x10^4$	$3.60x10^4$	$4.56x10^3$	$1.09x10^4$	0	$2.50x10^5$	$3.58x10^5$	$4.22x10^3$	$1.09x10^4$	$4.64x10^4$	$1.84x10^5$	$2.72x10^5$
Dihydrocaffeic acid	$C_9H_{10}O_4$	181.0500	4.46	PA	$6.56x10^4$	$1.23x10^5$	$2.74x10^5$	$3.84x10^5$	$3.47x10^5$	0	0	0	0	0	0	0	0	0	0
Ferulic acid*	$C_{10}H_{10}O_4$	193.0498	8.25	PA	$3.19x10^4$	$2.05x10^3$	$7.31x10^3$	$9.90x10^4$	$9.43x10^4$	$4.36x10^4$	$3.70x10^3$	$9.57x10^3$	0	$1.41x10^5$	0	$1.65x10^5$	$2.54x10^5$	$1.55x10^5$	$1.68x10^5$
Naringenin 7-O-glucoside*	$C_{21}H_{22}O_{10}$	433.1128	7.84	F	$1.33x10^4$	$2.52x10^4$	$5.29x10^4$	$8.95x10^4$	$7.92x10^4$	0	0	$4.75x10^4$	0	$4.83x10^4$	0	$3.80x10^4$	$6.59x10^4$	$5.78x10^4$	$6.90x10^4$
4-Hydroxybenzoic acid*	**$C_7H_6O_3$**	**137.0237**	**5.03**	**PA**	$1.37x10^4$	$3.06x10^4$	$7.30x10^4$	$1.83x10^5$	$1.44x10^5$	$8.51x10^3$	$4.56x10^3$	$2.79x10^3$	$1.12x10^4$	$1.35x10^4$	0	$7.17x10^3$	$4.79x10^3$	0	$3.59x10^4$
Dihydroxybenzoic acid isomer I*	$C_7H_6O_4$	153.0186	3.46	PA	$1.70x10^4$	$3.39x10^4$	$6.34x10^4$	$1.78x10^5$	$2.29x10^5$	0	0	0	$2.03x10^3$	$3.72x10^3$	$5.68x10^2$	$1.08x10^3$	0	$2.76x10^3$	0
Procyanidin dimer I	$C_{30}H_{26}O_{12}$	577.1338	5.42	F	$2.12x10^4$	$3.50x10^4$	$1.63x10^5$	$1.86x10^5$	$1.25x10^5$	0	0	0	0	0	0	0	0	0	0
Procyanidin trimer I	$C_{45}H_{38}O_{18}$	865.1963	5.76	F	$1.23x10^4$	$2.04x10^4$	$9.39x10^4$	$1.11x10^5$	$9.14x10^4$	0	0	0	0	0	0	0	0	0	0
Hesperidin*	$C_{28}H_{34}O_{15}$	609.1881	0.57	F	$4.00x10^3$	$6.16x10^3$	$1.17x10^4$	$1.98x10^4$	$2.36x10^4$	$4.36x10^3$	$1.53x10^4$	$2.16x10^4$	$2.25x10^4$	$2.29x10^4$	$4.74x10^3$	$1.50x10^4$	$1.79x10^4$	$2.11x10^4$	$2.21x10^4$
5-Caffeoylquinic acid*	$C_{16}H_{18}O_9$	353.0865	5.87	PA	$7.03x10^4$	0	$2.34x10^4$	$1.76x10^4$	$1.44x10^4$	0	$3.17x10^3$	0	$8.18x10^3$	0	$9.82x10^3$	$1.22x10^4$	$9.24x10^3$	0	$2.83x10^4$
3-Feruloylquinic acid*	$C_{17}H_{20}O_9$	367.1023	6.07	PA	1.17x104	0	$1.10x10^4$	0	$1.26x10^4$	$1.05x10^4$	$2.58x10^4$	0	$1.82x10^4$	0	$2.72x10^4$	$2.15x10^4$	0	$2.64x10^4$	$1.76x10^4$
Tetramethoxyflavone isomer III*	$C_{19}H_{18}O_6$	341.1019	10.12	F	$2.31x10^3$	0	0	0	$7.48x10^3$	$5.03x10^3$	$1.16x10^4$	$1.59x10^4$	$2.00x10^4$	$1.98x10^4$	$3.09x10^3$	$1.14x10^4$	$1.69x10^4$	$1.88x10^4$	$2.86x10^4$
Naringin 4'-O-glucoside*	$C_{21}H_{22}O_{10}$	433.1128	8.35	F	$2.82x10^3$	$5.73x10^3$	$1.18x10^4$	$1.99x10^4$	$1.58x10^4$	0	0	$1.13x10^4$	$1.28x10^4$	$8.59x10^3$	0	$8.83x10^3$	$1.42x10^4$	$1.20x10^4$	$1.35x10^4$
Vanillin*	**$C_8H_8O_3$**	**151.0393**	**5.80**	**OP**	$2.43x10^3$	$4.07x10^3$	$8.14x10^3$	$1.44x10^4$	$1.72x10^4$	$3.05x10^3$	0	0	0	$7.22x10^3$	0	0	0	$3.38x10^4$	$3.86x10^4$
Dihydroxy-trimethoxyflavone isomer II	$C_{18}H_{16}O_7$	343.0811	9.93	F	0	0	0	0	0	$2.11x10^3$	$3.70x10^3$	$8.89x10^3$	$2.08x10^4$	$2.61x10^4$	$1.61x10^3$	$6.37x10^3$	$1.60x10^4$	$2.19x10^4$	0

Tetramethoxyflavone isomer I*	$C_{19}H_{18}O_6$	341.1019	8.30	F	$2.08x10^3$	$3.92x10^3$	0	$7.93x10^3$	$7.42x10^3$	$2.40x10^3$	$4.57x10^3$	0	$1.45x10^4$	$1.31x10^4$	0	$6.56x10^3$	$1.12x10^4$	$1.44x10^4$	$1.48x10^4$
Eriodictyol 7-O-glucoside	$C_{21}H_{22}O_{11}$	449.1073	6.95	F	$2.65x10^3$	$5.41x10^3$	$2.07x10^4$	$3.57x10^4$	$2.96x10^4$	0	0	0	0	0	0	0	0	0	0
Procyanidin dimer B-type III	$C_{30}H_{26}O_{12}$	577.1335	7.48	F	$4.13x10^3$	$6.10x10^3$	$2.15x10^4$	$2.02x10^4$	$1.55x10^4$	0	0	0	0	0	0	0	0	0	0
Feruloyl glucose*	$C_{16}H_{20}O_9$	355.1023	6.46	PA	$9.20x10^2$	0	0	0	$8.41x10^3$	0	$6.28x10^3$	0	$4.16x10^3$	0	$1.59x10^3$	$1.26x10^4$	$2.44x10^4$	0	$4.42x10^3$
Quercetin 3-O-rutinoside	$C_{27}H_{30}O_{16}$	609.1447	8.07	F	$5.30x10^3$	$7.21x10^3$	$1.50x10^4$	$1.48x10^4$	$1.45x10^4$	0	0	0	0	0	0	0	0	0	0
Dihydroxy-trimethoxyflavone isomer III	$C_{18}H_{16}O_7$	343.0812	10.59	F	0	0	0	0	0	0	0	0	0	$1.28x10^4$	$4.29x10^2$	$1.96x10^3$	$6.73x10^3$	$1.12x10^4$	$1.80x10^4$
Isorhamnetin 3-O-glucoside	$C_{22}H_{22}O_{12}$	477.1026	8.85	F	$1.00x10^3$	$1.81x10^3$	$5.92x10^3$	$1.05x10^4$	$1.24x10^4$	0	0	0	0	0	0	0	$1.37x10^3$	0	$1.33x10^4$
Tetramethoxyflavone isomer II*	$C_{19}H_{18}O_6$	341.1018	8.67	F	$5.12x10^2$	0	$1.63x10^3$	$2.27x10^3$	$2.48x10^3$	$8.45x10^2$	$1.43x10^3$	$3.17x10^3$	$5.37x10^3$	$4.84x10^3$	0	$2.16x10^3$	$4.39x10^3$	0	$6.72x10^3$
Procyanidin trimer C-type II	$C_{45}H_{38}O_{18}$	865.1956	6.14	F	$1.03x10^3$	$1.57x10^3$	$8.07x10^3$	$1.08x10^4$	$6.29x10^3$	0	0	0	0	0	0	0	0	0	0
Catechol	$C_6H_6O_2$	109.0289	0.94	OP	$1.29x10^3$	$2.60x10^3$	$3.07x10^3$	$8.17x10^3$	$1.15x10^4$	0	0	0	0	0	0	0	0	0	0
Morin	$C_{15}H_{10}O_7$	301.0343	9.80	F	$8.09x10^2$	$1.22x10^3$	$1.21x10^3$	$2.93x10^3$	$3.96x10^3$	0	0	0	0	0	0	0	0	0	0
PC in common among all genotypes/growing seasons (n = 26)																			
Puerarin	$C_{21}H_{20}O_9$	415.1027	9.94	F	0	0	0	0	$7.13x10^5$	0	0	0	$2.76x10^6$	$2.26x10^6$	0	$1.18x10^3$	0	$2.60x10^6$	$2.64x10^6$
4-Hydroxybenzaldehyde*	**$C_7H_6O_2$**	**121.0289**	**6.14**	**OP**	$4.52x10^5$	$6.97x10^5$	$1.67x10^6$	$1.58x10^6$	0	$4.46x10^4$	$8.09x10^4$	$8.80x10^4$	$7.94x10^4$	$6.10x10^4$	$5.35x10^4$	$4.19x10^4$	$8.18x10^4$	0	$2.22x10^5$
Caffeic acid*	**$C_9H_8O_4$**	**179.0343**	**6.24**	**PA**	$9.42x10^4$	$8.05x10^4$	$1.81x10^5$	$4.14x10^5$	$5.29x10^5$	$1.17x10^4$	$3.26x10^4$	0	$4.46x10^5$	$6.08x10^5$	$2.02x10^4$	$4.08x10^4$	0	$5.10x10^5$	$3.63x10^5$
trans-ferulic acid*	**$C_{10}H_{10}O_4$**	**193.0499**	**8.00**	**PA**	$6.70x10^4$	$1.15x10^5$	$1.49x10^5$	$2.28x10^5$	$1.99x10^5$	$9.20x10^4$	$1.92x10^5$	$2.87x10^4$	$4.21x10^5$	$3.43x10^5$	$4.54x10^3$	$3.30x10^5$	$4.53x10^5$	$3.53x10^5$	$3.47x10^5$
p-coumaric acid*	**$C_9H_8O_3$**	**163.0393**	**7.48**	**PA**	$2.51x10^4$	$5.07x10^4$	$1.28x10^5$	$6.56x10^4$	$2.38x10^5$	$6.97x10^3$	$1.01x10^5$	$5.60x10^4$	$2.88x10^5$	$4.68x10^4$	$1.03x10^4$	$1.31x10^5$	$4.48x10^4$	$3.20x10^5$	$4.95x10^5$
Scutellarein*	$C_{15}H_{10}O_6$	285.0392	10.16	F	$8.49x10^2$	$1.21x10^4$	$1.66x10^4$	0	$1.76x10^3$	0	$2.54x10^4$	0	0	$2.96x10^5$	0	$4.65x10^4$	$1.71x10^5$	$4.09x10^5$	$5.54x10^5$
3'-Hydroxymelanettin*	$C_{16}H_{12}O_6$	299.0551	11.03	F	$2.98x10^2$	0	$4.96x10^3$	0	0	$5.83x10^3$	$1.03x10^3$	$2.45x10^3$	0	0	$7.92x10^3$	$3.10x10^4$	$2.19x10^5$	$3.35x10^5$	$6.93x10^5$
Esculetin*	$C_9H_6O_4$	177.0185	6.10	OP	$2.20x10^3$	$2.00x10^3$	$3.73x10^3$	$1.54x10^4$	$3.60x10^4$	$4.56x10^3$	$1.09x10^4$	0	$2.50x10^5$	$3.58x10^5$	$4.22x10^3$	$1.09x10^4$	$4.64x10^4$	$1.84x10^5$	$2.72x10^5$
Ferulic acid*	$C_{10}H_{10}O_4$	193.0498	8.25	PA	$3.19x10^4$	$2.05x10^3$	$7.31x10^3$	$9.90x10^4$	$9.43x10^4$	$4.36x10^4$	$3.70x10^3$	$9.57x10^3$	0	$1.41x10^5$	0	$1.65x10^5$	$2.54x10^5$	$1.55x10^5$	$1.68x10^5$
Naringenin 7-O-glucoside*	$C_{21}H_{22}O_{10}$	433.1128	7.84	F	$1.33x10^4$	$2.52x10^4$	$5.29x10^4$	$8.95x10^4$	$7.92x10^4$	0	0	$4.75x10^4$	0	$4.83x10^4$	0	$3.80x10^4$	$6.59x10^4$	$5.78x10^4$	$6.90x10^4$
4-Hydroxybenzoic acid*	**$C_7H_6O_3$**	**137.0237**	**5.03**	**PA**	$1.37x10^4$	$3.06x10^4$	$7.30x10^4$	$1.83x10^5$	$1.44x10^5$	$8.51x10^3$	$4.56x10^3$	$2.79x10^3$	$1.12x10^4$	$1.35x10^4$	0	$7.17x10^3$	$4.79x10^3$	0	$3.59x10^4$
Dihydroxybenzoic acid isomer I*	$C_7H_6O_4$	153.0186	3.46	PA	$1.70x10^4$	$3.39x10^4$	$6.34x10^4$	$1.78x10^5$	$2.29x10^5$	0	0	0	$2.03x10^3$	$3.72x10^3$	$5.68x10^2$	$1.08x10^3$	0	$2.76x10^3$	0
Hesperidin*	$C_{28}H_{34}O_{15}$	609.1881	0.57	F	$4.00x10^3$	$6.16x10^3$	$1.17x10^4$	$1.98x10^4$	$2.36x10^4$	$4.36x10^3$	$1.53x10^4$	$2.16x10^4$	$2.25x10^4$	$2.29x10^4$	$4.74x10^3$	$1.50x10^4$	$1.79x10^4$	$2.11x10^4$	$2.21x10^4$
5-Caffeoylquinic acid*	$C_{16}H_{18}O_9$	353.0865	5.87	PA	$7.03x10^4$	0	$2.34x10^4$	$1.76x10^4$	$1.44x10^4$	0	$3.17x10^3$	0	$8.18x10^3$	0	$9.82x10^3$	$1.22x10^4$	$9.24x10^3$	0	$2.83x10^4$
p-anisaldehyde	$C_8H_8O_2$	135.0444	7.07	OP	$3.48x10^2$	$7.94x10^2$	0	0	0	0	0	0	0	$8.75x10^3$	$9.99x10^2$	$1.35x10^4$	0	0	$1.63x10^5$

3-Feruloylquinic acid*	$C_{17}H_{20}O_9$	367.1023	6.07	PA	$1.17x10^4$	0	$1.10x10^4$	0	$1.26x10^4$	$1.05x10^4$	$2.58x10^4$	0	$1.82x10^4$	0	$2.72x10^4$	$2.15x10^4$	0	$2.64x10^4$	$1.76x10^4$
Tetramethoxyflavone isomer III*	$C_{19}H_{18}O_6$	341.1019	10.12	F	$2.31x10^3$	0	0	0	$7.48x10^3$	$5.03x10^3$	$1.16x10^4$	$1.59x10^4$	$2.00x10^4$	$1.98x10^4$	$3.09x10^3$	$1.14x10^4$	$1.69x10^4$	$1.88x10^4$	$2.86x10^4$
Naringin 4'-O-glucoside*	$C_{21}H_{22}O_{10}$	433.1128	8.35	F	$2.82x10^3$	$5.73x10^3$	$1.18x10^4$	$1.99x10^4$	$1.58x10^4$	0	0	$1.13x10^4$	$1.28x10^4$	$8.59x10^3$	0	$8.83x10^3$	$1.42x10^4$	$1.20x10^4$	$1.35x10^4$
Vanillin*	**$C_8H_8O_3$**	**151.0393**	**5.80**	**OP**	$2.43x10^3$	$4.07x10^3$	$8.14x10^3$	$1.44x10^4$	$1.72x10^4$	$3.05x10^3$	0	0	0	$7.22x10^3$	0	0	0	$3.38x10^4$	$3.86x10^4$
Tetramethoxyflavone isomer I*	$C_{19}H_{18}O_6$	341.1019	8.30	F	$2.08x10^3$	$3.92x10^3$	0	$7.93x10^3$	$7.42x10^3$	$2.40x10^3$	$4.57x10^3$	0	$1.45x10^4$	$1.31x10^4$	0	$6.56x10^3$	$1.12x10^4$	$1.44x10^4$	$1.48x10^4$
Feruloyl glucose*	$C_{16}H_{20}O_9$	355.1023	6.46	PA	$9.20x10^2$	0	0	0	$8.41x10^3$	0	$6.28x10^3$	0	$4.16x10^3$	0	$1.59x10^3$	$1.26x10^4$	$2.44x10^4$	0	$4.42x10^3$
Luteolin 7-O-rutinoside	$C_{27}H_{30}O_{15}$	593.1497	8.11	F	$1.09x10^3$	$2.56x10^3$	0	0	0	0	0	0	$6.86x10^3$	0	$2.01x10^3$	0	$1.56x10^4$	$9.81x10^3$	0
Tetramethoxyflavone isomer II*	$C_{19}H_{18}O_6$	341.1018	8.67	F	$5.12x10^2$	0	$1.63x10^3$	$2.27x10^3$	$2.48x10^3$	$8.45x10^2$	$1.43x10^3$	$3.17x10^3$	$5.37x10^3$	$4.84x10^3$	0	$2.16x10^3$	$4.39x10^3$	0	$6.72x10^3$
Procyanidin dimer B-type VII	$C_{30}H_{26}O_{12}$	577.1334	10.26	F	$4.26x10^2$	0	0	0	$2.20x10^3$	0	0	0	0	$1.23x10^4$	0	$1.32x10^3$	$4.91x10^3$	0	0
Phenylacetic acid	$C_8H_8O_2$	135.0444	6.51	PA	$3.73x10^2$	0	0	0	$2.69x10^3$	0	0	0	$5.41x10^3$	$3.80x10^3$	0	0	0	$5.26x10^3$	0
Procyanidin dimer B-type VI	$C_{30}H_{26}O_{12}$	577.1341	9.43	F	$4.57x10^2$	0	0	0	$2.33x10^3$	0	0	0	0	$5.80x10^3$	0	$9.60x10^2$	0	0	0

m/z = mass/charge; RT = retention time; F = flavonoids; PA = phenolic acids; OP = other polyphenols. PC in bold represents reference standards; and PC marked with the "*" symbol are in common both tables (DAF and genotype/GS). Mean of normalized relative compound abundance are shown for each sample (per grain).

Among the 20 common PC found in sorghum grains, regardless of developmental stage and genotype/GS, 7 had previously been reported as the main compounds in mature sorghum: trans-ferulic acid, caffeic acid, p-coumaric acid, 4-hydroxybenzoic acid, 4-hydroxybenzaldehyde acid, esculetin and ferulic acid (D'Almeida *et al.*, 2021b). Some derivatives of hydroxycinnamic and hydroxybenzoic acids have also been detected, e.g., dimeric hydroxybenzoic acid, ferulic acid glycosylated and esterified with quinic acid, and caffeic acid esterified with quinic acid. In addition to phenolic acids and other polyphenols class, 8 important flavonoids (5 aglycones and 2 glucones) were identified. They are synthesized by the central metabolite of flavonoid biosynthesis (naringenin) and have previously been reported in mature sorghum and other cereals, such as wheat (D'Almeida *et al.*, 2021b; Gallego-Jara *et al.*, 2021; Razgonova *et al.*, 2021).

Dimeric and trimeric procyanidins were reported among PC independent of development stage (Table 1, PC in common among all days after flowering). Sorghum is a potential source of procyanidins (condensed tannins) and they are located in the testa, the structure between the pericarp and the endosperm of the grain (Xiong *et al.*, 2019b). Its high bioactivity has already been well elucidated and is important for human health (Espitia-Hernández *et al.*, 2022a), but its negative impact on sorghum protein digestibility is still considered a problem. Although the abundance of these procyanidins increases progressively throughout the development of the grains analyzed in our study (respectively seven and six-fold higher between 10DAF and 40DAF), Table 1 confirms the presence of condensed tannins even in the initial stages of the grain.

In the case of PC in common among all genotypes/GS, it is important to highlight puerarin (Table 1, PC in common among all genotypes/GS), since it was previously reported as one of the main flavonoids in mature sorghum (D'Almeida *et al.*, 2021b; Xiong *et al.*, 2020). Synthesized by the isoflavonoid pathway via naringenin, we showed that the synthesis of puerarin occurs mainly at the final stages of grain maturation (33 and 40DAF); and it was identified as the most abundant PC in the mature grains of all genotypes/GS.

3.4. Evolution of phenolic profile during sorghum grain development

The knowledge of PC biosynthesis in sorghum is essential not only for fundamental research on phenolic characterization but also for improvement of grain quality and health benefits, *i.e.*, the elucidation of phenolic evolution mechanisms can

drive the sorghum harvest at the appropriate stages according to the needs. Globally, the number of PC identifications during the sorghum development stages exhibited low variation (between 52 and 69 PC) and irregular behavior (Figure 13A). The highest number of annotations refers to the earliest development stage (10DAF, 69 PC), mainly attributed to the 15 PC exclusively found at this stage. Among them, we found sinapic acid and quercetin (identified at level 1, confirmed with phenolic standards) and derivatives of hydroxycinnamic acids (caffeic, ferulic and p-coumaric acids).

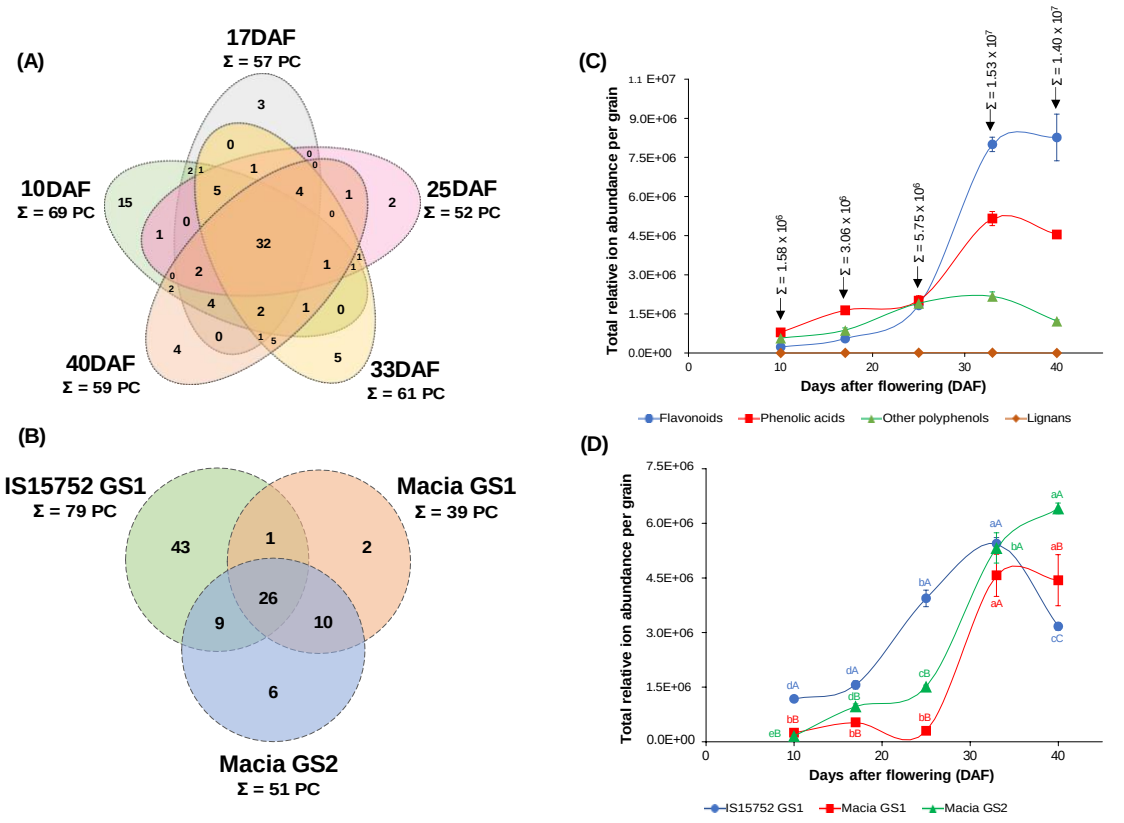


Figure 5. Metabolomic analysis. (A) Venn diagram with the number of identifications distribution in grains from different development stages. (B) Total relative ion abundance of phenolic compounds by class during grain development. (C) Total relative ion abundance of phenolic compounds in each sample during DAF. And (D) Venn diagram with the number of identifications distribution in each genotype/growing season (Wu *et al.*). Σ = sum of the total group value. Different lowercase and uppercase letters mean a significant difference ($p < 0.05$) between DAF and samples/genotypes, respectively. Bars represent standard deviation (n = 3).

The relative quantification of the annotated compounds through total ion abundance was evaluated by classes and by genotype (Figure 13C and D). Phenolic synthesis during grain development was evidenced, mainly between the initial grain developmental stage and mature grains (40DAF was nine-fold higher than 10DAF, based on the cumulative classes). In the initial stages (10 and 17DAF), the class of phenolic acids was more abundant (≅ 50%), followed by other polyphenols (≅ 30%) and flavonoids (≅ 20%). In the hard dough stage (25DAF), the abundance distribution of these three phenolic classes is equal (≅ 33% each). And in the mature stages (33 and 40DAF), flavonoids become the predominant class in this cereal (3/5 of total ion abundance). The abundance of lignans were inexpressive throughout the grain development (Supplementary Table 2).

As previously mentioned, PC are derived from the phenylpropanoid biosynthetic pathway, which has as its first step the conversion of phenylalanine and tyrosine to cinnamic and coumaric acid, respectively, by phenylalanine/tyrosine ammonia-lyase (Ma *et al.*, 2022). It can be hypothesized that due to the high presence of free amino acids at the early stages of the grain development (Ma *et al.*, 2024), PC synthesis will be favored throughout its development. The data presented here with sorghum and those previously reported on other developing cereal grains suggest a prevalence of the phenolic acid synthesis in early stages (Kim *et al.*, 2016; Lin & Lai, 2011; Santos *et al.*, 2019a). From 25DAF, the flavonoid biosynthesis pathway seems to be prioritized, which indicates the hypothesis of greater action of the enzyme 4-coumarate-CoA ligase. This enzyme converts cinnamic and coumaric acid, respectively, into cinnamoyl-CoA or p-coumaroyl-CoA, both precursors of this pathway.

To further explore the variation in the data set, the PCA biplot (scores representing the samples and loadings phenolic compounds) was applied to investigate the degree of similarity or dissimilarity between genotypes and during grain development (Figure 14A). Due to the large number of samples (two genotypes, five stages and two GS), this first PCA biplot was not efficient to visualize the phenolic variability (PC1 + PC2 axis = 45%), but some results can be clearly observed: 1) the distinction between genotypes (PC1 27%), where the tannin-rich and tannin-poor samples were distributed on the right and left side of the x-axis, respectively; 2) the separation between the samples of initial

stages of grain development (10, 17 and 25 DAF on the bottom side) and after physiological maturity (33 and 40 DAF on the upper side) (PC2 18%), and 3) the left side of the x-axis showed the low influence of GS on phenolic variability.

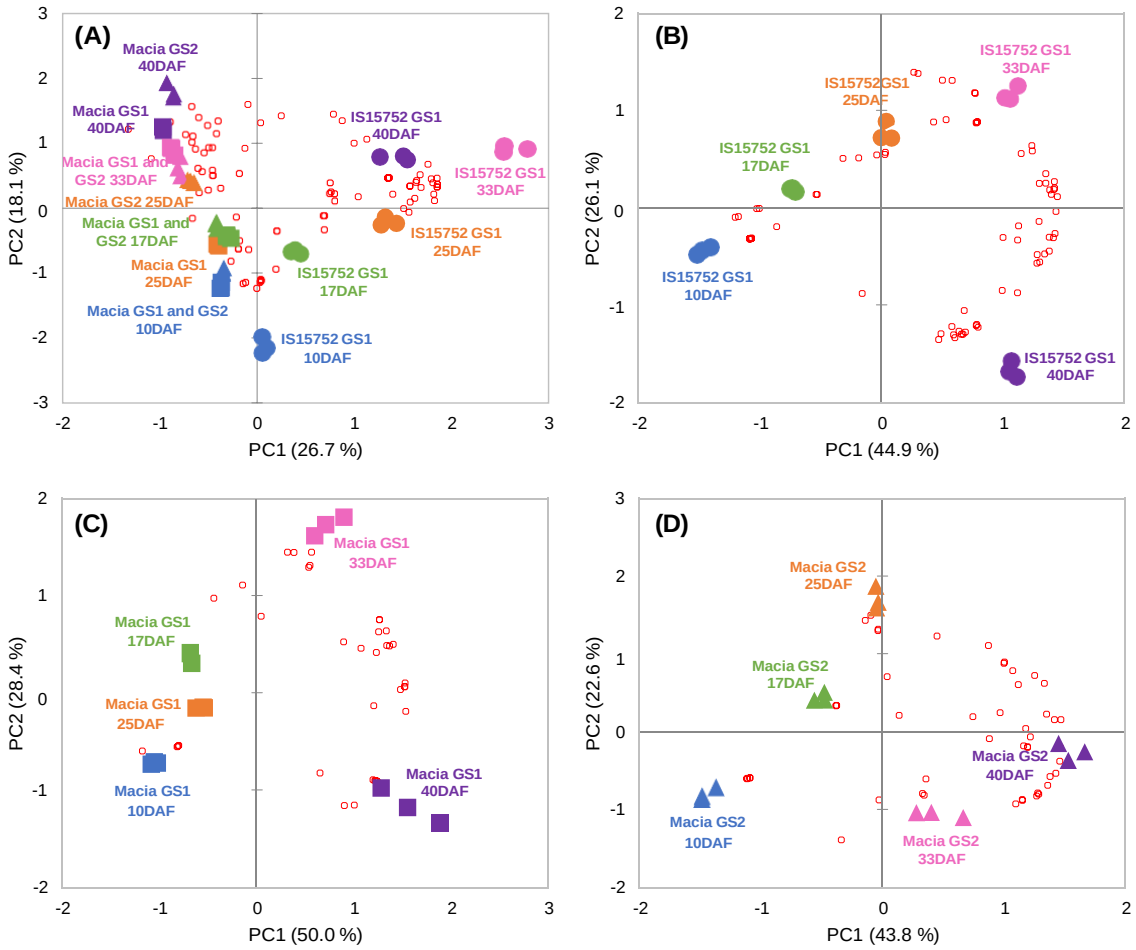


Figure 6. Principal component analysis (PCA) biplot: (A) of all sorghum samples, and (B-D) in each genotype/growing season (Wu *et al.*). The samples (symbols) are distributed according to relative intensity of phenolic compounds (red empty circles).

Each genotype was also evaluated separately, showing a clear trend towards distinction between grain development stages (Figure 14B-D). In this case, IS15752 GS1 (PC1 and PC2: 71%) and Macia GS2 (PC1 and PC2: 66%) showed the same behavior, i.e., half-moon distribution of scores, with earliest stages (10 and 17DAF) located in the left quadrant, mature stages (33 and 40DAF) in the right quadrant, and the intermediate stage (25DAF) centered on the x-axis (Figure 14B and 3D). Macia GS1 showed a similar profile with a higher variance explained by PC1 and PC2 (78%), however with a slight difference when compared to the other samples, the 25DAF score was grouped together immature stages (Figure 14C).

Hierarchical cluster analysis (HCA) with the correlation matrix (heatmap) were applied with all 97 annotated PC for a better visualization of the different stages and to identify which PC can discriminate each one (Figure 15). Firstly, the vertical HCA separated the samples into two large groups: early stages (10-25DAF) and mature stages (33-40DAF), which corroborates the distribution of scores previously found in the PCA (Figure 14B-D). In the PC characterization, HCA formed three large groups (horizontal axis, G1-G3; Supplementary Table 3) that can be further subdivided (a, b and c): 1) G1 is represented by 23 PC, more abundant in the early stages of grain development (10-17 DAF). Globally, these PC belong to the classes of phenolic acids (52%), followed by flavonoids (30%) and other polyphenols (9%), based on total relative ion abundance (Supplementary Table 3). These phenolic acids are present in G1a and presented an abundance reduction at the beginning of the soft dough stage, while G1b presented varied composition and also present an abundance reduction in the next stage (25DAF). 2) G2 corresponds to intermediate PC at grain maturation, i.e., an intersection between early stages (represented by 25DAF) and mature stages (represented by 33DAF). This group represents the beginning of the flavonoid synthesis (increase from 30% to 62%) and the reduction of the synthesis of phenolic acids (from 52% to 29%). These alterations in PC profile can occur at 25 DAF (G2a), at 33 DAF (G2b) or at both stages (G2c). 3) G3 had the highest number of PC (n=53, Supplementary Table 3) and is represented by the PC synthesized all along grain development, with greater accumulation in mature grains. These compounds are mostly flavonoids (60%), followed by phenolic acids (25%) and other polyphenols (15%). Furthermore, this large group presented three subdivisions, showing that these PC can be progressively synthesized (G3a); synthesized up to 33DAF, with stabilization at 40DAF (G3b); or they can be synthesized up to 33DAF, followed by

a reduction in 40DAF (G3c). The HCA multivariate analysis, therefore, corroborated the data reported in Figure 2B, and reinforced the hypothesis of alteration in the route of the phenylpropanoid biosynthetic pathway.

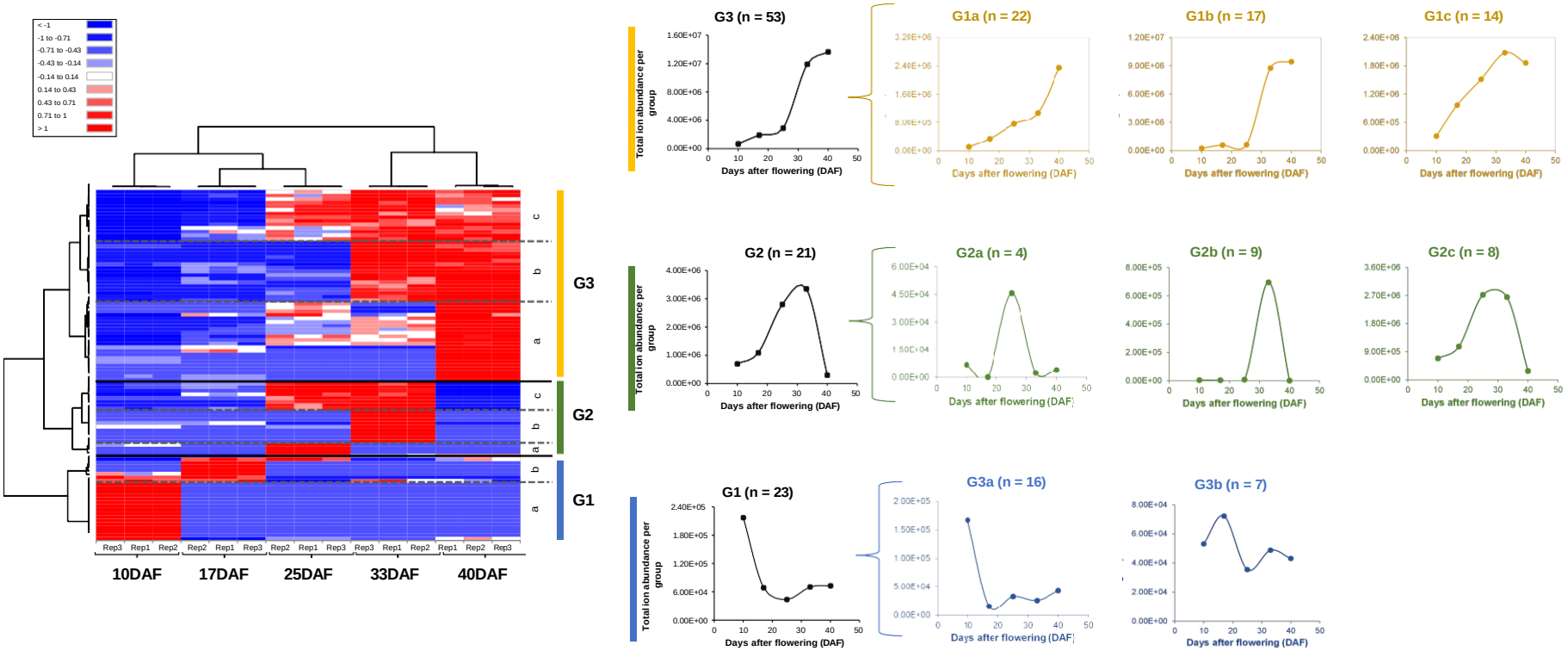


Figure 7. Hierarchical clustering (HCA) heat map of metabolomic data. Three cluster groups (G1-G3) and 3 sub-clusters (a, b and c) were generated using a Pearson correlation (ANOVA $p<0.05$) on the differentially abundant phenolic compounds during grain development. Different clusters and sub-clusters are expressed by the mean of the group total abundance.

Finally, the PC that made important contributions to the classification into two large groups formed by the PCA and HCA (early stages and mature grains) could be selected based on the projection of the importance of the variable (VIP) (Supplementary Figure 1). According to the online KEGG pathway database (www.genome.jp/kegg/pathway), a schematic diagram of phenylpropanoid and flavonoids pathway was created with VIP compounds to explain the main degradation/synthesis pathways of these compounds (Supplementary Figure 2). Quercetin, sinapic acid, ethyl gallate and phloridzin appear as relevant compounds in the early stages of grain development, the first three being specific to the milky stage. Phloridzin is a flavonoid widely reported in plants and has multiple pharmacological effects (Tian *et al.*, 2021). During its biosynthesis, the action of chalcone synthase produces the intermediate compound phloretin, which is converted into phloridzin by glucosylation (Tian *et al.*, 2021). The decrease of this flavonoid in mature sorghum grains may be related to new routes by p-coumaroyl-CoA during grain maturation, e.g., the synthesis of p-coumaroyl glucose (five-fold higher in mature stages) and/or synthesis of the flavonoids shown in Supplementary Figure 1 (twenty-two-fold higher in mature stages).

3.5. Growing season and genotype impact on phenolic compounds during sorghum grain growth

The number of identifications and the relative quantification by total ion abundance in each GS and/or genotype is shown in Figure 13B, D. As expected, tannin-rich sorghum grains had the highest number of PC, mainly due to its specific PC (54% of the total number of identifications). Among these compounds, several flavonoids have been reported, such as procyanidin dimers and trimers. Macia GS1 and GS2 have 36 PC in common, corroborating the low variability between GS or the low impact of GS on PC accumulation of the same genotype shown by PCA (Figure 14A).

The number of identifications showed strong correlation with the total relative quantification ($r = 0.84$, $p < 0.05$; data not shown). When analyzing the evolution of the relative abundance of these PC during grain growth (Figure 13D), a dissimilar behavior was observed between the variables: 1) Macia GS2 showed a significant and progressive synthesis of PC (40DAF forty-one-fold higher than 10DAF); 2) IS15752 showed a significant and progressive synthesis up to 33DAF, followed by decrease (42%) in the

final mature stage; 3) Macia GS1 showed the synthesis starting from 25DAF reaching the maximum at 33DAF (fifteen-fold higher than 25DAF).

The OPLS-DA was also applied in the phenolic profile between GS (Macia GS1 *vs* Macia GS2) and genotypes (IS15752 GS1 *vs* Macia GS1) impact (Figure 16). The OPLS-DA model parameters were robust in early stages (GS: R2Y = 0.998, Q2 = 0.987; genotype: R2Y = 1.000, Q2 = 0.998) and mature stages (GS: R2Y = 0.998, Q2 = 0.977; genotype: R2Y = 1.00, Q2 = 0.996) samples. Thus, a total of 10 PC in each S-plot, that made important contributions to the classification of the samples, could be selected based on the VIP and p-value of the OPLS-DA model. It is noteworthy that from these selected PC in GS and genotype variables, six compounds were found simultaneously on early and mature stages (Insert table – Figure 16): 4-hydroxybenzaldehyde, esculetin, naringenin 7-O-glucoside, scutellarein, 3'-hydroxymelanettin and ferulic acid for GS influence; and 4-hydroxybenzoic acid, dihydroxybenzoic acid isomer I, procyanidin dimer B-type I, dihydrocaffeic acid, (+)-catechin and 4-hydroxybenzaldehyde for genotype influence.

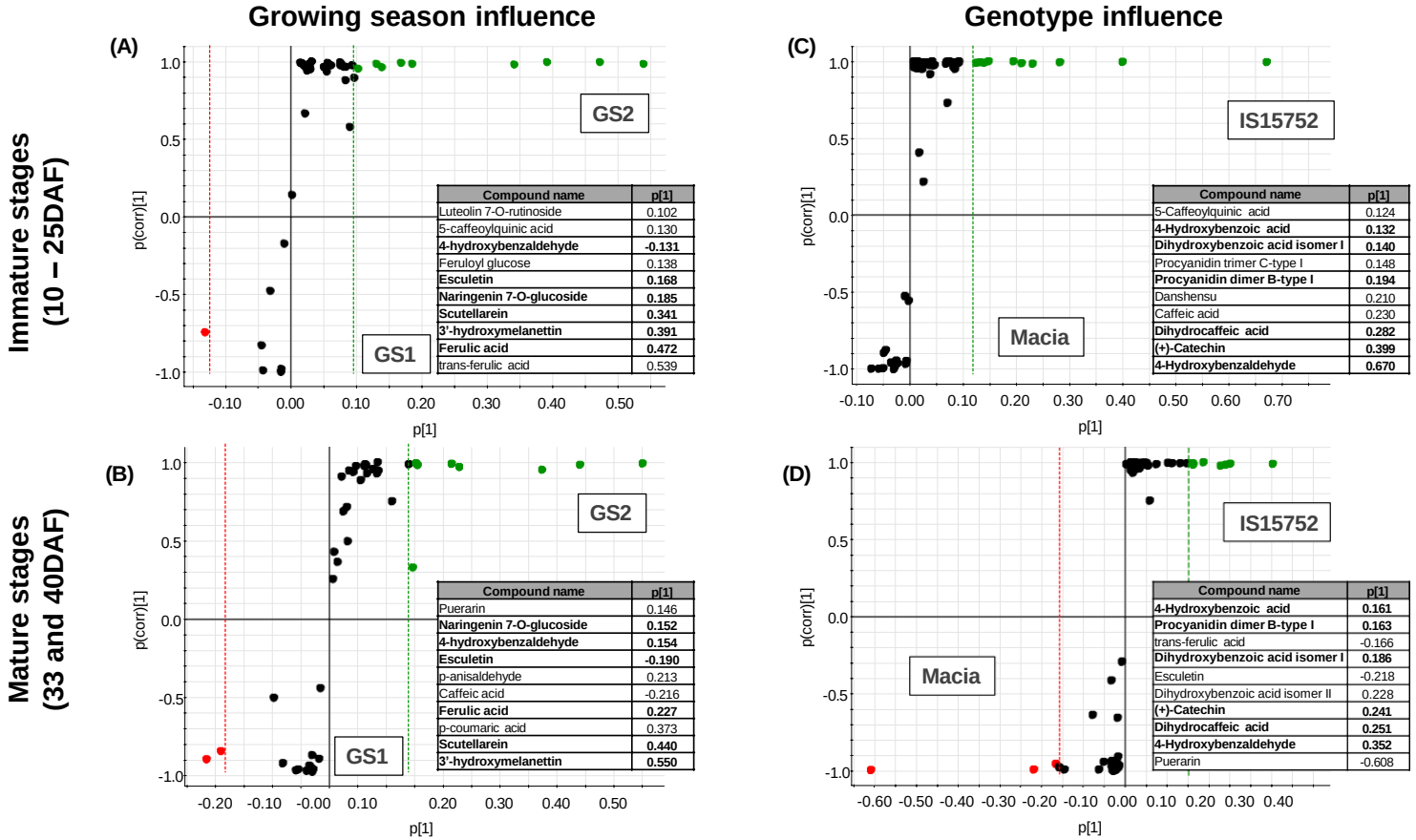


Figure 8. S-plot of orthogonal partial least squares discriminant analysis (OPLS-DA) between growing season (Wu *et al.*) influence (Macia GS1 *vs* Macia GS2) and genotype influence (Macia GS1 *vs* IS15752 GS1) in developing grains. In x-axis is represented the relative magnitude of variables (phenolic compounds), and in y axis, the confidence/reliability. Variables farthest from the origin in the plot have higher covariance (*p*[1]) and deemed significant markers. Insert tables shows the phenolic compound name in ascending order of covariance.

Quantitative and qualitative occurrence of plant PC can be associated with several agronomical important phenotypic, i.e., the phenolic diversity between the same genotype of a grain cultivated in different GS may indicate possible biotic and abiotic stresses (Hamany Djande *et al.*, 2021). Evaluating GS influence (Figure 16 A-B), four PC have similar behavior, regardless of the development stage analyzed (early or mature stages). These compounds were characteristic and of GS2 (green circles) (Fig. 5A): $[M-H]^-$ 7.84 *m/z* 433.1128, $[M-H]^-$ 8.25 *m/z* 193.0498, $[M-H]^-$ 10.16 *m/z* 285.0392 and $[M-H]^-$ 11.03 *m/z* 299.0551; respectively identified as naringenin 7-O-glucoside, ferulic acid, scutellarein and 3'-hydroxymelanettin. Šamec *et al.* (2021) emphasized the heightened susceptibility of the flavanone subclass (e.g., naringenin 7-O-glucoside) to heat stress, attributing this vulnerability to the presence of two hydroxyl groups in its B ring. Ferulic acid has also previously been pointed as a differentially abundant compound among GS, as its increase is inversely proportional to drought stress (Itam *et al.*, 2020). Although sorghum is known as a tolerant crop, the hypothesis is that Macia GS1 experienced an abiotic stress during its cultivation and that was enough to change the phenolic profile of this sample (Šamec *et al.*, 2021).

Finally, regarding the comparison between different genotypes of the same GS, phenolic acids were the major discriminant metabolites (57%), followed by flavonoids (36%) and other polyphenols (7%) (Figure 16 C-D). As expected, the presence of flavanols (characteristic of the tannin-rich genotype IS15752 GS1, green circles) was crucial for the differentiation between genotypes. Among them, we can mention monomer and oligomers of tannins such as catechin – the most commonly monomer reported in sorghum grains - and dimers and trimers of procyanidins. Although the literature mostly associates these compounds with antinutritional factors, tannins have high health-promoting ability and their dimers are well absorbed by the human body (Cardoso *et al.*, 2017). They also appear to play an important role in the food industry, particularly as high-value ingredients to naturally modify and expand protein functionality (Girard *et al.*, 2018; Girard *et al.*, 2016). Evidence indicates that polymeric sorghum tannins can drastically alter the rheological behavior of gluten in blended flours, being able to increase gluten-force (Girard *et al.*, 2016).

4. Conclusion

This is the most up-to-date work to show a comprehensive study of the synthesis of phenolic compounds in developing sorghum grain. Phenolic content increased during grain filling and showed a higher relative abundance in mature grains compared to developing immature grains. The flavonoids biosynthesis pathway seems to be prioritized from 25DAF and it may be the reason for mature grains higher antioxidant activity. Metabolomic approach also revealed the presence of hydroxycinnamic and hydroxybenzoic acids, as well as their derivatives, at all grain stages, except for lignans that was not identified in mature grains.

Chemometric analysis showed discriminatory compounds between stage, genotypes, and growing seasons. Genotype had more impact on phenolic profile variability, mainly due to the high presence of proanthocyanidins. Monomers, dimers and trimers of procyanidins specific of the tannin-rich genotype was also annotated in the tannin-free genotype. Growing season seems to less influence the polyphenol content but has important biomarkers in this differentiation, e.g., 4-hydroxybenzaldehyde. Untargeted metabolomics revealed the complex development of phenolic compounds in sorghum grains, which can contribute to open polyphenol databases and encourage greater exploitation by the agri-food industry to obtain health-promoting grains by selecting genotypes and developing stages with an optimized composition in bioactive compounds.

5. Acknowledgments and funding

The authors are grateful to Fabien DeBellis and David Pot for the help in setting up field trail. This work was financed by UNIRIO, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (427116/2018-0; 310343/2019-4; 407350/2023-3) and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (26/210.716/2021; E-26/201.317/2022; E-26/210.152/2023, E-26/210.008/2024).

6. Supplementary material

Supplementary Table 1. All phenolic compounds tentatively identified by UHPLC-MS^E in retention time order.

Please refer to the PDF file attached to the manuscript.

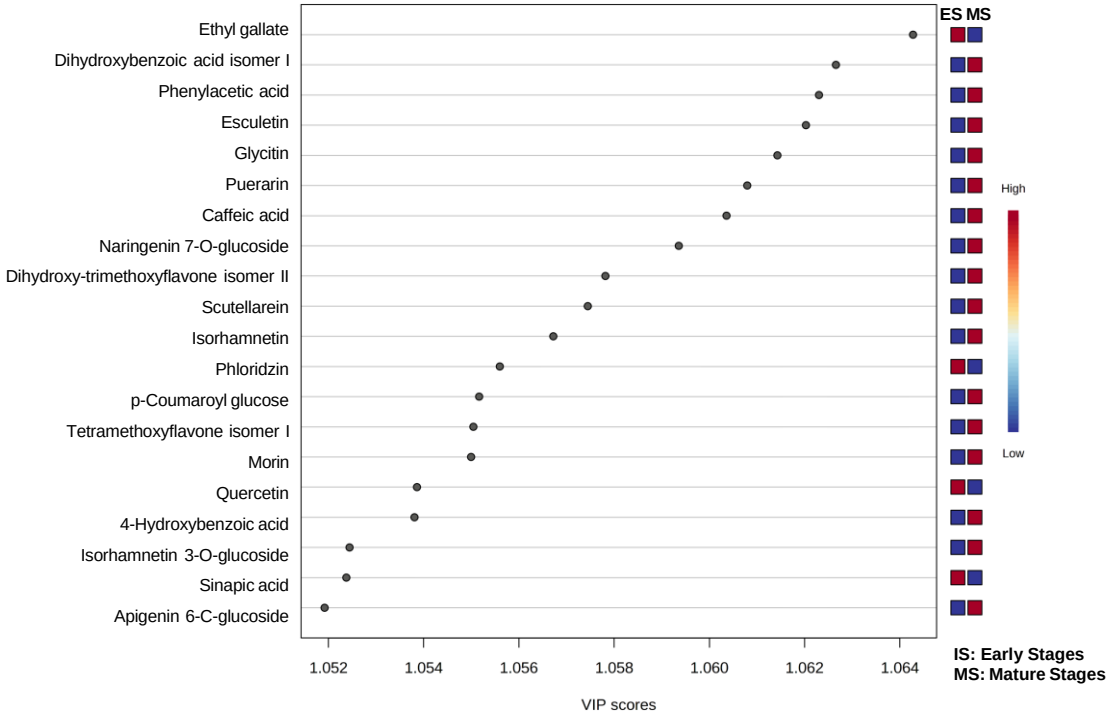
Supplementary Table 2. List of phenolic compounds grouped according to the hierarchical cluster analysis (HCA) presented in Figure 4.

Compound name	Class	*m/z*	RT (min)	10DAF	17DAF	25DAF	33DAF	40DAF
Group 1 (n = 23)								
Nortrachelogenin/7-Hydroxymatairesinol/Isohydroxymatairesinol	L	373.1299	0.60	7.65E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Ethyl gallate	PA	197.0445	2.10	2.94E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Caffeoylquinic acid isomer I	PA	353.0868	4.14	2.40E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00
3,4-Dihydroxyphenylglycol	OP	169.0497	4.21	0.00E+00	8.84E+02	0.00E+00	0.00E+00	0.00E+00
2,5-Dihydroxybenzoic acid	**PA**	**153.0186**	**5.05**	4.93E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Caffeoylquinic acid isomer II	PA	353.0867	5.46	8.34E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Paeonol	OP	165.0551	5.55	0.00E+00	1.12E+03	0.00E+00	0.00E+00	0.00E+00
p-Coumaric acid 4-O-glucoside	PA	325.0915	5.82	6.26E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00
5-Caffeoylquinic acid	PA	353.0865	5.87	8.02E+04	1.54E+04	3.27E+04	2.58E+04	4.27E+04
3-Feruloylquinic acid	PA	367.1023	6.07	4.95E+04	4.73E+04	1.10E+04	4.46E+04	3.02E+04
Feruloyl glucose	PA	355.1023	6.46	2.51E+03	1.89E+04	2.44E+04	4.16E+03	1.28E+04
Caffeoylquinic acid isomer III	PA	353.0863	6.64	2.82E+04	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Feruloylquinic acid isomer	PA	367.1024	6.96	1.06E+03	0.00E+00	0.00E+00	0.00E+00	0.00E+00
4-Hydroxyphenylacetic acid	**PA**	**151.0393**	**7.16**	3.71E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Dihydroquercetin 3-O-rhamnoside	F	449.1080	7.32	0.00E+00	1.13E+03	0.00E+00	0.00E+00	0.00E+00
Eriodictyol	F	287.0557	8.00	5.28E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Quercetin	**F**	**301.0347**	**8.01**	3.73E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Sinapic acid	**PA**	**223.0602**	**8.09**	3.15E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Phloridzin	F	435.1281	8.55	5.01E+02	6.98E+02	0.00E+00	0.00E+00	0.00E+00
Isorhamnetin 3-O-rutinoside	F	461.1071	8.59	5.53E+02	2.20E+03	0.00E+00	0.00E+00	0.00E+00
Neoeriocitrin	F	595.1658	8.69	4.39E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Trihydroxyisoflavanone isomer	F	271.0600	9.59	4.38E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00
NI	NI	191.0343	10.16	5.45E+02	0.00E+00	0.00E+00	0.00E+00	0.00E+00
Group 2 (n = 21)								
Dihydroformononetin	F	269.0835	0.55	0.00E+00	0.00E+00	1.26E+03	1.95E+03	0.00E+00
Danshensu	PA	197.0448	3.10	3.33E+04	5.94E+04	1.66E+05	2.53E+05	0.00E+00
(+)-Catechin	**F**	**289.0710**	**5.91**	9.35E+04	1.68E+05	6.62E+05	6.78E+05	0.00E+00
4-Hydroxybenzaldehyde	**OP**	**121.0289**	**6.14**	5.50E+05	8.20E+05	1.84E+06	1.66E+06	2.83E+05
4-Hydroxymandelic acid	PA	167.0342	6.14	7.98E+03	1.18E+04	2.95E+04	2.91E+04	0.00E+00
Dihydroquercetin	F	303.0505	6.34	0.00E+00	0.00E+00	1.31E+03	0.00E+00	0.00E+00
4-p-Coumaroylquinic acid	PA	337.0917	6.85	1.85E+03	0.00E+00	1.19E+03	7.42E+03	0.00E+00
Dihydroxybenzoic acid isomer II	PA	153.0187	7.05	9.52E+02	1.69E+03	5.95E+03	6.14E+05	0.00E+00
2,3,4-trihydroxybenzoic acid	PA	169.0134	7.05	0.00E+00	0.00E+00	0.00E+00	9.59E+03	0.00E+00
Cinnamoyl glucose	PA	309.0998	7.12	0.00E+00	0.00E+00	1.33E+03	1.97E+03	0.00E+00
Eriodictyol 7-O-glucoside	F	449.1081	7.50	0.00E+00	0.00E+00	2.13E+03	0.00E+00	0.00E+00
Luteolin 7-O-rutinoside	F	593.1497	8.11	3.10E+03	2.56E+03	1.56E+04	1.67E+04	0.00E+00
Glycitein	F	283.0601	8.43	0.00E+00	0.00E+00	0.00E+00	3.56E+03	0.00E+00
Hesperetin	F	301.0710	9.01	6.47E+03	0.00E+00	3.31E+04	0.00E+00	0.00E+00

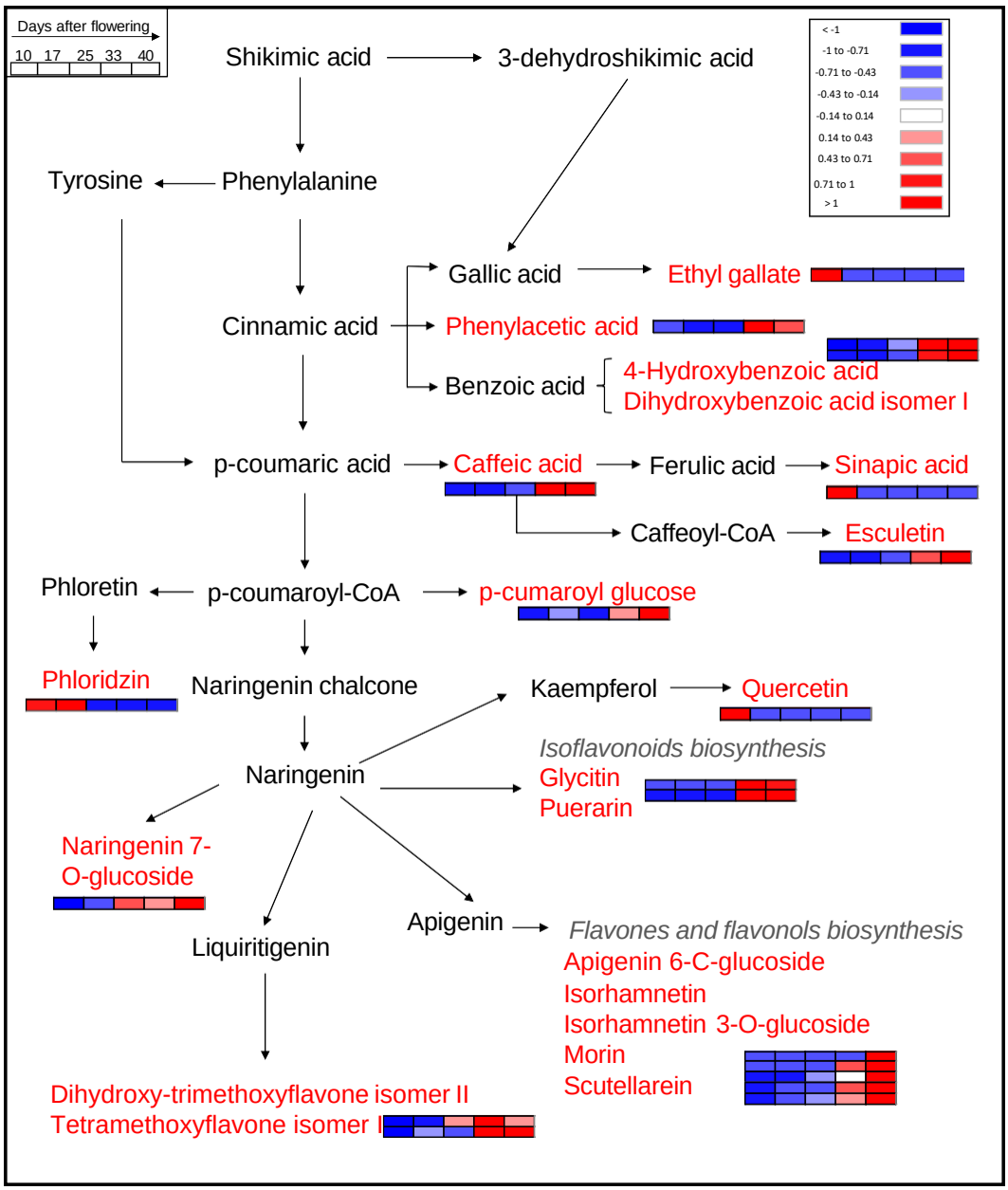
Procyanidin dimer B-type V	F	577.1548	9.01	0.00E+00	0.00E+00	0.00E+00	2.67E+04	7.69E+01
Homoeriodictyol	F	301.0702	9.64	0.00E+00	0.00E+00	0.00E+00	6.27E+03	0.00E+00
Sakuranetin	F	285.0757	9.89	0.00E+00	0.00E+00	0.00E+00	1.08E+04	0.00E+00
1-Acetoxypinoresinol	L	415.1394	10.50	0.00E+00	4.78E+03	9.91E+03	1.08E+04	0.00E+00
Jaceosidin	F	329.0655	10.96	9.53E+02	1.47E+03	0.00E+00	6.28E+03	0.00E+00
Nepetin	F	315.0500	11.19	0.00E+00	0.00E+00	0.00E+00	1.09E+04	0.00E+00
Violanone	F	315.0857	12.19	3.40E+02	0.00E+00	9.22E+03	2.40E+03	3.99E+03
Group 3 (n = 53)								
Hesperidin	F	609.1881	0.57	1.31E+04	3.65E+04	5.12E+04	6.34E+04	6.86E+04
Catechol	OP	109.0289	0.94	1.29E+03	2.60E+03	3.07E+03	8.17E+03	1.15E+04
Dihydroxybenzoic acid isomer I	PA	153.0186	3.46	1.75E+04	3.50E+04	6.34E+04	1.83E+05	2.33E+05
3,4-Dihydroxyphenylacetic acid	PA	167.0343	3.80	0.00E+00	5.32E+03	7.34E+03	1.18E+04	1.25E+04
Methylcatechol isomer	OP	123.0445	4.21	0.00E+00	2.14E+03	2.29E+03	3.69E+03	3.19E+03
Dihydrocaffeic acid	PA	181.0500	4.46	6.56E+04	1.23E+05	2.74E+05	3.84E+05	3.47E+05
Homovanillic acid	PA	181.0499	4.73	1.36E+04	2.59E+04	0.00E+00	7.82E+04	6.69E+04
4-Hydroxybenzoic acid	**PA**	**137.0237**	**5.03**	2.22E+04	4.23E+04	8.06E+04	1.94E+05	1.93E+05
Procyanidin dimer B-type I	F	577.1338	5.42	2.12E+04	3.50E+04	1.63E+05	1.86E+05	1.25E+05
Procyanidin trimer C-type I	F	865.1963	5.76	1.23E+04	2.04E+04	9.39E+04	1.11E+05	9.14E+04
Vanillin	**OP**	**151.0393**	**5.80**	5.48E+03	4.07E+03	8.14E+03	4.82E+04	6.30E+04
Coumaroyl Hexoside	PA	325.0910	5.96	7.78E+02	3.48E+03	0.00E+00	7.11E+03	1.55E+04
Esculetin	OP	177.0185	6.10	1.10E+04	2.38E+04	5.01E+04	4.49E+05	6.66E+05
Procyanidin trimer C-type II	F	865.1956	6.14	1.03E+03	1.57E+03	8.07E+03	1.08E+04	6.29E+03
Caffeic acid	**PA**	**179.0343**	**6.24**	1.26E+05	1.54E+05	1.81E+05	1.37E+06	1.50E+06
Phenylacetic acid	PA	135.0444	6.51	3.73E+02	0.00E+00	0.00E+00	1.07E+04	6.49E+03
Umbelliferone	OP	161.0236	6.60	0.00E+00	0.00E+00	0.00E+00	6.12E+03	5.04E+03
Eriodictyol 7-O-glucoside	F	449.1073	6.95	2.65E+03	5.41E+03	2.07E+04	3.57E+04	2.96E+04
p-coumaric acid	**PA**	**163.0393**	**7.48**	4.24E+04	2.83E+05	2.28E+05	6.73E+05	7.80E+05
Procyanidin dimer B-type III	F	577.1335	7.48	4.13E+03	6.10E+03	2.15E+04	2.02E+04	1.55E+04
trans-ferulic acid	**PA**	**193.0499**	**8.00**	1.63E+05	6.36E+05	6.30E+05	1.00E+06	8.88E+05
Quercetin 3-O-rutinoside	F	609.1447	8.07	5.30E+03	7.21E+03	1.50E+04	1.48E+04	1.45E+04
Tetramethoxyflavone isomer I	F	341.1019	8.30	4.49E+03	1.50E+04	1.12E+04	3.68E+04	3.54E+04
Naringin 4'-O-glucoside	F	433.1128	8.35	2.82E+03	1.46E+04	3.73E+04	4.47E+04	3.79E+04
Morin	F	301.0343	9.80	8.09E+02	1.22E+03	1.21E+03	2.93E+03	3.96E+03
Dihydroxy-trimethoxyflavone isomer II	F	343.0811	9.93	3.72E+03	1.01E+04	2.48E+04	4.27E+04	2.61E+04
Puerarin	F	415.1027	9.94	0.00E+00	1.18E+03	0.00E+00	5.36E+06	5.62E+06
Glycitin	F	445.1126	10.55	0.00E+00	0.00E+00	0.00E+00	3.23E+05	2.21E+05
Isorhamnetin	F	315.0501	10.92	0.00E+00	0.00E+00	0.00E+00	9.47E+03	1.47E+04
Koparin	F	299.0551	11.85	0.00E+00	0.00E+00	0.00E+00	8.18E+03	6.31E+03
Vanillactic acid	PA	211.0608	4.25	0.00E+00	0.00E+00	0.00E+00	0.00E+00	4.55E+03
Ferulaldehyde	OP	177.0185	5.58	1.34E+03	2.08E+03	6.58E+03	0.00E+00	1.37E+04
Methoxyphenylacetic acid	PA	165.0550	6.00	2.86E+02	3.06E+03	0.00E+00	0.00E+00	8.34E+03
p-anisaldehyde	OP	135.0444	7.07	1.35E+03	1.43E+04	0.00E+00	0.00E+00	1.72E+05
Naringenin 7-O-glucoside	F	433.1128	7.84	1.33E+04	6.32E+04	1.66E+05	1.47E+05	1.97E+05
Ferulic acid	PA	193.0498	8.25	7.55E+04	1.71E+05	2.71E+05	2.54E+05	4.03E+05

NI	F	771.1914	8.30	0.00E+00	2.49E+03	3.31E+03	1.83E+03	4.26E+03
6-hydroxyluteolin	F	301.0343	8.35	1.26E+03	0.00E+00	0.00E+00	0.00E+00	8.54E+03
Chrysoeriol 7-O-apiosyl-glucoside	F	593.1497	8.53	4.26E+02	7.86E+02	0.00E+00	0.00E+00	2.67E+03
Tetramethoxyflavone isomer II	F	341.1018	8.67	1.36E+03	3.59E+03	9.19E+03	7.64E+03	1.40E+04
Isorhamnetin 3-O-glucoside	F	477.1026	8.85	1.00E+03	1.81E+03	7.29E+03	1.05E+04	2.57E+04
Apigenin 6-C-glucoside	F	431.0976	8.87	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.31E+04
Tectoridin	F	461.1072	9.04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	3.36E+03
Dihydroxy-trimethoxyflavone isomer I	F	343.0811	9.05	0.00E+00	1.78E+03	3.28E+03	4.66E+03	1.26E+04
Procyanidin dimer B-type VI	F	577.1341	9.43	4.57E+02	9.60E+02	0.00E+00	0.00E+00	8.13E+03
Rhamnetin	F	315.0500	9.55	8.55E+02	0.00E+00	0.00E+00	0.00E+00	6.46E+03
Tetramethoxyflavone isomer III	F	341.1019	10.12	1.04E+04	2.30E+04	3.29E+04	3.88E+04	5.59E+04
Scutellarein	F	285.0392	10.16	8.49E+02	8.40E+04	1.88E+05	4.09E+05	8.51E+05
Procyanidin dimer B-type VII	F	577.1334	10.26	4.26E+02	1.32E+03	4.91E+03	0.00E+00	1.45E+04
Coumestrol	OP	267.0289	10.36	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.24E+03
Dihydroxy-trimethoxyflavone isomer III	F	343.0812	10.59	4.29E+02	1.96E+03	6.73E+03	1.12E+04	3.08E+04
Procyanidin dimer B-type VIII	F	577.1333	10.71	0.00E+00	0.00E+00	1.34E+03	0.00E+00	3.29E+03
3'-Hydroxymelanettin	F	299.0551	11.03	1.40E+04	3.21E+04	2.26E+05	3.35E+05	6.93E+05

m/z = mass/charge; RT = retention time; F = flavonoids; PA = phenolic acids; OP = other polyphenols; DAF = days after flowering. Bold represent reference standards.



Supplementary Figure 1. Variable importance in projection (VIP) scores generated from orthogonal partial least squares discriminant analysis (OPLS-DA). The twenty top important phenolic compounds (VIP score > 1.5) contributing to the separation of phenolic profile in early *vs.* mature stages. The relative abundance of phenolic compounds is indicated by a colored scale from blue to red representing the low and high, respectively.



Supplementary Figure 2. Schematic diagram of phenylpropanoid and flavonoids pathway associated with the synthesis and degradation of phenolic compounds during days after flowering (DAF). The compounds in red were selected by the Variable Importance in Prediction (VIP, see Supplementary Figure 3). The small heat maps represents normalized relative metabolite abundance of grains at a different stage from 10 to 40DAF, respectively. Red color represents higher abundance, and blue color represents lower abundance.

Part III

Effect of traditional processing methods on the protein and bioactive composition of sorghum

Chapter 4 – *"Enhancing the nutritional value and bioavailability of sorghum grains bred for northern Europe through processing: A perspective on phenolic and protein digestibility"*

Chapter 5 – *"Tracking the changes and bioaccessibility of phenolic compounds of sorghum grains (Sorghum bicolor (L.) Moench) upon germination and seedling growth by UHPLC-QTOF-MS/MS"*

Chapter IV - Enhancing the nutritional value and bioavailability of sorghum grains bred for northern Europe through processing: A perspective on phenolic and protein digestibility

Carolina Thomaz dos Santos D'Almeida [1,2], Joëlle Bonicel [3], Walter de Milliano [4], Trust Beta [5], Mariana Simões Larraz Ferreira [1,2] and Hamza Mameri [3*]

[1] Laboratory of Bioactives, Food and Nutrition Graduate Program, Federal University of the State of Rio de Janeiro, UNIRIO, Brazil

[2] Center of Innovation in Mass Spectrometry, Laboratory of Protein Biochemistry, UNIRIO, Brazil.

[3] UMR IATE, Univ Montpellier, INRAE, Institut-Agro Montpellier, F-34060 Montpellier, France.

[4] Maatschap de Milliano-Meijer, Oostburg, Zeeland, Netherlands

[5] Department of Food and Human Nutritional Sciences, University of Manitoba, Winnipeg, MB, Canada R3T 2N2

Highlights

- First-time Dutch sorghum characterization, valuing its northern Europe cultivation
- *In vitro* protein digestibility impact after cooking was milder in dehulled grains
- Chemometric analyses revealed key compounds impacted during processing stages
- Thermal processing led to the depolymerization of sorghum grain procyanidins
- Cooking enhanced sorghum phenolic content, while preserving its bioaccessibility

This chapter was submitted to *Food Chemistry* (Impact Factor: 8.5).

ABSTRACT

The impact of dehulling and cooking on protein content, *in vitro* digestibility, and phenolic profiles was evaluated for four Dutch sorghum varieties (HD7, HD19, HD100, HD101) produced above 48 degrees in the Netherlands. Grain characterization showed weight and size homogeneity, although black pericarp grains were more resistant to dehulling. Protein content ranged from 9 to 14%. Essential amino acids level, determined by HPLC, revealed that the lysine chemical score (~0.6) was lower than the required for adults. High-resolution untargeted metabolomics via UHPLC-ESI-QTOF/MS allowed the annotation of 219 phenolic compounds, showing flavonoids as the most representative class (91%). Multivariate statistical analyses showed a higher dehulling and genotype impact on the phenolic profiles compared to cooking; however, thermal treatment was essential for depolymerization of proanthocyanidin dimers and trimers. The combination of dehulling and cooking was effective for enhancing protein *in vitro* digestibility and increasing *in vitro* bioaccessibility of important phenolic compounds in sorghum. Dehulling and boiling have been demonstrated to be effective processes for sorghum-based products development with acceptable quality using Dutch varieties produced above 48 degrees in the Netherlands.

Keywords: Boiling process, dehulling, *in vitro* digestion, metabolomic, protein quality, *Sorghum bicolor.*

1. Introduction

Growing awareness of health benefits obtained from certain foods or food ingredients has increasingly influenced consumer decisions, reflecting a greater focus on overall well-being. Due to its inherent functional properties, sorghum (*Sorghum bicolor* L. Moench) has become a valuable food option for individuals with celiac disease and allergies. In 2023, world sorghum production reached 60 million tons, representing the fifth most produced cereal in the world (USDA, 2024a). Despite its per capita consumption remaining below that of conventional cereals (7.4 kg per year), there has been a notable increase (+24%) in sorghum cultivated area over the past five years (USDA, 2024a). This cultivation increase can be attribute to its agronomic/environmental advantages such as drought tolerance and less water requirement (Ragaee *et al.*, 2006; Stefoska-Needham *et al.*, 2015).

Sorghum is a starch (56-76%) and protein-rich (6-20%) grain, which generally contains noticeably higher levels of phenolic compounds, mainly flavonoid and tannin-type polyphenols than other cereals (Ragaee *et al.*, 2006; Xu *et al.*, 2021). Comprehensive research using *in vitro* and cell line assays has shown the potential health-promoting effects of sorghum, including antidiabetic, anti-inflammatory, cardiovascular disease prevention, and anticancer (Duodu & Awika, 2019; Taylor & Duodu, 2015; Xu *et al.*, 2021). However, a major issue limiting the sorghum consumption as food is its low protein digestibility.

Kafirins, the sorghum storage proteins, represent the most abundant fraction (~ 90% of total proteins in decorticated grains) and are naturally less digestible due to their synthesis into protein bodies and cross-linked disulfide bonds arrangement, in addition, protein-phenolic interactions further reduce the ability of digestive enzymes to proteolysis (Espinosa-Ramírez & Serna-Saldívar, 2016; Ozdal *et al.*, 2013). Polyphenols, mainly tannins, when exposed to oxygen in alkaline conditions, turn into quinones, leading to formation of protein cross-linkages. These active quinones can permanently bind with protein sulfhydryl and amino groups. These reactions between quinones and proteins are known to decrease protein digestibility and bioavailability (Ozdal *et al.*, 2013).

The fate of proteins and phenolic compounds along the gastrointestinal (GI) tract defines their biological activity; however, before intake, sorghum grains are processed, and it may interfere with their biological potential (Salazar-López *et al.*, 2018).

Technological processes change the biochemical and physicochemical properties of sorghum and can contribute to improve the population nutritional profile, especially in countries dependent on sorghum-, *e.g.*, semi-arid tropics (Motlhaodi *et al.*, 2018). Several technological processes have been tested to improve sorghum protein quality and/or phenolic compounds bioaccessibility (Abdelbost *et al.*, 2023; Cabrera-Ramírez *et al.*, 2020; D'Almeida *et al.*, 2021a; D'Almeida *et al.*, 2024).

Wet cooking, *i.e.* boiling the cereal in excess of water for variable time periods, is the traditional domestic method most commonly used to prepare different sorghum-based foods (Salazar-López *et al.*, 2018). Although, the negative impact of wet cooking on protein digestibility is well reported in the literature (Duodu *et al.*, 2002), its impact on phenolic bioaccessibility is still scarcely investigated particularly upon digestion. Previous studies suggested that wet cooking should be applied to sorghum rich in phenolic acids to help their solubilization and bioaccessibility, but with low flavonoid content, since they are heat-labile compounds (Salazar-López *et al.*, 2018). Otherwise, cooking can be enhanced by applying dehulling as a primary physical process to remove the outer pericarp and testa layer (Sruthi *et al.*, 2021). Due to sorghum grain hardness, this step is essential to remove its hard cellulosic cover and concentrate the phenolic compounds in the resulting bran (Sruthi *et al.*, 2021).

To assess how sorghum processing, such as cooking and dehulling affect protein digestibility and phenolic bioaccessibility in the GI tract, it is essential to accurately examine the phenolic profile. Untargeted metabolomics based on high-resolution mass spectrometry analysis is a valuable approach for unraveling changes and transformation mechanisms of chemical constituents following cereal technological processing (Yang *et al.*, 2024). Such detailed information on sorghum grains has not been previously reported, so an advanced analytical method focusing on sorghum specialized metabolites would address this gap. Therefore, this study aimed to provide comprehensive insights into how cooking and dehulling affects protein digestibility and phenolic compounds bioaccessibility in whole, dehulled, and bran sorghum fractions from Dutch varieties cultivated in Northern Europe.

2. Material and Methods

2.1. Plant material and cultivation conditions

Four Dutch sorghum varieties of different races and colors were grown in full fields (Maatschap de Milliano-Meijer, Oostburg, Zeeland, Netherlands). Two varieties from *Sorghum bicolor* race HD7 (brown-grained) and HD19 (white-grained) and two varieties from *Sorghum negricans* race HD100 and HD101 (black grains) possessing glume. These Dusormil sorghums were bred and selected for viable seed set, at Hoeve Dierkensteen, situated near the 51º Northern Latitude (NL), for regions above 48° (NL), and for all soils, since 2005. Dusormil HD7 and HD19 were also successfully tested in Denmark at 55º NL and accepted for the European Variety list in 2022. The sorghums are medium tall, 1.5- to 2.60 m and differ in the size of the chaff. Dusormil HD7 and HD19 have chaff partly covering the seeds, while Dusormil HD100 and HD101 have a chaff that covers the grain almost entirely.

The certified seeds for this trial of Dusormil HD7, HD19 and HD100 were produced in the Drome area of France (around 44° 97' NL). HD7 and HD19 crop cycle was of 108 days (sowing on 29 April and harvest on 15 May 2022), while HD100 crop cycle was of 106 days (sowing 27 May and harvest 10 September 2022). Dusormil HD 101 seeds was produced in Oostburg, the Netherlands (51º) with a crop cycle of 122 days (sowing on 29 May and harvest 28 September 2022). At 51° the crop cycles are between 120 (sowing late May) and 150 days (sowing early May). HD101 was the lowest plant height (about 1.6 m) and the shortest crop cycle. Crop cycles at 48° NL were considerably shorter than at 51° NL.

2.2. Grain characterization

Grain average weight was determined using a gravimetric method of 1000 grains. Water content was measured after 2 h at 135 °C, according to the AOAC official standard method 930.15 (AOAC, 2005a). For ash content, samples were kept in a muffle furnace at 600 °C during 2 h. The sorghum grain hardness was measured by AACC, method 55-30.01 (AACCI, 2010). The grain size analysis was performed using a set of seven sieves with different mesh sizes (0 – 3.5 µm). The grain density (mg/mL) was evaluated by the weight of a known volume of sample, according to Mariotti *et al.* (2006). The grain color was determined by a digital colorimeter (CR-400, Konica Minolta, Japan) with CIE L^x, a^x, and b^x color scale.

2.3. Dehulling and cooking process

Glumes from black varieties (HD100 and HD101) were manually removed in a quantity sufficient for further experiments. The dehulling procedure was adapted from Lestienne *et al.* (2007) to remove 15% of the total grain mass. Briefly, all sorghum whole-grain varieties (80 g) were dehulled into a laboratory-scale abrasive grain polisher (Model-TM 05C; Satake, Japan) at 800 rpm. Finally, four samples were obtained for each sorghum variety as follows: i) whole grain with glumes (WG+, exclusive to HD100 and HD101), ii) whole grain (WG), iii) dehulled grain (DG) and iv) bran (B) (Figure 17).

Sorghum whole grain and grain fractions (WG+, WG, DG and B) were boiled in a water bath (1:6, sample:water ratio) at 100 ºC during 30 min, with manual stirring every 5 min. After cooling time on the bench, the water was removed using sieves and grains were freeze-dried during 62 h. The four grain fractions in raw and cooked forms of each variety were cryogenically milled (25 Htz, min) using a ball mill (Retsch mixer mill MM 400, Germany) and stored at -40 °C until analysis.

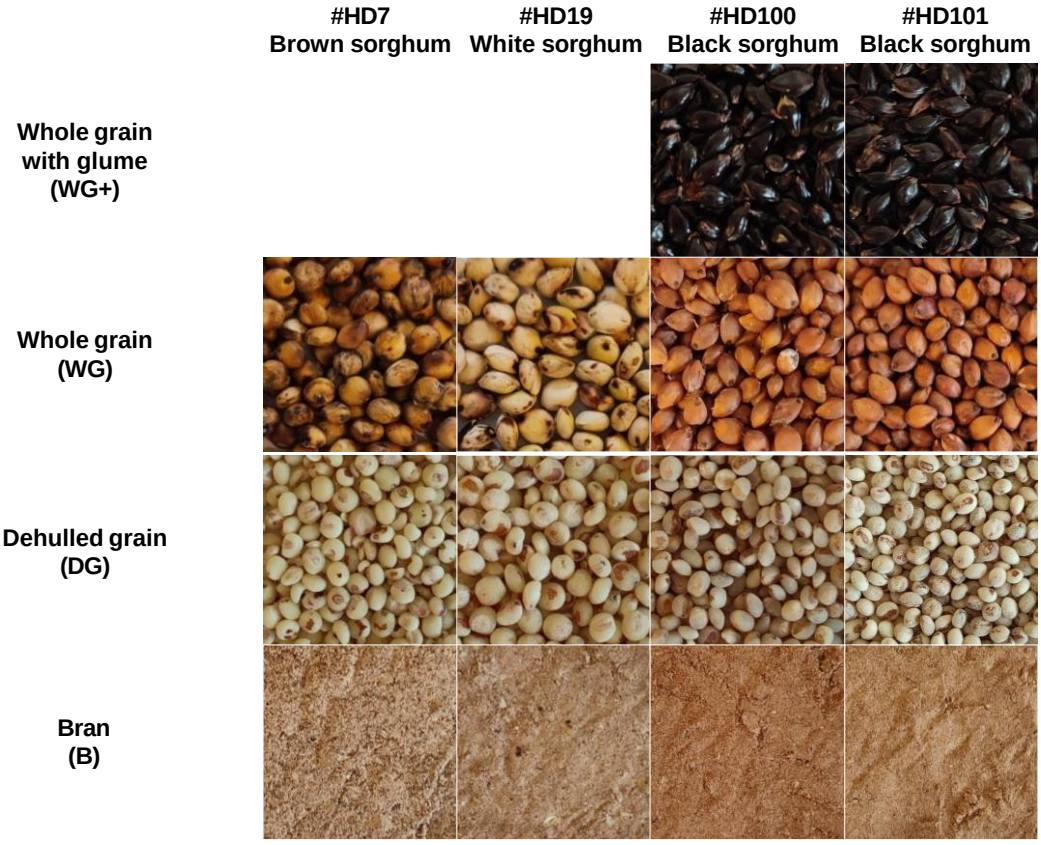


Figure 9. Photos of different fractions of sorghum grains with different pericarp colors.

2.4. Protein content and amino acids composition

The total nitrogen content of the different samples was determined using the micro-Kjeldahl method (AOAC, 2005a). Fifty milligrams of samples were transferred into a digestion tube containing 10 mL of sulfuric acid. After digestion (2 h, 400 ºC), samples were distilled and titrated using a Kjeltec 8400 fully automated Kjeldahl analyzer (FOSS, Hilleroed, Denmark). Protein content was calculated using a factor of 5.7 for nitrogen to protein conversion.

For amino acid (AA) composition, samples were first hydrolyzed with 6 N HCl at 110 ºC for 24 h as reported by (Abdelbost *et al.*, 2024). An ionic exchange chromatographic (Biochrom 30, Cambridge, UK) system using 0.2 M lithium citrate, pH 2.2 as eluent was applied. Norleucine was used as internal standard and analysis was performed in triplicate.

2.5. *In vitro* determination of protein digestibility and phenolic bioaccessibility

The *in vitro* digestion process was performed in triplicate based on the INFOGEST static method (Brodkorb *et al.*, 2019), with adaptations. This process consists of a 3-step procedure that mimics the physiological and pH conditions in the mouth, stomach, and small intestine, performed at 37 ºC. Briefly, 250 mg of each sample was mixed with 1 mL of simulated salivary fluid (pH 7) containing a-amylase (Sigma-Aldrich, cat. no. 1031) from human saliva rapidly added at 75 U/mL (final concentration). After 2 min under agitation, the a-amylase was inactivated under acidic conditions (6 M HCl) and the gastric phase was initiated. Simulated gastric fluid (pH 3, 4.82 mL) containing pepsin (Sigma-Aldrich, cat. no. P70000; final concentration: 2,000 U/mL) was added to the post-oral digestion tube (pellet + supernatant). The gastric phase was performed under agitation for 2 h and the pH was adjusted to 7 under alkaline conditions (3 M NaOH). Finally, the intestinal phase was performed by adding 7.2 mL of simulated intestinal fluid (pH 7) containing pancreatin (Sigma-Aldrich, cat. no. P7545, 100 U/mL) to the post-gastric digestion tube and stirring for 2 h. Sample controls (non-digestible samples) were obtained by replacing the volumes of the different enzymes with the corresponding volume of simulated fluids or replacing the samples with wheat starch.

For the determination of phenolic bioaccessibility, the tube containing the digested extract was centrifuged (13000 xg, 20 min, 4 °C) and the supernatant was collected. For improved recovery of the digested extract, 1 mL of MilliQ water was added

to the pellet prior to centrifugation (13000 xg, 5 min, 4 °C). Supernatants were pooled and analyzed by UHPLC-QTOF-MS/MS (See section 2.7). For protein analysis, enzymes were inhibited by adding TCA (12%, final concentration) and cooling on ice. *In vitro* protein digestibility (IVPD) was determined using total nitrogen content in the remaining pellet after digestion following the micro-Kjeldahl method (section 2.4) and then calculating according to the equation below:

$$IVPD = \frac{Ninitial - Nfinal}{Ntotal} \; x\,100$$

where $N_{initial}$ is the initial nitrogen in sorghum samples, N_{final} is the remaining nitrogen after digestion and N_{total} is the total nitrogen in sorghum samples.

2.6. Phenolic compounds extraction

Free and bound phenolic compounds extraction was performed in triplicate according to Santos *et al.* (2019b), with some modifications. Briefly, 70 mg of sorghum samples were extracted in 80% ethanol at a ratio 1:20 (w/v) to obtain free phenolic compounds (soluble). Samples were stirred (200 rpm, 25°C, 10 min) and centrifuged (5,000 xg, 25°C, 10 min). Extraction was performed twice, and the extracts were pooled. The pellets resulting from free phenolic extraction were submitted to alkaline hydrolysis with 1:70 (v/v) of 4M NaOH submerged in an ultrasonic bath (42 kHz, 90 min, 40 °C). Acid hydrolysis was then performed with concentrated HCl (~pH 2), and the samples were centrifuged (2000 x*g*, 5 min). The supernatant was washed three times with ethyl acetate (7 mL) and centrifuged between each wash (10000 x*g*, 5 min, 10 °C) to obtain the bound phenolic compounds (insoluble). Both extracts (free and bound) were evaporated (SpeedVac Savant, ThermoFisher Scientific, USA) and reconstituted in 1.5 mL of methanol, acetonitrile and Milli-Q water (2:5:93, v/v/v). The reconstituted extracts were filtered (0.22 µm, hydrophilic PTFE) and stored in vials at −80 °C.

2.7. Untargeted metabolomics analysis of phenolic sorghum extracts by using UHPLC-QTOF-MS/MS

Aliquots of 5 µL of each sample of free and bound phenolic extracts (section 2.6) and samples obtained following simulated digestion (section 2.5) were injected into an Ultra High Performance Liquid Chromatography (UHPLC) Acquity system (Waters, Milford, MA, USA) coupled to a XEVO G2-S Q-Tof (Waters, Milford, MA, USA)

equipped with electrospray ionization source. An UHPLC HSS T3 C18 column (100 x 2.1 mm, 1.8 µm particle diameter; Waters) at 30 °C and flow rate of 0.5 mL/min of ultra-pure water containing 0.3% formic acid and 5 mM ammonium formate (mobile phase A) and acetonitrile containing 0.3% formic acid (mobile phase B) was used for the gradient elution: 0 min - 97% A; 11.80 min - 50% A; 12.38 min - 15% A; 14.11 - 97% A. Data were acquired in triplicate in data-independent acquisition (DIA) MS^E mode, in negative and centroid mode between *m/z* 50 and 1200, collision energy ramp from 30 to 55 V, cone voltage 30 V, capillary voltage 3.0 kV, desolvation gas (N_2) 1,200 L/h at 600 °C, cone gas 50 L/h, source at 150 °C, and using leucine enkephalin (Leu-Enk, *m/z* 554.2615 [M-H]-) for calibration lock mass. A mixture containing 33 analytical standards of phenolic compounds (10 ppm) was prepared and injected in triplicate, prior to the injection of the samples, to ensure the reproducibility of the instrument and to confirm identity of phenolic compounds. Besides this, a set of quality control (QC) samples was also prepared by pooling equal volumes of each sample extract and injected after every batch of six runs to monitor the instrument's stability and the reliability and accuracy of mass spectrometry data analysis.

2.8. Bioinformatics and statistical analysis

MassLynx v 4.1 and Progenesis QI softwares (Waters, Milford, MA, USA) were used MS data acquisition and processing, respectively. Metabolites identification was based on parameters of standard runs including exact mass, retention time and MS/MS fragments spectra. Untargeted identification was performed according to the annotation levels proposed by the Metabolomics Standards Initiative (Sumner *et al.* (2007) from a customized database built from online PubChem and Phenol-Explorer databases. The following parameters were applied: exact mass error (<10 ppm); isotopic similarity (>80%); score (>30) and the highest score of fragmentation, generated by the software. Data from the literature and chemical characteristics of the molecules were also used to help the tentative identification of unknown compounds. In addition, only compounds present in all technical replicates (3/3) and showing CV < 30% were considered.

The processed data were exported to the XLSTAT software (Addinsoft, France) where the values of total ion counting (*i.e.* ion abundance) obtained from mass spectra were used for relative quantification and for statistical analysis (one-way Anova, and Tukey post-test, $p<0.05$). Multivariate statistical analysis, such as principal component analysis (PCA), Hierarchical clustering analysis (HCA) with heatmap and volcano plots

were performed by Metaboanalyst 6.0 (https://www.metaboanalyst.ca/). For volcano plot analysis, samples were compared by pairing and differentially abundant compounds were selected following the filters log2 fold change ±1.5 and ANOVA $p < 0.05$.

3. Results and Discussion

3.1. Physico-chemical characterization of the different sorghum grains

The characterization of Dutch sorghum genotypes is shown in Table 2. Grain water concentration (%) averaged 13.60±0.31 indicating grains were stored in good conditions for this study. The 1000-grain weight (GW) significantly varied between genotypes from 18.13±0.29 to 30.09±0.58 mg, with HD19 genotype being the heaviest grain, followed by HD7, HD101 and HD100. Some grain characteristics are crucial for its primary processing, such as milling and dehulling. Similar to Brazilian sorghum grains, Dutch grains were classified as very hard (HD7 and HD19) and extra hard (HD100 and HD101) (de Oliveira *et al.*, 2022). Grain hardness influences the energy consumption during the milling/dehulling process, the flour yield, and the level of starch grain damage. During the grain dehulling, HD100 and HD101 showed a higher resistance to the process and needed more time (30 minutes each) to remove 15% of the grain weight than HD7 and HD19 genotypes (11 minutes each) (data not shown). Grain tapped density, often related to its flowability, varied between 0.68±0.00 and 0.83±0.02 g/mL with HD7 and HD19 genotypes showing the highest values ($p<0.05$). Sorghum grains size varied slightly between genotypes averaging 2.34±0.04 µm. The grain size and homogeneity influence its processing and facilitate standardization of the equipment.

Table 2. Biochemical and physicochemical analyses of Dutch sorghum whole grains with different pericarp colors.

Genotypes	Pericarp color	Grain water (%)	Grain weight (g, d.m.)	Ash content (%, d.m.)	Protein content (%, d.m.)	Grain hardness index	Tapped density (g/mL)	Grain size (µm)	Colorimetric parameters		
									L*	a*	b*
HD7	Brown	13.89±0.05^{a}	26.84±0.25^{b}	1.43±0.05^{c}	9.30±0.06^{c}	9.82±1.34^{a}	832.50±18.86^{a}	2.32±0.17^{bc}	44.63±0.38^{b}	9.16±0.14^{a}	16.70±0.17^{a}
HD19	White	13.85±0.06^{a}	30.69±0.58^{a}	1.54±0.08^{c}	11.38±0.10^{b}	9.01±0.25^{ab}	831.34±7.32^{a}	2.38±0.14^{a}	57.82±0.23^{a}	5.98±0.23^{b}	16.80±0.30^{a}
HD100	Black	13.16±0.05^{c}	18.13±0.29^{d}	2.60±0.03^{b}	11.60±0.16^{b}	5.33±0.52^{c}	710.44±4.90^{b}	2.30±0.18^{c}	24.33±0.42^{d}	3.72±0.09^{d}	-2.65±0.26^{c}
HD101	Black	13.49±0.05^{b}	19.36±0.11^{c}	3.93±0.10^{a}	14.80±0.07^{a}	7.23±1.66^{bc}	681.05±3.61^{c}	2.37±0.16^{ab}	27.41±0.44^{c}	4.91±0.38^{c}	0.02±0.57^{b}

The values are expressed as mean ± standard deviation for at least 3 independent experiments. Different letters mean significantly difference between columns ($p < 0.05$ by one-way ANOVA).

The protein and ash contents of the different sorghum genotypes varied between 9.30±0.06 and 14.9±0.07% and 1.43±0.05 and 3.9±0.10%, respectively (Table 2). These contents were comparable in almost all genotypes with global germplasm collection at International Crops Research Institute for the Semi-Arid-Tropics (ICRISAT) (https://genebank.icrisat.org/). However, the ash content of HD101 found in the present study was 18.2% higher than recommended by FAO (1995). Although high-ash sorghum can be attractive from a nutritional perspective because it implies higher mineral content, the latter might negatively impact the technological quality and brightness ($r = 0.774$, $p < 0.005$) of sorghum-based products, resulting in limited consumer acceptance of these final product (Dube *et al.*, 2020).

Finally, sorghum whole grains showed different color parameters (Table 2). As expected, lightness (L*) values were significantly higher for white and brown genotypes than for black ones. L* values were lower than those previously reported on sorghum and only the white variety (HD19) showed similar values (Espinosa-Ramírez & Serna-Saldívar, 2016). Dusormil sorghums samples also showed higher a* values (more reddish) compared to the reported for white-regular, white-waxy, red-regular and high-tannin sorghums, suggesting a rich constitution in red colored pigments, mainly constituted by phenolics (Espinosa-Ramírez and Serna-Saldívar, 2016). The b* values showed the highest amplitude between genotypes ($p<0.05$). HD 7 and HD 19 showed similar and higher b* values (more yellowish) than black genotypes ($p<0.05$).

3.2. Amino acids profile and the chemical score

As reported above, the protein values of studied grains are in line with what is expected for sorghum, but its nutritional quality based on its AA profile needs to be evaluated. The seventeen AA detected and quantified in the four sorghum genotypes are presented in Figure 18. The sum of AA detected and quantified by HPLC is equivalent to 74% of the total protein content. The major AA found in grain sorghum were glutamic acid (Glu, 22%), leucine (Leu, 11%), proline (Pro, 10%), alanine (Ala, 9%), and aspartic acid (Asp, 7%). Aspartic and glutamic acids represented 29% of the total AA in the sorghum samples. However, according to the amino acid profile of cereals, these identified acids are products of glutamine and asparagine produced under acidic conditions (6 N HCl) applied during sample preparation.

The percentage of basic AA (lysine, arginine and histidine) found in our study (9%) was lower than that reported in other sorghum varieties (13-18%) (Mohapatra *et al.*,

2019; Mokrane *et al.*, 2010). This deviation can be explained by genetic or geographic variation among samples. The percentage of hydrophobic (alanine, isoleucine, leucine, methionine, phenylalanine, proline and valine) and uncharged polar AA (glycine, serine, threonine, tyrosine, and cysteine) were 47 and 15%, respectively, and are in accordance with what has been reported for sorghum (Mokrane *et al.*, 2010). To sum up, the most abundant AA were of the hydrophobic group, followed by acidic, uncharged polar and basic in decreasing order. This AA profile reinforces the hydrophobic character of kafirins.

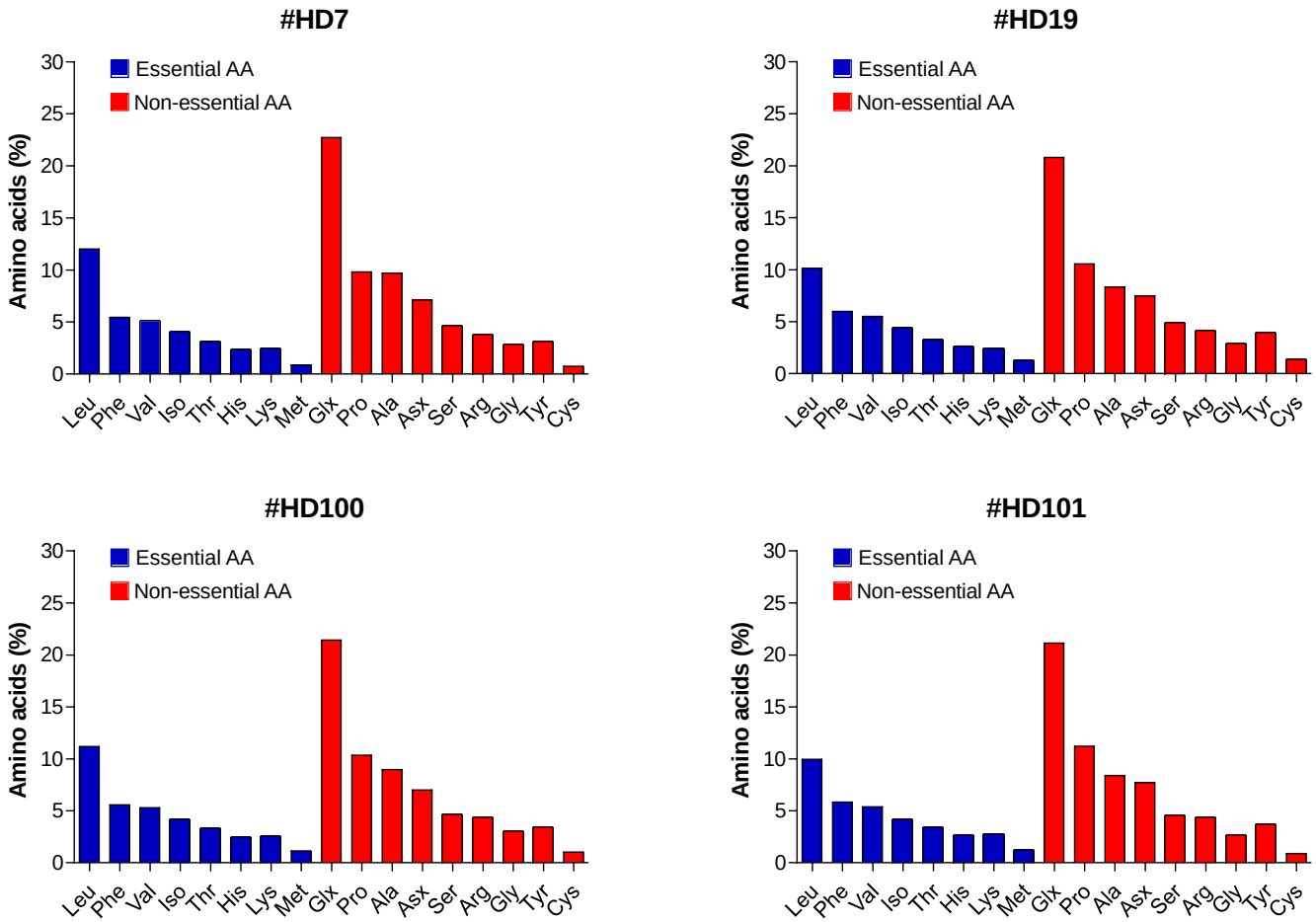


Figure 10. Amino acids profile of Dutch sorghum genotypes. Essential and non-essential amino acids are represented by blue and red bars, respectively.

From a nutritional perspective, the chemical score of essential amino acids (EAA) should ideally meet the daily intake needs of an adult (Consultation, 2011) (Supplementary table 4). Sorghum proteins presented seven EAA in sufficient amount (CS ≥ 100 %) in all samples: His, Ile, Leu, Phe, Tyr, The and Val. Tryptophan was not identified and quantified due to its probable degradation during acid hydrolysis in sample preparation. As expected, all analyzed sorghum genotypes were limiting in Lys, such as known for the most cereal species, where lysine is the main limiting AA (Eggum & Beames, 1983). Furthermore, the HD7, HD100 and HD101 genotypes also displayed sulfur amino acids (Met + Cys) as limiting AA in their proteins. Dietary EAA sulfur-containing is important for protein synthesis, sulfur metabolism, epigenetic modification, antioxidant defense and signaling (Ji *et al.*, 2023).

3.3. *In vitro* digestibility after sorghum processing

The four genotypes studied showed a large variability of *in vitro* protein digestibility (IVPD) between 9 and 81% for whole grain (WG) samples (Figure 2). This large variation was due to the HD100 genotype, which had an IVPD value lower than that reported in the literature (Correia *et al.*, 2010; Duodu *et al.*, 2003). Cooking sorghum grains significantly decreased the IVPD by -33% in HD19 and -39% in HD101. Boiling sorghum grains in excess water is known to be detrimental to the digestibility of kafirins due to the oxidation of sulfhydryl groups (i.e. promoting crosslinking through disulfide bonds) and, consequently, increasing the resistance to proteolysis (Abdelbost *et al.*, 2024; Duodu *et al.*, 2003; Hamaker *et al.*, 1986). In addition, starch gelatinization in wet-cooked sorghum may play an important role in the reduced accessibility of proteases to kafirins, as protease-resistant starch-kafirin complexes can be formed after cooling of cooked sorghum, (Duodu *et al.*, 2002).

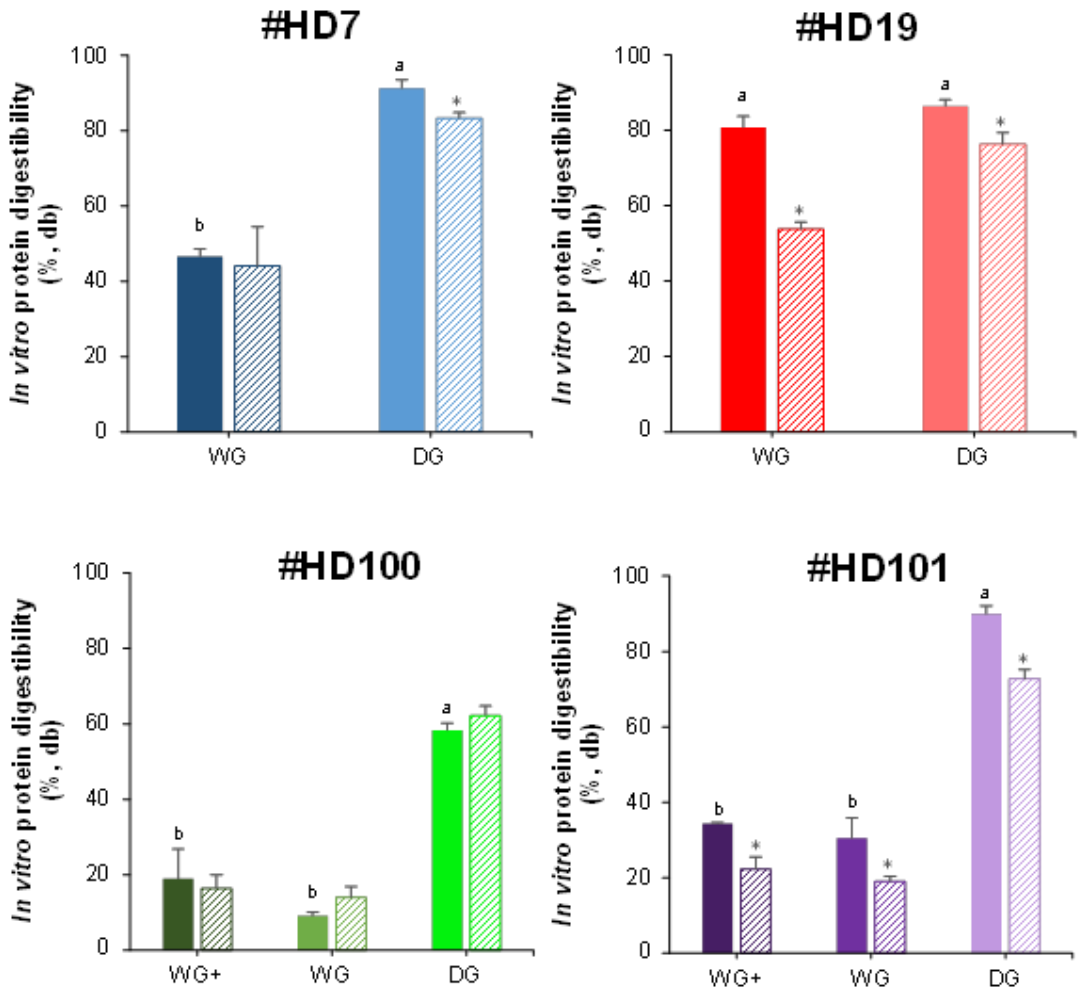


Figure 11. *In vitro* protein digestibility after thermal (cooking) and physical (dehulling) processing of Dutch sorghum grains. WG+ = whole grain with glume; WG = whole grain; DG = dehulled grain. Filled and hatched columns represent raw and cook samples, respectively. Different lowercase letters mean the significant difference ($p < 0.05$) between sorghum WG and grain fractions. * means significance level between raw and cooked samples. Bars represent standard deviation (n = 3).

Interestingly, in raw sorghum samples, IVPD was higher in dehulled grains (DG) than in whole grains (WG), except for the white-grained HD19 genotype, where there was no significant difference ($p>0.05$) (Figure 2). The phenolic compounds present in the pericarp form highly oxidizing agents, such as quinones, which react with the nucleophilic groups of the protein chain and alter it structure (Cabrera-Ramírez *et al.*, 2020; Duodu *et al.*, 2003; Jakobek, 2015). This reaction reduces protease activity, so removal of the pericarp can improve the digestibility of raw sorghum. Flavonoids (*e.g.*, quercetin) and phenolic acids (*e.g.*, chlorogenic acid) are examples of potent trypsin and intestinal chymotrypsin inhibitors that have already been reported to delay certain food proteolysis (Brás *et al.*, 2010; Cirkovic Velickovic & Stanic-Vucinic, 2018).

For dehulled grains, cooking reduced IVPD in the HD7 (-9%) HD19 (-12%) and HD101 (-19%) genotypes, and this reduction was considered to be mild. These data confirm the findings of Duodu *et al.* (2002), who demonstrated that the negative effect of cooking on IVPD can be reduced after grain dehulling, depending on the genotype. Therefore, dehulling can be considered as a viable option to increase protein digestibility by removing phenolic compounds and avoiding possible protein-phenolic interactions.

3.4. Analytical insights through phenolic identification and characterization by metabolomics

The overall count of features detected by the UHPLC-ESI-QTOF platform amounted to 13,355 ions. After filtering application and exclusion criteria in the metabolites annotation step (i.e. presence in the blank, ionic relative abundance below to 5,000 chromatogram baseline, isotopic similarity below 80%, and disparities in fragmentation data), 219 phenolic compounds were tentatively identified, among them 11 compounds were fully confirmed with commercial phenolic standards (Supplementary Table 5, compounds presented in bold).

The 10 main and most abundant phenolic compounds in our sorghum samples are listed below. Daidzin (RT 9.96) was the most abundant phenolic compound among the sorghum samples analyzed, corroborating previous studies (D'Almeida *et al.*, 2021a). Five compounds were confirmed with commercial phenolic standards (level 1 identification): 4-hydroxybenzoic acid (2^{nd} most abundant, RT 4.71), isoferulic acid (3^{rd}, 8.01), caffeic acid (4^{th}, 6.25), p-coumaric acid (5^{th}, 7.48) and kaempferol (6^{th}, 10.18). Other four most abundant compounds (7^{th} to 10^{th} position) were assigned to flavonoid

class and annotated according to level 2 as a quercetin glycosylate (quercetin 3-O-rhamnosyl-galactoside, RT 8.07) and three procyanidin isomers (one dimer, RT 5.42; and two trimers, RT 5.73 and 7.39).

Procyanidins *(i.e.*, condensed tannins) are flavonoids with a basic skeleton composed of catechin and two B-ring hydroxyl groups. Dimeric procyanidins are formed by an interflavan bond between C_4 of the B-ring and C_8 or C_6 of the C-ring while trimeric procyanidins are made up of three (-)-epicatechin units linked together by two successive (4β->8)-linkages, typically connected by C-C or C-O-C bonds (Rue *et al.*, 2018). Although procyanidins are associated with beneficial effects on human health (Ferni *et al.*, 2022), their hydroxyl groups can interact with sorghum components (*e.g.*, proteins and starch) and further reduce their digestibility (Duodu *et al.*, 2003). Eight dimeric and six trimeric procyanidins were tentatively identified and their ionic abundances within samples followed the decreasing order: HD100 > HD101 > HD7 > HD19. Furthermore, HD19 was the only genotype to exhibit a higher abundance of procyanidin dimers compared to trimers (dimer-to-trimer ratio: 3.3).

Procyanidin fragmentation pathways determined using MS^2 analysis is proposed in sorghum grains for the first time (Supplementary Figure 3A). For sorghum dimeric procyanidins ([M−H]- ion, *m/z* 577.1342), the quinone methide (QM) reaction appears to be prioritized, forming the *m/z* 287.0501 fragment (QM_2, 45% of relative abundance) (Supplementary Figure 3A). Similarly, trimeric sorghum procyanidins mainly formed the *m/z* 577.1270 (procyanidin dimer) fragment through the QM reaction (QM_1, 22%); however, in this case, other reactions occur simultaneously: 1) Retro-Diels-Alder (RDA) reaction formed the *m/z* 713.1417 fragment - characterized by an (Eseberri *et al.*)catechin unit loss (-152 Da). Again, the ion *m/z* 577.1270 from QM_1 reaction can undergo a new RDA-based fragmentation (*m/z* 425.0812, -152 Da loss). 2) Heterolytic ring fission (HRF) generated ions at *m/z* 739.1582 (-125 Da loss from *m/z* 865.2002 precursor ion) and 451.0980 (-126 Da loss from *m/z* 577.1270 fragment). Some fragments were formed by water (*m/z* 695.1310, 559.1187 and 407.0561; -18 Da) and hydrogen (*m/z* 575.1127, -2 Da) loss molecule (Rue *et al.*, 2018).

Finally, the heatmap depicted the relative abundance of total phenolic compounds across different whole-grain sorghum genotypes (Supplementary Figure 3B). Distinct patterns in phenolic profile emerged that allowed clear differentiation among samples:

numerous compounds were more (illustrated by red blocks), while others were less abundant (illustrated by blue blocks). From this, the four genotypes were clustered according to significant differences in phenolic levels, with the dark-colored pericarp sorghum grains possessing glumes (HD100 and HD101) being clearly distinguished from the lighter genotypes (HD7 and HD19). The observation evidenced the strong positive correlation between the total levels of phenolic compounds in the whole-sorghum grain and its colorimetric parameters: L* ($r = 0.78$, $p = 2.70 \times 10^{-3}$), a* ($r = 0.87$, $p = 2.39 \times 10^{-4}$) and b* ($r = 0.92$, $p = 1.86 \times 10^{-5}$), contrary to the report (Dykes *et al.*, 2005) that lighter, yellowish, and reddish grains have higher phenolic content.

3.5. Thermal and physical processing impact on sorghum grain phenolic profiles

Among the 219 phenolic compounds putatively identified for all sorghum samples, 87% were annotated as level 2 (MS/MS-based annotation) according the international Metabolomics Standards Initiative (Sumner *et al.*, 2007). MS/MS-based approach was valuable for assigning names and classes and facilitated the elucidation of changes and transformation mechanisms of diverse polyphenols in different sorghum varieties for raw and processed samples. The distribution in total relative ion abundance of these compounds in each sorghum grain fraction is shown in Figure 20A.

Black genotypes (HD100 and HD101) had their phenolic abundance negatively impacted when glumes were removed (-45% in WG+ when compared to WG). On the other hand, the ratio of free-to-bound phenolic compounds increased in HD100 (from 1.29 to 2.23) and HD101 (from 0.95 to 1.80) samples after threshing. According to Silva-Fernández *et al.* (2022), cereal glume is rich in insoluble fiber primarily composed of cellulose and hemicelluloses. These grain structural components are generally covalently bound to carboxylic groups of phenolic compounds by ester-type linkages, so removing this fibrous layer changes the phenolic profiles, with a preferential increase of free compounds and this alteration can directly impact phenolic bioaccessibility.

Grain dehulling is the main method to obtain bran, which concentrates phenolic compounds. The bran fraction from lighter genotypes (HD7 and HD19) represented, respectively, 80 and 94% of the phenolic content of the whole grain (WG); however, in the black-pericarp genotypes (HD100 and HD101), the phenolic abundance of bran surpassed the WG contents. HD100 and HD101 are harder grains (see section 3.1) and, consequently, had higher dehulling process resistance, requiring more time for abrasive

processing (~30 min) to obtain 15% of the grain. This prolonged time may have favored the transformation and/or extractability of free phenolic compounds in the bran fraction (3 and 2-fold higher in HD100 and HD101, respectively, when compared to their WG).

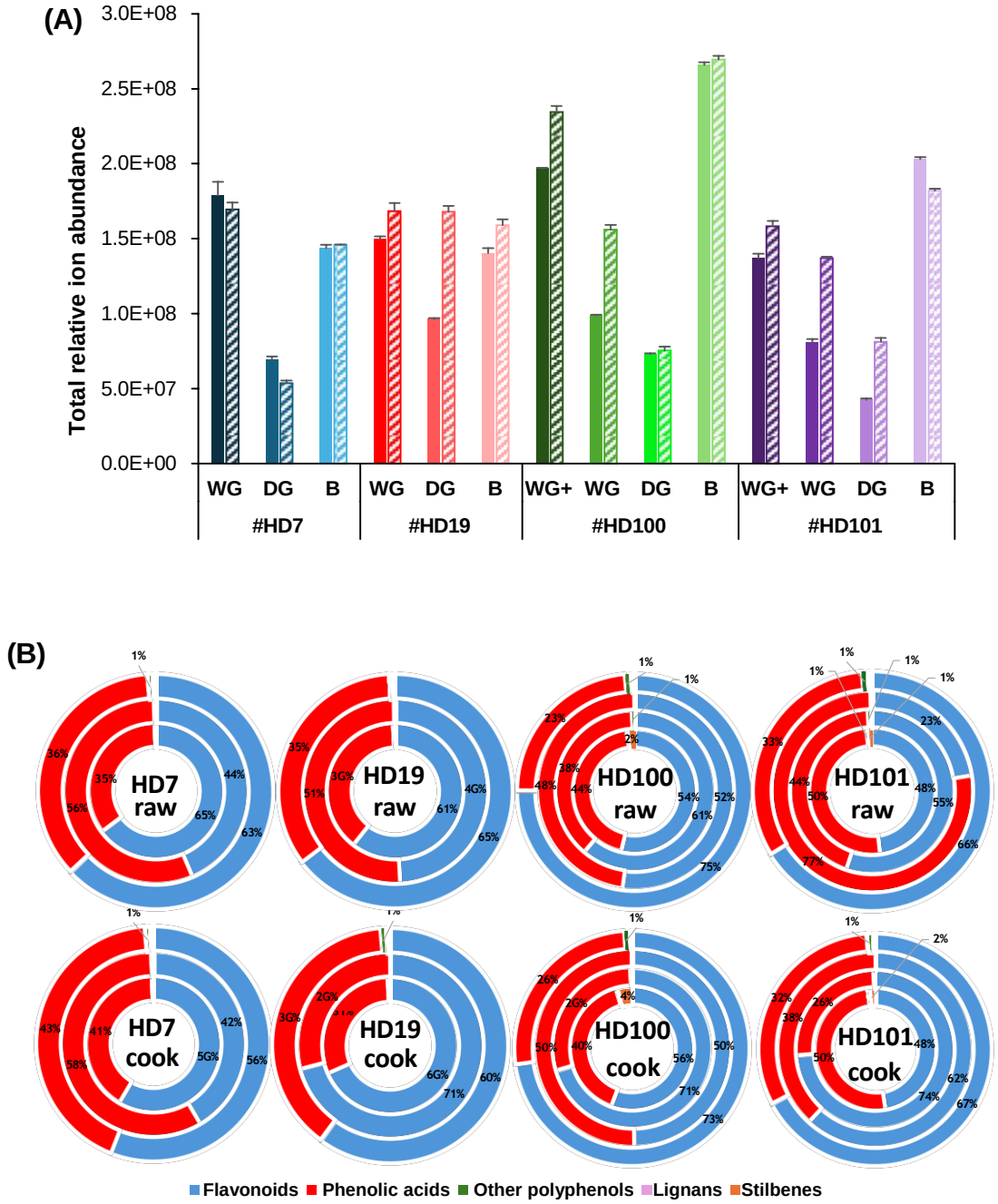


Figure 12. Phenolic profile of sorghum samples by metabolomics approach: (A) total relative ion abundance of phenolic compounds and (B) distribution of phenolic classes in samples. Filled and hatched columns represent raw and cook samples, respectively. Graph B is plotted as follows, from the internal to external axis: WG, DG and B for genotypes HD7 and HD19; and WG+, WG, DG and B for HD100 and HD101 genotypes. WG+ = whole grain with glume; WG = whole grain; DG = dehulled grain; B = bran.

Furthermore, although dehulling reduced total phenolic compounds in all genotypes (-61, -36, -26 and -48% in HD7, HD19, HD100 and HD101, respectively), DG samples still presented significant phenolic contents.

Globally, boiling positively impacted phenolic content in sorghum grain fractions but affected genotypes in a different way. HD7 genotype showed a 23% reduction in phenolic content for the dehulled fraction (Figure 20A), while dehulled HD19 and HD101 grains were positively impacted by cooking as seen by a major increase of +74 and +92%, respectively. Wet cooking also increased phenolic compounds in all fractions of HD100 genotype and WG had the highest increment (58%). This increase in phenolic abundance is directly associated with the total number of ID compounds ($r = 0.76$, $p = 1.70 \times 10^{-3}$) (Supplementary Figure 4), suggesting that high temperatures broke down high molecular-weight molecules or polymeric phenolic compounds into simple and more extractable molecules as suggested by Li *et al.* (2022). Wet heat treatment also appeared to damage the cell wall, breaking down cellular constituents and thereby releasing some bound phenolic compounds. The free-to-bound ratio increased from 1.99 to 2.46 after cooking and it increased 23% in the total free extract.

Using their foundational chemical attributes and structures as a guide, these phenolic compounds were categorized into 5 distinct classes (Figure 20B). Flavonoids and phenolic acids correspond together for 99% of the total phenolic abundance; however, the proportion of these two classes was variable in each fraction. Globally, WG+ raw samples had similar proportions of flavonoids and phenolic acids (~51% and 47%, respectively), while WG raw samples had ~61% flavonoids and ~39% phenolic acids. From this observation, we hypothesized that due to its fibrous composition, sorghum glume can interfere with the extractability of flavonoids, favored in WG after the threshing process (+81%).

Phenolic acids represent the most abundant class (~58%) in sorghum endosperm (i.e. dehulled grain, DG), since dehulling reduced total flavonoids by 15%, except for HD100 genotypes. The obtained brans, considered the most important cereal by-product, is actually a bioactive concentrated with 67% of flavonoids in its total phenolic composition. Although cooking increased sorghum phenolic relative abundance (as reported above), this wet heat processing did not significantly impact grain phenolic profiles (Figure 20B).

To evaluate the variables with the most impact on phenolic profile changes, a PCA biplot containing all samples (scores) was built based on abundance of compounds

(observations) (Figure 21A). Notably, PC1 and PC2 contributed to 66% of the total variance in the phenolic profiles. Dehulled samples, regardless of genotype, were grouped, suggesting bran as the most discriminant fraction of the grain. WG and B were grouped according to pericarp color, *i.e.*, lighter grains (HD7 and HD19) presented a similar profile between them; however, a dissimilar profile was found among black pericarp samples (HD100 and HD101). Cooked fractions were plotted near to their respective raw sample, corroborating with Figure 20B, *i.e.*, phenolic profile was not impacted by the wet thermal process. From Euclidean distance clustering, the most relevant variables for the phenolic profile were grain fraction > genotype > thermal treatment. This was confirmed with a subsequent PCA built by merging genotypes and containing only the different grain fractions , where a cumulative variation of almost 89% in the phenolic profiles was observed (Figure 21B).

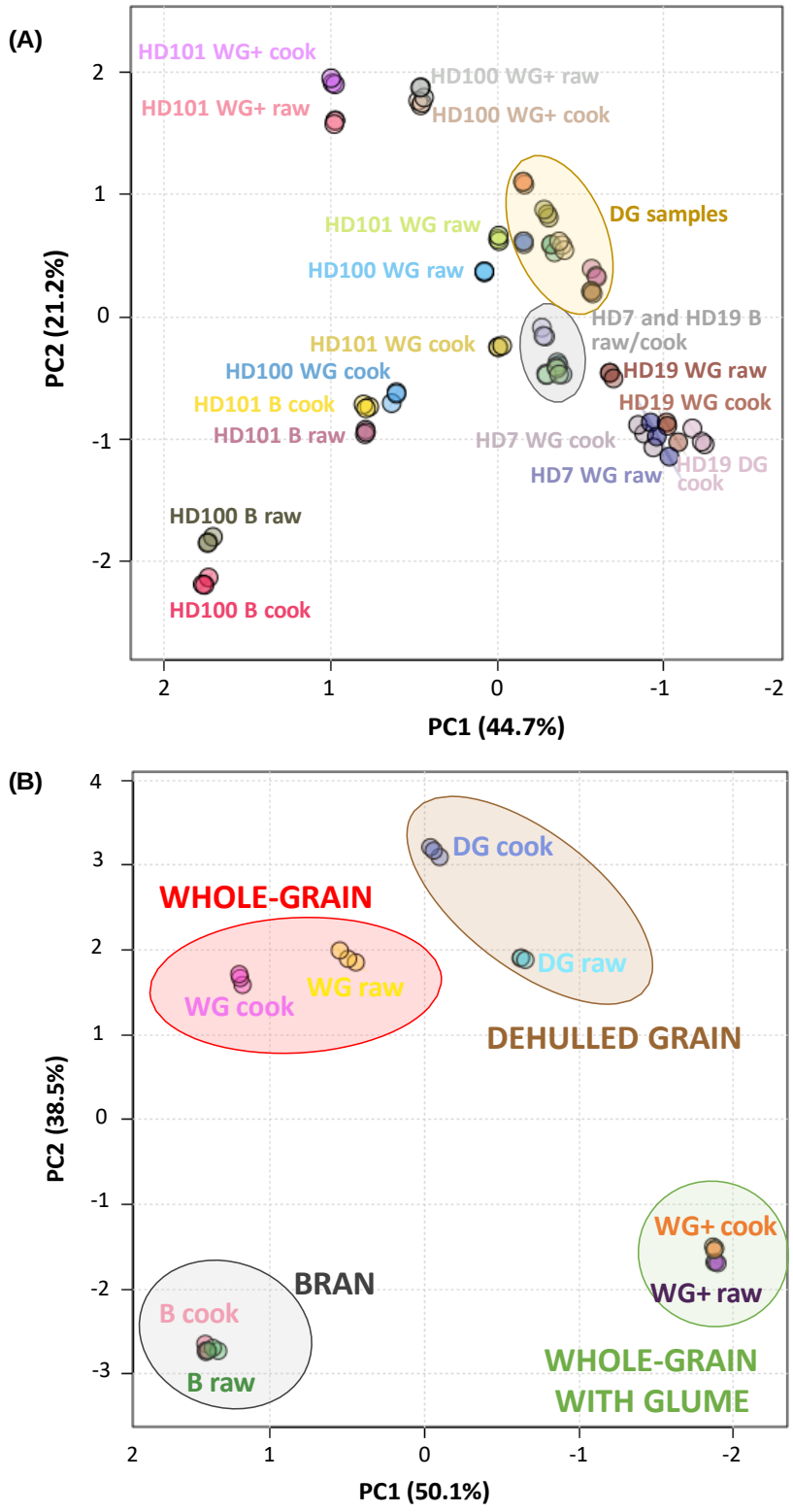


Figure 13. Principal component analysis (PCA) biplot of all samples (A) and only between sorghum grain fractions (B). The samples (symbols) are distributed according to relative intensity of identified phenolic compounds (observations). WG+ = whole grain with glume; WG = whole grain; DG = dehulled grain; B = bran.

3.6. Differentiation of phenolic compounds in sorghum grains under thermal and physical processing methods

To better understand the impact of sorghum processing on the differential abundance of phenolic compounds, volcano plots were built by paired comparison between compounds simultaneously present in different conditions (threshold of log2 fold change ± 2.0 and $p < 0.05$) (Figure 22). Firstly, the peeling effect was assessed by comparing WG+ *vs* WG, in which blue and red circles represent the decreased and increased phenolics, respectively (Figure 22A). Ninety-three compounds, mostly flavonoids (70%) decreased after glume removal. Among them, glycosylated flavonoids (mainly derived from quercetin, apigenin, daidzein and epicatechin) and plant pigments, such as theaflavin. Theaflavins are formed by the oxidation of a coupled catechin monomer and are widely known for their antioxidant and anticancer activity (Shan *et al.*, 2021). Normally present in black and green tea, theaflavins are not commonly found in sorghum grains. Although some studies have previously reported its derivative (theaflavin 3,3'-O-digallate) in this cereal (Espitia-Hernández *et al.*, 2024; Espitia-Hernández *et al.*, 2022b), to our knowledge, this is the first time that theaflavins have been reported in sorghum grain fractions.

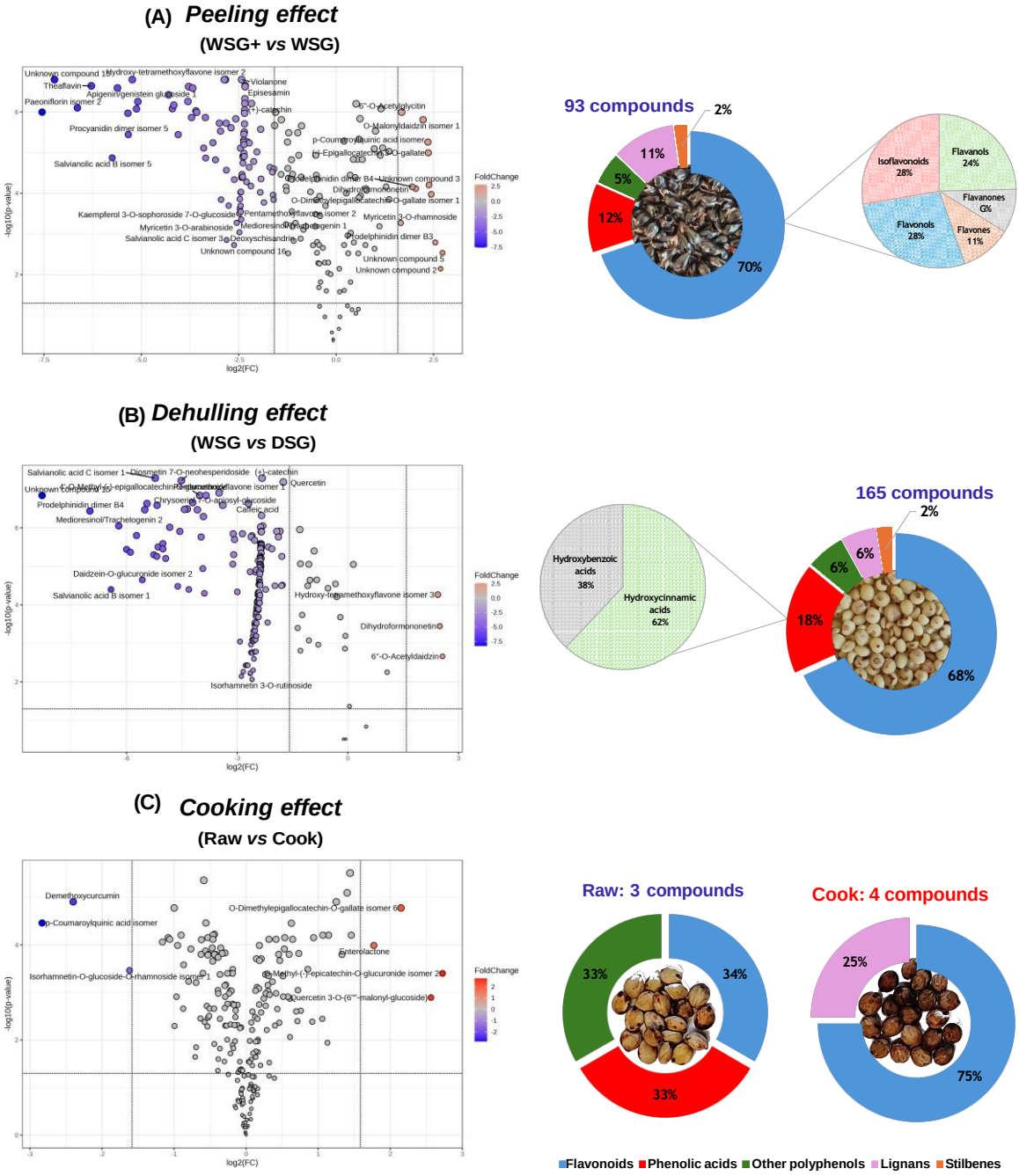


Figure 14. Differential phenolic compounds between Dutch sorghum grain processing: volcano plots of differential phenolic abundance (A) during peeling (WSG *vs* WSG+), (B) dehulling (DSG *vs* WSG) and (C) cooking (cook *vs* raw) process. WG+ = whole grain with glume; WG = whole grain; DG = dehulled grain. Red and blue dots indicate significantly increased and decreased metabolites, respectively (log2 fold change ± 1.5, $p < 0.05$). Gray dots represent the metabolites showing no significant difference.

Subsequently, the dehulling impact was evaluated based on differential compounds between DG *vs* WG (Figure 22B). A total of 165 phenolic compounds decreased after bran removal, such as procyanidin dimers (4) and trimers (5), corroborating the IVPD increase after dehulling. On the other hand, hydroxybenzoic and hydroxycinnamic acids identified at level 1 were reduced after this physical process (*e.g.*, caffeic, ellagic, p-coumaric, syringic, isoferulic and 4-hydroxybenzoic acids). These compounds are typically predominant in sorghum grain and are mainly found bound to cell wall polysaccharides (D'Almeida *et al.*, 2021a; Xiong *et al.*, 2020). In our samples, these six compounds taken together represent 86% of phenolic acids and 33% of phenolic total abundance in whole sorghum raw grains (WG). These data lead us to reflect on the use of dehulling for sorghum-based products, as this method must be tailored to meet consumer needs and the desired final product, *i.e.*, products with greater protein digestibility, less astringent flavor (less condensed tannin), but without considering the bioactive phenolic content.

Finally, the cooking effect pointed only 7 differentially abundant compounds (Figure 22C; 4 increased and 3 decreased), corroborating the previously observed in Figure 20. Although wet cooking promotes a total relative abundance increase, the cooking time-temperature binomial applied (100 ºC, 30 min) was not enough to promote phenolic transformations.

3.7. Processing methods impacted overall bioaccessibility of phenolic compounds

Sorghum grains are complex food matrices whose phytochemicals can establish synergistic and/or antagonistic reactions, thus modulating their overall nutritional properties. For this reason, it is crucial to evaluate the effects of technological processing on phenolic bioaccessibility. UHPLC-ESI-QTOF platform detected many features (5,820 ions), given enzyme residues and other food matrix components can be inadvertently detected. All detected features were plotted in a PCA analysis (Figure 23A, left-PCA), in which PC1 and PC2 contributed 61.7% and 13.3% of the total variability (75%), respectively, in the metabolite profiles. Non-digested (ND) and digested (D) samples can be distinguished on PC1-axis, while PC2-axis allowed a slight differentiation of the processing (dehulling and boiling), suggesting changes in the overall phenolic profile after GI digestion. Therefore, a new PCA was performed only with samples after digestion (Figure 23A, right-PCA). In this case, the metabolic profile variation

(PC1=54.7% and PC2=21.2%) and Euclidean distances clustering again demonstrated the greatest impact on phenolic bioaccessibility following the physical dehulling process versus hydrothermal treatment (boiling).

To further understand how processing individually affects phenolic bioaccessibility in sorghum grains, some compounds were selected based on identification level 1 reliability. Among these eleven compounds fully confirmed according to commercial standards (Supplementary table 5, compounds in bold), only seven were found in the digested extracts (Figure 23B). Among them, hydroxybenzoic and hydroxycinnamic acids showed greater bioaccessibility than (+)-catechin (flavonoid class) as determined by total relative abundance in digested extract. Procyanidin dimers and trimers were also not detected after digestion.

Threshing (glume removal) globally did not impact the phenolic bioaccessibility, except for 4-hydroxybenzoic acid (-61% in total relative abundance) (Figure 23B). On the other hand, dehulling showed multiple effects, depending on the compound analyzed: 1) syringic acid significantly increased (+203%), while 2) (+)-catechin, p-coumaric and 4-hydroxybenzoic acid were reduced (-51, -67 and -86%, respectively); 3) pyrogallol, caffeic and isoferulic acids remained unchanged after dehulling.

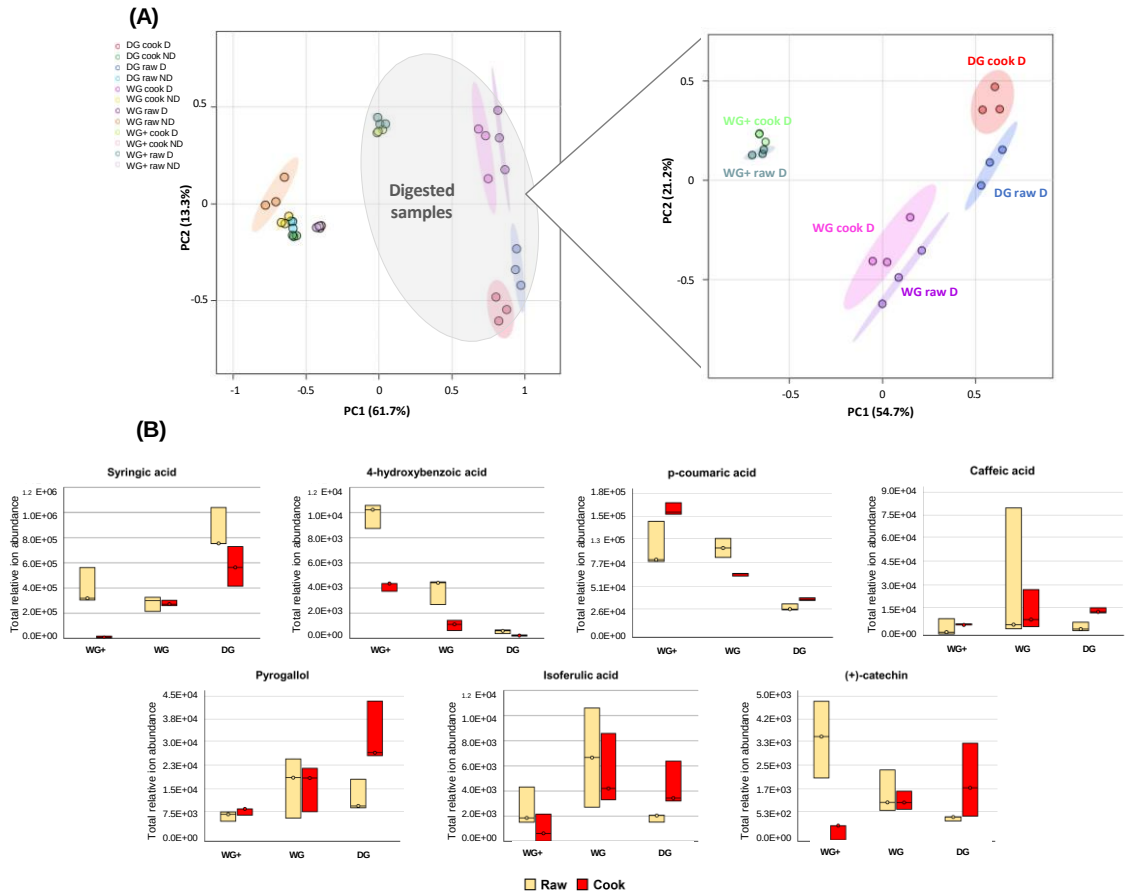


Figure 15. Overview of phenolic compounds profile after simulated *in vitro* digestion. (A) principal component analysis (PCA) biplot before and after INFOGEST digestion; and (B) variation in the expression levels of seven level 1 phenolic compounds during different processing methods. The samples (symbols) are distributed according to relative intensity of identified phenolic compounds. WG+ = whole grain with glume; WG = whole grain; DG = dehulled grain; B = bran.

After cooking, phenolic bioaccessibility of each fraction was also impacted differently. WG+ had reduced bioaccessibility of syringic acid, 4-hydroxybenzoic acid and (+)-catechin (-98, -58 and -89%, respectively), and increased bioaccessibility for p-coumaric acid (+42%) following heat treatment. In WG, cooking did not change the bioaccessibility of most compounds, except for 4-hydroxybenzoic and p-coumaric acid (-73 and -30%, respectively). Finally, in DG there was an increase in p-coumaric acids (+28%), caffeic (+202%), isoferulic (+130%), (+)-catechin (+156%) and pyrogallol (+140%).

It is important to note that bioaccessibility variation of each phenolic compound is influenced by the proportion of free and bound contents in the non-digested form. There was a significant increase in the free-to-bound ratio in caffeic (0.9 to 1.5), p-coumaric (0.8 to 1.3) and isoferulic acids (0.9 to 1.1), pyrogallol (0 to 1.2) and (+)-catechin (0.2 to 5.4) after cooking-dehulling combined process. According to the literature, soluble free phenolics may be released more easily from the food matrix during GI tract and may be more bioaccessible to human body (Gao *et al.*, 2024). So, the combination of physical and thermal processes can allow better absorption of these compounds and, therefore, their enhanced action on human health. Furthermore, the increase in the free-to-bound ratio of catechin suggests that cooking may lead to greater tannin depolymerization, since catechins are components of the procyanidin base (Rue *et al.*, 2018).

In contrast, 4-hydroxybenzoic acid was the only to exhibit a decrease in free-to-bound ratio after dehulling (1.0 to 0.7) and cooking (1.0 to 0.8). The proportional increase of conjugated 4-hydroxybenzoic acid via ester or ether bonds was strongly correlated with its reduction in bioaccessibility after both processes (negative correlation; $r = 0.98$, $p = 3.58 \times 10^{-6}$). These data corroborate Gao *et al.* (2024) report, which found that the release of phenolics bound to the food matrix during GI transit was scarce.

3.8. Correlation analysis

Sorghum nutritional components (protein and ash content) showed a strong positive correlation ($r = 0.92$, $p < 5 \times 10^{-4}$) (Supplementary Figure 5). Physicochemical characterization analyses of sorghum grains revealed moderate to strong positive correlations among grain weight (GW), grain density, grain hardness (GH), and colorimetric parameters (L*, a*, and b*). However, grain size (Wu *et al.*) showed no correlation with the other parameters. Sorghum phenolic compounds (PC) demonstrated a strong positive correlation with physicochemical characteristics (Supplementary Figure

5). However, flavonoids (no correlation, indicated by white rectangles) and phenolic acids (negative correlation, indicated by blue rectangles) did not appear to be responsible for this correlation. *In vitro* protein digestibility (IVPD) had a strong and moderate positive correlation with GH and PC, respectively; while procyanidins ($r = -0.87$, $p < 5x10^{-4}$) and phenolic acids contents ($r = -0.90$, $p < 5x10^{-4}$) had a negative impact on IVPD. Bioaccessibility was negatively correlated with protein content (strong; $r = -0.83$, $p < 5x10^{-4}$) and positively correlated with PC (moderate; $r = 0.61$, $p < 0.05$) and flavonoids (strong; $r = 0.83$, $p < 5x10^{-4}$). No correlation between IVPD and phenolic bioaccessibility was found.

Based on the correlation analyses, the physicochemical characteristics of sorghum grains may be associated with their phenolic content. Interestingly, procyanidins and phenolic acids had a similar impact on the protein digestibility of the sorghum grains studied. Furthermore, the phenolic profile significantly influenced bioaccessibility of the compounds, with flavonoid-rich grains being the most suitable.

4. Conclusion

This study showed the significant impact of sorghum processing on *in vitro* protein digestibility and phenolic bioaccessibility. Cooking negatively impacted protein digestibility, suggesting the formation of protease-resistant starch-protein complexes and protein disulfide cross-linking during the wet thermal process; however, the cooking-dehulling combination or glume presence in black pericarp grains partially mitigate this effect.

After cooking, the free-to-bound phenolic ratio increased, indicating the release of phenolics bound to the cell wall. However, the dehulling impacted on phenolic variability. Glume removal reduced glycosylated flavonoids and plant pigments, while dehulling also reduced the hydroxycinnamic and hydroxybenzoic acids contents (*e.g.* isoferulic and 4-hydroxybenzoic acids). Procyanidin dimers and trimers were also reduced after dehulling, which may contribute with the improved protein digestibility. Dehulling-cooking combination was effective for increasing bioaccessibility of several phenolics (*e.g.*, catechin, pyrogallol, isoferulic and caffeic acids), but this effect seems to depend on a high free-to-bound phenolic ratio.

Dutch sorghum grains were characterized for the first time, which could boost sorghum cultivation in cold Northern Europe. The combination of grain processing techniques demonstrates possible new and promising impacts for human nutrition. Future

research should continue to explore the complex interactions among other processing methods and grain composition to optimize sorghum-based foods and to maximize the health benefits of these new varieties of sorghum.

5. Acknowledgements and funding

This work was financed by UNIRIO, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (401053/2019-9; 407350/2023-3, 314100/2023-7); Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (E-26/201.317/2022; E-26/210.152/2023; E-26/210.008/2024); and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (code 001 and PDSE).

6. Supplementary material

Supplementary Table 3. Chemical scores of essential amino acids in different sorghum genotypes.

AMINO ACIDS (AA)	Essential AA (EAA) (mg/g protein)				Ideal AA profile[1]		EAA scores (%)			
	#HD7	#HD19	#HD100	#HD101	mg/g protein	Score (%)	#HD7	#HD19	#HD100	#HD101
Histidine	21	24	23	22	16	100	132	147	141	139
Isoleucine	36	40	38	35	30	100	122	133	127	117
Leucine	109	93	102	84	61	100	179	153	167	138
Lysine	22	22	23	23	48	100	47	47	48	49
Methionine + Cysteine	15	24	20	18	23	100	64	103	87	77
Phenylalanine + Tyrosine	77	91	82	81	41	100	188	222	200	197
Threonine	28	30	30	29	25	100	114	121	119	116
Valine	46	50	48	46	40	100	115	126	120	114
Total EAA	355	374	365	338	284	100	125	132	129	119

[1] Ideal profile contains all the essential AA in sufficient quantities compared to the needs of the human adult and their chemical score is 100% (FAO, 2011). An AA score below 100% is considered as limiting and does not meet the body's needs. The tryptophan was not analyzed.

Supplementary Table 4. Total putative identification of phenolic compounds in the sorghum extracts by UHPLC-MS^{E}.

Putative identification	ID level*	Molecular Formula	*m/z*	RT (min)	Score	FS (%)	Error (ppm)	IS (%)	Class	Subclass
Trihydroxy-prenylflavanone isomer 1	2	C20H20O5	339.1266	1.05	36.2	0.0	8.30	90.21	F	Flavanones
Daidzein O-glucuronide isomer 1	2	C21H18O10	429.0797	1.21	35.7	0.0	-6.96	86.20	F	Isoflavonoids
Myricetin 3-O-rhamnoside	2	C21H20O12	463.0848	1.32	36.7	0.0	-7.35	91.51	F	Flavonols
Ellagic acid arabinoside isomer 1	2	C19H14O12	433.0404	1.46	38.2	0.0	-1.99	93.36	PA	Hydroxybenzoic acids
Unknown compound 1	3	C22H20O11	459.0895	1.69	36.5	0.0	-8.17	91.77	F	Isoflavonoids
Pyrogallol	**1**	**C6H6O3**	**125.0236**	**1.80**	**37.1**	**0.0**	**-6.55**	**93.04**	**OP**	**Other polyphenols**
Trihydroxy-prenylflavanone isomer 2	2	C20H20O5	339.1261	1.82	35.9	2.4	6.88	84.94	F	Flavanones
Gallic acid	**1**	**C7H6O5**	**169.0129**	**1.84**	**37.0**	**0.0**	**-8.20**	**94.00**	**PA**	**Hydroxybenzoic acids**
Prodelphinidin dimer B3	2	C30H26O14	609.1225	2.00	35.6	0.0	-4.04	82.47	F	Flavanols
Unknown compound 2	3	C21H20O11	447.0899	2.16	36.3	0.0	-7.46	89.96	F	Flavones
Unknown compound 3	3	C23H22O11	473.1057	2.39	39.4	17.1	-6.76	87.48	F	Isoflavonoids
Protocatechuic acid 4-O-glucoside	2	C13H16O9	315.0691	2.40	35.8	0.0	-9.71	89.84	PA	Hydroxybenzoic acids
4-Hydroxybenzoic acid 4-O-glucoside	2	C13H16O8	299.0744	2.44	37.0	0.0	-9.58	95.74	PA	Hydroxybenzoic acids
Quercetin 3-O-(6"-malonyl-glucoside)	2	C24H22O15	549.0866	2.50	37.1	0.0	-3.57	89.79	F	Flavonols
Unknown compound 4	3	C21H20O6	367.1207	2.74	43.0	33.4	5.36	87.58	OP	Curcuminoids
6"-O-Acetylglycitin	2	C24H24O11	487.1220	3.10	38.5	2.0	-5.29	96.42	F	Isoflavonoids
(-)-Epigallocatechin 3-O-gallate	2	C22H18O11	457.0738	3.33	36.6	0.0	-8.45	92.61	F	Flavanols
Myricetin-O-arabinoside	2	C20H18O12	449.0698	3.48	36.3	0.0	-6.19	88.33	F	Flavonols
Unknown compound 5	3	C23H22O11	473.1059	3.48	37.1	0.0	-6.36	92.49	F	Isoflavonoids
O-Malonyldaidzin isomer 1	2	C24H22O12	501.1027	3.76	42.5	30.6	-2.34	84.90	F	Isoflavonoids
Procyanidin trimer isomer 1	2	C45H38O18	865.2012	3.80	38.4	11.9	3.04	83.49	F	Flavanols

(+)-Gallocatechin/(-)-Epigallocatechin	2	C15H14O7	305.0636	4.21	36.2	0.0	-9.92	92.07	F	Flavanols
Unknown compound 6	3	C22H22O12	477.1022	4.33	37.5	0.0	-3.46	91.79	F	Flavonols
O-Dimethylepigallocatechin-O-gallate isomer 1	2	C24H22O11	485.1060	4.46	37.3	0.0	-5.98	93.16	F	Flavanols
Caffeoylquinic acid isomer 1	2	C16H18O9	353.0844	4.49	37.1	0.0	-9.76	96.01	PA	Hydroxycinnamic acids
(+)-Catechin-O-glucose isomer	2	C21H24O11	451.1228	4.59	36.7	0.0	-4.01	88.19	F	Flavanols
Procyanidin trimer isomer 2	2	C45H38O18	865.1999	4.60	41.3	25.3	1.53	83.24	F	Flavanols
3'-O-Methyl-(-)-epicatechin 7-O-glucuronide	2	C22H24O12	479.1177	4.65	38.5	0.0	-3.68	96.81	F	Flavanols
Prodelphinidin dimer B4	2	C30H26O14	609.1233	4.70	37.0	0.0	-2.82	88.42	F	Flavanols
Quercetin 3,4'-O-diglucoside	2	C27H30O17	625.1372	4.70	35.7	0.0	-6.07	85.50	F	Flavonols
4-Hydroxybenzoic acid	**1**	**C7H6O3**	**137.0225**	**4.71**	**54.8**	**86.2**	**–9.04**	**97.96**	**PA**	**Hydroxybenzoic acids**
(+)-Catechin 3-O-glucose	2	C21H24O11	451.1225	4.78	36.9	0.0	-4.62	89.87	F	Flavanols
Procyanidin trimer isomer 3	2	C45H38O18	865.2010	4.80	36.7	0.0	2.86	86.72	F	Flavanols
Demethoxycurcumin	2	C20H18O5	337.1107	4.82	36.6	0.0	7.50	91.47	OP	Curcuminoids
Caffeic acid 4-O-glucoside/1-Caffeoyl-beta-D-glucose 1	2	C15H18O9	341.0849	4.87	37.4	0.0	-8.59	96.48	PA	Hydroxycinnamic acids
2,5-Dihydroxybenzoic acid	**1**	**C7H6O4**	**153.0175**	**5.02**	**50.2**	**37.9**	**-7.62**	**99.16**	**PA**	**Hydroxybenzoic acids**
Ellagic acid	2	C14H6O8	301.0012	5.10	36.9	0.0	7.45	92.73	PA	Hydroxybenzoic acids
O-Methyl-(-)-epicatechin-O-glucuronide isomer 1	2	C22H24O12	479.1174	5.14	36.8	0.0	-4.32	88.80	F	Flavanols
Dihydroxy-dimethoxyisoflavone	2	C17H14O6	313.0687	5.34	36.7	0.0	-9.63	93.83	F	Isoflavonoids
Procyanidin dimer isomer 1	2	C30H26O12	577.1342	5.42	39.1	0.0	-1.64	97.38	F	Flavanols
Esculin	2	C15H16O9	339.0692	5.60	36.4	0.0	-8.72	91.88	OP	Hydroxycoumarins
Ferulic acid 4-O-glucoside	2	C16H20O9	355.1010	5.62	36.9	0.0	-6.88	92.38	PA	Hydroxycinnamic acids

Caffeic acid 4-O-glucoside/1-Caffeoyl-beta-D-glucose 2	2	C15H18O9	341.0850	5.68	38.0	0.0	-8.22	99.30	PA	Hydroxycinnamic acids
Unknown compound 7	3	C22H18O12	473.0703	5.71	37.3	0.3	-4.79	91.48	PA	Hydroxycinnamic acids
Ellagic acid acetyl-arabinoside/Ellagic acid acetyl-xyloside	2	C21H16O13	475.0500	5.71	36.9	0.1	-3.71	88.54	PA	Hydroxybenzoic acids
Procyanidin trimer isomer 4	2	C45H38O18	865.2002	5.73	42.6	16.1	1.87	99.37	F	Flavanols
p-Coumaric acid 4-O-glucoside	2	C15H18O8	325.0897	5.80	36.8	0.0	-9.83	94.53	PA	Hydroxycinnamic acids
(-)-Epicatechin 3-O-gallate	2	C22H18O10	441.0792	5.82	37.7	0.0	-7.87	97.33	F	Flavanols
O-Dimethylepigallocatechin-O-gallate isomer 2	2	C24H22O11	485.1064	5.87	36.6	0.0	-5.21	89.20	F	Flavanols
Caffeoylquinic acid isomer 2	2	C16H18O9	353.0843	5.91	37.6	0.0	-9.78	98.65	PA	Hydroxycinnamic acids
Hydroxy-tetramethoxyflavone isomer 1	2	C19H18O7	357.0947	5.96	37.0	0.0	-9.20	94.98	F	Flavones
Ellagic acid arabinoside isomer 2	2	C19H14O12	433.0425	5.98	37.2	0.0	2.87	89.23	PA	Hydroxybenzoic acids
Glycitein-O-glucuronide isomer 1	2	C22H20O11	459.0898	6.08	37.1	0.0	-7.64	93.98	F	Isoflavonoids
Genistein-O-glucuronide isomer 1	2	C21H18O11	445.0740	6.10	38.2	7.4	-8.15	92.67	F	Isoflavonoids
4"-O-Methylepigallocatechin 3-O-gallate isomer	2	C23H20O11	471.0899	6.12	36.8	0.0	-7.17	92.31	F	Flavanols
Procyanidin trimer isomer 5	2	C45H38O18	865.1976	6.16	37.7	0.0	-1.09	89.98	F	Flavanols
1-Caffeoyl-5-feruloylquinic acid	2	C26H26O12	529.1313	6.18	36.7	3.5	-7.34	88.49	PA	Hydroxycinnamic acids
Caffeic acid	**1**	**C9H8O4**	**179.0331**	**6.25**	**50.7**	**61.9**	**–6.61**	**99.13**	**PA**	**Hydroxycinnamic acids**
Quercetin-O-diglucoside	2	C27H30O17	625.1407	6.32	38.4	0.0	-0.57	92.72	F	Flavonols
Kaempferol-O-triglucoside	2	C33H40O21	771.2005	6.35	39.6	2.1	2.09	98.19	F	Flavonols
4"-O-Methylepigallocatechin 3-O-gallate	2	C23H20O11	471.0913	6.39	37.9	0.5	-4.11	93.93	F	Flavanols
Salvianolic acid B isomer 1	2	C18H12O7	339.0477	6.44	36.9	0.0	-9.68	95.12	OP	Other polyphenols

O-Dimethylepigallocatechin-O-gallate isomer 3	2	C24H22O11	485.1062	6.46	37.6	0.6	-5.65	93.96	F	Flavanols
O-Methyl-(-)-epicatechin-O-glucuronide isomer 2	2	C22H24O12	479.1163	6.48	37.1	0.6	-6.71	92.62	F	Flavanols
Apigenin/genistein glucoside 1	2	C21H20O10	431.0953	6.52	36.2	0.1	-7.11	89.12	F	Flavanones
Isorhamnetin-O-glucoside-O-rhamnoside isomer 1	2	C28H32O16	623.1565	6.52	35.4	0.0	-8.45	86.35	F	Flavonols
Syringic acid	**1**	**C9H10O5**	**197.0438**	**6.53**	**37.0**	**0.0**	**–8.9**	**94.90**	**PA**	**Hydroxybenzoic acids**
p-coumaroyl tartaric acid isomer	2	C13H12O8	295.0431	6.53	35.8	0.0	-9.72	89.75	PA	Hydroxycinnamic acids
4'-O-Methyl-(-)-epigallocatechin-O-glucuronide	2	C22H24O13	495.1099	6.58	36.7	0.0	-9.12	93.62	F	Flavanols
Urolithin B 3-O-glucuronide	2	C19H16O9	387.0688	6.60	36.3	0.0	-8.57	91.14	OP	Hydroxycoumarins
6"-O-Malonylgenistin	2	C24H22O13	517.0949	6.71	37.0	0.0	-7.47	93.53	F	Isoflavonoids
Apigenin/genistein glucoside 2	2	C21H20O10	431.0951	6.76	36.2	0.0	-7.53	89.24	F	Flavanones
Glycitein-O-glucuronide isomer 2	2	C22H20O11	459.0909	6.79	37.9	0.2	-5.18	95.40	F	Isoflavonoids
Salvianolic acid B isomer 2	2	C18H12O7	339.0477	6.82	36.7	0.0	-9.78	94.04	OP	Other polyphenols
Myricetin 3-O-arabinoside	2	C20H18O12	449.0726	6.84	38.8	0.0	0.05	94.02	F	Flavonols
(+)-catechin	**1**	**C15H14O6**	**289.0710**	**6.87**	**57.0**	**92.9**	**-2.65**	**95.07**	**F**	**Flavanols**
Spinacetin 3-O-glucosyl-(1->6)-glucoside isomer	2	C29H34O18	669.1677	6.92	37.6	1.1	0.72	87.83	F	Flavonols
4'-O-Methylepigallocatechin	2	C16H16O7	319.0795	7.01	37.4	0.0	-8.90	96.88	F	Flavanols
Genistein-O-glucuronide isomer 2	2	C21H18O11	445.0746	7.08	37.5	1.9	-6.78	93.47	F	Isoflavonoids
Kaempferol 3-O-xylosyl-glucoside	2	C26H28O15	579.1340	7.11	38.8	0.0	-2.68	97.18	F	Flavonols
Eriodictyol 7-O-glucoside	2	C21H22O11	449.1053	7.12	37.3	0.0	-7.99	95.48	F	Flavanones
cis/trans-resveratrol-O-glucuronide 1	2	C20H20O9	403.1004	7.27	37.4	1.5	-7.64	94.11	S	Stilbenes

Salvianolic acid C isomer 1	2	C26H20O10	491.0956	7.28	37.6	1.6	-5.59	92.55	OP	Other polyphenols
Dicaffeoyl tartaric acid	2	C22H18O12	473.0703	7.30	37.2	0.0	-4.69	91.64	PA	Hydroxycinnamic acids
O-Methyl-(-)-epicatechin-O-glucuronide isomer 3	2	C22H24O12	479.1165	7.32	37.3	0.0	-6.35	93.50	F	Flavanols
Dihydroxy-trimethoxyflavone isomer 1	2	C18H16O7	343.0792	7.34	40.7	19.6	-9.14	93.97	F	Flavones
Phloretin 2'-O-glucuronide	2	C21H22O11	449.1055	7.34	38.0	3.1	-7.58	95.45	F	Dihydrochalcones
Procyanidin trimer isomer 6	2	C45H38O18	865.1985	7.39	36.9	0.0	-0.02	84.70	F	Flavanols
Quercetin 3-O-acetyl-rhamnoside	2	C23H22O12	489.1019	7.45	38.5	5.4	-4.02	91.84	F	Flavonols
p-coumaroyl tartaric acid	2	C13H12O8	295.0449	7.46	37.2	0.0	-3.56	90.40	PA	Hydroxycinnamic acids
Cinnamoyl glucose isomer	2	C15H18O7	309.0978	7.46	37.7	0.5	-0.66	88.77	PA	Hydroxycinnamic acids
p-coumaric acid	**1**	**C9H8O3**	**163.0381**	**7.48**	**57.1**	**95.6**	**–7.60**	**98.45**	**PA**	**Hydroxycinnamic acids**
Procyanidin dimer isomer 2	2	C30H26O12	577.1337	7.48	38.8	0.0	-2.50	96.80	F	Flavanols
Isorhamnetin-O-glucoside-O-rhamnoside isomer 2	2	C28H32O16	623.1602	7.51	37.8	0.0	-2.50	92.15	F	Flavonols
Diosmetin 7-O-rutinoside	2	C28H32O15	607.1610	7.53	35.4	0.0	-9.59	87.32	F	Flavones
O-Dimethylepigallocatechin-O-gallate isomer 4	2	C24H22O11	485.1066	7.55	42.2	23.2	-4.74	93.24	F	Flavanols
Unknown compound 8	3	C14H6O8	300.9978	7.57	37.3	0.0	-3.97	91.24	PA	Hydroxybenzoic acids
Unknown compound 9	3	C23H22O10	457.1117	7.61	37.9	0.0	-5.04	95.07	F	Isoflavonoids
Hydroxy-tetramethoxyflavone isomer 2	2	C19H18O7	357.0957	7.62	37.7	0.0	-6.34	95.88	F	Flavones
Unknown compound 10	3	C21H20O11	447.0892	7.64	36.5	0.0	-9.12	92.74	F	Flavones
Glycitein-O-glucuronide isomer 3	2	C22H20O11	459.0911	7.66	38.0	0.0	-4.67	95.56	F	Isoflavonoids

5,3',4'-Trihydroxy-3-methoxy-6:7-methylenedioxyflavone 4'-O-glucuronide	2	C23H20O14	519.0806	7.68	36.3	0.0	4.90	87.34	F	Flavonols
Catechin 5-O-gallate	2	C22H18O10	441.0797	7.69	36.0	0.0	-6.92	87.67	F	Flavanols
Rhoifolin 4'-O-glucoside	2	C33H40O19	739.2073	7.69	36.0	0.6	-2.47	82.51	F	Flavones
Equol-O-glucuronide	2	C21H22O9	417.1150	7.71	35.1	0.0	-9.77	86.40	F	Isoflavonoids
Daidzein-O-glucuronide isomer 2	2	C21H18O10	429.0803	7.73	37.2	0.0	-5.52	92.58	F	Isoflavonoids
Epicatechin-O-glucuronide	2	C21H22O12	465.1011	7.73	35.8	0.8	-5.92	85.21	F	Flavanols
Salvianolic acid B isomer 3	2	C18H12O7	339.0477	7.77	36.5	0.0	-9.71	93.22	OP	Other polyphenols
cis/trans-resveratrol-O-glucuronide 2	2	C20H20O9	403.1006	7.80	37.5	0.1	-7.00	95.30	S	Stilbenes
Deoxyschisandrin	2	C24H32O6	415.2135	7.82	38.6	0.0	2.13	95.55	L	Lignans
Unknown compound 11	3	C21H20O11	447.0901	7.84	37.5	0.7	-7.01	94.85	F	Flavones
Salvianolic acid C isomer 2	2	C26H20O10	491.0957	7.87	36.7	0.0	-5.37	89.81	OP	Other polyphenols
Theaflavin isomer	2	C29H24O12	563.1169	7.89	36.3	0.0	-4.61	86.62	F	Flavanols
Patuletin 3-O-(2"-feruloylglucosyl)(1->6)-[apiosyl(1->2)]-glucoside	2	C43H48O25	963.2353	7.91	43.5	41.1	-6.08	83.11	F	Flavonols
Quercetin 3-glucosyl-(1->2)-xyloside	2	C26H28O16	595.1300	7.93	38.2	0.0	-0.69	91.95	F	Flavones
Unknown compound 12	3	C21H20O11	447.0908	7.98	37.0	3.4	-5.54	88.13	F	Flavones
Kaempferol 3-O-(2"-rhamnosyl-galactoside) 7-O-rhamnoside	2	C33H40O19	739.2088	8.00	38.4	8.3	-0.41	84.34	F	Flavones
Isoferulic acid	**1**	**C10H10O4**	**193.0486**	**8.01**	**47.8**	**61.4**	**-5.41**	**99.04**	**PA**	**Hydroxycinnamic acids**
Apigenin 7-O-glucoside	2	C21H24O9	419.1344	8.05	38.9	0.0	-0.93	95.54	F	Flavones
Quercetin 3-O-rhamnosyl-galactoside	2	C27H30O16	609.1447	8.07	38.9	0.0	-2.30	97.28	F	Flavonols
Salvianolic acid B isomer 4	2	C18H12O7	339.0488	8.09	37.8	0.0	-6.49	96.21	OP	Other polyphenols
Chrysoeriol 7-O-apiosyl-glucoside	2	C27H30O15	593.1498	8.11	36.6	1.2	-2.40	84.65	F	Flavones

Taxifolin 3-O-rhamnoside	2	C21H22O11	449.1071	8.14	38.8	1.7	-4.03	96.76	F	US
1-Sinapoyl-2-feruloylgentiobiose	2	C33H40O18	723.2082	8.17	36.7	1.2	-8.31	91.36	PA	Hydroxycinnamic acids
Spinacetin 3-O-glucosyl-(1->6)-glucoside	2	C29H34O18	669.1656	8.19	38.8	0.7	-2.41	96.25	F	Flavonols
3'-O-Methylviolanone	2	C18H18O6	329.1004	8.23	37.8	0.0	-8.16	97.89	F	Isoflavonoids
Naringin 6'-malonate	2	C30H34O17	665.1694	8.25	46.6	49.0	-4.40	89.28	F	Flavanones
6-Geranylnaringenin	2	C25H28O5	407.1889	8.30	37.2	1.2	6.09	91.67	F	Flavanones
Quercetin 3-O-arabinoside/xyloside	2	C20H18O11	433.0744	8.30	35.5	0.0	-7.56	85.89	F	Flavonols
Isorhamnetin 3-O-rutinoside	2	C22H22O11	461.1067	8.34	38.1	1.7	-4.89	94.28	F	Flavonols
Hesperetin 7-O-rutinoside	2	C28H34O15	609.1775	8.35	35.7	0.0	-8.13	87.45	F	Flavanones
Paeoniflorin isomer 1	2	C23H28O11	479.1545	8.39	37.4	0.0	-2.81	90.41	PA	Hydroxybenzoic acids
Quercetin 3'-sulfate	2	C15H10O10S	380.9891	8.41	37.6	0.0	-8.19	97.37	F	Flavonols
(+)-Gallocatechin/(-)-Epigallocatechin 3-O-gallate	2	C22H18O11	457.0746	8.43	37.4	0.0	-6.56	94.49	F	Flavanols
Puerarin	2	C21H20O9	415.1005	8.44	37.5	0.0	-7.12	95.34	F	Isoflavonoids
Pentamethoxyflavone isomer 1	2	C20H20O7	371.1103	8.46	36.8	0.0	-8.86	94.03	F	Flavones
Quercetin 3-galactoside 7-rhamnoside	2	C27H30O16	609.1513	8.48	35.2	2.0	8.50	83.25	F	Flavonols
Isorhamnetin-O-galactoside	2	C22H22O12	477.1025	8.51	38.3	0.0	-2.81	94.89	F	Flavonols
Salvianolic acid C isomer 3	2	C26H20O10	491.0966	8.51	36.6	0.0	-3.67	87.43	OP	Other polyphenols
Theaflavin 3-O-gallate	2	C36H28O16	715.1366	8.53	35.9	7.8	8.51	81.31	F	Flavanols
1,2-Disinapoylgentiobiose	2	C34H42O19	753.2250	8.59	37.1	0.0	0.27	85.87	PA	Hydroxycinnamic acids
Tectoridin	2	C22H22O11	461.1073	8.61	38.0	0.0	-3.56	94.14	F	Isoflavonoids
Trihydroxy-dimethoxyflavone isomer 1	2	C17H14O7	329.0635	8.62	37.0	0.0	-9.56	95.64	F	Flavonols
Isorhamnetin-O-glucoside-O-rhamnoside isomer 3	2	C28H32O16	623.1600	8.64	38.0	1.1	-2.80	92.15	F	Flavonols

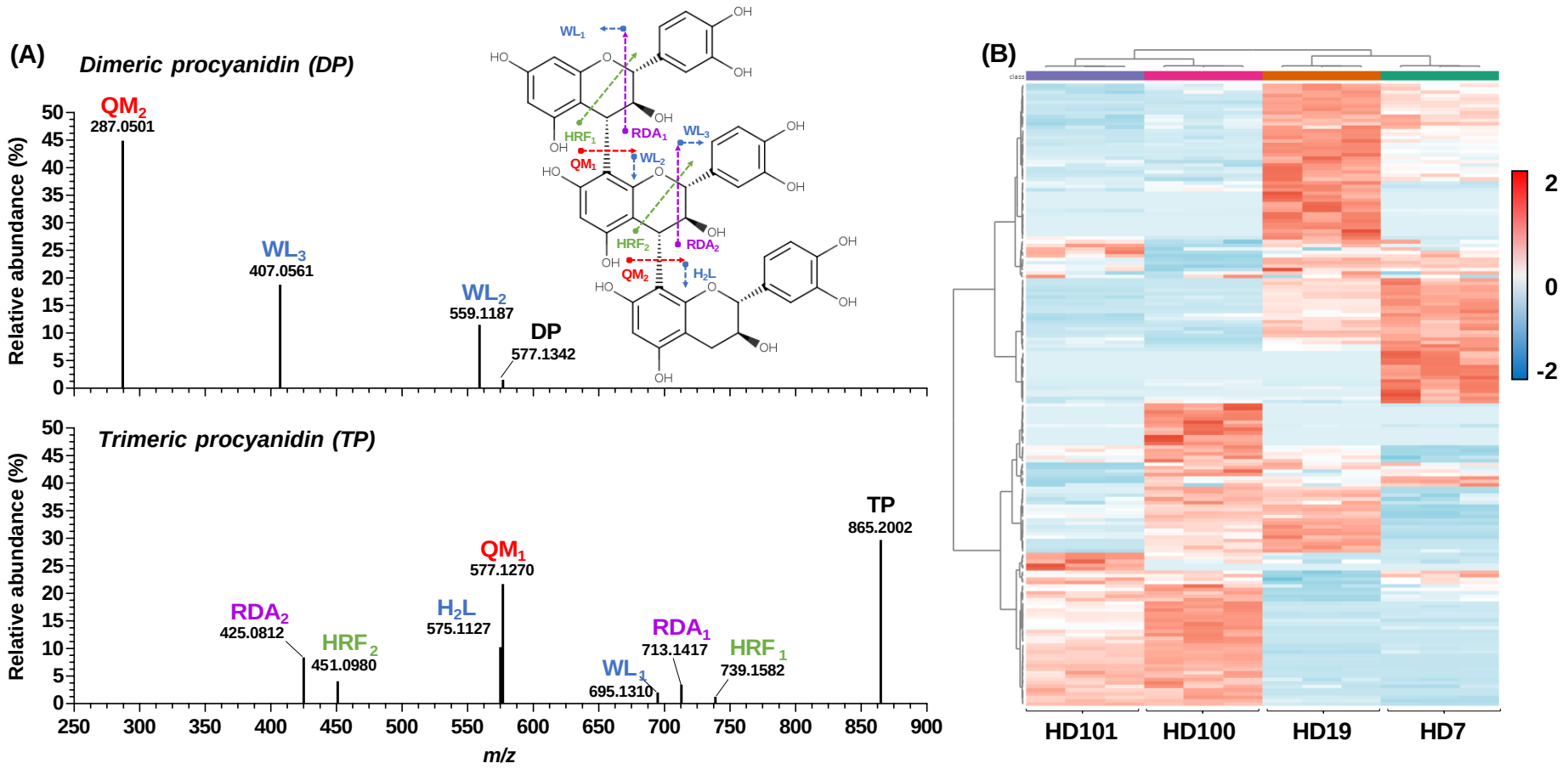
Eriodictyol 7-O-rutinoside	2	C27H32O15	595.1628	8.71	37.8	0.0	-6.86	96.95	F	Flavanones
Diosmetin 7-O-neohesperidoside	2	C28H32O15	607.1647	8.75	35.7	0.0	-3.59	82.66	F	Flavones
6"-O-Acetylgenistin	2	C23H22O11	473.1076	8.77	38.0	0.0	-2.71	93.41	F	Isoflavonoids
Unknown compound 13	3	C17H16O6	315.0843	8.80	37.5	0.0	-9.92	98.18	F	Isoflavonoids
Procyanidin dimer isomer 3	2	C30H26O12	577.1373	8.83	41.2	18.4	3.70	92.17	F	Flavanols
Quercetin 3-O-(6"-malonyl-glucoside) isomer	2	C24H22O15	549.0896	8.85	37.6	3.9	1.82	86.43	F	Flavonols
Prodelphinidin dimer B5	2	C30H26O14	609.1247	8.85	37.3	0.1	-0.47	86.76	F	Flavanols
Isorhamnetin-O-glucoside-O-rhamnoside isomer 4	2	C28H32O16	623.1616	8.88	38.1	0.0	-0.29	90.86	F	Flavonols
O-Dimethylepigallocatechin-O-gallate isomer 5	2	C24H22O11	485.1058	8.89	35.5	0.2	-6.54	84.85	F	Flavanols
Quercetin-O-glucoside	2	C21H20O12	463.0913	8.91	35.3	0.0	6.72	84.29	F	Flavonols
Dicaffeoylquinic acid isomer	2	C25H24O12	515.1155	8.96	35.7	5.6	-7.68	81.56	PA	Hydroxycinnamic acids
Kaempferol 3-O-sophoroside 7-O-glucoside	2	C33H40O21	771.1974	8.96	36.4	1.2	-1.94	83.13	F	Flavonols
Theaflavin	2	C29H24O12	563.1174	9.00	37.7	4.8	-3.80	88.26	F	Flavanols
Dihydroquercetin 3-O-rhamnoside	2	C21H22O11	449.1070	9.04	38.6	2.5	-4.26	95.54	F	Dihydroflavonols
Unknown compound 14	3	C22H22O11	461.1082	9.04	40.9	12.1	-1.59	94.24	F	Isoflavonoids
O-Methyl-(-)-epigallocatechin-O-glucuronide isomer	2	C22H24O13	495.1108	9.04	36.5	2.1	-7.38	88.93	F	Flavanols
Unknown compound 15	3	C23H22O10	457.1115	9.05	38.4	4.1	-5.53	94.36	F	Isoflavonoids
Hydroxy-tetramethoxystilbene isomer 1	2	C17H18O5	301.1064	9.07	37.0	2.6	-5.87	89.26	S	Stilbenes
Methoxynobiletin isomer 1	2	C22H24O9	431.1329	9.11	36.5	0.0	-4.36	87.46	F	Flavonols
Pentamethoxyflavone isomer 2	2	C20H20O7	371.1124	9.17	37.9	0.0	-3.32	93.65	L	Lignans

Paeoniflorin isomer 2	2	C23H28O11	479.1542	9.17	41.3	15.6	-3.44	94.76	PA	Hydroxybenzoic acids
Myricetin	**1**	**C15H10O8**	**317.0297**	**10.18**	**57.2**	**93.9**	**-1.71**	**94.12**	**F**	**Flavonols**
Quercetin-O-galactoside	2	C21H20O12	463.0888	9.21	36.9	0.2	1.36	86.12	F	Flavonols
Dehydrodiferulic acid isomer 1	2	C20H18O8	385.0901	9.25	37.4	0.0	-7.27	95.31	PA	Hydroxycinnamic acids
Procyanidin dimer isomer 4	2	C30H26O12	577.1344	9.28	37.8	0.0	-1.33	90.57	F	Flavanols
Sativanone isomer 1	2	C17H16O5	299.0906	9.33	37.2	0.0	-6.49	93.54	F	Isoflavonoids
Salvianolic acid B isomer 5	2	C18H12O7	339.0495	9.39	37.0	0.0	-4.59	90.56	OP	Other polyphenols
Procyanidin dimer isomer 5	2	C30H26O12	577.1343	9.44	37.0	0.1	-1.53	86.53	F	Flavanols
Methoxynobiletin isomer 2	2	C22H24O9	431.1332	9.46	37.9	0.3	-3.53	93.15	F	Flavonols
Prodelphinidin dimer B6	2	C30H26O14	609.1251	9.50	36.6	0.0	0.17	82.98	F	Flavanols
Tetramethylscutellarein	2	C19H18O6	341.1001	9.52	37.8	0.0	-8.75	98.58	F	Flavones
Unknown compound 16	3	C22H26O8	417.1591	9.55	35.8	3.2	8.69	85.64	L	Lignans
(+)-Gallocatechin 3-O-gallate	2	C22H18O11	457.0753	9.60	37.9	0.0	-5.16	95.25	F	Flavanols
Dehydrodiferulic acid isomer 2	2	C20H18O8	385.0904	9.62	38.0	0.0	-6.44	97.30	PA	Hydroxycinnamic acids
Daidzein-O-glucuronide	2	C21H18O10	429.0795	9.66	37.1	0.0	-7.49	94.03	F	Isoflavonoids
Trihydroxy-dimethoxyflavone isomer 2	2	C17H14O7	329.0635	9.68	36.6	0.0	-9.75	93.76	F	Flavonols
O-Dimethylepigallocatechin-O-gallate isomer 6	2	C24H22O11	485.1073	9.73	38.1	0.0	-3.27	94.47	F	Flavanols
Hydroxy-O-desmethylangolensin	2	C15H14O5	273.0777	9.75	36.4	0.0	2.96	85.74	F	US
Procyanidin dimer isomer 6	2	C30H26O12	577.1341	9.75	38.2	0.0	-1.85	93.40	F	Flavanols
O-Malonylglycitin isomer 1	2	C25H24O13	531.1141	9.80	38.2	0.0	-0.66	91.95	F	Isoflavonoids
Hydroxymatairesinol	2	C20H22O7	373.1261	9.82	37.4	0.5	-8.58	95.79	L	Lignans
Sativanone isomer 2	2	C17H16O5	299.0907	9.86	37.5	0.1	-6.00	94.47	F	Isoflavonoids

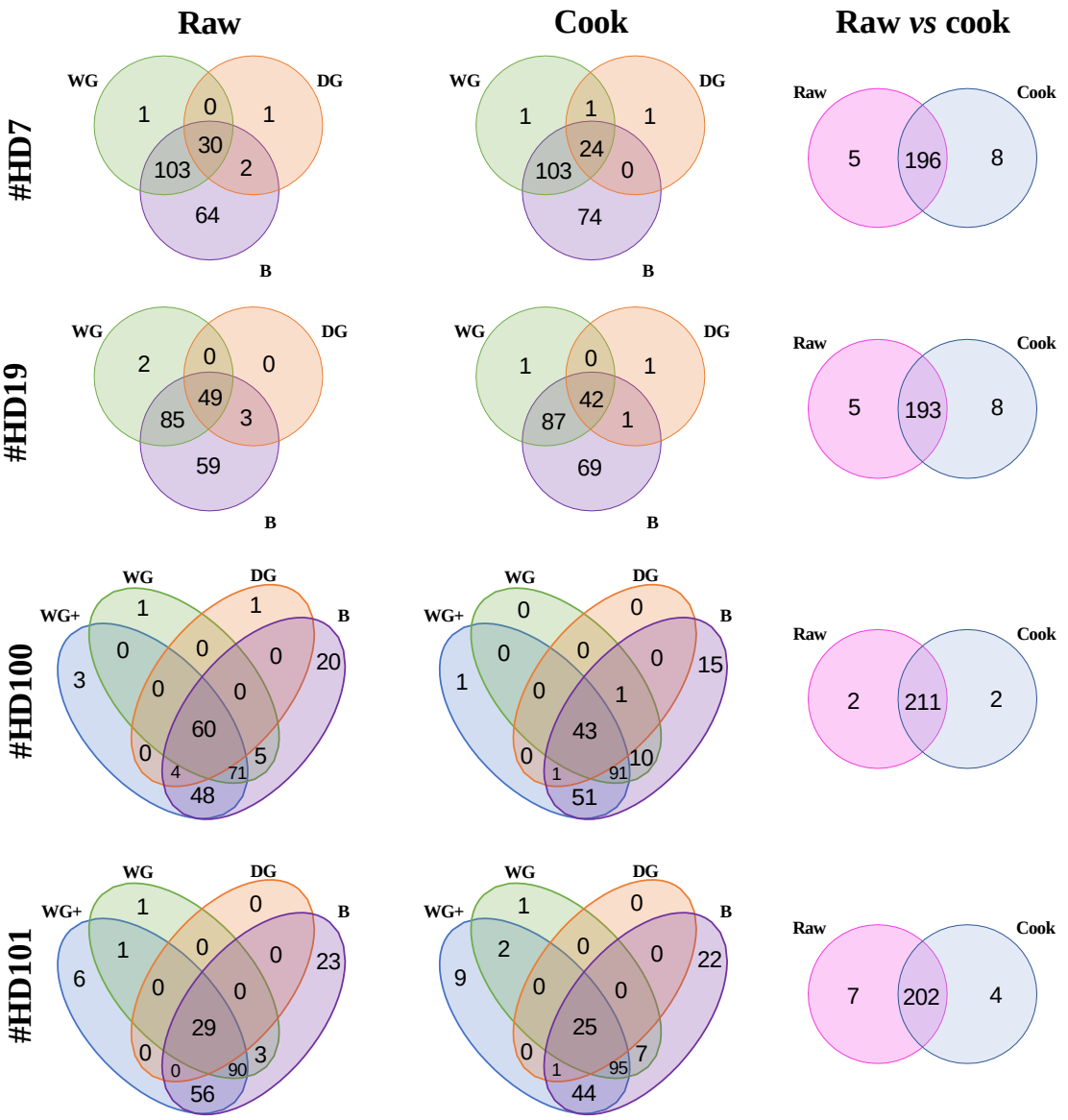
Daidzin	2	C21H20O9	415.1004	9.96	38.0	0.0	-7.28	98.01	F	Isoflavonoids
O-Malonylglycitin isomer 2	2	C25H24O13	531.1136	9.96	36.9	0.0	-1.45	86.41	F	Isoflavonoids
Patuletin 3-O-glucosyl-(1->6)-[apiosyl(1->2)]-glucoside	2	C33H40O22	787.1900	10.02	36.7	0.0	-4.83	88.95	F	Flavonols
Hydroxyenterolactone isomer	2	C18H18O5	313.1056	10.04	37.8	0.0	-8.14	98.18	L	Lignans
6'-Hydroxyangolensin/5'-Methoxy-O-desmethylangolensin	2	C16H16O5	287.0899	10.09	38.3	4.5	-9.12	97.22	F	Isoflavonoids
Kaempferol 3-O-sophoroside 7-O-glucoside/Kaempferol 3,7,4'-O-triglucoside/Quercetin 3-O-glucosyl-rhamnosyl-galactoside	2	C33H40O21	771.1931	10.14	37.1	0.3	-7.49	93.68	F	Flavonols
Kaempferol	**1**	**C15H10O6**	**285.0398**	**10.18**	**57.3**	**95.2**	**–2.31**	**94.04**	**F**	**Flavonols**
Procyanidin dimer isomer 7	2	C30H26O12	577.1344	10.20	38.5	0.0	-1.37	94.11	F	Flavanols
Quercetin	2	C15H10O7	301.0324	10.21	37.2	0.0	-9.91	96.76	F	Flavonols
Hesperetin	2	C16H14O6	301.0688	10.26	37.2	0.0	-9.75	96.64	F	Flavanones
p-Coumaroylquinic acid isomer	2	C16H18O8	337.0901	10.30	37.2	0.0	-8.12	95.21	PA	Hydroxycinnamic acids
Unknown compound 17	3	C22H22O12	477.1037	10.32	38.5	0.0	-0.28	92.82	F	Flavonols
Syringaresinol	2	C22H26O8	417.1524	10.36	39.7	20.4	-7.35	86.42	L	Lignans
O-Malonyldaidzin isomer 2	2	C24H22O12	501.1038	10.41	36.9	0.4	-0.14	84.08	F	Isoflavonoids
Violanone	2	C17H16O6	315.0846	10.50	37.7	0.0	-8.88	98.51	F	Isoflavonoids
Dicaffeoylquinic acid isomer 2	2	C25H24O12	515.1170	10.57	36.7	0.0	-4.89	89.14	PA	Hydroxycinnamic acids
Episesamin	2	C20H18O6	353.1000	10.62	37.6	0.0	-8.59	97.61	L	Lignans
Oleuropein-aglycone	2	C19H22O8	377.1209	10.66	40.5	18.3	-8.58	93.88	OP	Tyrosols
1-Acetoxypinoresinol	2	C22H24O8	415.1378	10.75	37.7	0.0	-4.89	94.18	L	Lignans
6"-O-Acetyldaidzin	2	C23H22O10	457.1117	10.78	38.4	0.2	-4.97	97.63	F	Isoflavonoids

Medioresinol/Trachelogenin 1	2	C21H24O7	387.1420	10.82	36.9	0.0	-7.46	92.90	L	Lignans
Cinnamoyl glucose	2	C15H18O7	309.0985	10.86	38.6	3.4	1.62	91.42	PA	Hydroxycinnamic acids
Enterolactone	2	C18H18O4	297.1104	10.91	36.1	0.0	-9.54	91.10	L	Lignans
Curcumin	2	C21H20O6	367.1157	11.02	37.9	0.0	-8.23	98.66	OP	Curcuminoids
Hydroxy-tetramethoxyflavone isomer 3	2	C19H18O7	357.0949	11.07	37.9	0.0	-8.65	99.06	F	Flavones
Medioresinol/Trachelogenin 2	2	C21H24O7	387.1413	11.10	43.5	32.1	-9.22	95.55	L	Lignans
Unknown compound 18	3	C17H16O6	315.0846	11.13	37.3	0.0	-8.75	95.98	F	Isoflavonoids
Trihydroxy-dimethoxyflavone isomer 3	2	C17H14O7	329.0634	11.26	36.6	0.0	-9.87	93.66	F	Flavonols
Procyanidin dimer isomer 8	2	C30H26O12	577.1338	11.48	37.7	0.0	-2.40	91.56	F	Flavanols
Hydroxy-tetramethoxystilbene isomer 2	2	C17H18O5	301.1058	11.71	38.1	0.0	-7.92	99.40	S	Stilbenes
Dihydroformononetin	2	C16H14O4	269.0819	11.75	39.1	7.1	-0.10	88.68	F	Isoflavonoids
Curcumin isomer	2	C21H20O6	367.1159	11.91	37.0	0.0	-7.63	93.52	OP	Curcuminoids
Dihydroxy-trimethoxyflavone isomer 2	2	C18H16O7	343.0794	12.19	36.4	0.0	-8.38	91.33	F	Flavones

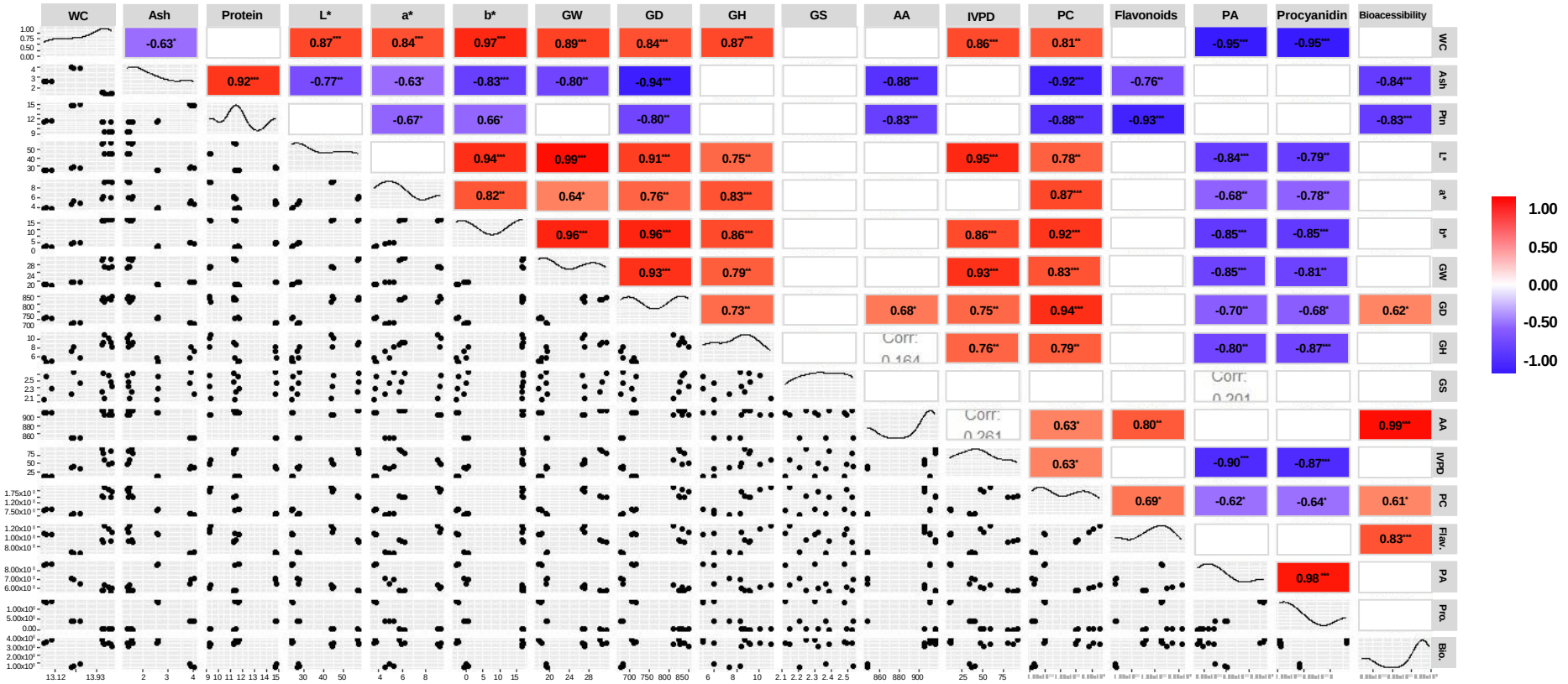
* Identification level is according to Sumner et al. (2007) where 1 means identified compounds with commercial standards; 2 means putatively annotated compounds base d upon physicochemical properties and/or spectral similarity; 3 means putatively characterized compound classes; and 4 are unknown compounds. *m/z* = mass/charge ratio; RT = retention time; FS = fragmentation score; IS = isotope similarity; PA = phenolic acids; F = flavonoids; OP = other polyphenols; L = lignans; S = stilbenes; and US = unassigned subclass.



Supplementary Figure 3. Overview of metabolomics data of sorghum grain samples: (A) MS/MS fragmentation ions of procyanidin dimer and trimer detected in MS^E negative mode; and (B) Hierarchical clustering analysis (HCA) heat map of the 219 annotated phenolic compounds in sorghum genotypes. Quinone methide (QM), Retro-Diels-Alder (RDA), Heterolytic ring fission (HRF), water loss (WL), hydrogen loss (H_2L).



Supplementary Figure 4. Venn diagrams with the number of identifications in each genotype, comparing WG and grain fractions when raw and when cooked and the overall cooking effect. WG+ = whole grain with glume; WG = whole grain; DG = dehulled grain; B = bran.



Supplementary Figure 5. Pairwise Pearson correlation between whole-grain analyses. Matrix correlation coefficients are represented by a color scale: white represents a correlation coefficient of zero, while darker blues to reds represent stronger negative and positive correlations, respectively. Histograms for all traits are displayed along the diagonal. To the left and below the diagonal are scatter plots comprising sorghum four genotypes. *** = 0.0005 significance level, ** = 0.005 significance level, * = 0.05 significance level. WC = water content; Ptn = protein content; GW = grain weight; GD = grain density; GH = grain hardness; GS = grain size; AA = amino acids; IVPD = *in vitro* protein digestibility; PC = phenolic compounds; Flav. = flavonoids; PA = phenolic acids; Pro. = procyanidins; and Bio. = bioaccessibility.

Chapter V - Tracking the changes and bioaccessibility of phenolic compounds of sorghum grains (*Sorghum bicolor* (L.) Moench) upon germination and seedling growth by UHPLC-QTOF-MS/MS

Carolina Thomaz dos Santos D'Almeida[1,2], Lynda Abdelbost[3], Hamza Mameri[3] and Mariana Simões Larraz Ferreira[1-2*]

[1] Laboratory of Bioactives (LABBIO), Food and Nutrition Graduate Program (PPGAN), Federal University of the State of Rio de Janeiro, UNIRIO, Brazil.

[2] Center of Innovation in Mass Spectrometry, Laboratory of Protein Biochemistry (LBP-IMasS), UNIRIO, Brazil.

[3] UMR IATE, Univ. Montpellier, INRAE, Institut-Agro Montpellier, F-34060 Montpellier, France.

Highlights

- Sorghum phenolic profile was revealed during germination and seedling growth
- Germination favored the decomplexation of compounds previously bound to cell-matrix
- Procyanidin dimers were found in tannin-free genotypes at low levels
- Germination increased phenolic acids *in vitro* bioaccessibility, mainly in oral phase
- Flavonoids bioaccessibility was progressively reduced towards seedling growth

This chapter has been approved for publication in *Food Research International* (Impact Factor: 7.0). DOI: 10.1016/j.foodres.2024.114854.

ABSTRACT

In this study, phenolic profile/content was analyzed by high-resolution untargeted metabolomics after short germination (72h) and seedling growth (144h), using three sorghum genotypes varying in tannin content (IS 29569, Macia and IS 30400) . *In vitro* antioxidant capacity and phenolic bioaccessibility were determined by microplate-based and INFOGEST methods, respectively. A total of 58% annotated compounds were found in all genotypes; and phenolic acids and flavonoids represent more than 80% of sorghum total abundance. PCA analysis showed higher phenolic variability in germination times (72%) than genotypes (51%). Germination reduced total ion abundance (-7%) and free:bound phenolic compounds ratio (2.4 to 1.1), but antioxidant capacity remained constant. These results indicate the cell matrix-phenolic decomplexation, with the free compounds were quickly consumed after radicle emergence. Germination increased phenolic bioaccessibility (mainly in oral phase) but reduces flavonoids contents in gastric/intestinal digestion steps. This work can stimulate seed germination as a viable option for sorghum-based foods development, with improved nutritional and bioactive properties.

Keywords: Seed germination, sorghum, polyphenols, *in vitro* digestion, antioxidant capacity

1. Introduction

Sorghum (*Sorghum bicolor* L. Moench) is a versatile cereal grain with a vital role in the agriculture and food production in different regions of the world. Currently, sorghum grain is the 5th most important cereal in the world, with an annual production of 61 million tons in 2021 (USDA, 2024a). Due to its resistance to adverse agro-climatic conditions, such as heat and drought stresses, sorghum has become a staple food in regions with limited water resources, *e.g.*, Africa and Asia which are responsible for more than 50% of sorghum production (Stefoska-Needham *et al.*, 2015). Recently, sorghum grains have also garnered growing attention in developed regions as a valuable food to be incorporated to the human diet, not only due to its gluten-free nature, but also to its substantial nutritional value and health-promoting potential (D'Almeida *et al.*, 2021b; de Morais Cardoso *et al.*, 2017).

Sorghum grains contain high levels of bioactive compounds and shows a wide range of phenolic compounds (Xiong *et al.*, 2019b). Phenolic compounds are known as plant-specialized metabolites with well reported bioactivities related to the inhibition or reduction of oxidative stress, inflammatory, mutagenic, and carcinogenic processes (Lutz *et al.*, 2019). Cereal phenolics, including those of sorghum grains, are also related to the modulation of the intestinal microbiota and the improvement of barrier function (Ed Nignpense *et al.*, 2021; Tiozon *et al.*, 2022). These phenolic compounds are found mainly bound to cell wall components (Dykes & Rooney, 2006; Santos *et al.*, 2019b) and, therefore, are less available to be absorbed by the human body (Ed Nignpense *et al.*, 2021). Merely a small fraction, (5–10%), can be absorbed in the small intestine and subsequently transported into the blood circulation system, thus reaching the target tissues (Ed Nignpense *et al.*, 2021; Tiozon *et al.*, 2022). The non-bioaccessible fraction moves directly to the colon (Ed Nignpense *et al.*, 2021). Therefore, to overcome this limitation and improve the accessibility of phenolic compounds, different technological processing has been applied to sorghum grain, obtaining satisfactory results (D'Almeida *et al.*, 2021b; Rashwan *et al.*, 2021).

Germination is a low-cost and traditional bioprocess used for several sorghum-based products (Arouna *et al.*, 2020). During this bioprocess, hydrolytic enzymes are activated and breakdown macronutrients into their simpler forms, improving the nutritional value and reducing anti-nutritional factors. From a functional point of view,

germination impacts the phenolic compounds profile, synthesizing new bioactive compounds involved in total antioxidant activity (Arouna *et al.*, 2020; Ghinea *et al.*, 2021; Singh *et al.*, 2019). The balance between free and bound phenolics can be changed by the enzymatic release, for instance, by endogenous esterase of compounds originally bound within seed tissues enhancing their bioaccessibility (Xu *et al.*, 2020). However, this effect depends on the germination binomial conditions (temperature and time) as well as the genotype (Pilco-Quesada *et al.*, 2020).

Metabolic pathways involved in the synthesis/biotransformations of these compounds, as well as the possible impact of their bioaccessibility on gastrointestinal tract, are scarcely documented particularly for sorghum grain. Despite the international INFOGEST network has proposed a static *in vitro* digestion model widely used to simulate human gastrointestinal conditions (Brodkorb *et al.*, 2019), a standardized method for measuring final phenolic bioaccessibility concentrations has not yet established. Due to the complexity of the sorghum metabolome, untargeted LC-MS metabolomics in combination with chemometric data analyzes could be a promising way to decipher the sorghum metabolites spectrum through a single protocol (Tiozon *et al.*, 2022). Untargeted metabolomics is an exploratory method that can reveal new information about the structures and functions of phenolic compounds in sorghum grains (Tiozon *et al.*, 2022); while the application of chemometric statistical analyzes elucidates possible biomarkers of cereal processing and its variations in phenolic bioaccessibility (D'Almeida *et al.*, 2021b; Rocchetti *et al.*, 2018).

Given the context, this work aimed to 1) evaluate the germination effect on phenolic compounds of three sorghum genotypes varying in tannin contents by MS-based metabolomics, and the possible changes in phenylpropanoids pathway involved in this bioprocess; 2) investigate the health-promoting role of sorghum seed germination by bioaccessibility and antioxidant potential, using *in vitro* models (INFOGEST simulated three-phases digestion and microplate-based analysis, respectively).

2. Material and methods

2.1. Plant material and germination conditions

Sorghum seeds (*Sorghum bicolor L.*) from three different races were grown in INRAE experimental field (Mauguio, Southern France) and kindly provided by UMR-

AGAP Institut (Montpellier): sorghum kafir (genotype IS 29569, pigmented testa, red pericarp), caudatum (cv. Macia, without pigmented testa, white pericarp) and kafir-caudatum hybrid (genotype IS 30400, without pigmented testa, white pericarp).

Germination was carried out according to Abdelbost *et al.* (2023). Seeds underwent a sterilization process for 20 minutes using sodium hypochlorite solution 5% (1/5, w/v), followed by three rinse steps with distilled water. Thirty-five seeds of each genotype were distributed into Petri dishes (Ø 10 cm) filled with 3 sheets of Whatman filter paper imbibed with 6 mL of deionized water. The dishes were maintained in a climate chamber (Memmert, France) at 16 °C and 60% relative humidity (RH). Samples of 35 seeds were collected at three different times (in hours), corresponding to different stages of germination: 0 h (representing ungerminated seeds, labeled as UG), 72 h and 144 h. These samples were rapidly frozen in liquid nitrogen and stored at -40°C for further analysis.

The seed samples were freeze-dried for 48 h (Freeze drying Christ ALPHA 1–4, Germany), until they reached a final moisture content of approximately 10%. Finally, the freeze-dried grains were cryogenically milled in a ball mill (Retsch MM 400) at 25 Hz during 4 min, and the resulting flours were stored at −20 °C for subsequent analyses.

2.2. Flavonoids extraction in acidified acetone

Flavonoids and proanthocyanidins from 30 mg of samples were extracted with acetone, water, and acetic acid (70:29.5:0.5, v/v/v) at a ratio of 1:20 (w/v) and stirred for 1 h at 25 °C (200 rpm) (modified from Rao *et al.*, 2018). The extracts were centrifuged (12,000 *xg*, 10 min, 25 °C) and extraction was repeated three times. The supernatants were pooled together before acetone evaporation (Savant, SpeedVac., ThermoFisher Scientific, USA). Dried extracts were resolubilized in 500 µL of methanol 50%, filtered (0.22 µm hydrophilic PTFE) (Analytical) and stored in vials at −80 °C until use.

2.3. Total proanthocyanidin (TPAC) and total flavonoid content (TFC)

The condensed tannin (proanthocyanidins) content was determined in triplicate based on the vanillin assay (Rao *et al.*, 2018), with some modifications. An aliquot of reconstituted sorghum acidified acetone extract (50 µL) was combined with 125 µL of 1% (w/v) vanillin and 125 µL of 25% sulphuric acid, both prepared in methanol, and incubated (37 °C, 15 min). For TFC assay, an aliquot of 100 µL of the acetone extract

was mixed in test tubes with 500 µL of Milli-Q water and 30 µL of 5% sodium nitrite. After 5 min incubation, 30 µL of 10% $AlCl_3$ was added and kept in the dark for 6 min. Then 200 µL of 1 M NaOH was added, followed by the addition of 200 µL of Milli-Q water. Absorbance was measured at 500 nm and 510 nm, respectively (FlexStation III microplate reader, Molecular Devices, USA), and results were expressed as mg of catechin equivalents (CE) per g of sample, in dry basis.

2.4. Free and bound phenolic compounds extraction

Phenolic compounds from sorghum ground grains were extracted in triplicate according to Santos *et al.* (2019b). **Free phenolic compounds (FPC, soluble)** – FPC were extracted using 70 mg of sample and 50 mg of celite, manually macerated, and then extracted with 1 mL of 80% ethanol. Samples were stirred (200 rpm, 10 min, 25 °C) and centrifuged (5000×G, 10 min, 25 °C). The extraction was repeated twice, and the collected supernatants were combined. **Bound phenolic compounds (BPC, insoluble)** - the remaining pellets were submitted to alkaline hydrolysis with 4 mL of NaOH 4M (submerged in an ultrasonic bath - 42 kHz, 90 min, 40 °C). After, the acid hydrolysis was performed with concentrated HCl (~pH 2), and the samples were centrifuged (2000×G, 5 min, 25 ºC). The supernatant was washed three times with ethyl acetate (7 mL) and centrifuged between each step (10000×G, 5 min, 10 °C). The extracts (FPC and BPC) were evaporated (SpeedVac Savant, ThermoFisher Scientific, USA) and resuspended in 1.5 mL of methanol, acetonitrile and Milli-Q water (2:5:93, v/v/v). The extracts were filtered (0.22 µm, hydrophilic PTFE) (Analytical) and stored in vials at −80 °C.

2.5. Total reducing capacity

The total reducing capacity (TRC) was determined based on Folin-Ciocalteu method adapted for microplates, according to D'Almeida *et al.* (2021b). Briefly, extracts (100 µL), Milli-Q water (700 µL), Folin-Ciocalteu reagent (50 µL) and 20% sodium carbonate (150 µL) were mix and incubated (30 min, 40°C). The absorbance reading was performed at 750 nm (FlexStation III microplate reader, Molecular Devices). Solvent blank and standard curve analyze were performed with gallic acid (5 to 130 µg/mL). Results were expressed in mg of gallic acid equivalents (GAE) per g of sample, in dry basis.

2.6. Determination of antioxidant capacity

The antioxidant capacity of samples was determined, in triplicate, by the DPPH (2,2-diphenyl-1-picrylhydrazyl) radical scavenging method, and the ferric reducing antioxidant power (FRAP) method adapted to microplates, according to D'Almeida *et al.* (2021b). **DPPH method** - a 20 µL aliquot of each extract was combined with 280 µL of the DPPH solution (32 µg/mL) and incubated (30 min, in the dark, 25 °C). **FRAP method** - the reagent was prepared in acetate buffer (0.3 M, pH 3.6), $FeCl_3.6H_2O$ (20 mM) and TPTZ (2,4,6-tri(2-pyridyl)-1,3,5-triazine) solution (10 mM) in a 10:1:1 ratio. A 20 µL aliquot of each extract was combined with 15 µL of Milli-Q water and 265 µL of FRAP reagent, gently vortexed and incubated (30 min, 37 °C).

Absorbance was measured using a microplate reader (FlexStation III, Molecular Devices, USA) at 715 nm and 595 nm for DPPH and FRAP assays, respectively, and results were expressed as µmol of trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid) equivalents (TE) per g of sample, in dry basis.

2.7. Simulated *in vitro* gastrointestinal digestion

The *in vitro* digestion process was performed in duplicate based on the INFOGEST method (Brodkorb *et al.*, 2019), with adaptations. This process consists of a 3-step procedure that simulates the physiological and pH conditions in the mouth, stomach, and small intestine, performed at 37 ºC. Briefly, 250 mg of each sample was mixed with 1 mL of simulated salivary fluid (pH 7) containing a-amylase (Sigma-Aldrich, cat. no. 1031) from human saliva rapidly added at 75 U/mL (final concentration). After 2 min under agitation (Multi Bio RS-24, Biosan, Brazil), the a-amylase was inactivated in acidic conditions (HCl 6 M) and the gastric phase was initiated. Simulated gastric fluid (pH 3, 4.82 mL) containing pepsin (Sigma-Aldrich, cat. no. P70000; final concentration: 1,000 U/mL) were added to the post-oral digestion tube (pellet + supernatant). The gastric phase was performed under agitation for 2 h and the pepsin was denatured under alkaline conditions (NaOH 3 M). Finally, the intestinal phase was performed adding 7.2 mL of simulated intestinal fluid (pH 7) containing pancreatin (Sigma-Aldrich, cat. no. P7545, 50 U/mL) to the post-gastric digestion tube; and stirring for 2 h. The tube containing the three digestion steps was centrifuged (13000 xg, 20 min, 4 °C) and the supernatant was obtained. Also, for improved recovery of the digested

extract, the pellet was added with 1 mL of MilliQ water, centrifuged (13000 xg, 5 min, 4 °C) and the supernatants were pooled.

2.8. Untargeted metabolomics of sorghum phenolic profile and bioaccessibility by UHPLC-QTOF-MS/MS

Ethanolic extracts (FPC and BPC) obtained from sorghum ground grains (section 2.4) and samples obtained by simulated digestion (oral, gastric, and intestinal steps; section 2.7) were injected (5 µL of each sample) into an Ultra High Performance Liquid Chromatography (UHPLC) Acquity system (Waters, USA) coupled with XEVO G2S Q-Tof (Waters, England) equipped with ionization source electrospray. An UHPLC HSS T3 C18 column (100 x 2.1 mm, 1.8 µm particle diameter; Waters) at 30 °C and flow rate of 0.5 mL/min of ultra-pure water containing 0.3% formic acid and 5 mM ammonium formate (mobile phase A) and acetonitrile containing 0.3% formic acid (mobile phase B) was used according to the gradient method: : 0 min - 97% A; 11.80 min - 50% A; 12.38 min - 15% A; 14.11 - 97% A. Data were acquired in triplicate in multiplex MS^E DIA (data independent analysis), negative and centroid modes between *m/z* 50 and 1200, collision energy ramp from 30 to 55 V; cone voltage 30 V; capillary voltage 3.0 kV; desolvation gas (N_2) 1,200 L/h at 600 °C; cone gas 50 L/h; source at 150 °C; and using leucine enkephalin (Leu-Enk, *m/z* 554.2615, [M-H]-) for calibration and MassLynx v. 4.1 software (Waters, USA) for MS data acquisition. A mix containing 33 analytical standards of phenolic compounds (10 ppm) was prepared and injected in triplicate, prior to the injection of the samples, to ensure the reproducibility of the instrument and to confirm phenolic compounds identification. Besides the injection of the chemical standards, a set of quality control (QC) samples was also prepared by pooling equal volumes of each sorghum extracts and were injected after each batch of six runs of sorghum samples to monitor the instrument's stability.

2.9. UHPLC-QTOF-MS/MS data processing

The raw data was first normalized by using Progenesis QI (Waters, Nonlinear Dynamics) to enable annotation/identification and quantitative comparisons among the different samples. The metabolites identification was based on chemical standards runs parameters, such as: isotope distribution of neutral mass, exact mass, retention time and MS/MS fragments spectra. To make clear the level of confidence in metabolite non-targeted identification, a level system was applied according to Sumner *et al.* (2007), starting with features matching commercial standards (level 1). Tentative annotation was

performed with customized database built from PubChem (www.pubchem.ncbi.nlm.nih.gov/) and online database Phenol-Explorer (www.phenol-explorer.eu/). In this case, compounds (level 2) and/or classes (level 3) were annotated following the parameters applied in descending order of importance: exact mass error (<10 ppm); isotopic similarity (>80%); score (>30) and the highest score of MS/MS fragmentation, obtained by the software. Data from the literature and chemical characteristics of the molecules were also used to help compounds annotation. Finally, reliable, and quantifiable features, but which were not assigned to any compound or class, were defined as "unknown compounds" (level 4).

2.10. Statistical analysis

All analyses were performed in triplicates, and the results were represented by mean and standard deviation. Statistical analysis was performed with Tukey's test ($p < 0.05$) and one-way ANOVA, using XLSTAT software (Addinsoft, France). Data generated from Progenesis QI were exported to perform a Principal Component Analysis (PCA) by XLSTAT. The differential abundance calculated from the total ion counting of compounds in the samples was performed by pairwise comparison. The volcano plots were built considering the compounds present in two compared conditions after application of two filters: log2 fold change ± 1.0 and ANOVA $p < 0.05$. The bioaccessibility of each phenolic compound was estimated by the total relative ion abundance in each digestion phase of raw (UG) and germinated sorghum.

3. Results and discussions

3.1. Phenolic compounds and antioxidant capacity

The TRC was used to estimate phenolic compounds in both, free and bound forms, in the different samples ungerminated (UG) and germinated at 72 h and 144 h for comparison purposes (Table 3). Among UG samples those of IS 29569 showed higher TRC (30.62±0.77 mg GAE/g), followed by Macia (21.39±0.49 mg GAE/g) and IS 30400 (18.07±0.54 mg GAE/g). The main difference was found between the free extracts of the different genotypes ($p<0.05$). The presence of phenolic acids and flavonoids, responsible for color in pigmented testa sorghum grains, can explain the superiority of IS 29569, in relation to Macia (+43.2%) and IS 30400 (+69.4%) genotypes. The presence of tannins in IS 29569 can also be an explanation of this result, especially due to the presence of oligomers of proanthocyanidins probably present in free extracts. Dykes et al. (2005)

showed that sorghum varieties with red and thick pericarp, but differing by the pigmented testa showed the highest TRC (averaging 8 mg GAE/g) and tannin content. In a new approach NIR-based, Rhodes *et al.* (2014) analyzed different accessions of the same sorghum races studied (kafir, kafir-caudatum and caudatum), and found lower results than those of the present study (6.02 ± 4.05 to 9.08 ± 5.86 mg GAE/g) but similar to the TRC values determined in the free extracts. They also showed a strong positive correlation between TRC and proanthocyanidins content, confirming that higher TRC contents can be related to tannins.

Globally, germination positively impacted TRC, but temporally was slightly dissimilar between genotypes. IS 29569 and Macia showed no statistical difference after 72h when compared to UG ($p > 0.05$), but in 144h both genotypes showed 20% increase in their TRC, due to the increase in free extracts. In IS 30400, the increase was progressive (+8% and +62% in 72h and 144h, respectively, both compared to UG). Again, the positive impact of germination and seedling growth was clearly observed in free extracts; especially in the latter genotype that showed a 3.5-fold increase after 144h. Germination increased the FPC/BPC ratio in IS 29569 (1.1 to 1.2), IS 30400 (0.3 to 1.1) and Macia (0.8 to 1.3) but IS 29569 was the only genotype with higher levels of free than bound phenolics (52%).

Table 3. Phenolic content and antioxidant activities determined by DPPH and FRAP methods in ungerminated sorghum extracts (UG) and after 72h and 144h of germination.

Samples	G (h)	Extract	TRC (mg GAE/g)	DPPH (µmol TE/g)	FRAP (µmol TE/g)	TPAC (mg CE/g)	TFC (mg CE/g)
IS 29569 *(kafir, pigmented testa)*	**UG**	Free	16.47±0.39^{b}	8.71±0.44^{a}	0.73±0.03^{b}	5.36±1.18^{a}	8.43±0.26^{a}
		Bound	14.60±0.74^{a}	5.26±0.86^{a}	0.59±0.01^{b}		
		Total	**30.62±0.77^{b}**	**13.97±0.85^{a}**	**1.32±0.04^{a}**		
	72h	Free	16.61±1.06^{b}	7.97±0.15^{a}	0.87±0.21^{a}	5.38±1.25^{a}	8.45±0.59^{a}
		Bound	16.22±0.92^{a}	6.23±0.17^{a}	0.68±0.04^{a}		
		Total	**32.83±1.74^{b}**	**14.21±0.30^{a}**	**1.54±0.04^{a}**		
	144h	Free	20.45±0.39^{a}	8.62±0.64^{a}	0.87±0.11^{a}	7.93±0.56^{a}	8.81±0.96^{a}
		Bound	16.47±0.67^{a}	6.23±0.24^{a}	0.65±0.02^{ab}		
		Total	**36.92±0.81^{a}**	**14.86±0.80^{a}**	**1.53±0.12^{a}**		
IS 30400 (*kafir-caudatum*, without pigmented testa)	**UG**	Free	4.35±0.26^{c}	3.33±0.96^{a}	0.07±0.02^{b}	nd	1.64±0.24^{a}
		Bound	13.72±0.45^{a}	4.20±1.22^{a}	0.69±0.12^{a}		
		Total	**18.07±0.54^{b}**	**7.53±2.18^{a}**	**0.77±0.13^{ab}**		
	72h	Free	5.99±0.09^{b}	3.41±1.09^{a}	0.09±0.02^{b}	nd	1.62±0.29^{a}
		Bound	13.45±0.31^{a}	5.74±0.19^{a}	0.48±0.01^{b}		
		Total	**19.44±0.15^{b}**	**9.15±1.04^{a}**	**0.57±0.03^{b}**		
	144h	Free	15.10±0.30^{a}	0.75±0.61^{b}	0.23±0.02^{a}	nd	1.66±0.45^{a}
		Bound	14.11±0.47^{a}	5.99±1.31^{a}	0.60±0.04^{a}		
		Total	**29.21±0.73^{a}**	**6.73±1.60^{a}**	**0.83±0.01^{a}**		
Macia (*caudatum*, without pigmented testa)	**UG**	Free	9.56±0.51^{b}	4.71±0.97^{a}	0.30±0.07^{ab}	nd	3.24±0.31^{a}
		Bound	11.83±0.15^{a}	6.91±1.03^{a}	0.45±0.06^{a}		
		Total	**21.39±0.49^{b}**	**11.62±1.99^{a}**	**0.74±0.10^{a}**		
	72h	Free	9.17±0.49^{b}	4.49±0.52^{a}	0.24±0.04^{b}	nd	2.85±0.22^{a}
		Bound	12.07±0.87^{a}	6.64±0.77^{a}	0.38±0.07^{a}		
		Total	**21.24±0.38^{b}**	**11.13±1.21^{a}**	**0.62±0.09^{a}**		
	144h	Free	14.43±1.28^{a}	5.81±0.38^{a}	0.45±0.10^{a}	nd	3.26±0.13^{a}
		Bound	11.28±1.03^{a}	6.34±0.21^{a}	0.34±0.01^{a}		
		Total	**25.72±2.19^{a}**	**12.14±0.24^{a}**	**0.78±0.09^{a}**		

Results are expressed on dry basis as average ± standard deviation (n=3). Different letters mean a significant difference ($p < 0.05$) in the same column between the same extracts, following a one-way ANOVA test. GAE (gallic acid equivalent), TE (Trolox equivalent), CE (catechin equivalent), TRC (total reducing content), DPPH (2-diphenyl-1-picrylhydrazyl), FRAP (ferric reducing antioxidant power), TPAC (total proanthocyanidin content) and TFC (total flavonoid content), nd (not detected).

According to Xu *et al.* (2020), germination can increase FPC by two potential routes: 1) during this bioprocessing, there is the activation of the endogenous hydrolytic enzymes and, consequentially, the breaking of bonds between phenolic compounds and macromolecules such as cellulose, protein, lignin, and sometimes due to covalent bonds (*e.g.*, ester, ether, and C−C bonds) in the grain cell wall. This enzymatic action would positively impact FPC and reduce BPC, however in our work no reduction was found in BPC content, the data from the present study showed a positive correlation between FPC and BPC ($r = 0.6927$, $p = 0.04$). 2) during seed germination, protein, starch, and lipids can be decomposed into amino acids, glucose, and acetyl CoA, substrates of phenolic synthesis. Due to the increase in intracellular mobility during germination (attributed to the water supply during the imbibition stage), enzymes, such as phenylalanine/tyrosine ammonia-lyase, are activated/synthesized and can convert these substrates into phenolic acids (El-Soud *et al.*, 2013). On the other hand, the accumulation of FPC active enzymes (such as glucosyltransferase, galactosyltransferase and acyltransferase) to esterifies phenolic acids with soluble carbohydrates, also increasing BPC (Ono *et al.*, 2010). According to Abdelbost *et al.* (2023), catalytic, lyases and transferase enzymes were three important subgroups in the molecular function category during sorghum germination, enzymes involved in the catabolism pathways of the seed storage macromolecules favor the degradation of complex macromolecules into simpler molecules, new compounds are synthesized (lyases) and esterified (transferases). From our results, this second route seems to be preferentially adopted; involving a greater extent of the synthesis of free phenolic compounds and a more discrete increase in the complexation of this newly synthetized phenolic compounds in insoluble structures.

The antioxidant capacity, measured by DPPH and FRAP methods, showed no significative correlation (0.6442, $p = 0.06$) and strong (0.8947, $p = 1.12 \times 10^{-3}$) correlation with TRC, respectively. These spectrophotometric results (TRC, DPPH and FRAP) indicate that, although the antioxidant capacity increase in both extracts after germination, in the FPC it occurs more expressively and it can have a positive impact on human health, due to the potential cellular anti-inflammatory and radical scavenging activity of the compounds phenolics. Furthermore, the hypothesis is that, with a higher FPC:BPC ratio and, consequently, less interaction between phenolic compounds and grain matrix constituents, they become more bioaccessible to the human body (Melini *et al.*, 2020).

3.2. Total proanthocyanidin and flavonoids content

Proanthocyanidins, also known as tannins, and flavonoids contents in the different genotypes and germination times are presented in Table 3. Among the UG samples, the IS 29569 genotype had the highest flavonoid content, surpassing Macia and IS 30400 by a factor of five and two, respectively. Furthermore, when comparing the proportion of flavonoids to TRC, IS 29569 genotype has the highest one (28%), followed by Macia (15%) and IS 30400 (9%), suggesting differences in their overall phenolic content, but also diversity. As expected, only the IS 29569 genotype presented detected levels of proanthocyanidins determined spectrophotometrically. The presence of condensed tannins is controlled by the B_1 and B_2 genes, which in their epistatic form ($B_1\_B_2\_$) result in grains with a pigmented testa layer (Earp *et al.*, 1983). Indeed, the tannin content in the UG IS 29569 genotype sample (5.36±1.18 mg EC/g) is within the range reported in the literature for pigmented testa sorghum grains (Rao *et al.*, 2018; Shen *et al.*, 2018).

It has been more recently reported in the literature that tannin content had a low correlation with pericarp and with testa color, which may be due to the relationship between tannin content and testa thickness. For tannin sorghum, varieties with lighter pigmented but thicker testa might have higher tannin content than those darker-pigmented, thinner testa (Zhang *et al.*, 2023). Corroborating our results, the levels of phenolics and antioxidant activity are highest when sorghums possess thick and colored (*e.g.* purple, dark red or black) pericarp, and a pigmented testa related to S gene (Dykes *et al.*, 2005).

Tannins (proanthocyanidins) are known as an antinutritional factor of sorghum grain, because, although it has high bioactivity, they can form insoluble complex with sorghum proteins (kafirins) particularly upon thermal processing and, therefore, can exert steric effects and prevent the access of enzymes during protein digestion (Duodu *et al.*, 2003). Technological processes can facilitate proanthocyanidin degradation, leading to the formation of flavonoid monomers and dimers (D'Almeida *et al.*, 2021b); however, in short-time germinated sorghum (24 and 72h), the reduction of proanthocyanidins has been attributed to novel hydrophobic interaction between tannins and kafirins (Saithalavi *et al.*, 2021). In contrast, our findings (Table 3) reveal that germination had no significant impact on tannin content in the IS 29569 genotype or on flavonoid content in all genotypes. In other words, the observed increase in phenolic content, as discussed in Section 3.1, can be not related with these two specific phenolic classes.

3.3. Overview of sorghum phenolic compound identifications by UHPLC-MS^{E}

The total number of features detected by the UHPLC-ESI-QTOF platform was 2,343 ions. To annotate the features which possibly represent phenolic compounds, we used a customized database built from PubChem and Phenol Explorer for chemical matches (section 2.9). Only 9% of detected ions had database matches, suggesting that many ions could be derived from uncharacterized or unknown metabolites. Furthermore, some compounds were filtered and excluded based on their presence in the blank, ionic relative abundance below 5,000 (considered our chromatogram baseline), isotopic similarity below 80%, and compounds with fragmentation data differing from the literature and/or databases. After assigning these filters, 66 phenolic compounds were putatively identified, among them 12% compounds were fully confirmed with commercial phenolic standards (Supplementary Table 6, compounds presented in bold).

The results presented in Figure 24A show the total chromatogram with detected peaks in the QC samples by UHPLC-ESI-QTOF platform, in which the 10 main and most abundant phenolic compounds are highlighted. Chromatographic peaks excluded after applying the filters mentioned above were not considered. Two of the most abundant compounds in our samples are from flavonoid class: apigenin and daidzin. Apigenin was identified at level 1 and was the most abundant compound (*m/z* 269.0436, RT 11.03) present in all samples (Supplementary Tables 6 and 7) and is reported as one of the main flavones in sorghum grains (Bradwell *et al.*, 2018). Although it is less stable than their counterparts, the 3-deoxyanthocyanidins (apigeninidin) due to the absence of the 3-OH group in the latter, sorghum apigenin extracts showed potential to be added to food systems such as fortified beverage with health-promoting properties (Bradwell *et al.*, 2018). The third most abundant compound among the sorghum samples is [M−H]- ion at *m/z* 415.1020 (RT 10.05) (Supplementary Table 7). This compound annotation was attributed to daidzin, due to the majority presence of *m/z* 253.0697 fragment ($C_6H_{10}O_5$), *i.e.*, the 7-O-glucoside bond loss and daidzein (aglycated form) formation (Zhao *et al.*, 2018). Furthermore, *m/z* 135.0431 fragment (C_8H_6O loss, 118 Da) was found in smaller quantities and it is a typical product of Retro-Diels-Alder rearrangement from the 1,4-position in daidzein C-ring (Zhao *et al.*, 2018). Daidzin/daidzein is an isoflavone with health benefits and versatile pharmacological properties, including its activities against cancer, cardiovascular disease, osteoporosis, diabetes, aging, oxidation, and inflammation

(Zhao *et al.*, 2018); and, although it is typically found in soybeans, it has previously been reported in sorghum grains (D'Almeida *et al.*, 2021b).

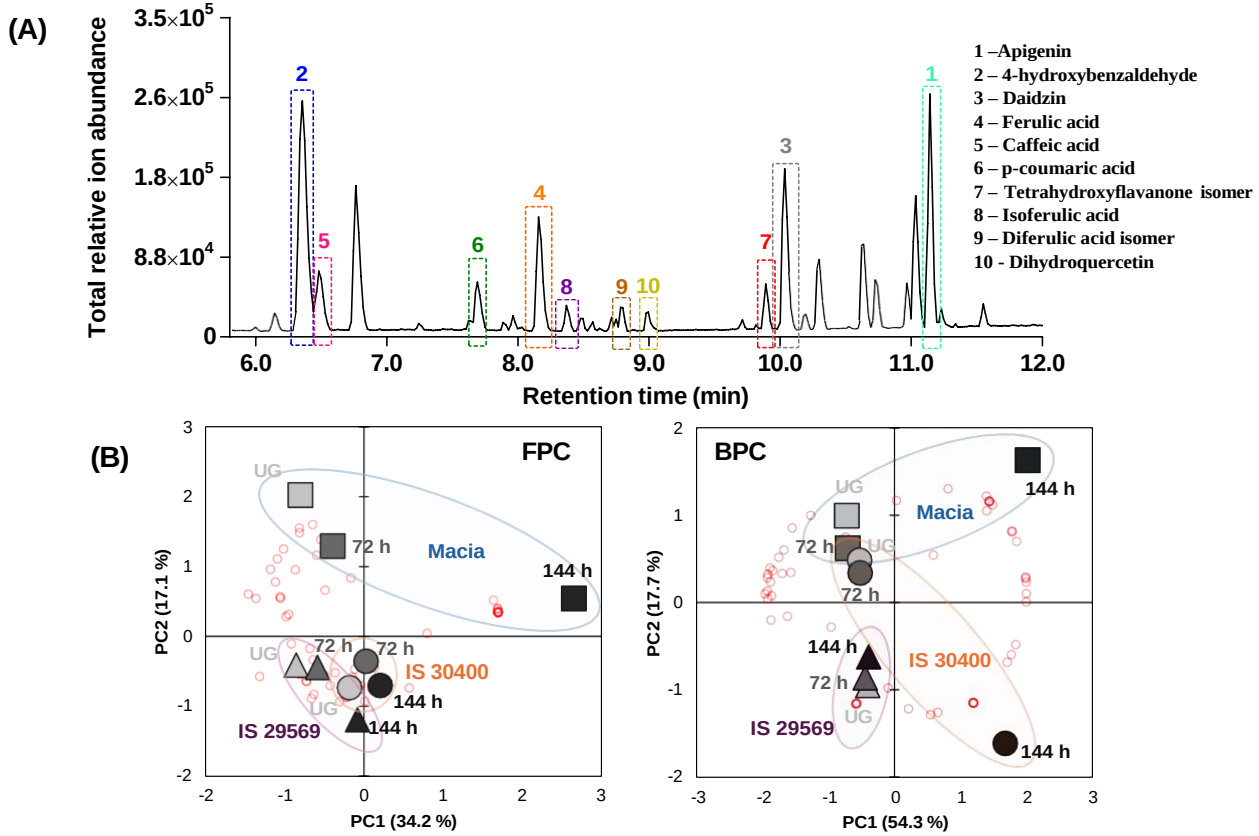


Figure 16. Overview of metabolomics data of sorghum grain samples: Representative chromatogram of QC samples analyzed by LC-QTOF-MS/MS, highlighting the ten most abundant peaks in sample (A); and Principal component analysis (PCA) biplot in free (FPC, left PCA) and bound (BPC, right PCA) extracts (B). The samples (symbols) are distributed according to relative intensity of identified phenolic compounds (red circles). UG: Ungerminated sorghum.

Among the other most abundant phenolic compounds, five were confirmed with commercial phenolic standards: 4-hydroxybenzaldehyde (RT 6.35), isoferulic acid (8.16), caffeic acid (6.48) and p-coumaric acid (7.69), occupying between the second and sixth positions, respectively. Isoferulic acid isomer (ferulic acid, RT 8.37) and its dimer (diferulic acid, 8.78), as well as two compounds from the flavonoid class (tetrahydroxyflavanone isomer and dihydroquercetin, retention time 9.82 and 9.05, respectively) were identified based on literature/fragmentation data and are among the seventh and tenth most abundant compounds in sorghum samples.

Hydroxycinnamic acids (peaks 3-4 and 6-8) are expected to be among the most important compounds in sorghum (Awika & Rooney, 2004; D'Almeida *et al.*, 2021b). In our sorghum samples, they represent 82% of the total abundance of phenolic acids class and are mainly bound (65%) to cell wall components. Conversely, among the most abundant flavonoids class (peaks 1, 5 and 9-10), 87% are free in plant cell vacuoles (free extract). It is important to highlight that this different phenolic profile in each extract, as well as the genotype used, will result in different behaviors during grain germination (Baranzelli *et al.*, 2023).

To explore the variation in the data set, the PCA biplot (scores-samples; loadings-phenolic compounds) was applied to investigate the degree of similarity or dissimilarity between the compounds relatively quantified , from the ion abundance in each sample (Figure 24). The left-PCA biplot indicated a clear distinction between Macia and others two genotypes in free extract, as Macia is the only genotype with good separation between germination times, mainly after 144h (Figure 24B). For bound phenolic compounds (right-PCA) the separation was greater (PC1 and PC2: 72%; Figure 24C), due to the great phenolic variability after 144 h of seedling growth in Macia and IS 30400 genotypes (PC1 = 54.3%). IS 29569 genotype did not have their bound compounds impacted by germination. To better understand these variations, each genotype and germination time were individually evaluated.

3.4. Genotype and germination time impacted phenolic compounds variability

Polyphenols, plant-specialized (secondary) metabolites perform multiple critical biological roles, but generally are considered involved in the physical and chemical plant

protection from pathogens, so genotypes and/or environmental growing conditions are the main factor influencing the phenolic profile of raw grains (Ceccaroni *et al.*, 2020). On the other hand, cereal processing can increase phenolic variability, due to the various parameters involved, *e.g.*, time-temperature binomial, pH, enzymatic action. Understanding the changes triggered in each of these variables is crucial to produce sorghum-based foods with preserved/increased levels of bioactive compounds.

Among the 66 phenolic compounds putatively identified for all genotypes, only 11 and 5 were exclusively found in the free and bound forms, respectively. Most of phenolics compounds were commonly identified in both extracts (50). The distribution of these compounds in each genotype (Figure 25A) showed the highest number of identifications in IS 29569 genotype (63), followed by IS 30400 (51) and Macia (46). The predominance of IS 29569 can be explained by the presence of exclusive compounds (10), among them, 90% were attributed to flavonoids class based on typical $[M-H]^-$ subunits loss (*e.g.*, cleavage of the C-C bonds) (Abrankó *et al.*, 2011). In flavonoid pathway, the basic structures are modified producing other phenolic compounds, *e.g.*, condensed tannins. It was confirmed according to section 3.3, in which tannins and flavonoids showed a strong positive correlation ($R = 0.96$, $p = 4.24 \times 10^{-5}$). Even with these discriminating compounds (Figure 25A), most of phenolics (58%) are not influenced by sorghum genotype. The variability was even lower after seed germination, *i.e.*, 83% of the phenolic compounds are present in the three different germination times that were assayed (Figure 25A). Only four compounds were synthesized at seedling growth (144 h), identified as flavonoids with hydroxy and methoxy groups ($C_{16}H_{12}O_4$, $C_{16}H_{14}O_5$ and its isomer, and $C_{22}H_{22}O_{10}$).

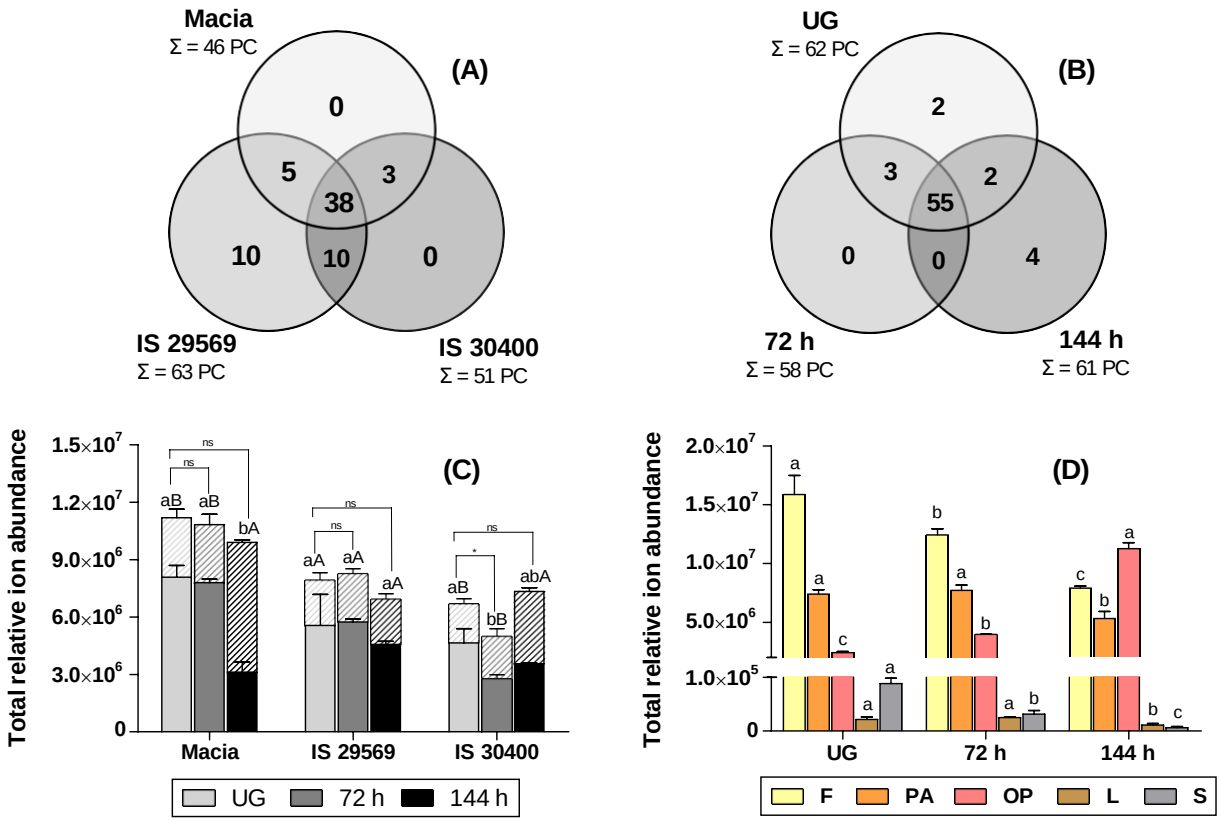


Figure 17. Impact of genetic variability and seed germination on phenolic content and profile: Venn diagrams with the number of identifications in each genotype (A) and germination time (B); total relative ion abundance of phenolic compounds (C); and distribution of phenolic classes in sample (D). UG: ungerminated sorghum; F: flavonoids; PA: phenolic acids; OP: other polyphenols; L: lignans; S: stilbenes. Different lowercase and uppercase letters in graph C mean a significant difference ($p < 0.05$) between free and bound extract samples, respectively; "ns" = no statistical difference and "*" = $p<0.05$. Different lowercase letters in graph D mean the significant difference ($p < 0.05$) of the same phenolic class throughout germination, following a one-way ANOVA test. Bars represent standard deviation (n = 3).

In addition to the number of phenolic compounds identified, the total relative ion abundance was also taken into account (Figure 25C). Globally, FPC represented 63% and was more abundant than BPC. After germination, the abundance of FPC decreased in Macia and IS 30400 genotypes, while BPC increased. In IS 29569 genotype, FPB and BPC appear not be impacted by the germination times applied in our experimental conditions. Some studies suggest that sprout is responsible for the increase in FPC and decrease in BPC, due to the production of enzymes able to hydrolyze the ester and glycosidic bonds between BPC and plant cell wall (Rasera *et al.*, 2020). Considering all phenolic classes together, especially in Macia genotype, the opposite behavior in this study can be attributed to some factors as reviewed by (Xu *et al.*, 2020)). 1) the increase in BPC may be due to enzymes activation involved in biosynthetic phenylpropanoids pathway and endogenous esterases; and/or the intense cell division phase favored by germination, which induces phenolic compounds to be transferred to the cell wall. 2) the decrease in FPC may be associated with its consumption in neutralizing reactive oxygen species (ROS) generated during germination, however, the most viable hypothesis is that 3) FPC can be transported to the cell wall within vesicles and form new BPC (Xu *et al.*, 2020). The last explanation would raise from the experimental approach used to conduct germination. In the studies of Rasera *et al.* (2020) they performed the germination at 25 °C during 48h-72h, it means 672 et 1008 °C.h considering the temperature base of 11 °C for sorghum germination. Our germination thermal time corresponded respectively to 350 and 720 °C.h, shorter than mentioned work.

Although total phenolic compounds, considering the sum of FPC and BPC, remains constant throughout germination (Figure 25C), the abundance of each phenolic class were variable at the different germination stages (Figure 25D). Besides, to be the most abundant, flavonoids appeared as the most susceptible class, showing a progressive reduction during grain germination stages ($p > 0.05$). Stilbenes, a polyphenolic allelochemical class, followed the same pattern; while phenolic acids and lignans remain almost constant for all genotypes, with a slight variation after 144h of seedling growth ($p < 0.05$). The ratio between flavonoids and the second most abundant class (phenolic acids) varied from 2.14 (UG) to 1.61 (72 h) and 1.49 (144 h), regardless the genotype. Other polyphenols showed an accumulation profile during the germination, contributing to mitigate the reduction presented by flavonoids in the total abundance of compounds.

The main compound of the other polyphenols which showed a significantly increase in all genotypes during germination steps is 4-hydroxybenzaldehyde.

To better understand which phenolic compounds impact this differentiation between germination times, volcano plots showing the differentially abundant compounds for each genotype were applied (Figure 26). There were 25 differentially abundant compounds (Supplementary Table 7), of which 17 are flavonoids, 7 are phenolic acids and 1 other polyphenols. The compound [M−H]- ion at m/z 121.0276, identified through commercial phenolic standards as 4-hydroxybenzaldehyde, was unanimously more abundant among the genotypes, mainly after 144 h of seedling growth. Conversely, 5 of the most abundant compounds in sorghum (Figure 24A) have decreased abundance in all genotypes after seed germination: daidzin, apigenin, caffeic acid, isoferulic acid and p-coumaric acid. One of the explanations is that germination induced the action of feruloyl/caffeoyl-CoA synthetase enzymes, which convert, respectively, ferulic and caffeic acid into 4-hydroxybenzaldehyde. This increase in enzymatic action favors a decrease in hydroxycinnamic acids levels (ferulic, caffeic and p-coumaric acids) (Supplementary Figure 6).

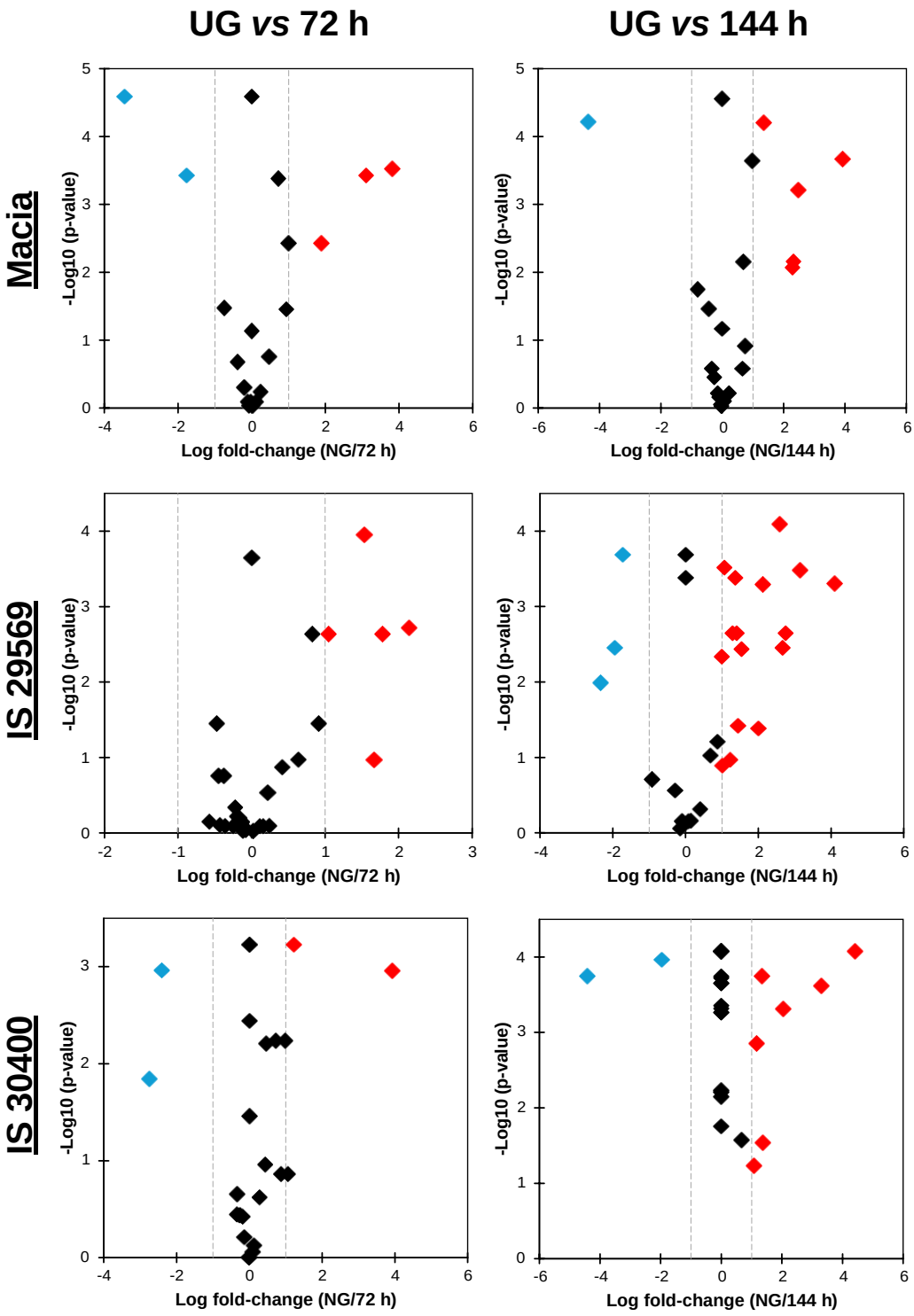


Figure 18. Volcano plots of differential abundance of metabolites between ungerminated sorghum (UG) and the different germination times (72 and 144 h) in each genotype. Red and blue dots indicate significantly increased and decreased metabolites, respectively (log2 fold change ± 1.0, $p < 0.05$). Black dots represent the metabolites showing no significant difference.

Finally, four [M−H]- ion at *m/z* 577.1332 were detected in all genotypes. The experimental fragmentation of these isomers is shown in Figure 27, in which the produced [M−H]- ions at *m/z* 289.0537 (purple rectangle) and 125.0223 (green rectangle) were the fragments with the highest intensity. Based on the literature, these are typical products of B-type procyanidin dimers fragmentation, also known as condensed tannins (Holt *et al.*, 2002). Procyanidins in sorghum grains can prevent the harmful effects caused by numerous enzymes, ROS and free radicals involved in the pathogenesis of oxidative origin, such as cardiovascular diseases and some forms of cancer (Rue *et al.*, 2018). The main fragmentation pathways of procyanidins include quinone methide fission (QMF) of the interflavanoid bond, forming two [M−H]- ions at *m/z* 289.0537 ($C_{15}H_{13}O_6$, (+)-catechin) (Holt *et al.*, 2002). Another characteristic fragment of B-type procyanidin dimer would be the compound *m/z* 451, formed by heterocyclic ring fission (HRF) (Holt *et al.*, 2002; Rue *et al.*, 2018). Although this was not detected in our samples, its product [M−H]- ions at *m/z* 125.0223 was present as the most intense fragment.

Interestingly, procyanidins were detected in all genotypes, even those classified as "tannin-free" (Figure 27) such as Macia cultivar. According to the classification by Dykes & Rooney (2006), sorghum is divided into three levels of tannin content: group I - low level tannin (0 – 1.8 mg); group II - medium level tannin (6.4 – 15.5 mg); and group III - high level tannin (15.5 – 56.3 mg catechin equivalents g^{-1}, dry mass), based on the vanillin/HCl method. Therefore, even if group I contains up to 1.8 mg of catechin equivalents, it can be considered “tannin-free” (Palacios *et al.*, 2021). In the present study this behavior was observed in Macia and IS 30400 genotypes (Figure 27). When the vanillin method was applied, both showed null catechin equivalents values (Table 3), and they are classified as tannin-free genotypes. However, when applying more sensitive techniques, such as metabolomics tools, some procyanidin dimers were detected, even in low abundance.

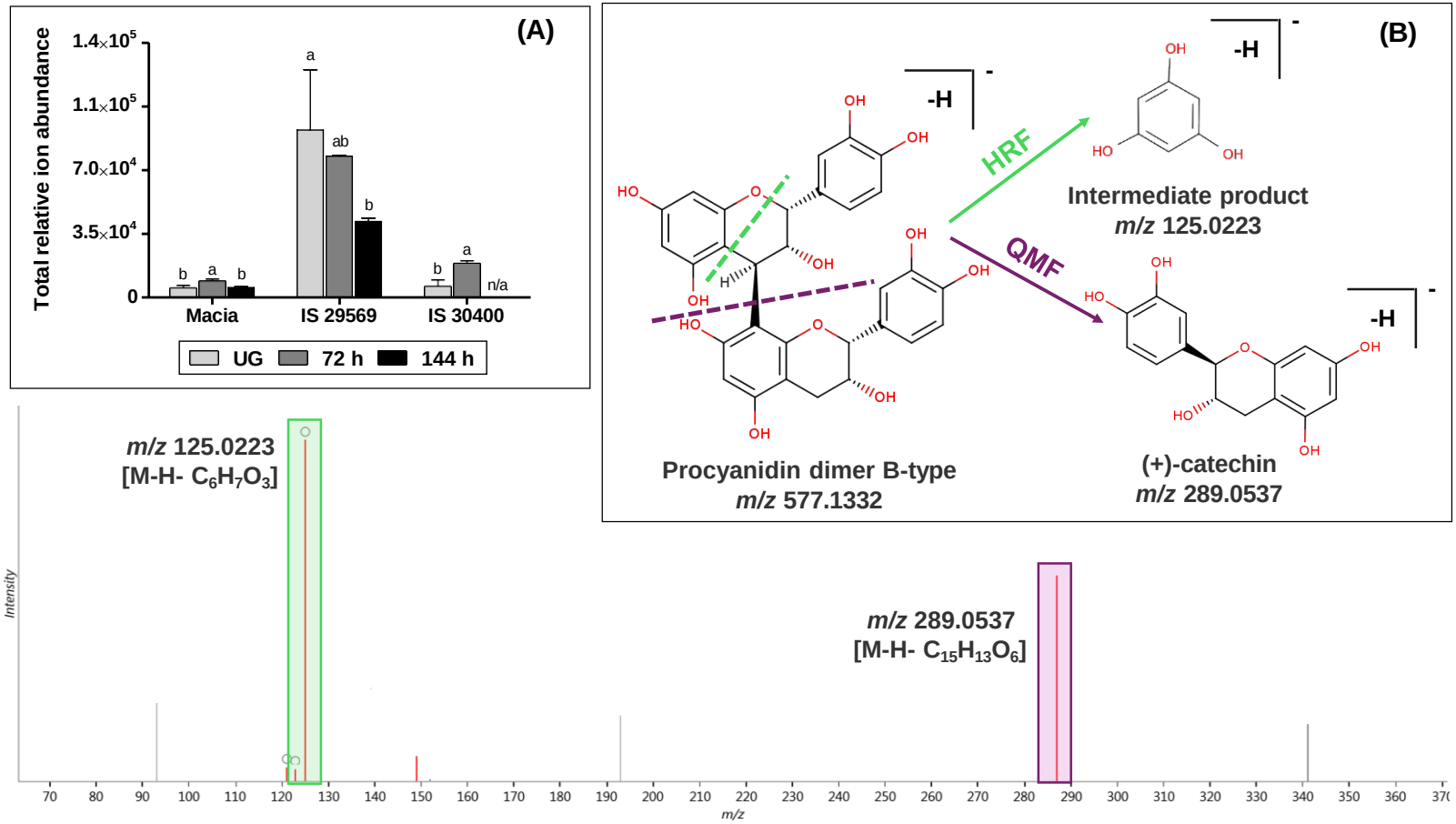


Figure 19. Fragmentation ions MS/MS in negative mode of procyanidin dimer B-type detected in sorghum extracts. Insert graph (A) represents the total relative ion abundance of these compound in each genotype and germination time; and (B) their fragmentation pathway showing the products formed by quinone methide fission (QMF) and heterocyclic ring fission (HRF). UG: ungerminated sorghum.

3.5. Germination increased overall phenolic compounds bioaccessibility during *in vitro* digestion

Bioaccessibility is the phenolic compound fraction that is released from food in the gastrointestinal tract, thus becoming available for intestinal absorption. Sorghum grain is a complex food matrix showing an outstanding composition of bioactive phenolic compounds; however, to exert their desired physiological effects on consumers, polyphenols need to be bioaccessible. Coupling untargeted omics approaches in modeling gastrointestinal digestion is still a challenge. In the case of our samples, the total number of features detected by UHPLC-ESI-QTOF platform was 5,379 ions, since enzyme residues and other food matrix components can be inadvertently detected. Therefore, we focused only in the bioacessibility of the most relevant phenolic compounds during sorghum germination after the simulated digestion (Figure 28).

The selection of compounds to be evaluated after *in vitro* digestion was based on the relative ion abundance shown in Figure 24A. Among the ten most abundant phenolic compounds in sorghum, only six were found in the digested extracts after applying the metabolomics filters described above. Procyanidin dimers were also not detected after digestion. In UG sorghum grains, compounds from the flavonoid class showed the highest total bioaccessibility, but each one with the maximum peak in a specific phase, *i.e.*, daidzin (UG total abundance: 2.21×10^6) and apigenin (UG total abundance: 2.34×10^6) have greater bioaccessibility in the gastric and intestinal phases, respectively (Figure 28).

Compounds belonging to hydroxybenzaldehyde and hydroxycinnamic acids groups had the greatest bioaccessibility in the oral phase (Figure 28). This result is unexpected, since the oral phase is reported to have little influence on the phenolic bioaccessibility due to the short interaction of the food bolus with oral enzymes (Wojtunik-Kulesza *et al.*, 2020). However, saliva can play an important role in the solubilization of these polyphenols present in sorghum samples, substantially increasing their availability (Ginsburg *et al.*, 2012). Among these compounds, 4-hydroxybenzaldehyde showed the highest bioaccessibility, followed by p-coumaric, isoferulic and caffeic acids.

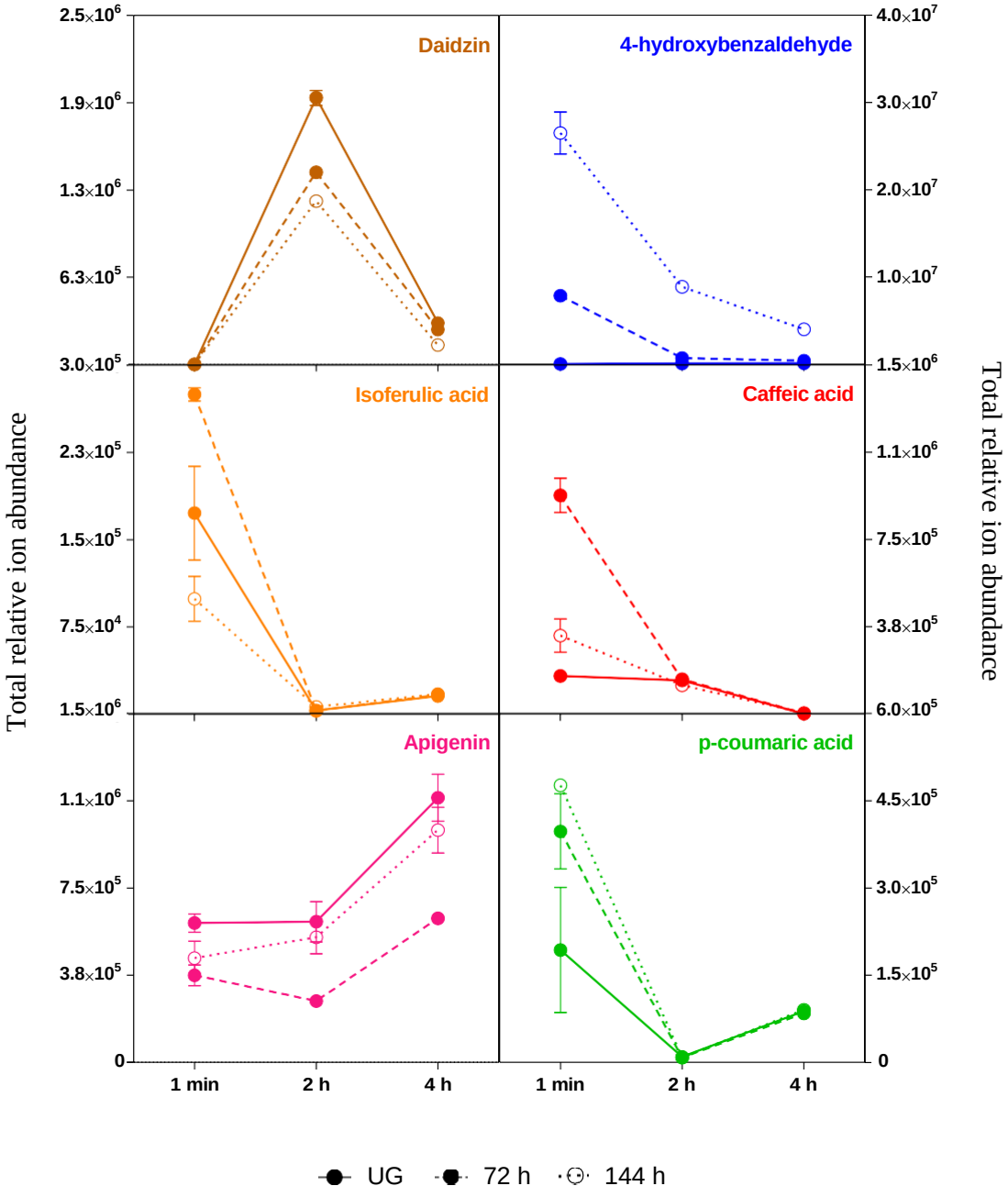


Figure 20. Total relative ion abundance of selected phenolic compounds released during the oral, gastric and intestinal phases of simulated *in vitro* digestion of ungerminated (UG) and germinated (72 and 144 h) sorghum samples.

In the gastric phase, phenolic compounds are usually stable at acidic pH and do not undergo degradation reactions (Ed Nignpense *et al.*, 2021); while the long time that samples are submitted to pH 7.0 (two hours) in intestinal phase is enough to be the most relevant step of the *in vitro* digestion in terms of flavonoids release (Wojtunik-Kulesza *et al.*, 2020). Although, this behavior can change for glycosylated flavonoids (daidzin). In the small intestine with high pH (~7), deglycosylation of flavonoids occurs, so the hypothesis is that the reduction of daidzin abundance in the intestinal phase was due to its deglycosylation and, consequently, its conversion into the aglycated form (daidzein) as suggested in (Wojtunik-Kulesza *et al.*, 2020).

Lastly, the impact of germination and seedling growth was evaluated in each compound, comparing UG, 72 and 144h (Figure 28). Globally, germination did not modify the digestion behavior in each phase of simulated gastrointestinal tract, the only variation was between values. Flavonoids (apigenin and daidzin) showed a reduction in total bioaccessibility after 72 h of germination (-46% and -26%) and 144 h (-15% and -41%, respectively) when compared to UG grains. Conversely, phenolic acids had their bioaccessibility increased after germination. Among them, p-coumaric acid and 4-hydroxybenzaldehyde showed an increase in total bioaccessibility both after 72 h (+68% and 31-times higher) and 144 h (+98% and 136-times higher, respectively), mainly in the oral phase. Germination saccharolytic enzymes hydrolyse the cell walls to release these phenolic compounds (previously bound) in free forms, consequently increasing its bioaccessibility (strong positive correlation between FPC and bioaccessibility of these compounds; $r = 0.92$; $p = 0.01$). Isoferulic and caffeic acids bioaccessibility was also favored by germination in the oral phase, but, for them, 72-hour bioprocessing time seems to be ideal (+54% and +253%, respectively, compared to UG).

4. Conclusion

The impact of sorghum grain germination on the phenolic profile and its bioaccessibility was revealed, for the first time, by untargeted metabolomics, considering contrasting tannin levels and testa pigmentation genotypes. Metabolomics approach revealed the (low) presence of procyanidin dimers even in genotypes classified as “tannin-free”, but the most of phenolic compounds were shown to be part of a nucleus of grain-specific compounds, regardless of these factors. Germination progressively affects the free/bound phenolic, indicating the release of those bound to the cell wall. Flavonoids

were the most impacted phenolic class during germination and seedling growth. Germination was an effective strategy to improve bioaccessibility of phenolic acids (*e.g.*, isoferulic, caffeic and p-coumaric acids) attaining the maximum in the oral phase of the TGI, while flavonoids bioaccessibility was reduced towards seedling growth. However, further research is needed to explore the optimal time-temperature binomial for phenolic compounds valorization, and to develop biostrategies to protect these compounds particularly susceptible to food processing and human digestion.

5. Acknowledgements and funding

The authors are grateful to "Genetic Improvement and Adaptation of Mediterranean and Tropical Plants" - AGAP Institute Joint Research Unit for providing and collecting samples. This work was financed by UNIRIO, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (401053/2019-9; 407350/2023-3); Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (26/010.100988/2018; E-26/210.716/2021; E-26/010.001294/2019; E-26/201.317/2022; E-26/210.152/2023); and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (code 001 and PDSE).

6. Supplementary material

Supplementary Table 5. Total putative identification of phenolic compounds in the sorghum extracts by UHPLC-MS^{E}.

Putative identification	ID level*	Molecular Formula	*m/z*	RT (min)	Score	FS (%)	Error (ppm)	IS (%)	Class	Subclass
Equol	2	C15H14O3	241.0893	0.49	34.8	0.0	9.31	84.35	F	Isoflavonoids
Unknown compound 1	3	C16H14O5	285.0793	0.49	36.8	0.0	8.60	93.77	F	Unknown subclass
3'-O-Methylviolanone	2	C18H18O6	329.1061	0.55	38.9	15.6	9.22	89.23	F	Isoflavonoids
Hesperidin/neohesperidin	2	C28H34O15	609.1867	0.55	48.4	61.9	6.87	88.00	F	Flavanones
Equol 4'-O-glucuronide/equol 7-O-glucuronide	2	C21H22O9	417.1224	0.58	38.1	11.6	7.86	87.66	F	Isoflavonoids
Glycitin	2	C22H22O10	445.1170	2.46	40.3	19.8	6.78	89.19	F	Isoflavonoids
4-Hydroxybenzoic acid 4-O-glucoside	2	C13H16O8	299.0746	2.58	51.8	76.9	-8.63	91.46	PA	Hydroxybenzoic acids
Caffeic acid glucoside	2	C15H18O9	341.0854	5.15	56.0	94.8	-7.08	93.14	PA	Hydroxycinnamic acids
4-Hydroxybenzoic acid	**1**	**C7H6O3**	**137.0225**	**5.30**	**54.8**	**86.2**	**–9.04**	**97.96**	**PA**	**Hydroxybenzoic acids**
Caffeoylquinic acid isomer	2	C16H18O9	353.0860	6.07	53.5	74.6	-5.10	98.63	PA	Hydroxycinnamic acids
4-Hydroxybenzaldehyde	**1**	**C7H6O2**	**121.0276**	**6.35**	**37.9**	**0.0**	**-9.12**	**99.32**	**OP**	**Hydroxybenzaldehydes**
Caffeic acid	**1**	**C9H8O4**	**179.0331**	**6.48**	**50.7**	**61.9**	**–6.61**	**99.13**	**PA**	**Hydroxycinnamic acids**
Ferulic acid glucoside	2	C16H20O9	355.1006	6.60	49.8	68.5	-8.15	89.45	PA	Hydroxycinnamic acids
Puerarin	2	C21H20O9	415.1014	7.08	54.9	90.9	-5.01	89.49	F	Isoflavonoids
Dihydroferulic acid 4-O-glucuronide	2	C16H20O10	371.0963	7.14	55.6	90.9	-5.65	93.52	PA	Hydroxyphenylpropanoic acids
Unknown compound 2	3	C21H22O11	449.1067	7.23	50.8	65.7	-4.86	94.05	F	Unknown subclass
2,4-Dihydroxybenzoic acid	2	**C7H6O4**	153.0175	7.32	50.2	37.9	-7.62	99.16	PA	Hydroxybenzoic acids
Apigenin glucoside	2	C26H28O14	563.1388	7.55	57.4	96.0	-3.17	94.55	F	Flavones
p-Coumaric acid	**1**	**C9H8O3**	**163.0381**	**7.69**	**57.1**	**95.6**	**–7.60**	**98.45**	**PA**	**Hydroxycinnamic acids**

Salvianolic acid D	2	C11H10O6	237.0381	7.89	37.3	0.0	-9.78	97.00	OP	Other polyphenols
Naringin glucoside 1	2	C21H22O10	433.1123	7.96	48.6	59.3	-4.06	88.37	F	Flavanones
Unknown compound 3	3	C15H10O5	269.0433	8.01	39.6	10.7	-8.19	96.42	F	Unknown subclass
Unknown compound 4	3	C21H20O10	431.0961	8.14	42.1	25.6	-5.29	90.83	F	Unknown subclass
Isoferulic acid	**1**	**C10H10O4**	**193.0486**	**8.16**	**47.8**	**61.4**	**-5.41**	**99.04**	**PA**	**Hydroxycinnamic acids**
Unknown compound 5	3	C27H30O15	593.1496	8.17	45.1	35.9	-2.68	93.02	F	Unknown subclass
Quercetin 3-O-rutinoside	**1**	**C27H30O16**	**609.1446**	**8.17**	**51.0**	**62.6**	**-2.43**	**95.11**	**F**	**Flavonols**
Naringin glucoside 2	2	C21H22O10	433.1116	8.19	44.2	34.7	-5.52	92.77	F	Flavanones
Ferulic acid	2	C10H10O4	193.0485	8.37	48.9	52.3	−5.44	98.46	PA	Hydroxycinnamic acids
Genistin	2	C19H18O6	341.1010	8.39	42.2	20.6	-6.04	97.11	F	Isoflavonoids
Dehydrodiferulic acid isomer 1	2	C20H18O8	385.0908	8.39	45.2	38.1	-5.33	94.13	PA	Hydroxycinnamic acids
Naringenin 4'-O-glucuronide	2	C21H20O11	447.0907	8.39	43.7	33.9	-5.74	91.38	F	Flavanones
Procyanidin dimer B-type 1	2	C27H30O14	577.1346	8.39	37.3	5.1	-2.95	84.92	F	Flavones
Methylgalangin	2	C16H12O5	283.0590	8.57	55.6	90.3	-7.87	96.62	F	Flavonols
Procyanidin dimer B-type 2	2	C27H30O14	577.1345	8.69	37.8	3.3	-3.03	89.52	F	Flavones
Dehydrodiferulic acid isomer 2	2	C20H18O8	385.0909	8.78	44.1	30.6	-5.12	95.89	PA	Hydroxycinnamic acids
Resveratrol glucuronide isomer 1	2	C20H20O9	403.1008	8.80	48.2	57.5	-6.57	91.06	S	Stilbenes
Unknown compound 6	3	C21H20O10	431.0964	8.96	56.0	94.4	-4.60	90.94	F	Isoflavonoids
Resveratrol glucuronide isomer 2	2	C20H20O9	403.1004	9.04	39.9	14.6	-7.55	93.18	S	Stilbenes
Dihydroquercetin	2	C15H12O7	303.0487	9.05	51.4	68.1	-7.65	97.72	F	Dihydroflavonols
Formononetin/dalbergin	2	C16H12O4	267.0640	9.07	45.4	44.1	-8.65	92.55	F	Isoflavonoids
Tectoridin	2	C22H22O11	461.1068	9.09	42.8	30.4	-4.69	88.83	F	Isoflavonoids
Procyanidin dimer B-type 3	2	C27H30O14	577.1349	9.11	54.0	83.4	-2.46	89.39	F	Flavones
Unknown compound 7	3	C16H14O6	301.0694	9.36	47.6	50.6	-7.84	96.41	F	Flavanones
Dehydrodiferulic acid isomer 3	2	C20H18O8	385.0906	9.43	36.4	0.0	-5.84	88.62	PA	Hydroxycinnamic acids
Methylepicatechin	2	C16H16O6	303.0850	9.52	55.0	85.9	-7.83	98.07	F	Flavanols
Butin	2	C15H12O5	271.0590	9.54	49.6	66.9	-8.08	89.89	F	Isoflavonoids

Naringenin glucoside	2	C21H22O10	433.1119	9.54	49.2	53.0	-4.78	98.32	F	Flavanones
Dehydrodiferulic acid isomer 4	2	C20H18O8	385.0907	9.71	41.3	14.9	-5.60	98.02	PA	Hydroxycinnamic acids
Procyanidin dimer B-type 4	2	C30H26O12	577.1332	9.80	46.8	52.7	-3.45	85.47	F	Flavanols
Tetrahydroxyflavanone isomer	2	C15H12O6	287.0539	9.82	50.3	61.5	-7.66	98.75	F	Flavanones
Butein	2	C15H12O5	271.0587	9.84	51.3	75.4	-9.10	90.92	F	Chalcones
Unknown compound 8	3	C16H14O5	285.0744	10.00	54.2	87.5	-8.55	93.14	F	Unknown subclass
Unknown compound 9	3	C18H16O7	343.0799	10.04	43.5	30.9	-7.12	94.51	F	Flavones
Daidzin	2	C21H20O9	415.1020	10.05	57.1	90.7	-3.51	99.15	F	Isoflavonoids
Dehydrodiferulic acid isomer 5	2	C20H18O8	385.0905	10.10	42.7	28.9	-6.07	91.42	PA	Hydroxycinnamic acids
Unknown compound 10	3	C15H12O6	287.0539	10.14	45.9	42.3	-7.75	95.70	F	Flavonoids
Unknown compound 11	3	C16H14O6	301.0692	10.34	47.4	51.2	-8.36	94.88	F	Unknown subclass
Isorhamnetin	2	C16H12O7	315.0486	10.36	56.1	95.3	-7.61	93.50	F	Flavonols
Unknown compound 12	3	C16H16O5	287.0901	10.36	41.6	25.8	-8.32	91.67	F	Isoflavonoids
Unknown compound 13	3	C18H16O7	343.0799	10.66	43.2	32.4	-7.11	91.65	F	Flavones
Quercetin	**1**	**C16H14O6**	**301.0693**	**10.78**	**55.8**	**92.6**	**-8.03**	**95.20**	**F**	**Flavonols**
1-Acetoxypinoresinol	2	C22H24O8	415.1364	10.78	36.1	0.0	-8.19	89.66	L	Lignans
Apigenin	**1**	**C15H10O5**	**269.0436**	**11.03**	**50.0**	**60.5**	**-7.35**	**97.66**	**F**	**Flavones**
Gardenin B	2	C19H18O7	357.0956	11.14	37.3	0.2	-6.71	93.71	F	Flavones
Scutellarein	2	C15H10O6	285.0386	11.18	47.6	46.0	-6.56	99.43	F	Flavones
Unknown compound 14	3	C16H12O6	299.0541	11.18	41.1	15.4	-6.59	97.39	F	Unknown subclass

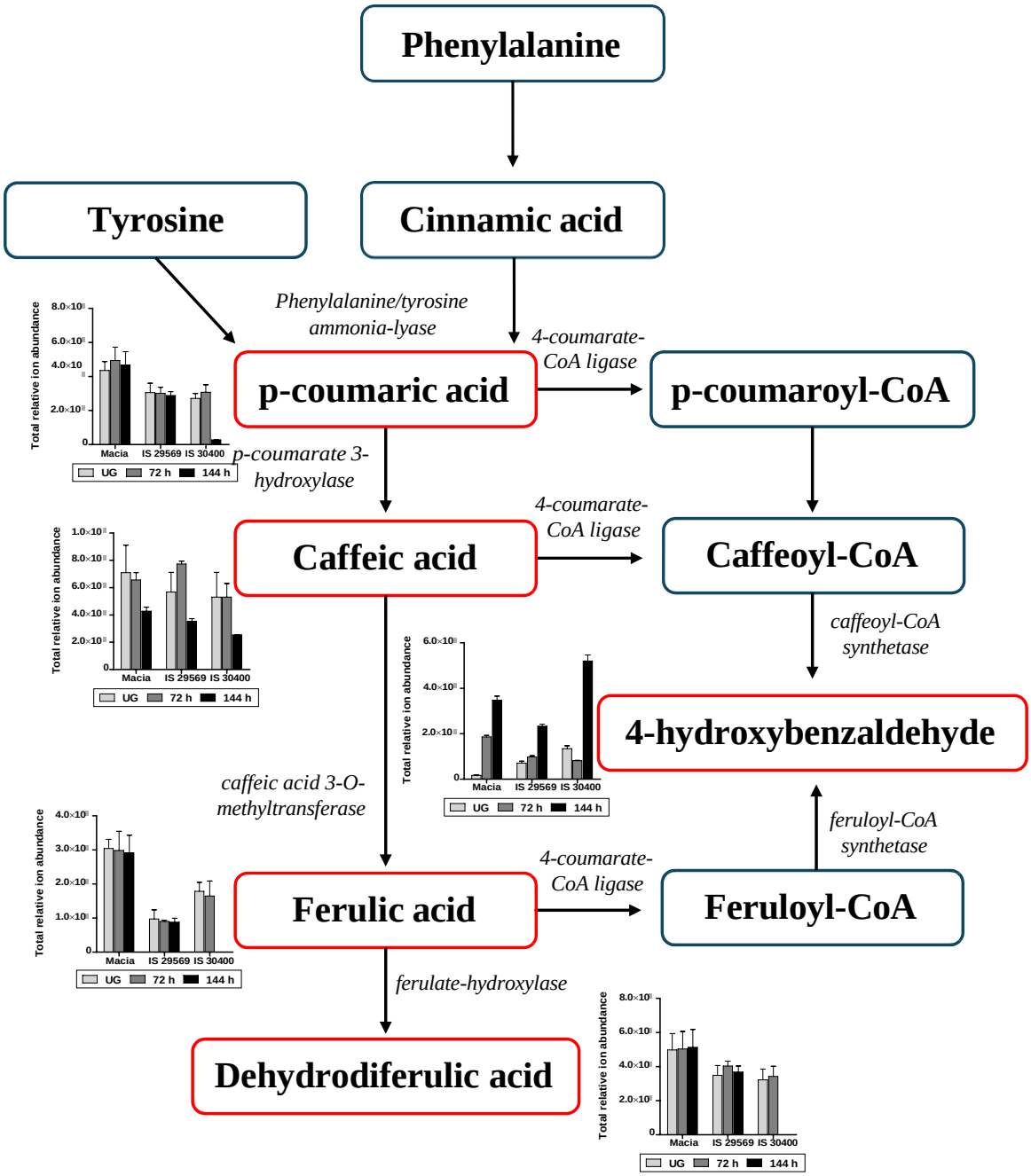
* Identification level according to Sumner et al. (2007). *m/z* = mass/charge; RT = retention time; FS = fragmentation score; IS = isotope similarity; PA = phenolic acids; F = flavonoids; OP = other polyphenols; L = lignans; and S = stilbenes.

Supplementary Table 6. Volcano plot data obtained for the phenolic compounds differentially abundant in the samples along sorghum germination.

Putative identification	*m/z*	RT (min)	Class	IS29569 abundance			IS30400 abundance			Macia abundance		
				UG	72 h	144 h	UG	72 h	144 h	UG	72 h	144 h
Hesperidin/neohesperidin	609.1867	0.55	F	$5.00x10^4$	$4.92x10^4$	$2.39x10^4$	$4.75x10^4$	$3.41x10^4$	$3.91x10^4$	$3.41x10^4$	$1.19x10^4$	$9.00x10^3$
Caffeic acid glucoside	341.0854	5.15	PA	$7.33x10^3$	$9.53x10^3$		$1.04x10^4$	$5.53x10^4$	$1.94x10^4$	$3.37x10^4$	$1.14x10^5$	$2.12x10^4$
4-Hydroxybenzoic acid	**137.0225**	**5.30**	**PA**	**$1.71x10^5$**	**$9.66x10^4$**	**$6.63x10^4$**	**$1.79x10^5$**	**$8.99x10^4$**	**$7.96x10^4$**	**$4.70x10^4$**	**$4.69x10^4$**	**$7.06x10^4$**
Caffeoylquinic acid isomer	353.0860	6.07	PA	$2.56x10^4$	$5.82x10^3$	$1.39x10^4$	$6.80x10^3$	$5.39x10^3$	$2.54x10^4$	$2.94x10^4$		$5.02x10^3$
4-Hydroxybenzaldehyde	**121.0276**	**6.35**	**OP**	**$7.01x10^5$**	**$9.73x10^5$**	**$2.33x10^6$**	**$1.34x10^6$**	**$8.07x10^5$**	**$5.20x10^6$**	**$1.68x10^5$**	**$1.85x10^6$**	**$3.48x10^6$**
Caffeic acid	**179.0331**	**6.48**	**PA**	**$5.67x10^5$**	**$7.73x10^5$**	**$3.52x10^5$**	**$5.29x10^5$**	**$5.31x10^5$**	**$2.53x10^5$**	**$7.11x10^5$**	**$6.55x10^5$**	**$4.27x10^5$**
Dihydroferulic acid 4-O-glucuronide	371.0963	7.14	PA	$6.22x10^4$	$2.14x10^4$	$1.03x10^4$	$6.11x10^4$	$2.62x10^4$	$2.39x10^4$	$8.69x10^4$	$5.29x10^4$	$3.39x10^4$
Apigenin glucoside	563.1388	7.55	F	$6.86x10^4$	$6.29x10^4$	$2.58x10^4$						
p-Coumaric acid	**163.0381**	**7.69**	**PA**	**$3.05x10^5$**	**$3.01x10^5$**	**$2.88x10^5$**	**$2.72x10^5$**	**$3.08x10^5$**	**$2.76x10^4$**	**$4.36x10^5$**	**$4.94x10^5$**	**$4.67x10^5$**
Naringin glucoside 1	433.1123	7.96	F	$2.12x10^4$	$6.67x10^3$	$5.27x10^3$	$5.82x10^3$	$7.49x10^3$		$1.29x10^4$	$1.10x10^4$	$1.33x10^4$
Unknown compound 4	431.0961	8.14	F	$3.21x10^4$	$2.46x10^4$	$1.10x10^4$						
Isoferulic acid	**193.0486**	**8.16**	**PA**	**$4.53x10^5$**	**$5.21x10^5$**	**$5.50x10^5$**	**$5.06x10^5$**	**$6.34x10^5$**	**$2.38x10^4$**	**$1.07x10^6$**	**$1.14x10^6$**	**$1.18x10^6$**
Unknown compound 5	593.1496	8.17	F	$4.04x10^4$	$4.07x10^4$	$1.49x10^4$				$1.04x10^5$		
Naringin glucoside 2	433.1116	8.19	F	$1.05x10^5$	$3.04x10^4$	$6.09x10^3$						
Naringenin 4'-O-glucuronide	447.0907	8.39	F	$3.14x10^5$	$1.67x10^5$	$4.66x10^4$	$1.09x10^4$	$1.89x10^4$	$1.88x10^4$	$1.04x10^5$	$7.36x10^3$	$6.87x10^3$
Procyanidin dimer B-type 2	577.1345	8.69	F	$6.44x10^4$	$5.46x10^4$	$2.77x10^4$				$5.42x10^3$	$6.31x10^3$	$1.16x10^4$
Unknown compound 6	431.0964	8.96	F	$4.94x10^4$	$2.39x10^4$	$5.59x10^3$				$6.25x10^4$	$3.14x10^4$	$1.12x10^4$
Dihydroquercetin	303.0487	9.05	F	$1.70x10^4$	$2.18x10^4$	$3.22x10^4$	$1.72x10^4$	$1.16x10^5$	$3.64x10^5$	$1.95x10^5$	$5.28x10^4$	$2.63x10^5$
Tectoridin	461.1068	9.09	F	$3.95x10^4$	$2.54x10^4$	$6.21x10^3$				$3.01x10^4$	$2.51x10^4$	$6.39x10^3$
Tetrahydroxyflavanone isomer	287.0539	9.82	F			$5.11x10^3$			$2.76x10^4$	$6.04x10^5$	$7.03x10^4$	$3.76x10^5$
Daidzin	415.1020	10.05	F	$3.31x10^6$	$3.51x10^6$	$1.65x10^6$	$2.52x10^6$	$1.38x10^6$	$9.68x10^5$	$3.45x10^6$	$3.50x10^6$	$1.74x10^6$
Unknown compound 11	301.0692	10.34	F	$3.85x10^4$	$3.13x10^4$	$2.30x10^4$	$5.77x10^4$	$4.89x10^4$	$1.40x10^4$	$2.17x10^4$	$1.60x10^4$	$1.50x10^4$
Isorhamnetin	315.0486	10.36	F	$6.25x10^4$	$4.66x10^4$	$1.43x10^4$	$6.48x10^4$			$1.71x10^4$	$8.88x10^3$	

Apigenin	**269.0436**	**11.03**	**F**	**$1.52x10^5$**	**$1.31x10^5$**	**$6.18x10^4$**	**$8.25x10^4$**	**$5.38x10^3$**	**$5.15x10^4$**	**$2.33x10^6$**	**$1.21x10^6$**	**$4.66x10^5$**
Unknown compound 14	299.0541	11.18	F	$9.17x10^4$	$9.05x10^4$	$4.59x10^4$	$2.11x10^4$		$5.35x10^3$	$4.13x10^5$	$2.97x10^5$	$8.45x10^4$

m/z= mass/charge; RT = retention time; PA = phenolic acids; F = flavonoids; OP = other polyphenols; UG = ungerminated sorghum; 72 h and 144 h are the germination times applied.



Supplementary Figure 6. Schematic depicting the phenylpropanoid pathway associated with the synthesis and degradation of some phenolic compounds during sorghum germination. The compounds in red were identified in our samples. UG = ungerminated grain.

Part IV

Interactions between protein and sorghum phenolic compounds: an unprecedented study

Chapter 6 – *"β-lactoglobulin and sorghum phenolic compounds molecular binding: interaction mechanism and thermal stability impact"*

Chapter VI - β-lactoglobulin and sorghum phenolic compounds molecular binding: interaction mechanism and thermal stability impact

Carolina Thomaz dos Santos D'Almeida [1,2], Anne Caroline de Albuquerque Sales [3], Ana Augusta O. Xavier [3], Hamza Mameri [4], Mariana Simões Larraz Ferreira [1,2], Guilherme M. Tavares [3]

[1] Laboratory of Bioactives, Food and Nutrition Graduate Program (PPGAN), Federal University of the State of Rio de Janeiro, UNIRIO, Brazil

[2] Center of Innovation in Mass Spectrometry, Laboratory of Protein Biochemistry, UNIRIO, Rio de Janeiro, 22290-240, Brazil.

[3] Department of Food Science and Nutrition, School of Food Engineering, University of Campinas, UNICAMP, Brazil

[4] UMR 1208 IATE, Univ. Montpellier, INRAE, L'Institut-Agro Montpellier, F-34060 Montpellier, France.

Highlights

- Sorghum-bran phenolic compounds and β-lg complexes were studied for the first time.
- Preferable binding sites changed according to the involved phenolic compounds.
- Antioxidant reactivity showed a synergistic effect between β-lg and phenolics.
- 72% of putatively identified phenolic compounds from sorghum are flavonoids.
- β-lg/catechin interaction improved the thermal stability of the phenolic compound.

This chapter was submitted to Food Chemistry (Impact Factor: 8.5).

ABSTRACT

The mechanism dictating the molecular interaction between β-lactoglobulin (β-lg) and sorghum bran phenolic compounds from 4 genotypes was studied, and the formed complexes were evaluated regarding their antioxidant capacity, and improvement of phenolic compounds' thermal stability. Catechin (CA) and ferulic acid (FA), representative compounds of sorghum grains, were used as model systems for comparison. Fluorescence quenching revealed higher affinity for β-lg:FA interaction (K_{sv} in the magnitude order of 10^5 M^{-1}) compared with β-lg:CA interaction (K_{sv} in the magnitude order of 10^4 M^{-1}). Nevertheless, regarding the molecular interaction between the proteins and the complex extracts of phenolic compounds, K_{sv} in the magnitude order of 10^4 M^{-1} were observed. Molecular docking indicated different preferable binding sites on the protein according to the phenolic compound chemical nature: flavonoid class bind near to a-helix region mainly by hydrophobic interactions, while phenolic acids bind close to the β-barrel (calix) region via hydrogen bonds/van der Waals force. Antioxidant capacity, determined by the FRAP and ORAC methods, progressively increased (+34% and +43%, respectively) after protein-phenolic interaction. Concerning the thermal stability of the phenolic compounds, HPLC-DAD-MS/MS technique identified epimerization as the primary response of CA to thermal treatment (90 ºC/10 min). The addition of β-lg exerted a protective effect against CA degradation (-7% in β-lg:CA complexes); however, proteins were not able to protect complex phenolic matrices (*e.g.* sorghum extracts). Our results may contribute to the sorghum-based food development with higher antioxidant capacity and thermal stability.

Keywords: HPLC-DAD-ESI-MS/MS, Molecular docking, phenolic profile, quenching fluorescence, *Sorghum bicolor* L., thermal treatment

1. Introduction

Global warming is considered one of the crucial current challenges in terms of food security (Giannini *et al.*, 2017). Based on this, climate-smart grains such as sorghum grains (*Sorghum bicolor* L.) have gained interest in recent years, due to their low need of input levels, high temperatures and drought tolerance (Hossain *et al.*, 2022). Sorghum is the fifth most produced cereal worldwide (58 million tons/year) and is the main staple crop for more than 500 million people in more than 30 countries (Khalifa & Eltahir, 2023; USDA, 2024a). From a nutritional perspective, sorghum is a gluten-free crop, rich in carbohydrates, proteins, vitamins, and minerals (Serna-Saldivar & Espinosa-Ramírez, 2019). Nonetheless, its focus on human consumption has been motivated by its health-promoting bioactive compounds, mainly phenolic compounds.

Phenolic compounds are specialized metabolites naturally biosynthesized by plants and are largely concentrated in the bran fraction. Flavonoids (*e.g.*, catechin) and phenolic acids (*e.g.*, ferulic acid) are the main sorghum phenolic classes, being equivalent to ~80% of total phenolic grain content (D'Almeida *et al.*, 2021b). Catechin ($C_{15}H_{14}O_6$; MW: 290.26 g/mol) are composed of two benzene rings (A- and B-rings), containing multiple hydroxyl groups, and a dihydropyran heterocyclic ring (the C-ring) (Shahidi *et al.*, 2021). Ferulic acid ($C_{10}H_{10}O_4$; MW: 194.18 g/mol) is a phenolic acid containing only a benzene ring, with an attached hydroxyl group (Shahidi *et al.*, 2021). These hydroxyl groups are essential for the phenolic antioxidant properties, which are linked to a reduced risk of developing several chronic diseases, including diabetes, obesity, cancer, and cardiovascular disease (Chen *et al.*, 2021; de Oliveira & de Alencar Figueiredo, 2024; Xu *et al.*, 2021).

Mechanistically, sorghum phenolic compounds can form complexes with kafirins (sorghum prolamins) via non-covalent and covalent interactions, in which soluble or insoluble complexes are formed with modified physicochemical attributes of both reactant molecules (Zhang *et al.*, 2021). Kafirin-phenolic interaction has been reported as one of the main factors responsible to reduce protein digestibility (Duodu *et al.*, 2003). However, due to the complexity of the kafirins protein structures, the mechanisms involving their molecular interaction are not yet well elucidated. Alternatively, several recent studies have used β-lactoglobulin (β-lg) as a model protein to simulate and understand possible interactions with phenolics (Liu *et al.*, 2022; Qie *et al.*, 2021; Rodrigues *et al.*, 2024; Zhang *et al.*, 2022). β-lg-phenolic interactions can be easily

induced during food formulation and processing, for instance, thus, understanding associated mechanisms and potential of the formed molecular complexes are of great interest from a scientific, industrial and economic perspectives (Zhang *et al.*, 2021).

β-lg is a small globular milk protein (18.3 kDa) characterized by a calyx structure formed by eight antiparallel β-strands and a single a-helix at outer surface of the β-barrel (Liang & Subirade, 2010). Due to its several binding sites, this protein can bind to multiple ligands and improves phenolic stability during food processing (Chen *et al.*, 2022b), which could be useful to improve the stability of phenolic compounds during sorghum processing. Phenolic compounds during thermal processing can be degraded, reducing the final product bioactive potential (Li *et al.*, 2022). Although some studies have already demonstrated the improvement of the thermal-stability for phenolic compounds after interaction with β-lg, they often addresses such behavior for binary systems composed by the protein and one isolated phenolic compound (Qie *et al.*, 2021; Zhang *et al.*, 2022). To the best of our knowledge, there are no studies considering the multi-binding properties of sorghum phenolic compounds in β-lg, which would be very relevant for the development of more complexes matrices, as well as to understand potential competitive-binding behaviors regarding different phenolic compounds.

In this context, we aimed to study the mechanisms dictating the molecular interaction between β-lg and sorghum bran phenolic compounds and the potential of the formed complexes regarding their antioxidant capacity, and improvement of thermal stability. Catechin and ferulic acid were selected as a model flavonoid and phenolic acid, respectively. These findings contribute to reduce the empiricism on optimizing the health benefits and functional properties of sorghum-based foods and allow us to hypothesize a possible complexation mechanism between phenolics and kafirins.

2. Material and methods

2.1. Material

β-lg was obtained by salting out (MatÉ & Krochta, 1994; Rodriguez *et al.*, 2023) from a commercial whey protein isolate (Glanbia Nutritionals, 94% protein, 0.9% fat, 2% lactose, and 0.4% minerals, according to the manufacturer). Ferulic acid (≥99% purity) and (+)-catechin (≥98% purity) were purchased from Sigma-Aldrich (Missouri, USA). HPLC-grade acetonitrile was obtained from J. T. Baker (Phillipsburg, USA). Ultra-pure

water was obtained through the Synergy® (Merck, Darmstadt, Germany) purification system. Other unmarked reagents were of analytical grade. For chromatographic analysis, solvents were filtered using 0.45 µm membranes from Millipore.

2.2. Sorghum samples and phenolic compounds extraction

Four Dusormil Dutch sorghum (*Sorghum bicolor*) with different pericarp colors (brown: HD7, white: HD19, and two black varieties: HD100 and HD101) were cultivated in Netherlands full fields and kindly provided by Maatschap de Milliano-Meijer (Oostburg, Netherlands). To obtain the sorghum bran (SB) fraction, whole-grains (80 g) were ≅15% dehulled (mass weight) into a laboratory-scale abrasive grain polisher (Model-TM 05C; Satake, Japan) at 800 rpm (Lestienne *et al.*, 2007, with some modification).

Phenolic compounds from SB were extracted in triplicate according to Santos *et al.* (2019b). Briefly, 210 mg of sample were extracted with 3 mL of 80% ethanol. Samples were stirred (200 rpm, 10 min, 25 °C) and centrifuged (5000×G, 10 min, 25 °C). The extraction was repeated twice, and the collected supernatants were combined. Sorghum phenolic extracts were evaporated (SpeedVac Savant, ThermoFisher Scientific, USA), resuspended in 1.5 mL of 20 mM phosphate buffer pH 7 and filtered (0.22 µm, hydrophilic PTFE) (Analytical). For a global estimative of sorghum extracts phenolic content, total reducing capacity was determined based on Folin-Ciocalteu method adapted for microplates (D'Almeida *et al.*, 2021b). Samples were stored at −80 °C until subsequent analyses.

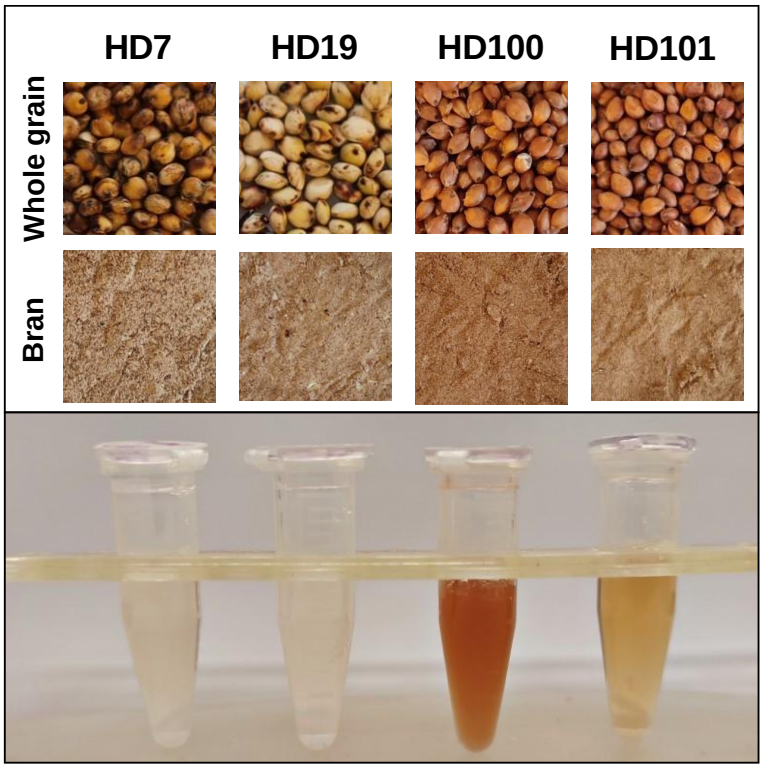


Figure 21. Image of whole grains, bran, and phenolic extracts from the Sorghum genotypes analyzed.

2.3. Protein-phenolic molecular interaction

β-lg stock dispersion was prepared by dispersing the lyophilized powder in a phosphate buffer (20mM, pH 7) with stirring for 30 min at room temperature ~25ºC (final concentration: 1 mM). Protein-phenolic complexes were obtained by mixing β-lg and (+)-catechin (CA), ferulic acid (FA) or SB extract phenolics (HD7, HD19, HD100 or HD101). β-lg concentration was fixed at 1 µM, whereas the phenolic compound concentration varied to obtain β-lg:CA, β-lg:FA and β-lg:SB mixtures in five different molar ratios (1:1; 1:2; 1:4; 1:6 and 1:7). Complexation was carried out under stirring (30 min) in the dark. Ligands free (without β-lg) in phosphate buffer were used as controls.

2.3.1. Fluorescence quenching

Protein-phenolic compound binding affinity was evaluated by fluorescence quenching according to (Vidotto & Tavares, 2022), in independent triplicate. Complexed solution at five different molar ratios (as mentioned above) were independently prepared in glass tubes and remained 15 min at 30 ºC in the dark, before analysis. After, solutions were transferred to quartz mini cuvettes, excited at 280 nm in a Synergy MX Multi-mode Reader (Biotek Instruments Inc. Winooski, USA), and fluorescence emission spectra was

read (300-450 nm). Absorption spectra were also obtained (230-450 nm) for inner filter effect correction as follows:

$$F_{corr} = F_{obs}\, x\, 10^{\frac{A_{exc} + A_{emi}}{2}} \quad (1)$$

where, F_{corr} is the fluorescence intensity after correction, F_{obs} is the read fluorescence intensity, and A_{exc} and A_{emi} are the absorption values at excitation and emission wavelengths.

The fluorescence quenching data for the protein-phenolic interaction were evaluated according to the Stern-Volmer equation (Lakowicz, 2006), as follows:

$$\frac{F_0}{F} = 1 + K_q\, t_0\, [Q] = 1 + K_{sv}\, [Q] \quad (2)$$

where, F_0 and F represent the fluorescence intensity in the absence and presence of ligand, respectively; [Q] is the ligand final concentration; K_q is the bimolecular quenching constant; K_{SV} indicates the Stern-Volmer quenching constant; and τ_0 is the lifetime of the quenched fluorophore ($t_0 = 10^{-8}$ for residue of tryptophan).

2.3.2. Molecular docking

Preferential binding sites between β-lg and phenolic compounds (CA and FA) were investigated by molecular docking. Protein structure was sourced from the RCSB Protein Data Bank (ID β-lg: 5IO5), while FA and CA structures were obtained from PubChem database (CID for CA: 9064 and CID for FA: 445858). Molecules preparations were conducted using Autodock4 software (Allouche, 2011), and potential binding sites were identified with Autodock Vina (Trott & Olson, 2010). Dynamics simulations were performed separately for each ligand and overlapping both ligands. Visualization of the results was done using PyMOL software (Schrödinger, 2015).

2.3.3. Antioxidant activity

The antioxidant capacity of samples was determined, in triplicate, by oxygen radical absorbance capacity (ORAC) and ferric reducing antioxidant power (FRAP) methods adapted to microplates, according to D'Almeida *et al.* (2021b) and Sompong *et al.* (2011).

For FRAP assays, the reagent was prepared in acetate buffer (0.3 M, pH 3.6), $FeCl_3.6H_2O$ (20 mM) and TPTZ (2,4,6-tri(2-pyridyl)-1,3,5-triazine) solution (10 mM) in a 10:1:1 ratio, respectively. A 20 µL aliquot of each extract was combined with 15 µL of

Milli-Q water and 265 µL of FRAP reagent, gently vortexed and incubated (30 min, 37 °C). Absorbance was measured using a microplate reader (FlexStation III, Molecular Devices, USA) at 595 nm. For ORAC method, 80 µL of the fluorescein solution (8 µM) and 80 µL of the sample or blank (20 mM phosphate buffer, pH 7) were added to the dark microplate. Then, 40 µL of 221 mM AAPH (2,2'-azobis(2-methyl-propanimidamide) dihydrochloride) was added. Fluorescence decrease was monitored using a Synergy MX Multi-mode Reader (Biotek Instruments Inc. Winooski, USA) operating with an excitation at 485 nm and an emission at 528 nm, every 1 min for 1 h. The antioxidant activity was determined by the area under fluorescence decay curve over time using GraphPad Prism software. The FRAP and ORAC results were expressed as µmol trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid) equivalents (TE) per g of sample, in dry basis.

2.4. Thermal stability of phenolic compounds assessed by HPLC-DAD-MS/MS

The thermal stability tests of protein-phenolic complexes were conducted according to the method of (Qin *et al.*, 2018), with some modifications. An aliquot of 1 mL of the different concentration and molar ratio of samples (β-lg:CA, β-lg:FA and β-lg:SB, and its single solution) was added to glass tubes and heated in a water bath at 90 □ for 10 min in the dark. After heating, the samples were quickly cooled to room temperature in an ice water bath for further analysis.

Before and after heat treatment phenolic compounds from protein-phenolic complexes (β-lg:CA, β-lg:FA and β-lg:SB) were injected into an Shimadzu HPLC (Kyoto, Japan) equipped with diode array detector (DAD) (SPD-M20A) and coupled with Bruker Daltonics Esquire 4000 spectrometer (Bremen, Germany), equipped with an ionization source electrospray and an ion-trap analyzer. An C18 Synergi Hydro column (4 µm, 250 × 4.6 mm, Phenomenex) at 29 °C and flow rate of 0.9 mL/min of ultra-pure water containing 0.5% formic acid (mobile phase A) and acetonitrile containing 0.5% formic acid (mobile phase B) was used according to the gradient method: A:B 99:1 to 50:50 in 50 min, following from 50:50 to 1:99 in 10 min. This latter ratio (1:99) was maintained for an additional 10 min. The column eluate was split to allow only 0.15 mL/min to enter the electrospray source interface.

MS/MS analysis was performed according to Chisté & Mercadante (2012), with modifications. The spectra were obtained between 200 and 600 nm, and the chromatograms were processed at 271 and 367 nm. The mass spectra were acquired with a scan range from *m/z* 100 to 800; the MS parameters were set as follows: ESI source in negative ion mode; 2000 V capillary voltage; −500 V end plate offset; −110 V capillary exit; 10 V skimmer 1; 5 V skimmer 2; 310 °C dry gas (N_2) temperature; 30 psi nebulizer, 1.4 V MS/MS fragmentation energy; and 1.8 V MS3 fragmentation energy.

2.5. MS-data processing and statistical analysis

All analyses were performed in triplicates, and the results were represented by mean and standard deviation. DataAnalysis v 4.2 software (Bruker, Germany) was used to acquire and process MS data. FA and CA were identified based on chemical standards runs parameters, such as: retention time and MS/MS fragments spectra. Phenolic compounds from SB extracts were annotated based on two online databases PubChem (www.pubchem.ncbi.nlm.nih.gov/) and Phenol-Explorer (www.phenol-explorer.eu/), on the literature and chemical characteristics of the molecules. Ligands free (without β-lg) in phosphate buffer were used as controls.

To annotation/identification of compounds, the Metabolomics Standards Initiative (MSI) level system was applied (Sumner *et al.* (2007): features matching with commercial standards (level 1), those annotated through MS/MS spectral data (level 2), putatively class annotation and chemical structure confirmed (level 3) and compounds with fragmentation data differing from the literature and/or databases were defined as "unknown compounds" (level 4). Peak area obtained from chromatograms were used to relative quantification.

Statistical analysis was performed with Tukey's test ($p < 0.05$) and one-way ANOVA, using XLSTAT software (Addinsoft, France). Multivariate Statistical Analysis was performed by Metaboanalyst 6.0 (https://www.metaboanalyst.ca/).

3. Results and discussion

3.1. Characteriztion of sorghum bran phenolic profile

Untageted HPLC-DAD-ESI-MS/MS metabolomics was applied as an exploratory method to characterize phenolic compounds in SB extracts of four genotypes possessing different pericarp colors. Initially, features were filtered and excluded based on their

presence in the blank and total absorbance below 15,000 mAU (considering the chromatograms baseline). After applying these filters, 18 phenolic compounds were putatively identified (Table 4) and the peak area intensities were highlighted on the total chromatogram of each SB genotypes (Supplementary Figure 9). Peaks **7** and **17** were respectively identified as catechin (*m/z* 289, RT 18.9) and ferulic acid (*m/z* 193, RT 28.1), based on coelution and comparison of mass spectra with the authentic standard (level 1). Catechin generated *m/z* 179 and 245 fragments, respectively characteristic of catechol group ($C_6H_6O_2$ loss, -110 Da) and C_2H_4O group (-44 Da) loses (Mohamed Yunus *et al.*, 2021); while ferulic acid MS/MS data showed th preferential CH_3 loss (-15 Da, *m/z* fragment 178), followed by CO_2 loss (-44 Da, *m/z* fragment 134) (Sinosaki *et al.*, 2020).

Five peaks (peaks **3**, **8**, **10**, **12**, and **13**) were annotated as catechin derivatives (level 2). Peak **10**, with a deprotonated precursor ion at *m/z* 289 (RT 21.4), exhibited a fragment ion at *m/z* 245 (a loss of 44 Da, corresponding to C_2H_4O group). This fragmentation pattern is similar to catechin, suggesting a tentative identification as its isomer, epicatechin (Mohamed Yunus *et al.*, 2021). Peak **8** had precursor ion $[M-H]^-$ at *m/z* 457 and fragments at *m/z* 305 $[M-H, -152\ Da]^-$ and 289 $[M-H, -168\ Da]^-$ indicating the initial loss of a galloyl unit (-152 Da, epigallocatechin formation) with subsequent loss of an oxygen (-16 Da, catechin formation), characteristic fragments of epigallocatechin gallate (Castro *et al.*, 2020). Peaks **12** and **13** displayed *m/z* 289 as main fragment and are confirmed as catechin derivatives, specifically catechin gallate and catechin glucoside, due to the loss of the galloyl (-152 Da) and glucoside (-162 Da) units, respectively (Chen *et al.*, 2022a).

Table 4. Annotation of phenolic compounds presents in sorghum bran extracts by HPLC-DAD-ESI-MS/MS.

Peak	RT (min)	$[M - H]^-$ *(m/z)*	MS/MS *(m/z)*	Putative identification	ID level*	Molecular formula	Class	Subclass	Reference
1	14.5	437	nd	Unknown compound 1	4	nd	nd	nd	nd
2	15.2	477	301; 178; 151	Quercetin 3-O-glucuronide	2	$C_{21}H_{18}O_{13}$	F	Flavonols	Kumar, *et al.* (2015)
3	16.8	577	451; 425; 407; 289; 287	Procyanidin dimer	2	$C_{30}H_{26}O_{12}$	F	Flavanols	Rue, *et al.* (2018)
4	17.3	463	301	Quercetin 3-O-glucoside	2	$C_{21}H_{20}O_{12}$	F	Flavonols	Kumar, *et al.* (2015)
5	17.9	179	135	Caffeic acid	2	$C_9H_8O_4$	PA	HCA	D'Almeida, *et al.* (2021)
6	18.6	433	301	Quercetin 3-O-arabinoside	2	$C_{20}H_{18}O_{11}$	F	Flavonols	Kumar, *et al.* (2015)
7	**18.9**	**289**	**245; 179**	**(+)-catechin**	**1**	**$C_{15}H_{14}O_6$**	**F**	Flavanols	**standard**
8	20.3	457	305; 289	Epigallocatechin gallate	2	$C_{22}H_{18}O_{11}$	F	Flavanols	Castro, *et al.* (2020)
9	20.9	253	235; 225; 195; 185	Daidzein	2	$C_{15}H_{10}O_4$	F	Isoflavonoids	Zhao, *et al.* (2018)
10	21.4	289	245	Epicatechin	2	$C_{15}H_{14}O_6$	F	Flavanols	Mohamed Yunus, *et al.* (2021)
11	21.9	447	301; 245; 151	Quercetin 3-O-rhamnoside	2	$C_{21}H_{20}O_{11}$	F	Flavonols	Kumar, *et al.* (2015)
12	23.0	441	289	Catechin gallate	2	$C_{22}H_{18}O_{10}$	F	Flavanols	Chen, *et al.* (2022)
13	24.3	451	289	Catechin glucoside	2	$C_{21}H_{24}O_{11}$	F	Flavanols	Chen, *et al.* (2022)
14	25.0	389	nd	Unknown compound 2	4	nd	nd	nd	nd
15	25.7	397	nd	Unknown compound 3	4	nd	nd	nd	nd
16	25.9	301	151	Quercetin	2	$C_{15}H_{10}O_7$	F	Flavonols	D'Almeida, *et al.* (2021)
17	**28.1**	**193**	**178; 134**	**Ferulic acid**	**1**	**$C_{10}H_{10}O_4$**	**PA**	**HCA**	**standard**
18	28.8	285	243	Kaempferol	2	$C_{15}H_{10}O_6$	F	Flavonols	March & Miao (2004)

Peak **3** (*m/z* 577, RT 16.8) generated *m/z* 451, 425, 407, 289 and 287 fragments. The fragment ion at *m/z* 451 resulted from heterolytic ring fission ($C_6H_6O_3$ loss, -126 Da). Retro-Diels-Alder reaction was also detected by fragment ion at *m/z* 425 ($C_8H_8O_3$ loss, -152 Da), followed by water loss (H_2O loss, -18 Da) to form the molecular ion at *m/z* 407. Finally, fragment ions at *m/z* 289 and 287 was consequent of the cleavage of quinone methide at the interflavan bond. The extensive MS/MS fragmentation observed for peak **3** enabled us to tentatively identify it as a procyanidin dimer (Rue *et al.*, 2018).

Peak **16** (RT 25.9) exhibited the $[M-H]^-$ at *m/z* 301 and produced the fragment ion at *m/z* 151, indicating a Retro-Diels-Alder reaction in compound C-ring and consequently a $C_7H_3O_4$ loss (-151 Da). Compound **16** was annotated as quercetin (Level 2) (D'Almeida *et al.*, 2021b). Four metabolites (peaks **2**, **4**, **6** and **11**) were identified as quercetin derivatives based on the presence of aglycone fragment ions at *m/z* 301. Peak **2** was identified as quercetin-3-O-glucuronide, due to its precursor ion $[M-H]^-$ at *m/z* 477 and a major fragment ion at *m/z* 301 (quercetin), indicating the loss of the glucuronic unit (-176 Da). Peak **4** presented precursor ion $[M-H]^-$ at *m/z* 463 ($C_{21}H_{20}O_{12}$) and was identified as quercetin 3-O-glucoside, due to its fragment ion (*m/z* 301), corresponding to glucosyl group loss (-162 Da). Peak **6** displayed the deprotonated ion $[M-H]^-$ at *m/z* 433 ($C_{20}H_{18}O_{11}$) with major MS/MS fragment at *m/z* 301 and 271, corresponding to the losses of glucuronyl (-176 Da) and glucosyl (-162 Da) moieties and this compound was annotated as quercetin 3-O-arabinoside. Peak **11** was annotated as quercetin 3-O-rhamnoside, due to its precursor ion $[M-H]^-$ at *m/z* 447 and fragments at *m/z* 301 (rhamnosyl loss, -146 Da) (Kumar *et al.*, 2015).

Catechin and quercetin derivatives account for 61% of annotated compounds and highlight the importance of these two compounds in SB; however, other characteristic metabolites were also detected in the SB samples. Peaks **9** and **18** present in all SB extracts were classified as flavonoids derivatives. For peak **9**, the precursor ion was observed at *m/z* 253 and further fragmentation at *m/z* 235 (H_2O loss), 225 (CO loss), 195 (CO loss from *m/z* 225 ion fragment) and 185 (C loss from *m/z* 195 ion fragment); hence, was tentatively identified as daidzein (level 2) (D'Almeida *et al.*, 2021b; Zhao *et al.*, 2018). The molecular ion of peak **18** was observed at *m/z* 285 and identified as kaempferol due to its MS/MS fragment at *m/z* 243 (March & Miao, 2004). Peak **5** was observed at *m/z* 179 (caffeic acid), and loss of CO_2 was observed as a characteristic ion at *m/z* 135 (D'Almeida *et al.*, 2021b). Three peaks (**1**, **14** and **15**) were detected in the SB

samples (HD100 and HD 101); however, their tentative identification could not be determined due to the lack of MS/MS information (unknown compounds, level 4).

To explore the variation in the data set, the PCA (scores-samples; loadings-phenolic compounds) was applied to investigate the degree of similarity or dissimilarity between the compounds relatively quantified, from peak area in each sample (Figure 30). The left-PCA (Figure 30A) revealed a grouping between CA and the SB genotypes (HD7, HD19, and HD101), indicating a similar phenolic profile in these extracts, particularly with flavonoids such as catechin and its derivatives. FA was separated from the other samples along the y-axis (PC2 = 35%), while HD100 showed distinct separation along the x-axis (PC1 = 55%). Despite HD100 having catechin derivatives as the major peaks in the chromatogram, the substantial presence of ferulic acid also contributes to its unique phenolic profile (Supplementary Figure 7).

A subsequent PCA, focusing solely on SB extracts (Figure 30B), revealed that the similarity between genotypes is independent of the sorghum grain pericarp color. HD7 (brown pericarp) and HD101 (black pericarp) demonstrated a high degree of similarity in their phenolic profiles. HD19 (white pericarp) was positioned in the same left-quadrant, separated only by the y-axis (PC2 = 16%), whereas HD100 (black pericarp) again exhibited a distinct phenolic profile, positioned in the right-quadrant (PC1 = 83%). Moreover, procyanidin dimer (peak 3) was found in all SB samples in the follow decreasing order of peak intensity: black pericarp HD101> white pericarp HD19 > brown pericarp HD7 > black pericarp HD100. This result can be explained by the fact that tannin content present low correlation with pericarp/testa coloration, indeed sorghum varieties with lighter pigmented but thicker testa might have higher tannin content than those darker pigmented, thinner testa (Zhang *et al.*, 2023).

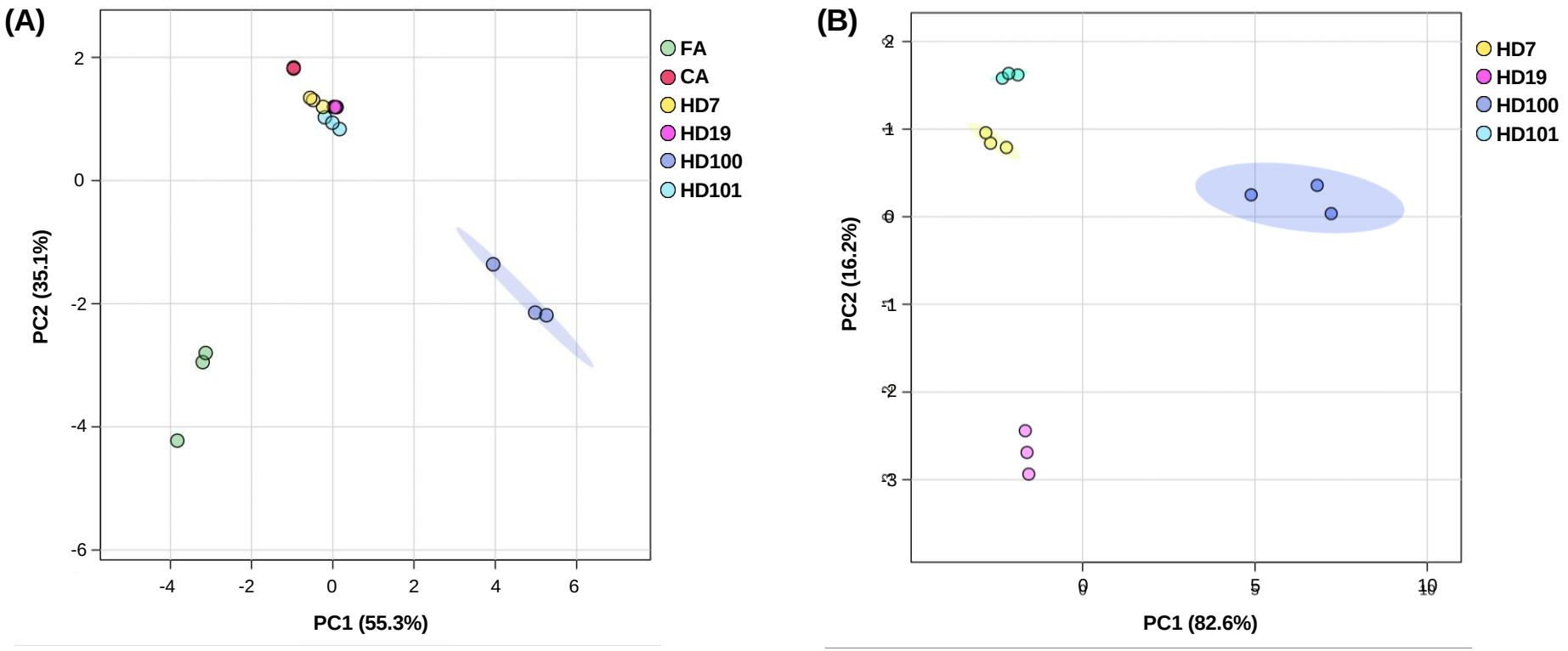


Figure 22. Principal component analysis (PCA) with all phenolic extracts (A) and only with sorghum samples (B). The samples (symbols) are distributed according to relative intensity of identified phenolic compounds. FA: ferulic acid; CA: catechin. Sorghum bran HD varieties.

3.2. Molecular binding of phenolic compound to β-lactoglobulin

In order to elucidate the binding mechanism of the different compounds to β-lg, fluorescence quenching experiments were conducted. Figure 31 (A-F) shows the fluorescence emission spectra (300–450 nm) of β-lg dispersions in the absence and in the presence of increasing concentrations of FA, CA and SB extracts, respectively. The β-lg λmax (325 nm) shifted during titration with CA and HD19 samples (11–13 nm); while the remaining β-lg fluorescence quenching spectra showed no significant shifts (Figure 31). According to the phenolic characterization carried out among SB samples (please refer to section 3.1), the flavonoid catechin is the most abundant phenolic compound of HD19 variety. Therefore, the significant peak shift of λmax suggested a conformational change in β-lg when interacting with CA and consequently, the water molecules penetration near some tryptophan residues, altering its microenvironment polarity (Zhang *et al.*, 2021).

Fluorescence data were analyzed by Stern-Volmer model and displayed satisfactory adjustment, thus K_{SV} and K_q could be obtained (Supplementary Table 8). For all tested samples, the K_q values (Supplementary Table 8) were higher than the maximum diffusion collision quenching constant (2.0×10^{12} M^{-1}), indicating that the static quenching played a dominant role in the interaction between β-lg and phenolic compounds. Under such conditions the Ksv values may be referred to as an association constant (Lakowicz, 2006). The β-lg-CA K_{SV} value found in our study (2.80×10^4 M^{-1}) is in agreement with the previously reported by Al-Shabib *et al.* (2020); but surprisingly β-lg-FA displayed Ksv values with a magnitude of order of 10^5 M^{-1}. The higher affinity between β-lg and FA can be explained by the small molecular size of this phenolic compound when compared to catechin, and, therefore, to a better access to the protein central cavity, an imporant binding site in β-lg native structure (Kontopidis *et al.*, 2004). SB samples showed similar K_{SV} values to CA, which suggests that both ligands may compete for similar binding sites on the protein structure. This hypothesis corroborates the data presented in Figure 30A, *i.e.*, sorghum phenolic profile has more similarity and were grouped together with CA.

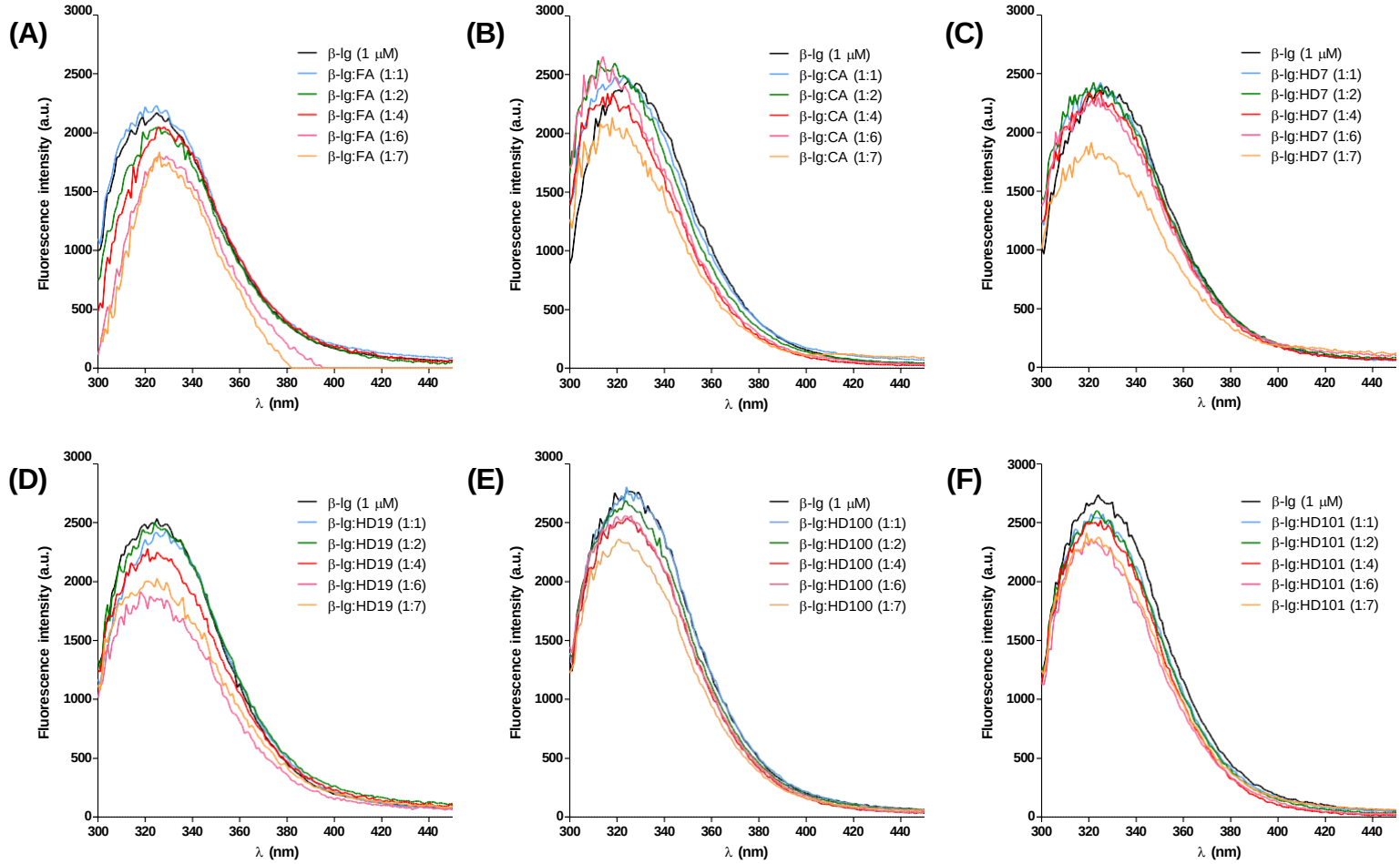


Figure 23. Fluorescence spectra of protein:phenolic compound complexes molar ratios 1:1 to 1:7 at the concentration of 1 μM of protein. (A) β-lg:FA; (B) β-lg:CA; (C-F) β-lg:sorghum samples (HD7, HD19, HD100 and HD101, respectively). β-lg: β-lactoglobulin; FA: ferulic acid; CA: catechin.

3.3. Effect of the molecular interaction on the antioxidant reactivity

The antioxidant capacity of the isolated and complexed samples determined by FRAP and ORAC methods are shown in Figure 32. Firstly, the isolated compounds (FA, CA and SB phenolic extracts) were evaluated at identical concentration of those used in the different mixtures (*e.g.* β-lg-FA, β-lg-CA and β-lg-SB) to achieve each molar ratio (Figure 32, red dots). Antioxidant capacity by FRAP method varied between samples in the following decreasing order: HD100 > HD101 > HD7 > HD19 > CA > FA; while ORAC analysis showed a slightly different behavior: HD100 > HD101 > HD19 > HD7 > FA > CA. This small variation between analyzes reinforces the importance of combining methods for the elucidation of antioxidant capacity between extracts; however, the good correlation between FRAP and ORAC ($r = 0.74$; $p = 2.46 \times 10^{-6}$) was not affected. Black sorghum pericarp phenolic superiority was demonstrated by Folin-Ciocalteu method (data not shown) and displayed a strong positive correlation with FRAP ($r = 0.90$; $p = 7.42 \times 10^{-5}$) and ORAC ($r = 0.99$; $p = 1.14 \times 10^{-11}$).

Alone, β-lg had the lowest FRAP and ORAC values (2.76±0.03 and 3.73±0.31 µmol TE/g sample). When complexed with the ligands, FRAP and ORAC values progressively increased as the molar ratio of phenolics increased. These findings highlight the great antioxidant potential of phenolic compounds, even when complexed with proteins. Among phenolic standards, CA had an increase of 308% and 96% between β-lg:CA 1:1 and 1:7 ratio in the FRAP and ORAC analyses, respectively; while FA had a less pronounced increase in FRAP (169%) and was similar in ORAC (96%). SB phenolic extracts showed an increase between 53-129% (FRAP method) and 110-156% (ORAC method). For both analyzes, the HD100 variety showed the highest antioxidant capacity but milder variation between molar ratios.

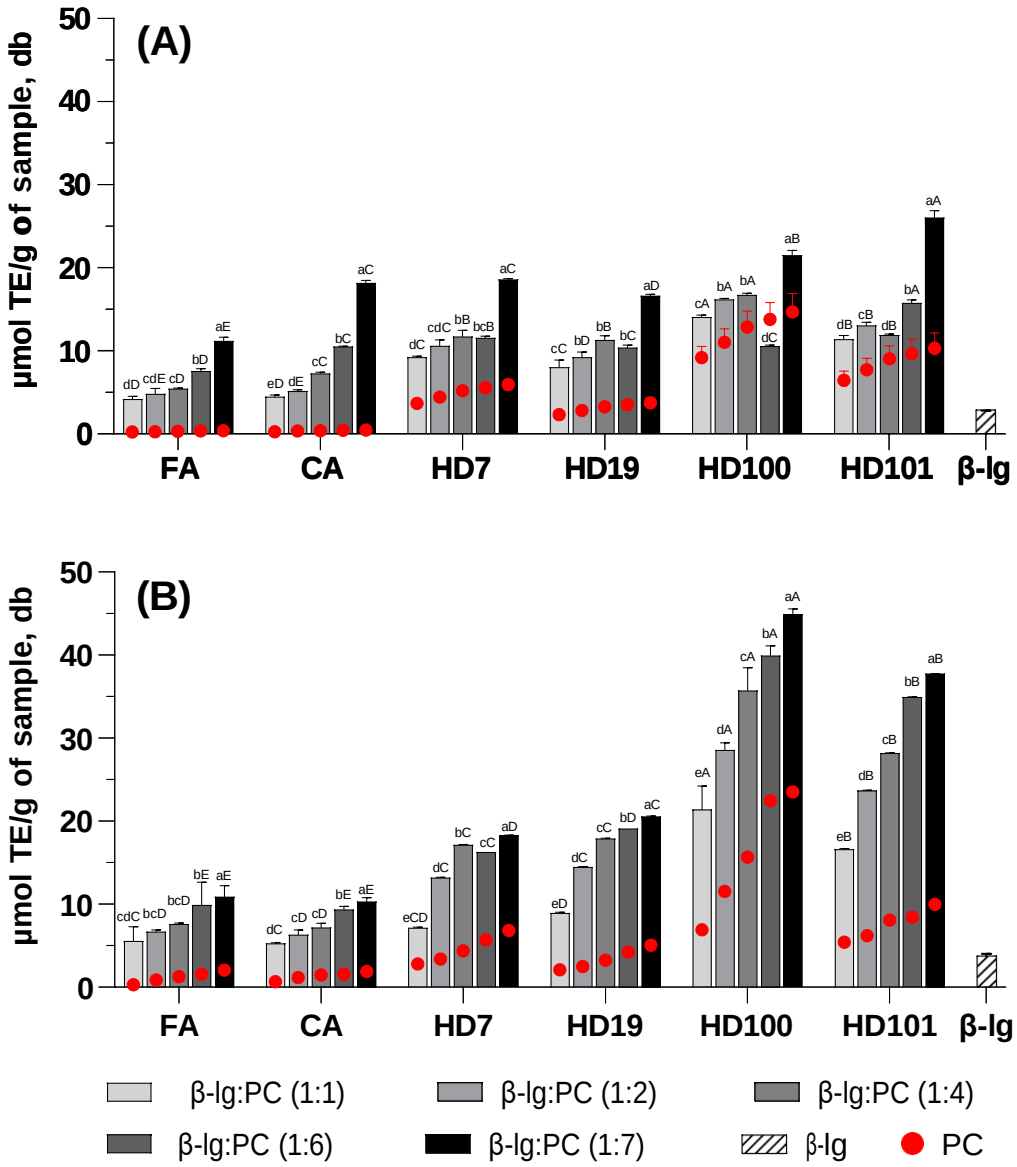


Figure 24. Antioxidant capacity measured by FRAP (A) and ORAC (B) assays of β-lg, β-lg:FA, β-lg:CA and β-lg:SB complexes. Complexes were obtained at protein:phenolic compound (β-lg:PC) molar ratios 1:1 to 1:7 at the concentration of 1 µM of protein. Values are expressed as mean ± standard deviation. Different lowercase and uppercase letters mean the significant difference ($p < 0.05$) between different molar ratio of the same ligand and the same molar ratio of different ligands, respectively, following a one-way ANOVA test. Red dots indicate the value for the isolated phenolic compound at the same concentration present in the complexes. β-lg: β-lactoglobulin; FA: ferulic acid; CA: catechin; PC: phenolic compounds. SB: sorghum bran varieties. TE: trolox equivalent.

Variation between methods can be explained by the preferential binding force between phenolic and protein. According to Al-Shabib *et al.* (2020), β-lg-CA interactions occur mainly due to hydrophobic interactions between the nonpolar aromatic ring of phenolics and hydrophobic sites of protein molecules; while hydrogen bonds and van der Waals forces were the main ones involvend in the interaction between β-lg and FA (Jia *et al.*, 2017). Phenolic hydroxyl groups donate electrons to ferric ions in the FRAP method, *i.e.*, higher available hydroxyls result in higher absorbance (antioxidant capacity). Due to the greater number of hydroxyl groups in CA molecules (five units) when compared to FA (one unit), CA naturally had greater antioxidant capacity in this method. When complexed with proteins, FA hydroxyl groups are partially occupied with amino groups of proteins, causing a more subtle antioxidant effect and less variation between 1:1 and 1:7 molar ratio.

To evaluate the effect of the molecular interaction in relation to the antioxidant reactivity of the isolated compounds (FA, CA and sorghum phenolic extracts), the sum of the values obtained for the isolated β-lg and compounds was calculated and compared with the complexes formed. All protein-phenolic complexes (ligands and molar ratio) showed values higher than the sum of the isolated compounds for FRAP (~34%) and ORAC (~43%) methods, reinforcing the previously mentioned hypothesis of protein conformational change and, consequently, better access to the protein nucleophilic centers previously unavailable due to steric hindrance (de Morais *et al.*, 2020a).

The results of antioxidant reactivity in this study differ from those reported by Arts *et al.* (2002), who observed a masking effect in antioxidant capacity following phenolic-protein interaction. Typically, the antioxidant capacity of the formed complexes is lower than the combined capacity of the phenolic compounds and proteins individually (Ozdal *et al.*, 2013). This reduction occurs because the interaction with β-lg partially occupies the active hydroxyl groups of the phenolic compounds, inhibiting their antioxidant capacity (Qie *et al.*, 2020). However, in the present study, phenolic reactivity increased even after interaction with the protein, indicating a potential synergistic effect.

3.4. Molecular modeling study

Molecular docking study was conducted to investigate the preferable binding sites of FA and CA (individually and overlapping) on the of β-lg protein structures. The eight most preferable interactions of each ligand were simulated and their molecular

configurations are shown in Figure 33. Furthermore, the binding energy and amino acids involved in each binding site were calculated (Supplementary Table 9).

CA binding energy varied between -5.8 and -5.6 kcal/mol value and globally involved 25 amino acids in the interaction: Ala 118, Ala-111, Ala-86, Asn-109, Asn-88, Asn-90, Asp-85, Glu-108, Glu-112, Ile-56, Ile-71, Ile-84, Leu-31, Leu-39, Leu-58, Leu-87, Leu-117, Lys-60, Lys-69, Lys-91, Met-107, Pro-38, Ser-116, Val-41, Val-92 (Supplementary Table 9). These amino acids had already been partially (36%) reported as characteristic of the β-lg:CA complex (Al-Shabib *et al.*, 2020), but our results showed 16 new amino acids residues participating to the molecular binding, *e.g.*, alanine, asparagine, aspartic acid, glutamic acids, lysine and serine. The profile of amino acids residues involved in the binding site indicates a predominance of hydrophobic interactions between the molecules, as described by Al-Shabib *et al.* (2020) and Dai *et al.* (2022). Figure 33A and Supplementary Table 9 showed a little variability among the eight most preferable binding sites of these phenolic compound, *i.e.*, CA has a preferred β-lg-bind region and the same amino acids are involved in this complex. CA bindings 2-8 showed 57-83% similarity with the main CA interaction (binding site 1).

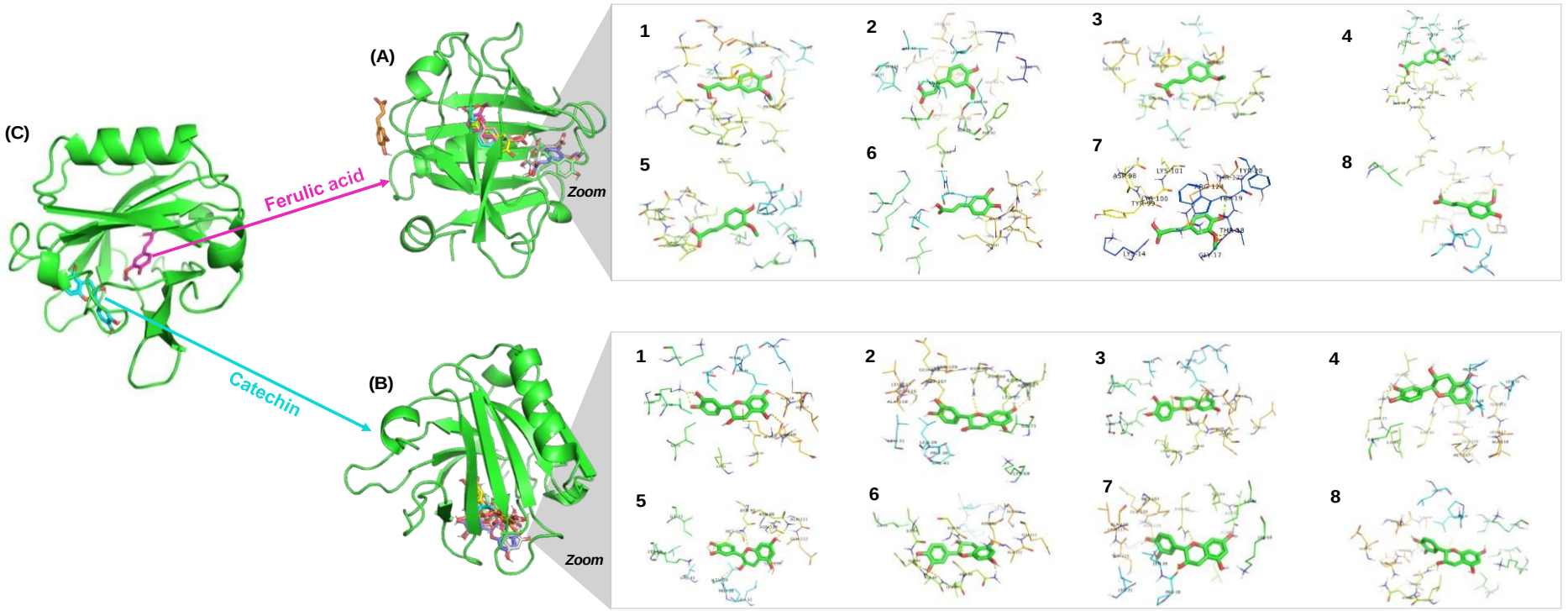


Figure 25. Preferable binding site according to rigid molecular docking simulations for the interaction between β-lg and catechin (A), ferulic acid (B) and both phenolics simultaneous binding (C). Amino acids involved in the eight preferable binding sites are detailed in the inserts.

Interestingly, the FA showed a different complexation behavior. β-lg:FA binding energy value varied between -6.0 and -4.3 kcal/mol and involved 49 amino acids in the interaction, presenting arginine, glutamine, glycine, phenylalanine, tryptophan and threonine as differential amino acids with CA (Supplementary Table 9). Binding energy values estimated by the molecular docking for β-lg:FA were higher than those found for β-lg:CA (Supplementary Table 9), confirming the stronger affinity between protein and FA previusly showed in fluorescence quenching experiments (Supplementary Table 8). In accordance with Jia *et al.* (2017) e Zhang *et al.* (2022), the profile of amino acids residues involved in binding site are mainly from hydrophilic and charged groups, suggesting hydrogen bonds and van der Waals force as the main ones dictating the moelecular interaction between β-lg and FA. Dissimilar to CA, FA showed great variability among the eight preferential interactions with β-lg (0-95%), suggesting the compound's affinity for different binding sites (Figure 33B).

Simultaneous binding of both phenolics was simulated and their molecular configurations are shown in Figure 33C. In the present study, CA and FA showed binding sites in different regions of the β-lg structure, *i.e.*, these phenolics do not compete for the same binding site on the protein. The hypothesis is that the slight change in protein conformation after interaction with CA, as evidenced by the shift in fluorescence quenching, may alter the accessibility of FA to its preferred binding sites. Our data proposes the binding to β-lg of the two main sorghum phenolic classes: flavonoid class bind near to a-helix protein region by hydrophobic interactions, while phenolic acids bind to β-barrel (calix) by hydrogen bonds and van der Waals force.

3.5. Effect of complexation on phenolic thermal stability

Futhermore, to investigate the protective effect of β-lg on the thermal stability of phenolic compounds, isolated extracts from CA, FA, and SB samples, as well as those complexed with β-lg, were subjected to heating at 90 °C for 10 min. Figure 34 displayes the comparative samples chromatograms at 25 ºC and 90 ºC. The use of two different commercial standards, CA and FA, allowed the observation of the distinct responses of the two main phenolic classes in sorghum grain: flavonoids and phenolic acids.

Flavonoids are well known for their thermosensitivity, which can result in its degradation, oxidation, or structural changes. For instance, CA (non-epimerized form) readily undergoes epimerization to epiCA (epimerized form) within a temperature range

of 34 to 100 °C (Kim *et al.*, 2022; Wang *et al.*, 2008). Comparing the chromatograms (Figure 34), we observed that approximately 12% of catechin at 25 °C decreased after 10 minutes of thermal process (90 °C), with two simultaneous reaction: the reduction in the CA peak was primarily due to epimerization (peak 10), while the remainder was attributed to the degradation of this thermolabile compound. In contrast, the FA peak remained unchanged, corroborating the greater phenolic acid class thermal stability.

Sorghum samples were also affected by thermal treatment based on their phenolic profile, particularly the proportion of flavonoids and phenolic acids in the SB. Due to the lower total phenolic content (data not shown) and the lower variability in its profile (see Section 3.1), the temperature impact on phenolic compounds was not noticeable on the genotype HD7. In white SB, CA – the main compound in HD19 – was drastically reduced after heat treatment (-79%). Notably, unlike the CA standard, there was no increase in peak 10 (epicatechin). This suggests that within the complex phenolic matrix of SB, the degradation of catechin is more favored than its epimerization.

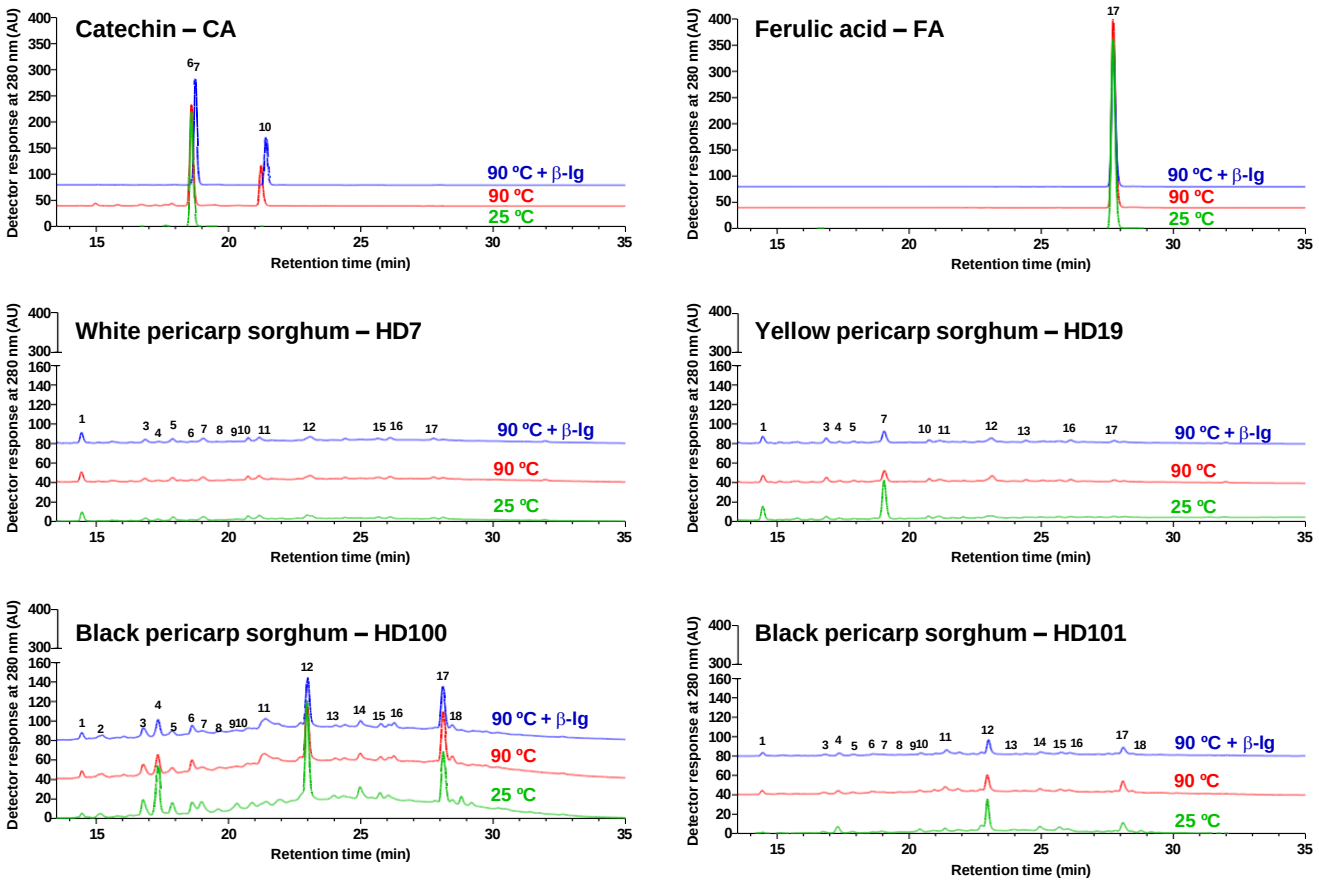


Figure 26. Chromatogram obtained by HPLC-DAD-ESI-MS/MS of catechin (CA), ferulic acid (FA) and phenolic compounds from SB extracts (HD samples) pre (25 ºC, green lines), post thermal treatment (90 ºC during 10 min; red lines) and post thermal treatment with protein (β-lg; blue lines). For peak characterization, please refer to Table 1.

Black pericarp genotypes (HD100 and HD101) exhibited analogous phenolic profiles and consequently responded similarly to thermal treatment. Catechin gallate and ferulic acid were identified as the two major phenolic compounds at 25 ºC. While ferulic acid remained unaffected at 90 ºC for 10 min, catechin gallate showed 38% and 42% degradation in the HD100 and HD101 genotypes, respectively. Additionally, the HD100 bran showed a 61% reduction in the quercetin 3-O-glucoside peak after heat treatment. According to Peng & Shahidi (2023) the degradation of glycosylated flavonoids typically results in their aglycone forms as the main products. However, the absence of an increase in peaks 7 and 16 (catechin and quercetin, respectively) in this study suggests the simultaneous degradation of these monomers and their glycosides forms.

Previous studies have indicated that the degradation of these phenolic compounds is more subtle when interacting with β-lg (Qie *et al.*, 2021; Zhang *et al.*, 2022). This hypothesis was tested in the present work, and the chromatograms of each protein-ligand complex after thermal treatment are presented in Figure 34 (blue lines). Although FA did not present isomerization or polymerization after heat treatment at 90 ºC for 10 min, the addition of β-lg resulted in a 28% reduction in its peak area. This reduction can indicates the proportion of FA that interacts with β-lg under the conditions analyzed. For CA, β-lg interaction reduced the epimerization reaction, with the epicatechin/catechin ratio remaining similar (0.40 without protein and 0.44 after interaction with β-lg). Moreover, the thermosensitivity of this compound was significantly reduced upon forming the β-lg complex, resulting in only 6% degradation.

Typically, the degradation of flavonoid molecules occurs through autoxidation induced by reactive oxygen species generated at elevated temperatures and/or the cleavage of the C-ring within the flavonoid structure (Peng & Shahidi, 2023). Both processes adversely affect this phenolic class bioactivity. Based on these mechanisms, some hypotheses can be proposed to explain the protective effect of β-lg against the thermal flavonoids degradation: 1) β-lg can form a protective barrier around compounds, reducing their exposure to heat and preventing direct thermal degradation. 2) β-lg:flavonoids interaction, mainly involving hydrophobic interactions (section 3.4), can stabilize the phenolic compounds and make them less susceptible to thermal degradation; 3) as noted in section 3.3, β-lg also has antioxidant properties, contributing to the stabilization of phenolic compounds by scavenging free radicals generated during heating.

In contrast, β-lg may not exhibit a significant protective effect against thermal degradation when incorporated into SB extracts (Figure 34, HD samples). Catechin in HD19 (-81%), catechin gallate in HD100 (-48%) and HD101 (-66%) and quercetin 3-O-glucoside in HD100 (-66%) extracts were degraded even with the protein presence. The complex composition of sorghum phenolic matrix, intensified by the presence of multiple catechin and quercetin derivatives in SB extracts, can lead to competition for protein binding sites. Despite β-lg has demonstrated substantial protective effects against the degradation of catechins in isolated scenarios, extracts containing a diversity of phenolics, such as SB extract, may modify the β-lg:flavonoids interactions and the effectiveness in stabilizing individual compounds. This hypothesis is supported by Sahihi & Ghayeb (2014) and Li *et al.* (2018), who demonstrated by molecular docking that different flavonoid compounds exhibit a preference for similar binding sites within the β-lg calyx.

4. Conclusion

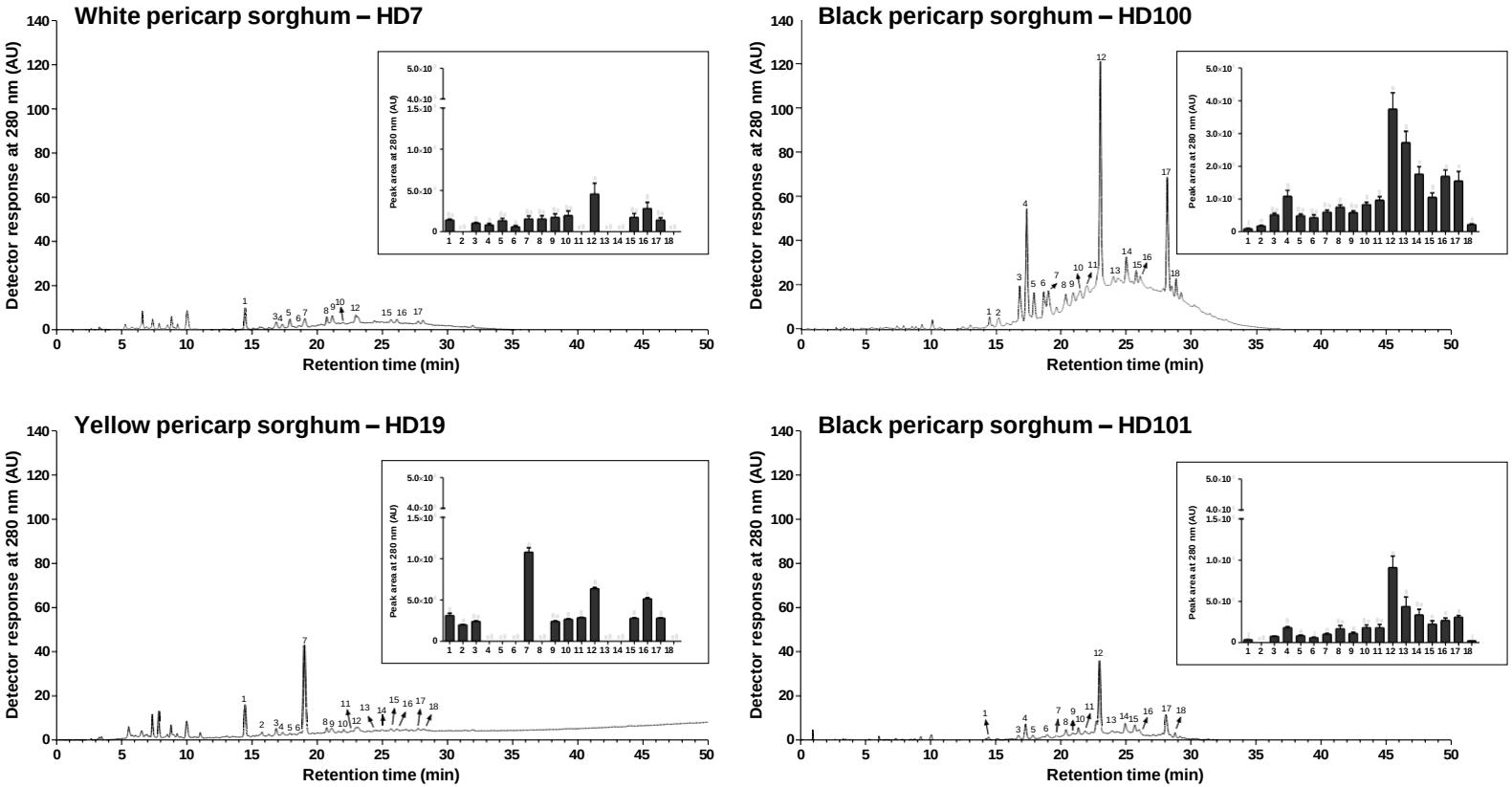
For the first time, the interactions mechanisms between β-lg and sorghum bran phenolic compounds was described. Our study can contribute to important gaps in the future understanding of kafirin-phenolic interactions, as well as their implications during heat treatment. Two effective experimental design strategies were employed: firstly, the use of two model phenolic compounds (CA and FA) to elucidate the interactions between β-lg within a simplified system; and secondly, the diverse range of sorghum genotypes with contrasting pericarp colors to understand the interactions between β-lg and sorghum phenolic compounds, independent of their profile variability.

Due to the predominance of flavonoids compounds, SB and CA model presented a similar interaction mechanisms with β-lg. CA and FA displayed distinct interaction mechanisms with β-lg, indicating that flavonoids and phenolic acids preferentially bind to different protein sites and their interactions are mainly driven by different molecular forces. The results also revealed the protective effect of protein against the thermal degradation of phenolic compounds in simplified model systems. However, increasing the complexity of the phenolic matrix (*e.g.* sorghum extracts) can alter the β-lg:phenolic interactions. Further studies are necessary to optimize β-lg application in complex food systems to fully leverage its protective capabilities.

5. Acknowledgements and funding

The authors are grateful to “Maatschap de Milliano-Meijer” and Walter de Milliano for providing and collecting samples. This work was financed by UNIRIO, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (401053/2019-9; 314100/2023-7; 407350/2023-3); Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (E-26/201.317/2022; E-26/210.152/2023; E-26/210.008/2024); Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Grants 2018/03822-5 and 2018/23752-1); and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (code 001).

6. Supplementary material



Supplementary Figure 7. Chromatogram obtained by HPLC-DAD-ESI-MS/MS of phenolic compounds from sorghum bran extracts, highlighting the eighteen putatively identified peaks. Insert graph represents the compound peak area in the respective sorghum genotypes. For peak characterization, please refer to Table 1.

Supplementary Table 7. The association constants (K_{sv}) and bimolecular quenching constants (K_q) for the interaction between β-lg and catechin (CA), ferulic acid (FA) or sorghum bran extracts (HD samples).

Protein	Ligand	K_{sv} ($\times 10^4$ M^{-1})	K_q ($\times 10^{12}$ M^{-1})
β-lg	CA	2.80 ± 1.78^{bc}	2.80 ± 1.78^{bc}
	FA	10.08 ± 2.53^{a}	10.08 ± 2.53^{a}
	HD7	4.45 ± 0.89^{b}	4.45 ± 0.89^{b}
	HD19	3.65 ± 1.36^{bc}	3.65 ± 1.36^{bc}
	HD100	2.34 ± 0.17^{c}	2.34 ± 0.17^{c}
	HD101	2.51 ± 0.81^{c}	2.51 ± 0.81^{c}

Values are expressed as the mean ± standard deviation. Different lowercase letters (a-e) in the same column indicate significant differences between ligands ($p < 0.05$). K_{sv} = Stern-Volmer constant; K_q = Bimolecular quenching constant.

Supplementary Table 8. Binding energies and amino acids involved in the protein-ligand binding site, according to the molecular docking simulations.

Protein	Ligand	Binding site number	Binding energy (kcal/mol)	Amino acids located in the preferable binding site
β-lg	Catechin	1	-5.8	Leu-31, Pro-38, Leu-39, Val-41, Leu-58, Lys-60, Lys-69, Ile-71, Ileu-84, Asn-90, Met-107, Glu-108, Asn-109, Ser-116, Leu-117, and Ala-118
		2	-5.8	Leu-31, Pro-38, Leu-39, Val-41, Lys-69, Ile-71, Ile-84, Asp-85, Ala-86, Leu-87, Asn-88, Asn-90, Met-107, Glu-108, Asn-109, Ser-116, Leu-117, and Ala-118
		3	-5.7	Leu-31, Pro-38, Leu-39, Val-41, Ile-56, Leu-58, Lys-69, Ile-71, Ile-84, Asn-90, Val-92, Met-107, Glu-108, Asn-109, Glu-112, Ser-116, Leu-117, and Ala-118
		4	-5.7	Leu-31, Pro-38, Leu-39, Ile-71, Ile-84, Asp-85, Ala-86, Leu-87, Asn-88, Asn-90, Met-107, Glu-108, Asn-109, Ser-116, Leu-117, and Ala-118
		5	-5.7	Leu-31, Pro-38, Leu-39, Val-41, Leu-58, Lys-69, Ile-71, Asn-88, Asn-90, Met-107, Asn-109, Ala-111, Glu-112, and Ser-116
		6	-5.6	Leu-31, Pro-38, Leu-39, Ile-71, Ile-84, Asp-85, Ala-86, Leu-87, Asn-88, Asn-90, Asn-109, Ala-111, Glu-112, and Ser-116
		7	-5.6	Leu-31, Pro-38, Leu-39, Lys-69, Ile-71, Ile-84, Asp-85, Ala-86, Leu-87, Asn-90, Met-107, Glu-108, Asn-109, Glu-112, Ser-116, Leu-117, and Ala-118
		8	-5.6	Leu-31, Pro-38, Leu-39, Val-41, Leu-58, Lys-60, Lys-69, Ile-71, Leu-87, Asn-88, Asn-90, Lys-91, Met-107, Glu-108, Asn-109, Glu-112, Ser-116, Leu-117, and Ala-118
	Ferulic acid	1	-6.0	Ile-12, Val-15, Val-41, Val-43, Leu-46, Leu-54, Ile-56, Leu-58, Ile-71, Phe-82, Ile-84, Val-92, Leu-93, Val-94, Leu-103, Phe-105, Met-107, Gln-120 and Leu-122
		2	-6.0	Ile-12, Val-15, Val-41, Val-43, Leu-46, Leu-54, Ile-56, Leu-58, Ile-71, Ala-73, Phe-82, Ile-84, Val-92, Val-94, Leu-103, Phe-105, Met-107, Gln-120 and Leu-122
		3	-6.0	Val-41, Val-43, Leu-46, Leu-54, Ile-56, Leu-58, Ile-71, Ile-84, Asn-90, Val-92, Val-94, Leu-103, Phe-105, Met-107, and Leu-122
		4	-5.1	Val-43, Leu-46, Leu-54, Ile-56, Leu-58, Ile-71, Ile-84, Asn-90, Lys-91, Val-92, Val-94, Leu-103, Phe-105, Met-107, and Leu-122
		5	-4.5	Pro-38, Leu-39, Val-41, Leu-58, Lys-60, Lys-69, Ile-71, Ile-84, Asp-85, Ala-86, Leu-87, Asn-88, Glu-89, Asn-90, and Met-107
		6	-4.4	Pro-38, Leu-39, Val-41, Lys-60, Lys-69, Ile-71, Asn-90, Met-107, Glu-108, Asn-109, Glu-122, Ser-116, Leu-117, and Ala 118
		7	-4.4	Lys-14, Gly-17, Thr-18, Trp-19, Tyr-20, Asp-98, Tyr-99, Lys-100, Lys-101, Arg-124, and Thr-125
		8	-4.3	Leu-31, Pro-38, Leu-39, Ile-71, Ile-84, Asn-88, Asn-90, Met-107, Glu-108, Asn-109, and Ser-116

Part V

General discussion, conclusions and perspectives

(Chapter 7)

Chapter 7 - General discussion, conclusions and perspectives

1. Biochemical transformations during sorghum grain maturation – a comprehensive study of protein accumulation, and phenolic synthesis

From an agronomic perspective, a crop development can result in-depth insights into factors influencing its production positively or negatively. Anticipating these impacts can enhance crop management decisions and increase the profitability. In sorghum cultivation, establishing the grain maturation stages is not a trivial task and cannot be done based only on visual transformations or calendar dates, requiring careful examination. The development stages of sorghum grain are influenced by various factors, including environmental conditions, management practices, and genetic factors. Temperature and light intensity affect the sorghum growth rate and grain filling process, while different sorghum varieties have varying developmental timelines and responses to environmental conditions. In Chapters 1 and 2 of this thesis, the study successfully delineates the complete development stages of sorghum grain and highlights the key biochemical transformations occurring from anthesis (flowering) to physiological maturation (full grain filling). The primary biochemical changes in sorghum grain are illustrated in Figure 35.

Three genotypes (cv. Macia, and genotypes IS30400 and IS15752) of developing sorghum grains were analyzed between 7 and 40 days after flowering (DAF) encompassing all four grain developing (or maturation) stages: milky (7-13 DAF), soft dough (~17 DAF), hard dough (21-29 DAF), and physiological maturity (33-40 DAF). The microscopic identification of protein bodies and the UHPLC-MS/MS detection of procyanidin dimers/trimers from the earliest stage of grain development (7 DAF) highlight significant limitations for using this grain in human nutrition, since protein-protein and protein-tanin complexes are the main factors contributing to sorghum's low protein digestibility (Duodu *et al.*, 2003). Notably, at the onset of the milky stage (7 DAF), the predominant amino acid group in the grain shifts from acidic to hydrophobic, a characteristic that persists until the grain reaches maturity. In fact, kafirins display a high hidrofobicity, also contributing to their low digestibility aspect (Duodu *et al.*, 2003). Globally, the initial stages (7-21 DAF) are characterized by minimal changes in sorghum grain. During this period, the grain is accumulating starch and proteins, its water content

remains stable (water plateau), and its phenolic profile is progressively synthesized predominantly composed of phenolic acids.

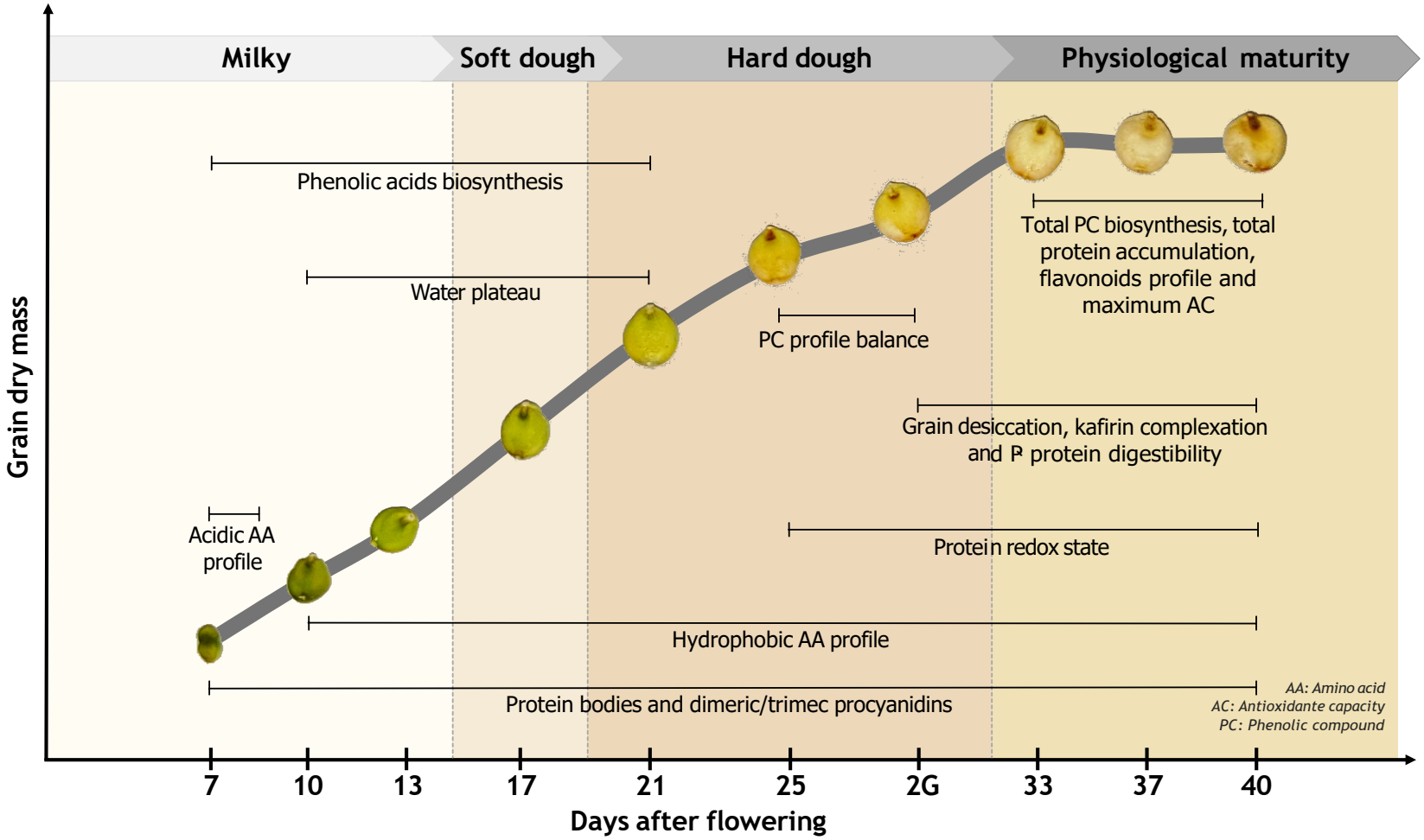


Figure 27. Sorghum developmental stages and key physiological and biochemical events observed during grain filling.

The desiccation of sorghum grain occurring at the final of hard dough stage (from 29 DAF) induces various stress responses leading to redox state modifications in the plant. The reduction in grain water content promotes the hydrophobic protein–protein interactions, as mentioned above. As well established in the literature and corroborated by the amino acid profile observed in the present study, orghum kafirins exhibit greater hydrophobicity compared to prolamins from other cereals (Duodu *et al.*, 2003). Simultaneously, protein complexation occurs via intermolecular disulfide bond formation among the cysteine residues of kafirins subunits, altering the redox state of sorghum proteins. These complexations hinder the complete action of proteases, thereby reducing the protein digestibility of sorghum.

Grain desiccation stage can also affects other enzymes. The 4-coumarate-CoA ligase is pivotal in the phenolic compounds biosynthesis in cereals, it protects the plant against oxidative damage and other stress-related factors (Sun *et al.*, 2013). Chen *et al.* (2019) demonstrated increased expression and activity of this enzyme in response to plant drought stress. Specifically, 4-coumarate-CoA ligase converts 4-coumaric acid and cinnamic acid into precursors for the flavonoid biosynthesis pathway and can be responsible for making flavonoids the predominant phenolic class in mature sorghum grains. Furthermore, the antioxidant capacity of sorghum grains peaks at physiological maturity. This phenomenon can be attributed to two factors: i) enzymatic activity increases throughout grain maturation, promoting the biosynthesis of compounds with antioxidant potential; and ii) at the beginning of desiccation phase, the expected generation of ROS are neutralized by phenolic compounds and other synthesized antioxidant molecules, and also by protein oxidation mechanisms, thereby protecting the grain from oxidative damage.

Elucidating the transformations occurring during sorghum grain growth is crucial for selecting earlier maturation stages with desirable characteristics for human consumption, such as enhanced digestibility. Additionally, understanding the accumulation dynamics of sorghum grain components, including kafirins and phenolic compounds, can inform the optimization of technological processing methods for the preparation of sorghum-based products.

2. Comparative phenolic profile of sorghum grains – from immature grains to post-processing methods (dehulling, cooking, and germination)

To obtain a comprehensive profile of phenolic compounds from initial stages to mature processed grains, an exploratory analysis was conducted. This analysis unified all compounds found by metabolomics approach presented in Chapters 2, 3, and 4. Overall, 318 phenolic compounds were tentatively identified, and their distribution across the samples is illustrated using a Venn Diagram (Figure 36).

The dataset for developing grains (Chapter 2, yellow circle) comprises 3 genotypes harvested at 10 different maturation stages and contained the highest number of unique compounds, assigned to the distinct phenolic profile of immature grains. As previously mentioned, immature grains exhibit a phenolic profile primarily composed of phenolic acids. Among the 65 unique compounds, the majority were esters of caffeic acid, ferulic acid, and p-coumaric acid with quinic acid. These findings contrast with those observed in developing wheat grains, where isomers of caffeoylquinic, feruloylquinic, and coumaroylquinic acids were predominantly found in mature grains (Ma *et al.*, 2022; Santos *et al.*, 2019b). In Chapters 3 and 4, the grains subjected to processing (dehulling, cooking, and germination) were obtained at harvest (i.e. mature stage). Germinated grains (Chapter 4, red circle) also exhibited unique compounds (27), with 70% being glycosylated flavonoids. During the metabolic activation phase of germination (phase II), flavonoids undergo glycosylation. This process transforms aglycones into stable, water-soluble, and non-reactive forms that can be stored in the plant's vacuole (Schendel, 2019).

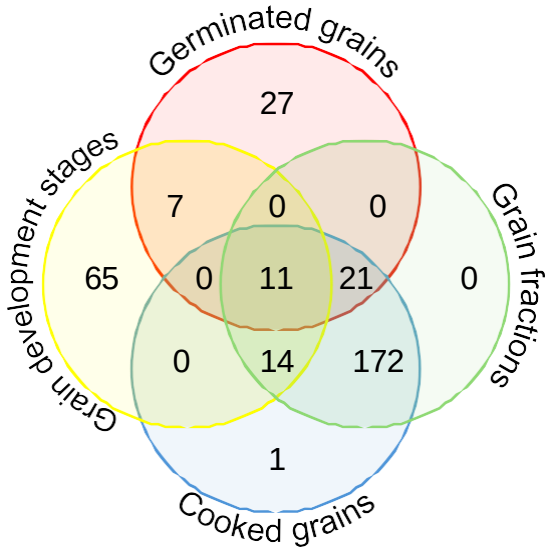


Figure 28. Comparison of the number of compounds identified across all sorghum datasets in this manuscript: from immature stages to mature grains post-processing.

However, the most part of compounds (172) were found at the intersection between fractionated (whole, decorticated at 15% and bran) and cooked grains (Chapter 3). The high similarity between these two datasets can be attributed to several factors: i) the mature grain samples used for both processes were from the same genotypes, cultivation location, and harvest; ii) prior to cooking, the samples were also decorticated to assess the thermal impact on each grain fraction; iii) the cooking parameters (time-temperature binomial) were not sufficient to alter the phenolic profile of the grain, only affecting the compound contents post-application of wet-heat.

Furthermore, eleven phenolic compounds were consistently found across all sorghum datasets. Four of these compounds were fully identified using phenolic standards (level ID 1), namely: 4-hydroxybenzoic, caffeic, isoferulic, and p-coumaric acids. These phenolic acids are well-documented in the literature as key compounds identified by LC-MS in the phenolic characterization of mature sorghum grains (D'Almeida *et al.*, 2021b; Ofosu *et al.*, 2021; Xiong *et al.*, 2020). It is worth mentioning that this thesis notably observes these compounds across a broad dataset, tracking their presence from the initial stages of grain development through their persistence after various processing methods. Additionally, four [M−H]- ions at *m/z* 577 were detected in all datasets and based on experimental fragmentation (level ID 2), these compounds were identified as procyanidin dimers. The presence of procyanidins, also known as condensed tannins, in all sorghum datasets reflects the sampling variability used throughout the thesis. Across all chapters, efforts were made to include grains with contrasting tannin levels (classified as tannin-rich and tannin-free/low tannin sorghums). In addition, it should be noted that this classification is often based on spectrophotometric techniques, thus emphasizing the effectiveness of metabolomic techniques to detect oligomers of tannins even in extractions more focused on polar compounds. This experimental design facilitated a more comprehensive understanding of phenolic variability among grains and the behavior of mature sorghum during processing.

Finally, based on all 318 metabolites and their relative ionic abundances, a hierarchical cluster analysis (HCA) was conducted (Figure 37), enabling the profiling of these specialized metabolites at pre- and post-harvest sorghum phases and elucidating the relationships between them. HCA revealed the formation of two distinct clusters: (1) whole Dutch sorghum grains (with HD100 and HD101 or without glume HD7, HD19) and isolated bran, both in their raw and cooked forms, and (2) immature, germinated, and

decorticated grains. This differentiation can be attributed to the breeding of Dusormil sorghum germplasm (Dutch sorghum), which has been optimized for high productivity during extended photoperiods in cool temperate environments.

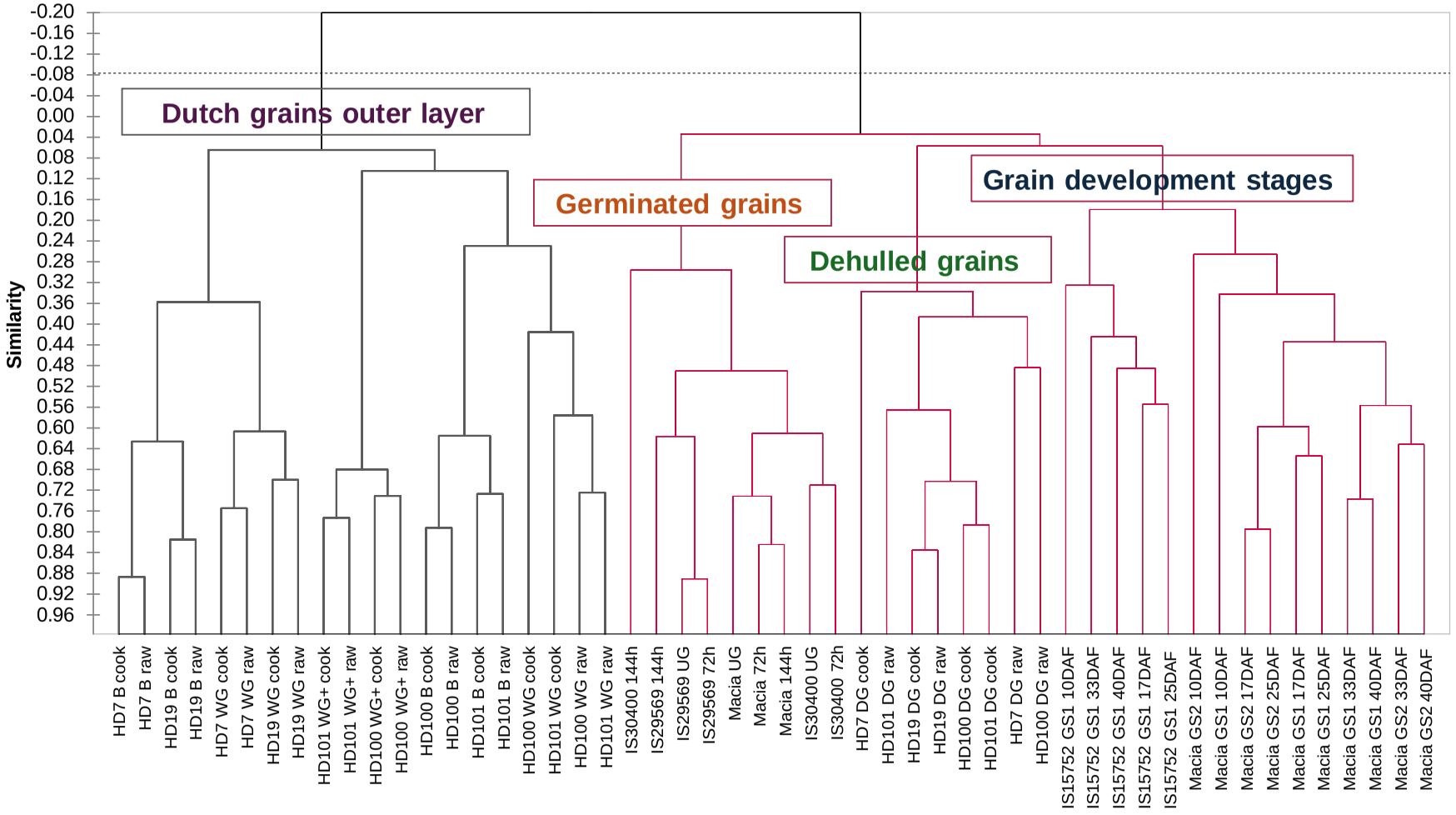


Figure 29. Dendrogram of cluster analysis based on total phenolic metabolomic profile in immature sorghum grains and post-processing methods.

Group 1 (Dutch sorghum grains) formed two subgroups: whole grains and bran from HD7 and HD19 (both raw and cooked) *vs* whole grains (with glume WG+ and without) and bran from HD100 and HD101. Similarly, in Group 2, within each subgroup (germinated, dehulled, and immature grains), greater dissimilarity was observed between contrasting genotypes than between maturation stages or processing methods. The clusters formed in Figure 3 support the hypothesis that sorghum phenolic profile variation is strongly influenced by the following grain parameters, in descending order of importance: race > genotype > maturation stage/processing methods > growing seasons.

3. Phenolic variability in sorghum grain – impact of race, genotype, and growing seasons

The ability of plants to synthesize phenolic compounds has evolved over time, enabling them to deal with various cultivation challenges. The intensity of phenolic compound synthesis is an inherited and genetically determined property (Nurzyńska-Wierdak, 2023). However, extrinsic factors related to cultivation conditions (agro-environmental factors, such as soil composition, climatic conditions (rainfall, temperature, dry periods, among others) and seasonal variations also influence this process (Eseberri *et al.*, 2022). To better understand this phenolic variability, this manuscript consistently employed contrasting grain genotypes, races, and growing seasons. This approach facilitated the elucidation of which factors have the most impact on the phenolic profile of sorghum.

According to the HCA (Figure 3), sorghum race has the greatest impact on phenolic variability. The subspecies *bicolor* includes the domesticated sorghum used for grain and can be divided based on floral morphology into five main interfertile races (Bicolor, Kafir, Caudatum, Durra, and Guinea) that can produce 10 intermediate races. Chapter 3 evaluated the phenolic profile of different sorghum races: *bicolor* (HD7 and HD19) and *nigricans* (HD100 and HD101). While *bicolor* is the most globally produced sorghum race, *Sorghum nigricans* has been characterized since 1967 as a typical Nigerian grain (Curtis, 1967). Visually, HD100 and HD101 are distinct from the *bicolor* due to their black pericarp color and the presence of glumes around their grains. Biochemical and metabolomic analyses revealed additional dissimilarities. *Sorghum nigricans* grains are harder and less dense, with higher ash content and lower protein digestibility. Regarding the phenolic profile, these grains showed greater variability and abundance of

compounds compared to sorghum HD7 and HD19. However, the most significant difference lies in the phenolic classes: while *Sorghum bicolor* has flavonoids as the predominant class, *Sorghum nigricans* has a balanced proportion (1:1) of flavonoids and phenolic acids.

Throughout Chapters 1, 2, and 4, genotypes with contrasting tannin levels were analyzed, specifically tannin-poor (T^-, Macia) and tannin-rich (T^+, IS 30400, IS 15752 and IS 29569) genotypes. Even during grain cultivation, a clear dissimilarity between genotypes was observed. T^- grains reached physiological maturity 33 DAF, at which point levels of insoluble kafirins became predominant over soluble ones, indicating protein complexation. In contrast, T^+ grains reached the maturity earlier, at just 25 DAF. From a bioactive perspective, T^+ genotypes presented higher levels of phenolic compounds and consequently greater antioxidant capacity compared to T^- sorghum. Chapter 2 identified several key compounds that differentiate between genotypes, including 4-hydroxybenzaldehyde, 4-hydroxybenzoic acid and its derivative (dihydroxybenzoic acid), and procyanidin dimer and its monomer (catechin). All these compounds were more abundant in the T^+ genotype. Chapter 4 further highlighted the differential impact of germination on each genotype: while T+ was not significantly affected, the T^- genotype experienced a drastic change in phenolic solubility, with reduced levels of free compounds and increased levels of bound compounds.

It is also important to highlight the similarities between the genotypes, which include: 1) both genotypes presented protein bodies from the initial stages of maturation; 2) there was no significant difference in the protein digestibility, indicating endogenous protein factors have a greater impact on the characteristics of sorghum kafirins; 3) despite differential abundance, procyanidins (dimers and trimers) were detected in both genotypes; 4) the presence of characteristic sorghum phenolic compounds, such as caffeic acid, ferulic acid and its isomer (isoferulic acid), and p-coumaric acid, was noted in both genotypes, and 5) the profile of phenolic classes remained similar between genotypes, with flavonoids being the predominant class in mature and phenolic acids in immature grains.

In a detailed evaluation of grains of the same race (*Sorghum bicolor*) and genotype (Macia), across different growing seasons (GS, 2017 and 2018), the impact of crop on phenolic compound variability was assessed (Chapter 2). The study found that crop year

had minimal influence on phenolic variability, with 67% of the annotated compounds remaining consistent regardless of growing season. Phenolic abundance was similar at the earliest (10 DAF) and physiological maturity stages (33 DAF) across both growing seasons. However, at the intermediate stages (17 and 25 DAF), there was a significant increase in abundance for GS2 samples, with levels from 83% to 400% higher, respectively, compared to GS1. Ferulic acid was also pointed as a discriminant compound presenting a differentially abundance and a greater ionic intensity in GS2.

4. Impact of mechanical, thermal and biological processing on the phenolic profile in sorghum grains

Food processing is considered a prerequisite for the production and/or consumption of sorghum-based food products, involving mechanical, thermal, and biological treatments. Previous scientific evidence has demonstrated that technological processing can enhance the shelf-life and sensory quality of sorghum-based foods, but also may alter the functional and nutritional properties in sorghum grains (Kasote *et al.*, 2021; Li *et al.*, 2022; Rashwan *et al.*, 2021). To assess the primary modifications in sorghum phenolic compounds, three processing methods were evaluated: dehulling (mechanical), cooking (thermal), and germination (biological) (Chapters 3 and 4).

An unsupervised multivariate analysis using principal component analysis (PCA) was conducted to provide an overview of all technological processes applied (Figure 38). The biplot PCA considered the relative ionic abundance of each putatively identified phenolic compound as variables (observations), and the scores for each sample were calculated. Figure 38 illustrates that across different treatments, PC1 and PC2 accounted together for ~60% of the total dataset variation. As expected, the mechanical process of separating the sorghum grain fractions (dehulling) had the most pronounced qualitative impact. In the PCA biplot, these samples exhibited a semi-moon distribution, with decorticated grains positioned at one extremity (superior left quadrant) and bran at the opposite (inferior right quadrant). The biological process (germination) clustered in the lower-left quadrant of the PCA biplot, showing greater proximity to the decorticated grains. Conversely, the thermal treatment (cooking) did not result in significant quantitative differentiation among its raw samples.

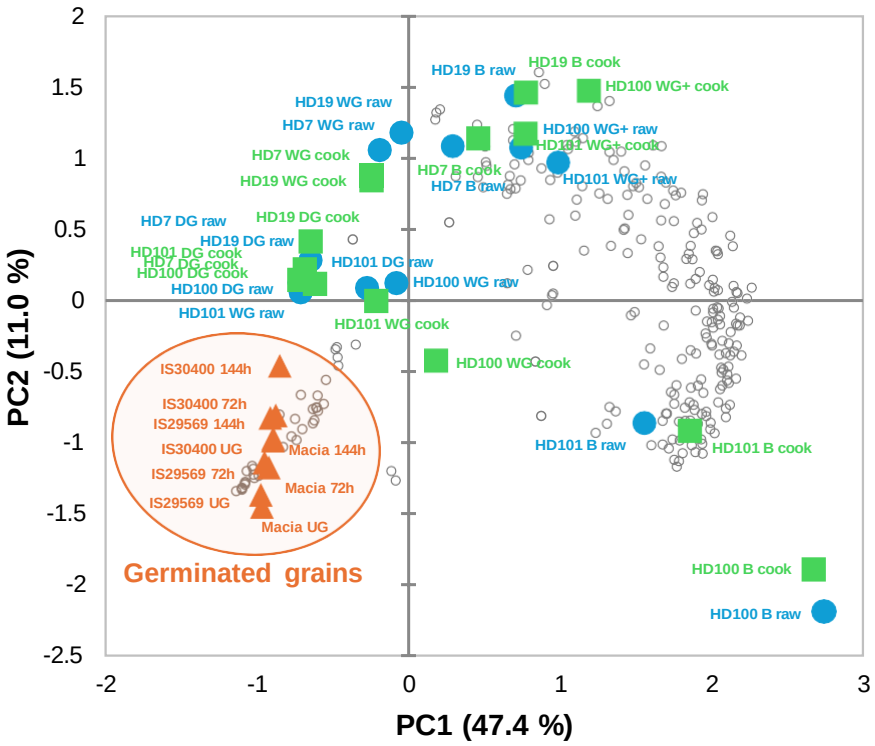


Figure 30. Principal Component Analysis (PCA) based on the phenolic profile post-processing methods. Germinated (orange triangle), dehulled (blue circles) and cooked (green squares) grains are distributed according to relative intensity of identified phenolic compounds (gray circles). UG = ungerminated grains; WG+ = wholegrains with glume; WG = wholegrains without glume; DG = dehulled grains; and B = bran.

To sum up, Figure 39 qualitatively illustrated the similarities and differences among the three processing methods. As expected, dehulling reduced both the number of identified compounds and their relative abundances, making it the only process that negatively affects phenolic content. Despite this reduction, dehulling is crucial for enhancing the organoleptic and nutritional properties of the grain (El Hanafi *et al.*, 2023). For instance, sorghum procyanidins, comprising nine molecular forms (dimers and trimers), were found differentially abundant (down-regulated) post-hulling. This reduction not only decreases the astringent flavor associated with tannins but may also improve protein digestibility (El Hanafi *et al.*, 2023). Furthermore, the total phenolic abundance in the decorticated grain and bran combined was higher than in the whole grain, suggesting that fractionation may enhance the extractability of phenolic compounds.

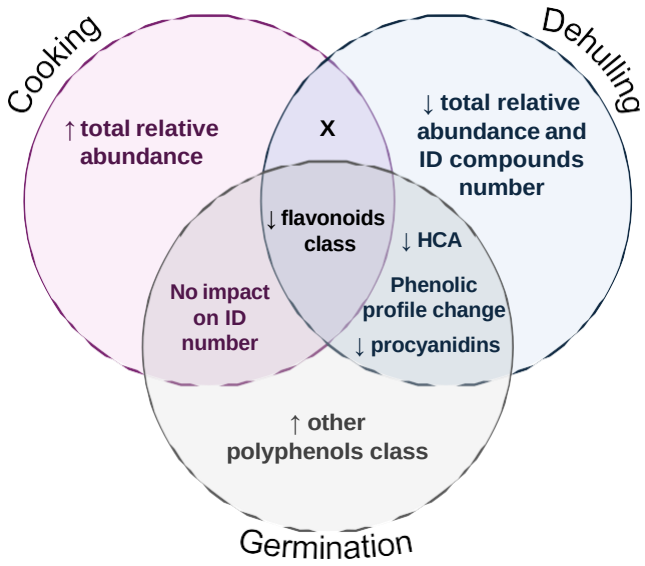


Figure 31. Main qualitative results of mature sorghum post-processing.

Dehulling showed notable similarities with germination, accounting for the close positioning of both methods in the PCA biplot (Figure 38). Germination also influenced procyanidin levels, indicating that this biological process promotes the depolymerization of these molecules into monomers. Furthermore, both mechanical (dehulling) and biological (germination) methods significantly impacted hydroxycinnamic acids (HCA), leading to a marked reduction in ferulic, caffeic, and p-coumaric acids. Phenolic profile was also impacted by grain processing. As previously mentioned throughout the manuscript, mature grains are primarily composed of flavonoids. Dehulling resulted in a 68% loss of flavonoids, making phenolic acids the predominant class. Germination also reduced flavonoid levels to the point of parity with phenolic acids, but a notable increase in other polyphenols made this the dominant class in germinated grains.

Germination also shared similarities with thermal treatment (cooking). Both methods did not significantly affect the total number of identified compounds: 83% and 95% of the compounds tentatively identified in raw grains persisted after biological and thermal treatment, respectively. However, cooking exhibited distinct characteristics: it increased the total abundance of phenolic compounds without altering the phenolic profile, with flavonoids remaining the majority class. This increase is due to heat-induced damage to the cellular structure of the grains, facilitating the release of bound phenolic compounds (Xiong *et al.*, 2019a).

The reduction in the total abundance of flavonoids was found in the three processing methods, as can be observed in the center of the Venn diagram (Figure 39). Flavonoids are the phenolic class most prone to oxidation and degradation during sorghum food processing, primarily due to their chemical structure, which contains multiple reactive hydroxyl groups (Ren *et al.* (2020). Chapter 5 demonstrated that the interaction between protein and an isolated flavonoid (catechin) exerts a protective effect on the phenolic compound against thermal processing. However, the synergistic interaction of multiple phenolics can limit their interaction with proteins, thereby reducing this protective effect. It is important to highlight that these results may vary depending on the genotype and the specific processing parameters applied.

5. Sorghum in human nutrition: investigating the relationship between protein digestibility and phenolic bioaccessibility

This manuscript has successfully characterized the phenolic and protein content/profiles in sorghum grains. However, the total amount of these phytochemicals and macronutrient detected does not correspond to the amount absorbed into the bloodstream. A significant portion of nutrients escapes during the gastrointestinal digestion due to intrinsic factors (*e.g.*, solubility and chemical structure) and extrinsic factors (*e.g.*, interaction with other grain components) (Duijsens *et al.*, 2021). To obtain insight into their post-gastrointestinal digestion and their potential beneficial effects on human health, we evaluated protein digestibility and phenolic bioaccessibility in sorghum grains by *in vitro* simulated digestion INFOGEST method. The main results are depicted in Figure 40.

Protein digestibility refers to the proportion of this polypeptide that is enzymatically hydrolyzed into smaller and absorbable units (amino acids). This is a key indicator of the grain's protein nutritional quality and biological value, reflecting the amount of protein available for body's metabolic needs (Duijsens *et al.*, 2021). Chapter 1 of the manuscript showed immature sorghum grains presented about 80% *in vitro* protein digestibility (IVPD). Following physiological maturation, this digestibility decreases, reaching values of approximately 60%. This reduction coincides with the onset of kafirin complexation via intermolecular disulfide bonds, which hide proteolytic cleavage sites and thereby limit proteases action.

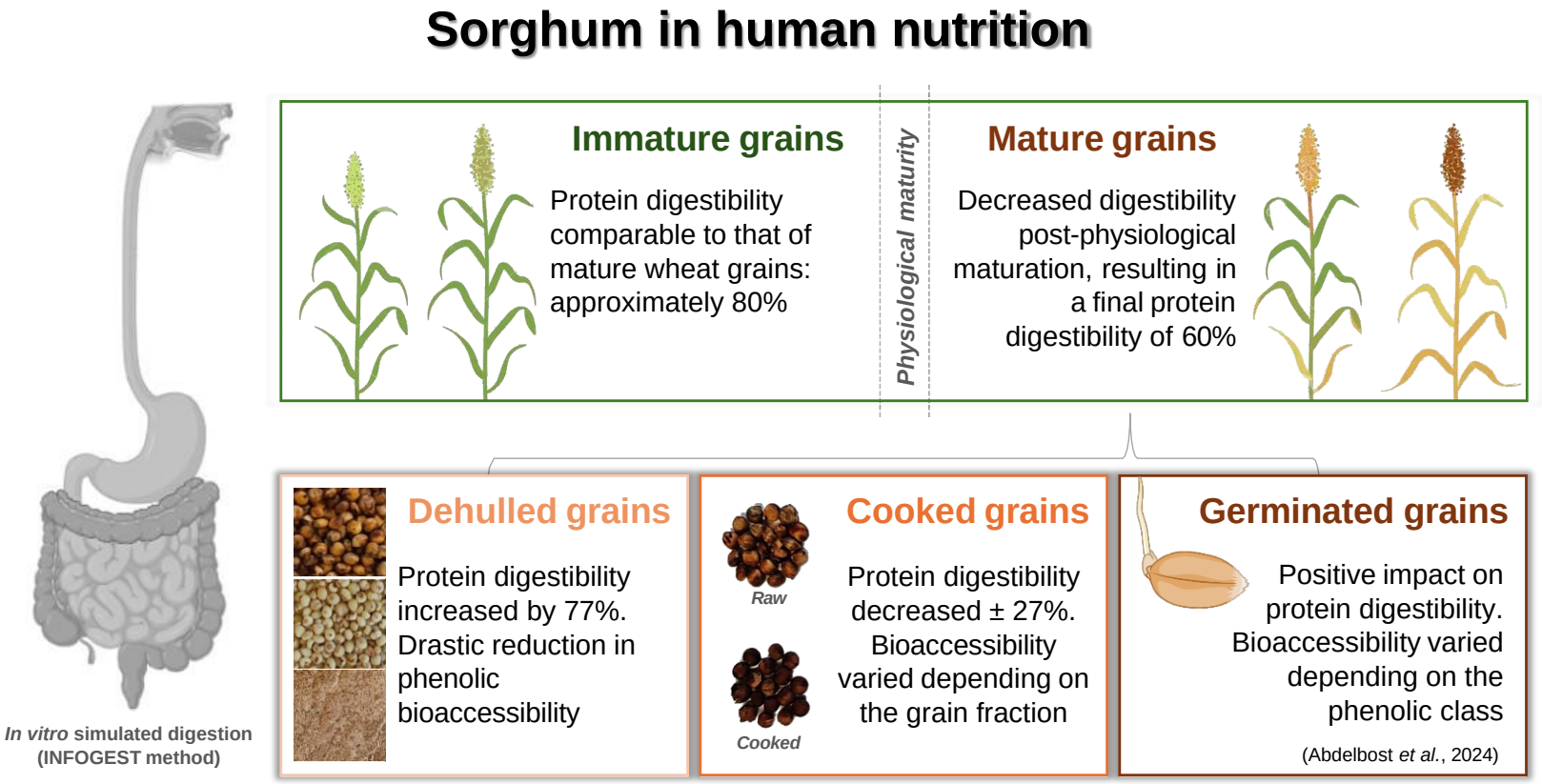


Figure 32. Key insights into sorghum in human health: protein digestibility and phenolic bioaccessibility.

In an effort to enhance protein availability in mature sorghum grains and to better understand its variation in sorghum-based products, IVPD was evaluated across three different processing methods. As expected, mechanical treatment (dehulling, Chapter 3) increased IVPD in sorghum grains by 77% due to the removal of the grain outer layers, which contains phenolic compounds and phytic acid. According to Duodu *et al.* (2003), these compounds represent negative extrinsic factors impairing protein digestibility. With a digestibility of approximately 83%, dehulled grains seems to be a viable option for human consumption.

On the other hand, but also expected thermal processing (cooking) negatively impacted the IVPD of sorghum grains (Duodu *et al.*, 2002). Overall, cooking whole grains reduced digestibility by 27%. Although the reduction was less severe in dehulled grains (only 17% decrease), cooking sorghum grains can not be considered a good strategy for enhancing protein digestibility for human consumption. Biological treatment (germination) impact was also assessed. Although Chapter 4 did not include IVPD analysis of germinated sorghum grains, Abdelbost *et al.* (2024) demonstrated the positive effects of germination on protein digestibility in genetically identical grains to those used in this thesis.

From a bioactive viewpoint, phenolic bioaccessibility varied depending on the applied processing method. Chapter 4 demonstrated the selective impact of germination on the bioaccessibility of specific phenolic classes. Flavonoids are primarily released during the gastrointestinal phase, with their bioaccessibility decreasing post-germination. Conversely, phenolic acids start to be released during the oral phase, with their bioaccessibility increasing after this biological treatment. Bioaccessibility was significantly reduced after dehulling due to the removal of the richest tissue in phenolic compounds (outer layers). Cooking presented an ambiguous effect: while it decreased bioaccessibility in whole grains, it increased bioaccessibility in decorticated grains. This observation is consistent with the protein digestibility data, suggesting that the protein-phenolic complex in sorghum grains not only inhibits protease activity but also limits the release of phenolic compounds in the gastrointestinal tract for absorption into the bloodstream. Chapter 5 demonstrated the significant interaction between these molecules.

To sum up, taking together the results of digestibility and bioaccessibility, germination is the most advantageous processing method for enhancing the health

benefits of sorghum grains. This traditional method is simple, easily reproducible; and preserves phenolic levels, decomplexes procyanidins (tannins), and improves protein digestibility. From an alternative perspective, encouraging the consumption of immature grains presents a novel approach. Although these grains have not yet completed their phenolic synthesis or reached their maximum antioxidant potential, they contain significant levels and a favorable composition of phenolic compounds, besides to show a good protein digestibility. Moreover, immature grains require less cultivation time, potentially reducing the overall planting cost. Nonetheless, careful monitoring of crop development and environmental conditions should be considered to determine the optimal harvest time for sorghum, such as as high humidity, or harvest losses due to shattering of grains.

6. Conclusions and perspectives

This study has advanced the understanding of sorghum grain development from the flowering stage through maturation and processing, elucidating critical physical and biochemical transformations. The robust phenolic profile and significant antioxidant potential of sorghum underscore its importance for human nutrition. Although the protein quality of mature grains presents challenges, innovative processing techniques and/or the consumption of grains at earlier developmental stages may enhance biological quality of sorghum for human consumption and geographical diffusion.

Looking ahead, sorghum holds substantial promise for enhancing global food security, particularly in regions vulnerable to climate change. The pressing challenge of rising global food demand *vs* climate change highlights a re-evaluation of underutilized crops such as sorghum. However, establishing sorghum as a major crop will require overcoming substantial challenges. Increased research efforts and governmental support are crucial for the widespread adoption of this resilient cereal.

Based on these findings, the following perspectives for future research are proposed:

Proteomics insights

The data presented in Chapter 1 provided novel insights into protein modifications throughout the development of sorghum grains. These findings revealed that agriculturally significant stages, such as desiccation and physiological maturation, are

associated with biochemical changes. Consequently, we were able to assess protein solubility and complexation throughout sorghum grain growth, as well as its digestibility at each developmental stage. To achieve a comprehensive understanding of these changes, we propose the application of proteomics approaches as the next step. A label-free proteomics approach applying Liquid Chromatography coupled to High Definition Mass Spectrometry (LC-HDMS) would be essential for mapping the dynamic expression profiles of proteins during sorghum seed development. This method would provide deep informations into the metabolic control of the corresponding protein functions.

Exploration of new processing methods and parameters

This thesis explored various food processing methods; however, further studies are needed to fully map phenolic and proteins changes in sorghum grains. There are several thermal (*e.g.*, extrusion), biological (*e.g.*, fermentation), and mechanical (*e.g.*, soaking, high pressure) methods that justify investigation. Such research could aggregate specific processing types to particular phenolic or protein changes, guiding the industry towards the optimal method based on the desired final product.

Moreover, two of the three methods utilized in this thesis involve the time-temperature binomial. It is well-established that phenolic compounds are sensitive to light, heat, and oxygen exposure (Li *et al.*, 2022). Consequently, the duration and intensity of the temperature applied during each processing method can differentially impact the phenolic compounds in sorghum. To achieve a comprehensive understanding of sorghum grain processing, it is essential to explore a range of temperatures and exposure times. Moreover, if immature grains are considered, there is a significant lack of research, and various parameters should be investigated.

The future of in vitro simulated digestion

In vitro INFOGEST method utilized in Chapters 3 and 4 of this thesis simulates gastrointestinal digestion using fluids and enzymes analogous to those present in the human body, allowing for an association with human ingestion of sorghum. Despite being well-described and relatively straightforward to use, the static method is not suitable for replicating the kinetics of human digestion (Wojtunik-Kulesza *et al.*, 2020). For example, during the gastric phase, pH and enzyme activity are maintained constant, which contrasts with the human body's gradual gastric addition and emptying throughout digestion. Another limitation is that the intestinal stage is treated as a single phase rather than

sequential duodenal, jejunal, and ileal phases. Given the relatively recent development of this protocol and its extensive application to cereals, we anticipate that future modifications and improvements will enhance its alignment with human digestive processes.

Furthermore, there is a significant gap regarding the optimal analytical method for assessing bioaccessibility. Initially designed to evaluate protein digestibility in foods of animal origin, the INFOGEST method is now widely applied to plant-based foods to analyse other nutrients and to estimate bioacessibility of phytochemicals such as phenolic compounds. However, the absence of a standardized analytical method complicates the achievement of simulated digestion results. This thesis employed metabolomic approaches for a comprehensive analysis of these compounds, but the presence of salts and denatured enzyme residues hindered the untargeted analysis of these metabolites. Developing a customized protocol for mass spectrometers, which indicates whether post-digestion extraction or cleaning steps are necessary to ensure compatibility of the post-digestion supernatant with the equipment.

Interaction between kafirins and phenolic compounds

To our knowledge, Chapter 5 is the first study to examine the interactions between a model protein (β-lactoglobulin) and the complex phenolic matrix of sorghum. The results were promising and innovative but highlight the need to explore the potential mechanisms of interaction between these phenolic compounds and the kafirins. Despite advances in kafirin extraction and purification methods, several limitations remain, including the low selectivity of extraction solvents and economic feasibility (Shah *et al.*, 2024). A comprehensive understanding of the extraction methodologies and their impact on the physicochemical properties of kafirin would provide the necessary knowledge to isolate this protein and evaluate its interactions with phenolics.

Publications and collaborations conducted during the thesis

1. 10.1016/j.foodres.2024.114854. **D'ALMEIDA, C. T. S.**; ABDELBOST, L.; MAMERI, H.; FERREIRA, M. S. L.. "Tracking the changes and bioacessibility of phenolic compounds of sorghum grains (*Sorghum bicolor* (L.) Moench) upon germination and seedling growth by UHPLC-QTOF-MS/MS". Food Research International, 2024.

2. 10.1016/j.foodres.2023.113739. PAES, L. T.; **D'ALMEIDA, C. T. S.**; DO CARMO, M. A. V.; CRUZ, L. S.; DE SOUZA, A. B.; VIANA, L. M.; MALTAROLLO, V. G.; MARTINO, H. S. D.; LIMA, G. D. A.; FERREIRA, M. S. L; AZEVEDO, L.; BARROS, F. A. R.. "Phenolic-rich extracts from toasted white and tannin sorghum flours have distinct profiles influencing their antioxidant, antiproliferative, anti-adhesive, anti-invasive, and antimalarial activities". Food Research International, 2023.

3. 10.3390/metabo13030386. MEDINA, T. S.; **D'ALMEIDA, C. T. S.**; DO NASCIMENTO, T. P.; DE ABREU, J. P.; DE SOUZA, V. R.; KALILI, D. C.; TEODORO, A. J.; CAMERON, L. C.; KOBLITZ, M. G.; FERREIRA, M. S. L.. "Food Service Kitchen Scraps as a Source of Bioactive Phytochemicals: Disposal Survey, Optimized Extraction, Metabolomic Screening and Chemometric Evaluation". Metabolites, 2023.

4. 10.1016/j.jcs.2023.103778. CAUDURO, T.; **D'ALMEIDA, C. T. S.**; BIDUSKI, B.; SANTOS, A.; SANTOS, M. C. B.; LIMA, L. R. S.; CAMERON, L. C.; BERTOLIN, T. E.; FERREIRA, M. S. L.; GUTKOSKI, L. C.. "Whole wheat flour replaced by sprouted wheat improves phenolic compounds profile, rheological and bread-making properties".

5. 10.1016/j.lwt.2021.112519. SANTOS, M. C. B.; LIMA, L. R. S.; **D'ALMEIDA, C. T. S.**; VICTORIO, V. C. M.; CAMERON, L. C.; BOURLIEU-LACANAL, C.; FERREIRA, M. S. L.. "Foodomics in wheat flour reveals phenolic profile of different genotypes and technological qualities". LWT - Food Science and Technology, 2022.

6. 10.1016/j.foodchem.2022.132515. NORONHA, M. C.; CARDOSO, R. R.; **D'ALMEIDA, C. T. S.**; CARMO, M. A. V.; AZEVEDO, L.; RIBEIRO JUNIOR, J. I.; ELLER, M. R.; CAMERON, L. C.; FERREIRA, M. S. L.; BARROS, F. A. R.. "Black tea kombucha: Physicochemical, microbiological and comprehensive phenolic profile changes during fermentation, and antimalarial activity". Food Chemistry, 2022.

7. 10.1016/j.foodres.2022.111635. SANTETTI, G. S.; LIMA, L. R. S.; BIDUSKI, B.; SANTOS, M. C. B.; **D'ALMEIDA, C. T. S.**; CAMERON, L. C.; GUTKOSKI, L. C.; FERREIRA, M. S. L.; AMBONI, R. D. M. C.. "Untargeted metabolomics analysis reveals improved phenolic profile in whole wheat bread with yerba mate and the effects of the bread-making process". Food Research International, 2022.

8. 10.1016/j.foodres.2021.110663. **D'ALMEIDA, C. T. S.**; MAMERI, H.; MENEZES, N. S.; DE CARVALHO, C. W. P.; QUEIROZ, V. A. V.; CAMERON, L.C.; MOREL, M. H.; TAKEITI, C. Y.; FERREIRA, M. S. L..

"Effect of extrusion and turmeric addition on phenolic compounds and kafirin properties in tannin and tannin-free sorghum". Food Research International, 2021.

9. 10.1016/j.lwt.2020.110136. MAIA, I. C.; **D'ALMEIDA, C. T. S.**; FREIRE, D. M. G.; CAVALCANTI, E. D. C.; CAMERON, L. C.; DIAS, J. F.; FERREIRA, M. S. L.. "Effect of solid-state fermentation over the release of phenolic compounds from brewer's spent grain revealed by UPLC-MS^{E}". LWT - Food Science and Technology, 2020.

10. *Submitted.* LACERDA, U. V.; DA COSTA, C. V. P.; CARDOSO, R. R.; **D'ALMEIDA, C. T. S.**; DO CARMO, M. A. V.; LIMA, A. S.; CRUZ, L. S.; DE SOUZA, A. B.; FERNANDES, P. O.; MALTAROLLO, V. G.; FERREIRA, M. S. L.; AZEVEDO, L.; ELLER, M. R.; BARROS, F. A. R.. "Green tea kombucha is rich in phenolic compounds and has antioxidant, antiproliferative, antibacterial and antimalarial activities".

Scientific communications

1. **D'ALMEIDA, C. T. S.**; MOREL, M.; TERRIER, N.; MILLIANO, W.; BETA, T.; MAMERI, H.; FERREIRA, M. S. L. Dynamic metabolomics reveals phenolic profiles in sorghum: from flowering to final product. In: 6º Congresso Brasileiro de Proteômica, 2024.

2. **D'ALMEIDA, C. T. S.**; MOREL, M.; TERRIER, N.; MILLIANO, W.; BETA, T.; MAMERI, H.; FERREIRA, M. S. L. From flowering to final product: a complex study of sorghum grains phenolic profile by metabolomic approach. In: 29º Congresso Brasileiro de Ciência e Tecnologia de Alimentos, 2024.

3. **D'ALMEIDA, C. T. S.**; MOREL, M.; BONICEL, J.; MILLIANO, W.; BETA, T.; FERREIRA, M. S. L.; MAMERI, H. Protein digestibility and phenolic compounds composition of sorghum grains grown in northern Europe are impacted by grain processing. In: VI Simpósio de Alimentos e Nutrição (SIAN), 2024.

4. FERREIRA, R. M.; ALVES, T. O.; **D'ALMEIDA, C. T. S.**; CARVALHO, M. G. O.; TAKEITI, C. Y.; FERREIRA, M. S. L. Efeito do processamento do trigo no perfil de compostos fenólicos determinados por UHPLC MS/MS. In: VI Simpósio de Alimentos e Nutrição (SIAN), 2024.

5. **D'ALMEIDA, C. T. S.**; MOREL, M.; BONICEL, J.; CAMERON, L. C.; MAMERI, H.; FERREIRA, M. S. L. Sorgo como uma alternativa sustentável para a alimentação humana: da síntese proteica à bioatividade. In: 8ª Jornada de Pós-Graduação, 2023.

6. **D'ALMEIDA, C. T.S.**; MOREL, M.; TERRIER, N.; CAMERON, L. C.; MAMERI, H.; FERREIRA, M. S. L. Evaluation of genotype and growing season impacts on the synthesis of phenolic compounds during sorghum (*Sorghum bicolor* (L.) Moench) grain development. In: Global Sorghum Conference - Sorghum in the 21st Century: Resiliency and Sustainability in the Face of Climate Change, 2023.

7. **D'ALMEIDA, C. T.S.**; MOREL, M.; BONICEL, J.; TERRIER, N;; FERREIRA, M. S. L.; MAMERI, H. Digestibility, biochemical changes and Kafirin polymerization analysis during sorghum grain development. In: Global Sorghum Conference - Sorghum in the 21st Century: Resiliency and Sustainability in the Face of Climate Change, 2023.

8. FERREIRA, R. M.; ALVES, T. O.; **D'ALMEIDA, C. T. S.**; CARVALHO, M. G. O.; TAKEITI, C. Y.; FERREIRA, M. S. L. Impacto do processamento na bioacessibilidade de compostos fenólicos em produtos derivados do trigo. In: 15° SLACAN - Simpósio Latino-Americano de Ciência de Alimentos e Nutrição, 2023.

9. **D'ALMEIDA, C. T. S.**; MOREL, M.; CAMERON, L. C.; MAMERI, H.; FERREIRA, M. S. L. Germination as a strategy to valorize sorghum grains (*Sorghum bicolor* (L.) Moench) in human food: phenolic compounds profile and bioaccessibility upon simulated digestion. In: 15° SLACAN - Simpósio Latino-Americano de Ciência de Alimentos e Nutrição, 2023.

10. FERREIRA, R. M.; **D'ALMEIDA, C. T. S.**; ALVES, T. O.; CARVALHO, M. G. O.; TAKEITI, C. Y.; FERREIRA, M. S. L. Efeito do processamento na bioacessibilidade de compostos fenólicos em produtos à base de trigo. Na: 22ª Jornada de Iniciação Científica, 2023.

11. **D'ALMEIDA, C. T. S.**; TERRIER, N.; POT, D.; MOREL, M.; CAMERON, L. C.; MAMERI, H.; FERREIRA, M. S. L. UHPLC-MS/MS-based metabolomics to assess genotype and crop effects in the synthesis of phenolic compounds during sorghum grain maturation. In: International Conference on Polyphenols, 2023.

12. **D'ALMEIDA, C.T.S.**; ABDELBOST, L.; MOREL, M. ; CAMERON, L. C. ; MAMERI, H. ; FERREIRA, M.SL. Metabolic fingerprinting applied to unravel the impact of germination on the phenolic profile of *Sorghum bicolor*. In: 3rd Iberoamerican Conference on Mass Spectrometry, 2022.

13. **D'ALMEIDA, C.T.S.**; TERRIER, N.; POT, D.; MOREL, M. ; CAMERON, L. C. ; MAMERI, H. ; FERREIRA, M. S. L. Dynamic changes in polyphenolic compounds and antioxidant activity in sorghum grains after germination. In: II CBCP Congresso Brasileiro de Tecnologia de Cereais e Panificação, 2022.

14. **D'ALMEIDA, C.T.S.**; MAMERI, H.; FERREIRA, M. S. L. A germinação como estratégia de valorização de grãos de sorgo na alimentação humana: perfil fenólico e potencial antioxidante. In: 7ª Jornada de Pós-Graduação da UNIRIO (JPG), 2022.

15. **D'ALMEIDA, C. T. S.**; SANTOS, M. C. B.; LIMA, L. R. S.; FERREIRA, M. S. L. Impacto do genótipo e da qualidade tecnológica no perfil fenólico de farinhas de trigo brasileiro: uma abordagem metabolômica. In: 7ª Jornada de Pós-Graduação da UNIRIO (JPG), 2022.

16. **D'ALMEIDA, C. T. S.**; MAMERI, H.; MOREL, M.; CAMERON, L. C.; FERREIRA, M. S. L. The synthesis of phenolic compounds during sorghum grain growth: chemometrics and metabolomic screening. In: Colloque grains, 2021.

17. **D'ALMEIDA, C. T. S.**; MAMERI, H.; DE CARVALHO, C. W. P.; QUEIROZ, V. A. V.; CAMERON, L. C.; MOREL, M. ; TAKEITI, C.Y.; FERREIRA, M. S.

L. Effect of extrusion and turmeric addition on phenolic compounds and kafirin properties in sorghum. In: Colloque grains, 2021.

18. MAIA, I. C.; **D'ALMEIDA, C. T. S.**; MEIRELES, M. C; FREIRE, D. M. G.; CAVALCANTI, E. D. C.; CAMERON, L. C.; DIAS, J. F.; FERREIRA, M. S. L. Avaliação metabolômica de compostos fenólicos após fermentação em estado sólido de diferentes bagaços de malte de cervejaria artesanal. In: Congresso Brasileiro de Ciência e Tecnologia de Alimentos, 2020.
19. **D'ALMEIDA, C. T. S.**; MAMERI, H.; MOREL, M.; CAMERON, L. C.; FERREIRA, M. S. L. The influence of crop on LC-MS-based phenolic profile of sorghum during grain growth. In: Congresso Brasileiro de Ciência e Tecnologia de Alimentos, 2020.
20. **D'ALMEIDA, C. T. S.**; MAMERI, H.; MOREL, M.; DE CARVALHO, C. W. P.; QUEIROZ, V. A. V.; CAMERON, L. C.; TAKEITI, C.Y.; FERREIRA, M. S. L. How does extrusion change the phenolic profile and impact protein digestibility in sorghum grains? In: CBCP Congresso Online Brasileiro de Tecnologia de Cereais e Panificação, 2020.

Participation in courses and events

1. VI Simpósio de Alimentos e Nutrição, 2024, Rio de Janeiro
2. Summer School 2023 - Seeds as a keystone for the transition to agroecology, 2023, Saclay Plant Sciences, SPS, França. (Carga horária: 29h).
3. First School of Integrated Metabolomics (FSIM), 2023, Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Brasil. (Carga horária: 50h).
4. 8ª Jornada de Pós-Graduação da UNIRIO, 2023, Rio de Janeiro.
5. Global Sorghum Conference - Sorghum in the 21st Century: Resiliency and Sustainability in the Face of Climate Change, 2023, Montpellier.
6. 15° SLACAN - Simpósio Latino Americano de Ciência de Alimentos e Nutrição, 2023, Campinas.
7. 7ª Jornada de Pós-Graduação da UNIRIO, 2022, Rio de Janeiro.
8. 3rd Iberoamerican Conference on Mass Spectrometry, 2022 , Rio de Janeiro.
9. II CBCP - Congresso Online Brasileiro de Tecnologia de Cereais e Panificação, 2022, online.
10. Colloque Graines, 2021, online.
11. Congresso Brasileiro de Ciência e Tecnologia de Alimentos (CBCTA), 2020, online.
12. I CBCP - Congresso Online Brasileiro de Tecnologia de Cereais e Panificação, 2020, online.

Awards

2024: Featured presentation at the Brazilian Metabolomics Summit for Young Researchers - 1st place. 6th BRProt Congress and Brazilian Metabolomics Summit.

2024: 2nd place for best oral presentation at the 6th Symposium on Food and Nutrition (SIAN).

2023: Three Minute Thesis competition - 1st place. Sorghum in the 21st Century - Global Sorghum Conference.

2023: Featured presentation at the 8th Postgraduate Conference, UNIRIO.

2020: Honorable Mention as the best paper in the thematic axis "Microbiological, chemical and biochemical aspects"., CBCP 2020 - Brazilian Online Congress of Cereal and Bakery Technology.

AACCI. (2010). *Approved Methods of Analysis* (Method 55-30.01. Particle Size Inde: for Wheat Hardness, Issue.

Abay, K. A., *et al*. (2023). The Russia-Ukraine war: Inplications for global and regional food security and potential policy responses. *Global Food Security*, *3C*, 100675. https://doi.org/https://doi.org/10.1016/j.gfs.2023.100675

Abdelbost, L., Bonicel, J., Morel, M.-H., C Maneri, H. (2024). Investigating sorghun protein solubility and in vitro digestibility during seed gernination. *Food chemistry*, *43S*, 138084. https://doi.org/10.1016/j.foodchen.2023.138084

Abdelbost, L., *et al*. (2023). Sorghun grain gernination as a route to inprove kafirin digestibility: Biochenical and label free proteonics insights. *Food chemistry*, *424*, 136407. https://doi.org/10.1016/j.foodchen.2023.136407

Abrankó, L., García-Reyes, J. F., C Molina-Díaz, A. (2011). In-source fragnentation and accurate nass analysis of nulticlass flavonoid conjugates by electrospray ionization tine-of-flight nass spectronetry. *Journal of Mass Spectrometry*, *4C*(5), 478-488. https://doi.org/10.1002/jns.114

Adiano, O. Q., Fawale, O. S., C Olawoye, B. (2018). Recent trends in the fornulation of gluten-free sorghun products. *Journal of Culinary Science & Technology*, *1C*(4), 311-325.

Afify, A. E.-M. M., El-Beltagi, H. S., Abd El-Salan, S. M., C Onran, A. A. (2011). Bioavailability of iron, zinc, phytate and phytase activity during soaking and gernination of white sorghun varieties. *Plos one*, *C*(10), e25512.

Afify, A. E.-M. M., El-Beltagi, H. S., Abd El-Salan, S. M., C Onran, A. A. (2012). Protein solubility, digestibility and fractionation after gernination of sorghun varieties. *Plos one*, *7*(2), e31154.

Al-Shabib, N. A., *et al*. (2020). Molecular interaction of tea catechin with bovine p-lactoglobulin: A spectroscopic and in silico studies. *Saudi Pharm J*, *3*, 238-245. https://doi.org/10.1016/j.jsps.2020.01.002

Alkay, Z., *et al*. (2022). The effect of flours of different innature cereal grains on sourdough and sourdough bread: Microbiological, rheological, te:tural and sugar profiles. *Journal of Food Processing and Preservation*, *4C*(11), e1707. https://doi.org/https://doi.org/10.1111/jfpp.1707

Allouche, A. R. (2011). Gabedit—A graphical user interface for conputational chenistry softwares. *Journal of Computational Chemistry*, *32*(1), 174-182. https://doi.org/10.1002/jcc.21600

Anoako, D., C Awika, J. M. (2016). Polyphenol interaction with food carbohydrates and consequences on availability of dietary glucose. *Current Opinion in Food Science*, *8*, 14-18. https://doi.org/https://doi.org/10.1016/j.cofs.2016.01.010

Ananda, G. K. S., *et al*. (2020). Wild Sorghun as a Pronising Resource for Crop Inprovenent [Review]. *Frontiers in plant science*, *11*. https://doi.org/10.338/fpls.2020.01108

AOAC. (2005a). *Association of official analytical chemists official methods of analysis of AOAC International*.

Official Methods of Analysis, (2005b).

Arbe:, P. M., *et al*. (2018). E:truded sorghun flour (Sorghun bicolor L.) nodulate adiposity and inflannation in high fat diet-induced obese rats. *Journal of Functional Foods*, *42*, 346-355. https://doi.org/https://doi.org/10.1016/j.jff.2018.01.010

Arouna, N., Gabriele, M., C Pucci, L. (2020). The Inpact of Gernination on Sorghun Nutraceutical Properties. *Foods*, *S*, 1218. https://doi.org/10.330/foods01218

Arts, M. J., *et al*. (2002). Interactions between flavonoids and proteins: effect on the total antio:idant capacity. *Journal of agricultural and food chemistry*, *50*(5), 1184-1187.

Awika, J. M. (2017). Sorghun: Its unique nutritional and health-pronoting attributes. In *Gluten-free ancient grains* (pp. 21-54). Elsevier.

Awika, J. M., McDonough, C. M., C Rooney, L. W. (2005). Decorticating Sorghun To Concentrate Healthy Phytochenicals. *Journal of agricultural and food chemistry*, *53*(16), 6230-6234. https://doi.org/10.1021/jf0510384

Awika, J. M., C Rooney, L. W. (2004). Sorghun phytochenicals and their potential inpact on hunan health. *Phytochemistry*, *C5*(), 11-1221. https://doi.org/10.1016/j.phytochen.2004.04.001

Awika, J. M., Yang, L., Browning, J. D., C Faraj, A. (200). Conparative antio:idant, antiproliferative and phase II enzyne inducing potential of sorghun (Sorghun bicolor) varieties. *LWT - Food*

Science and Technology, *42*(6), 1041-1046. https://doi.org/https://doi.org/10.1016/j.lwt.200.02.003

Bakari, H., *et al*. (2023). Sorghun (Sorghun bicolor L. Moench) and Its Main Parts (By-Products) as Pronising Sustainable Sources of Value-Added Ingredients. *Waste and Biomass Valorization*, *14*(4), 1023-1044. https://doi.org/10.1007/s1264 022-012-7

Baranzelli, J., *et al*. (2023). Grain Gernination Changes the Profile of Phenolic Conpounds and Benzo:azinoids in Wheat: A Study on Hard and Soft Cultivars. *Molecules*, *28*(2).

Bean, S. R., *et al*. (201). Structure and Conposition of the Sorghun Grain. In *Sorghum* (pp. 173-214). https://doi.org/https://doi.org/10.2134/agronnonogr58.c

Belton, P. S., Delgadillo, I., Halford, N. G., C Shewry, P. R. (2006). Kafirin structure and functionality. *Journal of Cereal Science*, *44*(3), 272-286. https://doi.org/https://doi.org/10.1016/j.jcs.2006.05.004

Bradwell, J., *et al*. (2018). Storage stability of sorghun phenolic e:tracts' flavones luteolin and apigenin. *LWT*, *S7*, 787-73. https://doi.org/10.1016/j.lwt.2018.08.006

Brás, N. F., *et al*. (2010). Inhibition of Pancreatic Elastase by Polyphenolic Conpounds. *Journal of agricultural and food chemistry*, *58*(1), 10668-10676. https://doi.org/10.1021/jf101734

Brodkorb, A., *et al*. (201). INFOGEST static in vitro sinulation of gastrointestinal food digestion. *Nature Protocols*, *14*(4), 1-1014. https://doi.org/10.1038/s4156-018-011-1

Buitinea-Cantúa, N. E., *et al*. (2013). Effect of defatting and decortication on distribution of fatty acids, phenolic and antio:idant conpounds in sorghun (S orghun bicolor) bran fractions. *International Journal of Food Science & Technology*, *48*(10), 2166-2175.

Cabrera-Ranírez, A. H., *et al*. (2020). Effect of the ni:tanalization process on the protein bioaccessibility of white and red sorghun flours during in vitro gastrointestinal digestion. *Food Research International*, *134*, 10234. https://doi.org/10.1016/j.foodres 2020.10234

Cardoso, L. d. M., Pinheiro, S. S., Martino, H. S. D., C Pinheiro-Sant'Ana, H. M. (2017). Sorghun (Sorghun bicolor L.): Nutrients, bioactive conpounds, and potential inpact on hunan health. *Critical Reviews in Food Science and Nutrition*, *57*(2), 372-30. https://doi.org/10.1080/1040838.2014.887057

Castro, C. B., *et al*. (2020). Metabolonics-Based Discovery of Bionarkers with Cytoto:ic Potential in E:tracts of Myracrodruon urundeuva. *Journal of the Brazilian Chemical Society*, *31*. https://doi.org/10.21577/0103-5053.2010242

Ceccaroni, D., *et al*. (2020). Effect of the tine and tenperature of gernination on the phenolic conpounds of Triticun aestivun, L. and Panicun niliaceun, L. *LWT*, *127*, 1036. https://doi.org/10.1016/j.lwt 2020.1036

Çetin-Babaoğlu, H., Arslan-Tontul, S., C Akın, N. (2020). Effect of innature wheat flour on nutritional and technological quality of sourdough bread. *Journal of Cereal Science*, *S4*, 103000. https://doi.org/https://doi.org/10.1016/j.jcs.2020.103000

Chen, G., Wang, J., C Li, Y. (2022a). E:tracts of sorghun bran, grape seed, and green tea: Chronatographic conparison of phenolic profiles and nitigation effect on acrylanide in antio:idant-fortified bread. *Food Chemistry Advances*, *1*, 100082. https://doi.org/10.1016/j.focha.2022.100082

Chen, X., *et al*. (2022b). Effect of whey protein isolate and phenolic copignents in the thernal stability of nulberry anthocyanin e:tract at an acidic pH. *Food chemistry*, *377*, 132005.

Chen, X., *et al*. (2021). Sorghun Phenolic Conpounds Are Associated with Cell Growth Inhibition through Cell Cycle Arrest and Apoptosis in Hunan Hepatocarcinona and Colorectal Adenocarcinona Cells. *Foods*, *5*, 3. https://doi.org/10.330/foods100503

Chen, X., *et al*. (201). Molecular cloning and functional analysis of 4-Counarate:CoA ligase 4(4CL-like 1)fron Fra:inus nandshurica and its role in abiotic stress tolerance and cell wall synthesis. *BMC Plant Biology*, *1S*(1), 231. https://doi.org/10.1186/s12870-01-1812-0

Chisté, R. C., C Mercadante, A. Z. (2012). Identification and quantification, by HPLC-DAD-MS/MS, of carotenoids and phenolic conpounds fron the Anazonian fruit Caryocar villosun. *J Agric Food Chem.*, *C0*(1520-5118). https://doi.org/10.1021/jf30104f

Cifuentes, A. (200). Food analysis and foodonics.

Cirkovic Velickovic, T. D., C Stanic-Vucinic, D. J. (2018). The Role of Dietary Phenolic Conpounds in Protein Digestion and Processing Technologies to Inprove Their Antinutritive Properties. *Comprehensive Reviews in Food Science and Food Safety*, *17*(1), 82-103. https://doi.org/10.1111/1541-4337.12320

Consultation, F. (2011). Dietary protein quality evaluation in hunan nutrition. *FAO Food Nutr. Pap*, *S2*, 1-66.

Correia, I., Nunes, A., Barros, A. S., C Delgadillo, I. (2010). Conparison of the effects induced by different processing nethods on sorghun proteins. *Journal of Cereal Science*, *51*(1), 146-151. https://doi.org/10.1016/j.jcs.200.11.005

Curtis, D. L. (167). The Races of Sorghun in Nigeria: their Distribution and Relative Inportance. *Experimental Agriculture*, *3*(4), 275-286. https://doi.org/10.1017/S00144770002216X

Czubinski, J., C Dwiecki, K. (2017). A review of nethods used for investigation of protein–phenolic conpound interactions. *International Journal of Food Science & Technology*, *52*(3), 573-585. https://doi.org/https://doi.org/10.1111/ijfs.1333

D'Alneida, C. T. d. S., *et al*. (2021a). Effect of e:trusion and turneric addition on phenolic conpounds and kafirin properties in tannin and tannin-free sorghun. *Food Research International*, *14S*, 110663. https://doi.org/10.1016/j.foodres.2021.110663

D'Alneida, C. T. S., *et al*. (2021b). Effect of e:trusion and turneric addition on phenolic conpounds and kafirin properties in tannin and tannin-free sorghun. *Food Research International*, *v. 14S*, 110663. https://doi.org/10.1016/j.foodres.2021.110663

D'Alneida, C. T. d. S., Abdelbost, L., Maneri, H., C Ferreira, M. S. L. (2024). Tracking the changes and bioacessibility of phenolic conpounds of sorghun grains (Sorghun bicolor (L.) Moench) upon gernination and seedling growth by UHPLC-QTOF-MS/MS. *Food Research International*, 114854. https://doi.org/https://doi.org/10.1016/j.foodres.2024.114854

Da Silva, L. S., Taylor, J., C Taylor, J. R. (2011). Transgenic sorghun with altered kafirin synthesis: kafirin solubility, polynerization, and protein digestion. *Journal of agricultural and food chemistry*, *5S*(17), 265-270.

Dai, Y., *et al*. (2022). Digestive stability and transport ability changes of p-lactoglobulin-catechin conple:es by M cell nodel in vitro. *Front Nutr.*, *S*. https://doi.org/10.338/fnut.2022.55135

Dawson, T. P., Perrynan, A. H., C Osborne, T. M. (2016). Modelling inpacts of clinate change on global food security. *Climatic Change*, *134*(3), 42-440. https://doi.org/10.1007/s10584-014-1277-y

De Mesa-Stonestreet, N. J., Alavi, S., C Bean, S. R. (2010). Sorghun Proteins: The Concentration, Isolation, Modification, and Food Applications of Kafirins. *Journal of Food Science*, *75*(5), R0-R104. https://doi.org/https://doi.org/10.1111/j.1750-3841.2010.01623.:

de Morais Cardoso, L., Pinheiro, S. S., Martino, H. S. D., C Pinheiro-Sant'Ana, H. M. (2017). Sorghun (Sorghun bicolor L.): Nutrients, bioactive conpounds, and potential inpact on hunan health. *Critical Reviews in Food Science and Nutrition*, *57*(2), 372-30. https://doi.org/10.1080/1040838.2014.887057

de Morais, F. P. R., *et al*. (2020a). Whey protein and phenolic conpound conple:ation: Effects on antio:idant capacity before and after in vitro digestion. *Food Research International*, *133*, 10104. https://doi.org/10.1016/j.foodres.2020.10104

de Morais, J. S., *et al*. (2020b). Antio:idant activity and bioaccessibility of phenolic conpounds in white, red, blue, purple, yellow and orange edible flowers through a sinulated intestinal barrier. *Food Research International*, *131*, 10046.

de Oliveira, L. d. L., *et al*. (2022). Physical, chenical, and antio:idant analysis of sorghun grain and flour fron five hybrids to deternine the drivers of liking of gluten-free sorghun breads. *LWT*, *153*, 112407. https://doi.org/10.1016/j.lwt.2021.112407

de Oliveira, L. L., C de Alencar Figueiredo, L. F. (2024). Sorghun phytonutrients and their health benefits: A systenatic review fron cell to clinical trials. *Journal of Food Science*, 1–25. https://doi.org/10.1111/1750-3841.17011

Ding, J., *et al*. (2018). Effects of controlled gernination on selected physicochenical and functional properties of whole-wheat flour and enhanced v-aninobutyric acid accunulation by ultrasonication. *Food chemistry*, *243*, 214-221. https://doi.org/https://doi.org/10.1016/j.foodchen.2017.0.128

Dou, Z., Chen, C., C Fu, X. (201). Bioaccessibility, antio:idant activity and nodulation effect on gut nicrobiota of bioactive conpounds fron Moringa oleifera Lan. leaves during digestion and fernentation in vitro. *Food & function*, *10*(8), 5070-507.

Dube, N. M., Xu, F., C Zhao, R. (2020). The efficacy of sorghun flour addition on dough rheological properties and bread quality: A short review. *Grain & Oil Science and Technology*, *3*(4), 164-171. https://doi.org/10.1016/j.gaost.2020.08.001

Duijsens, D., *et al*. (2021). How postharvest variables in the pulse value chain affect nutrient digestibility and bioaccessibility. *Comprehensive Reviews in Food Science and Food Safety*, *20*(5), 5067-506. https://doi.org/https://doi.org/10.1111/1541-4337.12826

Duodu, K. G., C Awika, J. M. (201). Chapter 8 - Phytochenical-Related Health-Pronoting Attributes of Sorghun and Millets. In J. R. N. Taylor C K. G. Duodu (Eds.), *Sorghum and Millets (Second Edition)* (pp. 225-258). AACC International Press. https://doi.org/10.1016/B78-0-12-811527-5.00008-3

Duodu, K. G., *et al*. (2002). Effect of Grain Structure and Cooking on Sorghun and Maize in vitro Protein Digestibility. *Journal of Cereal Science*, *35*(2), 161-174. https://doi.org/10.1006/jcrs.2001.0411

Duodu, K. G., Taylor, J. R. N., Belton, P. S., C Hanaker, B. R. (2003). Factors affecting sorghun protein digestibility. *Journal of Cereal Science*, *38*(2), 117-131. https://doi.org/10.1016/S0733-5210(03)00016-X

Dykes, L., C Rooney, L. W. (2006). Sorghun and nillet phenols and antio:idants. *Journal of Cereal Science*, *44*(3), 236-251. https://doi.org/10.1016/j.jcs.2006.06.007

Dykes, L., Rooney, L. W., Waniska, R. R., C Rooney, W. L. (2005). Phenolic conpounds and antio:idant activity of sorghun grains of varying genotypes. *J Agric Food Chem.*, *53*. https://doi.org/10.1021/jf05041è

Earp, C. F., Doherty, C. A., C Rooney, L. W. (183). Fluorescence nicroscopy of the pericarp, aleurone layer, and endospern cell walls of three sorghun cultivars. *Cereal Chemistry*, *C0*, 408-410.

Earp, C. F., McDonough, C. M., C Rooney, L. W. (2004). Microscopy of pericarp developnent in the caryopsis of Sorghun bicolor (L.) Moench. *Journal of Cereal Science*, *3S*(1), 21-27. https://doi.org/https://doi.org/10.1016/S0733-5210(03)00060-2

Ed Nignpense, B., Francis, N., Blanchard, C., C Santhakunar, A. (2021). Bioaccessibility and Bioactivity of Cereal Polyphenols: A Review. *Foods*, *7*(2304-8158), 155. https://doi.org/10.330/foods1007155

Eggun, B., C Beanes, R. (183). The nutritive value of seed proteins. In *Seed proteins: biochemistry*, *genetics*, *nutritive value* (pp. 4-531). Springer.

Egli, D. B., C TeKrony, D. M. (2008). Species differences in seed water status during seed naturation and gernination. *Seed Science Research*, *7*(1), 3-12. https://doi.org/10.1017/S060258500003305

El-Soud, W. A., *et al*. (2013). Ability of ellagic acid to alleviate osnotic stress on chickpea seedlings. *Plant Physiology and Biochemistry*, *71*, 173-183. https://doi.org/10.1016/j.plaphy.2013.07.007

El Hanafi, L., *et al*. (2023). Effects of Dehulling and Roasting on the Phytochenical Conposition and Biological Activities of Sesanun indicun L. Seeds. *Journal of Chemistry*, *2023*(1), 534315, Article 534315. https://doi.org/https://doi.org/10.1155/2023/534315

Eseberri, I., *et al*. (2022). Variability in the Beneficial Effects of Phenolic Conpounds: A Review. *Nutrients*, *14*. https://doi.org/10.330/nu140125

Espinosa-Ranírez, J., C Serna-Saldívar, S. O. (2016). Functionality and characterization of kafirin-rich protein e:tracts fron different whole and decorticated sorghun genotypes. *Journal of Cereal Science*, *70*, 57-65.

Espitia-Hernández, P., *et al*. (2022a). Sorghun (Sorghun bicolor L.) as a potential source of bioactive substances and their biological properties. *Critical Reviews in Food Science and Nutrition*, *C2*(8), 226-2280. https://doi.org/10.1080/1040838.2020.185238

Espitia-Hernández, P., *et al*. (2024). Assisted e:traction using green technologies and conventional solvent e:traction nethod: a conparison of the phenolic content and profile of sorghun and their antio:idant activity. *Journal of Food Measurement and Characterization*. https://doi.org/10.1007/s1164-024-02444-7

Espitia-Hernández, P., *et al*. (2022b). Solid-State Fernentation of Sorghun by Aspergillus oryzae and Aspergillus niger: Effects on Tannin Content, Phenolic Profile, and Antio:idant Activity. *Foods*, *11*(1). https://doi.org/10.330/foods1113121

Ezeogu, L. I., Duodu, K. G., Ennanbu:, M. N., C Taylor, J. R. N. (2008). Influence of Cooking Conditions on the Protein Matri: of Sorghun and Maize Endospern Flours. *Cereal Chemistry*, *85*(3), 37-402. https://doi.org/https://doi.org/10.104/CCHEM-85 3-037

FAO. (15). *Chemical composition and nutritive value* (orghun and Millets in Hunan Nutrition, Issue. http://www.fao.org/docrep/t0818e/T0818E0a.htn#Chapter%204%20-%20Chenical%20conposition%20and%20nutritive%20value

FAO. (2008). *An introduction to the basic concepts of food security food security information for action*. www.foodsec.org/docs/concepts_guide.pdf

FAO. (2023). *FAOSTAT: Suite of Food Security Indicators*. https://www.fao.org/faostat/en/#data/FS

FAO, *et al*. (2023). *The State of Food Security and Nutrition in the World 2023. Urbanization, agrifood systems transformation and healthy diets across the rural-urban continuum*.

Ferni, F., Alverina, C., C Fawzy, A. (2022). A Systenatic Review of Procyanidins: Updates on Current Bioactivities and Potential Benefits in Wound Healing. *02*, 718. https://doi.org/10.4711/ijnscrs/v2-i8-01

Ferreira, M. S., Sanson, M.-F., Bonicel, J., C Morel, M.-H. (2012a). Relationship between endospern cells redo: honeostasis and glutenin polyners assenbly in developing durun wheat grain. *Plant Physiology and Biochemistry*, *C1*, 36-45.

Ferreira, M. S. L., *et al*. (2012b). Physicochenical control of durun wheat grain filling and glutenin polyner assenbly under different tenperature regines. *Journal of Cereal Science*, *5C*(1), 58-66. https://doi.org/https://doi.org/10.1016/j.jcs.2011.11.001

Gallego-Jara, J., *et al*. (2021). Chapter 1 - Engineering of nicrobial cell factories for production of plant-based natural products. In V. Singh (Ed.), *Microbial Cell Factories Engineering for Production of Biomolecules* (pp. 381-32). Acadenic Press. https://doi.org/https://doi.org/10.1016/B78-0-12-821477-0.0001-2

Gallo, T. C. B., *et al*. (2024). Sorghun protein ingredients: Production, conpositional variability and enhancenent of aqueous dispersibility through honogenization. *Future Foods*, *S*, 100323. https://doi.org/https://doi.org/10.1016/j.fufo.2024.100323

Gao, Y., *et al*. (2024). Characterization of the active conponents and bioaccessibility of phenolics in differently colored fo:tail nillets. *Food chemistry*, *452*, 13355. https://doi.org/10.1016/j.foodchen.2024.13355

George, T. T., *et al*. (2022). Trends and progress in sorghun research over two decades, and inplications for global food security. *South African Journal of Botany*, *151*, 60-6. https://doi.org/https://doi.org/10.1016/j.sajb.2022.11.025

Gerik, T., Bean, B., C Vanderlip, R. (2003). Sorghun Growth and Developnent.

Ghinea, I. O., *et al*. (2021). HPLC-DAD Polyphenolic Profiling and Antio:idant Activities of Sorghun bicolor during Gernination. *Agronomy*, *11*(3). https://doi.org/10.330/agronony11030417

Giannini, A., *et al*. (2017). Clinate risk and food security in Mali: A historical perspective on adaptation. *Earth's Future*, *5*(2), 144-157. https://doi.org/10.1002/2016EF000404

Ginsburg, I., *et al*. (2012). Saliva increases the availability of lipophilic polyphenols as antio:idants and enhances their retention in the oral cavity. *Archives of Oral Biology*, *57*(10), 1327-1334. https://doi.org/10.1016/j.archoralbio.2012.04.01

Girard, A. L., C Awika, J. M. (2018). Sorghun polyphenols and other bioactive conponents as functional and health pronoting food ingredients. *Journal of Cereal Science*, *84*, 112-124. https://doi.org/https://doi.org/10.1016/j.jcs.2018.10.00

Girard, A. L., *et al*. (2018). Interaction nechanisns of condensed tannins (proanthocyanidins) with wheat gluten proteins. *Food chemistry*, *245*, 1154-1162. https://doi.org/https://doi.org/10.1016/j.foodchen.2017.11.054

Girard, A. L., *et al*. (2016). Effect of Condensed Tannin Profile on Wheat Flour Dough Rheology. *Journal of agricultural and food chemistry*, *C4*(3), 7348-7356. https://doi.org/10.1021/acs.jafc.6b02601

Guindo, D., *et al*. (2016). Pericarp thickness of sorghun whole grain is accurately predicted by NIRS and can affect the prediction of other grain quality paraneters. *Journal of Cereal Science*, *CS*, 218-227.

Guo, R., *et al*. (2017). Phenolic conpounds, antio:idant activity, antiproliferative activity and bioaccessibility of Sea buckthorn (Hippophaë rhannoides L.) berries as affected by in vitro digestion. *Food & function*, *8*(11), 422-4240.

Hagernan, A. E., *et al*. (18). High nolecular weight plant polyphenolics (tannins) as biological antio:idants. *Journal of agricultural and food chemistry*, *4C*(5), 1887-182.

Hanaker, B. R., Kirleis, A. W., Mertz, E. T., C A:tell, J. D. (186). Effect of cooking on the protein profiles and in vitro digestibility of sorghun and naize. *Journal of agricultural and food chemistry*, *34*(4), 647-64. https://doi.org/10.1021/jf00070a014

Hanany Djande, C. Y., *et al*. (2021). A Metabolonics Approach and Chenonetric Tools for Differentiation of Barley Cultivars and Bionarker Discovery. *Metabolites*, *11*(). https://ndpi-res.con/d_attachnent/netabolites/netabolites-11-00578/article_deploy/netabolites-11-00578-v2.pdf?version=1630484

Hanany Djande, C. Y., *et al*. (2020). Metabolonics: A Tool for Cultivar Phenotyping and Investigation of Grain Crops. *Agronomy*, *10*(6). https://ndpi-res.con/d_attachnent/agronony/agronony-10-00831/article_deploy/agronony-10-00831-v2.pdf?version=15337380

Herrero, M., *et al*. (2012). Foodonics: MS-based strategies in nodern food science and nutrition. *Mass spectrometry reviews*, *31*(1), 4-6.

Hithanani, G., C Srinivasan, K. (2014). Bioaccessibility of polyphenols fron wheat (Triticun aestivun), sorghun (Sorghun bicolor), green gran (Vigna radiata), and chickpea (Cicer arietinun) as influenced by donestic food processing. *Journal of agricultural and food chemistry*, *C2*(46), 11170-1117.

Hole, A. S., *et al*. (2012). Inproved bioavailability of dietary phenolic acids in whole grain barley and oat groat following fernentation with probiotic Lactobacillus acidophilus, Lactobacillus johnsonii, and Lactobacillus reuteri. *Journal of agricultural and food chemistry*, *C0*(25), 636-6375.

Holt, R., *et al*. (2002). Procyanidin diner B2 [epicatechin-(4p-8)-epicatechin] in hunan plasna after the consunption of a flavanol-rich cocoa–. *The American journal of clinical nutrition*, *7C*, 78-804. https://doi.org/10.103/ajcn/76.4.78

Hossain, M. S., *et al*. (2022). Sorghun: A prospective crop for clinatic vulnerability, food and nutritional security. *Journal of Agriculture and Food Research*, *8*, 100300. https://doi.org/10.1016/j.jafr.2022.100300

Itan, M., *et al*. (2020). Metabolic and physiological responses to progressive drought stress in bread wheat. *Scientific Reports*, *10*(1), 1718. https://doi.org/10.1038/s4158-020-74303-6

Jakobek, L. (2015). Interactions of polyphenols with carbohydrates, lipids and proteins. *Food chemistry*, *175*, 556-567. https://doi.org/10.1016/j.foodchen.2014.12.013

Janbunathan, R., Mertz, E., C A:tell, J. (175a). Fractionation of soluble proteins of high-lysine and nornal sorghun grain.

Janbunathan, R., Mertz, E. T., C A:tell, J. D. (175b). Fractionation of soluble proteins of high-lysine and nornal sorghun grain. *Cereal Chemistry*, *52*, 11-121.

Ji, M., *et al*. (2023). Methionine restriction-induced sulfur deficiency inpairs antitunour innunity partially through gut nicrobiota. *Nature Metabolism*, *5*(), 1526-1543. https://doi.org/10.1038/s42255-023-00854-3

Jia, J., Gao, X., Hao, M., C Tang, L. (2017). Conparison of binding interaction between p-lactoglobulin and three connon polyphenols using nulti-spectroscopy and nodeling nethods. *Food chemistry*, *228*, 143-151. https://doi.org/10.1016/j.foodchen.2017.01.131

Kang, J., *et al*. (2016). Identification and characterization of phenolic conpounds in hydronethanolic e:tracts of sorghun wholegrains by LC-ESI-MSn. *Food chemistry*, *211*, 215-226. https://doi.org/https://doi.org/10.1016/j.foodchen.2016.05.052

Kasote, D., *et al*. (2021). Food Processing Technologies to Develop Functional Foods With Enriched Bioactive Phenolic Conpounds in Cereals [Review]. *Frontiers in plant science*, *12*. https://doi.org/10.338/fpls.2021.771276

Khalifa, M., C Eltahir, E. A. B. (2023). Assessnent of global sorghun production, tolerance, and clinate risk [Original Research]. *Frontiers in Sustainable Food Systems*, *7*. https://doi.org/10.338/fsufs.2023.1184373

Khoddani, A., *et al*. (2023). Sorghun in foods: Functionality and potential in innovative products. *Critical Reviews in Food Science and Nutrition*, *C3*(), 1170-1186. https://doi.org/10.1080/1040838.2021.16073

Khoddani, A., Wilkes, M. A., C Roberts, T. H. (2013). Techniques for analysis of plant phenolic conpounds. *Molecules*, *18*(2), 2328-2375.

Kin, M. J., C Kin, S. S. (2017). Utilisation of innature wheat flour as an alternative flour with antio:idant activity and consuner perception on its baked product. *Food chemistry*, *232*, 237-244. https://doi.org/https://doi.org/10.1016/j.foodchen.2017.04.007

Kin, M. J., Yoon, W.-J., C Kin, S. S. (2016). Phytochenical Conpositions of Innature Wheat Bran, and Its Antio:idant Capacity, Cell Growth Inhibition, and Apoptosis Induction through Tunor Suppressor Gene. *Molecules*, *21*(10). https://ndpi-res.con/d_attachnent/nolecules/nolecules-21-0122/article_deploy/nolecules-21-0122.pdf?version=147473554

Kin, T. J., *et al*. (2022). Inproved quantification of catechin and epicatechin in red rice (Oryza sativa L.) using stable isotope dilution liquid chronatography-nass spectronetry. *Applied Biological Chemistry*, *C5*(1), 85. https://doi.org/10.1186/s13765-022-00754-2

Kontopidis, G., Holt, C., C Sawyer, L. (2004). Invited Review: p-Lactoglobulin: Binding Properties, Structure, and Function. *Journal of Dairy Science*, *87*(4), 785-76. https://doi.org/10.3168/jds.S0022-0302(04)73222-1

Kunar, S., *et al*. (2015). Rapid qualitative and quantitative analysis of bioactive conpounds fron Phyllanthus anarus using LC/MS/MS techniques. *Industrial Crops and Products*, *CS*. https://doi.org/10.1016/j.indcrop.2015.02.012

Labuschagne, M. T. (2018). A review of cereal grain proteonics and its potential for sorghun inprovenent. *Journal of Cereal Science*, *84*, 151-158. https://doi.org/https://doi.org/10.1016/j.jcs.2018.10.010

Lakowicz, J. R. (2006). Instrunentation for Fluorescence Spectroscopy. In *Principles of Fluorescence Spectroscopy* (pp. 27-61). Springer US. https://doi.org/10.1007/78-0-387-46312-4_2

Lazarides, M., Hacker, J., C Andrew, M. (11). Ta:onony, cytology and ecology of indigenous Australian sorghuns (Sorghun Moench: Andropogoneae: Poaceae). *Australian Systematic Botany*, *4*(4), 51-635. https://doi.org/https://doi.org/10.1071/SB1051

Lee, D. R., C Atkey, P. T. (184). Water loss fron the developing caryopsis of wheat (Triticun aestivun). *Canadian Journal of Botany*, *C2*(6), 131-1326. https://doi.org/10.113/b84-17

Lestienne, I., *et al*. (2007). Losses of nutrients and anti-nutritional factors during abrasive decortication of two pearl nillet cultivars (Pennisetun glaucun). *Food chemistry*, *100*(4), 1316-1323. https://doi.org/10.1016/j.foodchen.2005.11.027

Li, T., *et al*. (2018). Conparing the binding interaction between p-lactoglobulin and flavonoids with different structure by nulti-spectroscopy analysis and nolecular docking. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *201*, 17-206. https://doi.org/10.1016/j.saa.2018.05.011

Li, Z., Zhao, X., Zhang, X., C Liu, H. (2021). Bioactive Conpounds and Biological Activities of Sorghun Grains. *Foods*, *10*(11), 2868. https://www.ndpi.con/2304-8158/10/11/2868

Li, Z., Zhao, X., Zhang, X., C Liu, H. (2022). The Effects of Processing on Bioactive Conpounds and Biological Activities of Sorghun Grains. *Molecules*, *27*(10). https://doi.org/10.330/nolecules27103246

Liang, L., C Subirade, M. (2010). p-Lactoglobulin/folic acid conple:es: fornation, characterization, and biological inplication. *The Journal of Physical Chemistry B*, *114*(1), 6707-6712. https://doi.org/10.1021/jp10106r

Lina, L. R. d. S., *et al*. (2024). Overview of the Metabolite Conposition and Antio:idant Capacity of Seven Major and Minor Cereal Crops and Their Milling Fractions. *Journal of agricultural and food chemistry*.

Lina, L. R. S., *et al*. (2023). Onics data reveals the phenolic fingerprint of Brazilian whole wheat flours of different technological qualities. *Journal of food science and technology*, *C0*(2), 783-76. https://doi.org/10.1007/s1317-023-05665-8

Lin, P.-Y., C Lai, H.-M. (2011). Bioactive conpounds in rice during grain developnent. *Food chemistry*, *127*(1), 86-3. https://doi.org/https://doi.org/10.1016/j.foodchen.2010 12.02

Liu, C., *et al*. (2022). pH-dependent interaction nechanisns between p-lactoglobulin and EGCG: Insights fron nulti-spectroscopy and nolecular dynanics sinulation nethods. *Food Hydrocolloids*, *133*, 108022. https://doi.org/10.1016/j.foodhyd.2022.108022

Liu, F., Ma, C., McClenents, D. J., C Gao, Y. (2017). A conparative study of covalent and non-covalent interactions between zein and polyphenols in ethanol-water solution. *Food Hydrocolloids*, *C3*, 625-634. https://doi.org/https://doi.org/10.1016/j.foodhyd.2016.0.041

Lotze-Canpen, H. (2011). Clinate Change, Population Growth, and Crop Production: An Overview. In *Crop Adaptation to Climate Change* (pp. 1-11). https://doi.org/https://doi.org/10.1002/780470602.ch1

Lutz, M., *et al*. (201). Roles of Phenolic Conpounds in the Reduction of Risk Factors of Cardiovascular Diseases. *Molecules*, *24*(2). https://doi.org/10.330/nolecules24020366

Ma, D., *et al*. (2022). Dynanic Metabolonics and Transcriptonics Analyses for Characterization of Phenolic Conpounds and Their Biosynthetic Characteristics in Wheat Grain. *Front Nutr.*, *S*, 844337. https://doi.org/10.338/fnut.2022.844337

Ma, H., *et al*. (2024). Changes in free anino acid and protein polynerization in wheat caryopsis and endospern during filling after shading [Original Research]. *Frontiers in plant science*, *15*. https://doi.org/10.338/fpls.2024.134472

March, R. E., C Miao, X.-S. (2004). A fragnentation study of kaenpferol using electrospray quadrupole tine-of-flight nass spectronetry at high nass resolution. *International Journal of Mass Spectrometry*, *231*(2), 157-167. https://doi.org/10.1016/j.ijns.2003.10.008

Mariotti, M., Alanprese, C., Pagani, M. A., C Lucisano, M. (2006). Effect of puffing on ultrastructure and physical characteristics of cereal grains and flours. *Journal of Cereal Science*, *43*(1), 47-56. https://doi.org/10.1016/j.jcs.2005.06.007

MatÉ, J. I., C Krochta, J. M. (14). p-Lactoglobulin Separation fron Whey Protein Isolate on a Large Scale. *Journal of Food Science*, *5S*(5), 1111-1114. https://doi.org/https://doi.org/10.1111/j.1365-2621.14.tb08203.:

Melini, V., Melini, F., C Acquistucci, R. (2020). Phenolic Conpounds and Bioaccessibility Thereof in Functional Pasta. *Antioxidants*, *S*(4). https://doi.org/10.330/antio:040343

Mertz, E. T., *et al*. (184). Pepsin digestibility of proteins in sorghun and other najor cereals. *Proc Natl Acad Sci U S A*, *81*(0027-8424), 1-2. https://doi.org/10.1073/pnas.81.1.1

Mohaned Yunus, S. N., *et al*. (2021). Antio:idant and a-glucosidase inhibitory activities of eight neglected fruit e:tracts and UHPLC-MS/MS profile of the active e:tracts. *Food Sci Biotechnol*, *30*, 15-208. https://doi.org/10.1007/s10068-020-00856-:

Mohapatra, D., *et al*. (201). Effect of different processing conditions on pro:inate conposition, anti-o:idants, anti-nutrients and anino acid profile of grain sorghun. *Food chemistry*, *271*, 12-135. https://doi.org/10.1016/j.foodchen.2018 07.16

Mokrane, H., *et al*. (2010). Assessnent of Algerian sorghun protein quality [Sorghun bicolor (L.) Moench] using anino acid analysis and in vitro pepsin digestibility. *Food chemistry*, *121*(3), 71-723. https://doi.org/10.1016/j.foodchen.2010.01.020

Molotoks, A., Snith, P., C Dawson, T. P. (2021). Inpacts of land use, population, and clinate change on global food security. *Food and Energy Security*, *10*(1), e261. https://doi.org/https://doi.org/10.1002/fes3.261

Montini, L., *et al*. (2020). Matri:-assisted laser desorption/ionization-nass spectronetry inaging of netabolites during sorghun gernination. *Plant Physiology*, *183*(3), 25-42.

Morel, M.-H., Redl, A., C Guilbert, S. (2002). Mechanisn of Heat and Shear Mediated Aggregation of Wheat Gluten Protein upon Mi:ing. *Biomacromolecules*, *3*(3), 488-47. https://doi.org/10.1021/bn01563p

Motlhaodi, T., *et al*. (2018). Nutritional variation in sorghun [Sorghun bicolor (L.) Moench] accessions fron southern Africa revealed by protein and nineral conposition. *Journal of Cereal Science*, *83*, 123-12. https://doi.org/10.1016/j.jcs.2018.08.010

Muluneh, M. G. (2021). Inpact of clinate change on biodiversity and food security: a global perspective—a review article. *Agriculture & Food Security*, *10*(1), 36. https://doi.org/10.1186/s40066-021-00318-5

Nagar, E. E., *et al*. (2021). The inpact of chenical structure on polyphenol bioaccessibility, as a function of processing, cell wall naterial and pH: A nodel systen. *Journal of Food Engineering*, *28S*, 110304.

Neghliz, H., Cochard, H., Brunel, N., C Martre, P. (2016). Ear Rachis Xylen Occlusion and Associated Loss in Hydraulic Conductance Coincide with the End of Grain Filling for Wheat [Original Research]. *Frontiers in plant science*, *7*. https://doi.org/10.338/fpls 2016.0020

Neto, J. J. L., *et al*. (2017). Inpact of bioaccessibility and bioavailability of phenolic conpounds in biological systens upon the antio:idant activity of the ethanolic e:tract of Triplaris gardneriana seeds. *Biomedicine & Pharmacotherapy*, *88*, -1007.

Neupane, D., *et al*. (2022). Does Clinate Change Affect the Yield of the Top Three Cereals and Food Security in the World? *Earth*, *3*(1), 45-71. https://ndpi-res.con/d_attachnent/earth/earth-03-00004/article_deploy/earth-03-00004-v2.pdf?version=164318816

Nkhata, S. G., Ayua, E., Kanau, E. H., C Shingiro, J. B. (2018). Fernentation and gernination inprove nutritional value of cereals and legunes through activation of endogenous enzynes. *Food science & nutrition*, *C*(8), 2446-2458.

Nurzyńska-Wierdak, R. (2023). Phenolic Conpounds fron New Natural Sources-Plant Genotype and Ontogenetic Variation. *Molecules*, *28*(1420-304 (Electronic)). https://doi.org/10.330/nolecules28041731

Ofosu, F. K., *et al*. (2021). UHPLC-ESI-QTOF-MS/MS characterization, antio:idant and antidiabetic properties of sorghun grains. *Food chemistry*, *337*, 127788. https://doi.org/https://doi.org/10.1016/j.foodchen.2020.127788

Ono, E., *et al*. (2010). Functional differentiation of the glycosyltransferases that contribute to the chenical diversity of bioactive flavonol glycosides in grapevines (Vitis vinifera). *Plant Cell*, *8*, 2856-2871. https://doi.org/10.1105/tpc.110.074625

Oria, M. P., Hanaker, B. R., C Schull, J. M. (15). In vitro protein digestibility of developing and nature sorghun grain in relation to a-, p-, and v-kafirin disulfide crosslinking. *Journal of Cereal Science*, *22*(1), 85-3. https://doi.org/https://doi.org/10.1016/S0733-5210(05)80010-4

Ozdal, T., Capanoglu, E., C Altay, F. (2013). A review on protein–phenolic interactions and associated changes. *Food Research International*, *51*(2), 54-70. https://doi.org/10.1016/j.foodres.2013.02.00

Palacios, C. E., *et al*. (2021). Contents of tannins of cultivars of sorghun cultivated in Brazil, as determined by four quantification nethods. *Food chemistry*, *337*, 12770. https://doi.org/10.1016/j.foodchen.2020.12770

Paterson, A. H., *et al*. (200). The Sorghun bicolor genone and the diversification of grasses. *Nature*, *457*(722), 551-556. https://doi.org/10.1038/nature07723

Peng, H., C Shahidi, F. (2023). O:idation and degradation of (epi)gallocatechin gallate (EGCG/GCG) and (epi)catechin gallate (ECG/CG) in alkali solution. *Food chemistry*, *408*, 134815. https://doi.org/10.1016/j.foodchen.2022.134815

Pilco-Quesada, S., *et al*. (2020). Effects of gernination and kilning on the phenolic conpounds and nutritional properties of quinoa (Chenopodiun quinoa) and kiwicha (Anaranthus caudatus). *Journal of Cereal Science*, *S4*, 1026. https://doi.org/10.1016/j.jcs 2020.1026

Pontieri, P., *et al*. (2021). Conparison of content in phenolic conpounds and antio:idant capacity in grains of white, red, and black sorghun varieties grown in the Mediterranean area. *ACS Food Science & Technology*, *1*(6), 110-111.

Punia, H., *et al*. (2020). Proteone dynanics and transcriptone profiling in sorghun [Sorghun bicolor (L.) Moench] under salt stress. *3 Biotech*, *10*, 1-10.

Qie, X., *et al*. (2021). Interaction between p-lactoglobulin and chlorogenic acid and its effect on antio:idant activity and thernal stability. *Food Hydrocolloids*, *121*, 10705. https://doi.org/10.1016/j.foodhyd.2021.10705

Qie, X., *et al*. (2020). Analysis of p-lactoglobulin–epigallocatechin gallate interactions: The antio:idant capacity and effects of polyphenols under different heating conditions in polyphenolic–protein interactions. *Food & function*, *11*(5), 3867-3878.

Qin, X., *et al*. (2018). Maillard-reacted whey protein isolates enhance thernal stability of anthocyanins over a wide pH range. *Journal of agricultural and food chemistry*, *CC*(36), 556-564. https://doi.org/10.1021/acs.jafc.8b03125

Ragaee, S., Abdel-Aal, E.-S. M., C Noanan, M. (2006). Antio:idant activity and nutrient conposition of selected cereals for food use. *Food chemistry*, *S8*(1), 32-38. https://doi.org/10.1016/j.foodchen.2005.04.03

Ranathunga, R. A. A., C Suwannaporn, P. (2022). Young cereal grains as a new source of healthy and hypoallergenic foods: a review. *Journal of food science and technology*, *5S*(), 3336-3348. https://doi.org/10.1007/s1317-021-05228-

Rao, S., *et al*. (2018). Characterization of phenolic conpounds and antio:idant activity in sorghun grains. *Journal of Cereal Science*, *84*, 103-111. https://doi.org/10.1016/j.jcs.2018.07.013

Rasera, G. B., de Canargo, A. C., C de Castro, R. J. S. (2023). Bioaccessibility of phenolic conpounds using the standardized INFOGEST protocol: A narrative review. *Comprehensive Reviews in*

Food Science and Food Safety, *22*(1), 260-286. https://doi.org/https://doi.org/10.1111/1541-4337.13065

Rasera, G. B., Hilkner, M. H., C de Castro, R. J. S. (2020). Free and insoluble-bound phenolics: How does the variation of these conpounds affect the antio:idant properties of nustard grains during gernination? *Food Research International*, *133*, 10115. https://doi.org/10.1016/j.foodres 2020.10115

Rashwan, A. K., *et al*. (2021). Potential processing technologies for developing sorghun-based food products: An update and conprehensive review. *Trends in Food Science & Technology*, *110*, 168-182. https://doi.org/10.1016/j.tifs.2021.01.087

Ray, D. K., *et al*. (2012). Recent patterns of crop yield growth and stagnation. *Nature Communications*, *3*(1), 123. https://doi.org/10.1038/nconns226

Razgonova, M. P., *et al*. (2021). Phytochenical Analysis of Phenolics, Sterols, and Terpenes in Colored Wheat Grains by Liquid Chronatography with Tanden Mass Spectronetry. *Molecules*, *2C*(18). https://ndpi-res.con/d_attachnent/nolecules/nolecules-26-05580/article_deploy/nolecules-26-05580-v2.pdf?version=163168805

Ren, F., Nian, Y., C Perussello, C. A. (2020). Effect of storage, food processing and novel e:traction technologies on onions flavonoid content: A review. *Food Research International*, *132*, 10853. https://doi.org/https://doi.org/10.1016/j.foodres.201.10853

Rhodes, D. H., *et al*. (2014). Genone-Wide Association Study of Grain Polyphenol Concentrations in Global Sorghun [Sorghun bicolor (L.) Moench] Gernplasn. *Journal of agricultural and food chemistry*, *C2*(45), 1016-1027. https://doi.org/10.1021/jf503651t

Rocchetti, G., Giuberti, G., C Lucini, L. (2018). Gluten-free cereal-based food products: the potential of netabolonics to investigate changes in phenolics profile and their in vitro bioaccessibility. *Current Opinion in Food Science*, *22*, 1-8. https://doi.org/10.1016/j.cofs.2017.10.007

Rocchetti, G., *et al*. (2022). Functional inplications of bound phenolic conpounds and phenolics–food interaction: A review [https://doi.org/10.1111/1541-4337.1221]. *Comprehensive Reviews in Food Science and Food Safety*, *21*(2), 811-842. https://doi.org/https://doi.org/10.1111/1541-4337.1221

Rodrigues, R. M., Ferreira-Santos, P., Pereira, R. N., C Tei:eira, J. A. (2024). Accessing the thernal and electric effects in p-lactoglobulin denaturation and interaction with phenolic conpounds. *Food Hydrocolloids*, *150*, 10724. https://doi.org/10.1016/j.foodhyd 2023.10724

Rodriguez, M. D. A., *et al*. (2023). Does the protein structure of p-lactoglobulin inpact its conple: coacervation with type A gelatin and the ability of the conple:es to entrap lutein? *Food Hydrocolloids*, *140*, 108651. https://doi.org/10.1016/j.foodhyd.2023.108651

Roozeboon, K., Prasad, P. V. V., C Cianpitti, I. (2016). Sorghun Growth and Developnent. In. https://doi.org/10.2134/agronnonogr58.2014.0062

Roozeboon, K. L., C Prasad, P. V. V. (201). Sorghun Growth and Developnent. In *Sorghum* (pp. 155-172). https://doi.org/https://doi.org/10.2134/agronnonogr58.c8

Rue, E., Rush, M., C van Breenen, R. (2018). Procyanidins: a conprehensive review enconpassing structure elucidation via nass spectronetry. *Phytochemistry Reviews*, *17*. https://doi.org/10.1007/s11101 017-507-3

Sahihi, M., C Ghayeb, Y. (2014). An investigation of nolecular dynanics sinulation and nolecular docking: Interaction of citrus flavonoids and bovine p-lactoglobulin in focus. *Computers in Biology and Medicine*, *51*, 44-50. https://doi.org/10.1016/j.conpbioned.2014.04.022

Saithalavi, K. M., Bhasin, A., C Yaqoob, M. (2021). Inpact of sprouting on physicochenical and nutritional properties of sorghun: a review. *Journal of Food Measurement and Characterization*, *v. 15*(5), 410-4204. https://doi.org/10.1007/s1164-021-006-

Salazar-López, N. J., González-Aguilar, G., Rouzaud-Sández, O., C Robles-Sánchez, M. (2018). Technologies applied to sorghun (Sorghun bicolor L. Moench): changes in phenolic conpounds and antio:idant capacity. *Food Science and Technology*, *38*. https://doi.org/10.150/fst.16017

Šanec, D., *et al*. (2021). The Role of Polyphenols in Abiotic Stress Response: The Influence of Molecular Structure. *Plants (Basel)*, *10*. https://doi.org/10.330/plants10010118

Sánchez-Velázquez, O. A., *et al*. (2021). In vitro gastrointestinal digestion inpact on stability, bioaccessibility and antio:idant activity of polyphenols fron wild and connercial blackberries (Rubus spp.). *Food & function*, *12*(16), 7358-7378.

Santos, M. C. B., *et al*. (201à). Metabolonic approach for characterization of phenolic conpounds in different wheat genotypes during grain developnent. *Food Research International*. https://doi.org/https://doi.org/10.1016/j.foodres.2018.08.034

Santos, M. C. B., *et al*. (201b̈). Metabolonic approach for characterization of phenolic conpounds in different wheat genotypes during grain developnent. *Food Research International*, *124*, 118-128. https://doi.org/10.1016/j.foodres.2018.08.034

Schendel, R. R. (201ÿ). 10 - Phenol content in sprouted grains. In H. Feng, B. Nenzer, C J. W. DeVries (Eds.), *Sprouted Grains* (pp. 247-315). AACC International Press. https://doi.org/https://doi.org/10.1016/B78-0-12-811525-1.00010-5

Schrödinger, L. (2015). The PyMOL Molecular Graphics Systen, Version 1.8.

Serna-Saldivar, S. O., C Espinosa-Ranírez, J. (201ÿ). Chapter 5 - Grain Structure and Grain Chenical Conposition. In J. R. N. Taylor C K. G. Duodu (Eds.), *Sorghum and Millets (Second Edition)* (pp. 85-12ÿ). AACC International Press. https://doi.org/10.1016/B78-0-12-811527-5.00005-8

Shah, U., *et al*. (2024). Advances in E:traction, Structure, and Physiochenical Properties of Sorghun Kafirin for Bionaterial Applications: A Review. *Journal of Functional Biomaterials*, *15*(7). https://ndpi-res.con/d_attachnent/jfb/jfb-15-00172/article_deploy/jfb-15-00172-v2.pdf?version=17110617

Shahidi, F., Danielski, R., C Ikeda, C. (2021). Phenolic conpounds in cereal grains and effects of processing on their conposition and bioactivities: A review. *Journal of Food Bioactives*, *15*. https://doi.org/10.31665/JFB.2021.15281

Shan, Z., *et al*. (2021). [Retracted] Theaflavin Chenistry and Its Health Benefits. *Oxidative Medicine and Cellular Longevity*, *2021*, 6256618. https://doi.org/10.1155/2021/6256618

Shao, Y., *et al*. (2014). Phenolic acids, anthocyanins, and antio:idant capacity in rice (Oryza sativa L.) grains at four stages of developnent after flowering. *Food chemistry*, *143*, 0-6. https://doi.org/10.1016/j.foodchen.2013.07.042

Sharna, A., *et al*. (201ÿ). Response of Phenylpropanoid Pathway and the Role of Polyphenols in Plants under Abiotic Stress. *Molecules*, *24*(1420-304˙), 2452. https://doi.org/10.330/nolecules24132452

Shelenbe, J. S., *et al*. (2014). Effect of acidic condition on phenolic conposition and antio:idant potential of aqueous e:tracts fron sorghun (Sorghun bicolor) bran. *Journal of Food Biochemistry*, *38*(1), 110-118.

Shen, S., *et al*. (2018). Phenolic Conpositions and Antio:idant Activities Differ Significantly anong Sorghun Grains with Different Applications. *Molecules*, *23*(1420-304ÿ), 1203. https://doi.org/10.330/nolecules23051203. PMID: 2772811

Shull, J. M., Chandrashekar, A., Kirleis, A. W., C Ejeta, G. (10). Developnent of Sorghun (Sorghun bicolor (L.) Moench) Endospern in Varieties of Varying Hardness. *Food Structure*, *S*. https://digitalconnons.usu.edu/foodnicrostructure/vol/iss3/8

Shull, J. M., Watterson, J. J., C Kirleis, A. (12). Purification and innunocytochenical localization of kafirins in Sorghun bicolor (L. Moench) endospern. *Protoplasma*, *171*, 64-74.

Shull, J. M., Watterson, J. J., C Kirleis, A. W. (11). Proposed nonenclature for the alcohol-soluble proteins (kafirins) of Sorghun bicolor (L. Moench) based on nolecular weight, solubility, and structure. *Journal of agricultural and food chemistry*, *3S*(1), 83-87.

Silva-Fernández, S. E., García-Lara, S., C Serna-Saldívar, S. O. (2022). Physicochenical characterization of the anatonical structures of teosinte (Zea nays subsp. ne:icana) covered caryopses. *Journal of Cereal Science*, *103*, 103353. https://doi.org/10.1016/j.jcs.2021.103353

Singh, A., Sharna, S., Singh, B., C Kaur, G. (201ÿ). In vitro nutrient digestibility and antio:idative properties of flour prepared fron sorghun gerninated at different conditions. *Journal of food science and technology*, *5C*(6), 3077-308˙ https://doi.org/10.1007/s1317-01-03804-8

Singleton, V. L., Orthofer, R., C Lanuela-Raventós, R. M. (1ÿ). Analysis of total phenols and other o:idation substrates and antio:idants by neans of folin-ciocalteu reagent. In *Methods in Enzymology* (Vol. v. 2,˙ pp. 152-178). Acadenic Press. https://doi.org/https://doi.org/10.1016/S0076-687()017-1

Sinosaki, N. B. M., *et al*. (2020). Structural Study of Phenolic Acids by Triple Quadrupole Mass Spectronetry with Electrospray Ionization in Negative Mode and H/D Isotopic E:change.

Journal of the Brazilian Chemical Society, *31*. https://doi.org/10.21577/0103-5053.201017

Sonpong, R., Siebenhandl-Ehn, S., Linsberger-Martin, G., C Berghofer, E. (2011). Physicochenical and antio:idative properties of red and black rice varieties fron Thailand, China and Sri Lanka. *Food chemistry*, *124*(1), 132-140. https://doi.org/https://doi.org/10.1016/j.foodchen.2010.05.115

Sruthi, N. U., Rao, P. S., C Rao, B. D. (2021). Decortication induced changes in the physico-chenical, anti-nutrient, and functional properties of sorghun. *Journal of Food Composition and Analysis*, *102*, 104031. https://doi.org/10.1016/j.jfca.2021.104031

Stefoska-Needhan, A., Beck, E. J., Johnson, S. K., C Tapsell, L. C. (2015). Sorghun: An Underutilized Cereal Whole Grain with the Potential to Assist in the Prevention of Chronic Disease. *Food Reviews International*, *31*(4), 401-437. https://doi.org/10.1080/875512.2015.1022832

Strauss, G., C Gibson, S. M. (2004). Plant phenolics as cross-linkers of gelatin gels and gelatin-based coacervates for use as food ingredients. *Food Hydrocolloids*, *18*(1), 81-8. https://doi.org/https://doi.org/10.1016/S0268-005X(03)00045-6

Sunner, L. W., *et al*. (2007). Proposed nininun reporting standards for chenical analysis Chenical Analysis Working Group (CAWG) Metabolonics Standards Initiative (MSI). *Metabolomics.*, *3*(1573-3882). https://doi.org/10.1007/s11306-007-0082-2

Sun, H., *et al*. (2013). Analysis of five rice 4-counarate:coenzyne A ligase enzyne activity and stress response for potential roles in lignin and flavonoid biosynthesis in rice. *Biochemical and Biophysical Research Communications*, *430*(3), 1151-1156. https://doi.org/https://doi.org/10.1016/j.bbrc.2012.12.01

Tanni, R., Männistö, S., Maukonen, M., C Kaartinen, N. E. (2024). Whole grain intake, diet quality and risk factors of chronic diseases: results fron a population-based study in Finnish adults. *European Journal of Nutrition*, *C3*(2), 37-408.

Taylor, J., Bean, S. R., Ioerger, B. P., C Taylor, J. R. N. (2007). Preferential binding of sorghun tannins with v-kafirin and the influence of tannin binding on kafirin digestibility and biodegradation. *Journal of Cereal Science*, *4C*(1), 22-31. https://doi.org/https://doi.org/10.1016/j.jcs.2006.11.001

Taylor, J. R. N., C Duodu, K. G. (2015). Effects of processing sorghun and nillets on their phenolic phytochenicals and the inplications of this to the health-enhancing properties of sorghun and nillet food and beverage products. *Journal of the Science of Food and Agriculture*, *S5*(2), 225-237. https://doi.org/10.1002/jsfa.6713

Taylor, J. R. N., C Duodu, K. G. (2023). Resistant-Type Starch in Sorghun Foods—Factors Involved and Health Inplications. *Starch - Stärke*, *75*(-10), 210026. https://doi.org/https://doi.org/10.1002/star.20210026

Tian, L., *et al*. (2021). The Bioavailability, E:traction, Biosynthesis and Distribution of Natural Dihydrochalcone: Phloridzin. *International Journal of Molecular Sciences*, *22*(2). https://ndpi-res.con/d_attachnent/ijns/ijns-22-0062/article_deploy/ijns-22-0062.pdf?version=1611064481

Tian, Y., *et al*. (2012). High nolecular weight persinnon tannin is a potent antio:idant both e: vivo and in vivo. *Food Research International*, *45*(1), 26-30. https://doi.org/https://doi.org/10.1016/j.foodres.2011.10.005

Tiozon, R. J. N., *et al*. (2022). Metabolonics based inferences to unravel phenolic conpound diversity in cereals and its inplications for hunan gut health. *Trends in Food Science & Technology*, *127*, 14-25. https://doi.org/10.1016/j.tifs.2022.06.011

Toné-Sánchez, I., *et al*. (2021). Bioprocessed wheat ingredients: Characterization, bioaccessibility of phenolic conpounds, and bioactivity during in vitro digestion. *Frontiers in plant science*, *12*, 7088.

Trott, O., C Olson, A. J. (2010). AutoDock Vina: Inproving the speed and accuracy of docking with a new scoring function, efficient optinization, and nultithreading. *Journal of Computational Chemistry*, *31*(2), 455-461. https://doi.org/10.1002/jcc.21334

UN-DESA. (2022). *World Population Prospects 2022: Summary of Results*. .

USDA. (2024a). *Crop Production*. https://fas.usda.gov/data/production

USDA. (2024b). *Crop Production Historical Track Records - April 2024* (US Departnent of Agriculture; National Agricultural Statistics Service, Issue.

Van Hung, P. (2016). Phenolic Conpounds of Cereals and Their Antio:idant Capacity. *Crit Rev Food Sci Nutr.*, *5C*, 25-35. https://doi.org/10.1080/1040838.2012.7080

Venkateswaran, K., Elangovan, M., C Sivaraj, N. (201). Chapter 2 - Origin, Donestication and Diffusion of Sorghun bicolor. In C. Aruna, K. B. R. S. Visarada, B. V. Bhat, C V. A. Tonapi (Eds.), *Breeding Sorghum for Diverse End Uses* (pp. 15-31). Woodhead Publishing. https://doi.org/https://doi.org/10.1016/B78-0-08-10187-8.00002-4

Vidotto, D. C., C Tavares, G. M. (2022). Sinultaneous binding of folic acid and lutein to p-lactoglobulin and a-lactalbunin: A spectroscopic and nolecular docking study. *Food Bioscience*, *50*, 102107. https://doi.org/https://doi.org/10.1016/j.fbio.2022.102107

Vilarinho, A., dos SANTOS, F., C RODRIGUES, J. (2007). Reconendacão da Cultivar de Sorgo Granífero BRS 310 para Cultivo no Cerrado de Roraina.

Wang, R., Zhou, W., C Jiang, X. (2008). Reaction Kinetics of Degradation and Epinerization of Epigallocatechin Gallate (EGCG) in Aqueous Systen over a Wide Tenperature Range. *Journal of agricultural and food chemistry*, *5C*(8), 264-2701. https://doi.org/10.1021/jf0730338

Waniska, R. D., C Rooney, L. W. (2000). Structure and chenistry of the sorghun caryopsis. *Sorghum: Origin, history, technology, and production*, *2*, 64-67.

Wojtunik-Kulesza, K., *et al*. (2020). Influence of In Vitro Digestion on Conposition, Bioaccessibility and Antio:idant Activity of Food Polyphenols—A Non-Systenatic Review. *Nutrients*, *12*(5). https://doi.org/10.330/nu12051401

Wu, G., *et al*. (2017). Phenolic profile and content of sorghun grains under different irrigation nanagenents. *Food Research International*, *S7*, 347-355.

Wu, Y., *et al*. (2012). Presence of tannins in sorghun grains is conditioned by different natural alleles of Tannin1. *Proceedings of the National Academy of Sciences*, *10S*(26), 10281-10286.

Wu, Y., *et al*. (2013). Mutation in the seed storage protein kafirin creates a high-value food trait in sorghun. *Nature Communications*, *4*(1), 2217.

Xiong, Y., *et al*. (201à). Effect of processing on the phenolic contents, antio:idant activity and volatile conpounds of sorghun grain tea. *Journal of Cereal Science*, *85*, 6-14. https://doi.org/https://doi.org/10.1016/j.jcs.2018.10.012

Xiong, Y., Zhang, P., Warner, R. D., C Fang, Z. (201b̀). Sorghun Grain: Fron Genotype, Nutrition, and Phenolic Profile to Its Health Benefits and Food Applications [https://doi.org/10.1111/1541-4337.12506]. *Comprehensive Reviews in Food Science and Food Safety*, *18*(6), 2025-2046. https://doi.org/https://doi.org/10.1111/1541-4337.12506

Xiong, Y., *et al*. (2020). Conprehensive profiling of phenolic conpounds by HPLC-DAD-ESI-QTOF-MS/MS to reveal their location and forn of presence in different sorghun grain genotypes. *Food Research International*, *137*, 10671. https://doi.org/10.1016/j.foodres.2020.10671

Xu, J., Wang, W., C Zhao, Y. (2021). Phenolic Conpounds in Whole Grain Sorghun and Their Health Benefits. *Foods*, *10*. https://doi.org/10.330/foods1008121

Xu, M., Rao, J., C Chen, B. (2020). Phenolic conpounds in gerninated cereal and pulse seeds: Classification, transfornation, and netabolic process. *Crit Rev Food Sci Nutr.*, *v. C0*, 740-75. https://doi.org/10.1080/1040838.2018.1550051

Yang, X., *et al*. (2024). An investigation into the effects of various processing nethods on the characteristic conpounds of highland barley using a widely targeted netabolonics approach. *Food Research International*, *180*, 114061. https://doi.org/10.1016/j.foodres.2024.114061

Zarei, M., *et al*. (2022). Sorghun as a Potential Valuable Aquafeed Ingredient: Nutritional Quality and Digestibility. *Agriculture*, *12*(5). https://ndpi-res.con/d_attachnent/agriculture/agriculture-12-0066/article_deploy/agriculture-12-0066-v2.pdf?version=1652088850

Zhang, L., *et al*. (2023). GWAS of grain color and tannin content in Chinese sorghun based on whole-genone sequencing. *Theoretical and Applied Genetics*, *13C*(4), 77. https://doi.org/10.1007/s00122-023-04307-z

Zhang, Q., Cheng, Z., Wang, Y., C Fu, L. (2021). Dietary protein-phenolic interactions: characterization, biochenical-physiological consequences, and potential food applications. *Crit Rev Food Sci Nutr.*, *C1*, 358-3615. https://doi.org/10.1080/1040838.2020.18031

Zhang, S., *et al*. (2020). Changes in the phenolic conpound content and antio:idant activity in developnental naize kernels and e:pression profiles of phenolic biosynthesis-related genes. *Journal of Cereal Science*, *SC*, 103113. https://doi.org/https://doi.org/10.1016/j.jcs.2020.103113

Zhang, S., *et al*. (2022). Binding of p-lactoglobulin to three phenolics inproves the stability of phenolics studied by nultispectral analysis and nolecular nodeling. *Food Chemistry: X*, *15*, 10036. https://doi.org/10.1016/j.foch:.2022.10036

Zhang, Y.-J., *et al*. (2015). Antio:idant phytochenicals for the prevention and treatnent of chronic diseases. *Molecules*, *20*(12), 21138-21156.

Zhao, W., *et al*. (2018). Rapid Screening and Identification of Daidzein Metabolites in Rats Based on UHPLC-LTQ-Orbitrap Mass Spectronetry Coupled with Data-Mining Technologies. *Molecules*, *23*(1). https://doi.org/10.330/nolecules23010151