



**Universidade Federal de Santa Catarina
Departamento de Microbiologia, Imunologia e Parasitologia.
Programa de Pós-Graduação em Biotecnologia e Biociências**

Rodolfo Moresco

**BIOPROSPECÇÃO, METABOLÔMICA E CARACTERIZAÇÃO
AMILÁCEA E DE CAROTENOIDES EM GERMOPLASMA
BIOFORTIFICADO DE MANDIOCA (*Manihot esculenta* Crantz),
VIA FERRAMENTAS DE BIOINFORMÁTICA**

Florianópolis
2017

Rodolfo Moresco

**BIOPROSPECÇÃO, METABOLÔMICA E CARACTERIZAÇÃO
AMILÁCEA E DE CAROTENOIDES EM GERMOPLASMA
BIOFORTIFICADO DE MANDIOCA (*Manihot esculenta* Crantz)
VIA FERRAMENTAS DE BIOINFORMÁTICA.**

Tese submetida ao Programa de Pós-
Graduação em Biotecnologia e
Biociências do Centro de Ciências
Biológicas da Universidade Federal de
Santa Catarina para a obtenção do
Grau de Doutor em Biotecnologia e
Biociências.

Orientador: Prof. Dr. Marcelo
Maraschin

Coorientador: Prof. Dr. Miguel
Francisco de Almeida Pereira Rocha

Florianópolis
2017

Ficha de identificação da obra elaborada pelo autor através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Moresco, Rodolfo

Bioprospecção, metabolômica e caracterização amilácea e de carotenoides em germoplasma biofortificado de mandioca / Rodolfo Moresco ; orientador, Marcelo Maraschin, coorientador, Miguel Francisco de Almeida Pereira Rocha, 2017.
144 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Ciências Biológicas, Programa de Pós-Graduação em Biotecnologia e Biociências, Florianópolis, 2017.

Inclui referências.

1. Biotecnologia e Biociências. 2. Mandioca (Manihot esculenta Crantz). 3. Amido. 4. Carotenoides. 5. Bioinformática. I. Maraschin, Marcelo. II. Francisco de Almeida Pereira Rocha, Miguel. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Biotecnologia e Biociências. IV. Título.

Rodolfo Moresco

**BIOPROSPECÇÃO, METABOLÔMICA E CARACTERIZAÇÃO
AMILÁCEA E DE CAROTENOIDES EM GERMOPLASMA
BIOFORTIFICADO DE MANDIOCA (*Manihot esculenta* Crantz)
VIA FERRAMENTAS DE BIOINFORMÁTICA.**

Esta tese foi julgada adequada para obtenção do Título de “Doutor em Biotecnologia e Biociências” e aprovada em sua forma final pelo Programa de Pós-Graduação em Biotecnologia e Biociências

Florianópolis, 30 de novembro de 2017.

Prof. Mário Steindel, Dr.
Coordenador do Curso de Pós-Graduação em Biotecnologia e
Biociências

Banca Examinadora:

Prof. Marcelo Maraschin, Dr.
Orientador
Universidade Federal de Santa Catarina

Prof. Miguel Francisco de Almeida Pereira Rocha, Dr.
Coorientador
Universidade do Minho – Braga - Portugal

Prof. Leonardo Rörig, Dr.
Universidade Federal de Santa Catarina

Prof. Rosendo Augusto Yunes, Dr.
Universidade do Vale do Itajaí - UNIVALI

Prof.^a Edna Regina Amante, Dra.
Universidade Federal de Santa Catarina

Este trabalho é dedicado aos meus pais Tarcísio Moresco e Ana Maria Alves Moresco. Por todo amor, carinho, compreensão e estímulo.

AGRADECIMENTOS

A Deus por ter me dado força e coragem para nunca desistir, por me proteger, por ter iluminado meus pensamentos para que eu sempre tomasse a melhor decisão e por ter colocado em meu caminho pessoas tão admiráveis que me ajudaram de muitas formas até a conclusão desse trabalho.

Aos meus pais, Tarcísio Moresco e Ana Maria Alves Moresco, por garantirem sempre o melhor para mim, pelo apoio e estímulo aos estudos, pelo imensurável amor.

Ao meu Professor Orientador Dr. Marcelo Maraschin, ao qual devo muito da minha formação profissional e pessoal. Um exemplo de ser humano e mestre, inspiração que levarei por toda minha vida profissional.

Ao meu coorientador Dr. Miguel Rocha da Universidade do Minho pela convivência harmoniosa de sempre, pela oportunidade de realização do doutorado sandúiche no exterior, durante o período de seis meses em Portugal e a todos os colegas do Bisbii (Bioinformatics and Systems Biology Interdisciplinary Initiative) que também me auxiliaram durante a minha estadia na Universidade do Minho em Braga – Portugal.

Aos meus irmãos, Catarina Moresco Bianchessi, César Moresco e Simoni Moresco. Agradeço muito a Deus pela família que formamos e pelo nosso companheirismo.

Aos anjos da minha família (Gabriel, Miguel, Davi, Mariana e Estevão) que mesmo sem entenderem muitas coisas, forneceram-me algo muito importante, seus sorrisos e carinhos.

Ao amor da minha vida, Caroline Venâncio, pela sua (fundamental) ajuda, pelo bem que me faz, por me entender e me respeitar e pelo seu amor, carinho e aconchego.

Ao meu avô professor Domingos Moresco (Dindinho Moresco) (in memoriam), pelo orgulho que me faz sentir em carregar seu nome, pela sua magnífica história de vida, pelo seu exemplo de amor à profissão e à família, que me servem de inspiração.

Aos meus grandes amigos Irineu Meschke Junior e Franciele Eccel Martinenghi pelos momentos de descontração tão importantes durante esta jornada, pela amizade espontânea e sincera.

À Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina. Sinceros agradecimentos aos pesquisadores Dr. Luiz Augusto Martins Peruch, Dr. Alexander Luís Moreto e Msc. Enilto de Oliveira Neubert, pelo essencial suporte e disponibilidade.

Ao grupo do Laboratório de Morfogênese e Bioquímica Vegetal (LMBV), ao qual tenho honra em pertencer, em especial aos amigos Fernanda Ramlov, Eva Regina de Oliveira Rodrigues, Beatriz Veleirinho, Letícia Mazzarino, Thaís Alberti, Ana Voytena, Manuel del Prá, Cláudia Bauer, que compartilharam a rotina laboratorial e muito mais que isto, amizade, ideias, críticas e risadas!

À Professora Edna Regina Amante, pela sua fundamental orientação neste trabalho. Agradeço sua disponibilidade e prontidão. Ao grupo do Laboratório de Frutas e Hortaliças (CAL-UFSC), pela receptividade e ajuda e por compartilharem sua experiência.

Ao Professor Carlos Eduardo Maduro Campos, (Laboratório de Difração de raio-X, Departamento de física – UFSC) porque, além da aquisição dos difratogramas e espectros Raman, esteve fortemente envolvido na orientação das análises destes dados. Agradeço pelo tempo dedicado e principalmente pelo seu interesse e sua observação crítica a este trabalho.

Ao professor Leonardo Rorig, pela boa receptividade de sempre e abertura em seu laboratório para outros projetos além desta tese, pelos conselhos e amizade.

Aos colegas professores, coordenadores e aos alunos da Universidade do Vale do Itajaí (UNIVALI) que me acolheram na instituição e que tanto valorizam o meu trabalho, que proporcionaram eu descobrir o amor pela docência, além do amor pela pesquisa.

À equipe técnica do Laboratório Central de Microscopia Eletrônica (LCME-UFSC), pelo auxílio durante as análises de microscopia eletrônica de varredura.

À Universidade Federal de Santa Catarina e todos os professores do Programa de Pós-Graduação em Biotecnologia e Biociências, discentes, pessoal técnico e administrativo.

À Capes pela concessão da Bolsa para realização deste trabalho e ao Governo Brasileiro (CNPq e MCT) pelo financiamento desta pesquisa científica (Edital – CNPq/Repensa. (Desenvolvimento da Cadeia Produtiva da Mandioca na Região Centro-Sul do Brasil.

“Rir muito e com frequência; ganhar o respeito de pessoas inteligentes e o afeto das crianças; merecer a consideração de críticos honestos e suportar a traição de falsos amigos; apreciar a beleza, encontrar o melhor nos outros; deixar o mundo um pouco melhor, seja por uma saudável criança, um canteiro de jardim ou uma redimida condição social; saber que ao menos uma vida respirou mais fácil porque você viveu: isto é ter sucesso. ”

(Ralph Waldo Emerson)

LISTA DE FIGURAS

CAPITULO I

Figura 1 - Estrutura da amilose e amilopectina.....	41
Figura 2 - Curva de viscosidade típica do RV.....	47
Figura 3 - Estrutura química dos principais carotenoides encontrados em raízes de <i>M. esculenta</i> Crantz.....	48
Figura 4 - Vias biossintéticas de carotenoides em plantas superiores..	50
Figura 5 - Diagrama representativo do espaço de cores CIE $L^*a^*b^*$..	54
Figura 6 - Localização dos carotenoides no plano CIE $L^*a^*b^*$. ..	55
Figura 7 - Fluxograma da abordagem de estudos do metaboloma.....	56

CAPITULO II

Figure 1: Typical UV-Vis spectrophotometric profiles ($\lambda = 200-700$ nm, acetone: petroleum ether - v/v) of root parenchymal tissues of ten cassava genotypes cultivated in southern Brazil.	75
Figure 2: Fold change analysis of the UV-Vis data set ($\lambda = 200-700$ nm, acetone: petroleum ether - v/v).....	76
Figure 3: Volcano plot analysis (Fold change threshold = 2 and $p \leq 0.001$)... ..	77
Figure 4 - Factorial distribution (principal components 1 and 2) of the UV-Vis spectral data set (200-700 nm) of the organosolvent extract of roots of ten cassava genotypes (A). Graphical demonstration of factorial distribution (principal components 1 and 2) according to the root flesh color (B).	78
Figure 5: A - Factorial distribution (principal components 1 and 2) of the spectral data set of the fingerprint region of carotenoids (UV-Vis 400-500 nm, acetone: petroleum ether - v/v, 1:1). B - Graphical demonstration according to the root flesh color.	79
Figure 6: A - Scores scatter plot (PC1 and PC2) of the quantitative data of carotenoids determined by RP-HPLC in root samples of ten cassava	

genotypes (n = 3). B - Magnification to the overlapping samples at the PCA..... 80

Figure 7: Dendrogram of cassava genotypes in respect to their carotenoid composition determined by RP-HPLC, followed by hierarchical clustering analysis (UPGMA method - 97.61% of cophenetic correlation).....81

CAPITULO III

Figure 1 - Structural formula of vitamin A (A) and of the carotenoids β -carotene (B), α -carotene (C) and β -cryptoxanthin (D), the major carotenoid sources in the human diet 87

Figure 2: Representation of the CIE L* a* b* color space. 89

Figure 3: Machine learning approach used. Three different datasets were used as input to the models, namely the UV-visible, CIELAB and fusion datasets..... 93

Figure 4: Typical UV-visible spectrophotometric profiles ($\lambda = 200\text{-}700\ \eta\text{m}$, acetone: petroleum ether (v/v)) of root parenchymal tissues of three cassava samples: A - sample 5, B - sample 23 and C - sample 74. The 400-500 ηm region of the spectrum is highlighted in cases B and C..... 95

Figure 5: Concentration of total carotenoids ($\mu\text{g.g}^{-1}$ dry weight \pm standard deviation, n = 3) in samples of roots of fifty *M. esculenta* genotypes, determined by UV-visible spectrophotometry (450 ηm , $\varepsilon = 2592\ \text{M}^{-1}\text{cm}^{-1}$). 96

Figure 6: Location of the cassava samples in the CIELAB color space according to their root pulp colors. 97

Figure 7: Scores plot with the distribution of the fifty samples on the first and second PCA components resulting from the UV-visible spectrophotometric data (400-500 ηm) (n= 3 replicates).....98

CAPITULO IV

Fig. 1 - Micrographs (SEM, A-J) of starch granules of cassava genotypes roots. (A) *Mesa*, (B) *Pioneira*, (C) *Oriental*, (D) *Amarela*, (E) *Catarina*, (F) *IAC*, (G) *Salézio*, (H) *Estação*, (I) *Videira*, and (J) *Rosada*. 122

Fig. 2 - Typical X-ray powder diffractogram of a starch sample of the studied cassava genotypes. The sharp peaks correspond to reflections of the B-type crystalline polymorph. 124

Fig. 3 – ATR-FTIR spectra of starches of ten cassava genotypes in the 3000–600cm⁻¹ wavenumber region. 126

Fig. 4 A- Factorial distributions of PC1 and PC2 for the spectral data (ATR-FTIR, 3000-600cm⁻¹) of cassava starches produced in southern Brazil. B- Factorial distribution of PC1 and PC2 for the ATR-FTIR (1200-950.cm⁻¹) carbohydrate fingerprint region dataset..... 128

LISTA DE TABELAS

CAPÍTULO III

Table 1: Performance values (RMSE and R^2) obtained for the different machine learning models trained with UV-visible spectrophotometry data (400-500 ηm)..... 99

Table 2: Performance values (RMSE and R^2) obtained for a random forest model trained with UV-visible spectrophotometry data (400-500 ηm), applying several pre-processing methods to the data..... 101

Table 3: Performance values (RMSE and R^2) obtained for the different machine learning models trained with CIELAB data. 102

Table 4: Performance values (RMSE and R^2) obtained for the different machine learning models trained with a fusion between UV-visible spectrophotometry and CIELAB data..... 103

CAPÍTULO IV

Table 1 - Physicochemical and functional properties of starches of ten cassava genotypes cultivated in southern Brazil.....119

Table 2 - Pasting characteristics of starch samples of ten cassava genotypes cultured in the southern Brazil.125

LISTA DE ABREVIATURAS E SIGLAS

ANN - *Artificial Neural Network*

CEASA/SC - Centrais de Abastecimento do Estado de Santa Catarina S/A

CG – Cromatografia Gasosa

CIAT - Centro Internacional de Agricultura Tropical

CIE L*a*b* /**CIELAB** - *Commision Internationale L'eclairage*

CLAE - Cromatografia Líquida de Alta Eficiência

CNPQ - Centro Nacional de Pesquisa de Mandioca e Fruticultura

DRX –Difração de Raios-X

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária

EPAGRI - Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina

FAO - *Food and Agriculture Organization*

FTIR - Espectroscopia Vibracional de Infravermelho com Transformada de Fourier (*Fourier Transform Infrared Spectroscopy*)

HCA – Análise Hierárquica de cluster (*Hierarchical Clustering Analysis*)

HCN - Ácido cianídrico

IAC - Instituto Agronômico de Campinas

IBGE - Instituto Brasileiro de Geografia e Estatística

KNN - *K-Nearest Neighbors*

LDA - *Linear Discriminant Analysis*

MEV – Microscopia Eletrônica de Varredura

MS - Espectrometria de Massas (*Mass Spectrometry*)

PCA – Análise de Componentes Principais

PLS - *Partial Least Squares*

RAE - Atividade Equivalente de Retinol (*retinol activity equivalents*)

RMN - Ressonância Magnética Nuclear (*Nuclear Magnetic Resonance*)

RP-HPLC - *Reverse Phase-High Performance Liquid Chromatography*

RVA - *Rapid Visco Analyser*

SIMCA - *Soft Independent Modelling by Class Analogy*

SVM - *Support Vector Machines*

UV-vis. – Ultravioleta – visível

SUMÁRIO

CAPÍTULO I	27	
1	INTRODUÇÃO	29
1.1	OBJETIVOS.....	33
1.1.1	<i>Objetivo geral</i>	33
1.1.2	<i>Objetivos específicos</i>	33
2	REVISÃO DE LITERATURA	36
2.1	<i>MANIHOT ESCULENTA</i> CRANTZ - BOTÂNICA, ORIGEM E DOMESTICAÇÃO.....	36
2.2	IMPORTÂNCIA ECONÔMICA.....	37
2.3	IMPORTÂNCIA ALIMENTAR E CULTURAL.....	38
2.4	MELHORAMENTO GENÉTICO.....	39
2.5	COMPOSIÇÃO AMILÁCEA.....	40
2.5.1	<i>Propriedades dos amidos</i>	42
2.5.2	<i>O amido de mandioca</i>	43
2.5.3	<i>Características físico-químicas e morfológicas dos grânulos de amido</i>	44
2.5.4	<i>Cristalinidade dos grânulos de amido</i>	44
2.5.5	<i>Características de pasta do amido</i>	45
2.6	CAROTENOIDES.....	47
2.6.1	<i>Manihot esculenta</i> como fonte de carotenoides provitamina A.....	49
2.6.2	<i>Biofortificação</i>	51
2.6.3	<i>Biofortificação de Manihot esculenta</i>	52
2.7	ESPAÇO DE CORES CIELAB E CAROTENOIDES.....	53
2.8	METABOLÔMICA.....	55
2.9	FERRAMENTAS DE BIOINFORMÁTICA.....	57
2.9.1	<i>Machine learning</i>	58
2.9.2	<i>Data fusion</i>	58
2.9.3	<i>Specmine</i>	59
3	REFERÊNCIAS	59
CAPÍTULO II	69	
UV-visible scanning spectrophotometry and chemometric analysis as tools for carotenoids analysis in cassava genotypes (<i>Manihot esculenta</i> Crantz)		
1	INTRODUCTION	72
2	MATERIAL AND METHODS	73

2.1	SELECTION OF CASSAVA GENOTYPES	73
2.2	EXTRACTION AND ANALYSIS OF CAROTENOID COMPOUNDS 74	
2.3	STATISTICAL AND CHEMOMETRIC ANALYSIS	74
3	RESULTS AND DISCUSSION.....	75
4	CONCLUSIONS.....	82
5	REFERENCES	83
CAPÍTULO III		85
UV-vis and CIELAB based chemometric characterization of <i>Manihot esculenta</i> carotenoid contents		
1	INTRODUCTION	87
2	MATERIALS AND METHODS.....	91
2.1	SELECTION OF CASSAVA GENOTYPES	91
2.2	CAROTENOID EXTRACTION AND QUANTIFICATION.....	91
2.3	STATISTICAL ANALYSIS	92
2.3.1	<i>Machine Learning</i>	<i>92</i>
2.3.2	<i>UV-visible dataset</i>	<i>93</i>
2.3.3	<i>CIELAB dataset.....</i>	<i>94</i>
2.3.4	<i>Fusion dataset</i>	<i>94</i>
2.3.5	<i>Tools and reporting.....</i>	<i>94</i>
3	RESULTS AND DISCUSSION.....	94
3.1	DETERMINATION OF CAROTENOID CONTENTS	94
3.2	CIELAB COLOR SPACE INTERPRETATION	97
3.3	CAROTENOID CONTENT PREDICTION USING UV-VISIBLE DATA	99
3.4	CAROTENOID CONTENT PREDICTION USING CIELAB DATA	101
3.5	CAROTENOID CONTENT PREDICTION USING FUSION DATA	103
4	CONCLUSIONS.....	105
5	REFERENCES	106
CAPÍTULO IV		110
Screening of physicochemical traits and pasting properties of starches of Brazilian cassava genotypes (<i>Manihot esculenta</i> Crantz)		

1	INTRODUCTION.....	112
2	MATERIAL AND METHODS.....	113
2.1	SELECTION OF GENOTYPES	113
2.2	STARCH ISOLATION	114
2.3	PHYSICOCHEMICAL AND FUNCTIONAL PROPERTIES	114
2.3.1	<i>Amylose content</i>	114
2.3.2	<i>Absolute density</i>	114
2.3.3	<i>Oil and water absorption</i>	115
2.3.4	<i>Swelling power of starch</i>	115
2.3.5	<i>Scanning Electron Microscopy (SEM)</i>	115
2.3.6	<i>Size of the starch granules</i>	115
2.4	X-RAY POWDER DIFFRACTION (XRPD)	115
2.4.1	<i>Crystallinity Index (CI)</i>	115
2.5	FOURIER TRANSFORM INFRARED SPECTROSCOPY (FTIR)	116
2.6	PASTING PROPERTIES	116
2.7	STATISTICAL ANALYSIS	116
3	RESULTS AND DISCUSSION.....	117
3.1	PHYSICOCHEMICAL AND FUNCTIONAL PROPERTIES	117
3.1.1	<i>Amylose content</i>	117
3.1.2	<i>Absolute density</i>	118
3.1.3	<i>Oil and water absorption</i>	120
3.1.4	<i>Swelling power (SP)</i>	120
3.1.5	<i>Shape and granule size</i>	120
3.2	X-RPD MEASUREMENTS AND CRYSTALLINITY INDEX ..	123
3.3	PASTING PROPERTIES – RAPID VISCO ANALYSER (RVA)	124
3.4	INFRARED SPECTROSCOPY (FTIR)	126
4	CONCLUSIONS	130
5	REFERENCES.....	131
	CAPÍTULO V.....	142
	CONSIDERAÇÕES FINAIS	142
1	CONCLUSÕES GERAIS.....	143
1	PRINCIPAIS CONTRIBUIÇÕES CIENTÍFICAS E/OU BIOTECNOLÓGICAS.....	143
2	PUBLICAÇÕES VINCULADAS À TESE DE DOUTORAMENTO	144

2.1	ARTIGOS COMPLETOS PUBLICADOS EM PERIÓDICOS.....	144
2.2	ARTIGOS COMPLETOS ACEITOS PARA PUBLICAÇÃO 146	
2.3	CAPÍTULOS DE LIVROS INTERNACIONAIS PUBLICADOS:	146
2.4	APRESENTAÇÕES DE TRABALHO:.....	148
3	AGRADECIMENTOS	148

Bioprospecção, metabolômica e caracterização amilácea e de carotenoides em germoplasma biofortificado de mandioca (*Manihot esculenta* Crantz) via ferramentas de bioinformática.

APRESENTAÇÃO

Este trabalho está dividido em 4 capítulos. O primeiro capítulo considera uma breve introdução sobre o tema, o referencial teórico e a definição dos objetivos deste trabalho.

O segundo capítulo refere-se ao artigo científico intitulado *UV-visible scanning spectrophotometry and chemometric analysis as tools for carotenoids analysis in cassava genotypes (Manihot esculenta Crantz)*, publicado no periódico *Journal of Integrative Bioinformatics*. Este artigo trata da caracterização metabolômica, focando na composição de carotenoides de genótipos de *M. esculenta*, cultivados no sul do Brasil, através das técnicas analíticas, e.g., UV-visível e CLAE. O conjunto de dados gerados por estas técnicas permitiu construir um modelo de classificação a partir de análises quimiométricas.

O Capítulo III trata da validação de um método de quantificação de carotenoides em amostras de *M. esculenta*, a partir de dados de colorimetria, utilizando o sistema CIE L*a*b*. Através de técnicas estatísticas de análises de prognósticos como o aprendizado de máquina (*Machine Learnig*) foram correlacionados dados colorimétricos, de fácil obtenção a campo, com os teores obtidos via técnicas tradicionais à quantificação destes compostos, como a espectrofotometria UV-visível e a CLAE. A partir disto, foram construídos modelos de predição de carotenoides para este tipo de biomassa. Este trabalho refere-se a um artigo publicado no *Journal of Integrative Bioinformatics*.

O quarto capítulo refere-se ao artigo intitulado *Screening of physicochemical traits and pasting properties of starches of Brazilian cassava genotypes (Manihot esculenta Crantz)*, recentemente submetido ao periódico internacional *International Journal of Food Science and Technology*. Este estudo trata da caracterização de genótipos de *M. esculenta*, com ênfase na fração amilácea. Foram estudadas as características físico-químicas e propriedades de pasta de amidos, sugerindo usos específicos para estas biomassas, sejam alimentares ou industriais.

Por fim, o capítulo V traz a conclusão geral da tese, as considerações finais, além das contribuições tecnológicas e científicas

geradas por este trabalho até o momento da elaboração deste manuscrito.

CAPÍTULO I

Bioprospecção, metabolômica e caracterização amilácea e de carotenoides em germoplasma biofortificado de mandioca (*Manihot esculenta* Crantz) via ferramentas de bioinformática.

1 INTRODUÇÃO

A mandioca (*Manihot esculenta*, Crantz) é uma planta perene, arbustiva, pertencente à família Euforbiácea. Originária do continente americano, já era cultivada no Brasil pelos indígenas, quando os portugueses aqui chegaram. Esta espécie tem uma importância social pois serve como alimento energético básico na dieta de mais de 800 milhões de pessoas, em diversos países (FAO, 2013).

Produtiva em solos deficientes em fertilidade é muito tolerante às doenças, à seca e às pragas, além de ser flexível quanto à época de colheita (UARROTA et al., 2017). Por ser uma cultura rústica, adaptada a variados ambientes, sem grandes exigências em agroquímicos e sem depender de sementes, despertou pouco interesse das corporações controladoras da “agricultura moderna” permanecendo por longo tempo à margem das atenções dos governos e do agronegócio (PINTO-ZEVALLOS; PAREJA; AMBROGI, 2016). Atualmente, devido à sua importância cultural e econômica frente à agricultura familiar e às preocupações crescentes relacionadas à segurança alimentar, conservação dos recursos fitogenéticos e aspiração por modelos de produção agrícolas mais sustentáveis, têm-se buscado resolver os principais gargalos tecnológicos e organizacionais gerados por essa histórica desconsideração.

Em *M. esculenta*, a parte mais importante da planta é a raiz, rica em amido, utilizada na alimentação humana e animal, ou como matéria-prima industrial. A fração amilácea é formada basicamente por dois polissacarídeos, i.e., amilose e amilopectina, sendo a relação de conteúdos destas duas macromoléculas variável de acordo com a fonte botânica e também entre genótipos de uma mesma espécie, conferindo características específicas à pasta de amido (SINGH et al., 2003).

A fonte de amido pode influenciar de modo determinante diversos processos tecnológicos da indústria alimentícia, como a textura e a retenção de água de determinados alimentos, assim como processos metabólicos vitais da nutrição humana, como a resposta glicêmica ao alimento ingerido. Esses eventos estão intimamente relacionados às características estruturais do amido, como teor de amilose, distribuição

de comprimento das cadeias de amilopectina e cristalinidade no grânulo, sendo também associados a algumas etapas dos processos de gelatinização e retrogradação, como inchamento do grânulo, lixiviação de amilose e/ou amilopectina, etc. (ZHU, 2015). Tais características demonstram que estudos sobre a caracterização estrutural de frações amiláceas podem ser usados como marcadores bioquímicos para avaliar a diversidade química entre variedades, e.g. genótipos de mandioca, e também são importantes a fim de dirigir e otimizar suas aplicações tecnológicas e nutricionais. Isso é extremamente relevante, pois, a maior parte dos amidos nativos não possui as propriedades funcionais requeridas e por isso tem um uso limitado na indústria.

Em uma segunda abordagem, estudos indicam o potencial de *M. esculenta* como fonte de compostos de interesse à saúde humana e nutrição, dada à presença de carotenoides provitamina A (β -caroteno, e.g.) e licopeno nas raízes de coloração amarela e rosada, respectivamente (NASSAR et al., 2007). Estudos epidemiológicos sugerem uma correlação positiva entre uma dieta rica em carotenoides e a redução na incidência de alguns tipos de câncer (CARVALHO et al., 2012; HUGHES, 2001), sendo também amplamente reconhecido o efeito antioxidante e inibidor da ação de radicais livres desses metabólitos (PALOZZA, 2005). No organismo humano, os carotenoides possuem uma função importante como precursores da vitamina A, sendo o β -caroteno, α -caroteno e a β -criptoxantina os principais carotenoides que apresentam esta atividade. A deficiência de vitamina A é um grande problema nutricional em países em desenvolvimento, causando cegueira, xerofthalmia e mortes prematuras, especialmente em crianças (WEDNER; ROSS, 2017). Assim, a presença desses metabólitos secundários indica uma funcionalidade adicional à matéria-prima e aos produtos derivados de mandioca, sendo esta considerada um alimento tradicional para muitas populações com risco de deficiência de vitamina A. Tais características evidenciam a importância de pesquisas de caracterização das biomassas de genótipos de *M. esculenta*, subsidiando o desenvolvimento de novas aplicações da matéria-prima dessa espécie junto às indústrias alimentícia, farmacêutica e cosmética, por exemplo.

Sendo assim, a presente proposta visa iniciar um processo de bioprospecção e desenvolver protocolos de análise de metabólômica, com ênfase na composição amilácea e carotenóidica de genótipos de mandioca de mesa pigmentados (raízes com polpa de coloração amarelada e rosada) e genótipos industriais para produção de amido, que possibilitem sua utilização como ferramentas auxiliares nos processos

de identificação e seleção de genótipos de mandioca, oriundos do programa de melhoramento, já lançados no Brasil por projetos de biofortificação e outros genótipos promissores identificados para este desígnio. Este trabalho abordou a exploração e integração de dados a partir de técnicas experimentais analíticas distintas (típicos de uma plataforma metabolômica), incluindo técnicas cromatográficas (Cromatografia Líquida – LC), espectroscópicas (Espectroscopia Vibracional de Infravermelho – FTIR), além da Espectrofotometria UV-visível (UV-vis.) e da Difractometria (Difração de Raios-X – DRX), associadas às ferramentas de bioinformática. A combinação de dados metabolômicos com o conhecimento bioquímico existente é raramente descrita na literatura, consistindo em uma lacuna a ser preenchida.

Assume-se que o uso concomitante destas tecnologias analíticas seja uma ferramenta eficiente na busca e definição de genótipos com teores superiores de metabólitos-alvo, de importância em programas de melhoramento genético, subsidiando a seleção de genótipos parentais promissores e/ou potenciais. Além disso, o acesso às peculiaridades químicas destes genótipos que determinam seus potenciais industriais, nutricionais e de eventuais benefícios à saúde humana amplia suas possibilidades, podendo agregar-lhes valor e amenizar o quadro de baixo retorno econômico associado ao seu cultivo. Consequências positivas sobre a qualidade de vida das populações rurais, permitindo a manutenção das mesmas nessa atividade e a conservação daqueles genótipos são esperadas neste cenário.

1.1 OBJETIVOS

1.1.1 Objetivo geral

Desenvolver, através de ferramentas de bioinformática e quimiometria, protocolos de análises do metaboloma de *M. esculenta* que possibilitem a geração de uma ferramenta de discriminação eficiente à identificação e seleção de genótipos de interesse, no contexto bioquímico.

1.1.2 Objetivos específicos

- Caracterizar a fração amilácea quanto à concentração de amilose/amilopectina, densidade absoluta, capacidades de absorção lipídica e de água e propriedades de pasta;
- Analisar a morfologia via microscopia eletrônica de varredura (MEV) e o tamanho dos grânulos de amido por difração laser (Master Sizer);
- Determinar o índice de cristalinidade dos grânulos de amido por difração de Raio-X (DRX);
- Caracterizar a fração amilácea (amilose e amilopectina) por Espectroscopia Vibracional de Infravermelho Médio com Transformada de Fourier (FTIR);
- Determinar a concentração de carotenoides totais em amostras de raízes dos genótipos de mandioca via espectrofotometria UV-visível;
- Caracterizar o perfil carotenóidico por cromatografia líquida de alta eficiência (CLAE) e subsequente construção de modelo descritivo, via análise de componentes principais (PCAs).
- Determinar os valores de cores das raízes no espaço CIELAB das amostras através de colorimetria;
- Validar um método de quantificação de teores de carotenoides em amostras de *M. esculenta* a partir de dados de colorimetria, utilizando o sistema CIE $L^*a^*b^*$.
- Construir modelos estatísticos preditivos e descritivos, de classificação e de reconhecimento de padrões, a partir dos dados de perfis metabólicos, via mineração de dados e técnicas de estatística multivariada não supervisionada e supervisionada.

2 REVISÃO DE LITERATURA

2.1 *Manihot esculenta* Crantz - BOTÂNICA, ORIGEM E DOMESTICAÇÃO.

A mandioca é uma planta heliófila, perene e arbustiva, pertencente à classe das Dicotiledôneas, à ordem Malpighiales, à família Euphorbiaceae, ao gênero *Manihot* e à espécie *Manihot esculenta* Crantz. O gênero *Manihot* é composto por 98 espécies com diferentes hábitos de desenvolvimento, arbustivos ou arbóreos. *M. esculenta* é a única espécie desse gênero que produz raízes comestíveis, considerando o acúmulo de amido naqueles órgãos (OLSEN; SCHAAL, 1999).

A mandioca pertence ainda a um grupo de plantas reconhecidas como espécies cianogênicas, devido à presença de glicosídeos cianogênicos em sua composição, sendo considerada a espécie cianogênica mais importante do Brasil (AMORIM et al., 2006). Tais glicosídeos, conhecidos como linamarina e lotaustralina, presentes naturalmente nas estruturas da planta, são determinantes para a sua toxicidade. Quando as raízes de *M. esculenta* são cortadas, raladas ou prensadas, a estrutura celular das raízes é quebrada, permitindo o contato da linamarina e da lotaustralina com a linamarase, a qual catalisa a reação de hidrólise dos glicosídeos, gerando ácido cianídrico (BLAGBROUGH et al., 2010). As variedades de mandioca podem ser classificadas de acordo com os teores de ácido cianídrico presentes nas suas raízes. O grupo de variedades de mandioca “mansa”, conhecida também como “doce”, caracteriza-se, principalmente, por apresentar teores de cianeto abaixo de 100 mg/kg de polpa nas raízes frescas, enquanto que as variedades com concentrações de cianeto na raiz fresca acima de 100 mg/kg de peso fresco são denominadas “bravas” ou “venenosas”. Estas variedades são impróprias para consumo fresco, sendo, portanto, indicadas apenas para a indústria, onde a toxicidade da raiz é reduzida durante o processamento (BOLHUIS, 1954).

Apesar da sua importância econômica, as origens evolutivas e geográficas da cultura de *M. esculenta* ainda são controversas (OLSEN; SCHAAL, 1999). É originária das Américas e possivelmente do Brasil Central, mas há evidências de seu cultivo por civilizações Incas há 4000 anos na América Central, Venezuela, Colômbia e Peru. Estudos evidenciam que a cultura da mandioca (*Manihot esculenta* Crantz spp. *esculenta*) possivelmente foi domesticada de populações selvagens de *M. esculenta* da subespécie *flabellifolia* (HOWELER; OATES; ALLEM, 2001), na extensão da fronteira sul da bacia Amazônica (OLSEN; SCHAAL, 1999). Sabe-se, contudo, que *M. esculenta* foi

domesticada e cultivada por nativos americanos e levada à África e Ásia após a chegada dos europeus no século XVI. Seu cultivo em regiões tropicais estende-se em uma faixa de 30° de latitude norte e a sul. De interesse, destaca-se que o Brasil é o país com o maior centro de diversidade de espécies de *Manihot* (NASSAR; HASHIMOTO; FERNANDES, 2008).

2.2 IMPORTÂNCIA ALIMENTAR E CULTURAL

A mandioca é comumente cultivada por pequenos produtores ao longo do Brasil, sendo a organização destes importante à manutenção da tradição de populações locais, as quais dependem economicamente desta atividade (DENARDIN; KOMARCHESKI, 2015). Baseado nesta ideia de valorizar os produtos artesanais de qualidades especiais de produção local e de manter as tradições regionais, o *Slow Food*, movimento surgido na década de 1980 em oposição ao *fast-food*, tem agido localmente e promovido eventos e encontros para disseminar a importância do prazer à alimentação saudável e consciente, a fim de incentivar pequenos setores negligenciados. Como exemplo, a farinha de mandioca de Santa Catarina tem sido destacada pelo movimento *Slow Food*, porque apresenta características singulares que a diferenciam de outras encontradas no Brasil, como a cor branca, textura fina e macia. Isto resulta principalmente das variedades de mandioca utilizadas e da fina peneiração após forneada, simbolizando, portanto, a troca cultural entre índios e açorianos, sendo que os primeiros apresentaram a cultura da mandioca e os modos rústicos de elaborar a farinha, enquanto os últimos aperfeiçoaram a seleção de variedades, os maquinários e as técnicas utilizadas. Esse processo foi em parte influenciado pela busca de uma farinha semelhante àquela de trigo, muito utilizada nas receitas tradicionais das Ilhas dos Açores (CASTELLANO, 2015). No litoral Catarinense, o processamento industrial destina-se principalmente à produção de farinha e polvilho azedo. Ressalta-se que nessa região há inúmeras casas de farinha artesanais, onde a produção é efetuada manualmente, obtendo-se uma farinha de melhor qualidade e melhores preços. Dessa forma, o segmento mandioqueiro catarinense, por apresentar características culturais históricas, técnicas de manejo e qualidade de produto únicas, busca por meio da Denominação de Origem Controlada ou de outras certificações que garantem a procedência e a forma tradicional de produção, valorizar os produtos derivados do beneficiamento das raízes de mandioca na região, garantindo assim a preferência por parte dos consumidores.

2.3 IMPORTÂNCIA ECONÔMICA

A mandioca ocupa a posição da terceira cultura mais importante, como fonte de calorias, do mundo, dentre o grupo das culturas alimentícias básicas, que inclui o arroz e o milho (FAO, 2013).

Nos últimos cinco anos, configuram-se como os maiores produtores mundiais de mandioca a Nigéria (480.10⁶ t), Tailândia (268.10⁶ t), Brasil (238.10⁶ t) e Indonésia (236.10⁶ t). O continente africano lidera a produção de mandioca, representando 55,3% do volume total produzido, seguido do continente asiático com 32,1% e o americano com 12,5%. O Brasil é o terceiro maior produtor mundial de mandioca com 10% do volume total de produção. Estima-se que 65% da produção seja utilizada para consumo humano, 25%, para uso industrial (principalmente como amido) e 19% para alimentação animal (FAOSTAT, 2014). Segundo o Instituto Brasileiro de Geografia e Estatística – IBGE, a produção média de mandioca no Brasil nos últimos cinco anos (2012-2016) foi de 22.846.652 t, com uma produtividade média de 14.615 kg/ha, sendo que em 2016, a espécie foi cultivada em mais de 1,5 milhão de hectares (IBGE, 2017). Ainda de acordo com aquele Instituto, estima-se uma pequena queda na safra de 2017 em relação aos dois anos anteriores, onde 20,8 milhões de t deverão ser produzidas até o final desse ano (IBGE, 2017).

A região Sul do Brasil é responsável pela segunda maior produção de mandioca do País, atrás apenas da região Norte. Em particular, o Estado de Santa Catarina, tradicional produtor de farinha e fécula, apresentou em 2016 uma produção de 385.875 t, representando assim a décima terceira maior produção desta raiz no Brasil. Naquele ano, o Estado colheu 20.713 hectares, com um rendimento médio de 18.630 kg/ha (“Relatório de Mercado Agrícola CEASA/SC”, 2017).

A farinha e o amido de mandioca são os principais produtos de interesse à indústria, sendo o amido e seus derivados utilizados em produtos para a alimentação humana e para fins industriais, e.g., embalagens, colas, mineração, têxtil e farmacêutica. Mais da metade do amido de mandioca produzido no Brasil é utilizado por indústrias alimentícias na forma nativa do amido, destacando-se as indústrias de biscoitos, iogurtes, panificação, chocolates e bombons, conservas entre outras (CEREDA et al., 2001). Além disso, boa parte da mandioca de “mesa” é consumida na forma *in natura* e já vem sendo comercializada nas formas pré-cozida e congelada. O mercado da mandioca de mesa tem crescido nos últimos anos com o advento de novas formas de comercialização, a partir do surgimento de indústrias de congelados e pré-cozidos congelados, que oferecem um produto de qualidade durante todo o ano. Estas novas formas de comercialização vêm promovendo a

sofisticação dos mercados de mandioca de mesa, cada vez mais exigentes quanto à qualidade do produto (“Relatório de Mercado Agrícola CEASA/SC”, 2017).

2.4 MELHORAMENTO GENÉTICO

Por ser uma cultura perene e devido à sua rusticidade, *M. esculenta* pode ser mantida *on farm*, constituindo bancos de germoplasma. Como a mandioca é cultivada em quase todo o território brasileiro, uma alta variabilidade genética é suposto existir. No entanto, a maior parte desta ainda não foi utilizada em programas de melhoramento genético. Além disto, pelo fato dessas variedades terem sido cultivadas e sofrerem seleções ao longo do tempo pelos “agricultores-melhoristas”, pode haver variedades com características de interesse agrônomo e industrial no âmbito das pequenas propriedades, as quais ainda não foram identificadas e introduzidas em bancos de germoplasma (CEREDA et al., 2001).

Em programas de melhoramento genético de mandioca, os principais métodos utilizados são a introdução e seleção de variedades e as hibridizações intraespecíficas. A introdução e seleção de variedades é o método mais comum por ser menos oneroso e por apresentar grande chance de êxito (IGLESIAS et al., 1997). As propriedades buscadas com essas ações são maior valor nutricional, maior concentração de amido e menor teor de ácido cianídrico, dentre outros atributos que tornem as novas variedades mais completas, competitivas e atraentes ao mercado. Um exemplo de melhoramento do valor nutricional é o trabalho com o genótipo Rosada, desenvolvido no Centro Internacional de Agricultura Tropical (CIAT) de Cali, Colômbia – o detentor do maior banco mundial de germoplasma de *M. esculenta*. Por meio de pesquisas genéticas foi possível aumentar o teor de licopeno nas raízes, um antioxidante recomendado na prevenção do câncer, principalmente o de próstata (FUKUDA, 2006).

No Brasil, cerca de sete mil variedades estão disponíveis ao melhoramento genético, concentradas nos principais bancos de germoplasma do país: o Centro Nacional de Pesquisa de Mandioca e Fruticultura (CNPQ, da Embrapa), o Cenargen (da divisão Recursos Genéticos e Biotecnologia da Embrapa), a Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina (Epagri), o Instituto Agrônomo de Campinas (IAC) e a Escola Superior de Agricultura Luiz de Queiroz (ESALQ-USP).

2.5 COMPOSIÇÃO AMILÁCEA

A sacarose e o amido são sintetizados a partir de triose fosfato gerados pelo ciclo de Calvin e representam as duas principais formas de armazenamento de carbono e energia nos vegetais superiores. A sacarose, sintetizada no citosol, é a principal forma de carboidrato translocável pela planta, através do floema. O amido, carboidrato de reserva energética, é insolúvel em água e sintetizado nos cloroplastos e amiloplastos (TAIZ; ZEIGER, 2009).

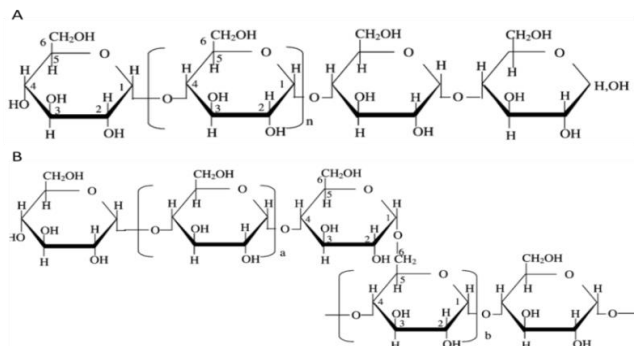
No processo de tuberização, a sacarose produzida pela fotossíntese nas folhas é armazenada nos cloroplastos na forma de amido transitório, o qual é depois degradado para nova síntese de sacarose. Esta é então translocada pelo floema para o interior das células dos órgãos de reservas subterrâneos, acumulando-se nos vacúolos, onde é degradada por invertases, liberando glucose e frutose, que são fosforiladas por suas respectivas quinases, entrando no amiloplasto e iniciando a formação dos grânulos de amido.

A síntese de amido é similar nos cloroplastos e amiloplastos, envolvendo a presença das enzimas pirofosforilase ADP-glucose, uma enzima chave na síntese daquele polissacarídeo, promovendo o alongamento das cadeias $\alpha(1\rightarrow4)$; e das enzimas ramificantes, responsáveis pela formação da microestrutura do grânulo do amido (TAIZ; ZEIGER, 2009).

Dentre os polissacarídeos, somente o amido se organiza na forma de grânulos, que são densamente empacotados em uma estrutura semicristalina. A cristalinidade varia de 15 a 45% e a densidade em torno de $1,5 \text{ g/cm}^3$. Na sua forma granular, o amido pode ser facilmente isolado pela gravidade, sedimentação, centrifugação e filtração. Os grânulos se desenvolvem ao redor do ponto de nucleação, denominado de hilum, sendo a forma e o tamanho dependentes da espécie botânica (LIU, 2005).

Estruturalmente, o amido é um homopolissacarídeo composto por cadeias de amilose e amilopectina. A amilose é formada por unidades de glucose unidas por ligações glicosídicas $\alpha(1\rightarrow4)$, originando uma cadeia linear, podendo apresentar pequenas ramificações. Por sua vez, a amilopectina é formada por unidades de glucose unidas em $\alpha(1\rightarrow4)$, apresentando em torno de 5% pontos de ramificações $\alpha(1\rightarrow6)$ (PÉREZ; BERTOFT, 2010).

Figura 1– Estrutura da amilose (a) e amilopectina (b).



Fonte: Tester; Karkalas e Qi, (2004).

A amilose é um dos componentes responsáveis pela estrutura do grânulo e sua quantificação é relevante ao entendimento do comportamento físico-químico do amido. Conforme relatado por Hu; Burton e Yang, (2010), a pequena ramificação química na estrutura da amilose contribui ao seu valor nutricional. Enzimas de degradação podem agir simultaneamente nas extremidades não redutoras (existentes nas ramificações externas da amilopectina) e assim aceleram a conversão deste polímero em monossacarídeos). Logo, a amilose é menos susceptível à ação enzimática sendo mais adequada ao armazenamento de energia, devido à sua estrutura não ramificada com menos extremidades não redutoras (NELSON; COX, 2006). A amilose é digerida mais lentamente, fornecendo benefícios à saúde humana. Estudos clínicos indicam que a amilose é importante na redução do impacto glicêmico e da insulina (BEHALL; SCHOLFIELD, 2005) e também no aumento da capacidade do corpo em metabolizar gorduras (HIGGINS et al., 2004).

A amilopectina tem um papel principal na cristalinidade dos grânulos de amido, sendo estrutural e funcionalmente a mais importante das duas frações, pois sozinha é suficiente para formar o grânulo, como ocorre em mutantes que são desprovidos de amilose. A região cristalina do grânulo de amido é formada por cadeias laterais de amilopectina, enquanto a região amorfa é composta por pontos adjacentes da amilopectina e por amilose. Assim, a cristalinidade do grânulo está, principalmente, relacionada às duplas hélices formadas pelas ramificações de amilopectina (HOOVER, 2001).

A funcionalidade do amido, assim como a organização física da estrutura granular é, em grande parte, atribuída à razão amilose/amilopectina. Por estar ligada ao desenvolvimento da planta, a estrutura do amido está intimamente relacionada às condições do ambiente de

cultivo. Alterações de clima, por exemplo, podem influenciar na formação do grânulo e na deposição do amido (AMANTE, 1986).

2.5.1 Propriedades dos amidos

A estrutura macromolecular do amido interfere em suas características físico-químicas, e.g., solubilidade, viscosidade, poder de gelatinização e de adesão. Amidos naturais podem não apresentar propriedades adequadas ao uso industrial, de modo que as empresas de alimentos, cosmética e farmacêutica recorrem, em muitos casos, ao uso de amidos modificados. As razões que levam à modificação são: otimizar as características de cozimento (gelatinização), diminuir a retrogradação, reduzir a tendência das pastas a formarem géis, aumentar a estabilidade das pastas ao resfriamento e congelamento, aumentar a transparência das pastas ou géis, melhorar a formação de filmes, aumentar a adesividade, adicionar grupamentos hidrofóbicos e introduzir poder emulsificante (BEMILLER; LAFAYETTE; IN, 1997).

Neste contexto, as indústrias de alimentos e os produtores agrícolas estão interessados na identificação e no desenvolvimento de espécies que produzam amidos nativos com características físico-químicas especiais e propriedades semelhantes à dos amidos modificados. Esses amidos poderiam substituir amidos modificados quimicamente, ou ainda abrir novos mercados para espécies ainda não valorizadas (KIM et al., 1995).

As propriedades do amido envolvem suas características físicas (densidade e tamanho do grânulo, cristalinidade, morfologia, condutibilidade térmica e elétrica e atividade ótica), químicas (teores de amilose, amilopectina e outros constituintes e.g. lipídeos e proteínas) e funcionais (poder de inchamento, gelatinização, retrogradação e hidrólise) (CEREDA et al., 2001).

2.5.2 O amido de mandioca

As fontes de amido mais utilizadas mundialmente são de cereais, como milho e trigo. O amido de mandioca apresenta vantagens, pois é facilmente extraído, já que as raízes contêm pequenas quantidades de proteínas, gorduras e outros componentes. Dessa forma, o processo de extração é simples e o amido obtido é puro e branco. De modo geral, o amido de mandioca é constituído do polissacarídeo quase puro (99,26 %), contendo aproximadamente 0,34 % de proteína, 0,22 % de gordura e 0,06 % de cinzas (AMANTE, 1986). Por apresentar baixo teor de lipídios, o amido de mandioca e seus derivados não apresentam sabor e aroma de cereais, o que é desejável para muitos produtos alimentícios.

Na indústria alimentícia, principalmente no setor frigorífico, é um importante substituto para o amido de milho (DEMIATE et al., 2005).

O amido de mandioca apresenta cerca de 18 % de amilose, enquanto que os amidos de cereais possuem em torno de 22 %. Devido a essa diferença, os géis de amidos de cereais são mais rígidos, comparativamente aqueles de espécies tuberosas que são mais viscosos e transparentes (PERONI; ROCHA; FRANCO, 2006).

O amido de mandioca apresenta forma arredondada, com uma das extremidades truncada e tamanho de grânulos que variam de 5 a 40 μm (JANE et al., 1994). Quando aquecido em água, este amido apresenta alta viscosidade e baixa tendência a retrogradação (SWINKELS, 1985). Dentre os amidos de raízes e tubérculos, é um dos que apresenta menor temperatura inicial de gelatinização, i.e., valores médios em torno de 60°C (HOOVER, 2001). Somadas a estas características, a mandioca é ainda uma das culturas agrícolas mais cultivadas nos trópicos, com a produtividade superior a outras culturas agrícolas fontes de amido, necessitando de baixos *inputs* de energia para expressar seu potencial produtivo (MONTAGNAC; DAVIS; TANUMIHARDJO, 2009).

2.5.3 Características físico-químicas e morfológicas dos grânulos de amido

Segundo Tester; Karkalas e Qi, (2004).a forma (redondo, oval, poliédrico), o tamanho de partícula (2 a 100 μm) e a distribuição de tamanho da partícula dos grânulos são características da origem botânica dos amidos. Amidos com morfologias diversas têm mostrado suscetibilidades peculiares à hidrólise enzimática.

Os grânulos de amido de mandioca apresentam diâmetros similares aos de amido de milho, porém são morfológicamente diferentes. São ovais ou redondos com alguns côncavo-convexos característicos. Os grânulos de amido são insolúveis em água fria; no entanto, se a continuidade da membrana exterior tiver sido destruída pela moagem, a absorção de água pode ser facilitada. Por tratamento do grânulo intacto com água quente, uma parte solúvel do amido difunde através da parede do grânulo. Em água quente, os grânulos incham e evoluem para o rompimento (TESTER; MORRISON, 1990).

2.5.4 Cristalinidade dos grânulos de amido

Na planta, o amido é armazenado como corpos intracelulares parcialmente cristalinos denominados grânulos. Por meio da difração de raios-X, podem-se distinguir três tipos de grânulos que, dependendo da sua forma e estrutura cristalina, denominam-se A, B e C (ZOBEL;

YOUNG; ROCCA, 1988). Os grânulos de amido tipo A apresentam picos de intensidade nos ângulos de difração 2θ em $15,3^\circ$, $17,8^\circ$ e $23,5^\circ$. Por sua vez, grânulos do tipo B apresentam picos de intensidade nos ângulos de difração 2θ em $5,6^\circ$, $14,4^\circ$, $17,2^\circ$, $22,2^\circ$ e 24° , enquanto picos de intensidade nos ângulos de difração 2θ em $5,6^\circ$, $15,3^\circ$, $17,3^\circ$ e $23,5^\circ$ são encontrados aqueles do tipo C (ZOBEL; YOUNG; ROCCA, 1988),

As cadeias externas relativamente curtas das moléculas de amilopectina (entre 23 e 29 unidades de glucose) favorecem a formação de polimorfo cristalino tipo A, comumente encontrado em cereais. Já as cadeias externas maiores de amilopectina (entre 33 e 44 unidades de glucose) favorecem a formação de polimorfo tipo B, encontrados em amido de banana, amidos retrogradados e amidos ricos em amilose. O polimorfo tipo C é composto por moléculas de amilopectina com 26 a 29 moléculas de glucose, é considerado um intermediário entre A e B, sendo característico de amido de leguminosas e sementes. Geralmente, amidos de tuberosas e raízes exibem padrão de difração tipo B. (ELIASSON, 2004).

2.5.5 Características de pasta do amido

2.5.5.1 Poder de inchamento

O poder de inchamento é uma medida da capacidade de hidratação dos grânulos. Sua determinação consiste na medida do peso do grânulo intumescido (inchado) e da água oclusa (AMANTE, 1986).

O inchamento e a solubilidade do grânulo variam de acordo com a fonte botânica do amido, fornecendo evidências da interação entre as cadeias nos domínios amorfos e cristalinos dos polissacarídeos. A extensão destas interações é influenciada pela proporção amilose/amilopectina e pelas características dessas moléculas (distribuição e massa molecular, grau e comprimento de ramificações e conformação) (SINGH et al., 2003).

Quando os grânulos de amido intactos são colocados em água fria não são solúveis, mas podem reter pequenas quantidades de água, devido à difusão e absorção de água nas regiões amorfas, um processo reversível pela secagem (WHISTLER; DANIEL, 2000). No entanto, quando o amido é aquecido em excesso de água, acima da temperatura de gelatinização, a estrutura cristalina é rompida devido ao relaxamento das ligações de hidrogênio e as moléculas de água interagem com os grupos hidroxilas da amilose e da amilopectina. Isso causa um aumento do tamanho do grânulo devido ao inchamento, com solubilização parcial (HOOVER, 2001).

Segundo Ratnayake e Jackson (2006), a absorção de água pelos grânulos de amido nas regiões amorfas desestabiliza sua estrutura cristalina, resultando na perda da birrefringência, que é uma das definições de gelatinização. Sob aquecimento contínuo, os grânulos incham mais, levando à perda da estrutura cristalina, o que pode causar a completa separação da amilose e amilopectina. Tester; Morrison, (1990) reportaram que o inchamento do amido é primariamente uma propriedade da amilopectina, porque amidos cerosos usualmente incham mais, comparativamente a amidos com baixo teor daquele polissacarídeo. Por sua vez, Gomand et al., (2010) relatam que durante o inchamento a amilose atua como diluente. De fato, amidos ricos em amilose mostram inchamento e solubilidade restritos, mesmo após um período prolongado de aquecimento (LEACH; SCHOCH, 1962).

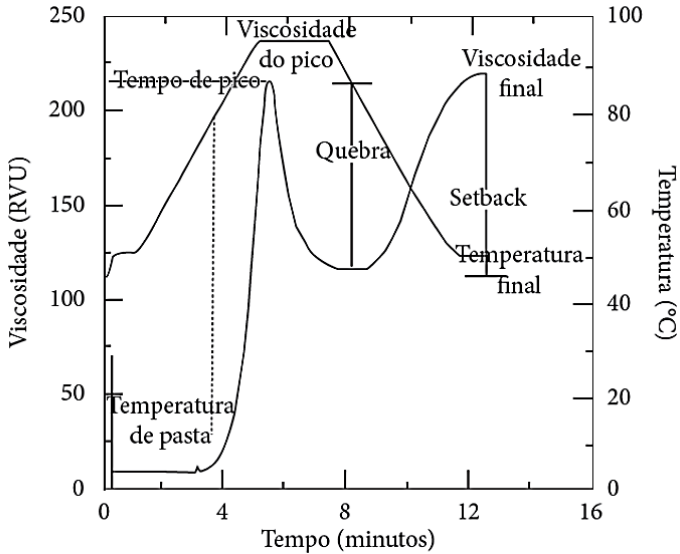
2.5.5.2 Viscosidade

O momento em que desaparece a birrefringência é conhecido como ponto ou temperatura de gelatinização, que normalmente ocorre dentro de uma faixa de temperaturas, gelatinizando primeiramente os grânulos maiores e posteriormente os menores. Já o empastamento é o fenômeno seguinte à gelatinização na dissolução de amido. Envolve inchamento granular, separação dos componentes moleculares e, eventualmente, total rompimento dos grânulos. É nesta etapa que ocorre o aumento da viscosidade do gel (WHISTLER; DANIEL, 2000).

Segundo Dongowski, (1997), a formação da pasta ocorre porque com o aumento da temperatura da suspensão amido/água as moléculas constituintes do grânulo daquele polissacarídeo vibram e giram, fazendo com que as ligações de hidrogênio intermoleculares sejam quebradas e substituídas com moléculas de água, o que produz uma hidratação extensiva. O efeito de mobilidade é percebido primeiramente nas regiões amorfas, as quais estão em estado vítreo, não cristalino. Quando os grânulos aquecidos na presença de até 60% de água, que é plasticizante para o amido, alcançam a sua temperatura de transição vítrea, as regiões amorfas plasticizadas sofrem uma transição de fase, passando do estado vítreo para o estado pastoso.

A construção das curvas de viscosidade representa importante ferramenta às observações do comportamento do gel de amido e de suas relações com as condições de processamento: aquecimento, agitação e resfriamento. O perfil de empastamento de amidos obtidos pelo RVA (*Rapid Visco Analyser*) inclui pico de viscosidade, tempo para atingir este pico, quebra, viscosidade final, temperatura de pasta e tendência à retrogradação (*setback*) (WICKER, 2004) (Figura 2). Dependendo do tipo de amido, diferentes perfis de empastamento podem ser gerados.

Figura 2 - Curva de viscosidade típica do RVA mostrando as variáveis comumente avaliadas.



Fonte: Newport Scientific (2010).

2.5.5.3 Retrogradação

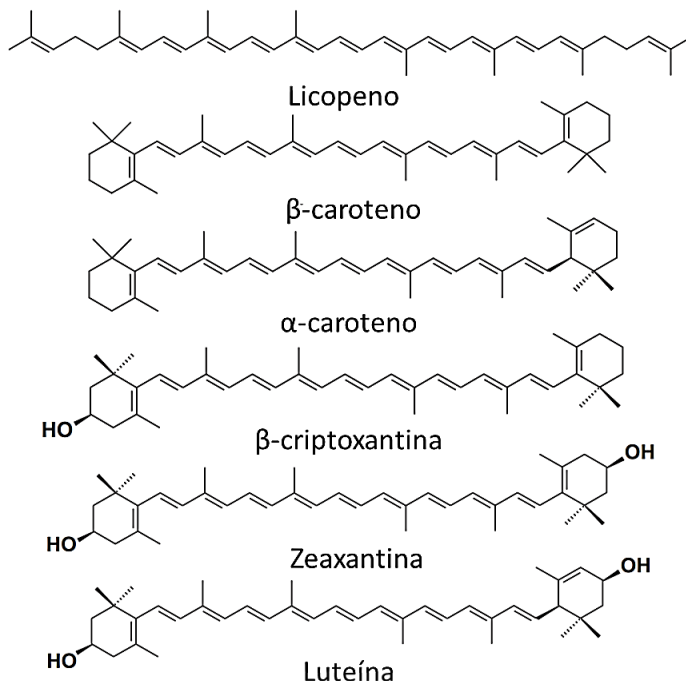
Com o resfriamento da pasta de amido, após sua gelatinização, as cadeias de amilose e amilopectina perdem energia e as ligações de hidrogênio tornam-se mais fortes, induzindo a associação, via ligações de hidrogênio, de moléculas de amilose, formando uma estrutura ordenada; um processo conhecido como retrogradação ou *setback*. Com o envelhecimento do gel, ou em ciclos de congelamento e descongelamento, as cadeias tendem a interagir mais fortemente entre si e, assim, expulsam a água do sistema, um fenômeno denominado sinérese (AMANTE, 1986).

2.6 CAROTENOIDES

Os carotenoides são grupos de pigmentos com mais de 700 componentes conhecidos, geralmente tetraterpenoides (C40), constituídos pela união de oito unidades isoprenoides (C5). Os terpenos apresentam uma composição molecular básica $C_{10}H_{15}$ e estão subdivididos em subclasses, em função do número de unidades de carbonos em suas estruturas, a saber: monoterpenos (C10), sesquiterpenos (C15), diterpenos (C20), triterpenos (C30) e tetraterpenos (C40). Há dois grandes grupos de carotenoides de acordo com suas estruturas químicas, i.e., os carotenos, constituídos apenas de

átomos de carbono e hidrogênio e, portanto, de natureza apolar, e as xantofilas, que são derivados oxigenados dos carotenos. As xantofilas são mais polares em relação aos carotenos, porque apresentam em suas estruturas químicas grupos funcionais como cetonas, aldeídos, epóxi e hidroxilas (Figura 3) (RODRIGUEZ-AMAYA, 2001).

Figura 3 - Estrutura química dos principais carotenoides encontrados em raízes de *M. esculenta* Crantz.



Fonte: Landrier; Marcotorchino e Tourniaire, (2012).

Dos mais de 700 carotenoides conhecidos, aproximadamente 50 são precursores da vitamina A, apresentando como característica estrutural em comum pelo menos um anel β -ionona não substituído, com cadeia lateral poliênica, com um mínimo de 11 carbonos. Entre os carotenoides, o β -caroteno é o mais abundante em alimentos e com maior atividade provitamina A. Embora aproximadamente 40 carotenoides possam ser absorvidos, metabolizados e usados pelo organismo humano, apenas 6 são usualmente detectados no plasma humano. Este grupo inclui α - e β -caroteno, licopeno, β -criptoxantina e luteína (FERNÁNDEZ-GARCÍA et al., 2012).

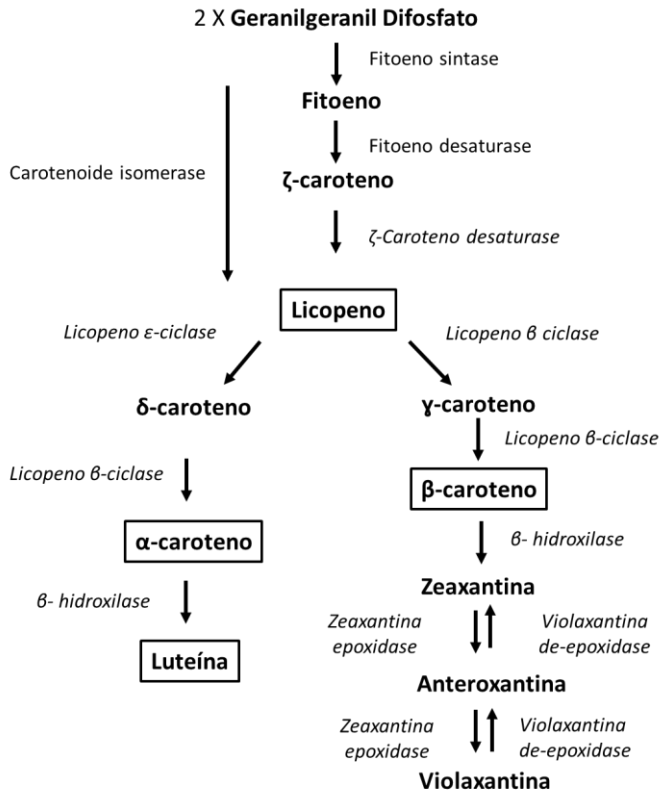
Tanto os carotenoides precursores de vitamina A quanto os não precursores, como a luteína e o licopeno, apresentam ação protetora contra o câncer. Possíveis mecanismos que explicam tal atividade referem-se ao efeito de sequestro de radicais livres, modulação do metabolismo do carcinoma, inibição da proliferação celular, aumento da diferenciação celular via retinoides e aumento da resposta imune (STAHL; SIES, 2003).

2.6.1 *Manihot esculenta* como fonte de carotenoides provitamina A

Os primeiros dados de variabilidade no conteúdo de carotenoides em mandioca foram publicados por autores indianos (MOORTHY et al., 1990). Os valores reportados por aqueles autores não excederam a 8 μg de β -caroteno/g de raízes frescas. Iglesias et al., (1997) realizaram no CIAT (Centro Internacional de Agricultura Tropical – Colômbia), o primeiro *screening* do conteúdo de carotenoides em banco de germoplasma mundial de mandioca. Ao analisar 632 acessos da coleção, verificaram uma significativa correlação ($r=0,82$) entre a cor da raiz e o conteúdo de carotenoides, tendo sido detectada uma ampla variação de concentração (1 $\mu\text{g g}^{-1}$ a 24 $\mu\text{g g}^{-1}$) em raízes frescas. Os cinco genótipos que apresentaram as maiores concentrações de β -caroteno foram coletados na região da floresta amazônica do Brasil e da Colômbia. Neste mesmo estudo, verificou-se que a presença de carotenoides nas raízes é determinada por dois pares de genes, sendo um responsável pelo transporte dos carotenoides às raízes e outro pelo acúmulo dos pigmentos. Estes dados são relevantes a estudos subsequentes de melhoramento genético da espécie utilizando técnicas de engenharia genética.

Estudos de biologia molecular têm identificado os promotores específicos do armazenamento de carotenoides nas raízes. Em tecidos de plantas que exibem baixos níveis de produção de carotenoides, o fator limitante é comumente a expressão no nível da fitoeno sintetase (PHY), que catalisa a primeira reação específica da carotenogênese, formando os precursores na biossíntese de carotenoides provitamina A, fitoeno e fitoflueno (CEBALLOS et al., 2012). As etapas finais da via biossintética dos carotenoides, destacando a formação dos principais carotenoides, são apresentadas na Figura 4.

Figura 4 - Vias biossintéticas de carotenoides em plantas superiores.



Fonte: Adaptado de Clotault et al. (2008).

Além do β -caroteno, *M. esculenta* contém outros carotenoides como a β -criptoxantina e outros não pró-vitamina A, e.g., licopeno e luteína, de reconhecidos benefícios à saúde humana, atuando principalmente como agentes antioxidantes (MONTAGNAC; DAVIS; TANUMIHARDJO, 2009, STAHL; SIES, 2003).

Em raízes de mandioca com polpa de coloração rosada, o licopeno destaca-se como o carotenoide majoritário. Genótipos com teores consideráveis deste pigmento têm sido cultivados em pequena escala por agricultores indígenas na região amazônica. Estes genótipos apresentam mutações que impedem a conversão de licopeno a β -caroteno (NASSAR, 2007).

2.6.2 Biofortificação

Três estratégias têm sido tradicionalmente usadas à prevenção da hipovitaminose A: diversificação alimentar, fortificação de alimentos e suplementação. Estas estratégias têm sido pouco efetivas e têm falhado na erradicação completa do problema em diversas populações e etnias no mundo (SALTZMAN et al., 2013). Nos últimos anos, programas e instituições de pesquisas têm iniciado a implementação de uma quarta estratégia, i.e., a biofortificação, para o incremento de micronutrientes nas espécies alimentares de cultivo local, apresentando como vantagem o fato de não modificar drasticamente os hábitos alimentares das populações com deficiências nutricionais. O objetivo da biofortificação é buscar alimentos mais nutritivos, mas que também tenham boas condições agrônomicas e aceitação no mercado (CEBALLOS et al., 2012).

A biofortificação, elo entre o conhecimento técnico-científico da agronomia e da saúde, oferece um meio viável para atingir populações rurais malnutridas, que podem ter acesso limitado às dietas diversificadas, suplementos e alimentos fortificados comercialmente. A principal estratégia consiste em aumentar os teores de micronutrientes em alimentos que já são da preferência agrônômica e de consumo das populações. A etapa inicial consiste na identificação de genótipos, preservados por estas populações e/ou mantidos em bancos de germoplasma, com teores superiores de metabólitos-alvo, para que possam ser explorados via melhoramento genético (SALTZMAN et al., 2013).

Da mesma maneira, não se espera que a biofortificação venha a tratar completamente ou eliminar a deficiência de micronutrientes em todos os grupos populacionais. Usualmente, uma intervenção isolada não resolve o problema da desnutrição de micronutrientes, mas a biofortificação complementa as intervenções existentes, colaborando por aumentar a adequação diária de ingestão de micronutrientes às populações mais vulneráveis (SALTZMAN et al., 2013).

2.6.3 Biofortificação de *Manihot esculenta*

A mandioca é considerada um alimento tradicional para muitas populações com risco de deficiência de vitamina A, sendo produzida predominantemente em pequena escala por agricultores com recursos limitados. Na mandioca, o β -caroteno é o carotenoide majoritário com atividade provitamina A (RODRIGUEZ-AMAYA, 2010), porém, concentrações de carotenoides em raízes de mandiocas tipicamente brancas (mais comumente encontradas nestas populações) são baixas, i.e., $\sim 1 \mu\text{g g}^{-1}$ de peso fresco e $3 \mu\text{g g}^{-1}$ de peso seco (IGLESIAS et al.,

1997). Sabe-se, contudo, que a coloração amarela de suas raízes apresenta alta correlação com o teor de carotenoides totais e na busca por genótipos com maior atividade provitamina A, valores superiores a $10 \mu\text{g g}^{-1}$ de β -caroteno/g de peso fresco têm sido encontrados (CHÁVEZ et al., 2005).

O progresso em incrementar o conteúdo de carotenoides nas raízes de mandioca foi significativo nos últimos anos. Níveis superiores de carotenoides totais foram reportados por Montagnac; Davis e Tanumihardjo (2009) em mandioca biofortificada ($20 \mu\text{g g}^{-1}$).

Pesquisas exploratórias realizadas pelo CIAT (Centro Internacional de Agricultura Tropical) identificaram, na América do Sul, fontes de germoplasma com altos teores de carotenoides. A partir desses genótipos variedades adaptadas ao ambiente africano foram obtidas, inclusive resistentes ao vírus do mosaico da mandioca, e lançadas na Nigéria, em 2011. Estas variedades apresentam atividade provitamina A suficiente para proporcionar aproximadamente 25 % de atividade equivalente de retinol (RAE) em mulheres e crianças com idade escolar, naquela região (SALTZMAN et al., 2013).

Welsch et al. (2010) mostraram em indivíduos de *M. esculenta* modificadas geneticamente que a superexpressão de genes *PHY* favorece o fluxo de carbono à carotenogênese. Comparativamente a genótipos de raízes brancas ($0,65 \pm 0,06 \mu\text{g. g}^{-1}$ de carotenoides), foi demonstrado que plantas *PHY* mutantes produziam raízes de polpa amarela, com concentrações daqueles pigmentos bastante superiores ($21,84 \pm 0,84 \mu\text{g. g}^{-1}$).

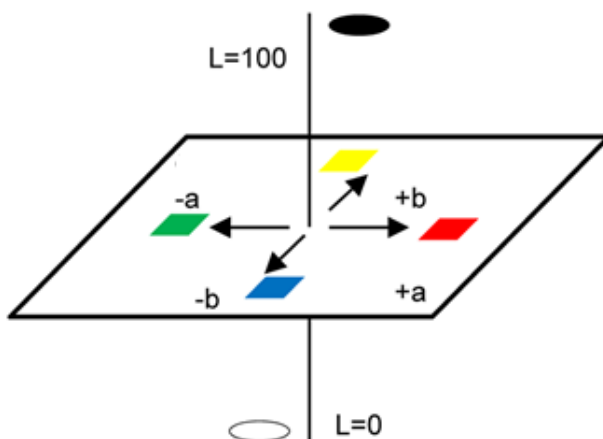
2.7 ESPAÇO DE CORES CIELAB E CAROTENOIDES

A cultura da mandioca caracteriza-se pela grande variabilidade de coloração das raízes, as quais variam do branco ao rosado. A definição da cor está relacionada à presença de diversos pigmentos carotenoides e ao balanço entre eles. Como anteriormente mencionado, a intensidade de cor amarela das raízes de mandioca relaciona-se positivamente com o conteúdo de carotenoides. Entretanto, a possibilidade de se adotar a cor de raízes como critério indireto para seleção visando maior conteúdo de caroteno é questionável, já que a cor é uma característica de difícil avaliação visual, com subjetividades inerentes.

Com o objetivo de normalizar a medição da cor, em 1931 a CIE (*Commission Internationale de L'Eclairage*) adotou os seguintes métodos para medição e especificação de cor: uso de fontes de luz-padrão definidas pela CIE, condições exatas à observação ou medição da cor, uso de unidades matemáticas apropriadas para expressar a cor e

definição do observador-padrão. Em 1976, a CIE recomendou o uso da escala de cor CIE $L^*a^*b^*$ ou CIELAB. Este sistema é baseado em três elementos: a luminosidade ou claridade, a tonalidade ou matiz e a saturação ou cromaticidade. O valor de a^* caracteriza a coloração nas regiões do vermelho ($+a^*$) ao verde ($-a^*$), o valor b^* indica coloração no intervalo do amarelo ($+b^*$) ao azul ($-b^*$). O valor L fornece a luminosidade, variando do branco ($L=100$) ao preto ($L=0$). A Figura 5 apresenta um diagrama representativo do espaço de cores CIELAB.

Figura 5 - Diagrama representativo do espaço de cores CIELAB.



Fonte: Liu et al. (2014).

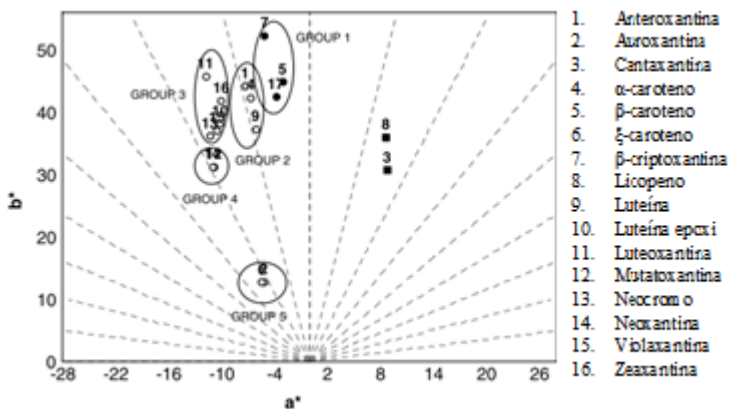
A escala de cores CIELAB pode ser usada em qualquer objeto cuja cor pode ser medida. É atualmente o sistema mais utilizado à descrição quantitativa da cor de um objeto devido a sua uniformidade, facilidade de aquisição e por ser uma técnica de muito baixo custo.

As avaliações químicas que consistem na extração, caracterização e quantificação dos pigmentos carotenoides, especialmente a análise espectrofotométrica e a cromatografia líquida de alta eficiência (CLAE), são bastante precisas, mas de maior custo e consumo de tempo. A medição objetiva da cor é uma técnica não destrutiva e muito rápida que permite obter uma série de variáveis em poucos segundos, o que facilita sua medida à campo e evita a degradação de compostos de interesse quando da extração destes.

Meléndez-Martínez et al. (2006) estudaram as localizações das soluções dos pigmentos carotenoides em acetona no plano $a^* b^*$ (Figura 6). Os carotenoides avermelhados, licopeno e cantaxantina, apresentam-se localizados no primeiro quadrante do plano (valores

positivos de a^* e b^*). Os outros carotenoides foram agrupados pelo número de ligações duplas conjugadas em suas moléculas. Para carotenoides que apresentam cor amarela é observado que valores de a^* claramente diminuem em amostras com 11 duplas ligações, como o β -caroteno, a β -criptoxantina e a zeaxantina, em relação às amostras com 9 ligações duplas conjugadas.

Figura 6 - Localização dos carotenoides no plano CIE $L^*a^*b^*$.

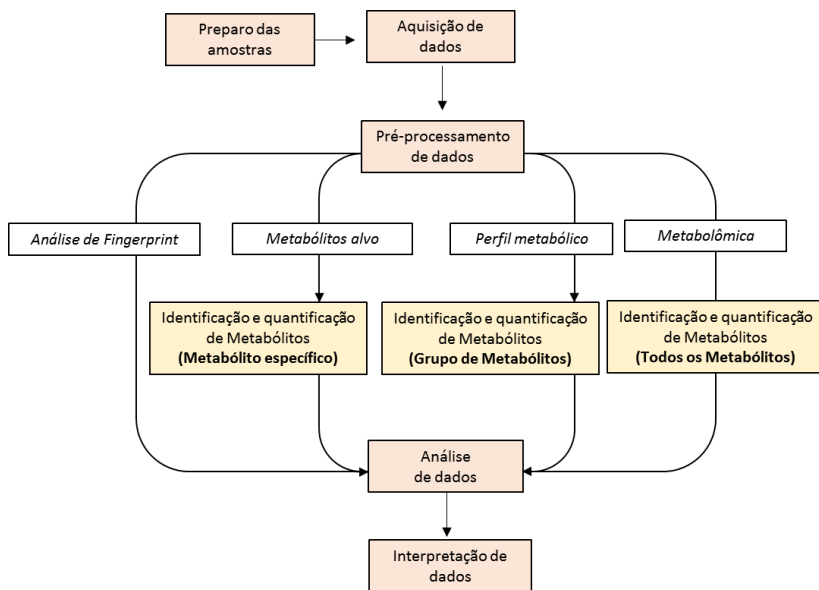


Fonte: Adaptado de Meléndez-Martínez et al. (2006).

2.8 METABOLÔMICA

Os termos 'genoma', 'transcriptoma' e 'proteoma' referem-se ao DNA total, ao conjunto de transcritos de RNAm e de proteínas de uma célula, respectivamente. De forma análoga, 'metaboloma' é o termo que designa o conjunto de metabólitos de baixa massa molecular (i.e., < 1000Da) presentes em uma célula, necessários à sua manutenção, crescimento e funcionamento normal (GROMSKI et al., 2015). Devido ao alto grau de diversidade dos metabólitos primários e secundários, ainda não é tecnologicamente possível identificá-los e quantificá-los em sua totalidade utilizando uma única técnica analítica (ZHANG et al., 2012). Por isto, algumas estratégias têm sido utilizadas na determinação do metaboloma, a saber: análise de metabólitos-alvo (*target analysis*), perfil metabólico (*metabolite profiling*) e análise de *fingerprint* (*metabolic fingerprinting*) (FIEHN, 2002). Um esquema de trabalho típico de estudos metabolômicos é mostrado na Figura 7.

Figura 7 - Fluxograma geral da abordagem de estudos do metaboloma



Fonte: Adaptado de Afonso (2017).

O estudo de metabólitos-alvo considera a análise qualitativa e quantitativa de um ou poucos compostos relacionados a uma via metabólica específica. A análise do perfil de metabólitos visa à identificação e quantificação de diversos grupos de metabólitos, geralmente utilizando as cromatografias líquida (LC) e gasosa (GC) muitas vezes associadas à espectrometria de massas (MS), ou à ressonância magnética nuclear (NMR). Esta abordagem aumenta o poder de detecção e de resolução da análise, além de possibilitar a investigação concomitante de vários grupos de metabólitos. Por sua vez, a análise do *fingerprint* de metabólitos utiliza métodos analíticos mais rápidos, como as técnicas espectroscópicas e.g. FTIR, Raman e UV-vis e consiste em uma análise global que usualmente dispensa etapas complexas de preparação das amostras e, eventualmente, a necessidade de reagentes. Nesta abordagem, de uma maneira geral é possível analisar tanto metabólitos primários (carboidratos, aminoácidos, ácido graxos) quanto secundários (polifenóis, carotenoides, alcaloides, entre outros (DUNN; ELLIS, 2005).

2.9 FERRAMENTAS DE BIOINFORMÁTICA

Os recentes avanços em dados ômicos e, em particular, em metabolômica, através de técnicas experimentais de elevado rendimento e métodos de análise de dados têm fornecido novas abordagens para muitas questões de pesquisa na área biológica e impulsionado o desenvolvimento tecnológico em biomedicina e biotecnologia. De fato, a capacidade de identificar e quantificar um grande número de compostos em amostras biológicas complexas proporciona avanços significativos na genômica funcional, identificação de biomarcadores, caracterização da amostra e controle de qualidade (e.g. alimentos e bebidas), ou a descoberta e desenvolvimento de medicamentos.

Independentemente da estrutura dos dados, a complexidade de dados metabolômicos requer sofisticados métodos de análises, geralmente de natureza multivariada. Neste contexto, inúmeras ferramentas de bioinformática têm sido desenvolvidas para gerenciar, processar, minerar, integrar, analisar estatisticamente e até criar modelos matemáticos de redes metabólicas (ROESSNER; BOWNE, 2009).

Dentro da análise multivariada, duas famílias de métodos são frequentes nos estudos metabolômicos: a análise não-supervisionada e a análise supervisionada (ou análise discriminante). A análise não-supervisionada considera o uso de amostras sobre as quais nenhum conhecimento/informação prévio(a) é requerido(a) e busca agrupá-las pela similaridade. Dentre os métodos de análise não-supervisionada mais utilizados na metabolômica estão as análises de componentes principais (PCA - *Principal Component Analysis*), de clusters hierárquicos (HCA- *Hierarchical Clustering Analysis*) e *k-means*.

O uso de técnicas não supervisionadas é recomendável como um ponto de partida na avaliação de dados, gerando “insights” relevantes quanto à estrutura de dados (grupos, e/ou *outliers*), importantes no processo seguinte de desenvolvimento de modelos de classificação e/ou calibração (TAYLOR; RATON, 2009). Em geral, os resultados de PCA e HCA são complementares e quando utilizados em “tandem” constituem uma ferramenta interessante à construção de modelos de reconhecimento confiáveis.

Quando os grupos ou as classes já estão definidos, o passo seguinte envolve a identificação de importantes informações (que justifiquem e expliquem tais agrupamentos), como a identificação de metabólitos ou picos. Na análise discriminante, inicialmente um conjunto de amostras rotuladas é utilizado pelo algoritmo estatístico como treino, para que este seja capaz de classificar corretamente novas amostras a partir das amostras conhecidas (JAIN; MURTY; FLYNN,

1999). Este grupo de análises é conhecido como *machine learning*, ou "aprendizado de máquina".

2.9.1 Machine learning

Técnicas de *machine learning* são frequentemente adotadas em estudos metabolômicos onde a determinação do *fingerprint* metabólico é pretendida. Para tal, algoritmos são desenvolvidos, os quais aprendem e criam previsões sobre dados, via construção de um modelo a partir da adição (*inputs*) de amostras conhecidas. Para construir um modelo preditivo, um conjunto de dados de treinamento é fornecido e o algoritmo deverá ter a capacidade de generalizar a partir destas experiências, realizando previsões precisas sobre exemplos novos e não computados anteriormente.

Entre os mais populares métodos supervisionados à construção de modelos de classificação (i.e., dados de natureza discreta) em estudos metabolômicos citam-se: *Partial Least Squares-Discriminant Analysis* (PLS-DA), *Linear Discriminant Analysis* (LDA), *Soft Independent Modelling by Class Analogy* (SIMCA) e *Decision Trees*. De outra forma, quando os dados são de natureza contínua utilizam-se modelos de regressão que incluem: *Partial Least Squares Regression* (PLS-r) e *Regression Trees*. Métodos como *K-Nearest Neighbors* (KNN), *Support Vector Machines* (SVM), *Artificial Neural Network* (ANN) e *Random Forests* podem ser usados como modelos de classificação e regressão (VARMUZA; FILZMOSER, 2009).

2.9.2 Data fusion

A fusão de dados (*data fusion*) é um processo de combinação de dados de diferentes fontes que busca melhorar o desempenho de modelos de previsão. Assim, os modelos de análise buscam a associação, detecção, correlação e estimativa de dados para gerar uma melhor informação do estado do sistema. Nesta abordagem, os dados são coletados por "n" tipos de fontes distintas, e.g. IR, Raman e UV-vis.

Os dados podem então ser pré-processados para extrair um vetor de características que represente os dados observados, seguido de uma abordagem de aprendizado de máquina usando os objetos observados. A resultante deste processo deve ser particionada em grupos que representam observações pertencentes à mesma categoria. Os algoritmos de fusão combinam dados de fontes múltiplas para obter um resultado mais preciso em relação ao uso destes individualmente.

2.9.3 Specmine

Vários métodos e ferramentas à análise de dados espectrais têm surgido, sendo que uma das maiores limitações enfrentadas é a falta de estruturas integradas que permitam a extração de informações relevantes do conjunto de dados e sua integração com o conhecimento bioquímico prévio. A falta de reprodutibilidade em muitos processos de análise é também um forte obstáculo na descoberta biológica, sendo comum a carência de *pipelines* de análise nos trabalhos atualmente publicados.

Visando diminuir esta lacuna, um *package* em linguagem R foi desenvolvido, a saber o *Specmine - Metabolomics and Spectral Data Analysis and Mining* - (COSTA; MARASCHIN; ROCHA, 2015). O *Specmine* constitui uma ferramenta à análise e extração de informações de dados espectrais, de forma dedicada aos estudos metabolômicos. O *package* possibilita o carregamento de dados em diferentes formatos, pré-processamento, identificação de metabólitos, análise estatística univariada e multivariada de dados, aprendizado de máquina (*machine learning*) e a identificação de variáveis (*feature selection*) relevantes. A ferramenta utiliza diversas funções em código R, integrando vários pacotes através de uma interface única. O desenvolvimento do *specmine* com seu conjunto integrado de funções, permite obter um *pipeline* robusto e confiável de análise de dados, suportando a extração de conhecimento relevante de dados de metabolômica.

3 REFERÊNCIAS

AFONSO, P. T. **Development of web-based tools for spectral data analysis and mining**. Master Degrees in Bioinformatics - Minho University - Braga - Portugal. October, 2017.

AMANTE, E. R. **Caracterização de amidos de variedades de mandioca (*Manihot esculenta*, Crantz) e da batata doce (*Ipomoea batatas*)**. Dissertação. Ciência e Tecnologia de Alimentos. Universidade de Viçosa, Minas Gerais. Brasil, 1986.

BEHALL, K. M.; SCHOLFIELD, D. J. Food Amylose Content Affects Postprandial Glucose and Insulin Responses. **Cereal Chemistry Journal**, v. 82, n. 6, p. 654–659, 1 nov. 2005.

BEMILLER, J. N.; LAFAYETTE, W.; IN, U. S. A. 1 U bersichtsbeitrag / **Review Starch Modification : Challenges and Prospects**. v. 49, n. 4, p. 127–131, 1997.

BLAGBROUGH, I. S. et al. Cassava: an appraisal of its phytochemistry and its biotechnological prospects. **Phytochemistry**, v. 71, n. 17–18, p. 1940–51, dez. 2010.

BOLHUIS, G. G. The toxicity of cassava roots. **Netherlands Journal of Agricultural Science**, v. 2, p. 176–186, 1954.

CARVALHO, L. M. J. et al. Retention of total carotenoid and β -carotene in yellow sweet cassava (*Manihot esculenta* Crantz) after domestic cooking. **Food & nutrition research**, v. 56, p. 1–8, 2012.

CASTELLANO, F. **Território e desenvolvimento rural: o caso dos engenhos de farinha de mandioca do Litoral Centro-sul de Santa Catarina**. 2015. 120 f. Dissertação (Mestrado em Agroecossistemas) – Programa de Pós-Graduação em Agroecossistemas, Universidade Federal de Santa Catarina, Florianópolis, 2015.

CEBALLOS, H. et al. Spatial distribution of dry matter in yellow fleshed cassava roots and its influence on carotenoid retention upon boiling. **Food Research International**, v. 45, n. 1, p. 52–59, 2012.

- CEREDA, M. P. et al. **Propriedades gerais do amido**. Serie: culturas de tuberosas amiláceas latino-americanas, v. 1. São Paulo, Brasil: Fundação Cargill, 2001.
- CHÁVEZ, A. L. et al. Variation of quality traits in cassava roots evaluated in landraces and improved clones. **Euphytica**, v. 143, n. 1–2, p. 125–133, 2005.
- CLOTAULT, J. et al. Expression of carotenoid biosynthesis genes during carrot root development. **Journal of Experimental Botany**, v. 59, n. 13, p. 3563–3573, 2008.
- COSTA, C.; MARASCHIN, M.; ROCHA, M. An R package for the integrated analysis of metabolomics and spectral data. **Computer Methods and Programs in Biomedicine**, v. 129, p. 117–124, 2015.
- DEMIATE, I. M. et al. Características viscoamilográficas de amidos de mandioca quimicamente modificados. Publication UEPG: **Ciências Exatas e da Terra, Ciências Agrárias e Engenharias**, v. 11, n. 1, p. 7–17, 2005.
- DENARDIN, V. F.; KOMARCHESKI, R. FARINHEIRAS DO BRASIL: **Tradição, cultura e perspectivas da Produção Familiar de Farinha de Mandioca**, v. 1, 297 p., 2015.
- DONALD, A. M. et al. Scattering Studies of the Internal Structure of Starch Granules. **Starch - Stärke**, v. 53, n. 10, p. 504–512, 2001.
- DONGOWSKI, G. R. L. WHISTLER AND J. N. BEMILLER: Carbohydrate Chemistry for Food Scientists., Eagan Press, St. Paul, MN, USA, 1997. **Food / Nahrung**, v. 41, n. 5, p. 318, 1997.
- DUNN, W. B.; ELLIS, D. I. Metabolomics: Current analytical platforms and methodologies. **TrAC - Trends in Analytical Chemistry**, v. 24, n. 4, p. 285–294, 2005.
- ELIASSON, A. C. Starch in food: structure, function and applications. 605 p. **CRC Press, Boca Raton, FL**, 2004.
- FAO. Save and Grow: Cassava. A Guide to Sustainable Production Intensification. [s.l.] **Food and Agriculture Organisation**, 2013.

FERNÁNDEZ-GARCÍA, E. et al. Carotenoids bioavailability from foods: From plant pigments to efficient biological activities. **Food Research International**, v. 46, n. 2, p. 438–450, maio 2012.

FIEHN, O. Metabolomics - the link between genotypes and phenotypes. **Plant Molecular Biology**, v. 48, p. 155–171, 2002.

FUKUDA W.M.G.; PEREIRA, M. E. C. BRS Rosada – Mandioca de mesa com raiz colorida e mais nutritiva. **Folder BRS Rosada - Embrapa Mandioca e Fruticultura Tropical**, Cruz das Almas, Bahia, 2006.

GOMAND, S. V. et al. Structural properties and gelatinisation characteristics of potato and cassava starches and mutants thereof. **Food Hydrocolloids**, v. 24, n. 4, p. 307–317, jun. 2010.

GROMSKI, P. S. et al. A tutorial review: Metabolomics and partial least squares-discriminant analysis - a marriage of convenience or a shotgun wedding. **Analytica Chimica Acta**, v. 879, p. 10–23, 2015.

HIGGINS, J. A et al. Resistant starch consumption promotes lipid oxidation. **Nutrition & metabolism**, v. 1, p. 8, 2004.

HOOVER, R. Composition, molecular structure, and physicochemical properties of tuber and root starches: A review. **Carbohydrate Polymers**, v. 45, n. 3, p. 253–267, 2001.

HOWELER, R. H.; OATES, C. G.; ALLEM, A. C. **Impact of cassava processing on the environment**. Proceedings of the validation forum on the global cassava development strategy strategic environmental assessment, p. 136, 2001.

HU, G.; BURTON, C.; YANG, C. Efficient measurement of amylose content in cereal grains. **Journal of Cereal Science**, v. 51, n. 1, p. 35–40, 2010.

HUGHES, D. A. Dietary carotenoids and human immune function. **Nutrition**, v. 17, n. 10, p. 823–827, 2001.

IGLESIAS, C. et al. Genetic potential and stability of carotene content in cassava roots. **Euphytica**, v. 94, n. 3, p. 367–373, 1997.

- JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **ACM Computing Surveys**, v. 31, n. 3, p. 264–323, 1999.
- JANE, J.-L. et al. Anthology of Starch Granule Morphology by Scanning Electron Microscopy. **Starch - Stärke**, v. 46, n. 4, p. 121–129, 1994.
- KIM, Y. S. et al. Screening Potato Starch for Novel Properties Using Differential Scanning Calorimetry. **Journal of Food Science**, v. 60, n. 5, p. 1060–1065, 1995.
- LANDRIER, J. F.; MARCOTORCHINO, J.; TOURNIAIRE, F. Lipophilic micronutrients and adipose tissue biology. **Nutrients**, v. 4, n. 11, p. 1622–1649, 2012.
- LEACH, H. W.; SCHOCH, T. J. Starch Solubility DMSO. **Cereal chemistry**, v. 39, 318-326, 1962.
- LIU, Q. Understanding Starches and Their Role in Foods. In: Food Carbohydrates. [s.l.] **CRC Press**, 2005.
- LIU, W. et al. Optimal color design of psychological counseling room by design of experiments and response surface methodology. **PLoS ONE**, v. 9, n. 3, p. 1–9, 2014.
- AMORIM, S. et al. Intoxicações por plantas cianogênicas no brasil (Intoxication by cyanogenic plants in Brazil). **Ciência Animal**, v. 16, n. 161, p. 17–2617, 2006.
- MATA, A. P. et al. Levantamento Sistemático Da Produção Agrícola. IBGE, p. 115, 2017.
- MELÉNDEZ-MARTÍNEZ, A. J. et al. Relationship between the colour and the chemical structure of carotenoid pigments. **Food Chemistry**, v. 101, n. 3, p. 1145–1150, 2006.
- MONTAGNAC, J. A.; DAVIS, C. R.; TANUMIHARDJO, S. A. Nutritional Value of Cassava for Use as a Staple Food and Recent Advances for Improvement. **Comprehensive Reviews in Food Science and Food Safety**, v. 8, 2009.

- MOORTHY, S. et al. Variability of β -carotene in cassava germplasm. **Food Chemistry**, v. 36, p. 233–236, 31 dez. 1990.
- NASSAR, N. et al. Cassava diversity in Brazil: the case of carotenoid-rich landraces. **Genetics and molecular research: GMR**, v. 6, n. 1, p. 116–21, jan. 2007.
- NASSAR, N. M. A. Wild and indigenous cassava, *Manihot esculenta* Crantz diversity: An untapped genetic resource. **Genetic Resources and Crop Evolution**, v. 54, n. 7, p. 1523–1530, 11 abr. 2007.
- NASSAR, N. M. A.; HASHIMOTO, D. Y. C.; FERNANDES, S. D. C. Wild *Manihot* species: Botanical aspects, geographic distribution and economic value. **Genetics and Molecular Research**, v. 7, n. 1, p. 16–28, 2008.
- NELSON, D. L.; COX, M. M. **Princípios de Bioquímica**, Quarta edição. Artmed Editora S.A. Porto Alegre. Brasil, 2011.
- OLSEN, K. M.; SCHAAL, B. A. Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 10, p. 5586–91, 1999.
- PALOZZA, P. Can β -carotene regulate cell growth by a redox mechanism? An answer from cultured cells. **Biochimica et Biophysica Acta - Molecular Basis of Disease**, v. 1740, n. 2, p. 215–221, 2005.
- PÉREZ, S.; BERTOFT, E. The molecular structures of starch components and their contribution to the architecture of starch granules: A comprehensive review. **Starch - Stärke**, v. 62, n. 8, p. 389–420, 5 jul. 2010.
- PERONI, F. H. G.; ROCHA, T. S.; FRANCO, C. M. L. Some Structural and Physicochemical Characteristics of Tuber and Root Starches. **Food Science and Technology International**, v. 12, n. 6, p. 505–513, 2006.
- PINTO-ZEVALLOS, D. M.; PAREJA, M.; AMBROGI, B. G. Current knowledge and future research perspectives on cassava (*Manihot esculenta* Crantz) chemical defenses: An agroecological view. **Phytochemistry**, v. 130, p. 10–21, 2016.

R CORE TEAM. R: A Language and Environment for Statistical Computing, 2014. Disponível em: <<http://www.r-project.org/>>

RATNAYAKE, W. S.; JACKSON, D. S. Gelatinization and Solubility of Corn Starch during Heating in Excess Water: New Insights. **Journal of Agricultural and Food Chemistry**, v. 54, n. 10, p. 3712–3716, 2006.

Relatório de Mercado Agrícola CEASA/SC. **Relatório de Mercado Agrícola CEASA/SC**, v. 5, n. Abril/2017, p. 32, 2017.

RODRIGUEZ-AMAYA, D. B. Quantitative analysis, in vitro assessment of bioavailability and antioxidant activity of food carotenoids—A review. **Journal of Food Composition and Analysis**, v. 23, n. 7, p. 726–740, nov. 2010.

ROESSNER, U.; BOWNE, J. What is metabolomics all about? **BioTechniques**, v. 46, n. 5 SPEC. ISSUE, p. 363–365, 2009.

SALTZMAN, A. et al. Biofortification: Progress toward a more nourishing future. **Global Food Security**, v. 2, n. 1, p. 9–17, 2013.

SINGH, N. et al. Morphological, thermal and rheological properties of starches from different botanical sources. **Food Chemistry**, v. 81, n. 2, p. 219–231, 2003.

STAHL, W.; SIES, H. Antioxidant activity of carotenoids. **Molecular Aspects of Medicine**, v. 24, n. 6, p. 345–351, 2003.

SWINKELS, J. J. M. Composition and Properties of Commercial Native Starches. **Starch - Stärke**, v. 37, n. 1, p. 1–5, 1985.

TAIZ, L.; ZEIGER, E. **Fisiologia Vegetal**. Porto Alegre, Editora Artmed, 4ª. ed., 2009.

TESTER, R. F.; KARKALAS, J.; QI, X. Starch—composition, fine structure and architecture. **Journal of Cereal Science**, v. 39, n. 2, p. 151–165, mar. 2004.

TESTER, R. F.; MORRISON, W. R. Swelling and gelatinization of cereal starches. I. Effects of amylopectin, amylose, and lipids. **Cereal Chemistry**, v. 67, n. 6, p. 551–557, 1990.

UARROTA, V. G. et al. Advances in understanding cassava growth and development. 2017. **Burleigh Dodds Series in Agricultural Science**. 1ed. Cambridge: Burleigh Dodds Science Publishing, 2017, p. 3-35.

VARMUZA, K.; FILZMOSER, P. **Introduction to multivariate statistical analysis in chemometrics**. CRC press, Boca Raton, Florida, 1 edition. 2009.

WEDNER, S. H.; ROSS, D. A. **Vitamin A Deficiency and Its Prevention**. A2 - Quah, Stella R. BT - International Encyclopedia of Public Health (Second Edition). In: Oxford: Academic Press, 2017. p. 378–382.

WELSCH, R. et al. Provitamin A accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. **The Plant cell**, v. 22, n. 10, p. 3348–56, out. 2010.

WHISTLER, R. L.; DANIEL, J. R. Starch. In: Kirk-Othmer Encyclopedia of Chemical Technology. [s.l.] John Wiley & Sons, Inc., 2000.

WICKER, L. Hydrocolloids – practical guides for the food industry. **Journal of Texture Studies**, v. 35, n. 5, p. 570–571, 2004.

ZEEMAN, S. C. et al. Starch Synthesis in Arabidopsis . Granule Synthesis , Composition , and Structure 1. **Plant physiology**, v. 129, n. June, p. 516–529, 2002.

ZHANG, A. et al. Modern analytical techniques in metabolomics analysis. **The Analyst**, v. 137, n. 2, p. 293–300, 2012.

ZHU, F. Composition, structure, physicochemical properties, and modifications of cassava starch. **Carbohydrate Polymers**, v. 122, p. 456–480, 2015.

ZOBEL, H. F.; YOUNG, S. N.; ROCCA, L. A. Starch Gelatinization: An X-ray Diffraction Study?. **Cereal chemistry**, v. 65, n. 6, p. 443–446, 1988.

CAPÍTULO II

UV-VISIBLE SCANNING SPECTROPHOTOMETRY AND CHEMOMETRIC ANALYSIS AS TOOLS FOR CAROTENOIDS ANALYSIS IN CASSAVA GENOTYPES (*Manihot esculenta* Crantz)

**Artigo completo publicado no *Journal of Integrative
Bioinformatics*. Doi: 10.2390/biecoll-jib-2015-280**

UV-visible scanning spectrophotometry and chemometric analysis as tools for carotenoids analysis in cassava genotypes (*Manihot esculenta* Crantz)

Rodolfo Moresco^{1*}, Virgílio G. Uarrota¹, Aline Pereira¹, Máira Tomazzoli¹, Eduardo da C. Nunes², Luiz Augusto Martins Peruch², Jussara Gazzola³, Christopher Costa⁴, Miguel Rocha⁴ and Marcelo Maraschin¹

¹Plant Morphogenesis and Biochemistry Laboratory, Federal University of Santa Catarina, Florianópolis, Brazil <http://ufsc.br>

²Santa Catarina State Agricultural Research and Rural Extension Agency (EPAGRI), Experimental Station of Urussanga, Urussanga, Brazil <http://www.epagri.sc.gov.br>

³Experimental Nutrition Laboratory, Federal University of Santa Catarina, Florianópolis, Brazil, <http://ufsc.br>

⁴Centre of Biological Engineering, University of Minho, Braga, Portugal, <http://www.uminho.pt>

Summary

In this study, the metabolomics characterization focusing on the carotenoid composition of ten cassava (*Manihot esculenta*) genotypes cultivated in southern Brazil by UV-visible scanning spectrophotometry and reverse phase-high performance liquid chromatography was performed. Cassava roots rich in β -carotene are an important staple food for populations with risk of vitamin A deficiency. Cassava genotypes with high pro-vitamin A activity have been identified as a strategy to reduce the prevalence of deficiency of this vitamin. The data set was used for the construction of a descriptive model by chemometric analysis. The genotypes of yellow-fleshed roots were clustered by the higher concentrations of *cis*- β -carotene and lutein. Inversely, cream-fleshed roots genotypes were grouped precisely due to their lower concentrations of these pigments, as samples rich in lycopene (red-fleshed) differed among the studied genotypes. The analytical approach (UV-Vis, HPLC, and chemometrics) used showed to be efficient for understanding the chemodiversity of cassava genotypes, allowing to classify them according to important features for human health and nutrition.

Keywords: Chemometrics, descriptive models, partial metabolome, cassava genotypes, carotenoids, RP-HPLC, UV-vis.

*To whom correspondence should be addressed. Email:

1 INTRODUCTION

Cassava (*Manihot esculenta* Crantz, 1766) currently ranks as the third most important species as a source of calories in the world among the group of staple food crops, including rice and maize. It is primarily consumed in places where there are prevailing drought, poverty, and mal-nutrition. Diseases related to vitamin A deficiency are among the major nutritional problems in developing countries. It is estimated that 190 million children in preschool age have low retinol activity in plasma ($\leq 0.70 \mu\text{mol.L}^{-1}$), subclinical symptom of this deficiency [1]. Cassava is considered a staple food for many populations with risk of vitamin A deficiency and is predominantly produced by small-scale farmers with limited resources. Carotenoids come from plants secondary metabolism and belong to the class of terpenes, which are classified by the number of C5 units and also known as tetraterpenes (with 40 carbons or 8 units of C5). The carotenoids are more stable in *trans* isomeric form and could be divided in carotenes and xanthophylls. The former are hydrocarbons carotenoids, such as α - and β -carotene and lycopene, while xanthophylls are oxygenated derivatives of carotenes and include the compounds lutein and β -cryptoxanthin, for instance [2].

Some carotenoids are precursors to vitamin A [3]. The enzymatically conversion of pro-vitamin A carotenoids into vitamin A occurs mainly in the intestinal mucosa during the absorption of those precursors, protecting against vitamin A deficiency. Several health-promoting effects of carotenoids such as immune-enhancement and reduction of the risk of developing degenerative diseases, cancer, cardiovascular diseases, cataract, and macular degeneration have been claimed [4] [5].

Pro-vitamin A carotenoids contrast in their vitamin A activities because of their distinct chemical structures, being important to identify which carotenoid compounds form the food matrix and to check their actual activity in conversion to vitamin A. In cassava, β -carotene is the majoritarian pro-vitamin A carotenoid [6], but the amounts found in white and cream cassava roots (the most commonly consumed by populations) are usually low in comparison to yellow ones [3].

It is known that roots with yellow flesh are highly correlated with the concentration of carotenoids and the search for materials with higher pro-vitamin A activity is recognized as a strategy to reduce the prevalence of this deficiency [7]. Besides β -carotene, *M. esculenta* also

contains small amounts of other carotenoids, e.g., lycopene and the xanthophylls, lutein and β -cryptoxanthin.

Because of the high importance of cassava crops in Brazil, genebanks of cassava collections have been established and maintained for the purpose of preserving wild genotypes, traditional landraces and commercial varieties. Thus, the identification and preservation of genotypes rich in carotenoids is thought to be relevant for the Brazilian and global food security and nutrition.

Spectroscopic methods using UV-Visible (UV-Vis) wavelengths are rapid, cheap, and provide metabolic fingerprints that can be processed enabling pattern recognition between samples. UV-Vis scanning spectrophotometry requires little sample amounts and preparation, and rapidly provides valuable and robust information about the presence of particular classes of metabolites, such as carotenoids. Chromatographic methods (e.g., reverse phase-high performance liquid chromatography – RP-HPLC) are also largely used for the identification and quantification of plant extracts' compounds. These compounds can be used as biochemical markers assisting to discriminate samples with peculiar characteristics.

This study, in connection with the metabolomics characterization of the genebank's cassava accesses, emphasizes the carotenoid profile in root samples using a typical analytical platform, i.e., UV-Vis spectrophotometry and reverse phase-high performance liquid chromatography (RP-HPLC). The data set afforded (i.e., the spectrophotometric profiles and the chromatographic quantification of each carotenoid compound) was used to build descriptive and classification models by calculation of the principal components and cluster analysis. Such an analytical approach allows the rapid and effective extraction of relevant and non-redundant information from a set of complex data, enabling a more detailed and robust understanding of possible differences and/or similarities in the studied samples, as well as their improved discrimination. In practical terms, this study develops and applies biotechnological approaches, by coupling the use of biochemical markers with bioinformatics tools to gain insights to more rationally support genetic breeding programs of cassava.

2 MATERIAL AND METHODS

2.1 SELECTION OF CASSAVA GENOTYPES

Roots of ten genotypes of *M. esculenta* (2012/2013 season) from the EPAGRI's germplasm bank (Urussanga Experimental Station, 28°31'18"S, 49°19'03"W, Santa Catarina, southern Brazil) were used in

this study. Traditionally, they have been called *Apronta mesa*, *Pioneira*, *Oriental*, *Amarela*, *Catarina*, *IAC576-70* - (*Instituto Agronômico de Campinas*), hereafter designed as *IAC5760-70*, *Salézio*, *Estação*, *Videira*, and *Rosada* and were selected based on their economic and social importance. The samples were classified considering a visual analysis of colors of root parenchymal tissues. Cream-fleshed and yellow-fleshed, with lower and higher intensity of the yellow color, respectively, and red-fleshed, samples with reddish color.

2.2 EXTRACTION AND ANALYSIS OF CAROTENOID COMPOUNDS

Carotenoids were extracted from fresh roots as described in 2004 by Rodriguez-Amaya & Kimura [2] using an Ultra-Turrax (Janke & Kunkel IKA - T25 basic) and organic solvents: acetone and petroleum ether (v/v, 1:1). The absorbances of the organosolvent extracts were collected on a UV-visible spectrophotometer (Gold Spectrum lab 53 UV-Vis spectrophotometer, BEL photonics, Brazil) using a spectral window from 200 to 700nm. Aliquots (10 μ l) of the extracts were also injected into a liquid chromatograph (LC-10A Shimadzu) system equipped with a C18 reversed-phase column (Vydac 201TP54, 250mm x 4.6mm, 5 μ m \emptyset , 35°C) coupled to a pre-column (C18 Vydac 201TP54, 30mm x 4.6mm, 5 μ m \emptyset) and a spectrophotometric detector (450nm). Methanol: acetonitrile (90: 10, v/v) was used for elution at a rate of 1 ml/min. The identification and quantification of compounds of interest was carried out via co-chromatography and comparison of retention times of samples with those of standard compounds (Sigma–Aldrich, USA) under the same experimental conditions.

2.3 STATISTICAL AND CHEMOMETRIC ANALYSIS

All procedures were performed in triplicate (n = 3). Data were collected, summarized, and submitted to analysis of variance (ANOVA) followed by post-hoc Tukey's test ($p \leq 0.05$) for mean comparison. The processing of the spectrophotometric profile considered the definition of the spectral window of interest (200-700nm), baseline correction, normalization, and optimization of the signal/noise ratio (smoothing). The processed data set was initially subjected to multivariate statistical analysis, by applying Principal Component Analysis (PCA) and clustering methods. Further, these data were also used for fold-change and Volcano plot with t-test univariate analyses, considering samples with pigmented roots (red and yellow fleshed) and non-pigmented roots (cream-fleshed) as the two groups to compare. Furthermore, the spectral data set and the amounts of the target carotenoids determined by RP-

HPLC were used for calculation of the Principal Components, supported by scripts written in R language (v. 3.1.1) [8] using tools defined by our research group and some functions from the packages Chemospec [9], HyperSpec [10], and ggplot2 [11]. Finally, in order to extract latent information from the UV-Vis dataset, classification models were built by applying supervised classification and feature selection methods, e.g., PLS-DA (Partial Least Squares Discriminant Analysis) and RF (Random Forest). All R scripts, raw data, and additional chemometric analysis are available in supplementary material, in <http://darwin.di.uminho.pt/metabolomicspackage/> as well as the data analysis report automatically generated from the R scripts using the features provided by R Markdown <http://darwin.di.uminho.pt/metabolomicspackage/cassava-carotenoids.html>. This allows anyone to fully reproduce and document the experiments.

3 RESULTS AND DISCUSSION

Carotenoids typically show maximum absorption at 450nm] and as depicted in Fig 1, all the spectral profiles (200-700nm) of the yellow and red cassava root extracts revealed prominent absorbance peaks between 400-500nm, indicating that the organosolvent system used was efficient to extract the target metabolites. Lower absorbance values were found in cream-fleshed roots, precisely because they have low concentrations of carotenoids as the *Rosada* genotype (red-fleshed) showed the highest absorbance values at 450nm.

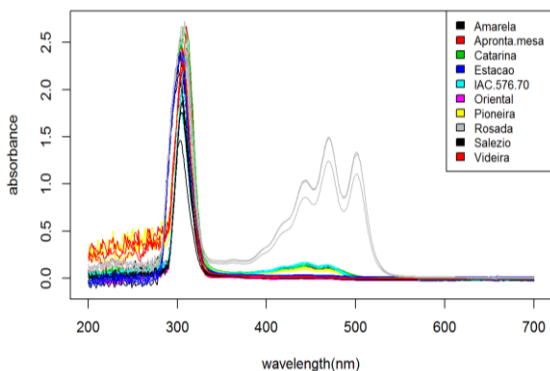


Figure 1: Typical UV-Vis spectrophotometric profiles ($\lambda = 200-700$ nm, acetone: petroleum ether - v/v) of root parenchymal tissues of ten cassava genotypes cultivated in southern Brazil.

Fold change analysis applied to samples with pigmented roots (red and yellow-fleshed) and non-pigmented roots (cream-fleshed) discriminated important points between 400 and 500 nm (carotenoids fingerprint region). Besides, some signals around 300 nm, a typical region of phenolic compounds absorption, also showed to be significant in their relative signal intensity defined as a 2-fold change (Fig. 2). This approach is consistent with the univariate analysis expressed through statistical significance of the descriptive p-value model. Indeed, wavelengths with lower p-values and therefore representing more significant differences between the samples studied were detected in the 440-470 nm spectral window (data shown on the analysis report).

Because of this, a volcano plot analysis was applied, associating the statistical differences found by the t-test and fold change analysis (Fold change threshold = 2 and p value ≤ 0.001).

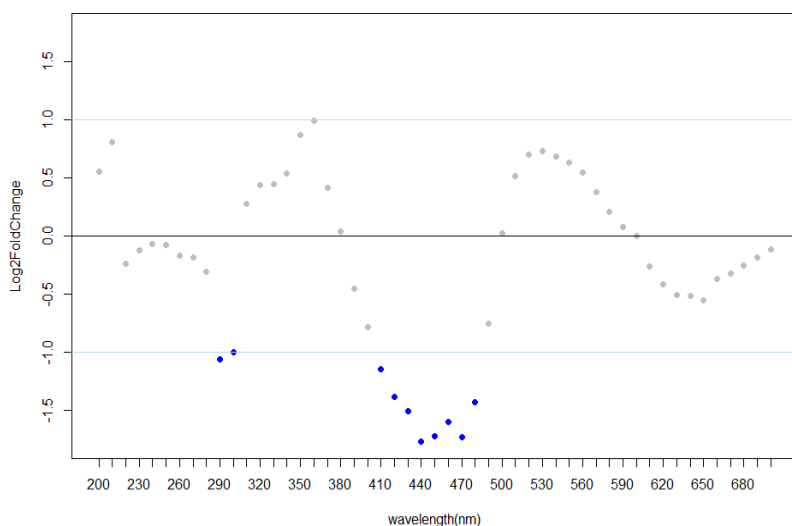


Figure 2: Fold change analysis of the UV-Vis data set (λ = 200-700 nm, acetone: petroleum ether - v/v). Significant differences (blue symbols) in the relative intensity (defined as a 2-fold change) of absorbance signals important for discriminating the samples where detected between 400 and 500 nm (carotenoids fingerprint region).

This analysis is used to quickly identify differences in large metabolomics datasets. In this case, once again, the values that display large magnitude fold changes, as well as high statistical significance,

occurred at 440-470 nm wavelengths, being related to typical maximum absorption bands of carotenoids (Fig. 3).

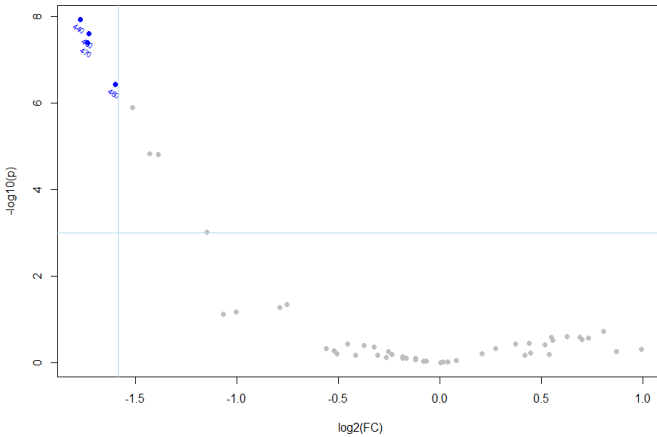


Figure 3: Volcano plot analysis (Fold change threshold = 2 and p 0.001). The larger magnitude fold changes and the higher statistical significance refer to wavelengths between 440-470 nm. (Blue symbols).

In a second series of experiments, the principal components were calculated from the full spectrophotometric ($\lambda = 200-700\text{nm}$) data matrix. PC1 and PC2 contributed to explain 78.9% of the total variance of the data set, but a clear discrimination of the samples according to the carotenoid concentrations was not found. Only the *Rosada* genotype distinguished from the others by grouping in PC1 + / PC2 -. Genotypes with high (yellow) and low carotenoid contents (cream) were spread out over the factorial distribution plane, making difficult any discrimination (Fig.4).

Such findings prompted us to build a second analytical model by applying PCA to the carotenoid fingerprint region of the UV-Vis (400-500nm). In this case, PC1 and PC2 accounted for 99.97% of the variance, clearly revealing three groups according to their similarities (Fig. 5). Interestingly, the samples were grouped according to their carotenoids contents determined by RP- HPLC and distributed according to the root flesh color. Cassava genotypes with yellow-fleshed roots (*Pioneira*, *Amarela*, *Catarina* and *IAC576-70*) were clustered along the PC2 + axis. Genotypes with cream-fleshed roots and lower carotenoid content (*Apronta mesa*, *Oriental*, *Salézio*, *Estação*, and *Videira*) were grouped in PC1/PC2 -. In its turn, the *Rosada* genotype

(red) seems to have a metabolic profile occurring away from all the other samples.

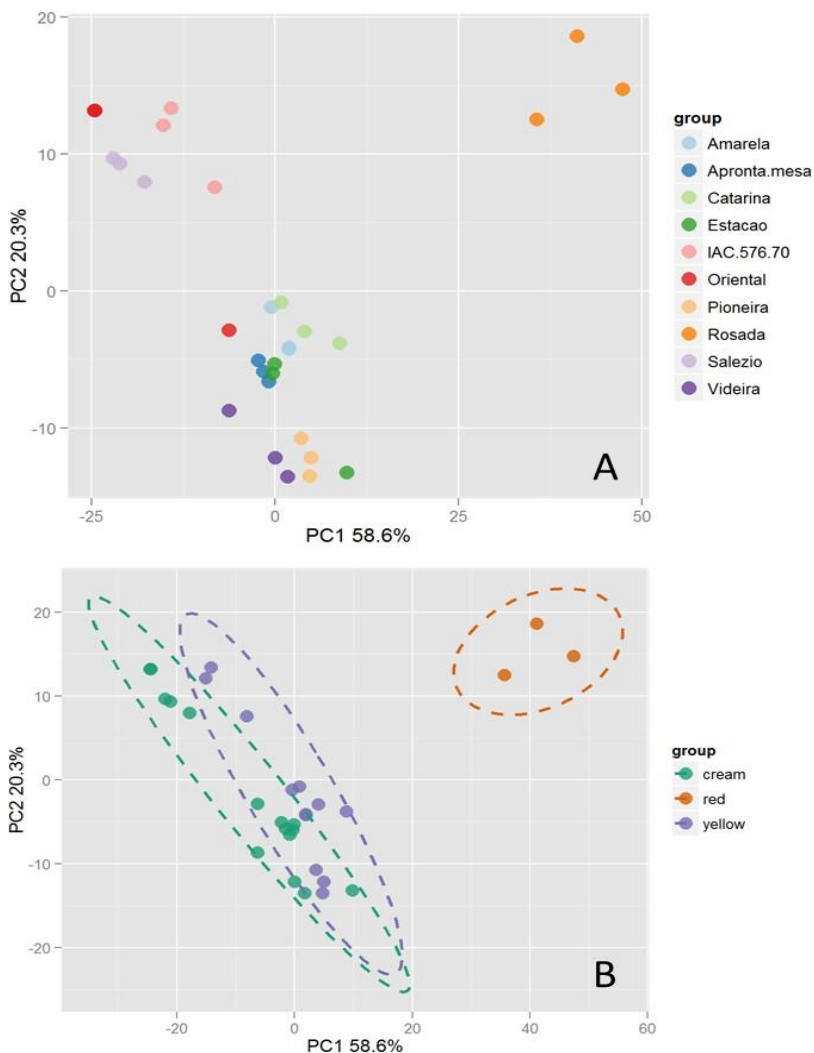


Figure 4 - Factorial distribution (principal components 1 and 2) of the UV-Vis spectral data set (200-700 nm) of the organosolvent extract of roots of ten cassava genotypes (A). Graphical demonstration of factorial distribution (principal components 1 and 2) according to the root flesh color (B).

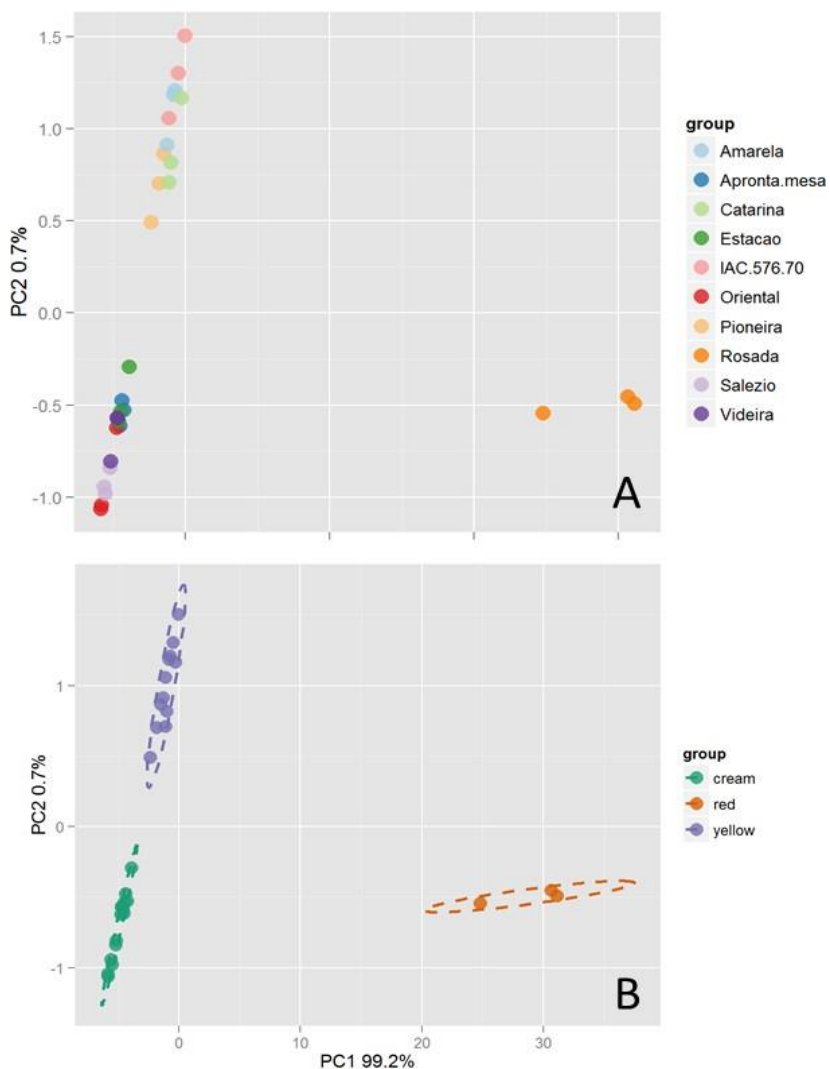


Figure 5: A - Factorial distribution (principal components 1 and 2) of the spectral data set of the fingerprint region of carotenoids (UV-Vis 400-500 nm, acetone: petroleum ether - v/v, 1:1). B - Graphical demonstration according to the root flesh color.

The chromatographic profiles of the organosolvent extracts identified *cis*- and *trans*- β -carotene, α -carotene, lutein, and β -cryptoxanthin in all the cassava genotypes analyzed. The presence of lycopene, a common precursor of the carotenoids above mentioned, was detected only in *Rosada* genotype, a fact that led us to speculate this is an important reason for its clear discrimination in respect to other genotypes. The isomer *trans* of β -carotene was the major compound regardless of the sample analyzed. In a second series of experiments, PCA was applied to the chromatographic data set revealing patterns of similarity of carotenoid composition among the studied genotypes.

These findings corroborate the results previously found by UV-Vis scanning spectrophotometry taking into account the fingerprint region of carotenoids (i.e. 400-500nm). Fig. 6 depicts the grouping of genotypes after calculation of the principal components from the RP-HPLC quantification of carotenoids. PC1 and PC2 explain 97.8% of the total variance of the sample population under study.

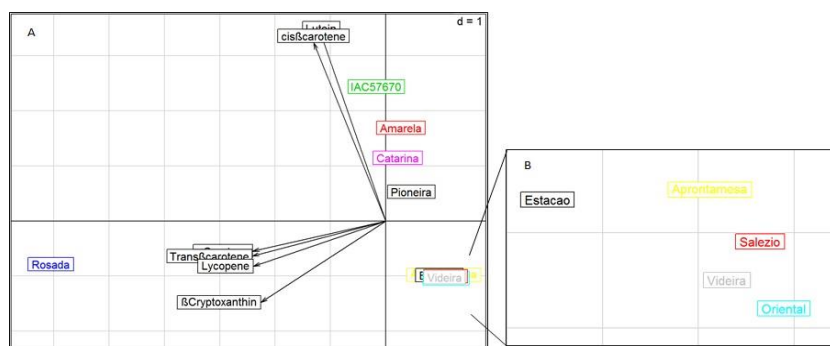


Figure 6: A - Scores scatter plot (PC1 and PC2) of the quantitative data of carotenoids determined by RP-HPLC in root samples of ten cassava genotypes (n = 3). B - Magnification to the overlapping samples at the PCA.

The genotypes with yellow-fleshed roots (*Pioneira*, *Amarela*, *Catarina*, and *IAC-576-70*) were grouped in PC2+, influenced by their higher concentration of *cis*- β -carotene and lutein. Inversely, the genotypes with cream-fleshed roots (*Apronta mesa*, *Oriental*, *Sale'zio*, *Estaco*, and *Videira*) grouped in PC1 +/PC2 - due to their lower amounts of these pigments. Samples of red- fleshed roots (i.e., *Rosada*) showed higher dissimilarity among the studied genotypes, grouping into PC1/PC2 -. This result seems to be directly influenced by the presence of lycopene and the higher concentrations of *trans*- β -carotene, α -

carotene, and β -cryptoxanthin. Finally, hierarchical cluster analysis was applied to the chromatographic data, affording similar results to the UV-Vis scanning spectrophotometry for the fingerprint region of carotenoids. Genotypes with the highest similarity in their carotenoid composition are represented by cluster hierarchical analysis in Fig. 7. The similarities were defined based on the Euclidean distance between two samples, using the arithmetic average (UPGMA) to compute the resulting tree, the obtained cophenetic correlation was of 97.61%.

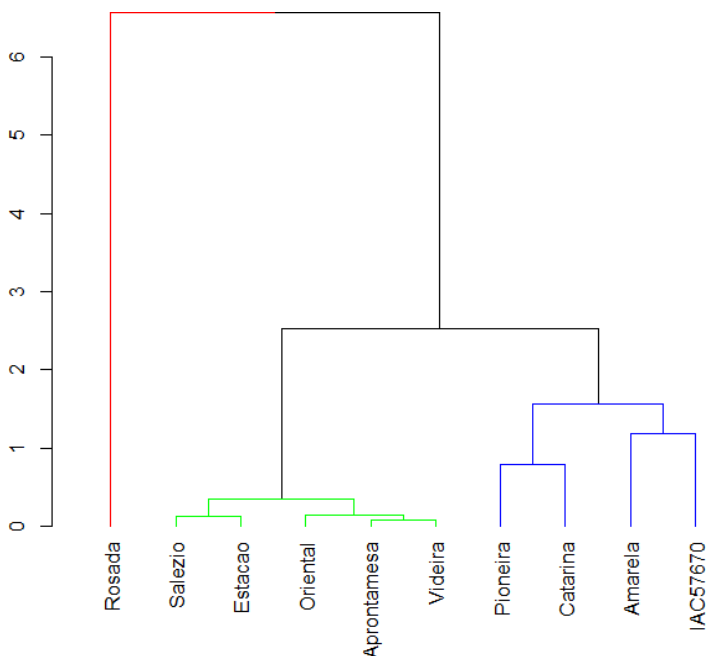


Figure 7: Dendrogram of cassava genotypes in respect to their carotenoid composition determined by RP-HPLC, followed by hierarchical clustering analysis (UPGMA method - 97.61% of cophenetic correlation). The similarity among genotypes of the same cluster is statistically significant (p 0.05) when the branches in the dendrogram show the same color. Significance determined by Simprof analysis (Similarity Profile Analysis) from R Clustsig package in accordance with Clarke, Somerfield & Gorley (2008). [12].

In order to extract latent information from the UV-visible dataset, classification models were built by applying supervised classification and feature selection methods. The methods used in this work were PLS-DA (Partial Least Squares Discriminant Analysis) and

RF (Random Forest). The quantitative measure of the performance for PLS-DA and RF classification models showed accuracy values around 82% and 66%, respectively, revealing a good performance of the methods, mainly PLS-DA. The cross-validation for the PLS and RF models (10 fold, repeated 10 times) expressed by the confusion matrix showed that the method is able to differentiate the genotypes studied. Besides, taking into consideration the influence of the variables used by the models in the classification, the wavelength characteristic of carotenoids were again the most expressive to validate and distinguish the studied materials (data shown on report). Such findings demonstrate that the analytical approach used in this study is useful for distinguishing the genotypes of a population sample in a germplasm bank, for example.

4 CONCLUSIONS

The data set obtained by the analytical techniques employed in this work allowed a better understanding of the chemical variability associated with roots' carotenoid composition of the cassava genotypes. The large disparity in carotenoid contents reveals the chemical variability among the genotypes analyzed. Substantial amounts of carotenoids were detected in the cassava genotypes, indicating their potential as source of interesting compounds to human health and nutrition, given the presence of pro-vitamin A carotenoids (β -carotene, e.g.) and lycopene in roots of yellow and red color, respectively. The *Rosada* genotype was found to be discrepant because its richness in the carotenoids, in addition to the presence of lycopene in relevant amounts. The information obtained by coupling the analysis of biochemical markers for pro-vitamin A in cassava genotypes to bioinformatics tools revealed to be relevant for the rational design of biochemically-assisted cassava breeding programs. Indeed, the analytical approach adopted (i.e., UV-Vis/RP-HPLC/chemometrics) allowed to discriminate and classify the claimed genetic variability of the studied samples based on their biochemical traits, helping to identify/select cassava genotypes of interest to human health and nutrition.

Acknowledgements

To FAPESC (Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina) and CNPq (Conselho Nacional de

Desenvolvimento Científico e Tecnológico) for financial support. The research fellowship from CNPq on behalf of the last author is acknowledged. The work is partially funded by Project PropMine, funded by the agreement between Portuguese FCT and Brazilian CNPq. The authors also thank the FCT Strategic Project of UID/BIO/04469/2013 unit.

5 REFERENCES

- [1] Who. Global prevalence of vitamin A deficiency in populations at risk 1995-2005. *World Health Organization Global Database on Vitamin A Deficiency*, 2009.
- [2] D. Rodriguez-Amaya and M. Kimura. HarvestPlus Handbook for Carotenoid Analysis. *HarvestPlus Technical Monographs*, page 59, 2004.
- [3] A. C. Kurilich and J. A. Juvik. Quantification of carotenoid and tocopherol antioxidants in *Zea mays*. *Journal of Agricultural and Food Chemistry*, 47(5):1948–1955, 1999.
- [4] H. Tapiero, D. M. Townsend and K. D. Tew. The role of carotenoids in the prevention of human pathologies. *Biomedicine and Pharmacotherapy*, 58(2):100–110, 2004.
- [5] S. Voutilainen, T. Nurmi, J. Mursu and T. H. Rissanen. Carotenoids and cardiovascular health 1 – 3. *The american journal of clinical nutrition*, 83:1265–1271, 2006.
- [6] D. B. Rodriguez-Amaya. *A Guide to Carotenoid Analysis in Foods*. 2001.
- [7] C. Iglesias, J. Mayer, L. Chavez and F. Calle. Genetic potential and stability of carotene content in cassava roots. *Euphytica*, 94(3):367–373, 1997.
- [8] R Core Team. R : A Language and Environment for Statistical Computing. 1, 2014. URL .
- [9] B. A. Hanson. ChemoSpec : An R Package for Chemometric Analysis of Spectroscopic Data and Chromatograms (Package Version 2.0-2). pages 1–41, 2014.

- [10] C. Beleites. Import and Export of Spectra Files. pages 1–20, 2011.
- [11] H. Wickham and W. Chang. Package ‘ggplot2’. *The Comprehensive R Archive Network*, 2015.
- [12] K. R. Clarke, P. J. Somerfield and R. N. Gorley. Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. *Journal of Experimental Marine Biology and Ecology*, 366(1-2):56–69, 2008.

CAPÍTULO III

DESENVOLVIMENTO DE FERRAMENTAS DE CLASSIFICAÇÃO E QUANTIFICAÇÃO DE CAROTENOIDES EM *M. esculenta* VIA METABOLÔMICA E MINERAÇÃO DE DADOS

**Artigo publicado no *Journal of Integrative Bioinformatics*:
UV-vis and CIELAB based chemometric characterization of
Manihot esculenta carotenoid contents**

UV-vis and CIELAB based chemometric characterization of *Manihot esculenta* carotenoid contents

Rodolfo Moresco², Telma Afonso¹, Virgilio G. Uarrota²,
Bruno Bachiega Navarro², Eduardo da C. Nunes³, Marcelo
Maraschin², Miguel Rocha.^{1, *1}

¹Centre Biological Engineering, School of Engineering, University of Minho, Braga, Portugal, <https://www.ceb.uminho.pt/>

²Plant Morphogenesis and Biochemistry Laboratory, Federal University of Santa Catarina, Florianopolis, Brazil, <http://www.ufsc.br/>

³Santa Catarina State Agricultural Research and Rural Extension Agency (EPAGRI), Experimental Station of Urussanga, Urussanga, Brazil, <http://www.epagri.sc.gov.br/>

Abstract

Vitamin A deficiency is a prevalent health problem in many areas of the world, where cassava genotypes with high pro-vitamin A content have been identified as a strategy to address this issue. In this study, we found a positive correlation between the color of the root pulp and the total carotenoid contents and, importantly, showed how CIELAB color measurements can be used as a non-destructive and fast technique to quantify the amount of carotenoids in cassava root samples, as opposed to traditional methods. We trained several machine learning models using UV-visible spectrophotometry data, CIELAB data and a low-level data fusion of the two. Best performance models were obtained for the total carotenoids contents calculated using the UV-visible dataset as input, with R^2 values above 90%. Using CIELAB and fusion data, values around 60% and above 90% were found. Importantly, these results demonstrated how data fusion can lead to a better model performance for prediction when comparing to the use of a single data source. Considering all these findings, the use of colorimetric data associated with UV-visible and HPLC data through statistical and machine learning methods is a reliable way of predicting the content of total carotenoids in cassava root samples.

¹ To whom correspondence should be addressed. Email:

Keywords: Carotenoids, Cassava genotypes, Chemometrics, CIELAB, Machine learning

1 INTRODUCTION

Carotenoids refer to the most important natural pigments, having already been described more than 700 phytochemicals of this class of secondary metabolites, with a broad range of structures and polarities. About 50 carotenoids are vitamin A precursors, however, only three of these represent major sources in the human diet (β -carotene, α -carotene and β -cryptoxanthin) [1].

In terms of chemical structure, vitamin A is essentially a half of the β -carotene molecule with a water molecule at the end of the lateral polyene chain (Figure 1), making β -carotene a potent vitamin A precursor, with 100% of activity assigned to it [2]. Having a broad range of colors, varying from yellow to dark red, carotenoids are responsible for the color of many plant leaves, fruits and flowers, as well as birds, insects, fish, and crustaceans. However, they are only produced by plants, bacteria, fungi and algae, while other organisms can only incorporate them through their diets.

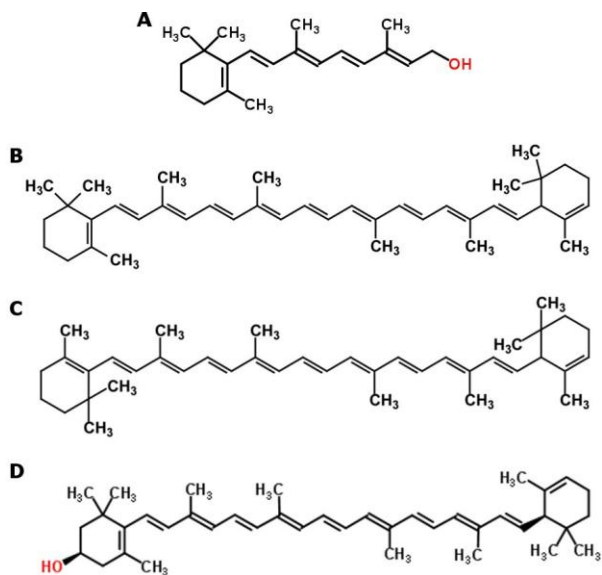


Figure 1 Structural formula of vitamin A (A) and of the carotenoids β -carotene (B), α - carotene (C) and β -cryptoxanthin (D), the major carotenoid sources in the human diet [3].

The recognized benefits of carotenoid consumption include their correlation with a diminished risk of several degenerative disorders, including various types of cancer, cardiovascular or ophthalmological diseases, as well as their preventive effect associated with their antioxidant activity, protecting cells and tissues from oxidative damage [4].

Though sources of provitamin A carotenoids around the world are abundant, vitamin A deficiency is a leading cause of morbidity and mortality, especially in young children and pregnant and lactating women in less favored countries, affecting around 250 million people worldwide. Therefore, food-based interventions focused on alleviating vitamin A deficiency in susceptible populations have advantages over supplementation and fortification programs, especially in rural areas, because they can provide a sustainable source of a variety of nutrients and other phytochemicals without the recurring transport and administration costs of these other methods [1].

Cassava is the commonly used term to designate the *Manihot esculenta* species. This tuberous-root plant species offers a wide variety of agronomic advantages, being resilient to droughts, inexpensive, resistant to major diseases and pests, easy to grow and having flexible harvest times, allowing farmers to harvest the roots as needed. It is, therefore, a valuable source of energy for people living in the poorest regions. However, this crop is also a poor source of provitamin A carotenoids, whose deficiency is a major problem in those regions. Thus, the identification of cassava genotypes with high contents of provitamin A carotenoids as a strategy to reduce the prevalence of deficiency of this vitamin is of most importance for the global food security and nutrition [5, 6].

A typical trait of cassava is its starchy root, whose color can vary from white to red. The color is strongly correlated to the presence and contents of several carotenoid pigments and their associations [7]. However, the possibility of adopting the color of roots as an indirect criterion for selection of higher carotene content is questionable, since color is a characteristic of difficult visual evaluation.

Therefore, the use of a standardized color measurement technique is of most importance. One such technique was adopted by

the *Commission Internationale de Leclairage* (CIE) and is based on the Lab color space, which describes mathematically all perceivable colors in three dimensions: L^* for lightness and a^* and b^* for the color opponents green-red and blue-yellow, hence the CIE $L^* a^* b^*$ or CIELAB color scale designation (Figure 2). The values of these three variables are usually absolute, with the L^* value representing the darkest black at $L^* = 0$, and the brightest white at $L^* = 100$. On the other hand, the a^* value represents red and green opponents at positive and negative values, respectively, while the b^* value represents yellow and blue opponents at positive and negative values, respectively. Both color channels, a^* and b^* , will represent true neutral gray values at $a^* = 0$ and $b^* = 0$ [8, 9].

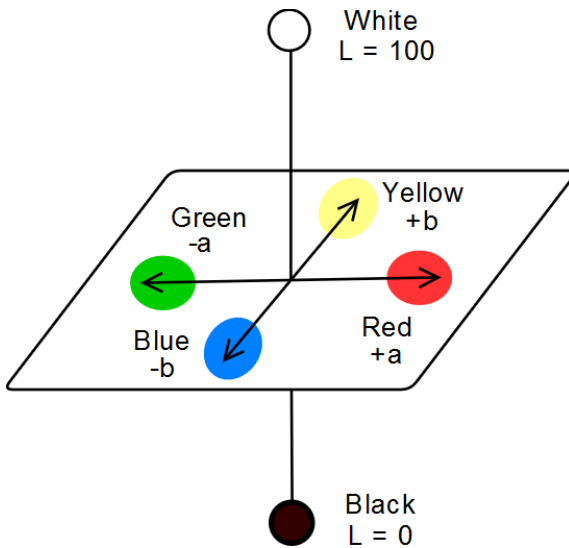


Figure 2: Representation of the CIE $L^* a^* b^*$ color space.

CIELAB is currently the most used system for quantitative color description of an object, due to its uniformity, ease of acquisition, very low cost and device independence, having been used for instance in the unique identification of skin color for clinical and scientific purposes [10] and as an optimal color design approach for transforming patients' perception into color elements [11].

Chemical extraction, followed by the identification and quantification of carotenoid pigments, especially by UV-visible

spectrophotometry and high performance liquid chromatography (HPLC) are very accurate, but extremely expensive, also requiring a long time for the analysis. Since the CIELAB color measurement is a non-destructive and very fast technique, that facilitates the acquisition of measurements in the field, while also avoiding the degradation of the compounds, it becomes an appealing approach in comparison to traditionally used methods.

Throughout the last years, there has been a massive increase in available scientific data. While contributing to a more informed community, the large amounts of data can pose a problem when it comes to the analysis, which needs to be automated so that information can be more efficiently retrieved. Machine learning provides systems with the ability to automatically learn and improve from experience without being explicitly programmed, focusing on the development of computer programs that can access and learn for data. The learning process begins with data that is provided and searched for patterns, so that better decisions can be made in the future. This is often a very attractive alternative to manually look for such patterns, and in the last decade the use of machine learning has spread rapidly throughout computer science and beyond [12,13].

Data fusion is a process of combining data from different sources to improve the performance of prediction models. It deals with association, detection, correlation, and estimation of data to achieve a better information of the system's state. Low-level fusion is made on a data level, by direct association and combination of raw data, representing measures of the same physical phenomena. After data combination, a feature vector is extracted and used in a machine learning process. It aims to provide more accurate results, assuming proper data association [14].

The aim of this work is to validate a quantification method for carotenoid content estimation in roots of *M. esculenta* from colorimetric data using the CIELAB color system, assuming that the statistical and machine learning techniques can correlate colorimetric data easily obtained in the field with the contents obtained through traditional techniques such as UV-visible spectrophotometry and HPLC.

The present study employs machine learning, as well as other statistical techniques and bioinformatics tools to detect genotypes of *M. esculenta* with high contents of carotenoids, while also providing tools that can support the plant-breeding program at Epagri (Agricultural Research Company and Rural Extension of the State of Santa Catarina,

(<http://www.epagri.sc.gov.br/>) that aims to obtain genotypes with high levels of pro-vitamin A carotenoids and superior nutritional traits.

2 MATERIALS AND METHODS

2.1 SELECTION OF CASSAVA GENOTYPES

Roots of fifty *M. esculenta* genotypes harvested in 2015/2016 season from the Epagri's germplasm bank (Urussanga Experimental Station, 28°31'18"S, 49°19'03"W, Santa Catarina, southern Brazil) were used in this study due to their economic and social importance.

All genotypes were cultivated under the same soil, climatic conditions and agricultural treatments. Importantly, the investigated genotypes were pre-selected according to their relevance for biofortification projects, due to the presence of carotenoids with provitamin A activity and lycopene (visual selection), low levels of cyanogenic glycosides and suitable agronomic traits (e.g., high yield, resistance to drought and to pests and diseases), being widely cultivated in southern Brazil. In fact, the Epagri plant breeder team indicated fifty genotypes from the germplasm bank that were also preferred by local small farmers for commercial production due to their physiochemical variability.

2.2 CAROTENOID EXTRACTION AND QUANTIFICATION

Carotenoids were extracted from fresh cassava roots as described by Rodriguez-Amaya & Kimura (2004) [15] using an Ultra-Turrax (Janke & Kunkel IKA - T25 basic) and mixture of acetone: petroleum ether (v/v) as extraction solution.

The absorbances of the organosolvent extracts were then recorded on an UV-visible spectrophotometer (Gold Spectrum lab 53 UV-Vis spectrophotometer, BEL photonics, Brazil) using a spectral window from 200 to 700 nm. Aliquots (10 µl) of the extracts were also injected into a liquid chromatograph (LC-10A Shimadzu) system equipped with a C18 reversed-phase column (Vydac 201TP54, 250mm x 4.6mm, 5µm φ, 35°C) coupled to a pre-column (C18 Vydac 201TP54, 30mm x 4.6mm, 5µm φ) and a spectrophotometric detector (450 nm). A mixture of methanol: acetonitrile (90:10, v/v) was used for elution at a flow rate of 1 mL/min. The identification and quantification of compounds of interest was carried out via co-chromatography and comparison of retention times of samples with those of standard

compounds (Sigma-Aldrich, USA) under the same experimental conditions.

The color measurements of the root samples were made immediately after harvest using a colorimeter (CR-400, Minolta®, Japan) and the results expressed according to the CIELAB color space scale [5]. For all fifty samples, three readings were performed at different sites.

2.3 STATISTICAL ANALYSIS

Data relating to the quantification of carotenoids were expressed as the mean (μg carotenoids /g root - dry weight) \pm standard deviation and submitted to an analysis of variance (ANOVA) followed by post-hoc Tukey's honest significant difference test ($p < 0.05$) for mean comparison.

Spectrophotometric data and the amounts of the target carotenoids determined by HPLC were treated using multivariate statistical analysis and chemometrics techniques supported by scripts written in R language (v. 3.3.1) [16].

All data analysis were supported and structured using the R *specmine* package [17]. This package was developed within our research group and allows the integrated analysis of metabolomics and spectral data, providing a set of methods for metabolomic data analysis, including the loading of data in different formats, preprocessing, metabolite identification, univariate and multivariate data analysis, machine learning, and feature selection. It is a package that allows conducting a data analysis pipeline, exploring and applying different methods of analysis, while also having functions that provide abundant options for graphical visualization of the results.

2.3.1 Machine Learning

In order to obtain machine learning models capable of accurately predicting the carotenoid contents in cassava roots, we used regression-derived statistical and machine learning models, such as least absolute shrinkage and selection operator (Lasso), ridge and linear regression, regression trees, random forests, elastic network, partial least squares (PLS), support vector machines (SVM), and K-nearest neighbors models [12,13].

Data from UV-visible spectrophotometry, CIELAB, as well as a fusion of the two were used as inputs to each of the referred machine learning models. The total carotenoid content (TCC) determined by

spectrophotometry using the Lambert-Beer law and the TCC determined by HPLC were used as response variables for the prediction. The content of trans- β -carotene determined by HPLC, the most abundant carotene in cassava roots, was also used as an alternative response variable for the models (Figure 3).

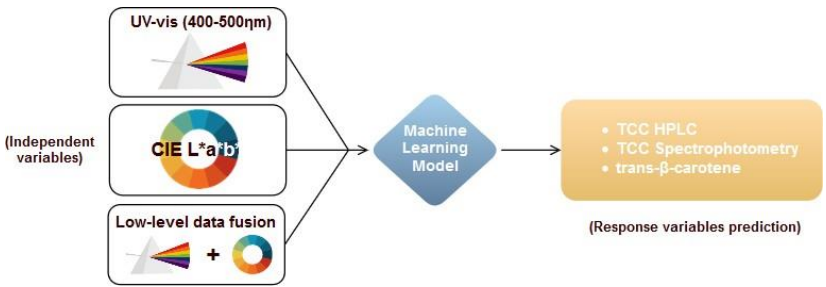


Figure 3: Machine learning approach used. Three different datasets were used as input to the models, namely the UV-visible, CIELAB and fusion datasets. The response variables used for prediction were the total carotenoid content (TCC) determined by spectrophotometry (Lambert-Beer law) and the TCC and trans- β -carotene content determined by HPLC.

This being a regression problem, the chosen evaluation metrics to compare model performance were the root mean square error (RMSE) and the coefficient of determination (R^2), since they explicitly show how much the model predictions deviate, on average, from the actual values in the dataset.

2.3.2 UV-visible dataset

Considering that most carotenoids exhibit absorption in the visible region of the spectrum, between 400 to 500 ηm , a subset of the original UV-visible dataset was used, with samples belonging to this wavelength interval (101 features). Additionally, missing values contained within this dataset were replaced with the mean of the variables' values.

Using the different response variables for prediction, we selected the models that showed best performance and the variable importance was calculated. A set of pre-processing methods was applied to the datasets to see whether model performance could be improved, using the models that showed best performance with raw

data. These pre-processing methods included smoothing interpolation, scaling, multiplicative scatter correction (MSC), first derivative calculation and background, offset and baseline corrections. The data was also subject to filter-based feature selection (40%, 60% and 80% data filtering) to determine if it could improve model performance.

2.3.3 CIELAB dataset

The analysis pipeline was similar for the CIELAB dataset, however, linear regression models with feature selection and the data pre-processing and filtering processes were excluded from the analysis pipeline, as it did not make sense to perform these, considering there are only 3 features in the dataset (L^* a^* and b^* parameters), while pre-processing was meant for spectral data.

2.3.4 Fusion dataset

For the fusion dataset, which contained 104 variables (absorbance values + L^* a^* b^* parameters), the analysis pipeline was similar to that of the CIELAB dataset, while data filtering was also performed similarly to the UV-visible dataset.

2.3.5 Tools and reporting

All R scripts, raw data and additional analysis pipelines reports are available as supplementary material at <http://darwin.uminho.pt/pacbb2017/cassava-carotenoids/>, allowing full reproducibility of the experiments. These were made using the already mentioned *specmine* R package, which provides a number of functions to train, use and evaluate machine learning methods, mostly based in the R package *caret* [18].

3 RESULTS AND DISCUSSION

3.1 DETERMINATION OF CAROTENOID CONTENTS

The UV-visible spectrophotometric profiles measured between 200-700 ηm clearly allow us to discriminate samples according to their carotenoid content. This is evident when comparing the typical UV-visible spectrophotometric profiles of cassava samples 5, 23 and 74 (Figure 4).

These three samples vary greatly in color, with sample 5 having a cream color, sample 23 a yellow one and the sample 74 a reddish color. In fact, the spectrophotometric profiles differ from each other

only at 400-500 ηm region of the spectrum, which is the region where carotenoids typically show absorbance peaks.

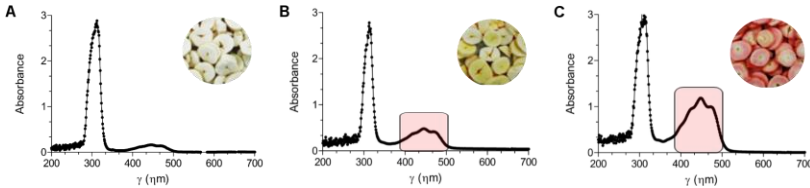


Figure 4: Typical UV-visible spectrophotometric profiles ($\lambda = 200-700 \eta m$, acetone: petroleum ether (v/v) of root parenchymal tissues of three cassava samples: A - sample 5, B - sample 23 and C - sample 74. The 400-500 ηm region of the spectrum is highlighted in cases B and C.

The cream colored sample profile (Figure 4A) shows an absence of absorbance peaks between the 400-500 ηm region. On the other hand, the yellow colored sample profile (Figure 4B) shows more noticeable peaks in this region, while the reddish colored sample (Figure 4C) presents three peaks of great absorption in this region of the spectrum. It is, therefore, expected that the more colored the root the higher carotenoid content it possesses.

To confirm this possibility, the total carotenoid content was determined by UV-visible spectrophotometry, using the Lambert-Beer formula, and is shown in Figure 5 for each of the fifty fresh root samples.

The wide disparity in the observed carotenoid contents reveals the chemical variability among the analyzed genotypes. In the present study, the cream-colored roots showed the lowest concentrations of total carotenoids, with values around $0.57 \mu g.g^{-1}$, while higher concentration values were measured in yellow and reddish pigmented roots i.e, $54.93 \mu g.g^{-1}$. The most abundant carotenoids, trans- β -carotene and cis- β -carotene, had concentration values that ranged from 1.82 to $42.82 \mu g.g^{-1}$ and from 1.19 to $28.86 \mu g.g^{-1}$, respectively. The results from the HPLC carotenoid quantification are available as supplementary material in the metadata file.

These findings altogether are consistent with data reported in the literature that observe a positive correlation between the color of the root pulp and the total carotenoid content [19, 20, 21].

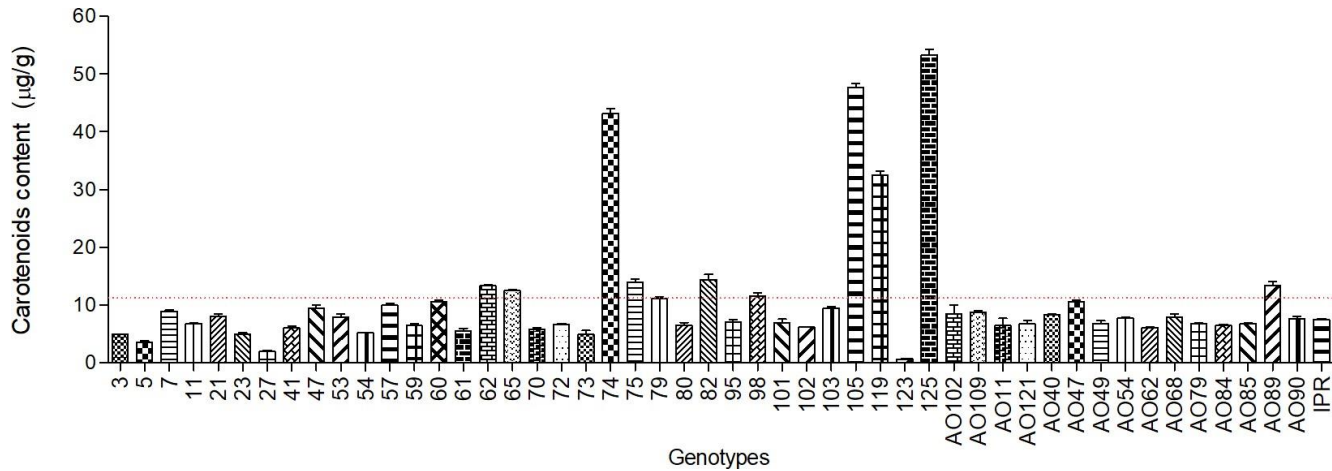


Figure 5: Concentration of total carotenoids ($\mu\text{g}\cdot\text{g}^{-1}$ dry weight \pm standard deviation, $n = 3$) in samples of roots of fifty *M. esculenta* genotypes, determined by UV-visible spectrophotometry (450 nm , $\epsilon = 2592 \text{ M}^{-1}\text{cm}^{-1}$).

3.2 CIELAB COLOR SPACE INTERPRETATION

To better understand the correlation between samples and the different types of carotenoids with the CIELAB color space we projected the observed values of L^* , a^* and b^* for each root sample into the CIELAB plane [22]. Figure 6 shows the samples location according to the color of roots in the CIELAB color space, whose visual interpretation is sufficient to verify which samples possess higher carotenoid amounts. Samples 105, 119 and 125 (Figure 6, ellipse I) show high b^* values, which stands for the coloration in the yellow range, and these are in fact the samples with the highest carotenoid contents, as it can be observed in Figure 5, Interestingly, sample 74 (Figure 6, ellipse IV) is deviated into the positive axis of a^* , which corresponds to the red coloration. In fact, this sample is a reddish root, mostly due to its lycopene content, which confers reddish coloration to the biomass [23]. It is one of the samples with the highest carotenoid concentration (Figure 5).

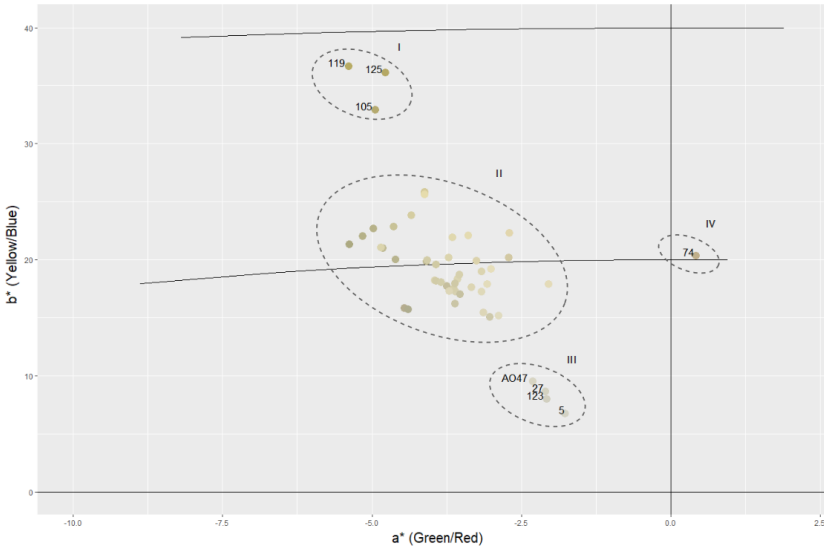


Figure 6: Location of the cassava samples in the CIELAB color space according to their root pulp colors. The a^* value characterizes the coloration in the regions of red (+ a^*) to green (- a^*). The value b^* indicates coloring in the range of yellow (+ b^*) to

blue (-b*). Sample identifiers in ellipse II were omitted for easier interpretation of the plot.

Samples 123, 27, 05, and AO47 (Figure 6, ellipse III) were grouped in values of b* closer to zero, these being the samples with the lowest carotenoid content (Figure 5). The remaining samples had medium and more similar carotenoid content, being grouped together in a* negative and b* positive values (Figure 6, ellipse II). The same similarity patterns of carotenoid composition were found among the evaluated genotypes through a PCA analysis (Figure 7). In their set, PC1 and PC2 explain about 99.5% of the total variance of the sample population data under this study.

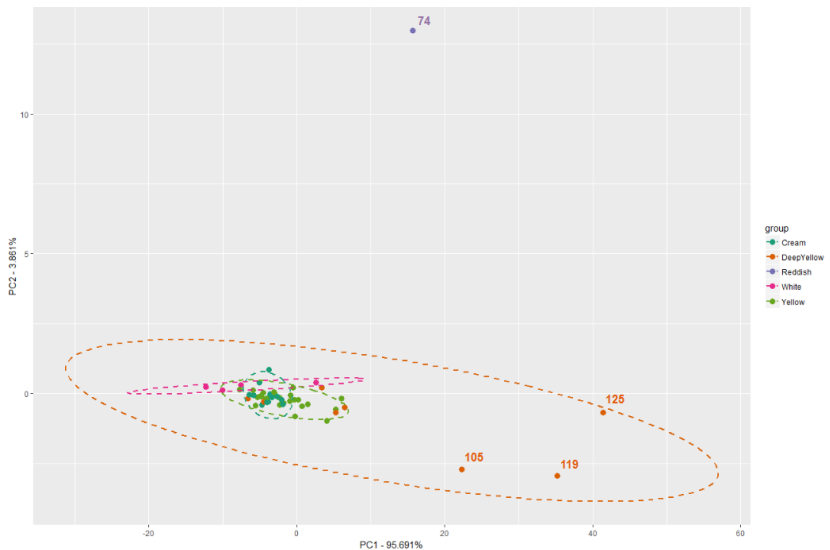


Figure 7: Scores plot with the distribution of the fifty samples on the first and second PCA components resulting from the UV-visible spectrophotometric data (400-500 nm) (n= 3 replicates).To facilitate the interpretation of the plot, only the sample identifiers for the most relevant samples are shown.

This analysis revealed genotype grouping according to their root pulp coloration, as well as carotenoid quantification, similarly to the CIELAB data, with samples 74, 105, 119, and 125 being the most discrepant within this sample universe.

These results are in accordance with the findings in subsection 3.1 that positively correlate the carotenoid content with the color of the cassava roots.

3.3 CAROTENOID CONTENT PREDICTION USING UV-VISIBLE DATA

In Table 1, the performance values obtained by the ML models using UV-visible data (400-500 ηm) as inputs and the total carotenoid content (TCC) determined by spectrophotometry (Lambert-Beer formula), the TCC determined by HPLC and the total content of trans- β -carotene (the most abundant carotene in cassava roots) as response prediction variables are shown.

Table 1: Performance values (RMSE and R^2) obtained for the different machine learning models trained with UV-visible spectrophotometry data (400-500 ηm). The total carotenoid content (TCC) determined by spectrophotometry (Lambert-Beer formula), the TCC determined by HPLC and the total content of trans- β -carotene (the most abundant carotene in cassava roots) were used as response prediction variables. The parenthesis indicate the package specific method chosen for the simulation, with exception to the linear regression models

	UV-visible (400-500 ηm)					
	TCC Spectrophotometry		TCC HPLC		trans- β -carotene	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
Partial Least Squares (simpls)	3.492	0.9208	5.789	0.5721	4.309	0.36296
Support Vector Machines (e1071)	3.709	0.9316	5.844	0.5975	4.218	0.39924
Partial Least Squares (widekernelpls)	3.732	0.9238	5.779	0.5701	4.324	0.45308
Random Forest	3.768	0.9483	7.275	0.3596	5.753	0.23993
Elastic Net	3.793	0.9185	5.934	0.6340	4.191	0.41274
Partial Least Squares (pls)	3.800	0.9529	5.643	0.5971	4.265	0.47090
Ridge Regression (w/ FS)	3.855	0.9478	5.880	0.6038	4.159	0.35640
Ridge Regression	3.877	0.9283	7.282	0.6163	4.407	0.31655
Support Vector Machines (kernelab)	3.928	0.9409	5.907	0.5892	4.230	0.46608
Partial Least Squares (kernelpls)	4.096	0.8962	5.878	0.5661	4.211	0.42217
Linear Regression (w/ Stepwise Selection)	4.158	0.9192	8.341	0.5265	6.135	0.20603
Linear Regression (w/ Forward Selection)	4.178	0.8883	8.783	0.4716	5.142	0.31153
Linear Regression (w/ Backwards Selection)	4.392	0.8711	6.373	0.5226	5.355	0.27887
K-Nearest Neighbors	4.732	0.9224	6.277	0.4451	4.597	0.22467
Lasso	5.207	0.8174	17.508	0.2494	16.145	0.18959
Conditional Inference Random Forest	6.713	0.7917	6.806	0.5588	4.703	0.36963
Conditional Inference Tree	7.363	0.7114	6.916	0.4805	4.894	0.28851
Decision Trees	7.582	0.6833	6.795	0.4736	5.189	0.05344

It is clear that the highest R^2 performance values (above 90%) and lowest RMSE values were obtained when using the TCC determined using spectrophotometric data as response prediction variable. This was expected considering that both input and response data used employ the same physical phenomenon of detection of compounds (absorbance). The models that best performed in this case were partial least squares (PLS) using both *simpls* and *widekernelpls* methods, support vector machines (SVMs) and random forests with RMSE performance values of 3.492, 3.732, 3.709 and 3.768, respectively.

Using the TCC determined by HPLC as the response variable, a small decrease in performance values is observed. Here, PLS (*widekernelpls* and *pls* methods) and elastic network showed best performance with RMSE values of 5.779, 5.643 and 5.934, respectively, and R^2 values around 60%.

The worst results were obtained by using trans- β -carotene as response variable, with best performance models being PLS (*widekernelpls* and *pls* methods) and SVMs, with RMSE values of 4.324, 4.265 and 4.230, respectively, and R^2 values around 46%.

When observed, the values of VIPs (Variable Importance in the Projection) for this analysis (supplementary material), which identify the most relevant variables for the validation of the method, it can be detected that the wavelengths 449, 448 and 450 nm (precisely the wavelength that is used for the quantification of β -carotene through the Lambert-Beer formula) were used in 100%, 99.93% and 99.76% of cross-validation training performance. This result is important because it attests to the robustness of the models in predicting the contents of these compounds in cassava samples.

Applying pre-processing methods to the data, as well as feature selection, showed an overall increase in model performance for most models used (supplementary material). In Table 2 one such case is shown, where using pre-processed UV-visible data as input to Random Forest model that showed best performance when using raw data, increased even further model performance. By applying smoothing interpolation, background and offset corrections, or background correction alone, RMSE values decreased from 6.194 to 5.773, 5.936 and 6.175, respectively. R^2 values also increased from 55% to around 60% in each case.

Table 2: Performance values (RMSE and R^2) obtained for a random forest model trained with UV-visible spectrophotometry data (400-500 ηm), applying several pre-processing methods to the data. The total carotenoid content (TCC) determined by HPLC was used as response prediction variable.

UV-visible (400-500 ηm) + Preprocessing, Random Forest		
	TCC Spectrophotometry	
	RMSE	R^2
Smoothing Interpolation	5.773	0.6053
Background and Offset corrections	5.936	0.5927
Background correction	6.175	0.5956
No preprocessing	6.194	0.5581
Scaling	6.447	0.5740
Background, Baseline and Offset corrections	9.397	0.4780
First derivative	10.774	0.4482
Multiplicative Scatter Correction	11.621	0.3245

3.4 CAROTENOID CONTENT PREDICTION USING CIELAB DATA

Table 3 shows the performance values obtained by using CIELAB data as input to the various machine learning models and using the TCC determined by spectrophotometry (Lambert-Beer formula), the TCC determined by HPLC and the total content of trans- β -carotene (the most abundant carotene in cassava roots) as response prediction variables.

Similarly to the results obtained in subsection 3.3, the highest R^2 performance values and lowest RMSE values were obtained when using the TCC determined using spectrophotometric data as a response prediction variable. However, there is a noticeable overall decrease in model performance when using all three prediction variables. This is easily explained by the number of features present in the data, considering that in this case only three features are present (L^* , a^* and b^*), while in the previous case there were far more features, about 101 (data measured from 400 to 500 ηm).

Using the TCC determined by spectrophotometry as response variable, the models that showed best performance were linear and ridge regressions and elastic network with RMSE values of 6.295, 6.417 and 6.456, respectively, with R^2 values around 60%.

For the second variable, TCC determined by HPLC, the best models were elastic network, ridge regression and SVMs with RMSE

values of 6.534, 6.584 and 6.534, respectively, and R^2 values around 40%.

Table 3: Performance values (RMSE and R^2) obtained for the different machine learning models trained with CIELAB data. The total carotenoid content (TCC) determined by spectrophotometry (Lambert-Beer formula), the TCC determined by HPLC and the total content of trans- β -carotene (the most abundant carotene in cassava roots) were used as response prediction variables. The parenthesis indicate the package specific method chosen for the simulation.

	CIELAB data					
	TCC		TCC HPLC		trans- β -carotene	
	Spectrophotometry					
	RMSE	R^2	RMSE	R^2	RMSE	R^2
Partial Least Squares (simpls)	6.990	0.6022	6.789	0.4142	4.731	0.2371
Support Vector Machines (e1071)	7.015	0.5350	6.645	0.3840	4.829	0.1506
Partial Least Squares (widekernelpls)	7.125	0.6221	6.696	0.3960	4.551	0.2794
Random Forest	6.647	0.5124	7.571	0.2938	5.148	0.1532
Elastic Net	6.456	0.5785	6.534	0.4129	4.787	0.1840
Partial Least Squares (pls)	6.939	0.5916	6.622	0.3946	4.667	0.2446
Ridge Regression (w/ FS)	6.638	0.5628	6.653	0.3895	4.802	0.2020
Ridge Regression	6.417	0.5681	6.584	0.4213	4.886	0.1774
Support Vector Machines (kernelb)	7.294	0.5040	6.534	0.3662	4.878	0.2043
Partial Least Squares (kernelpls)	7.121	0.5827	6.756	0.4319	4.785	0.2278
Linear Regression	6.295	0.5933	6.749	0.4004	4.937	0.2424
K-Nearest Neighbors	6.636	0.5336	7.278	0.2569	4.997	0.2036
Lasso	6.412	0.5503	6.669	0.4110	4.826	0.1539
Conditional Inference Random Forest	8.162	0.4385	6.930	0.4085	4.667	0.2066
Conditional Inference Tree	9.388	0.3063	7.307	0.3842	4.934	0.1105
Decision Trees	9.990	0.2679	7.641	0.3534	5.015	0.2880

Lower RMSE values were observed when using trans- β -carotene as response variable, with best performance models being PLS (*widekernelpls* and *pls* methods) and conditional inference random forests, with RMSE values of 4.551, 4.667 and 4.667, respectively. However, models showed a decrease in the fitting of the data with an R^2 around 25%.

Looking at the VIPs (supplementary material) it becomes clear which variables played the most important role in the prediction of carotenoid content in the cassava samples. The b^* parameter was relevant in 100% of the cases, which was somewhat expected, considering that the samples are widely distributed across the y axis in Figure 6, which corresponds to the b^* parameter. Looking at the same plot we can see that the a^* interval in which samples are distributed is not as wide, however, this parameter was relevant in about 56% of the

predictions. The L^* parameter was the least relevant of the three, with a VIP of 0%.

The only pre-processing method applied to CIELAB data was scaling, as the other methods would not make much sense considering they are aimed at spectral data. Scaling the data showed an increase in model performance, however quite limited (results are shown in supplementary material).

3.5 CAROTENOID CONTENT PREDICTION USING FUSION DATA

In Table 4 the performance values obtained by using a low-level fusion between UV- visible (400-500 ηm) and CIELAB data as input to the various machine learning models are shown. Similarly to the previous cases, the response prediction variables used were the total carotenoid content (TCC) determined by spectrophotometry (Lambert-Beer formula), the TCC determined by HPLC and the total content of trans- β -carotene.

Table 4: Performance values (RMSE and R^2) obtained for the different machine learning models trained with a fusion between UV-visible spectrophotometry and CIELAB data. The total carotenoid content (TCC) determined by spectrophotometry (Lambert-Beer formula), the TCC determined by HPLC and the total content of trans- β -carotene (the most abundant carotene in cassava roots) were used as response prediction variables. The parenthesis indicate the package specific method chosen for the simulation, with exception to the linear regression models.

	UV-visible (400-500 nm) + CIELAB data					
	TCC		TCC HPLC	trans- β -carotene		
	Spectrophotometry					
	RMSE	R^2	RMSE	R^2	RMSE	R^2
Partial Least Squares (simpls)	3.706	0.8746	6.082	0.5032	4.685	0.2545
Support Vector Machines (e1071)	3.875	0.8887	6.379	0.5477	4.353	0.3551
Partial Least Squares (widekernelpls)	4.017	0.9247	6.031	0.4448	4.592	0.2804
Random Forest	3.758	0.9444	7.114	0.3527	6.101	0.1621
Elastic Net	3.775	0.9179	6.160	0.6031	4.450	0.3256
Partial Least Squares (pls)	3.682	0.8931	6.187	0.4834	4.652	0.2530
Ridge Regression (w/ FS)	3.570	0.9298	5.981	0.5781	4.730	0.3430
Ridge Regression	4.839	0.8510	8.469	0.4723	5.627	0.2783
Support Vector Machines (kernelab)	4.612	0.8800	6.299	0.5249	4.436	0.4241
Partial Least Squares (kernelpls)	3.804	0.8312	6.010	0.5263	4.681	0.2745
Linear Regression (w/ Stepwise Selection)	4.718	0.7973	8.052	0.5295	4.909	0.2406
Linear Regression (w/ Forward Selection)	4.829	0.8743	8.279	0.4734	4.860	0.2492
Linear Regression (w/ Backwards Selection)	4.479	0.8020	6.385	0.5419	5.179	0.2966
K-Nearest Neighbors	6.412	0.6320	7.355	0.2622	4.996	0.1359
Lasso	4.983	0.8076	18.784	0.2545	13.821	0.1487
Conditional Inference Random Forest	6.663	0.7671	6.531	0.5158	4.645	0.3540
Conditional Inference Tree	7.566	0.6697	6.923	0.4351	4.870	0.2706
Decision Trees	8.021	0.6997	7.789	0.3427	5.221	0.2181

The results obtained using fusion data are similar to those explained in subsection 3.3 and subsection 3.4 in the sense that highest R^2 performance values and lowest RMSE values were obtained when using the TCC determined using spectrophotometric data as response prediction variable. Overall, there is an increase in model performance when comparing to the results obtained for UV-visible data alone.

The best model performance when using the TCC determined by spectrophotometry as response variable was achieved by ridge regression (with feature selection) and PLS (*pls* and *simpls* methods) models with RMSE values of 3.570, 3.682 and 3.706, respectively, and R^2 values around 90%.

Using the second variable, TCC determined by HPLC, the models that best performed were ridge regression (with feature selection) and PLS (*kernelpls* and *widekernelpls*) models, having RMSE values of 5.981, 6.010 and 6.031, respectively, with R^2 values around 50%.

Lower RMSE values were observed when using trans- β -carotene as response variable, similarly to previous cases, with best performance models being SVMs (*e1071* and *kernelab* methods) and elastic network with RMSE values of 4.353, 4.436 and 4.450, respectively, and R^2 values around 30%.

The VIPs computed for this case (supplementary material) showed that the variables which presented the most important role in the prediction of carotenoid content in the cassava samples were those of wavelength around 170nm (VIPs > 99%). Here, the CIELAB b^* parameter was relevant in about 65% of predictions, while the a^* and L^* parameters had a VIP close to zero.

The only preprocessing method applied to the fusion data was scaling, as the methods employed in subsection 3.3 are aimed at spectral data. Data filtering was also applied to the data. Both methods contributed to an overall increase in model performance when compared to the performance obtained using raw UV-visible data (supplementary material).

4 CONCLUSIONS

The present study has shown how CIELAB color measurement can be used as a fast and non-destructive method to calibrate for the total carotenoid content of cassava genotypes roots with acceptable prediction error. By performing a low-level fusion between UV-visible spectrophotometry and CIELAB data we demonstrated how data fusion

can lead to a better model performance for prediction when comparing to the use of a single data source, having similar results already been published [24]. Moreover, the UV-visible spectrophotometric profiles measured between 400-500nm and the consequent carotenoid content determination allowed the observation of a positive correlation between the color of the root pulp and the total carotenoid content, which is in accordance with data reported in the literature [19, 20 and 21]. This finding was more explicit when observing the projection of the fifty cassava root samples in the CIELAB color space plane, having several clusters been formed, where the highest values of b* (which stands for the yellow coloration) and a* (which stands for the red coloration) were associated to the samples with highest carotenoid content.

In addition, the information obtained by coupling the analysis of pro-vitamin A biochemical markers to bioinformatics tools helps supporting the rational design of biochemically-assisted breeding programs of *M. esculenta*, that aim to obtain cultivars with high levels of pro-vitamin A carotenoids and superior nutritional traits.

Acknowledgements

To CNPq (National Counsel of Technological and Scientific Development) for financial support (Process n 407323/2013-9), to CAPES (Coordination for the Improvement of Higher Education Personnel (CAPES)), and EPAGRI (Agricultural Research and Rural Extension Company of Santa Catarina). The research fellowship from CNPq on behalf of M. Maraschin is acknowledged. The work is partially funded by Project PropMine, funded by the agreement between Portuguese FCT (Foundation for Science and Technology) and Brazilian CNPq.

5 REFERENCES

- [1] Tanumihardjo SA, Palacios N, Pixley KV, et al. Provitamin A carotenoid bioavailability: what really matters? International Journal for Vitamin and Nutrition Research. 2010; 80(4):336.
- [2] Rodriguez-Amaya DB, et al. A guide to carotenoid analysis in foods. ILSI press Washington, DC; 2001.

- [3] Pence HE, Williams A. ChemSpider: an online chemical information resource. *Journal of Chemical Education*. 2010; 87(11):1123–1124.
- [4] Stahl W, Sies H. Antioxidant activity of carotenoids. *Molecular aspects of medicine*. 2003;24(6):345–351.
- [5] La Frano MR, Woodhouse LR, Burnett DJ, Burri BJ. Biofortified cassava increases β -carotene and vitamin A concentrations in the TAG-rich plasma layer of American women. *British Journal of Nutrition*. 2013;110 (2):310–320.
- [6] Sánchez T, Ceballos H, Dufour D, Ortiz D, Morante N, Calle F, et al. Prediction of carotenoids, cyanide and dry matter contents in fresh cassava root using NIRS and Hunter color techniques. *Food chemistry*. 2014;151:444–451.
- [7] Sánchez T, Cha´vez AL, Ceballos H, Rodriguez-Amaya DB, Nestel P, Ishitani M. Reduction or delay of post-harvest physiological deterioration in cassava roots with higher carotenoid content. *Journal of the Science of Food and Agriculture*. 2006; 86 (4):634–639.
- [8] Brockes A. Commission Internationale de LEclairage.
- [9] Schanda J. *Colorimetry: understanding the CIE system*. John Wiley & Sons; 2007.
- [10] Weatherall IL, Coombs BD. Skin color measurements in terms of CIELAB color space values. *Journal of investigative dermatology*. 1992; 99(4):468–473.
- [11] Liu W, Ji J, Chen H, Ye C. Optimal color design of psychological counseling room by design of experiments and response surface methodology. *PLoS one*. 2014;9(3):e90646.
- [12] Singh Y, Bhatia PK, Sangwan O. A review of studies on machine learning techniques. *International Journal of Computer Science and Security*. 2007;1(1):70–84.

- [13] Domingos P. A few useful things to know about machine learning. *Communications of the ACM*. 2012;55(10):78–87.
- [14] Fourati H. *Multisensor Data Fusion: From Algorithms and Architectural Design to Applications*. vol. 1. Boca Raton, FL: CRC Press, Taylor & Francis Group LLC; 2016.
- [15] Rodriguez-Amaya DB, Kimura M, et al. *HarvestPlus handbook for carotenoid analysis*. vol. 2. International Food Policy Research Institute (IFPRI) Washington; 2004.
- [16] [16] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria; 2017. Available from: .
- [17] Costa C, Maraschin M, Rocha M. An R package for the integrated analysis of metabolomics and spectral data. *Computer methods and programs in biomedicine*. 2016; 129:117–124.
- [18] Kuhn M. *Caret package*. *Journal of Statistical Software*. 2008; 28(5):1–26.
- [19] Champagne A, Bernillon S, Moing A, Rolin D, Legendre L, Lebot V. Carotenoid profiling of tropical root crop chemotypes from Vanuatu, South Pacific. *Journal of food composition and analysis*. 2010; 23(8):763–771.
- [20] Chávez AL, Sánchez T, Jaramillo G, Bedoya J, Echeverry J, Bolanõs E, et al. Variation of quality traits in cassava roots evaluated in landraces and improved clones. *Euphytica*. 2005; 143 (1):125–133.
- [21] Iglesias C, Mayer J, Chavez L, Calle F. Genetic potential and stability of carotene content in cassava roots. *Euphytica*. 1997; 94 (3):367–373.
- [22] Kljak K, Grbe s̃a D, Karolyi D. Reflectance colorimetry as a simple method for estimating carotenoid content in maize grain. *Journal of Cereal Science*. 2014; 59 (2):109–111.

- [23] Meléndez-Martínez AJ, Britton G, Vicario IM, Heredia FJ. Relationship between the colour and the chemical structure of carotenoid pigments. *Food Chemistry*. 2007; 10 (3):1145–1150.
- [24] Botwey RH, Daskalaki E, Diem P, Mougiakakou SG. Multi-model data fusion to improve an early warning system for hypo-/hyperglycemic events. In: *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE; 2014. p. 4843–4846.

CAPÍTULO IV

SCREENING OF PHYSICOCHEMICAL TRAITS AND PASTING PROPERTIES OF STARCHES OF BRAZILIAN CASSAVA GENOTYPES (*Manihot esculenta* Crantz)

Screening of physicochemical traits and pasting properties of starches of Brazilian cassava genotypes (*Manihot esculenta* Crantz)

Rodolfo Moresco¹, Virgílio G. Uarrotta¹, Eduardo da C. Nunes², Edna Regina Amante³, Carlos Eduardo M. Campos⁴, Christopher Costa⁵, Miguel Rocha⁵ and Marcelo Maraschin¹

¹Plant Morphogenesis and Biochemistry Laboratory, Department of Plant Science, Federal University of Santa Catarina, Florianópolis, Brazil.

²Santa Catarina State Agricultural Research and Rural Extension Agency (EPAGRI), Experimental Station of Urussanga, Urussanga, Brazil

³Laboratory of Fruits and Vegetables, Department of Food Science and Technology, Federal University of Santa Catarina, Florianópolis, Brazil.

⁴Laboratory of X-ray Diffraction, Department of Physics, Federal University of Santa Catarina, Florianópolis, Brazil.

⁵Centre Biological Engineering, School of Engineering, University of Minho, Braga, Portugal.

Abstract. This study aimed to characterize cassava genotypes cultivated in southern Brazil, emphasizing their starch composition and physicochemical and pasting properties. A wide range of variation in amylose contents among genotypes was detected, with amounts ranging from 15.30 to 27.22%. All samples of cassava starch analyzed exhibited X-ray diffraction pattern type B, with values of crystallinity index dispersed between 37.3 to 49.1%. The RVA profile curves showed initial pasting temperatures varying from 63.25 to 65.35°C, with maximum viscosity temperatures between 98.31 and 99.74°C. The setback ranged from 627 to 965 mPa s. A further chemometric analysis was applied to a Fourier Transform Infrared Spectroscopy (FTIR) dataset of the cassava starch samples. The classification of genotypes via PCA using the carbohydrate fingerprint region clearly showed differences in their carbohydrate composition. The analytical techniques employed, associated with chemometrics allowed distinguishing the genotypes according to their physicochemical and functional peculiarities. Thus, industrial and food applications can be thought according to the genotypes' traits herein described.

Keywords: Cassava; Starch; Chemometrics; X-ray diffractometry; IR spectroscopy.

1 INTRODUCTION

Cassava (*Manihot esculenta* Crantz) ranks as the third more important crop as source of calories in human diet worldwide, along with the basic food crops including rice and corn. The most important part of the plant is the root, rich in starch, used in food or as raw material for various industries. This species has a very important nutritional aspect, as it serves as basic energy food in the diet of more than 800 million people in many countries (Howeler, R., Litaladio, N., and Thomas, 2013). Cassava affords important yields in fertility-deficient soils and is tolerant to plant diseases, drought, and pests (Cock, 1982; Pinto-Zevallos *et al.*, 2016). In this context, exploiting the potential of cassava starch has proved a highly remunerative activity, despite the technological barriers such as the demand for more productive genetic materials with industrial and nutritional superior quality. The expansion of the cassava starch market depends on varieties with good cooking traits and that meet the requirements of the industry. This scenario emphasizes the importance of seeking adapted genetic material, with better nutritional and technological characteristics increasing the possibility of to add value to cassava roots.

The starch fraction is basically composed of two polysaccharides, i.e., amylose [composed of glucose units connected by α -(1,4) glycosidic linkages, forming a linear chain] and amylopectin [composed of α -(1,4) linked glucose units with α -(1,6) branching points forming a branched structure] (Whistler & Daniel, 1984). The ratio of contents of these two macromolecules, that ultimately determines specific starch paste characteristics, is variable according to the botanical source and even among genotypes of the same species (Singh *et al.*, 2003). Amylose and amylopectin differ in the degree of hydration and hydrothermal behavior under heating, showing functional properties peculiar to different types of starches of industrial interest (Zhu, 2015).

In southern Brazil, the Santa Catarina Agricultural Research and Rural Extension Company (EPAGRI - Urussanga county) maintains a cassava germplasm bank with hundreds of genetic accesses coming from other research institutions in Brazil and abroad, to support its ongoing cassava genetic breeding program. However, the almost total lack of information on the chemical composition and nutritional value of the genotypes is thought to be a major drawback for the development of new cultivars.

Thus, in connection with a wider investigation on the metabolomics of *M. esculenta*, this study aimed the characterization of

ten genotypes of the EPAGRI's germplasm bank with emphasis on their starch composition. The claimed genetic variability of the genotypes is inferred to be associated to the eventual physicochemical and functional diversity of their root's starchy fraction, which could ultimately addresses optimized industrial applications for that raw material.

2 MATERIAL AND METHODS

2.1 SELECTION OF GENOTYPES

Roots of ten *M. esculenta* genotypes belonging to the EPAGRI's germplasm bank (Experimental Station of Urussanga, latitude 28°31'18'' S, longitude 49°19'03'' W, Santa Catarina State, southern Brazil) were sampled in the 2014/2015 season. Three commercial-size roots per clone were harvested. The genotypes have been traditionally called *Mesa*, *Pioneira*, *Oriental*, *Amarela*, *Catarina*, *IAC*, *Salézio*, *Estação*, *Videira*, and *Rosada*. The ten genotypes were indicated by EPAGRI plant breeders team because they are the most important and preferred by local small farmers for the commercial production of starch, containing an important physicochemical variability of the genotypes present in the germplasm bank.

2.2 STARCH ISOLATION

The roots, after being properly washed to remove soil were peeled and ground with the aid of a manual grater. Two liters distilled / deionized water were added to the biomass, followed by stirring and filtering on two overlapping sieves (0,250 mm) containing a filter element (cotton cheesecloth) in bilayer. The biomass retained was collected and washed with distilled water, for three times, to obtain a colorless filtrate, i.e. high purity starch. The filtrate was transferred to a 5 L beaker and allowed to stand for 5 h for starch decantation. Subsequently, the supernatant was removed and the starch deposited was re-suspended in 2 L distilled / deionized water, repeating the procedure twice for complete removal of the soluble material and increased purity of the starch fraction. The starch was dried in an oven with air circulation, at 45°C, for 48 h. Starches samples corresponding to each genotype after drying were standardized to a 0.250 mm particle size, packed in sealed vials, and stored at room temperature, protected from light and humidity until further analysis.

2.3 PHYSICOCHEMICAL AND FUNCTIONAL PROPERTIES

2.3.1 Amylose content

The amylose content of starch samples was determined using the colorimetric procedure described by Williams *et al.*, (1970). The absorbance of the solutions was recorded in a UV-visible spectrophotometer (Gold Spectrum lab 53 UV-Vis spectrophotometer, BEL photonics, Italy) at 625 nm. The contents of amylose were calculated through an external standard curve ($r^2 = 0.99$). Amylose and amylopectin were fractionated from cassava starch according to (McCready and Hassid, 1943), previously to building the external standard curve.

2.3.2 Absolute density

The granule size test for density determination was performed as proposed by Chang (1988), with the aid of a helium gas pycnometer (Quantachrome®, model Multipycnometer MVP - 4DC).

2.3.3 Oil and water absorption

The absorption of oil and water by the starch samples was determined as described by Beuchat (1977).

2.3.4 Swelling power of starch

The starch swelling power of the genotypes in study was determined by the method of Leach & Schoch (1962).

2.3.5 Scanning Electron Microscopy (SEM)

The size and shape of the starch granules were analyzed by scanning electron microscopy (SEM). To this end, samples (~10 mg) were deposited on double face carbon tapes fixed on aluminum sample holders (stubs) and coated with a thin gold layer (350Å, Polaron E5001). The micrographs were recorded in a scanning electron microscope model JEOL JSM-6390LV (JEOL, Tokyo, Japan), operating at 10 kV, with a magnification range of 500 x to 2000 x

2.3.6 Size of the starch granules

The granule size was measured from scanning micrographs with magnification of 500x and 2000x (5 images per sample and 10

measurements per image), with the aid of the ImageJ® software (Wayne Rasband, National Institute of Health, USA).

2.4 X-RAY POWDER DIFFRACTION (XRPD)

The X-ray powder diffraction measurements of the starch samples were carried out in a X'Pert PRO Multi-Purpose diffractometer (PanAnalytical, The Netherlands), equipped with Cu K α radiation ($\lambda = 1,5418\text{\AA}$) and a fast linear detector (Xcelerator) from 4° to 50° in the 2 θ range with 0.334° step size and 10s step time.

2.4.1 Crystallinity Index (CI)

The relative crystallinity index was determined based on the model described by Matveev *et al.*, 2001. For that, a reference baseline was used for the diffraction pattern from amorphized cassava starch sample. The amorphized starch sample was obtained via mechanical grinding, using a high energy ball mill (SPEX 8000). The milling time required to complete amorphization of the sample was 90 min. The crystallinity index (CI) was defined as the ratio between the crystalline area (A_c) and the total area covered by the curve, composed by the sum of the areas of the crystalline region (A_c) and amorphous region (A_a), according to the equation:

$$CI(\%) = \frac{A_c}{(A_c + A_a)} \times 100$$

The estimated crystallinity of starch samples was also calculated through FTIR spectroscopy as previously described by Sevenou *et al.* (2002) and Kuhnén *et al.* (2010).

2.5 FOURIER TRANSFORM INFRARED SPECTROSCOPY (FTIR)

FTIR spectra of the starch samples from the cassava genotypes were collected on a Bruker IFS-55 spectrometer (Model Opus. v 5.0, Bruker Biospin, Germany) equipped with a DGTS detector and a golden gate single reflection diamond attenuated total reflectance accessory (45° incidence-angle). A background spectrum of the clean crystal was acquired previously collecting five spectra per sample (128 co-added scans before Fourier transform) in a 4000-400 cm^{-1} spectral window, with a resolution of 4 cm^{-1} .

2.6 PASTING PROPERTIES

Variations of the consistency of the starch suspensions during pasting under the effect of temperature and agitation were recorded on a Rapid Visco Analyser (RVA - Newport Scientific, Narabeen, NSW, Australia). The starch suspensions were prepared with 140 g/L (dry weight basis) with distilled-deionized water. The suspensions were heated from 50°C to 95°C at a rate of 3°C/min, remaining at 95°C for 5 min, followed by cooling at 50°C at the same ratio. The rotation speed was fixed at 3g. From the graphic profiles the pasting temperature, maximum viscosity, temperature at maximum peak viscosity, time to achieve the maximum peak viscosity, minimal viscosity, final viscosity at cooling cycle, and setback were calculated. The setback is the measure of retrogradation tendency or syneresis upon cooling of cooked starch pastes. This variable was determined by subtracting the values of the final viscosity from the minimal one.

2.7 STATISTICAL ANALYSIS

Data were collected, summarized and submitted to analysis of variance (ANOVA), followed by Tukey's test ($P < 0.05$). All procedures were performed at least in triplicate ($n = 3$). Furthermore, the ATR-FTIR spectral dataset was subjected to multivariate statistical analysis using the methods of principal component analysis (PCA) and clustering analysis. For that, scripts were written in R language (v. 3.1.1) (R Core Team, 2008) using tools defined in our group and available in the R *specmine* package (Costa *et al.*, 2015).

All R scripts, raw data, and additional analysis are available in supplementary material, as well as the data analysis report automatically generated from the R scripts using the features provided by R Markdown. Furthermore, they are also available at [_](#). This allows fully reproducing and documenting the experiments.

3 RESULTS AND DISCUSSION

3.1 PHYSICOCHEMICAL AND FUNCTIONAL PROPERTIES

3.1.1 Amylose content

The cassava genotypes showed to differ ($p < 0.05$) in their starch composition, especially as to their amylose and amylopectin amounts in the roots. A wide range of amylose content was detected, ranging from 15.30% (genotype *IAC*) to 27.22% (genotype *Videira*) (**Table 1**).

Similar findings have been observed in cassava genotypes through the amylopectin precipitation technique via complexation reaction with concanavalin A, e.g., 0% by waxy genotypes and 30% in high-amylose genotypes (Rolland-Sabaté *et al.*, 2012), 15.2 to 26.5% (Sanchez *et al.*, 2009) and even lower values, 6 to 18.4% (Gomand *et al.*, 2010). On the other hand, amylose contents from 25.95 to 32.09% have been reported by Nunes & Jesus (2009) using the same methodology employed in this work.

Starches with superior amylose contents such as found in *Videira*, *Amarela*, and *Mesa* genotypes are known for their properties in the formation of films and for presenting a lower digestibility compared to those with higher concentrations of amylopectin (Rindlav-Westling *et al.*, 1998). In fact, high amylose content is desired when the product needs to be crisp and sturdy, while the amylopectin allows better maintenance of the gel shape, giving greater resistance to additional treatments, such as cutting and drying (Sajilata and Singhal, 2005).

3.1.2 Absolute density

In this study, the average density of the starches was 1.53 g.cm⁻³, with a range of 1.43 (genotype *Mesa*) to 1.64 g.cm⁻³ (genotype *IAC*). Statistical differences ($p < 0.05$) for this variable have been found among the genotypes as shown in **Table 1**. Cassava starches produced in the semiarid region of Bahia state (northeastern Brazil) presented absolute density values varying from 1.50 to 1.58 g.cm⁻³ (Nunes and Jesus, 2009). Absolute density is an important characteristic of the granule to predict its susceptibility to enzymatic hydrolysis when the purpose is produce maltodextrin. The arrangement of amylose and amylopectin chains in the granules leads to the formation of zones of high or low density. The region where amylopectin is found is more dense or crystalline. Because it is more compact, this region is less hydrated and more resistant to the hydrolysis process (Kim *et al.*, 2008). The amorphous regions of the starch granules, mainly formed by amylose, are less dense, absorbing more water at temperatures below the gelatinization temperature (Singh *et al.*, 2003).

Table 1 - Physicochemical and functional properties of starches of ten cassava genotypes cultivated in southern Brazil.

Genotypes	Amylose content (%)	Granule density (g/cm³)	Granule size (μm)	Swelling power at 60°C (%)	Lipid absorption (mL/g)	Water absorption (mL/g)	Crystallinity (X-RPD - %)
Mesa	25.14 ± 0.12b	1.43 ± 0.02e	12.98 ± 0.98ab	8.82 ± 1.14bcde	1.29 ± 0.17ab	3.36 ± 0.24abcd	38.47 ± 2.95a
Pioneira	18.00 ± 0.03g	1.47 ± 0.002de	11.99 ± 1.53ab	7.80 ± 0.44de	1.46 ± 0.18a	2.92 ± 0.26cd	45.07 ± 3.19a
Oriental	20.87 ± 0.06f	1.49 ± 0.003de	13.61 ± 1.44a	9.35 ± 0.92abcde	1.37 ± 0.05ab	3.62 ± 0.001a	41.09 ± 5.96a
Amarela	25.20 ± 0.13b	1.49 ± 0.004d	11.40 ± 3.11ab	7.69 ± 0.12e	1.47 ± 0.16a	3.53 ± 0.15ab	42.90 ± 1.83a
Catarina	24.30 ± 0.02c	1.56 ± 0.01bc	12.20 ± 0.92ab	8.50 ± 0.59cde	1.20 ± 0.02ab	3.00 ± 0.19bcd	43.04 ± 6.84a
IAC	15.30 ± 0.06h	1.64 ± 0.08a	13.19 ± 3.12a	9.43 ± 0.54abcd	1.27 ± 0.06ab	3.36 ± 0.01abcd	49.19 ± 3.43a
Salézio	22.75 ± 0.19e	1.54 ± 0.004bcd	10.82 ± 1.02ab	9.06 ± 0.08abcde	1.27 ± 0.04ab	3.49 ± 0.10abc	37.32 ± 6.60a
Estação	20.79 ± 0.14f	1.50 ± 0.002cd	10.26 ± 2.52ab	10.21 ± 0.18ab	1.32 ± 0.06ab	2.85 ± 0.45d	45.70 ± 3.98a
Videira	27.22 ± 0.15a	1.48 ± 0.01de	10.05 ± 2.71ab	10.65 ± 0.52a	1.26 ± 0.03ab	2.95 ± 0.08bcd	41.07 ± 6.43a
Rosada	23.47 ± 0.14d	1.59 ± 0.02ab	8.64 ± 1.99b	9.53 ± 0.13abc	1.12 ± 0.12b	3.86 ± 0.01a	42.17 ± 3.02a

3.1.3 Oil and water absorption

The investigated cassava genotypes showed values of oil and water absorption varying from 1.12 to 1.47 mL.g⁻¹ and from 2.85 to 3.86 mL.g⁻¹, respectively (**Table 1**). Interestingly, the *Rosada* genotype showed higher ($p < 0.05$) water absorption and lower oil absorption values. Low water absorption is an important trait in starches for the production of biscuits, while the formulation of sausage, bread dough, and processed cheese requires starches with high water absorption capacity, (Adebowale *et al.*, 2005). Otherwise, the ability of starch to adsorb oil may be indicative of its emulsifying potential being an important feature since fats act as taste retainers, making these starches more suitable for enhancing the flavor in food formulation (Kaushal *et al.*, 2012). Furthermore, it has been described that starches with high amylose content are suitable to reduce oil absorption in fried products due to its film-forming properties (Brannan *et al.*, 2014; Boettger, Marotta, & Murray, 1971).

3.1.4 Swelling power (SP)

Differences in the swelling power of starches indicate discrepancies in the intra-molecular organization of the granules of the genotypes investigated. In the present study, SP values of 7.69 (genotype *Amarela*) to 10.65% (genotype *Videira*), at 60°C, were detected (**Table 1**). These values corroborate the findings pointing out that cassava starches exhibit weak intra-molecular organization compared to others starch sources, indicated by the low relaxation temperature of hydrogen bonds and rapid increase in granules (swelling) at low temperatures (60°C, for instance) (Nwokocha *et al.*, 2009).

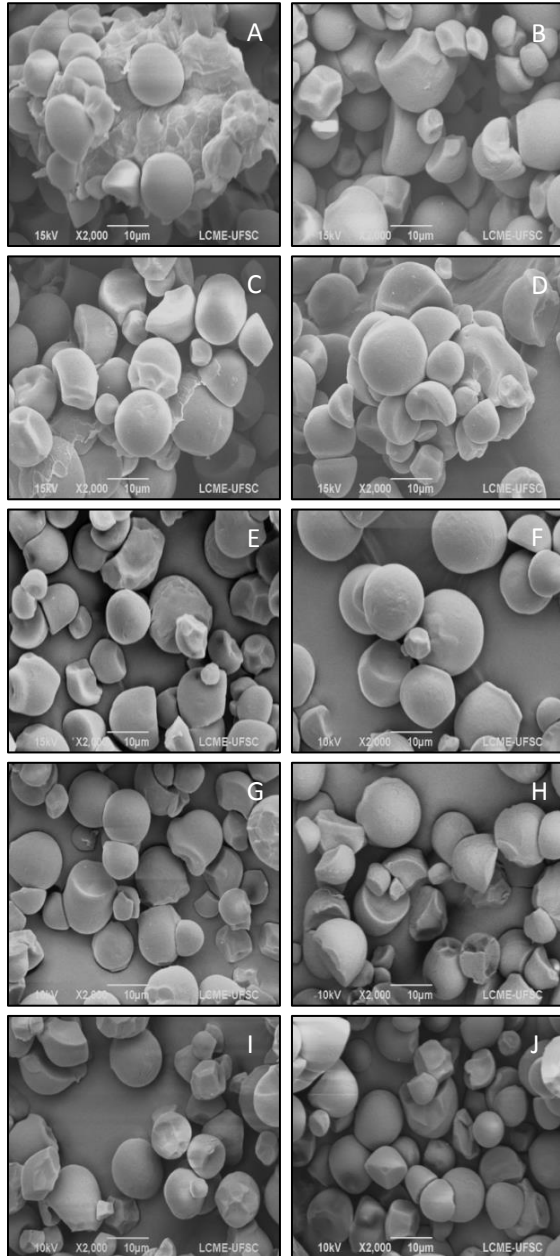
3.1.5 Shape and granule size

Scanning electron microscopy analysis revealed morphological details of the starch granules (**Fig. 1 A-J**), showing typical shapes of cassava starch, i.e., rounded, oval, convex or biconcave granules and of varying sizes (Zhu, 2015). It was also possible to detect a certain proportion of the starch granules damaged, probably due to the extraction method adopted, and the presence of a fine parenchymal structure involving granules (**Figs. 1A and 1C**), rich in pectin and cellulosic materials. The sizes of granules ranged from 8.64 (*Rosada*) to 13.62 μm (*Oriental*) (**Table 1**), a narrower amplitude as compared with the granule diameter observed for cassava starch by (Charoenkul *et al.*,

2011), varying from 4.8 to 22.8 μm , as other authors reported a range from 7 – 20 μm (Rolland-Sabaté *et al.*, 2012).

These variables are important in determining the potential usage of starches, e.g., small granules may be used as fat substitutes due to their size being similar to oil droplets. Other applications concern the production of biodegradable films and paper. A study of the properties of cohesion and packaging has shown that cassava starch, which is usually rounded and regular in shape, exhibited lower form factor promoting a better particle packing, being useful in the production of capsules for pharmaceutical purposes (Itiola and Odeku, 2005). Chitedze (2012) also showed that cassava variety is a critical choice for the ultimate quality of ibuprofen tablets, affecting its mechanical binding, disintegration, friability, and hardness characteristics.

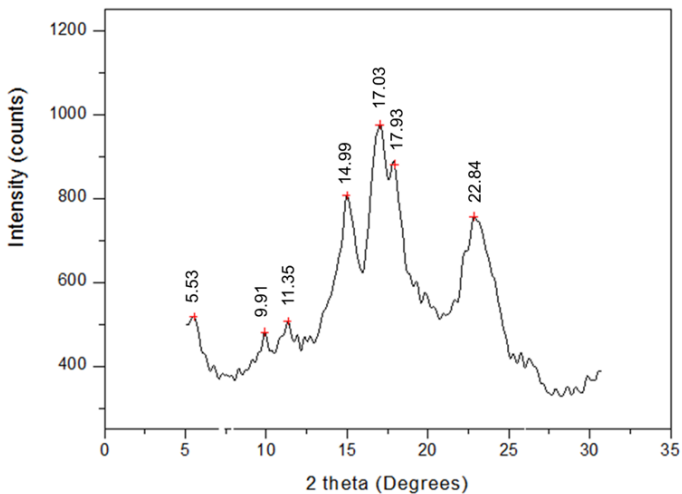
Fig. 1 - Micrographs (SEM, A-J) of starch granules of cassava genotypes roots. (A) Mesa, (B) Pioneira, (C) Oriental, (D) Amarela, (E) Catarina, (F) IAC, (G) Salézio, (H) Estação, (I) Videira, and (J) Rosada.



3.2 X-RPD MEASUREMENTS AND CRYSTALLINITY INDEX

The XRPD data of the studied cassava genotypes showed to be quite similar. **Fig. 2** shows a typical XRPD pattern of the studied genotypes. Three most intense and sharpest Bragg peaks located at about $2\Theta \approx 15^\circ$, 17° , 18° were detected, as well as a large and also intense peak at 22.9° . The last one seems to have shoulders at low and high angles. Moreover, some additional peaks at $2\Theta \approx 5.5^\circ$ (incomplete), 9.9° , and 11.3° could also be observed. All cassava starch samples exhibited the type B diffraction pattern, characterized by three most intense Bragg peaks located at about $2\Theta \approx 5^\circ$, 15° , 17° , 18° , and 23° and some additional peaks at $2\Theta \approx 9.9^\circ$ and 11.3° (Lopez-Rubio *et al.*, 2008). **Table 1** presents the values for crystallinity index calculated from the XRPD data of the cassava genotypes. The values dispersed between 37.3% (*Salézio*) to 49.1% (*IAC*), with an average crystallinity of 42.6%. Mutungi, Passauer, Onyango, Jaros, & Rohm (2012) reported a range of crystallinity indexes for cassava starch from 33.4% to 58.2%, taking into consideration an amorphous model as reference to the baseline diffraction pattern. Amylopectin plays a major role in the starch granules crystallinity (Pérez & Bertoft, 2010; Rolland-Sabaté *et al.*, 2012; Zhu, 2015). The results herein found are consistent with this fact as the *IAC* genotype, the richest in amylopectin content, showed the highest crystallinity index. In the pharmaceutical industry, for example, starch is widely used as an excipient. A certain degree of crystallinity is required for starch maintaining the specificity of drug release and other functional properties (Pifferi *et al.*, 1999). According to Airaksinen *et al.* (2005), crystalline excipients added to pharmaceutical formulations can accelerate the polymorphic transformations of the active substances, while the amorphous ones can delay such transformations.

Fig. 2 - Typical X-ray powder diffractogram of a starch sample of the studied cassava genotypes. The sharp peaks correspond to reflections of the B-type crystalline polymorph.



3.3 PASTING PROPERTIES – RAPID VISCO ANALYSER (RVA)

The starch viscoamylographic profiles of the genotypes under study showed typical curves for tuberous starches and were similar in shape (**Supplementary material – Fig. S1**). The pasting properties of the starch samples are shown in **Table 2**. The viscosity curves show that the initial paste temperature varied from 63.25° to 65.35°C, with maximum viscosity temperatures at 98.31-99.74°C. Maximum and minimum viscosity intervals from 2951 to 3260 mPa s and from 1462 to 1704 mPa s were found to occur, respectively. Sánchez, Dufour, Moreno, & Ceballos (2010) noted values of maximum viscosity of 876-1006 mPa s in cassava starches and paste formation temperature of 63.3 to 64.8°C. Higher pasting temperatures were found in cereal starches such as maize (84°C) and rice (89°C). Finally, at the end of cooling time, a wide range of viscosity (2211 mPa s to 2669 mPa s) and setback (retrogradation tendency of amylose in a starch paste) (627 mPa s to 965mPa s) were detected among the studied genotypes.

Table 2²³ - Pasting characteristics of starch samples of ten cassava genotypes cultured in the southern Brazil.

Genotypes	Pasting temperature (°C)	Peak viscosity (mPa s)	Peak time (min)	Temperature at peak viscosity (°C)	Minimum viscosity (mPa s)	Final viscosity (mPa s)	Breakdown (mPa s)	Setback (mPa s)
Mesa	65.25	3260	5.93	99.48	1704	2699	1556	965
Pioneira	64.15	2996	5.27	99.74	1462	2275	1534	813
Oriental	63.25	3170	5.20	99.61	1578	2211	1592	633
Amarela	65.30	3179	5.13	99.48	1497	2310	1682	813
Catarina	64.25	3029	5.87	98.70	1588	2484	1441	896
IAC	63.55	2951	5.73	98.96	1608	2235	1343	627
Salézio	65.15	3100	5.67	99.74	1541	2408	1559	867
Estação	65.35	2999	6.07	99.87	1693	2550	1306	857
Videira	63.40	3001	6.47	98.31	1655	2571	1346	916
Rosada	65.15	3037	5.93	98.70	1588	2379	1449	791

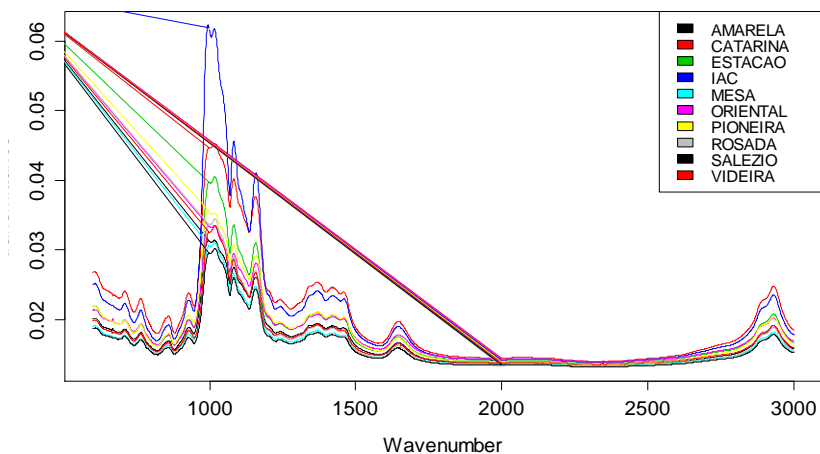
²Values expressed as mean \pm standard deviations of three independent experiments (n=3). Distinct letters in column represent statistical differences for the mean values at $P < 0.05$, by Tukey test.

³ Pasting temperature (°C), peak viscosity (mPa s), peak time (min), temperature at peak viscosity (°C), minimum viscosity (mPa s), final viscosity (mPa s), breakdown (mPa s, peak viscosity - minimum viscosity), and setback (mPa s, final viscosity - minimum viscosity).

3.4 INFRARED SPECTROSCOPY (FTIR)

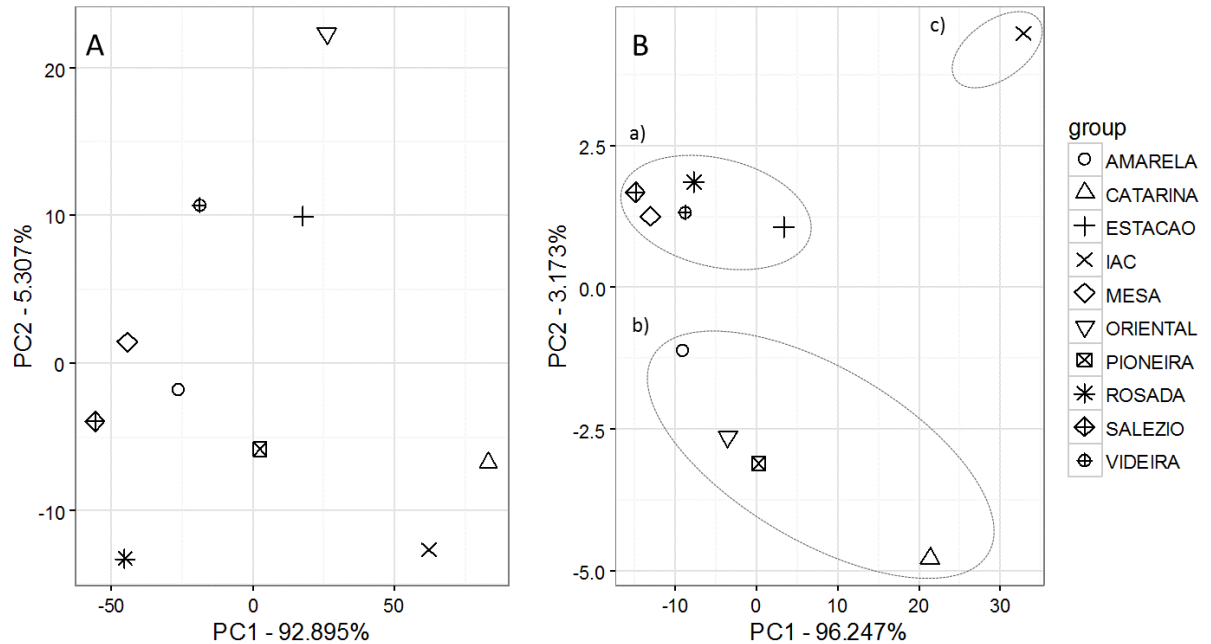
The FTIR spectral profiles (**Fig. 3**) of the starch samples revealed the presence of several chemical constituents in the region of $3000\text{ cm}^{-1} - 700\text{ cm}^{-1}$, where signs commonly associated to proteins, oils, amylose, and amylopectin were clearly identified. In general, carbohydrates show high absorbance values in the infrared spectral window between $1200\text{ cm}^{-1} - 950\text{ cm}^{-1}$, a fingerprint region of these primary metabolites (Černá *et al.*, 2003). In the studied samples, the highest intensity peaks detected (**Fig. 3**) were associated to the carbohydrate compounds, an expected finding since starch is the major constituent of cassava roots. Besides, signals associated to protein constituents were also detected at 1650 cm^{-1} and 1550 cm^{-1} related to the primary amines. The presence of carboxyl groups in the samples, associated to the axial deformation of the $\text{C}=\text{O}$ functional group typically found in fatty acids was detected at 1740 cm^{-1} . Additionally, bands at $2927\text{ cm}^{-1} - 2885\text{ cm}^{-1}$, corresponding to signals of aliphatic hydrocarbons were identified, which are possibly related to fatty acids or oils in the samples, and derived from the axial and angular deformation of the methyl and/or methylene group(s) of the alkyl portion of these compounds.

Fig. 3 – ATR-FTIR spectra of starches of ten cassava genotypes in the $3000\text{--}600\text{ cm}^{-1}$ wavenumber region.



Only the visual analysis of the FTIR spectra profiles is not efficient to extract all the information available due to the amount and complexity of this data. A typical FTIR spectrum often contains 4000 data points and such a dimensionality makes difficult the extraction of additional information relevant to the objectives of the present study. Thus, as a strategy to reduce data dimension without losing relevant information, principal component analysis was applied to the FTIR dataset (3000-600 cm^{-1} – **Fig. 4 - A**). PC1 and PC2 contributed to explain 98.6 % of the variance present in the spectral dataset. However, a clear discrimination of the samples was not found, rather samples occurred spread out by all the sections of the factorial distribution graph. The factorial contributions analysis revealed that mostly the bands at 970 cm^{-1} , 1180 cm^{-1} , and 2925 cm^{-1} contributed to the sample dispersion along PC1, revealing distinct chemical compositions of the starch fraction of the cassava genotypes. The eigenvectors values revealed a greater effect of carbohydrates (970 cm^{-1} and 1180 cm^{-1}) and oils (2925 cm^{-1}) on the classification model of genotypes built via PCA..

Fig. 4 - A- Factorial distributions of PC1 and PC2 for the spectral data (ATR-FTIR, 3000-600cm⁻¹) of cassava starches produced in southern Brazil. PC1 and PC2 contributed to explain 98.6 % of the variance present in the spectral dataset. B- Factorial distribution of PC1 and PC2 for the ATR-FTIR (1200-950.cm⁻¹) carbohydrate fingerprint region dataset. PC1 and PC2 explained 99.6% of the existing variance



Although the whole spectra provided a good starting point for an exploratory data analysis in a metabolic fingerprint approach, it is often necessary to analyze specific regions of the spectra to understand the discrimination obtained by PCA. Therefore, a complementary experimental procedure for the calculation of principal components considered the carbohydrate fingerprint region dataset, i.e., 1200 cm^{-1} - 950 cm^{-1} . In this case, the samples could be clearly discriminated by PCA potentially reflecting differences in their carbohydrate composition. **Fig. 4 - B** shows that PC1 and PC2 accounted for 99.6% of the existing variance, clearly revealing the existence of two genotype groups according to their similarities, namely the genotypes *Rosada*, *Estação*, *Videira*, *Salézio*, and *Mesa* in PC2+ (**Fig. 4 - B** – ellipse **a**), as the *Amarela*, *Pioneira*, *Oriental*, and *Catarina* genotypes grouped in PC2- (**Fig. 4 - B** – ellipse **b**). Furthermore, the genotype *IAC* (**Fig. 4 - B** – ellipse **c**) clearly differed from the others, suggesting a peculiar metabolic profile and chemical composition. The eigenvectors values correlated to PC2 refer to the bands at 1015 cm^{-1} and 970 cm^{-1} (data not show), commonly assigned to amylose and amylopectin, respectively. It is clear that the classification and grouping of genotypes via PCA using the fingerprint region for carbohydrates is consistent with the results obtained by calculating the crystallinity through FTIR spectroscopy, as previously described (Sevenou, Hill, Farhat, & Mitchell, 2002; Kuhnen *et al.*, 2010) (**Supplementary material - Table S1**) and XRD (**Table 1**). The genotypes *Mesa*, *Videira*, *Rosada*, *Salézio*, and *Estação*, with higher amylose content and lower ratio $\text{IR } 999\text{ cm}^{-1}/1118\text{ cm}^{-1}$ were shown to be grouped into PC2+ (**Fig. 4 - B** - ellipse **a**), suggesting that the IR spectroscopy is a rapid and effective analytical tool for the characterization of cassava starches. In a second series of experiments, cluster analysis was also applied to the FTIR dataset. The similarities were defined based on the Euclidean distance between two samples using the arithmetic average (UPGMA) method. Genotypes with the highest similarity in their FTIR spectral dataset are represented by a hierarchical clustering analysis in **Fig. S2** for carbohydrate fingerprint region and in **Fig. S3** for full FTIR spectral dataset. In this approach, the genotype *IAC* is presented as a branch grouped into the cluster with the genotype *Catarina*, revealing similarities between these two genotypes. Further PCA and cluster analysis for the spectral windows typical of oils ($3000\text{-}2800\text{ cm}^{-1}$, **Fig. S4** and **Fig. S5**) and proteins ($1650\text{-}1500\text{ cm}^{-1}$, **Fig. S6** and **Fig. 7**) also allowed discriminating the samples, especially the genotypes *IAC* and *Catarina*, proving they have distinct features

regarding those primary metabolites from the other samples of the population.

4 CONCLUSIONS

The methodological approach used in this study allowed us to identify and characterize the chemical variability of the cassava genotypes. Thus, industrial and food applications can be thought according to the genotypes' traits herein described. For example, *Videira*, *Amarela*, and *Mesa* genotypes, with higher levels of amylose, have the properties required when a product needs to be crunchy, sturdy, and with low oil absorption, e.g., the manufacture of chips. Furthermore, it is well known that starches with high amylose contents are associated with lower glycemic response, lower blood cholesterol levels, and slower emptying of the gastrointestinal tract, important traits for the food industry pursuing nutraceutical properties in edible biomasses. Similarly, cassava starches with more regular and rounded shape granules promote a better packing of the particles, being useful to the pharmaceutical industry in the formulation of capsules and tablets containing bioactive compound(s). In addition, a certain degree of crystallinity is required for the starch to maintain the specificity of drug release and other functional properties. A lower retrogradation tendency (e.g., *IAC*, *Oriental*, and *Rosada* genotypes) confers to the starch the property of forming a more stable paste, being advantageous in food products such as soups and sauces, which are subject to loss of viscosity and precipitation.

The vibrational infrared spectroscopy coupled to chemometrics revealed discrepancies in the chemical profiles of genotypes and seems to be an interesting analytical approach for a fast and effective characterization of cassava starch. The ATR-FTIR spectra of the *IAC* and *Catarina* genotypes showed to be discrepant in respect to the other ones, suggesting having peculiar metabolic profiles. It is assumed, therefore, that the concomitant use of the metabolomics technologies employed in this work is of interest for investigating the chemical and structural characteristics of the starch fractions originating from *M. esculenta* germplasm banks. Taken together, the results can be applied as support tools for cassava biochemically assisted genetic breeding programs, optimizing the selection and maintenance process of genetic diversity in germplasm banks, for instance.

5 REFERENCES

Adebowale, K.O., Olu-owolabi, B.I., Olayinka, O.O. & Olayide, S. (2005). Effect of heat moisture treatment and annealing on physicochemical properties of red sorghum starch. *African Journal of Biotechnology*, **4**, 928–933.

Airaksinen, S., Karjalainen, M., Shevchenko, A., Westermarck, S., Leppänen, E., Rantanen, J. & Yliruusi, J. (2005). Role of water in the physical stability of solid dosage formulations. *Journal of Pharmaceutical Sciences*, **94**, 2147–2165.

Beuchat, L.R. (1977). Functional and electrophoretic characteristics of succinylated peanut flour protein. *Journal of Agricultural and Food Chemistry*, **25**, 258–261.

Boettger R.M., Marotta N.G. & Murray D.G. (1971). Novel amylose coatings for deep fried potato products. *US Patent 3,597,227*, 1–4. Retrieved from <http://www.google.com/patents/US3269586>

Brannan, R.G., Mah, E., Schott, M., Yuan, S., Casher, K.L., Myers, A. & Herrick, C. (2014). Influence of ingredients that reduce oil absorption during immersion frying of battered and breaded foods. *European Journal of Lipid Science and Technology*, **116**, 240–254.

Černá, M., Barros, A.S., Nunes, A., Rocha, S.M., Delgadillo, I., Čopíková, J. & Coimbra, M.A. (2003). Use of FT-IR spectroscopy as a tool for the analysis of polysaccharide food additives. *Carbohydrate Polymers*, **51**, 383–389.

Chang, C.S. (1988). Measuring density and porosity of grain kernels. *Cereal Chemistry*, **65**, 13–15.

Charoenkul, N., Uttapap, D., Pathipanawat, W. & Takeda, Y. (2011). Physicochemical characteristics of starches and flours from cassava varieties having different cooked root textures. *LWT - Food Science and Technology*, **44**, 1774–1781.

Chitedze, J. (2012). Binding effect of cassava starches on the compression and mechanical properties of ibuprofen tablets. *Journal of Applied Pharmaceutical Science*, **2**, 31–37.

Cock, J.H. (1982). Cassava: a basic energy source in the tropics. *Science*, **218**, 755 LP-762.

Costa, C., Maraschin, M. & Rocha, M. (2015). An R package for the integrated analysis of metabolomics and spectral data. *Computer Methods and Programs in Biomedicine*, **129**, 117–124.

Gomand, S.V., Lamberts, L., Derde, L.J., Goesaert, H., Vandeputte, G.E., Goderis, B., Visser, R.G.F. & Delcour, J.A. (2010). Structural properties and gelatinisation characteristics of potato and cassava starches and mutants thereof. *Food Hydrocolloids*, **24**, 307–317.

Howeler, R., Lutaladio, N., and Thomas, G. (2013). *Save and grow cassava: a guide to sustainable production intensification*, FAO, p. 142.

Itiola, O.A. & Odeku, O.A. (2005). Packing and cohesive properties of some locally extracted starches. *Tropical Journal of Pharmaceutical Research*, **4**, 363–368.

Kaushal, P., Kumar, V. & Sharma, H.K. (2012). Comparative study of physicochemical, functional, antinutritional and pasting properties of taro (*Colocasia esculenta*), rice (*Oryza sativa*) flour, pigeonpea (*Cajanus cajan*) flour and their blends. *LWT - Food Science and Technology*, **48**, 59–68.

Kim, J.Y., Park, D.J. & Lim, S.T. (2008). Fragmentation of waxy rice starch granules by enzymatic hydrolysis. *Cereal Chemistry*, **85**, 182–187.

Kuhnen, S., Ogliari, J.B., Dias, P.F., Boffo, E.F., Correia, I., Ferreira, A.G., Delgadillo, I. & Maraschin, M. (2010). ATR-FTIR spectroscopy and chemometric analysis applied to discrimination of landrace maize flours produced in southern Brazil. *International Journal of Food Science & Technology*, **45**, 1673–1681.

Leach, H.W. & Schoch, T.J. (1962). Leach 1962 - Starch Solubility DMSO. *Cereal chemistry*, **39**, 318–326.

Lopez-Rubio, A., Flanagan, B.M., Gilbert, E.P. & Gidley, M.J. (2008). A novel approach for calculating starch crystallinity and its correlation with double helix content: a combined XRD and NMR study. *Biopolymers*, **89**, 761–8.

Matveev, Y.I., Soest, J.J.G. Van, Nieman, C., Wasserman, L.A., Protserov, V.A., Ezernitskaja, M. & Yuryev, V.P. (2001). Relationship between thermodynamic and structural properties of low and high amylose maize starches. *Carbohydrate Polymers*, **44**, 151–160.

McCready, R. & Hassid, W. (1943). The separation and quantitative estimation of amylose and amylopectin in potato starch. *Journal of the American Chemical Society*, **65**, 1154–1157.

Mutungi, C., Passauer, L., Onyango, C., Jaros, D. & Rohm, H. (2012). Debranched cassava starch crystallinity determination by Raman spectroscopy: Correlation of features in Raman spectra with X-ray diffraction and ¹³C CP/MAS NMR spectroscopy. *Carbohydrate Polymers*, **87**, 598–606.

Nunes, L.B. & Jesus, W. (2009). Rendimento de extração e caracterização química e funcional de féculas de mandioca de região do semi-árido baiano. *Alimentos e Nutrição Araraquara*, **20**, 129–134.

Nwokocha, L.M., Aviara, N.A., Senan, C. & Williams, P.A. (2009). A comparative study of some properties of cassava (*Manihot esculenta*, Crantz) and cocoyam (*Colocasia esculenta*, Linn) starches. *Carbohydrate Polymers*, **76**, 362–367.

Pérez, S. & Bertoft, E. (2010). The molecular structures of starch components and their contribution to the architecture of starch granules: A comprehensive review. *Starch - Stärke*, **62**, 389–420.

Pifferi, G., Santoro, P. & Pedrani, M. (1999). Quality and functionality of excipients. *Farmaco*, **54**, 1–14.

Pinto-Zevallos, D.M., Pareja, M. & Ambrogi, B.G. (2016). Current knowledge and future research perspectives on cassava (*Manihot esculenta* Crantz) chemical defenses: An agroecological view. *Phytochemistry*, **130**, 10–21.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.

Rindlav-Westling, A., Stading, M., Hermansson, A.-M. & Gatenholm, P. (1998). Structure, mechanical and barrier properties of amylose and amylopectin films. *Carbohydrate Polymers*, **36**, 217–224.

Rolland-Sabaté, A., Sánchez, T., Buléon, A., Colonna, P., Jaillais, B., Ceballos, H. & Dufour, D. (2012). Structural characterization of novel cassava starches with low and high-amylose contents in comparison with other commercial sources. *Food Hydrocolloids*, **27**, 161–174.

Sajilata, M.G. & Singhal, R.S. (2005). Specialty starches for snack foods. *Carbohydrate Polymers*, **59**, 131–151.

Sánchez, T., Dufour, D., Moreno, I.X. & Ceballos, H. (2010). Comparison of pasting and gel stabilities of waxy and normal starches from potato, maize, and rice with those of a novel waxy cassava starch under thermal, chemical, and mechanical stress. *Journal of Agricultural and Food Chemistry*, **58**, 5093–9.

Sanchez, T., Salcedo, E., Ceballos, H., Dufour, D., Mafla, G., Morante, N., Calle, F., Pérez, J.C., Debouck, D., Jaramillo, G. & Ximena Moreno, I. (2009). Screening of starch quality traits in cassava (*Manihot esculenta* Crantz). *Starch - Stärke*, **61**, 12–19.

Sevenou, O., Hill, S.E., Farhat, I. a & Mitchell, J.R. (2002). Organisation of the external region of the starch granule as determined by infrared spectroscopy. *International Journal of Biological Macromolecules*, **31**, 79–85.

Singh, N., Singh, J., Kaur, L., Sodhi, N.S. & Gill, B.S. (2003). Morphological, thermal and rheological properties of starches from different botanical sources. *Food Chemistry*, **81**, 219–231.

Whistler, R.L. & Daniel, J.R. (2000). Starch. In: *Kirk-Othmer Encyclopedia of Chemical Technology*. John Wiley & Sons, Inc.

Williams, P.C., Kuzina, F.D. & Hynka, I. (1970). A rapid colorimetric procedure for estimating the amylose content of starches and flours. *Cereal Chemistry*, **47**, 411–420.

Zhu, F. (2015). Composition, structure, physicochemical properties, and modifications of cassava starch. *Carbohydrate Polymers*, **122**, 456–480.

SUPPLEMENTARY FIGURES

Fig. S1 - Typical RVA pasting curves under Rapid Visco Analyser (RVA) of cassava starch samples of genotype cultivated in southern Brazil.

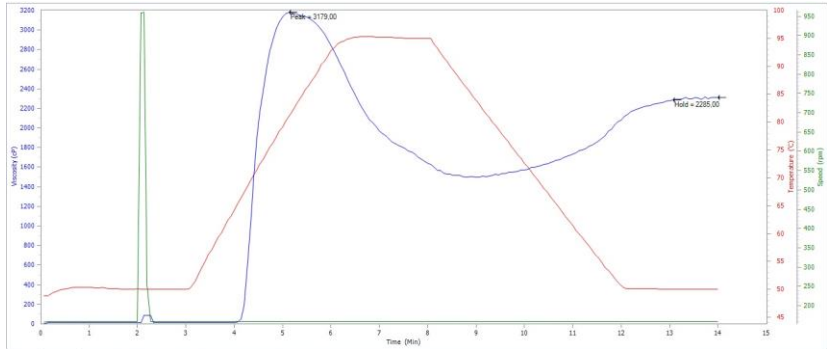


Table S1 - Infrared (IR) ratio of absorbances at 999 cm^{-1} and 1018 cm^{-1} for the starches of the studied cassava genotypes.

Samples	IR ratio 999/1018 (cm^{-1})
Mesa	0.986
Videira	0.977
Rosada	0.981
Salézio	0.981
Estação	0.986
Catarina	0.991
Oriental	0.998
Amarela	0.998
IAC	1.003
Pioneira	1.005

Fig.S2 - Similarity of cassava genotypes starches in respect to their carbohydrate fingerprint region in the ATR-FTIR spectra (1200-950.cm⁻¹). Hierarchical cluster dendrogram analysis (UPGMA method).

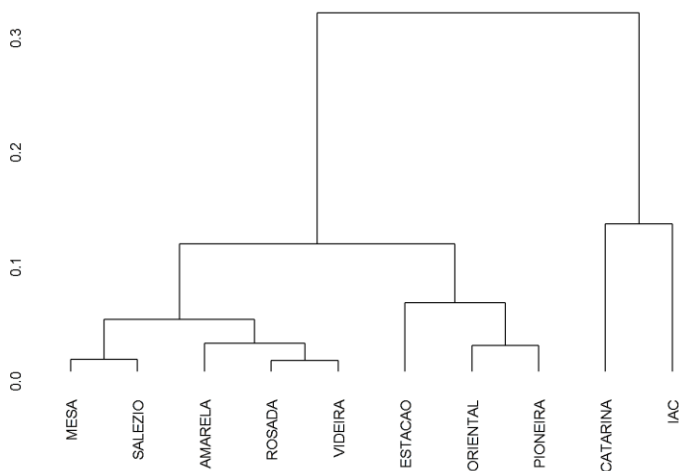


Fig. S3 - Similarity of cassava genotypes starches in respect to their ATR-FTIR spectral dataset. Hierarchical cluster dendrogram analysis (UPGMA method).

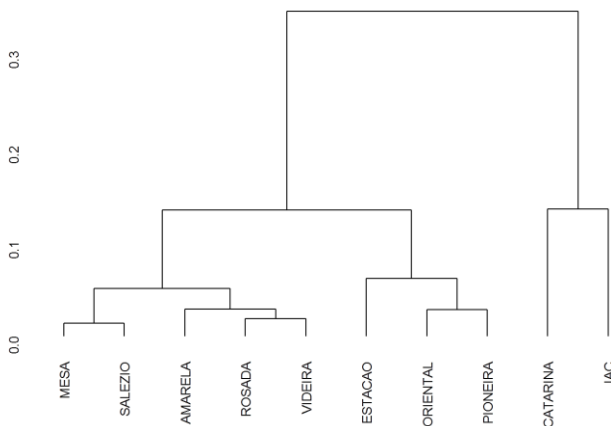


Fig. S4 - Factorial distributions of PC1 and PC2 for the ATR-FTIR spectroscopic data at the lipid fingerprint region (2800-3000cm⁻¹).

PC1 and PC2 contributed to explain 99.97 % of the variance present in the spectral dataset.

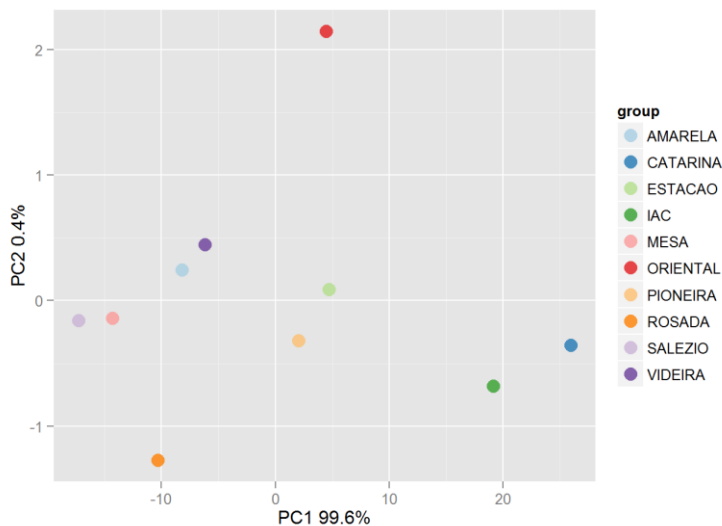


Fig. S5 - Similarity of cassava genotypes starches in respect to their ATR-FTIR spectroscopic dataset at the 2800-3000cm⁻¹ (lipid fingerprint region). Hierarchical cluster dendrogram analysis (UPGMA method).

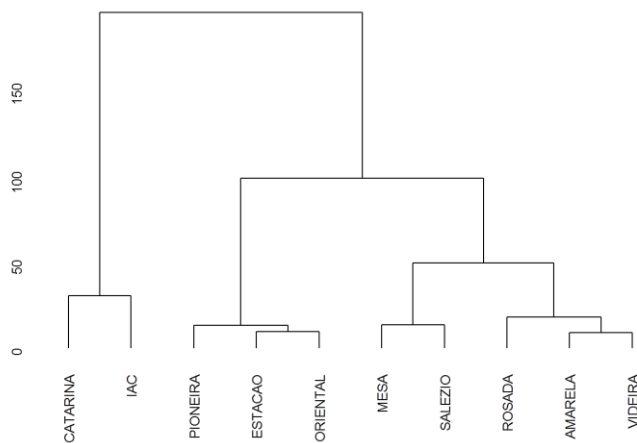


Fig. S6 - Factorial distributions of PC1 and PC2 for the ATR-FTIR spectral data in the protein fingerprint region, 1500-1650cm⁻¹. PC1 and PC2 contributed to explain 99.8 % of the variance present in the spectral dataset.

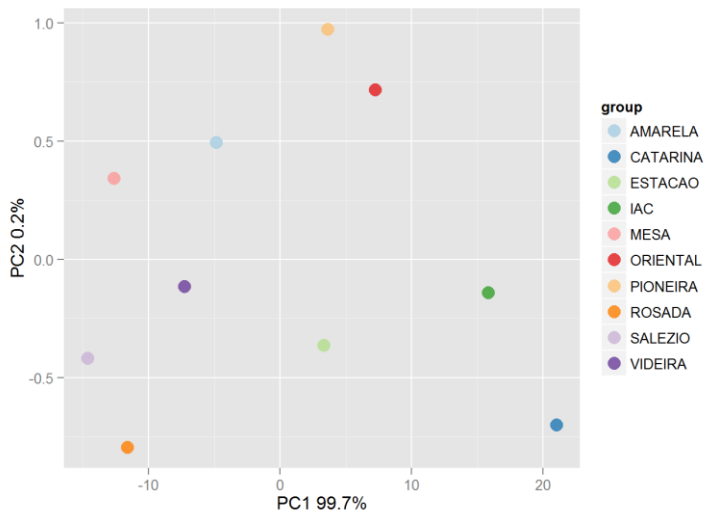
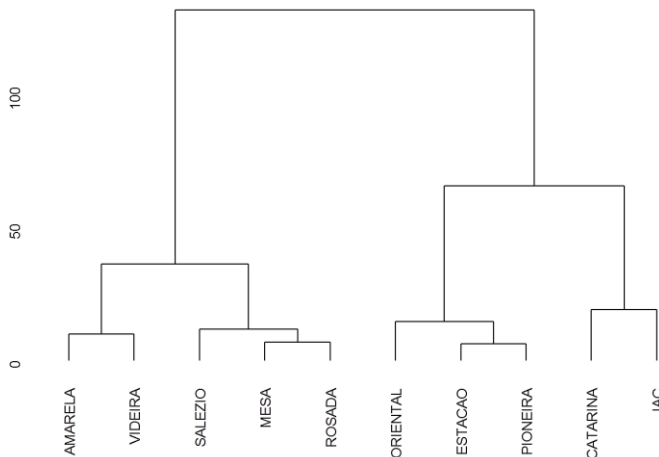


Fig. S7 - Similarity of cassava genotypes starches in respect to their ATR-FTIR spectral dataset at 1500-1650cm⁻¹ (protein fingerprint region). Hierarchical cluster dendrogram analysis (UPGMA method)



CAPÍTULO V

CONSIDERAÇÕES FINAIS

1 CONCLUSÕES GERAIS

A ampla disparidade na variabilidade e no conteúdo de carotenoides (i.e. *cis*- e *trans*- β -caroteno, α -caroteno, luteína, β -criptoxantina e licopeno) e na composição amilácea (i.e. amilose/amilopectina) observada, revela a variabilidade química dos genótipos analisados.

Genótipos do banco de germoplasma da Epagri e cultivados por agricultores familiares no estado de Santa Catarina exibiram quantidades apreciáveis de carotenoides, indicando seu potencial como fonte de compostos de interesse à saúde humana e nutrição, dada à presença de carotenoides provitamina A (β -caroteno, e.g.) e licopeno nas raízes de coloração amarela e rosada, respectivamente.

A abordagem metodológica utilizada neste trabalho permitiu identificar e caracterizar a variabilidade química dos genótipos analisados. Além disto, as técnicas analíticas empregadas e os ensaios físico-químicos realizados, associados às análises quimiométricas e ferramentas de bioinformática, permitiu um melhor entendimento da variabilidade química associada à composição carotenóidica e amilácea das raízes dos genótipos de mandioca (*Manihot esculenta* Crantz) estudados.

Assume-se, portanto, que o uso concomitante das tecnologias empregadas neste trabalho seja de grande interesse à investigação das características estruturais da fração amilácea e carotenóidica oriunda de bancos de germoplasma de *M. esculenta*. Tomados em conjunto, os resultados obtidos podem ser empregados como ferramentas de apoio aos programas de melhoramento genético de mandioca assistidos bioquimicamente, otimizando o processo de seleção e também de manutenção da diversidade genética existente em bancos de germoplasma e em poder dos produtores rurais. Adicionalmente, estes resultados enfatizam a importância da manutenção *on farm* destes genótipos pelos pequenos agricultores familiares do sul do Brasil e por bancos de germoplasma, com vistas à sua preservação, já que representam uma importante fonte de recursos fitogenéticos com características nutricionais e industriais promissoras.

1 PRINCIPAIS CONTRIBUIÇÕES CIENTÍFICAS E/OU BIOTECNOLÓGICAS

Bioprospecção de linhagens de genótipos mantidos em bancos de germoplasma), com teores superiores de metabólitos-alvo (β -caroteno e licopeno), para exploração racional via programas de melhoramento genético em projetos de biofortificação;

Caracterização de perfis de metabólitos primários (amido) e secundários (carotenoides) de genótipos de mandioca, permitindo a identificação de candidatos a marcadores bioquímicos de interesse à seleção assistida e de avaliação de genótipos;

Geração de uma ferramenta de discriminação de genótipos com base em fenótipos bioquímicos (i.e., perfis espectrais de FTIR, Raman, UV-vis, DRX e cromatográficos - carotenoides) e análise quimiométrica, caracterizada por seu baixo custo, rapidez e elevada precisão.

O conjunto de perfis espectrais dos genótipos em estudo viabilizará a criação de um banco de dados de referência à análise comparativa de outros acessos/variedades, bem como a discriminação daqueles de maior interesse nutricional;

Tomados em conjunto, espera-se que os resultados obtidos possam ser usados como ferramentas de apoio essencial ao programa de melhoramento genético clássico de mandioca, no sentido de torná-lo mais eficiente no processo de seleção e também de manutenção da diversidade genética. Adicionalmente, os resultados alcançados fornecerão informações importantes quanto a aspectos nutracêuticos que podem agregar valor à matéria-prima e produtos derivados de determinados genótipos da espécie alvo.

2 PUBLICAÇÕES VINCULADAS À TESE DE DOUTORAMENTO

2.1 ARTIGOS COMPLETOS PUBLICADOS EM PERIÓDICOS

- **MORESCO, R.; UARROTA, V. G.; PEREIRA, A.; TOMAZZOLI, M.; NUNES, E. D. C., PERUCH, L. A. M.; GAZZOLA, J.; COSTA, C.; ROCHA, M., MARASCHIN, M.** UV-visible scanning spectrophotometry and chemometric analysis as tools for carotenoids analysis in cassava genotypes (*Manihot esculenta* Crantz). **Journal of Integrative Bioinformatics**, 12(4): 280, 2015. [http:// doi:10.2390/biecoll-jib-2015-280](http://doi:10.2390/biecoll-jib-2015-280)

- MARQUEZI, M.; GERVIN, V. M.; WATANABE, L. B.; **MORESCO, R.**; AMANTE, E. R. CHEMICAL AND FUNCTIONAL PROPERTIES OF DIFFERENT COMMON BRAZILIAN BEAN (*Phaseolus vulgaris* L.) CULTIVARS. **Brazilian Journal of Food and Technology**, v. 20, p. 1-6, 2017.
- UARROTA, V. G., **MORESCO, R.**, COELHO, B., NUNES, E. D. C., PERUCH, L. A. M., NEUBERT, E. D. O., MARASCHIN, M. Metabolomics combined with chemometric tools (PCA, HCA, PLS-DA and SVM) for screening cassava (*Manihot esculenta* Crantz) roots during postharvest physiological deterioration. **Food Chemistry**, 161, 67–78, 2014.
- TOMAZZOLI, M.; PAI NETO, R. D.; **MORESCO, R.**; WESTPHAL, L.; SOMENSI-ZEGGIO, A.; SPECHT, L.; COSTA, C., ROCHA, M.; MARASCHIN, M. Discrimination of Brazilian propolis according to the seasoning using chemometrics and machine learning based on UV-Vis scanning data. **Journal of Integrative Bioinformatics**, 12(4): 279, 2015. <http://doi:10.2390/biecoll-jib-2015-279>
- UARROTA, V. G., NUNES, E. D. C., PERUCH, L. A. M., NEUBERT, E. D. O., COELHO, B., **MORESCO, R.**, ... MARASCHIN, M. Toward better understanding of postharvest deterioration: biochemical changes in stored cassava (*Manihot esculenta* Crantz) roots: **Food Science & Nutrition**, n/a–n/a, 2015. <http://doi.org/10.1002/fsn3.303>
- UARROTA, V. G., **MORESCO, R.**, SCHMIDT, E. C., BOUZON, L. Z., NUNES, E. D. C., NEUBERT, E. D. O., PERUCH, L. A. M., ROCHA, M., MARASCHIN, M. Data supporting the role of enzymes and polysaccharides during cassava postharvest physiological deterioration. **Food Chemistry Data in Brief**, v. 6, p. 503-506, 2016. <http://doi:10.1016/j.dib.2015.12.043>
- UARROTA, V. G., **MORESCO, R.**, SCHMIDT, E. C.,

BOUZON, L. Z., NUNES, E. D. C.; NEUBERT, E. D. O., PERUCH, L. A. M, ROCHA, M., MARASCHIN, M. The role of ascorbate peroxidase, guaiacol peroxidase, and polysaccharides in cassava (*Manihot esculenta* Crantz) roots under postharvest physiological deterioration. **Food Chemistry**, v. 197, p. 737-746, 2015.

2.2 ARTIGOS COMPLETOS ACEITOS PARA PUBLICAÇÃO

- **MORESCO, RODOLFO**; AFONSO, TELMA; UARROTA, VIRGÍLIO G.; NAVARRO, BRUNO BACHIEGA; NUNES, EDUARDO DA C.; ROCHA, MIGUEL; MARASCHIN, MARCELO. UV-vis and CIELAB based chemometric characterization of *Manihot esculenta* carotenoid contents. **Journal of Integrative Bioinformatics**, v na, p na, 2017.

2.3 CAPÍTULOS DE LIVROS INTERNACIONAIS PUBLICADOS:

- **MORESCO, RODOLFO**; AFONSO, TELMA; UARROTA, VIRGÍLIO G.; NAVARRO, BRUNO BACHIEGA; NUNES, EDUARDO DA C.; ROCHA, MIGUEL; MARASCHIN, MARCELO. Classification Tools for Carotenoid Content Estimation in *Manihot esculenta* via Metabolomics and Machine Learning. **Advances in Intelligent Systems and Computing**. 616 ed.: Springer International Publishing, 2017, v. 616, p. 280-288.
- **MORESCO, R.**; UARROTA, VIRGÍLIO G.; PEREIRA, ALINE.; TOMAZZOLI, M.; NUNES, E. D. C., PERUCH, L, A. M.; COSTA, C.; ROCHA, M., MARASCHIN, M. Carotenoid Analysis of Cassava Genotypes Roots (*Manihot esculenta* Crantz) Cultivated in Southern Brazil Using Chemometric Tools. **Advances in Intelligent Systems and Computing**. 1ªed. Springer International Publishing, 2015, v., p. 11-18.
- OLIVEIRA, EVA REGINA; WICKERT, ESTER; FERNANDA; **MORESCO, RODOLFO**; SIMÃO, LARISSA; NAVARRO, BRUNO B.; BAUER, CLAUDIA; CABRAL,

DÉBORA; ROCHA, MIGUEL; MARASCHIN, MARCELO. Influence of Solar Radiation on the Production of Secondary Metabolites in Three Rice (*Oryza sativa*) Cultivars. **Advances in Intelligent Systems and Computing**. 1ed. Berlin: Springer International Publishing, 2017, v. 1, p. 297-304

- LOPES, SUSANE; MORESCO, RODOLFO; PERUCH, LUIZ AUGUSTO MARTINS; ROCHA, MIGUEL; MARASCHIN, MARCELO. UV-Vis Spectrophotometry and Chemometrics as Tools for Recognition of the Biochemical Profiles of Organic Banana Peels (*Musa* sp.) According to the Seasonality in Southern Brazil. In: Fdez-Riverola, Florentino; Mohamad, Mohd Saberi; Rocha, Miguel; De Paz, Juan F; Pinto, Tiago. (Org.). **Advances in Intelligent Systems and Computing**. 1ed. Berlin: Springer International Publishing, 2017, v. 1, p. 289-296
- UARROTA, VIRGÍLIO GAVICHO; STEFEN, DEIVID L. V.; DE SOUZA, CLOVIS ARRUDA; MEDEIROS COELHO, CILEIDE MARIA; MORESCO, RODOLFO; MARASCHIN, MARCELO; SÁNCHEZ-MORA, FERNANDO DAVID; DA COSTA NUNES, EDUARDO; DE OLIVEIRA NEUBERT, ENILTO; MARTINS PERUCH, LUIZ AUGUSTO. **Advances in understanding cassava growth and development**. Burleigh Dodds Series in Agricultural Science. 1ed. Cambridge: Burleigh Dodds Science Publishing, 2017, p. 3-35.
- PEREIRA, A. MORESCO, R., MARASCHIN, M. Bioactive Compounds of Banana as Health Promoters. **Natural Bioactive Compounds from Fruits and Vegetables as Health Promoters**, 2015, p. 220-235.
- TOMAZZOLI, M.; PAI NETO, R. D.; MORESCO, R.; WESTPHAL, L.; SOMENSI-ZEGGIO, A.; SPECHT, L., COSTA, C., ROCHA, M.; MARASCHIN, M. UV-Visible Scanning Spectrophotometry and Chemometric Analysis as Tools to Build Descriptive and Classification Models for Propolis from Southern Brazil. **Advances in Intelligent Systems and Computing**. 1ª ed.: Springer International Publishing, 2015, v., p. 19-27.

2.4 APRESENTAÇÕES DE TRABALHO:

- 9th International Conference on Practical Applications of Computational Biology & Bioinformatics. Carotenoid Analysis of Cassava Genotypes Roots (*Manihot Esculenta* Crantz) Cultivated in Southern Brazil Using Chemometric Tools. PACBB'15 - University of Salamanca (Spain), 3rd to 4th June 2015.

3 AGRADECIMENTOS

Ao CNPq e Capes pelo apoio financeiro. (Edital 22/2010 – CNPq/Repensa. Desenvolvimento da Cadeia Produtiva da Mandioca na Região Centro-Sul do Brasil).

À FAPESC (Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina).

Este trabalho também é parcialmente financiado pelo projeto Propmine (Processo CNPq n°202772/2015-2), financiado pelo acordo entre FCT Português (UID/BIO/04469/2013) e CNPq/Brasil.