



**UNIVERSIDADE ESTADUAL DO CEARÁ
REDE NORDESTE DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA
DOUTORADO EM BIOTECNOLOGIA**

YNAYARA COLARES DE LIMA

**AVALIAÇÃO DO PERFIL METABÓLICO DE CAJU PROVENIENTE DE ACESSOS
DE *Anacardium occidentale* L. USANDO UPLC-QToF-MS, NIR E QUIMIOMETRIA**

**FORTALEZA - CEARÁ
2018**

YNAYARA COLARES DE LIMA

AVALIAÇÃO DO PERFIL METABÓLICO DE CAJU PROVENIENTE DE ACESSOS
DE *Anacardium occidentale* L. USANDO UPLC-QToF-MS, NIR E QUIMIOMETRIA

Tese apresentada ao Curso de Doutorado em Biotecnologia, do Programa de Pós-graduação em Biotecnologia da Rede Nordeste de Biotecnologia – RENORBIO, da Universidade Estadual do Ceará, como requisito parcial à obtenção do título de Doutor em Biotecnologia. Área de Concentração: Recursos Naturais.

Orientador: Prof. Dr. Edy Sousa de Brito
Co-orientadora: Dra. Ana Cecília Ribeiro de Castro

FORTALEZA - CEARÁ

2018

Dados Internacionais de Catalogação na Publicação

Universidade Estadual do Ceará

Sistema de Bibliotecas

Lima, Ynayara Colares de.

Avaliação do Perfil Metabólico de Caju Proveniente de Acessos de *Anacardium occidentale* L. Usando UPLC-QToF-MS, NIR e Quimiometria [recurso eletrônico] / Ynayara Colares de Lima. - 2018.

1 CD-ROM: il.; 4 ¼ pol.

CD-ROM contendo o arquivo no formato PDF do trabalho acadêmico com 83 folhas, acondicionado em caixa de DVD Slim (19 x 14 cm x 7 mm).

Tese (doutorado) - Universidade Estadual do Ceará, Rede Nordeste de Biotecnologia, Doutorado em Biotecnologia, Fortaleza, 2018.

Área de concentração: Recursos Naturais.

Orientação: Prof. Dr. Edy Sousa de Brito.

Coorientação: Prof.^a Dra. Ana Cecília Ribeiro de Castro.

1. Metabolômica. 2. Quimiometria. 3. Caju. I.
Título.




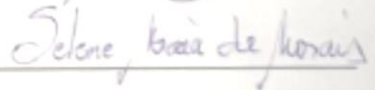

YNAYARA COLARES DE LIMA

AVALIAÇÃO DO PERFIL METABÓLICO DE CAJU PROVENIENTE DE ACESSOS DE *Anacardium occidentale* L. USANDO UPLC-QToF-MS, NIR E QUIMIOMETRIA

Tese apresentada ao Curso de Doutorado em Biotecnologia, do Programa de Pós-graduação em Biotecnologia da Rede Nordeste de Biotecnologia – RENORBIO, da Universidade Estadual do Ceará, como requisito parcial à obtenção do título de Doutor em Biotecnologia. Área de Concentração: Recursos Naturais.

Aprovada em: 31 de janeiro de 2018.

BANCA EXAMINADORA

BANCA EXAMINADORA:	CONCEITO	CPF	ASSINATURA
Prof. Dr. Edy Sousa de Brito, EMBRAPA (Presidente)	Satisfatório	565061554-34	
Dr. Delane da Costa Rodrigues, UFC (Examinadora)	Satisfatório	014914973-51	
Dr. Elenilson de Godoy Alves Filho, UFC (Examinador)	Satisfatório	886932951-87	
Dr. Selene Maia de Moraes, UECE (Examinadora)	Satisfatório	051679653-49	
Dr. Guilherme Julião Zocolo, EMBRAPA (Examinador)	Satisfatório	036611846-41	

AGRADECIMENTOS

A Embrapa pela oportunidade de desenvolver meu trabalho com todo o aporte necessário.

Ao meu orientador Dr. Edy Brito, pela confiança e por todo o ensinamento que pôde me transmitir.

A minha co-orientadora Dra. Ana Cecília Castro e a minha “mãe científica” Profa. Selene Morais por todo o apoio e encorajamento no decorrer do trabalho.

Ao Dão, que foi essencial nas coletas em Pacajus.

Aos pesquisadores e analistas que sempre estiveram disponíveis para me ajudar e sanar minhas dúvidas: Dr. Ebenézer Silva, Dr. Guilherme Zocolo, Dr. Kirley Canuto, Dr. Paulo Riceli, Dra. Lorena Silva e Dr. Elenilson Godoy.

A Dra. Izabel Florindo, a Maria Helena M. Viana e a Rafaela Cajado, por sempre me auxiliarem tão gentilmente sempre que precisei.

Aos professores da Renorbio por todo o conhecimento adquirido nas disciplinas.

A Banca Examinadora, que prontamente aceitou o convite e por todas as contribuições a minha tese.

Às companheiras de turma Delane Rodrigues e Roseane Procópio, por caminharem comigo nesse árduo trajeto.

A Luciana Bertini, por contribuir há anos pro meu crescimento pessoal e profissional.

A Gislene Mendes (Gigi), a quem conheço a tão pouco tempo, mas que se mostrou uma pessoa incrível, com quem eu posso contar em todos os momentos.

Ao Dearlan Pedrosa por ter me incentivado a fazer o doutorado desde o início.

Aos meus amigos e familiares que sempre me colocaram pra cima e me mostram, a cada passo que eu dou, que eu sou capaz.

A Carolina Dutra, por todas as ferramentas que me ensinou para concluir essa tese.

Às minhas tias Márcia e Lindalva por estarem presentes, ajudando sempre nos cuidados com minha filha.

A Deus, por ser meu refúgio nos momentos mais difíceis.

A minha filha, minha Lavínia, por ser meu incentivo a vencer todas as batalhas.

A minha mãe, Terezinha, e a minha “tia-mãe” Auxiliadora, por estarem sempre ao meu lado, me incentivando, me acolhendo, me amando e, principalmente, cuidando do meu bem mais precioso enquanto eu não podia estar presente.

“Quando tudo parecer dar errado em sua vida, lembre-se que o avião decola contra o vento, e não a favor dele”.

Henry Ford

RESUMO

O cajueiro (*Anacardium occidentale* L.) é uma frutífera de grande relevância no cenário socioeconômico do Brasil. Seus produtos (principalmente a castanha) possuem alto valor nutritivo e comercial. Devido a sua importância para o país e, principalmente, para o Nordeste, a Embrapa Agroindústria Tropical conta com o Banco de Germoplasma do Caju, o qual é composto por mais de 600 acessos diferentes. A fim de se ter um maior conhecimento da variabilidade metabólica dos acessos contidos no Banco de Germoplasma, genótipos do Banco de Germoplasma e do Programa de Melhoramento Genético do cajueiro foram analisados por meio de Cromatografia Líquida acoplada a Espectrometria de Massas (LC-MS) e Espectroscopia de Infravermelho Próximo. Foi criada uma matriz numérica utilizando os dados cromatográficos e as ferramentas quimiométricas de Análise de Grupamento Hierárquico (HCA), Análise de Componentes Principais (PCA) e análise discriminante por mínimos quadrados parciais (PLS-DA) foram aplicadas. Em HCA e PCA, observou-se a formação de quatro grupamentos, e os eixos PC1 e PC2 explicaram 36,7% da variância. O método PLS-DA foi empregado para melhorar a associação da variabilidade química de acordo com o genótipo do caju e, com três Variáveis Latentes (LVs), o modelo apresentou uma capacidade de predição de 88,05%. Os principais compostos responsáveis pela separação dos grupos foram identificados. São eles: galoilhexose I, digaloilhexosídeo I, éster etílico do ácido hidroxibutanóico-hexosídeo, miricetina-3-O-glucosídeo, e miricetina-3-O-rhamnosídeo. O espectrômetro portátil de Infravermelho Próximo (NIR) foi utilizado a fim de avaliar seu potencial para criar um modelo de classificação utilizando os teores de vitamina C, sólidos solúveis e acidez total. Porém, os resultados mostraram que outros parâmetros devem ser levados em conta para uma melhor compreensão da variabilidade química e física do caju do banco de germoplasma da Embrapa.

Palavras-chave: Metabolômica. Quimiometria. Caju.

ABSTRACT

Cashew tree (*Anacardium occidentale* L.) is an edible plant of great relevance in the socioeconomic scenario of Brazil. The cashew products (mainly cashewnut) have high nutritional and commercial value. Due to the relevance of cashew products to the country and mainly to the Northeast, Embrapa Agroindústria Tropical keeps a Germplasm Bank, which is composed of more than 600 different cashew accessions. In order to obtain a better knowledge of the metabolic variability of the accesses contained in the Germplasm Bank through metabolomic approaches. Twenty four genotypes of the Germplasm Bank and the Genetic Improvement Program of cashew were analyzed by liquid chromatography coupled to mass spectrometry (LC-MS). A numerical matrix was created using the chromatographic data and the chemometric tools of hierarchical grouping (HCA), main component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were applied. In HCA and PCA, the formation of four groups was observed, and the PC1 and PC2 axes explained 36.7% of the variance. The PLS-DA method was used to improve the association of chemical variability according to the cashew genotype and, with three latent variables (LVs), the model presented a prediction capacity of 88.05%. The main compounds responsible for the separation of the groups were galloylhexose I, digalloylhexoside I, hydroxybutanoic acid hexoside ethyl ester, myricetin-3-O-glucoside, and myricetin-3-O-rhamnoside. The near infrared portable spectrometer (NIR) was used in order to evaluate its potential to create a classification model using the levels of vitamin C, ° Brix, and total acidity. However, the results presented that other parameters should be taken into account for a better understanding of the chemical and physical variability of the cashew of the Embrapa germplasm bank.

Keywords: Metabolomics Chemometrics. Cashew.

LISTA DE FIGURAS

Figura 1 -	Cajueiro comum (esquerda) e anão precoce (direita).....	19
Figura 2 -	Evolução da produção mundial de castanha de caju, período de 2004 a 2014.....	21
Figura 3 -	Evolução da área com cajueiros colhida no Brasil, período de 1990 a 2015.....	21
Figura 4 -	Estruturas de aminoácidos identificados no caju.....	23
Figura 5 -	Estruturas de compostos voláteis identificados no caju.....	24
Figura 6 -	Estruturas de flavonóides identificados no caju.....	25
Figura 7 -	Estruturas de alquil fenóis identificados no caju.....	26
Figura 8 -	Estruturas de carotenóides identificados no caju.....	27
Figura 9 -	Banco de Germoplasma de Caju da Embrapa.....	28
Figura 10 -	Esquema de cromatografia líquida de alta eficiência.....	30
Figura 11 -	Sistema de UPLC.....	30
Figura 12 -	Esquema da espectrometria de massas.....	32
Figura 13 -	NIR portátil.....	33
Figura 14 -	Ilustração de um dendrograma.....	34
Figura 15 -	Representação de duas PC's no espaço.....	36
Figura 16 -	Total de publicações na área de metabolômica associada a análise multivariada.....	38
 Capítulo 1		
Figure 1 -	Total ion chromatograms from 24 different genotypes of cashew juices acquired under negative ionization mode.....	56
Figure 2 -	Dendrogram representing the chemical composition similarity relationships among the cashew.....	57
Figure 3 -	a) PC1 vs PC2 scores coordinate system for the cashew juices from different genotypes; b) respective loadings plotted in lines form. Samples were symbolized according to the clustering tendency: blue triangles at negative values of PC1 and PC2; red squares at positive values of PC1 and negative of PC2; black stars at positive values of PC1 and	

	PC2; green circles only at positive values of PC2. Samples that did not presented relevant results were symbolized by gray circle.....	58
Figure 4 -	Relative contributions of the isotopic forms in chromatograms for the compounds at retention times 1.60, 2.82, 3.23, 3.82, and 4.25 min.....	60
Figure 5 -	LV1 vs. LV2 vs. LV3 scores 3-D plot (a) and the respective loadings potted in lines form (b) from PLS-DA for the cashew juices classification. Legend: compounds 1 – galloylhexose I (1.60 min); 2 – digalloylhexoside I (2.82 min); 3 – hydroxybutanoic acid ethyl ester-hexoside (3.28 min); 4 – myricetin-3-O-glucoside (3.83 min); 5 – myricetin-3-O-rhamnoside (4.25 min).....	62
 Capítulo 2		
Figure 1 -	a) Raw absorbance spectra from 135 cashew apple obtained using the portable NIR spectrometer, and the same spectra after the follow treatment: b) MSC; c) SNV; and d) second derivative using Savitzky-Golay filter with a 74 second order polynomial for five points.....	73
Figure 2 -	3-D dendrogram (sample × wavelength × intensity) representing chemical composition similarity relationships among the samples.....	75

LISTA DE TABELAS

Tabela 1 -	Comparação de parâmetros gerais entre HPLC e UPLC.....	31
 Capítulo 1		
Table 1 -	Genotype of the 24 cashew fruits, respective accession numbers, and morphoagronomic characteristics: plant size; tree appearance; fruits color and shape; origin (county-state); and illustration.....	50
Table 2 -	Parameters obtained by multivariate calibration using PLS-DA model	62
 Capítulo 2		
Table 1 -	Concentrations of ascorbic acid (in mg.100 g⁻¹ of cashew apple), °Brix values, and acidity in 40 randomly selected samples (real or reference values), and those predicted values by each regression model	77
Table 2 -	Parameters obtained by multivariate calibration from each PLS model	78

LISTA DE ABREVIATURAS E SIGLAS

1D	Unidimensional
2D	Bidimensional
ANOVA	Analysis of Variance
AOAC	Association of Official Analytical Chemists
ASCII	American Standard Code for Information Interchange
BPI	Base Peak Intensity
CONAB	Companhia Nacional de Abastecimento
COW	Correlation Optimized Warping
DFI	2,6-dicloro-indofenol
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
ESI	Electrospray Ionization
FAO	Food and Agriculture Organization of the United Nations
FT-IR	Fourier-transform infrared spectroscopy
FW	Fresh Weight
HCA	Hierarchical Cluster Analysis
HPLC	High Performance Liquid Chromatography
IAL	Instituto Adolfo Lutz
ICS	International Chemometrics Society
IUPAC	International Union of Pure and Applied Chemistry
IV	Infravermelho
LC	Liquid Chromatography
LCC	Líquido da Casca da Castanha do Caju
LC-MS	Liquid Chromatography - Mass Spectrometry
LV	Latent variable
MS	Mass Spectrometry
MSC	Multiplicative Scatter Correction
NIPALS	Nonlinear Iterative Partial Least Squares
NIR	Near Infrared
PC	Principal Component
PCA	Principal Component Analysis
PDA	Photodiode Array Detector

PLS	Partial Least Squares
PLS-DA	Partial Least Squares – Discriminant Analysis
RMN	Ressonância Magnética Nuclear
SEC	Square Error of Calibration
SEV	Square Error of Validation
SNV	Standard Normal Variate
SVD	Singular Value Decomposition
TTA	Total Titratable Acidity
UV	Ultra Violeta

SUMÁRIO

1	INTRODUÇÃO	15
2	OBJETIVOS	17
2.1	OBJETIVOS ESPECÍFICOS.....	17
3	REVISÃO DE LITERATURA	18
3.1	O CAJUEIRO (<i>Anacardium occidentale</i> L.).....	18
3.2	IMPORTÂNCIA SOCIOECONÔMICA DO CAJU.....	19
3.3	COMPOSIÇÃO QUÍMICA E APLICAÇÕES DE PRODUTOS ORIUNDOS DO CAJU.....	22
3.4	BANCO DE GERMOPLASMA.....	28
3.5	INSTRUMENTAÇÃO.....	29
3.5.1	Cromatografia Líquida de Alta Eficiência	29
3.5.2	Espectrometria de Massas	31
3.5.3	Espectroscopia na Região do Infravermelho Próximo	32
3.6	QUIMIOMETRIA.....	33
3.6.1	Análise de Grupamento Hierárquico (HCA)	34
3.6.2	Análise de Componentes principais (PCA)	35
3.6.3	Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) ...	36
3.7	APLICAÇÕES QUIMIOMÉTRICAS NA ÁREA DA METABOLÔMICA.....	37
	REFERÊNCIAS BIBLIOGRÁFICAS	39
4	CAPÍTULO 1 - METABOLIC VARIATION ON JUICES FROM DIFFERENT GENOTYPES OF CASHEW APPLE BY NON-TARGETED UPLC-QToF-MS	46
4.1	ABSTRACT.....	47
4.2	INTRODUCTION.....	48
4.3	EXPERIMENTAL.....	49
4.3.1	Sampling	49
4.3.2	Sample preparation and UPLC-qTOF-MS	53
4.3.3	Chemometric analyses	54
4.3.4	Relative contribution	55
4.4	RESULTS AND DISCUSSION.....	55
4.5	CONCLUSIONS.....	63

4.6	ACKNOWLEDGEMENTS.....	63
4.7	REFERENCES.....	64
5	CAPÍTULO 2 - PORTABLE MICRO NEAR INFRARED SPECTROSCOPY COUPLED TO CHEMOMETRIC ANALYSIS IN ASSESSING THE COMPOSITION VARIABILITY OF DIFFERENT GENOTYPES OF CASHEW APPLE.....	67
5.1	ABSTRACT.....	68
5.2	INTRODUCTION.....	69
5.3	EXPERIMENTAL.....	70
5.3.1	Sampling.....	70
5.3.2	Portable NIR analysis.....	70
5.3.3	Quantification of ascorbic acid, °Brix, and determination of the acidity.....	70
5.3.4	Chemometric analysis.....	71
5.4	RESULTS AND DISCUSSION.....	72
5.4.1	Exploratory evaluation.....	72
5.4.2	Classification analysis.....	76
5.5	CONCLUSIONS.....	79
5.6	ACKNOWLEDGEMENTS.....	79
5.7	REFERENCES.....	80
6	CONSIDERAÇÕES FINAIS.....	82

1 INTRODUÇÃO

O cajueiro (*Anacardium occidentale* L.) é uma das principais frutíferas cultivadas no Brasil, devido ao alto valor nutritivo e comercial dos seus produtos. Só no ano de 2016, no Nordeste, as exportações dos produtos oriundos do caju, principalmente a castanha, movimentaram mais de 103 milhões de dólares (LIMA, 2017). Devido à importância socioeconômica do caju, em especial para região Nordeste, e a expansão do cultivo em diversos ecossistemas do Brasil, a Embrapa Agroindústria Tropical, localizada no Ceará, conserva um Banco de Germoplasma (BAG) de Caju, cujo principal objetivo, dentre outros, é a caracterização do acervo para potencializar o uso da base genética disponível, além de subsidiar o programa de melhoramento genético do cajueiro.

O caju é rico em compostos fenólicos (BRITO *et al.*, 2007) e diversos estudos têm sido realizados a fim de identificar os metabólitos bioativos dessa planta. Os metabólitos podem ser definidos como o produto do metabolismo de um organismo. O campo da ciência que realiza a análise abrangente e quantitativa dos metabólitos de um sistema biológico é chamada de metabolômica (FIEHN, 2002). Os estudos metabolômicos são realizados a partir de análises comparativas sobre os perfis metabólicos individuais por meio de métodos analíticos capazes de gerar informações sobre as propriedades de determinadas classes de compostos, criando um perfil metabólico único da amostra. Além disso, é aplicável a diversas matrizes biológicas, como extratos vegetais, microrganismos, fluidos, dentre outros. (FUNARI *et al.*, 2013).

A metabolômica aplicada à bioprospecção de extratos vegetais pode ser realizada em duas etapas. Na primeira etapa, os perfis metabólicos das amostras são avaliados e relacionados às suas respectivas atividades biológicas. Na segunda etapa, caso existam diferenças nos perfis, são realizadas análises de constituintes das amostras por meio de técnicas de separação acopladas a técnicas espectrométricas com o objetivo de conhecer e quantificar o maior número possível de metabólitos. Observando-se a presença de moléculas diferentes em amostras bioativas e inativas, pode-se inferir quais delas são responsáveis pela atividade biológica alvo (SIMÕES *et al.*, 2016).

Nos últimos anos, o desenvolvimento de hardware, software e novas metodologias, como a cromatografia líquida de ultra eficiência e o uso de técnicas

hifenizadas, proporcionaram grandes progressos na detecção de metabólitos. Em análises do tipo *untargeted* (análises globais), milhares de metabólitos podem ser detectados mesmo sem um conhecimento prévio do perfil químico da amostra (DUNN *et al.*, 2013).

As técnicas de separação que mais se destacam são a cromatografia gasosa de alta resolução (*High Resolution Gas Chromatography*, HRGC) e a cromatografia líquida (*Liquid Chromatography*, LC) de alta e de ultra-alta eficiência (HPLC e UHPLC, respectivamente) (FUNARI *et al.*, 2013). A espectrometria de massas (*Mass Spectrometry*, MS) e a ressonância magnética nuclear (RMN) são as técnicas de detecção mais utilizadas na metabolômica (VILLAS-BÔAS *et al.*, 2004).

Técnicas espectroscópicas, como a espectroscopia na região do infravermelho próximo (*Near Infrared*, NIR), tem bastantes aplicações no campo da metabolômica. Por ser uma técnica não-destrutiva, estudos sugerem que a espectroscopia de NIR seja utilizada para avaliação da qualidade de amostras ainda no campo (RIBEIRO *et al.*, 2016).

Essas técnicas permitem explorar e conhecer a variabilidade metabólica de germoplasmas conservados (ARANHA *et al.*, 2017) e quanto melhor for o conhecimento biológico de um banco de germoplasma, maior e mais eficiente será o direcionamento de estratégias de melhoramento e uso do acervo conservado. Diante disto, este trabalho teve como objetivo explorar a variabilidade de acessos de caju provenientes do Banco de Germoplasma da Embrapa por meio de um estudo metabolômico, utilizando ferramentas quimiométricas, técnicas espectrométricas e espectroscópicas.

2 OBJETIVOS

Avaliar a variabilidade metabólica de genótipos do banco de Germoplasma e do Programa de Melhoramento Genético do cajueiro.

2.1 OBJETIVOS ESPECÍFICOS

- Analisar genótipos de caju do Banco de Germoplasma e do Programa de Melhoramento Genético por meio de técnicas hífenadas (LC-MS);
- Investigar tendências de agrupamentos dos genótipos, por meio de ferramentas quimiométricas;
- Identificar as moléculas responsáveis pela separação dos grupos;
- Quantificar os teores de vitamina C, acidez total e sólidos solúveis de sucos de caju;
- Avaliar o uso de espectros de Micro-NIR na classificação de pseudofrutos com base na quantificação de vitamina C, acidez total e sólidos solúveis.

3 REVISÃO DE LITERATURA

3.1 O CAJUEIRO (*Anacardium occidentale* L.)

O cajueiro (*Anacardium occidentale* L.) é uma planta pertencente ao gênero *Anacardium*. Este gênero foi descrito em 1735 por Lineu, e, na revisão do gênero, apenas 10 espécies foram reconhecidas (MITCHELL; MORI, 1987). A maioria das espécies de *Anacardium* são pouco exploradas comercialmente e o aproveitamento delas se restringe ao consumo local. O *Anacardium occidentale* L., é a única espécie cultivada, explorada comercialmente e distribuída em todo o mundo tropical (BARROS; CRISÓSTOMO, 1995).

Embora a origem do cajueiro ainda não esteja bem esclarecida, é fato que o Brasil é o mais importante centro de diversidade do gênero *Anacardium*, destacando-se a região amazônica como principal centro, e a região do Planalto Central como centro secundário (MITCHELL; MORI, 1987). A espécie *A. occidentale* L. tem como centro de dispersão o Nordeste brasileiro (BARROS; CRISÓSTOMO, 1995).

Além de países da África e da Ásia, encontra-se espalhado por quase todo o território nacional, e, apesar de estar presente em praticamente todos os estados brasileiros, adapta-se melhor às condições ecológicas do litoral do Nordeste. Foi introduzido pelos portugueses na África e Índia no século XVI e, atualmente, pode ser encontrado espontaneamente ou cultivado em alguns países africanos (Gana, Angola, Nigéria, Madagascar, Quênia, Tanzânia, e Moçambique), na América Central (México), nos Estados Unidos (Florida) na América do Sul (Brasil e Peru), no Médio Oriente e Ásia (SABE, 2016).

No Brasil, a variabilidade de cajueiros é agrupada em dois tipos, cajueiro comum e anão precoce, que são diferenciados pelo porte. O cajueiro comum apresenta porte elevado com altura variando de oito a quinze metros e é o mais difundido. O anão precoce é considerado um ecótipo do cajueiro comum e tem como características o baixo porte (média de altura de 4m) e florescimento mais cedo em relação ao comum (BARROS, 1988). A Figura 1 ilustra os dois tipos de cajueiro.

Figura 1 - Cajueiro comum (esquerda) e anão precoce (direita).



Fonte: <http://poderdasfrutas.com/a-historia-do-caju/>. Fonte: Castro (2017).

Os produtos mais importantes oriundos do cajueiro são: a castanha, o líquido da casca da castanha (LCC) e o pedúnculo, também chamado de pseudofruto. O LCC é muito utilizado em indústrias químicas para fabricação de tintas, lubrificantes, resinas fenólicas e pós de fricção para a indústria automotiva, por exemplo. A castanha fornece também a amêndoa, que é o principal produto por ser muito apreciada no consumo humano, além da película que envolve a amêndoa, a qual é rica em taninos utilizados na indústria de vernizes. A maior parte da produção do LCC e da amêndoa é destinada à exportação. O pedúnculo pode ser consumido de diversas formas: in natura, em forma de polpas, cajuína, doces (USAID, 2006; SANTOS *et al.*, 2007; EMBRAPA, 2016).

Segundo Moura *et al.* (2001), os pedúnculos de coloração laranja a vermelha são os mais apreciados no mercado brasileiro, pois esta cor está associada a frutos mais maduros. Porém, a coloração trata-se apenas de uma característica do genótipo; as características como acidez, vitamina C e açúcar, por exemplo, são o que determinam a qualidade do caju.

3.2 IMPORTÂNCIA SOCIOECONÔMICA DO CAJU

O cajueiro é uma das frutíferas de destaque no contexto socioeconômico do Brasil devido ao alto valor nutritivo e comercial dos seus produtos. A planta possui considerável capacidade adaptativa a solos de baixa fertilidade, a temperaturas elevadas e ao estresse hídrico, tornando-se importante fonte de renda para os estados

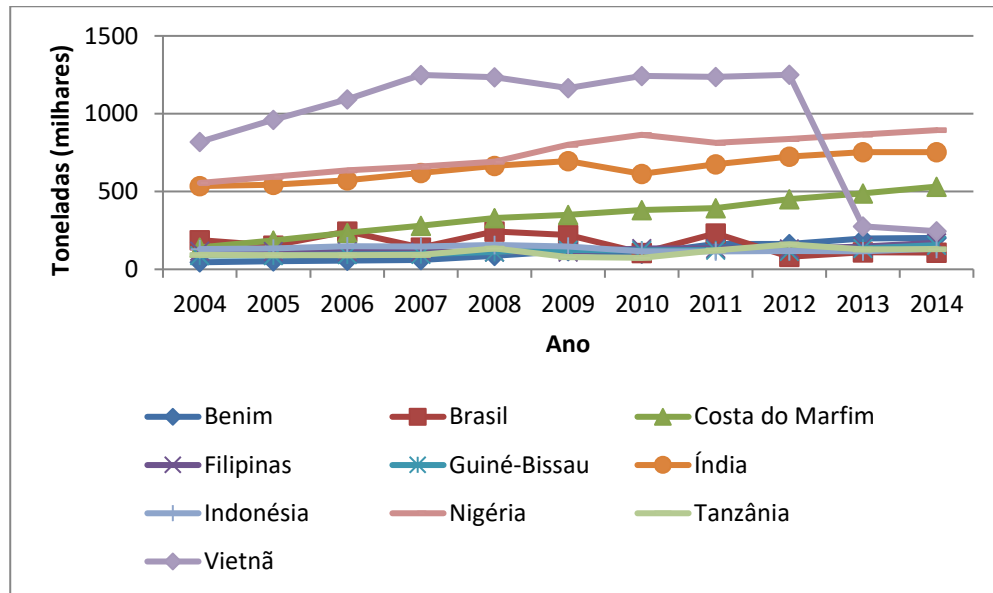
nordestinos. Além disso, produz em período seco, na entressafra das culturas anuais e viabiliza diversas agroindústrias através do beneficiamento da castanha, produção de sucos, doces, cajuína, além de gerar matérias-primas para indústrias químicas de tintas, vernizes, lubrificantes, aditivos, entre outros (EMBRAPA, 2016; SILVA *et al.*, 2014).

A cajucultura é uma indústria de multiprodutos que podem ser obtidos por meio do beneficiamento do caju, com uma grande perspectiva de crescimento no mercado brasileiro e mundial (ALMEIDA *et al.*, 2017). Tem grande importância socioeconômica no nordeste brasileiro, com destaque para o estado do Ceará que, somente no ano de 2016, gerou, pela exportação da amêndoa, cerca de 103 milhões de dólares mesmo com chuvas abaixo da média histórica (LIMA, 2017).

A produção brasileira de caju atingiu 104.475 toneladas em 2015, correspondendo a uma área colhida de 585.966 hectares, sendo a região Nordeste responsável por 99,4% da produção (ALMEIDA *et al.*, 2017). Valores mais expressivos poderão ser alcançados com o incremento da produção de amêndoas e do mercado dos subprodutos.

Com base nas estatísticas mais recentes da *Food and Agriculture Organization of the United Nations* (FAO), há produção de castanha de caju em 34 países ao redor do mundo. O Brasil está entre os dez maiores produtores mundiais, porém, até 2009, ocupava a quinta posição (Figura 2). Em síntese, houve uma expressiva redução nas participações da Índia e do Brasil na produção mundial. Essas perdas de participação podem ser explicadas pela diminuição do plantio de novas áreas na Índia e no Brasil e pelo acelerado crescimento da produção no Vietnã e em alguns países africanos. Além disso, o Brasil, dentre os maiores produtores de castanha, é o que apresenta mais baixa produtividade devido a fatores como declínio natural da produção dos pomares, longas secas e pouca utilização de tecnologia na exploração dos cajueiros (PESSOA e LEITE, 2013; BNB, 2016).

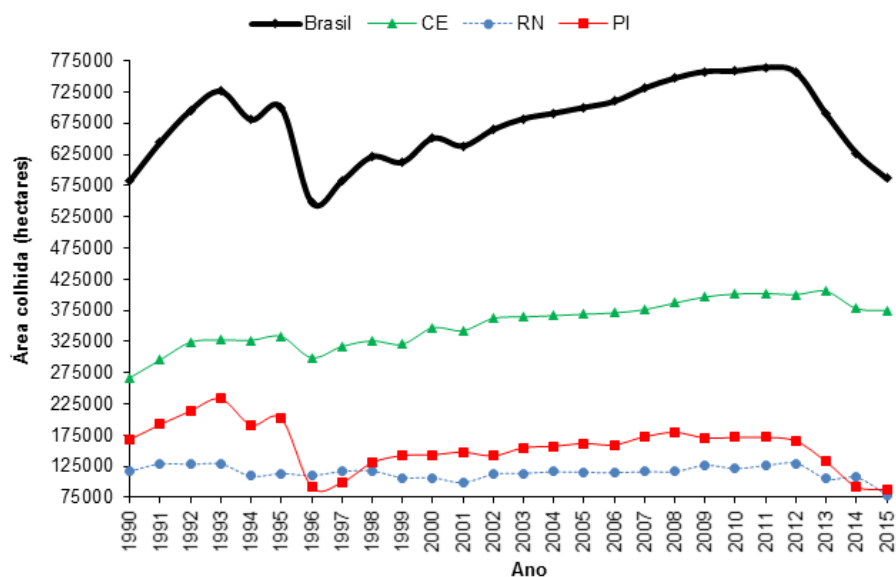
Figura 2 - Evolução da produção mundial de castanha de caju, período de 2004 a 2014



Fonte: elaborada pela autora. Dados: FAO (2014).

Entre os anos de 1990 e 2012, a área destinada à cajucultura no Brasil apresentou um acréscimo de 30%, passando de 582 mil hectares para 756 mil hectares (Figura 3). A redução da área colhida observada o período entre 2013 e 2015 pode ser atribuído pelo déficit hídrico e pela alta mortalidade de cajueiros (EMBRAPA, 2016).

Figura 3 - Evolução da área com cajueiros colhida no Brasil, período de 1990 a 2015



Fonte: EMBRAPA (2016).

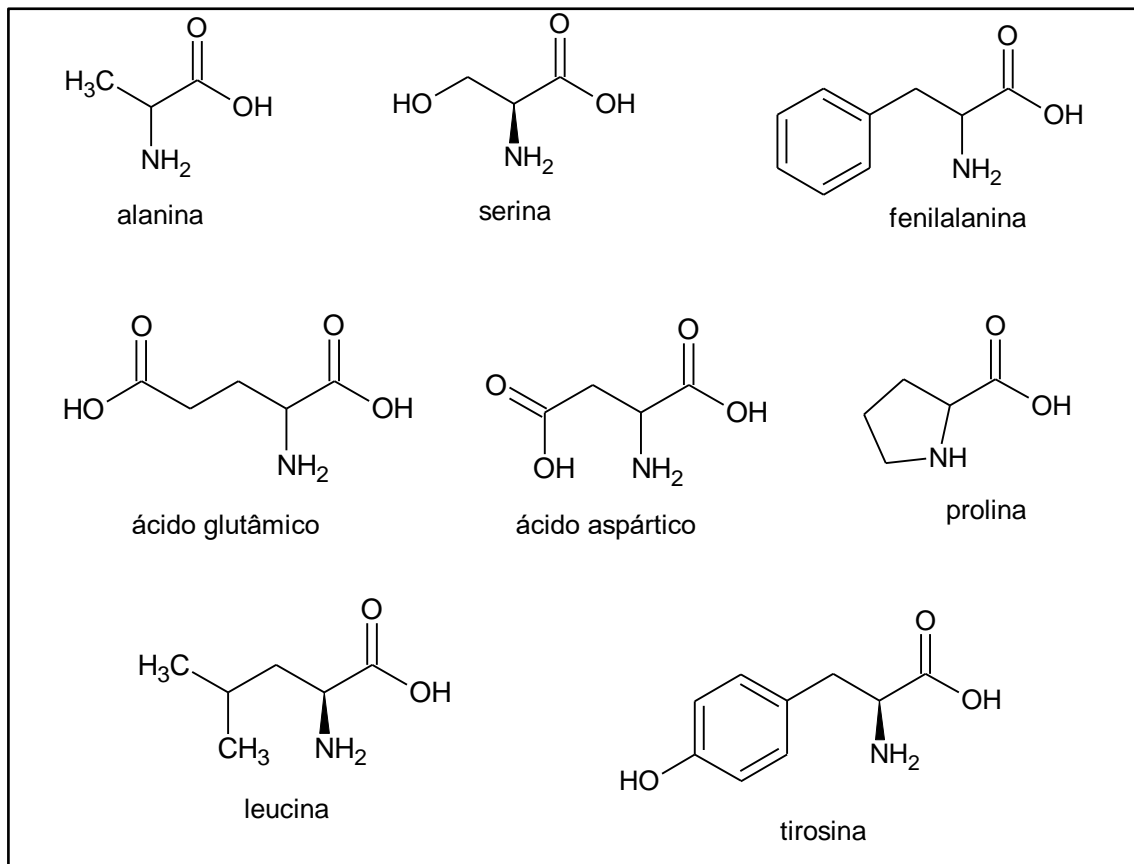
Apesar da redução da área colhida, observou-se um acréscimo na diferença entre área colhida e área plantada, indicando um possível aperfeiçoamento dos pomares na região. Como exemplo, temos o estado do Ceará, que desde 2008 vem apresentando um aumento da área cultivada com cajueiro anão precoce e redução de cajueiro comum, apontando uma provável substituição dos pomares antigos por pomares novos de cajueiros clonais (EMBRAPA, 2016).

3.3 COMPOSIÇÃO QUÍMICA E APLICAÇÕES DE PRODUTOS ORIUNDOS DO CAJU

O caju apresenta quantidades significativas de vitaminas, taninos, sais minerais, ácidos orgânicos, carboidratos (LAVINAS *et al.*, 2006) e açúcares redutores (DAMASCENO *et al.*, 2008). Possui quantidade de vitamina C cerca de quatro vezes superior ao suco de laranja (AKINWALE, 2000). Lowor e Agylent-Badu (2009) relatam a presença de minerais como potássio, cálcio, zinco, ferro, fósforo e magnésio. Os taninos estão concentrados na camada cerosa da “pele” do caju, a qual contém aproximadamente oito vezes mais taninos que o mesocarpo (MICHODJEHOUN-MESTRES *et al.*, 2009).

No geral, cajus de diferentes regiões apresentam variação em sua composição, que pode ser atribuída a fatores como o tipo de solo e as condições climáticas da região, dentre outros fatores (DABONNE *et al.*, 2016). Estudos mostram que o suco de caju de cor amarela contém uma maior quantidade de ácido ascórbico, enquanto que o suco de caju de cor vermelha apresenta mais taninos e aminoácidos (DAS; ARORA, 2017). Oliveira *et al.* (2002) relatam a presença de oito aminoácidos livres majoritários no suco de caju: alanina, serina, leucina, fenilalanina, prolina, ácido glutâmico, tirosina e ácido aspártico (Figura 4). Os dois primeiros podem ser considerados aminoácidos-chave para a identificação do suco de caju, pois permaneceram majoritários, apesar da concentração dos aminoácidos terem sofrido grandes variações durante o período avaliado.

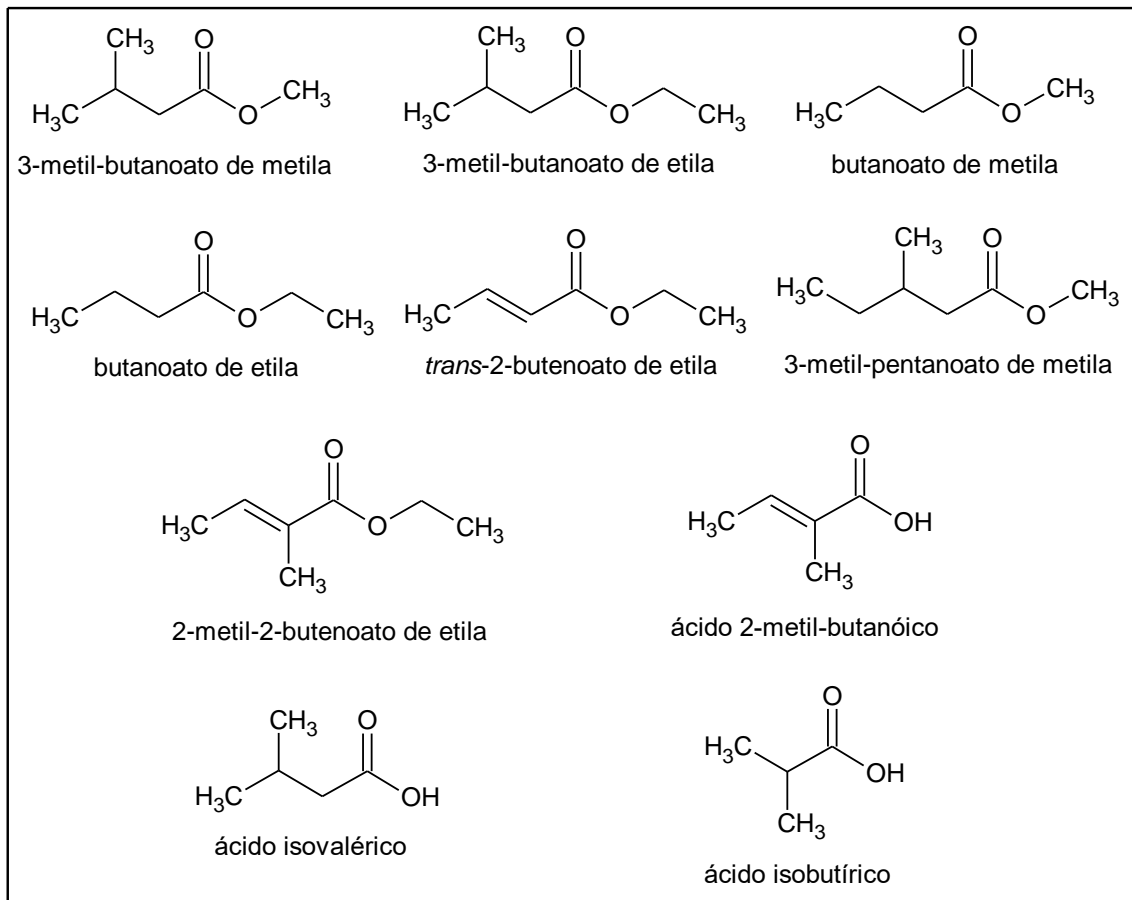
Figura 4 - Estruturas de aminoácidos identificados no caju.



Fonte: elaborada pela autora.

Garruti *et al.* (2003) estudaram os compostos voláteis responsáveis pelo aroma do caju e detectaram 58 compostos utilizando cromatografia gasosa de alta resolução. Os compostos identificados eram predominantemente ésteres (cerca de 42% dos compostos totais), seguidos de aldeídos (14% dos compostos totais). Os aromas dos compostos 3-metil-butanoato de metila, 3-metil-butanoato de etila, butanoato de metila, butanoato de etila, *trans*-2-butenato de etila, 2-metil-butanoato de etila, 3-metil-pentanoato de metila e 2-metil-2-butenato de etila foram descritos como caju, doce, floral e frutado; o ácido 2-metil-butanóico mostrou um desagradável odor forte e prolongado, e os ácidos isovalérico e isobutírico contribuíram para o cheiro pungente/ácido do suco de caju (Figura 5).

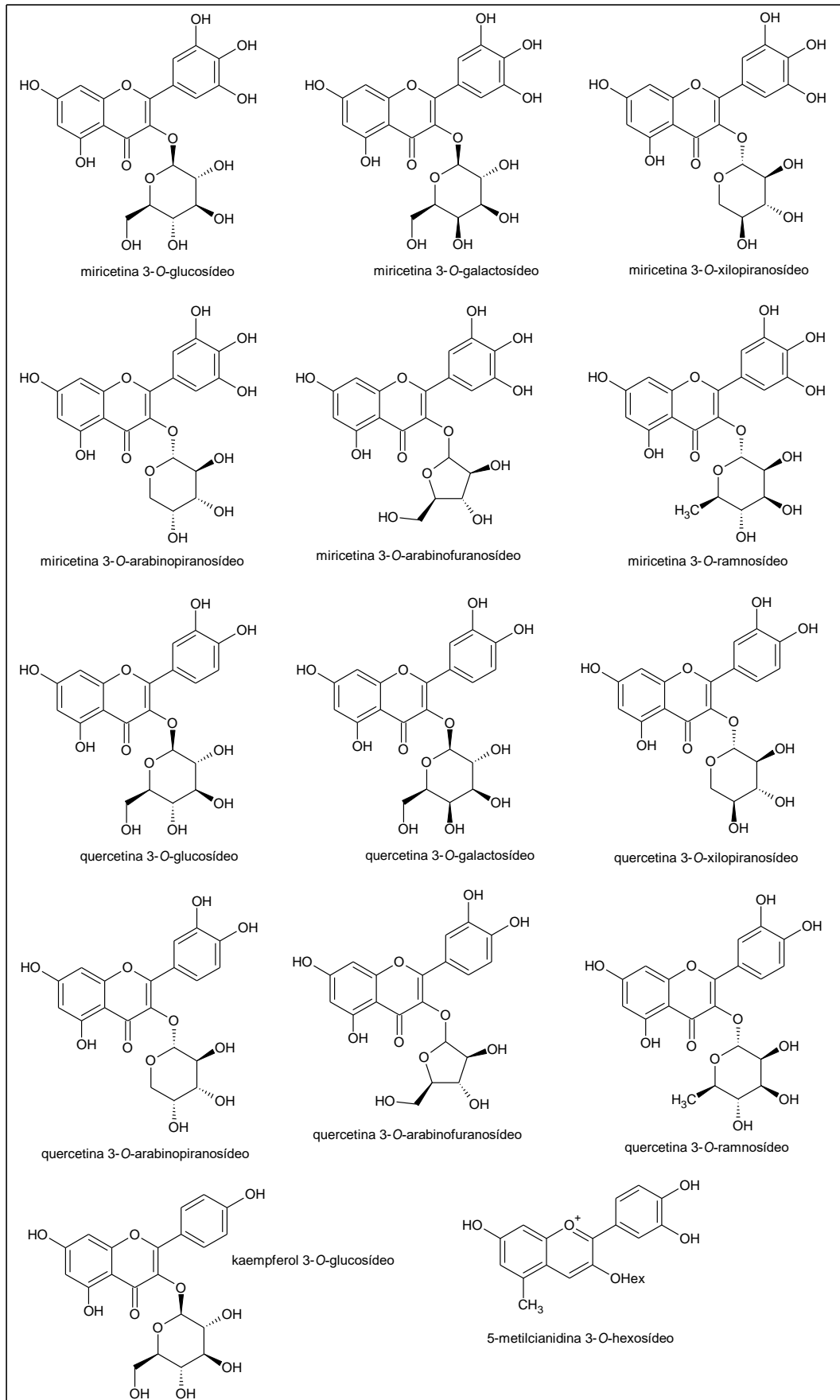
Figura 5 - Estruturas de compostos voláteis identificados no caju.



Fonte: elaborada pela autora.

Brito *et al.* (2007) identificaram treze flavonóides glicosilados e uma antocianidina (Figura 6) no extrato metanol-água do pseudofruto do caju. Os compostos identificados foram: 3-*O*-galactosídeo, 3-*O*-glucosídeo, 3-*O*-ramnosídeo, 3-*O*-xilopiranosídeo, 3-*O*-arabinopiranosídeo e 3-*O*-arabinofuranosídeo da quercetina e da miricetina, kaempferol 3-*O*-glucosídeo e 3-*O*-hexosídeo da metil-cianidina. Esses compostos pertencem a classe dos flavonóides, a qual é conhecida por suas propriedades antioxidante, antiviral e antibacteriana, dentre outras (KOZŁOWSKA; SZOSTAK-WEGIEREK, 2014).

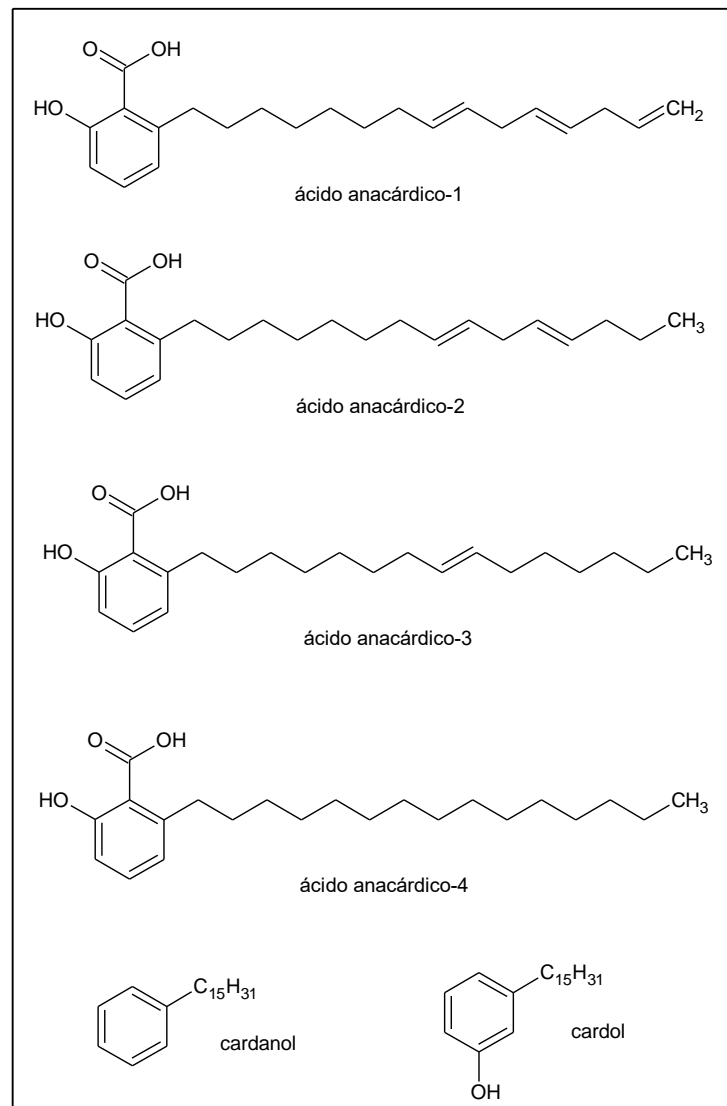
Figura 6 - Estruturas de flavonóides identificados no caju



Fonte: elaborada pela autora.

O líquido da casca da castanha de caju (LCC) extraído por aquecimento apresenta alquil fenóis. Estes consistem de ácidos anacárdicos, cardanóis e cardóis. Trevisan *et al.* (2006) identificaram ácidos anacárdicos, cardanóis e cardóis (Figura 7) no LCC, os quais, na mistura bruta extraída, demonstraram ter uma potente capacidade antioxidante e de inibição da xantina oxidase. Dentre os compostos isolados, o ácido anacárdico-1, que apresenta três ligações duplas na cadeia lateral, confere maior potencial antioxidante e de inibição da enzima do que os outros ácidos que contêm 1 ou 2 ligações duplas, pois exerce uma supressão da geração de superóxido e da inibição da xantina oxidase, ao invés da eliminação de espécies reativas de oxigênio (MASUOKA; KUBO, 2004; TREVISAN *et al.*, 2006).

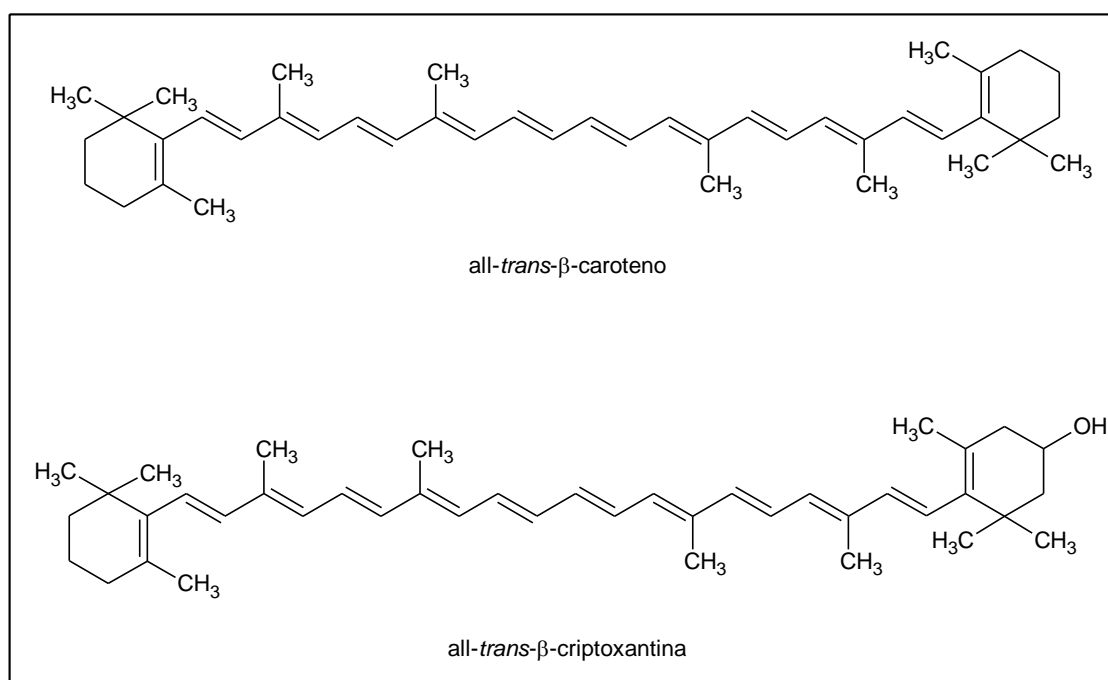
Figura 7 - Estruturas de alquil fenóis identificados no caju



Fonte: elaborada pela autora.

Cerca de 20 a 25% do pseudofruto processado na indústria se transforma em fibra residual que é descartada ou utilizada como suplemento na alimentação animal (ABREU et al., 2013). Muitos estudos têm sido direcionados para o aproveitamento dos subprodutos da indústria da castanha e do processamento do pedúnculo do caju. Abreu et al. (2013) sugere o reaproveitamento desse subproduto como fonte de carotenóides, os quais podem ser utilizados como corantes alimentícios. Barreto et al. (2007) estudaram bagaço do caju submetido a vários ciclos de maceração, prensagem e filtragem, chamado de retentado, o qual apresentou teor de compostos fenólicos totais, carotenóides totais e a atividade anti-radical livre encontrada expressivamente maior que nos extratos bruto e concentrado. Assim, o retentado mostrou-se ser uma rica fonte de compostos fenólicos e carotenoides, dentre os quais se destacam all-trans- β -criptoxantina e all-trans- β -caroteno, ilustrados na Figura 8.

Figura 8 - Estruturas de carotenóides identificados no caju



Fonte: elaborada pela autora.

Pinho *et al.* (2011) avaliaram a inclusão de bagaço de pedúnculo de caju em pó em hambúrgueres de carne bovina. A adição de até 10,7% de resíduo de pedúnculo de caju não causou grandes alterações nas características sensoriais dos hambúrgueres e melhorou a qualidade nutricional do produto. O bagaço, o extrato bruto concentrado e suas frações de ácidos fenólicos livres apresentaram significante

capacidade antioxidante, que pode ser atribuída ao seu conteúdo de compostos fenólicos. Esses resultados abrem perspectiva para um melhor reaproveitamento dos subprodutos (BROINIZI *et al.*, 2007).

3.4 BANCO DE GERMOPLASMA

Figura 9 – Banco de Germoplasma de Caju da Embrapa



Fonte: Castro (2017).

Banco de Germoplasma (Figura 9) é conceituado como sendo o repositório onde se armazena a variabilidade genética de uma ou de várias espécies. As principais atividades do banco de germoplasma são a conservação, coleta, caracterização e documentação dos acessos da espécie-alvo (FERREIRA, 2011).

A importância da agroindústria do caju para a economia da região Nordeste e a expansão do cultivo em diversos ecossistemas do Brasil ampliam a necessidade de se conhecer, avaliar e usar a base genética disponível e conservada no Banco Ativo de Germoplasma de Caju. Atualmente, são conservados 657 acessos de cajueiro, dos quais a maioria é da espécie *Anacardium occidentale* L. A coleção é mantida em campo, na Estação Experimental de Pacajus-CE, com coordenadas geográficas de 4° 11' 07" S, 38° 30' 07" O e altitude de 70 metros. A região apresenta tipo climático tropical quente, semiárido brando e tropical quente sub-úmido, com pluviosidade média de 791,4 mm ao ano e temperaturas médias de 26°C a 28°C (CASTRO *et al.*, 2013).

O conhecimento biológico do acervo possibilita o direcionamento de estratégias de melhoramento e uso do germoplasma conservado. A caracterização

morfológica é realizada a partir de lista de descritores estabelecidos para a cultura do cajueiro, que incluem características vegetativas, caracteres da planta, indicadores agroindustriais do fruto (CASTRO *et al.*, 2013) e, além disso, sempre são incluídos descritores para novas demandas potenciais.

Existe tendência dos bancos de germoplasma utilizarem ao máximo seus recursos, procurando dentro das coleções genótipos com outras vocações de uso. Alguns bancos ativos de germoplasma já vêm desenvolvendo trabalhos de identificação de genótipos para alimentar a dinâmica do setor químico, como a avaliação da quantificação e qualificação do LCC.

3.5 INSTRUMENTAÇÃO

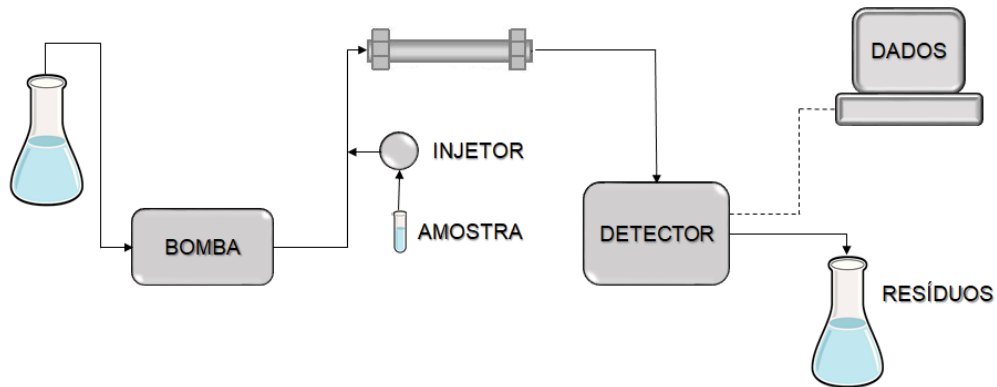
3.5.1 Cromatografia Líquida de Alta Eficiência

As técnicas cromatográficas modernas são altamente eficientes na análise de produtos naturais. Vários métodos podem ser empregados e, dentre eles, a cromatografia líquida de alta eficiência (*High-Performance Liquid Chromatography*, HPLC) merece destaque. São várias as aplicações da cromatografia líquida de alta eficiência, especialmente em análises ambientais, farmacêuticas, forenses, de alimentos, dentre outros (ZHANG; LI; YAN, 2008).

Segundo a *International Union of Pure and Applied Chemistry* (IUPAC), “cromatografia é o método físico de separação no qual os componentes a serem separados se distribuem entre duas fases, uma das quais estacionária, enquanto a outra se movimenta numa direção definida”. Assim, a separação dos compostos em análise de HPLC, ocorre por meio da injeção da amostra na coluna cromatográfica (fase estacionária), a qual está recheada com partículas porosas que vão interagir química ou fisicamente com as moléculas da amostra. A fase móvel (solvente) também interage com as moléculas, carregando primeiro, com a ajuda de uma bomba de alta pressão, aquelas que tem mais afinidade com o solvente. Desta forma, ocorre a separação dos compostos. Após a separação, os compostos passam para o detector, o qual gera um sinal de intensidade proporcional à quantidade do composto analisado. Por fim, um software computacional gera um cromatograma com os dados

obtidos. A Figura 10 mostra um esquema de funcionamento de um HPLC (KUPIEC, 2004).

Figura 10 - Esquema de cromatografia líquida de alta eficiência



Fonte: elaborada pela autora.

A busca por análises mais rápidas, eficientes e de custo menor resultou no surgimento da cromatografia líquida de ultra eficiência (*Ultra-Performance Liquid Chromatography*, UPLC). A UPLC baseia-se nos mesmos princípios de separação da HPLC, porém aquela utiliza colunas cromatográficas de dimensões reduzidas com partículas da fase estacionária de tamanho inferior a $\leq 2 \mu\text{m}$ e altas velocidades lineares de fase móvel, resultando em uma maior pressão de trabalho. Um sistema de UPLC está ilustrado na Figura 11.

Figura 11 – Sistema de UPLC



Fonte: waters.com

As principais vantagens da UPLC são: menores tempos de análise, menores volumes de amostra e de fase móvel, melhor resolução e detectabilidade e

diminuição na geração de resíduos (MALDANER; JARDIM, 2009; MALDANER; JARDIM, 2012). A tabela 1 abaixo mostra as principais diferenças entre HPLC e UPLC.

Tabela 1 – Comparação de parâmetros gerais entre HPLC e UPLC

Características	HPLC	UPLC
Tamanho da partícula	3-5 μm	Menor que 2 μm
Fluxo da fase móvel	1 mL/min	0,25 mL/min
Volume de injeção	10-20 μL	1-10 μL
Dimensões da coluna	(100,0 x 2,1 mm, 5 μm)	(100,0 x 2,1 mm, 1,7 μm)
Pressão	1-2000 psi	1-15000 psi
Limite de detecção	0,01 $\mu\text{g/mL}$	0,01 ng/mL

Fonte: Khan e Ali (2015) adaptado.

A UPLC vem sendo aplicada em detecção de impurezas, identificação de metabólitos, estudos farmacocinéticos, análises toxicológicas, estudos metabolômicos, análise de produtos naturais, dentre outras utilizações (KHAN; ALI, 2015).

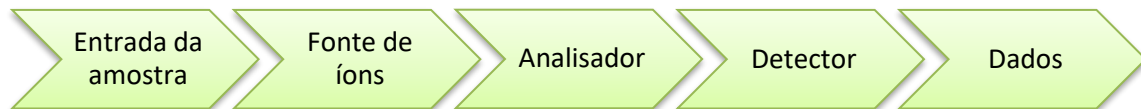
3.5.2 Espectrometria de Massas

A espectrometria de massas é uma técnica analítica utilizada para converter moléculas de uma amostra em íons, podendo gerar informações quantitativas e qualitativas da amostra analisada. Ao serem inseridas no espectrômetro de massas (*Mass Spectrometer*, MS), as moléculas serão ionizadas, adquirindo cargas positivas ou negativas. Os íons passam pelo analisador de massas e, em seguida, dirigem-se ao detector, onde são gerados sinais de acordo com a relação massa/carga (m/z), criando um espectro de massas de acordo com a abundância relativa dos sinais (HO *et al.*, 2003). A grande vantagem desta técnica é sua alta sensibilidade na detecção de compostos (VILLAS-BOAS, 2005).

A ionização por *electrospray* (*Electrospray Ionization*, ESI) trabalha em pressão atmosférica cuja formação de íons em fase gasosa ocorre sem uma excessiva fragmentação. É a técnica mais usada para análise de moléculas polares (WU *et al.*, 2013). O detector de arranjos lineares de fotodiodos (*diode array detector*,

DAD) pode ser combinado com LC-MS para fornecer informações de cada pico do cromatograma utilizando uma pequena quantidade de amostra e realizando a análise em um curto espaço de tempo (WU *et al.*, 2013). Esse detector faz uma varredura em todos os comprimentos de onda da região do ultravioleta (UV). A Figura 11 abaixo ilustra o esquema de funcionamento de um EM.

Figura 12 - Esquema da espectrometria de massas



Fonte: elaborada pela autora.

A hifenização LC-MS é capaz de fornecer informações estruturais detalhadas através do padrão de fragmentação das moléculas. Deste modo, compostos podem ser rapidamente identificados, evitando o isolamento repetitivo de compostos conhecidos ou indesejados (WU *et al.*, 2013).

3.5.3 Espectroscopia na Região do Infravermelho Próximo

Infravermelho (IV) próximo é o nome dado à região do espectro eletromagnético imediatamente superior à região visível em termos de comprimento de onda, ou seja, trata-se da região do infravermelho “mais próxima” da região visível. Essa região se estende de 750 a 2500 nm. A espectroscopia no IV é um tipo de espectroscopia de absorção, que usa a região do IV-próximo do espectro eletromagnético da luz visível (LIMA; BAKKER, 2011).

As principais características da espectroscopia na região do IV-próximo são: rapidez, análise não-destrutiva e não-invasiva, além de necessitar de preparação mínima da amostra (PASQUINI, 2003). NIR pode ser usado para analisar materiais como alimentos, produtos farmacêuticos, produtos petroquímicos, dentre outros. (TSUCHIKAWA; KOBORI, 2015).

Ribeiro *et al.* (2016) realizaram um experimento utilizando um NIR portátil (Figura 12) a fim de elaborar um modelo de predição de características de qualidade do pedúnculo do caju. O modelo se mostrou eficiente, pois, além de possuir baixo

custo, é capaz de avaliar a qualidade do fruto antes mesmo da colheita, o que pode ajudar na seleção de cultivares de caju para fins comerciais.

Figura 13 – NIR portátil.



Fonte: astro34.com.br

3.6 QUIMIOMETRIA

A Sociedade Internacional de Quimiometria (*International Chemometrics Society* - ICS) define quimiometria como a ciência que relaciona medições obtidas em um sistema ou processo químico, obtendo informações sobre o estado do sistema, por meio do emprego de métodos matemáticos ou estatísticos (BANSAL *et al.*, 2014). O melhoramento na obtenção dos dados ocasionado pelo aparecimento de novas técnicas analíticas induziu o surgimento desta ciência entre as décadas de 70 e 80 (KUMAR *et al.*, 2014). Ela investiga, interpreta, classifica e prevê conjuntos de dados de interesse químico (SOUZA; POPPI, 2012).

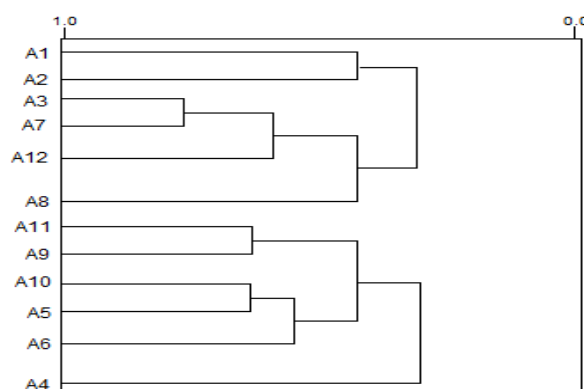
A quimiometria se divide em três áreas principais: planejamento e otimização de experimentos, reconhecimento de padrões e, por fim, a calibração multivariada (POPPI; PASQUINI, 1993). A área de planejamento de experimentos, a mais conhecida dentre as áreas, trabalha com a avaliação dos efeitos de variáveis sobre um determinado processo e do quanto essas variáveis são significativas. No reconhecimento de padrões, avalia-se uma ampla quantidade de dados (cromatogramas, por exemplo) com o objetivo de encontrar similaridades entre eles, formando, assim, agrupamentos e determinando tendências. Por fim, a calibração multivariada procura sugerir um modelo que relacione uma série de medidas com uma determinada propriedade de amostras (SOUZA; POPPI, 2012).

Métodos quimiométricos aliados a técnicas espectrométricas demonstram ser ferramentas eficazes para avaliar padrões de qualidade de produtos (WEI *et al.*, 2017; DU *et al.*, 2018), construir perfis metabólicos de plantas (GAD e BOUZABATA, 2017), aprimorar métodos de identificação de espécies patogênicas de *Candida* (WOHLMEISTER *et al.*, 2017), analisar perfil de amostras de interesse forense (MATERAZZI *et al.*, 2017), verificar a composição química de compostos orgânicos voláteis microbianos em ambientes externos (GARCIA-ALCEGA *et al.*, 2017), detecção de medicamentos falsificados (CUSTERS *et al.*, 2016) dentre outras inúmeras finalidades.

3.6.1 Análise de Grupamento Hierárquico (HCA)

A Análise de Agrupamento Hierárquico (do inglês *Hierarchical Cluster Analysis*, HCA) é uma técnica multivariada que ordena dados baseada em suas características e agrupa os dados obtidos pela afinidade entre eles no espaço, enfatizando os seus agrupamentos e padrões naturais (VANDEGINSTE *et al.*, 1998). O método consiste em calcular a distância entre os pontos (variáveis) no espaço, constituindo-se uma matriz de forma que as amostras mais semelhantes sejam agrupadas entre si. Baseado nesta matriz de proximidade entre as amostras, se constrói um diagrama de similaridade denominado dendrograma, ilustrado na figura 13 (MOITA NETO, 2004; KUMAR *et al.*, 2014). Cada nível do dendrograma representa uma partição do conjunto de dados em um número específico de clusters (grupamentos). Com base no dendrograma, é possível definir o número de clusters (DOWNS; BARNARD, 2002).

Figura 14 – Ilustração de um dendrograma



Fonte: elaborada pela autora.

Na metabolômica, por exemplo, HCA e PCA foram utilizados com sucesso por Madala *et al.* (2014) para avaliar a distribuição de metabólitos pelas células das plantas *Nicotiana tabacum* e *Sorghum bicolor* quando expostas a um indutor de ativação da defesa das plantas. Da mesma forma, Jing *et al.* (2015) discriminaram o perfil de COVs de 20 diferentes espécies e variedades de *Citrus* por meio de HCA, permitindo o emprego de um novo procedimento de classificação taxonômica de espécies de plantas.

3.6.2 Análise de Componentes principais (PCA)

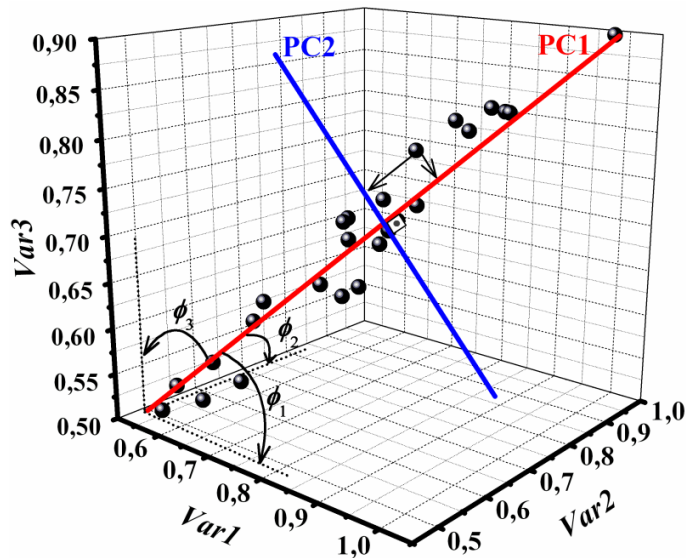
A análise de componentes principais (*Principal Component Analysis*, PCA) é um método multivariado exploratório, que formula hipóteses a partir dos dados obtidos, e é capaz de segregar a informação relevante da redundante e aleatória (PANERO *et al.*, 2009). A PCA é provavelmente o método mais difundido na quimiometria. Com ele, obtém-se arranjos de variáveis que descrevem as principais tendências em um conjunto de dados (WISE *et al.*, 1999).

Em uma análise de componentes principais, os dados são projetados nos eixos ortogonais chamados de PCs (Figura 13). Os *scores* representam os valores das PCs em relação às amostras, enquanto os *loadings* representam a influência de cada variável original nas PCs. O estudo do conjunto de *scores* e *loadings* permite ainda estimar a influência de cada variável em cada amostra (PANERO *et al.*, 2006). A informação contida numa PC é exclusivamente sua e, portanto, não está presente nas outras PCs. A primeira componente principal, PC1, apresenta a máxima variância dos dados. A PC seguinte (PC2) mostra a segunda maior variância dos dados em um eixo ortogonal à PC1. Da mesma maneira, as PCs seguintes apresentam a máxima variância restante e são ortogonais às componentes anteriores (RUSCHEL, 2017).

Segundo Teófilo (2013),

Em termos matemáticos, o cosseno do ângulo entre o eixo da variável com o eixo da componente representa a contribuição de cada eixo original para uma componente principal. Se uma variável individual e um componente principal apontam para a mesma mesma direção, o ângulo entre elas é zero e o cosseno é 1. Assim, a componente apresenta toda a variação no eixo daquela variável (TEÓFILO, 2013).

Figura 15 – Representação de duas PC's no espaço



Fonte: Teófilo (2013).

HCA e PCA são técnicas multivariadas com embasamentos diferentes. São técnicas independentes, mas que podem ser complementares na análise do conjunto de dados, gerando uma percepção mais geral das amostras (CAZAR, 2003).

3.6.3 Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA)

A Análise Discriminante por Mínimos Quadrados Parciais (*Partial Least Squares – Discriminant Analysis*, PLS-DA) é um método multivariado que relaciona a matriz de *scores* e *loadings* com uma matriz de referência, capaz de gerar as variáveis latentes necessárias para a análise classificatória dos dados (SIMÕES, 2008). Essas variáveis latentes são análogas aos PCs da PCA, porém sem a característica da ortogonalidade, e são responsáveis por explicar a máxima variância entre as matrizes analisadas (CIOSEK *et al.*, 2005). Enquanto a PCA é uma análise do tipo exploratória guiada pela variabilidade total das amostras, o PLS-DA é um método supervisionado guiado pela variabilidade encontrada nos grupos segregados (CIOSEK *et al.*, 2005; BARKER; RAYENS, 2003). São diversas as aplicações do método PLS como, por exemplo, na indústria de alimentos, de produtos farmacêuticos, petroquímica, em análises clínicas, de produtos naturais, dentre outras (TEÓFILO, 2013).

3.7 APLICAÇÕES QUIMIOMÉTRICAS NA ÁREA DA METABOLÔMICA

As abordagens da metabolômica podem ser classificadas em targeted (em português, metabolômica alvo) e untargeted (em português, metabolômica global). A primeira, refere-se à análise quantitativa de determinados metabólitos previamente escolhidos; os resultados obtidos são quantitativos e, na maioria dos casos, o estudo é orientado por hipóteses. A segunda, baseia-se na análise qualitativa da maior quantidade possível de metabólitos contidos no material em estudo; é uma abordagem geradora de hipóteses e tem como meta comparar as “impressões digitais” dos constituintes das amostras analisadas. (CANUTO *et al.*, 2017; DETTMER; ARONOV; HAMMOCK , 2007).

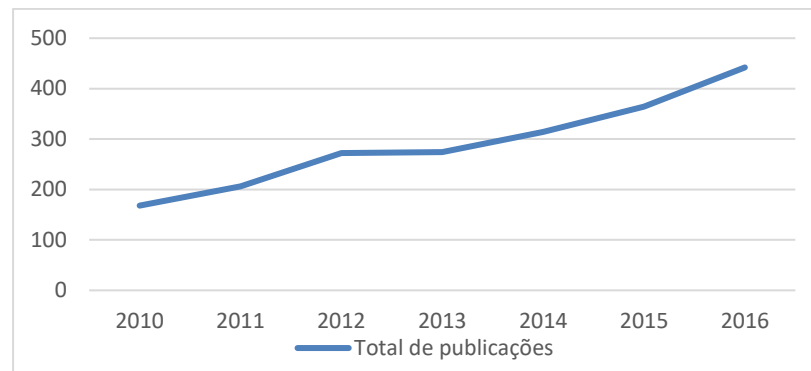
Diversos estudos envolvendo as abordagens da metabolômica vem sendo desenvolvidos. Uarrota *et al.* (2014) utilizaram métodos quimiométricos como PCA e PLS-DA para rastrear a composição química de raízes de mandioca durante a deterioração fisiológica pós-colheita. Grupamentos foram separados com êxito e os pesquisadores concluíram que a deterioração fisiológica pós-colheita “parece estar relacionada negativamente com compostos fenólicos, carotenóides e positivamente com flavonóides e antocianinas” (UARROTA *et al.*, 2014).

Khalil, Fekry e Farag (2017) avaliaram o perfil de voláteis de 13 variedades de tâmaras. Foram identificados 89 compostos voláteis e os mesmos foram agrupados pela sua origem biossintética. Por meio de análise multivariada, os dados foram classificados. Após agrupamento, os compostos 2,3-butanodiol, hexanal, hexanol e cinamaldeído foram apontados como os responsáveis pela separação dos grupos.

Zaitsu *et al.* (2014) utilizou análise do tipo untargeted para avaliar a dependência de drogas em amostras de plasma e de urina de ratos. O estudo demonstrou que a metabolômica é capaz de descrever os diferentes estados biológicos na dependência de drogas e é capaz de prever o grau de dependência.

A base de dados Scopus aponta, só em 2017, 422 trabalhos cujas na área de metabolômica associada a análise multivariada. A Figura 16 mostra um gráfico com a produção bibliográfica na área desde 2010.

Figura 16 - Total de publicações na área de metabolômica associada a análise multivariada



Fonte: elaborada pela autora. Dados: Scopus (2017).

REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, F. P.; DORNIER, M.; DIONISIO, A. P.; CARAIL, M.; CARIS-VEYRAT, C.; DHUIQUE-MAYER, C. Cashew apple (*Anacardium occidentale* L.) extract from by-product of juice processing: A focus on carotenoids. **Food chemistry**, v. 138, n. 1, 2013.

AKINWALE, T. O. Cashew apple juice: its use in fortifying the nutritional quality of some tropical fruits. **European Food Research and Technology**, v. 211, n. 3, p. 205-207, 2000.

ALMEIDA, L. H. F.; CORDEIRO, S. A.; PEREIRA, R. S.; COUTO, L. C.; DE SOUZA LACERDA, K. W. Viabilidade econômica da produção de caju (*Anacardium occidentale* L.). **Nativa**, v. 5, n. 1, 2017.

ARANHA, B. C.; HOFFMANN, J. F.; BARBIERI, R. L.; ROMBALDI, C. V.; CHAVES, F. C. Untargeted Metabolomic Analysis of *Capsicum* spp. by GC-MS. **Phytochemical Analysis**, v. 28, n. 5, p. 439-447, 2017.

BANSAL, A.; CHHABRA, V.; RAWAL, R. K.; SHARMA, S. Chemometrics: A new scenario in herbal drug standardization. **Journal of Pharmaceutical Analysis**, v. 4, n. 4, 2014.

BARRETO, G. P. M.; SOUZA, A. C. R.; AZEREDO, H. M. C.; MERCADANTE, A. Z. Compostos bioativos em sub-produtos da castanha de caju. **Alimentos e Nutrição Araraquara**, v. 18, n. 2, 2007.

BARKER, M.; RAYENS, W. Partial least squares for discrimination. **Journal of chemometrics**, v. 17, n. 3, 2003.

BARROS, L. M. Melhoramento. In: LIMA, V. M. S. P. e RAMOS, A. D. (Ed.). **A cultura do cajueiro no Nordeste do Brasil**. Fortaleza: ETENE, 1988.

BARROS, L. D. M.; CRISÓSTOMO, J. R.; PAIVA, W. D.; PAIVA, J. D. MELHORAMENTO genético do cajueiro. **Cajucultura: modernas técnicas de produção**. Fortaleza: Embrapa-CNPAT, 1995.

BRITO, E. S.; ARAÚJO, M. C. P.; LIN, L. Z.; HARNLY, J. Determination of the flavonoid components of cashew apple (*Anacardium occidentale*) by LC-DAD-ESI/MS. **Food chemistry**, v. 105, n. 3, p. 1112-1118, 2007. BROINIZI, P. R. B.; ANDRADE-WARTHA, E. R. S.; OLIVEIRA E SILVA, A. M. d.; NOVOA, A. J. V.; TORRES, R. P.; AZEREDO, H. M. C.; ALVES, R. E.; MANCINI-FILHO, J. Avaliação da atividade antioxidante dos compostos fenólicos naturalmente presentes em subprodutos do pseudofruto de caju (*Anacardium occidentale* L.). **Ciência e Tecnologia de Alimentos**, v. 27, n. 4, 2007.

CABRAL, E. C. **Utilização da Técnica de *Fingerprinting* por Espectrometria de Massas para a Análise de Extratos de Produtos Naturais**. Instituto de Química, Universidade de São Paulo, São Paulo, 2010.

CANUTO, G. A.; COSTA, J. L.; CRUZ, P. L. R.; SOUZA, A. R. L.; FACCIO, A. T.; KLASSEN, A.; RODRIGUES, K. T.; TAVARES, M. F. M. Metabolômica: Definições, Estado-da-Arte e Aplicações Representativas. **Química Nova**, v. XY, n. 00, 2017.

CASTRO, A. C. R. d.; VIDAL, R. F.; BARROS, L. d. M.; VIDAL NETO, F. d. C.; BORDALLO, P. d. N.; ARAGÃO, F. A. S. d. Introdução, coleta e conservação de recursos genéticos do cajueiro. In: ARAÚJO, J. P. P. D. (Ed.). **Agronegócio caju: práticas e inovações**. Brasília, DF: Embrapa, v.7, cap. 1, 2013.

CAZAR, R. A. An exercise on chemometrics for a quantitative analysis course. **Journal of Chemical Education**, v. 80, n. 9, 2003.

CIOSEK, P.; BRZÓZKA, Z.; WRÓBLEWSKI, W.; MARTINELLI, E.; DI NATALE, C.; D'AMICO, A. Direct and two-stage data analysis procedures based on PCA, PLS-DA and ANN for ISE-based electronic tongue—Effect of supervised feature extraction. **Talanta**, v. 67, n. 3, 2005.

CUSTERS, D.; COURSELLE, P.; APERS, S.; DECONINCK, E. Chemometrical analysis of fingerprints for the detection of counterfeit and falsified medicines. **Reviews in Analytical Chemistry**, 2016.

DABONNE, S.; NAKA, T.; MARTIN, D. K.; PATRICE, K. L. Assessment of some biochemical parameters of apple juices from two cashew varieties as affected by three regions of Côte d'Ivoire. **Journal of Advances in Agriculture**, n. 2, v. 5, 2016.

DAMASCENO, L.; FERNANDES, F.; MAGALHAES, M.; BRITO, E. Evaluation and optimization of non enzymatic browning of cajúina during thermal treatment. **Brazilian Journal of Chemical Engineering**, v. 25, n. 2, 2008.

DAS, I.; ARORA, A. Post-harvest processing technology for cashew apple—A review. **Journal of Food Engineering**, v. 194, 2017.

DETTMER, K.; ARONOV, P. A.; HAMMOCK, B. D. MASS SPECTROMETRY-BASED METABOLOMICS. **Mass spectrometry reviews**, v. 26, n. 1, 2007.

DOWNS, GEOFF M.; BARNARD, JOHN M. Clustering methods and their uses in computational chemistry. **Reviews in computational chemistry**, v. 18, 2002.

DU, L.; LU, W.; CAI, Z.; BAO, L.; HARTMANN, C.; GAO, B.; YU, L. Rapid detection of milk adulteration using intact protein flow injection mass spectrometric fingerprints combined with chemometrics. **Food Chemistry**, v. 240, n. Supplement C, 2018.

DUNN, W. B.; ERBAN, A.; WEBER, R. J. M.; CREEK, D. J.; BROWN, M.; BREITLING, R.; HANKEMEIER, T.; GOODACRE, R.; NEUMANN, S.; KOPKA, J.; VIANT, M. R. Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. **Metabolomics**, v. 9, n. 1, 2013..

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA (EMBRAPA). Cultivo do Cajueiro. **Embrapa Agroindústria Tropical**. Sistemas de Produção, 1. ISSN 1678-8702 Versão eletrônica. Jul/2016. Disponível em: <https://www.spo.cnptia.embrapa.br/conteudo?p_p_id=conteudoportlet_WAR_sistemaasdeproducaolf6_1ga1ceportlet&p_p_lifecycle=0&p_p_state=normal&p_p_mode=vie>

w&p_p_col_id=column-2&p_p_col_count=1&p_r_p_-76293187_sistemaProducaoid=7705&p_r_p_-996514994_topicoid=10308>. Acesso em: 15 jun. 2017.

EVANS, A. M.; DEHAVEN, C. D.; BARRETT, T.; MITCHELL, M.; MILGRAM, E. Integrated, Nontargeted Ultrahigh Performance Liquid Chromatography/Electrospray Ionization Tandem Mass Spectrometry Platform for the Identification and Relative Quantification of the Small-Molecule Complement of Biological Systems. **Analytical Chemistry**, v. 81, n. 16, 2009.

FAO - Food and Agriculture Organization of the United Nations. **FAOSTAT**. Production. Crops. 2014. Disponível em:<<http://www.fao.org/faostat/en/#compare>>. Acesso em: 15 de junho de 2017.

FERREIRA, F. R. Germplasm of fruit crops. **Revista Brasileira de Fruticultura**, v. 33, n. SPE1, 2011.

FIGUEIREDO JUNIOR; H. S. Análise da Indústria de castanha de caju inserção das micro e pequenas empresas no mercado internacional. **Revista USAID – Brasil**. v. 1,out. 2006.

FUNARI, C. S.; CASTRO-GAMBOA, I.; CAVALHEIRO, A. J.; BOLZANI, V. d. S. Metabolômica, uma abordagem otimizada para exploração da biodiversidade brasileira: estado da arte, perspectivas e desafios. **Química Nova**, 2013.

GAD, H. A.; BOUZABATA, A. Application of chemometrics in quality control of Turmeric (*Curcuma longa*) based on Ultra-violet, Fourier transform-infrared and ¹H NMR spectroscopy. **Food Chemistry**, v. 237, n. Supplement C, 2017.

GARCIA-ALCEGA, S.; NASIR, Z. A.; FERGUSON, R.; WHITBY, C.; DUMBRELL, A. J.; COLBECK, I.; GOMES, D.; TYRREL, S.; COULON, F. Fingerprinting outdoor air environment using microbial volatile organic compounds (MVOCs) – A review. **TrAC Trends in Analytical Chemistry**, v. 86, n. Supplement C, 2017.

GARRUTI, D. S.; FRANCO, M. R.; SILVA, M. A.; JANZANTTI, N. S.; ALVES, G. L. Evaluation of volatile flavour compounds from cashew apple (*Anacardium occidentale* L) juice by the Osme gas chromatography/olfactometry technique. **Journal of the Science of Food and Agriculture**, v. 83, n. 14, 2003.

HO, C. S.; LAM, C. W. K.; CHAN, M. H. M.; CHEUNG, R. C. K.; LAW, L. K.; LIT, L. C. W.; NG, K. F.; SUEN, M. W. M.; TAI, H. L. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. **The Clinical Biochemist Reviews**, v. 24, n. 1, 2003.

KHALIL, M. N.A.; FEKRY, M. I.; FARAG, M. A. Metabolome based volatiles profiling in 13 date palm fruit varieties from Egypt via SPME GC–MS and chemometrics. **Food Chemistry**, v. 217, 2017.

KHAN, H.; ALI, J. UHPLC/Q-TOF-MS Technique: Introduction and Applications. **Letters in Organic Chemistry**, v. 12, 2015.

KUMAR, N.; BANSAL, A.; SARMA, G.; RAWAL, R. K. Chemometrics tools used in analytical chemistry: An overview. **Talanta**, v. 123, 2014.

KUPIEC, T. Quality-Control Analytical Methods: High-Performance Liquid Chromatography. **International Journal of Pharmaceutical Compounding**, v. 8, n. 3, 2004.

LAVINAS, F. C.; ALMEIDA, N. C. d.; MIGUEL, M. A. L.; LOPES, M. L. M.; VALENTE-MESQUITA, V. L. Estudo da estabilidade química e microbiológica do suco de caju in natura armazenado em diferentes condições de estocagem. **Food Science and Technology**, v. 26, 2006.

LIMA, G. A. S. **Programa da Cajucultura no Ceará**. CONAB - Companhia nacional de abastecimento. Disponível em: http://www.conab.gov.br/OlalaCMS/uploads/arquivos/17_02_09_14_35_39_cajucultura_no_ceara_-_2017.pdf. Acesso em: 15 jun. 2017.

LOWOR, S.; AGYENTE-BADU, C. Mineral and proximate composition of cashew apple (*Anacardium occidentale* L.) juice from northern savannah, forest and coastal savannah regions in Ghana. **American Journal of Food Technology**, v. 4, n. 4, 2009.

MALDANER, L.; JARDIM, I. C. S. F. UHPLC - Uma abordagem atual: desenvolvimentos e desafios recentes. **Scientia Chromatographica**, v. 4, n. 3, 2012.

MALDANER, L.; JARDIM, I. C. S. F. O estado da arte da cromatografia líquida de ultra eficiência. **Química Nova**, São Paulo, v. 32, n. 1, 2009.

MASUOKA, N; KUBO; I. Characterization of xanthine oxidase inhibition by anacardic acids. **Biochimica et Biophysica Acta**, v. 1688, 2004.

MATERAZZI, S.; GREGORI, A.; RIPANI, L.; APRICENO, A.; RISOLUTI, R. Cocaine profiling: Implementation of a predictive model by ATR-FTIR coupled with chemometrics in forensic chemistry. **Talanta**, v. 166, n. Supplement C, 2017.

MICHODJEHOUN-MESTRES, L.; SOUQUET, J.-M.; FULCRAND, H.; BOUCHUT, C.; REYNES, M.; BRILLOUET, J.-M. Monomeric phenols of cashew apple (*Anacardium occidentale* L.). **Food Chemistry**, v. 112, n. 4, 2009.

MITCHELL, J. D.; MORI, S. A. The Cashew and Its Relatives. **Memoirs of the New York Botanical Garden**, v. 42, 1987.

MOITA NETO, J. M. **Estatística multivariada**: uma visão didática-metodológica. Revista Filosofia da Ciência, 2004.

MOURA, C. F. H.; Alves, R. E.; INNECCO, R.; FILGUEIRAS, H. A. C.; MOSCA, J. L.; PINTO, S. A. A. Características físicas de pedúnculos de cajueiro para comercialização *in natura*. **Revista Brasileira de Fruticultura**, v. 23, n. 03, 2001.

OLIVEIRA, M. E. B.; DE OLIVEIRA, G. S. F.; MAIA, G. A.; MOREIRA, R. d. A.; MONTEIRO, A. C. d. O. Aminoácidos livres majoritários no suco de caju: variação ao longo da safra. **Revista Brasileira de Fruticultura**, v. 24, n. 1, 2002.

PANERO, F. S.; VIEIRA, M. F. P.; CRUZ, Â. M. F.; MOURA, M. F. V.; SILVA, H. E. B. d. Aplicação da análise exploratória de dados na discriminação geográfica do quiabo do Rio Grande do Norte e Pernambuco. **Eclética Química**, v. 34, 2009.

PANERO, S.; FRANCISCO, S.; PANERO, S.; JOAO, S.; DA SILVA, B.; EBSEDUARDO, H. Application of HCA and PCA in the discrimination of not polluted tubular wells from tubular wells with incidence of anthropogenic pollution in Western Amazon Region. **Chinese Journal of Geochemistry**, v. 25, n. 1, , 2006.

PASQUINI, C. Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. **Journal of the Brazilian Chemical Society**, São Paulo, v. 14, n. 2, 2003 .

PATRAS, A.; BRUNTON, N. P.; DOWNEY, G.; RAWSON, A.; WARRINER, K.; GERNIGON, G. Application of principal component and hierarchical cluster analysis to classify fruits and vegetables commonly consumed in Ireland based on in vitro antioxidant activity. **Journal of Food Composition and Analysis**, v. 24, n. 2, p. 250-256, 2011.

PESSOA, P. F. A. d. P.; LEITE, L. A. d. S. Desempenho do agronegócio caju brasileiro. In: ARAÚJO, J. P. P. D. (Ed.). **Agronegócio caju: práticas e inovações**. Brasília: Embrapa, v.1, cap. 1, 2013.

PINHO, L. X.; AFONSO, M. R. A.; CARIOCA, J. O. B.; COSTA, J. M. C. d.; RAMOS, A. M. The use of cashew apple residue as source of fiber in low fat hamburgers. **Food Science and Technology (Campinas)**, v. 31, n. 4, 2011.

POPPI, R. J.; PASQUINI, C. Spectrophotometric determination of a mixture of weak acids using multivariate calibration and flow injection analysis titration. **Chemometrics and intelligent laboratory systems**, v. 19, n. 2, 1993.

RIBEIRO, L. P. D.; SILVA, A. P. M.; LIMA, A. A.; SILVA, E. O.; RINNAN, Â.; PASQUINI, C. Non-Destructive Determination of Quality Traits of Cashew Apples (*Anacardium Occidentale*, L.) Using a Portable near Infrared Spectrophotometer. **Journal of Near Infrared Spectroscopy**, V. 24, n. 1, 2016.

RUSCHEL, C. F. C. **Aplicação de ferramentas quimiométricas e técnicas espectroscópicas na análise de combustíveis**. Tese (Doutorado) - Instituto de Química, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 2017.

SABE, I. L. S. **Avaliação das características físico-químicas e sensoriais de diferentes formulações do néctar de caju (*Anacardium Occidentale* L.)**. Tese (Doutorado) - Departamento de Engenharia Química, Universidade Eduardo Mondlane, Maputo, Moçambique, 2016.

SANTOS, R. P.; SANTIAGO, A. A. X.; GADELHA, C. A. A.; CAJAZEIRAS, J. B.; CAVADA, B. S.; MARTINS, J. L.; OLIVEIRA, T. M.; BEZERRA, G. A.; FREIRE, V. N. Production and characterization of the cashew (*Anacardium occidentale* L.) peduncle bagasse ashes. **Journal of Food Engineering**, v. 79, n. 4, 2007.

SILVA, Ana Veruska Cruz; MUNIZ, E. N.; Ledo, A. S.; Rabbani, A. R. C.; Amorim, J. A. E.; Vitória, M. F. Diversidade genética em germoplasma de *Anacardium occidentale*. **Scientia Plena**, v. 10, n. 11, 2014.

SIMÕES, C. M. O.; SCHENKEL, E. P.; DE MELLO, J. C. P.; MENTZ, L. A.; PETROVICK, P. R. **Farmacognosia: do Produto Natural ao Medicamento**. Porto Alegre: Artmed, 2016.

SIMÕES, S. **Desenvolvimento de métodos validados para a determinação de captopril usando espectrofotometria NIR e calibração multivariada**. Tese (Doutorado em Ciências) - Universidade Federal da Paraíba, João Pessoa, 2008.

SOUZA, A. M. d.; POPPI, R. J. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte I. **Química Nova**, 2012.

TEÓFILO, R. **Métodos Quimiométricos: uma Visão Geral: conceitos básicos de quimiometria**. Viçosa: UFV, v. 1, 2013.

TREVISAN, M.T.S.; PFUNDSTEIN, B.; HAUBNER, R.; WÜRTELE, G.; SPIEGELHALDER, B.; BARTSCH, H.; OWEN, R. W. Characterization of alkyl phenols in cashew (*Anacardium occidentale*) products and assay of their antioxidant capacity. **Food and Chemical Toxicology**, v. 44, n. 2, 2006.

TSUCHIKAWA, S.; KOBORI, H. A review of recent application of near infrared spectroscopy to wood science and technology. **Journal of Wood Science**, v. 61, n. 3, p. 213-220, 2015.

UARROTA, V. G.; MORESCO, R.; COELHO, B.; NUNES, E. C.; PERUCH, L. A. M; NEUBERT, E. O.; ROCHA, M.; 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**, v. 161, 2014.

VANDEGINSTE, B. G. M.; MASSART, D. L.; BUYDENS, L. M. C.; DE JONG, S.; LEWI, P. J.; SMEYERS-VERBEKE, J. **Handbook of Chemometrics and Qualimetrics**. Amsterdam: Elsevier, 1998.

VIDAL, M. F. **Situação da cajucultura nordestina após a seca**. Fortaleza: ETENE, 2016.

VILLAS-BÔAS, S. G.; MAS, S.; ÅKESSON, M.; SMEDSGAARD, J.; NIELSEN, J. Mass spectrometry in metabolome analysis. **Mass Spectrometry Reviews**, v. 24, n. 5, 2005.

WEI, Y.; AN, Q.; LI, J.; SONG, H.; JI, X.; YANG, Y.; YUE, G. Fingerprint and multicomponent quantitative analysis for the quality evaluation of *Sibiraea angustata* leaves by HPLC-DAD and HPLC-ESI-MS/MS combined with chemometrics. **Journal of Liquid Chromatography & Related Technologies**, v. 40, n. 9, 2017.

WISE, B. M.; GALLAGHER, N. B.; BUTLER, S. W.; WHITE, D. D.; BARNA, G. G. A comparison of principal component analysis, multiway principal component analysis,

trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process. **Journal of Chemometrics**, v. 13, n. 3-4, 1999.

WOHLMEISTER, D.; VIANNA, D. R. B.; HELFER, V. E.; CALIL, L. N.; BUFFON, A.; FUENTEFRIA, A. M.; CORBELLINI, V. A.; PILGER, D. A. Differentiation of *Candida albicans*, *Candida glabrata*, and *Candida krusei* by FT-IR and chemometrics by CHROMagar™ *Candida*. **Journal of Microbiological Methods**, v. 141, n. Supplement C, 2017.

WU, H.; GUO, J.; CHEN, S.; LIU, X.; ZHOU, Y.; ZHANG, X.; XU, X. Recent developments in qualitative and quantitative analysis of phytochemical constituents and their metabolites using liquid chromatography–mass spectrometry. **Journal of pharmaceutical and biomedical analysis**, v. 72, 2013.

ZAITSU, K.; MIYAWAKI, I.; BANDO, K.; HORIE, H.; SHIMA, N.; KATAGI, M.; TATSUNO, M.; BAMBA, T.; SATO, T.; ISHII, A.; TSUCHIHASHI, H.; SUZUKI, K.; FUKUSAKI, E. Metabolic profiling of urine and blood plasma in rat models of drug addiction on the basis of morphine, methamphetamine, and cocaine-induced conditioned place preference. **Analytical and bioanalytical chemistry**, v. 406, n. 5, 2014.

ZHANG, B.; LI, X.; YAN, B. Advances in HPLC detection—towards universal detection. **Analytical and Bioanalytical Chemistry**, v. 390, n. 1, 2008.

4 CAPÍTULO 1 - METABOLIC VARIATION ON JUICES FROM DIFFERENT GENOTYPES OF CASHEW APPLE BY NON-TARGETED UPLC-QToF-MS

Ynayara C. de Lima^a; Elenilson G. Alves Filho^a; Lorena Mara A. Silva^a; Ana Cecília R. Castro^a; Guilherme J. Zocolo^a; Kirley M. Canuto^a; Selene M. Morais^b; Edy S. de Brito^a

^a Embrapa Agroindústria Tropical, Fortaleza-CE, Brazil.

^b Centro de Ciências e Tecnologia, Universidade Estadual do Ceará, Fortaleza-CE, Brazil.

* Corresponding author

E-mail: edy.brito@embrapa.br

Phone: +55 85 33917393

Fax: +55 85 33917100

4.1 ABSTRACT

Cashew apple (*Anacardium occidentale* L.), presents beneficial characteristics to human health as antitumor, antimicrobial, urease inhibitory and lipoxygenase activity due its flavonoids content. Embrapa Cashew Germplasm Bank develop researches of cashew quality through genetic enhancement of cashew. Twenty four genotypes of the Germplasm Bank were analysed using hyphenated techniques (LC-MS). Cashew juices were analyzed in the negative ionization mode. A numerical matrix was created using the chromatographic data and Hierarchical Cluster Analysis (HCA) method was applied to segregate samples in groups in relation to their similarity. A trend of formation of four cashew clusters with the similarity index of 0.362 was observed. The clustering observed in Hierarchical HCA was important for interpreting the Principal Component Analysis (PCA) plot. Galloylhexose I (1.60 min, 331.0550 m/z), digalloylhexoside I (2.82 min, 483.0741 m/z), hydroxybutanoic acid hexoside ethyl ester (3.28 min, 293.1224 m/z), myricetin-3-O-glucoside (3.83 min, 479.0826 m/z), myricetin-3-O-rhamnoside (4.25 min, 463.0875 m/z) and a mixture of unknown compounds (4.82 min) were the main responsible for the separation of juices. The PLS-DA method was used to improve the association of chemical variability according to the cashew genotype and presents a prediction capacity of 88.05% with three latent variables (LVs). LC-MS associated with chemometric tools is a good way to classify cashew samples into groups by similarity among them. The compounds identified as responsible for the separation of the groups can be the “key-metabolites” for the identification of cashew juice samples, but more reseraches are needed to confirm this hypothesis.

Keywords: cashew juice; UPLC-MS; chemometrics

4.2 INTRODUCTION

The development of studies of flavonoid content in fruits and vegetables is well documented since the metabolomic profiling may influence significantly the juice characteristics such as organoleptic properties (flavor and aroma), stability, and/or microbiologic control [1, 2]. In addition, biological activity and effects of food derivative flavonoids are extensively reported as anti-allergenic, anti-inflammatory, antimicrobial, anti-oxidant and others [3, 4]. Particularly, cashew apple presents beneficial characteristics to human health as antitumor, antimicrobial, urease inhibitory and lipoxygenase activity [5-8]. Genetic enhancement of cashew quality is still in development in Brazil reaching knowledge of genetic diversity and construction of germplasm banks. Most of cashew breeding programs are based on traditional selection approaches as size and weight of nuts or yield to evaluate cashew tree performance, but efforts are also applied in the prospection of dwarf genotypes with better fruit quality and resistance to pests and disease [9]. In this context, genotypes of cashew (*Anacardium occidentale*) are been cultivated in experimental fields by Embrapa at Pacajus county, Ceará state (Brazil). At the same time, Embrapa maintains a cashew germplasm bank also located in Pacajus. This collection holds more than six hundred accessions with great variability, however, there are few information regarding the metabolites variations on cashew apple composition according to the genotype and genetic enhancement [10]. The identification of the characteristics with economic interest related to conserved or cultivated accessions may be appropriated by breeding programs aiming to launching of high performance cultivars for specific demands, such as cultivars with high levels of bioactive compounds.

Traditionally, liquid chromatography coupled to photodiode array detector (PDA) and/or mass spectrometry (MS) has been used as a tool for structure characterization and elucidation of organic compounds in cashew juice matrices, driving by the needs of multiple screening applications. For cost reason, less time and solvent expenditure, cashew apple juice profiling based on ultra performance liquid chromatography (UPLC) combined to MS detection is being optimally achieved. In such foodstuff studies, a large number of samples have to be screened looking for identify metabolites and their classification according to the geographic origin, climate conditions, genotype, and cultural practices using multivariate statistics [5, 11-14].

Usually, multivariate analyses are applied to explore complex matrices as foodstuff to determine variations and relationships among the samples composition [15, 16]. This untargeted approach is advantageous when the compounds and degradation products are not always known.







Therefore, due to the need of the investigation of plants that produce more and more beneficial compounds to the health, the aim of this study was evaluate the metabolites profiling variability on juices from 24 different genotypes of cashew apples.









4.3 EXPERIMENTAL








4.3.1 Sampling




The juice samples were obtained from different genotypes of cashew apples cultivated at Embrapa Experimental Station (Pacajus, Ceará, Brazil) with the coordinates 4°11'07"S; 38°30'07"W and an altitude of 70 m above the sea level. Table 1 presents 24 genotypes of the cashew samples with their respective accession numbers and morphoagronomic characteristics. The region has a tropical climate, average temperatures of 26-28 °C and 791,4 mm average annual rainfall. The soil, classified as Ultisol, has a sandy/medium texture and low organic matter content. The cashew samplings were collected between August and December in two harvests (2012 and 2013).

Table 1 - Genotype of the 24 cashew fruits, respective accession numbers, and morphoagronomic characteristics: plant size; tree appearance; fruits color and shape; origin (county-state); and illustration.

Accession number	Plant size	Tree appearance	Pseudofruit color	Pseudofruit shape	Sampling origin	Illustration
CP 76	tall	open erect	orange	pyriform	crop* / Maranguape-CE	
Clone 98/101	semi tall	compact erect	orange	pyriform	breeding program* / Pacajus-CE	
Progeny 2005/127	semi tall	compact erect	dark red	pyriform	breeding program/ Beberibe-CE	
Progeny 2005/133	semi tall	open erect	orange	spherical	breeding program/ Cruz-CE	
BRS 226	dwarf	compact erect	orange	pyriform	crop / Pio IX-PI	
Clone 2005/102	tall	compact erect	orange	pyriform	breeding program/ Beberibe-CE	

CP 09	semi tall	compact erect	orange	pyriform	crop / Maranguap e-CE	
B 393	tall	compact erect	light red	spherical	germplasm * / Aracati-CE	
BRS 275	semi tall	open erect	orange	pyriform	crop / Pacajus-CE and Maranguap e-CE	
B 963	tall	open erect	yellow orange	pyriform	germplasm / Maranguap e-CE	
Hybrid 2001/3	semi tall	compact erect	orange	pyriform	breeding program/ Maranguap e-CE e Pio IX-PI	
Hybrid 2001/6	semi tall	compact erect	yellow orange	pyriform	breeding program/ Maranguap e-CE e Pio IX-PI	
B 967	tall	open erect	orange	cylindrical	germplasm / Pacajus-CE	
CP 06	tall	open erect	yellow	conical obovate	crop / Pacajus-CE	

Progeny 2005/122	semi tall	open erect	yellow	spherical	breeding program/ Beberibe- CE	
Hybrid 2001/13	semi tall	open erect	orange	pyriform	breeding program/ Pacajus- CE	
Clone 2005/111	semi tall	open erect	orange	pyriform	breeding program/ Serra do Mel-RN	
Clone 98/116	semi tall	open erect	orange	pyriform	breeding program/ São Luiz do Curu- CE	
B741	semi tall	compact erect	orange	pyriform	breeding program (CP 76 x <i>A. microcarpu m</i>) / Maranguap e-CE	
Progeny 2005/223	semi tall	open erect	orange	pyriform	breeding program/ Beberibe- CE	
Embrapa 51	semi tall	open erect	orange	pyriform	crop / Pacajus- CE	

M886	tall	open erect	yellow	spherical	breeding program/ Beberibe- CE	
1001	tall	open erect	orange	pyriform	crop/ Pacajus- CE	
Clone 2003/102	semi tall	compact erect	orange	pyriform	breeding program/ Pio IX -PI	

* Legends: crop means registered product on market; breeding program means plant before crossbreed; germplasm means plant collected and conserved in germoplasm bank.

The cashew apple juices were obtained under the same circumstances. Cashews of each genotype were manually pressed to obtain the juice and then centrifuged for 5 min at 605 g (IEC Clinical Centrifuge, Damon/IEC Division, Needham, MA). The samples were conserved in freezer at -18 °C until the analysis.

4.3.2 Sample preparation and UPLC-qTOF-MS

The cashew juice samples were filtered through PTFE membranes of 0.22 μm . The UPLC-qTOF-MS analysis was performed on Acquity system (Waters) coupled with quadrupole / TOF (Waters) equipped with an ESI source operated in the negative ion mode. The chromatographic separation was performed using a Waters Acquity UPLC BEH (150.0 x 2.1 mm, 1.7 μm) column with temperature set at 40 °C. Water and acetonitrile were used for mobile phase, both with 0.1 % of formic acid. The gradient ranged from 2 to 95 % of water in 15 min in a flow of 0.4 mL.min⁻¹ and injection volume of 5.0 μL per sample. N₂ was used as the desolvation gas. The desolvation temperature was set at 350 °C with a flow rate of 350 L.h⁻¹ and a source temperature of 120 °C. The capillary voltage was set to 3,200 V. The collision energies/cone voltages were set at 6 eV/15 V (low) and 30-50 eV/30 V (high) to achieve sufficient

fragmentation. Data were collected using negative ionization mode between 100 Da and 1180 Da, and the mode tandem was MS^E.

4.3.3 Chemometric analyses

Chemometric analyses were performed in the cashew juice matrix acquired in negative ionization mode from the 24 samples harvested with duplicate of year (24 in 2012 and 24 in 2013) and triplicate of the preparation of cashew juice samples, which result in 144 chromatograms. The region of all the chromatograms between 0.65 and 7.12 min were used for the analysis, resulting in a matrix with dimensionality of 104,112 data points (144 chromatograms × 723 variables). The samples were named according to the accession numbers (Table 1).

All the chromatograms data were converted to American Standard Code for Information Interchange (ASCII) files to matrix construction, which was exported for chemometric analysis by Hierarchical Cluster Analysis (HCA), Principal Component Analysis (PCA), and Partial Least Square – Discriminant Analysis (PLS-DA) using The Unscrambler XTM program (version 10.4). Before the application of the chemometric algorithms, the noise and imperfection regions were removed. For HCA, the matrix was mean-centered and the incremental linkage method (sum-of-squares approach in calculating intercluster distances) was applied. The Euclidian distance was used for distance metric. The PCA was performed using Singular Value Decomposition (SVD) algorithm and the mean-centered processing was applied over the variables since this treatment provides better difference among the samples. In addition, normalize scaling parameter and baseline correction using linear fit algorithms were applied over the samples to reduce noises and minor analytical errors [17, 18]. In order to improve the identification of chemical constituents according to the genotype of the cashew, the supervised method PLS-DA was employed and NIPALS (Nonlinear Iterative Partial Least Squares) algorithm was used to construct the model. The number latent variables (LVs) were selected in accordance with the statistical parameters: SEC (Square Error of Calibration); SEV (Square Error of Validation); coefficient of correlation between the real concentration and the concentration predicted during the calibration; coefficient of correlation between the real concentration and the concentration predicted during the validation; and similarity criterion SEC/SEV ratio above 0.75 [15, 19].

4.3.4 Relative contribution

All single peak detected as exactly as possible in both m/z and retention time was used for determining peak area correctly for successful determinates the relative contribution of the compounds with no-overlapped signals in the chromatograms. The relative contribution peaks areas were calculated based on total ion abundance from the peaks in the samples, since the relative amplitude of the peaks measures provides the relative abundance of the isotopic forms in the chromatograms. Therefore, the normalized mean in the base peak intensity (BPI) at the retention times 1.60 min, 2.82 min, 3.28 min, 3.83 min, 4.25 min were determined.

The results were evaluated using the analysis of variance ANOVA single factor (significance level of 0.05; means comparison using Tukey test; Levene to test the homogeneity of variance) in order to statistically certify the differences among the relative contributions. The deviation of the method was estimated based on null hypothesis (p-value) from the three replicates of sampling during two years – totaling of 6 samples for each cashew genotype.

4.4 RESULTS AND DISCUSSION

Commonly, in liquid chromatography coupled to mass spectrometry analysis some organic compounds may preferentially ionize using the positive or negative ionization mode, as phenolic and carboxylic derivatives ionized well in negative ionization while flavonoids and alkaloids ionized better in positive ionization mode. Therefore, the cashew juices were analyzed under negative ionization mode in order to screen the metabolites and to differentiate each juice according to the genotypes of the cashew apple.

Previously the chemometric analysis, a valuable feature of chromatograms acquisition must be taken into account: the retention times of the chromatographic peaks are sensitive to minor fluctuations in temperature, pH, flow, and pump operation. In order to solve the problem of small peaks shifts related to the retention times equivalent to a same compound into the chromatograms, some different peak alignment methods may be established. This alignment practice can be performed manually, or using the COW (Correlation Optimized Warping) available in some

chemometric software [20], or overcome using the bucketing method, which reduce the chromatogram dimensionally slicing it one in equally sized regions [21], making it easier the respective loadings analysis [22]. Therefore, all the chromatogram peaks presented in Figure 1 were previously aligned using the COW method available in the software. Furthermore, Figure 1 shows the complexity of the chromatographic data for a visual differentiation of the samples composition and the necessity of a multivariate evaluation since the use o certified standard compounds and the construction of calibration curves are not required.

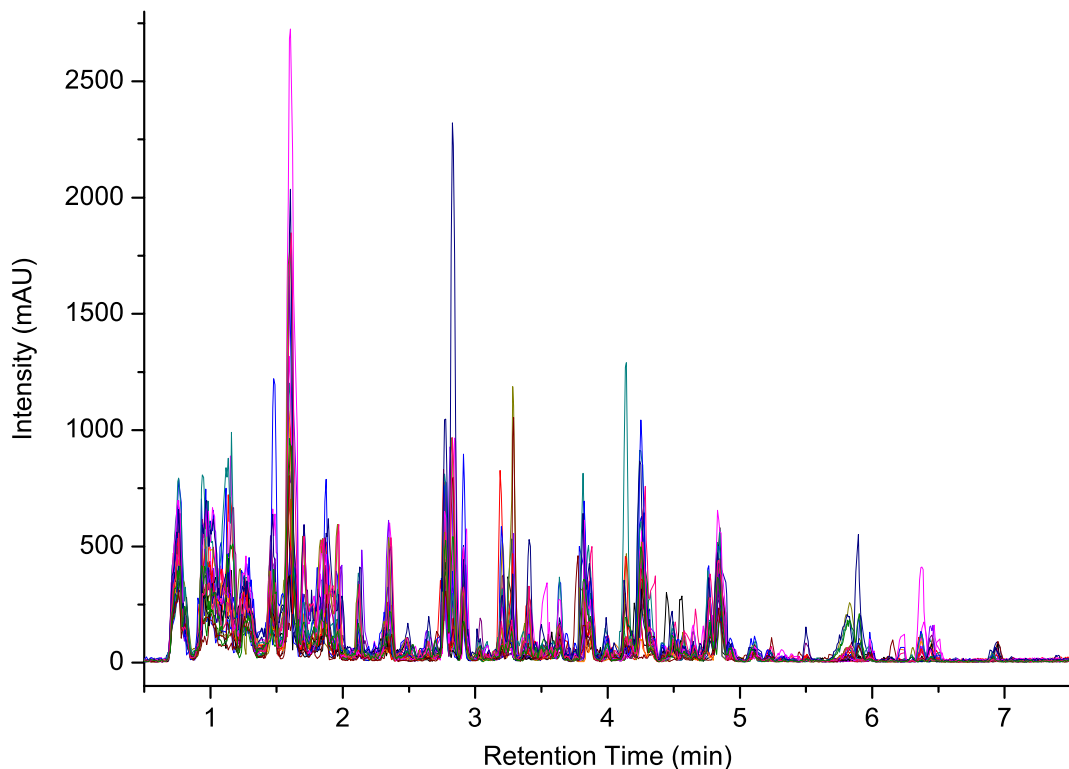


Figure 1 - Total ion chromatograms from 24 different genotypes of cashew juices acquired under negative ionization mode.

In order to comprehend the variability of the organic compounds in cashew juices according to their genotype, a numerical matrix was created using the chromatographic data and a non-targeted multivariate analysis was applied. Initially, the unsupervised method based on HCA was applied to segregate the samples in groups regarding their similarity. Important tendencies of four clustering were observed in the dendrogram according to the genotype of the cashew at the similarity index

0.362 (Figure 2). The cluster 4 presented the most distant samples included in the study, zero of similarity index. The results reflected the natural differences among the samples groups, which were formed as a function of juice composition dependent of the cashew genotype.

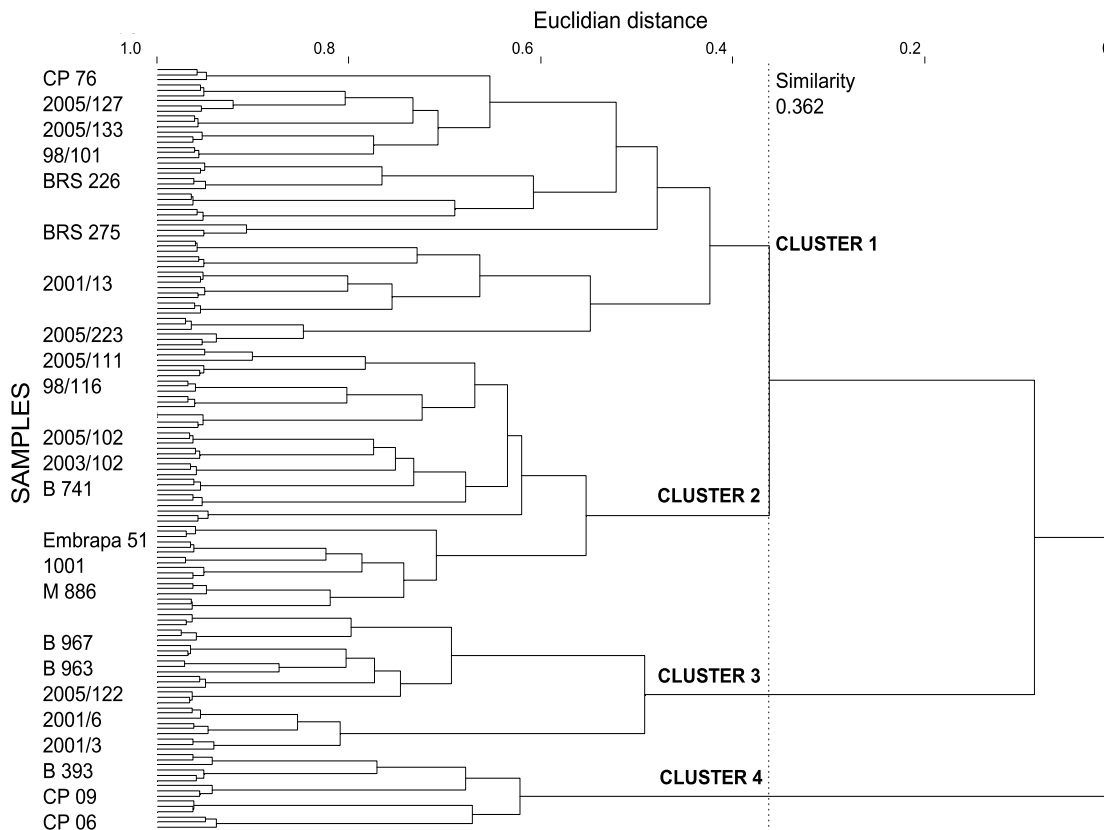


Figure 2 - Dendrogram representing the chemical composition similarity relationships among the cashew juices.

Additionally, PCA analysis was applied to assist the modeling and interpretation of the multivariate data, with the scores graph presented in Figure 3a and the respective loadings plotted in lines in Figure 3b. The clustering observed in HCA were important to interpret the resultant scores graph, which exhibited the main separation tendencies of the samples in respect to PC1 and PC2 axes, with a total explained variance of 36.7 %. Despite of the low total variance cumulated in the first two principal components (PCs), significant chemometric variations were observed in the composition profile detected in the cashew juices prepared from different genotypes. Furthermore, the subsequent PCs presented same relevant results

retained in both PC1 and PC2 axes. In general, the samples were symbolized according to the clustering tendency and the scores region: blue triangles at negative values of PC1 and PC2; brown squares at positive values of PC1 and negative of PC2; black stars at positive values of PC1 and PC2; green circles only at positive values of PC2. Each sample with no relevant result was symbolized by gray circle, since the replicates of the sampling according to the year were not representative.

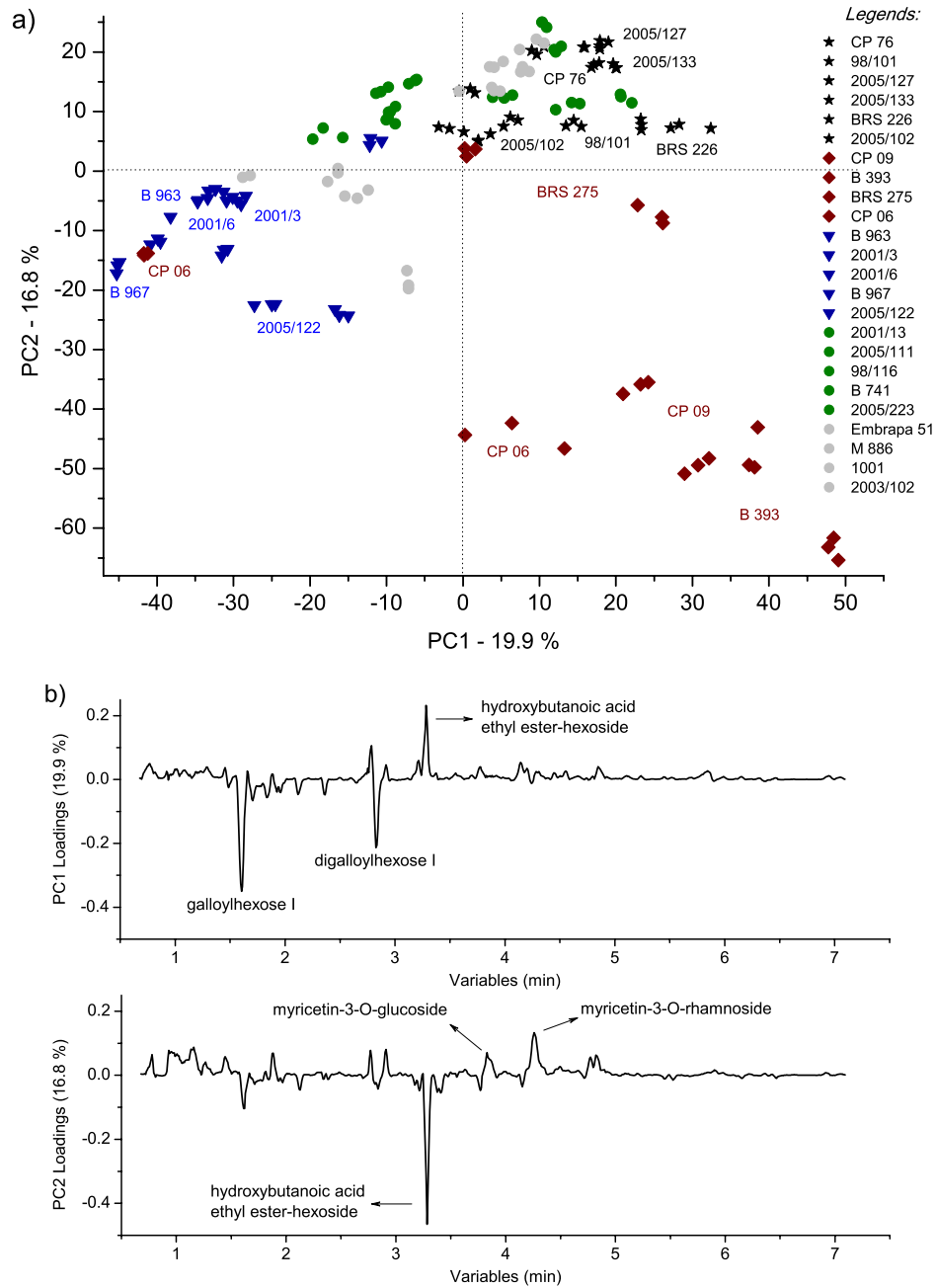


Figure 3 - a) PC1 vs PC2 scores coordinate system for the cashew juices from different genotypes; b) respective loadings plotted in lines form. Samples were symbolized according to the clustering tendency: blue triangles at negative values of PC1 and PC2; brown squares at positive values of PC1 and negative of PC2; black stars at positive values of PC1 and PC2; green circles only at positive values of PC2. Samples that did not presented relevant results were symbolized by gray circle.

An examination of the PC1 and PC2 loadings plotted in lines are represented by Figure 4, which provided clear evidence of the variables responsible for the separations or clustering observed in the scores plot. Signals from galloylhexose I (at 1.60 min, 331.0650 m/z), digalloylhexoside I (at 2.82 min, 483.0741 m/z), hydroxybutanoic acid ethyl ester-hexoside (at 3.28 min, 293.1242 m/z), myricetin-3-O-glucoside (at 3.83 min, 479.0826 m/z), myricetin-3-O-rhamnoside (at 4.25 min, 463.0875 m/z), and an mixture of unknown compounds (at 4.82 min) were the main responsible for the placement of the juices from different genotypes. Minor compounds were not relevant in this study due to the pretreatments of mean centering applied over the variables and normalization over the samples [23]. The negative loadings of PC1 are presenting mainly the juice samples symbolized by blue triangles with higher loadings of galloylhexose I and digalloylhexoside I. This later compound is a trihydroxybenzoic acid derivative, which provides astringent flavor and can contribute to immature cashew apple flavor characteristic [24].

The samples from the juices symbolized by brown squares and black stars presented tendencies to have higher amounts of hydroxybutanoic acid ethyl ester-hexoside at retention time of 3.28 min according to positive loadings of PC1. Hydroxybutanoic acid ethyl ester-hexoside has been previously described in melon fruit and it is considered a precursor of the volatile compounds. The presence of this compound has been associated with some amino acids such as alanine, glutamine, isoleucine, phenylalanine, tryptophan or tyrosine [25]. Alanine has been reported as one of the key amino acids of the characteristic profile of cashew apple juice. Phenylalanine and tyrosine have also been detected but in smaller amounts [26].

Positive loadings of PC2 presented tendency of the cashew juices symbolized by green circle and black stars to have higher amounts of myricetin-3-O-glucoside, myricetin-3-O-rhamnoside, and an unknown mixture of compounds between the retention times 4.79-4.85 min. The presence of myricetin-derived and others flavonoids compounds in fruits and corresponding juices offer biological benefits as the reductions of cardiovascular disease and risks of cancer [5], besides anti-hyperglycaemic property [27]. In addition, these compounds have been reported in methanol-water extract [5].

Compounds with significant changes according to the chemometric evaluation (Figure 4) and did not exhibit overlapping signals in the chromatograms were semi-quantified and their variations were expressed as relative contribution (see

experimental section 2.4). The relative peaks areas were calculated for quantitative expression of the chemical properties and the differences were evaluated by ANOVA single factor. Figure 5 illustrates (same color of Figure 3) the relative contributions from the total ion abundance of the peaks in the samples, since the relative amplitude of the peaks measures provides the relative abundance of the isotopic forms in the chromatograms. The results from the signals areas of the base peak intensity (BPI) chromatograms corroborated the PCA results considering the deviation of the method from the three replicates of sampling during two years, which totaled 6 samples for each cashew genotype. These results were deduced from the loadings plotted in Figure 4, with the hydroxybutanoic acid ethyl ester-hexoside (at 3.28 min) responsible for clustering the samples at brown group, and galloylhexose I (at 1.60 min) and digalloylhexoside I (at 2.82 min) for clustering the samples in the blue group.

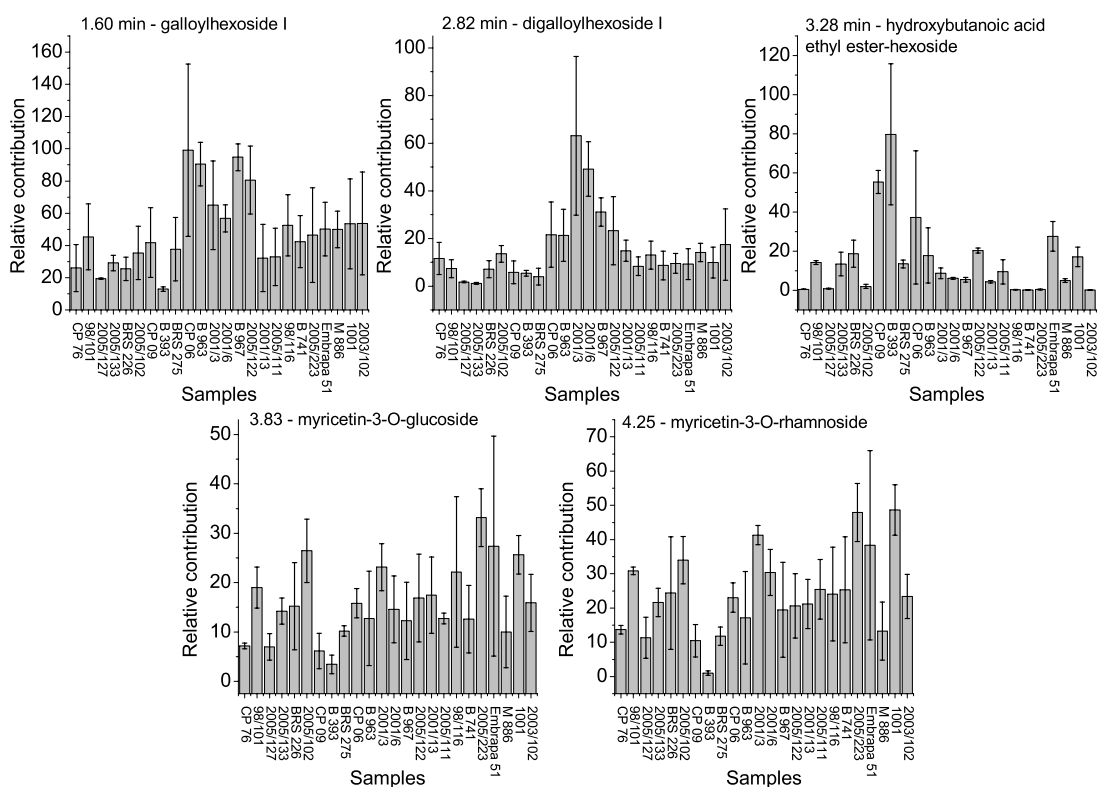


Figure 4 - Relative contributions of the isotopic forms in chromatograms for the compounds at retention times 1.60, 2.82, 3.23, 3.82, and 4.25 min.

The corroboration between quantitative and chemometric results confirms the advantageous for untargeted analysis because the compounds and degradation products are not always known, it is sometimes difficult to find certified standard

compounds, and being simpler than the classical univariate analysis as quantification by construction of calibration curves for each compound.

Based on the non-targeted chemometric evaluation (Figure 3) and the relative quantification results (Figure 4), a classification model using Partial Least Squares – Discriminant Analysis (PLS-DA) was employed to improve the association of the chemical variability according to the cashew genotype, which is illustrated at Figure 5. With three latent variables (LVs) the model presented a prediction ability of 88.05 % considering the replicate of the year sampling and triplicate of chromatograms acquisition. The statistical parameters used to assess the quality of the classification method (Table 2) indicated a well-adjusted model, since the SEC/SEV ratio was above 0.75 (similar values) (BEEBE et al., 1998; WOLD et al., 2001). The low SEC and SEV values express the suitable predictive performance of the model estimated as a function of the global model error, samples leverage, and the sample residual X-variance. The low errors values indicated similarity between the sample used for prediction and those used to make the calibration model.

Table 2 - Parameters obtained by multivariate calibration using PLS-DA model.

<i>Model</i>	LV1+LV2+LV3 ^a	r ² cal ^b	SEC ^c	r ² val ^d	SEV ^e	SEC/SEV ^f
<i>PLS-DA</i>	88.05 %	0.88	0.298	0.85	0.341	0.874

^a Total variance percent in X matrix refer to three Latent Variables (LVs).

^b Coefficient of correlation between the real concentration and the concentration predicted during the calibration.

^c Standard Error of Calibration.

^d Coefficient of correlation between the real concentration and the concentration predicted during the validation.

^e Standard Error of Cross Validation.

^f Similarity criterion.

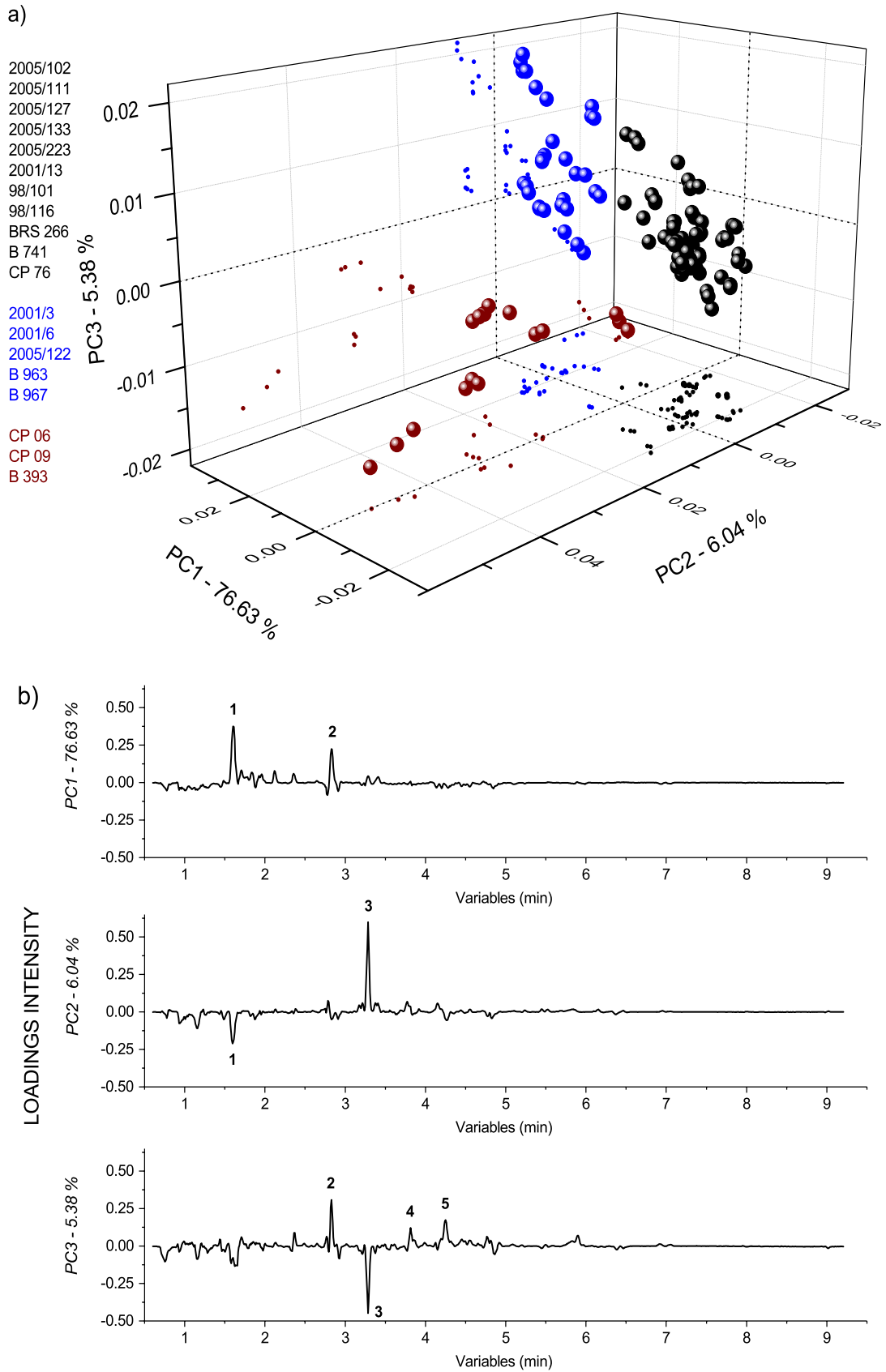


Figure 5 - LV1 vs. LV2 vs. LV3 scores 3-D plot (a) and the respective loadings plotted in lines form (b) from PLS-DA for the cashew juices classification. Legend: compounds 1 – galloylhexose I (1.60 min); 2 – digalloylhexoside I (2.82 min); 3 – hydroxybutanoic acid ethyl ester-hexoside (3.28 min); 4 – myricetin-3-O-glucoside (3.83 min); 5 – myricetin-3-O-rhamnoside (4.25 min).

In general, the LV1 axis was most relevant to cluster the cashew juice samples in black color (2005/102, 2005/111, 2005/127, 2005/133, 2005/223, 2001/13, 98/101, 98/116, BRS 266, B 741, CP 76) the their separation from those in brown (CP 06, CP 09, B 393) and blue (2001/3, 2001/6, 2005/122, B 963, B 967). In addition, the LV2 axis was important to clustering the brown samples at positive scores, and LV3 present relevance to separation the brown and blue cashew juice samples. The interpretation of the loadings presented the samples in black with higher amount of galloylhexose I and digalloylhexoside I than the brown and blue samples. The samples in brown color presented the higher amount of hydroxybutanoic acid ethyl ester-hexoside than the black and blue samples, beyond to present the opposite behavior between the hydroxybutanoic acid ethyl ester-hexoside and galloylhexose I. Lastly, LV3 axis presented the higher amount of the flavonoids myricetin-3-O-glucoside and myricetin-3-O-rhamnoside in the blue samples beyond to presents the opposite behavior between this flavonoids and hydroxybutanoic acid ethyl ester-hexoside. Therefore, the statistics data indicated that the model could be acceptable to classify new or unknown cashew juice samples.

4.5 CONCLUSIONS

Using hyphenated techniques it was possible to obtain the metabolic fingerprint of the studied genotypes. The chemical profiles allowed the classification of the samples into groups by similarity among them using chemometric tools. The identification of the major compounds responsible for the separation of the groups can be the “key-metabolites” for the identification of cashew juice samples, but new researches in these area are essential to confirm this hypothesis.

4.6 ACKNOWLEDGEMENTS

The authors thank CNPq and FUNCAP for the financial support and the award of scholarship (303791/2016-0).

4.7 REFERENCES

1. Garruti, D.S., et al., *Evaluation of volatile flavour compounds from cashew apple (Anacardium occidentale L) juice by the Osme gas chromatography/olfactometry technique*. Journal of the Science of Food and Agriculture, 2003. **83**(14): p. 1455-1462.
2. Zepka, L.Q., et al., *Aroma compounds derived from the thermal degradation of carotenoids in a cashew apple juice model*. Food Research International, 2014. **56**: p. 108-114.
3. Sultana, B. and F. Anwar, *Flavonols (kaempferol, quercetin, myricetin) contents of selected fruits, vegetables and medicinal plants*. Food Chemistry, 2008. **108**(3): p. 879-884.
4. Manach, C., A. Mazur, and A. Scalbert, *Polyphenols and prevention of cardiovascular diseases*. Current opinion in lipidology, 2005. **16**(1): p. 77-84.
5. de Brito, E.S., et al., *Determination of the flavonoid components of cashew apple (Anacardium occidentale) by LC-DAD-ESI/MS*. Food chemistry, 2007. **105**(3): p. 1112-1118.
6. Cavalcante, A.A.M., et al., *Cashew (Anacardium occidentale) apple juice lowers mutagenicity of aflatoxin B1 in S. typhimurium TA102*. Genetics and Molecular Biology, 2005. **28**(2): p. 328-333.
7. Kubo, I., K.-i. Nihei, and K. Tsujimoto, *Antibacterial action of anacardic acids against methicillin resistant Staphylococcus aureus (MRSA)*. Journal of Agricultural and Food Chemistry, 2003. **51**(26): p. 7624-7628.
8. Ha, T.J. and I. Kubo, *Lipoxygenase inhibitory activity of anacardic acids*. Journal of Agricultural and Food Chemistry, 2005. **53**(11): p. 4350-4354.
9. Chaves, N., et al. *Characterization of cashew accessions for ornamental purposes*. in *X International Symposium on Postharvest Quality of Ornamental Plants 1060*. 2012.
10. Castro, A., et al. *Brazilian cashew germplasm bank*. in *XXVIII International Horticultural Congress on Science and Horticulture for People (IHC2010): III International Symposium on 918*. 2010.
11. Scherer, R., et al., *Validation of a HPLC method for simultaneous determination of main organic acids in fruits and juices*. Food Chemistry, 2012. **135**(1): p. 150-154.
12. Zepka, L.Q. and A.Z. Mercadante, *Degradation compounds of carotenoids formed during heating of a simulated cashew apple juice*. Food Chemistry, 2009. **117**(1): p. 28-34.
13. Pereira, A.L.F., et al., *Storage stability and acceptance of probiotic beverage from cashew apple juice*. Food and Bioprocess Technology, 2013. **6**(11): p. 3155-3165.

14. Assunção, R.B. and A.Z. Mercadante, *Carotenoids and ascorbic acid composition from commercial products of cashew apple (Anacardium occidentale L.)*. Journal of Food Composition and Analysis, 2003. **16**(6): p. 647-657.
15. Beebe, K.R., R.J. Pell, and M.B. Seasholtz, *Chemometrics: a practical guide*. Vol. 4. 1998: Wiley-Interscience.
16. Hotelling, H., *Analysis of a complex of statistical variables into principal components*. Journal of educational psychology, 1933. **24**(6): p. 417-441.
17. Silva, L.M.A., et al., *¹H HRMAS NMR spectroscopy and chemometrics for evaluation of metabolic changes in citrus sinensis Caused by Xanthomonas axonopodis pv. citri*. Journal of the Brazilian Chemical Society, 2012. **23**: p. 1054-1061.
18. Costa, D.P., et al., *Influence of fruit biotypes on the chemical composition and antifungal activity of the essential oils of eugenia uniflora leaves*. Journal of the Brazilian Chemical Society, 2010. **21**: p. 851-858.
19. Wold, S., M. Sjöström, and L. Eriksson, *PLS-regression: a basic tool of chemometrics*. Chemometrics and intelligent laboratory systems, 2001. **58**(2): p. 109-130.
20. Nielsen, N.-P.V., J.M. Carstensen, and J. Smedsgaard, *Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping*. Journal of Chromatography A, 1998. **805**(1): p. 17-35.
21. Alves Filho, E.G., et al., *¹H qNMR and Chemometric Analyses of Urban Wastewater*. Journal of the Brazilian Chemical Society, 2015. **26**(6): p. 1257-1264.
22. Larsen, F.H., F. van den Berg, and S.B. Engelsen, *An exploratory chemometric study of ¹H NMR spectra of table wines*. Journal of Chemometrics, 2006. **20**(5): p. 198-208.
23. Rinnan, Å., F.V.D. Berg, and S.B. Engelsen, *Review of the most common pre-processing techniques for near-infrared spectra*. TrAC Trends in Analytical Chemistry, 2009. **28**(10): p. 1201-1222.
24. Cunha, A.G., et al., *UPLC–qTOF-MS/MS-based phenolic profile and their biosynthetic enzyme activity used to discriminate between cashew apple (Anacardium occidentale L.) maturation stages*. Journal of Chromatography B, 2017. **1051**: p. 24-32.
25. Bernillon, S., et al., *Metabolomic and elemental profiling of melon fruit quality as affected by genotype and environment*. Metabolomics, 2013. **9**(1): p. 57-77.
26. Okpanachi, U., S. Attah, and D. Shaahu, *A Comparative study between vitamins and amino acid profile of sun-dried red and yellow cashew pulp*. Int. J. An. Biol, 2015. **1**(5): p. 23.

27. Arumugam, B., et al., *Potential antihyperglycaemic effect of myricetin derivatives from Syzygium malaccense*. *Journal of Functional Foods*, 2016. **22**: p. 325-336.

**5 CAPÍTULO 2 - PORTABLE MICRO NEAR INFRARED SPECTROSCOPY
COUPLED TO CHEMOMETRIC ANALYSIS IN ASSESSING THE COMPOSITION
VARIABILITY OF DIFFERENT GENOTYPES OF CASHEW APPLE**

Ynayara C. de Lima^{a*}; Elenilson G. Alves Filho^a; Gislane Mendes de Morais^a;
Ebenézer O. Silva^a; Ana Cecília R. Castro^a; Edy S. de Brito^a

^a Embrapa Agroindústria Tropical, Fortaleza-CE, Brazil.

* Corresponding author

E-mail: edy.brito@embrapa.br

Phone: +55 85 33917393

Fax: +55 85 33917100

5.1 ABSTRACT

Embrapa (a Brazilian agricultural research corporation) is developing a genetic improvement program in the attempt to enhance the quality of cashew apple. Therefore, the aim of the present study was to evaluate the potential of the micro NIR spectrometer to explore the cashew apple composition from different genotypes and develop classification models considering the concentrations of ascorbic acid (vitamin C), °Brix, and acidity. It was demonstrated that NIR spectroscopy provided a non-destructive, low-cost, and fast results of the cashew apple composition. Important composition tendencies were observed with main four clustering according to the genotype, even considering the different fruits morphology (shape, size, and color). The unexpected weak classification performances of the models using the concentrations of ascorbic acid (vitamin C), °Brix value, and acidity show that further parameters related to the fruit composition and environmental factors on the instrument performance must be taken into account for a better comprehension of the chemical and physical variability of the cashew apple from the germplasm bank of Embrapa.

Keywords: cashew apple; Micro-NIR; chemometrics

5.2 INTRODUCTION

The fruit consumption has increased significantly over the years with the improvement of living standards. Meanwhile, consumers are looking for fruits with higher quality related to the taste, flavor and, consequently, the organic composition as vitamins and sugars [1]. In this regard, the increasing success of cashew production is mainly related to the genetic improvement of cultivars in order to develop varieties adapted to the general edaphoclimatic characteristics and cultivation conditions of each geographic region [2, 3]. Additionally, new cashew cultivars circumvent the issues of pathogen attacks that may limit their production. Since the phenotypic variation of a crop species may occur due to the environment and genetics, Embrapa (a Brazilian agricultural research corporation) is developing genetic improvement programs, which are the focus for improvement of food. Therefore, in the attempt to enhance the quality of cashew apple, a substantial attention has been dedicated to the study of physicochemical and nutritional attributes.

In general, researchers have investigated the potential of various technologies to develop a simple, precise, and low-cost method to study the physicochemical and nutritional features of fruits. Among all these technologies, the near infrared (NIR) spectroscopy is considered an economical, fast, and efficient technique for the evaluation of foodstuffs [4-6]. The NIR analysis produces an electromagnetic spectrum (reflectance or transmittance) between 780 and 2,500 nm, where the wavelength depends of scattering and absorption processes according to the chemical composition (molecular functional groups, C-H, N-H, S-H or O-H bonds). The micro NIR spectrometer is a small and portable device for real-time, non-intrusive chemical and physical analysis [7]. However, due to elevated number of samples and variations that should be taken into account in fruit evaluation and the difficulty of spectral interpretation, exploratory and classificatory multivariate analysis are necessary to extract the required information from the total matrix [2, 8, 9].

From the foregoing, the aim of the present study was to evaluate the potential of the portable NIR spectrometer to explore the cashew apple composition from different genotypes from Embrapa germplasm bank and develop classification models considering the concentrations of ascorbic acid (vitamin C), °Brix, and acidity.

5.3 EXPERIMENTAL

5.3.1 Sampling

Seven hundred and sixty four cashew apple (*Anacardium occidentale*, L.) from different accessions were randomly harvested at Embrapa Experimental Station (Pacajus, Ceará, Brazil) – coordinates 4°11'07"S; 38°30'07"W, altitude 70 m.

5.3.2 Portable NIR analysis

The NIR experiments were performed on portable NIR spectrometer (Micro-NIR 1700, Viavi, Milpitas, CA, USA) operated in the 1150-2170 nm range, spectral resolution of 10 nm, dimensions of 45 mm diameter × 42 mm high, two tungsten radiation sources to illuminate samples for reflectance measurements, continuous monochromator based on a linear variable filter. The parameters for spectral data acquisition were set at 50 ms integration time and averaging of 100 scans for each sample. The reference spectra for absorbance calculation were obtained from a piece of Spectralon™, and the dark signal was obtained by pointing the measurement window of the instrument to the ambient environment.

5.3.3 Quantification of ascorbic acid, °Brix, and determination of the acidity

Previously the chemometric evaluation, the concentrations of ascorbic acid (vitamin C), the °Brix value, and the acidity of each cashew apple were determined. The propose of these results was to create categorical variables (Y column) to develop the regression analysis using Partial Least Square (PLS), to find the latent variables (LVs) in the X matrix (samples data) to prediction models maximizing the covariance between X and Y.

The total titratable acidity (TTA) was performed by titration with NaOH solution (0.1 N) in a sample of 1 g of the pulp diluted in approximately 50 mL of distilled water containing 3 drops of 1 % phenolphthalein until it reached pink coloration. The results were expressed as percentage of malic acid according to IAL (1985) [10].

Soluble solids (SS) content was determined by refractometry as described by AOAC (2005) using a digital refractometer (ATAGO™ N1, Kirkland-USA) with automatic temperature compensation [11]. The results were expressed in °Brix (concentration of sucrose w/w).

The total vitamin C was determined by titration with 0.02% 2,6-dichloro-indophenol (DFI) as method by Strohecker and Henning (1967) [12]. 1 g of pulp was diluted to 100 mL of 0.5% oxalic acid and homogenized. Then, 5 mL of this solution was diluted to 50 mL with distilled water and titrated and results were expressed as mg.100 g⁻¹ FW (fresh weight).

5.3.4 Chemometric analysis

Chemometric analyses were performed in the numerical matrix from 764 cashew apple fruits harvested in different years. After that, the averaging method was applied on those cashew apples from the same accession (same tree), which result in 135 spectral dataset. The spectral region between 1150 and 2170 nm was used for the analysis, resulting in a numerical matrix with dimensionality of 16,875 data points (135 spectra × 125 variables). The samples were named according to the accession numbers and year of harvest.

The spectra data were converted to American Standard Code for Information Interchange (ASCII) files to matrix construction through Origin™ 9.4 program, which was exported for chemometric analysis using Hierarchical Clustering algorithm for heat map evaluation. To reduce the dimensionality of the original data and to assist the modeling and interpretation of multivariate data, the numerical matrix was averaged along variables by a factor of 2 using The Unscrambler X™ program (version 10.4), and imported into GENE-E program (software.broadinstitute.org/GENE-E) for pattern recognition and the classification was performed on the columns (samples). The Euclidean distance method was used to measure the proximity between samples and the average linkage method (sum-of-squares approach in calculating intercluster distances) was applied. The results were presented as 3-D dendrogram (heat map) form where the deeper red color represents the higher relative intensity on the line (wavelength), the deeper blue color the lower relative intensity, and the intermediary intensity in white color.

Besides the unsupervised analysis, to improve the identification of chemical changes due to the different cashew apple genotypes, the supervised PLS (Partial Least Squares) method was employed using the concentrations of ascorbic acid, °Brix values, and acidity of the cashew apple previously calculated (section 2.3) as categorical variables (Y column). The NIPALS (Nonlinear Iterative Partial Least Squares) algorithm was applied to build the models and the latent variables were selected in accordance to statistical parameters: SEC (Square Error of Calibration); SEV (Square Error of Validation), and correlation coefficient (R^2). The full cross-validation method was applied to evaluate the performance of the PLS models: each samples of cashew apple was left out from the calibration data set and the sub-models were calibrated on the remaining data points.

5.4 RESULTS AND DISCUSSION

5.4.1 Exploratory evaluation

The NIR spectroscopy provides non-destructive, low-cost, non-invasive, and fast results of the samples composition (any molecule containing C-H, N-H, S-H or O-H bonds), but information from a comparison of many samples can be partly obscured by spectral imperfections. If the distribution of the variables is highly skewed, basic pretreatments of the original data are essential to assess important feature regarding the composition, minimizing or eliminating not relevant variations within spectra. Clearly, the NIR analyses are highly undesirable source of variation for multivariate studies and pretreatments can correct many systematic or random errors. Therefore, transformations of the compositional data should be applied correctly to avoid false correlations among the variables, since the modifications are directly applied in raw data. It is know that there are imperfections intrinsic from the NIR evaluation and, consequently, these are often limiting factor for determining the amount and quality of information obtainable from the experiments. In this regarding, previously the chemometric analysis the main pretreatments of the numerical matrix from NIR data analysis were tested reaching the increase of the information quality, which are illustrated in Fig. 1: MSC (Multiplicative Scatter Correction), SNV (Standard

Normal Variate), second derivative using Savitzky-Golay filter with a second order polynomial for five points.

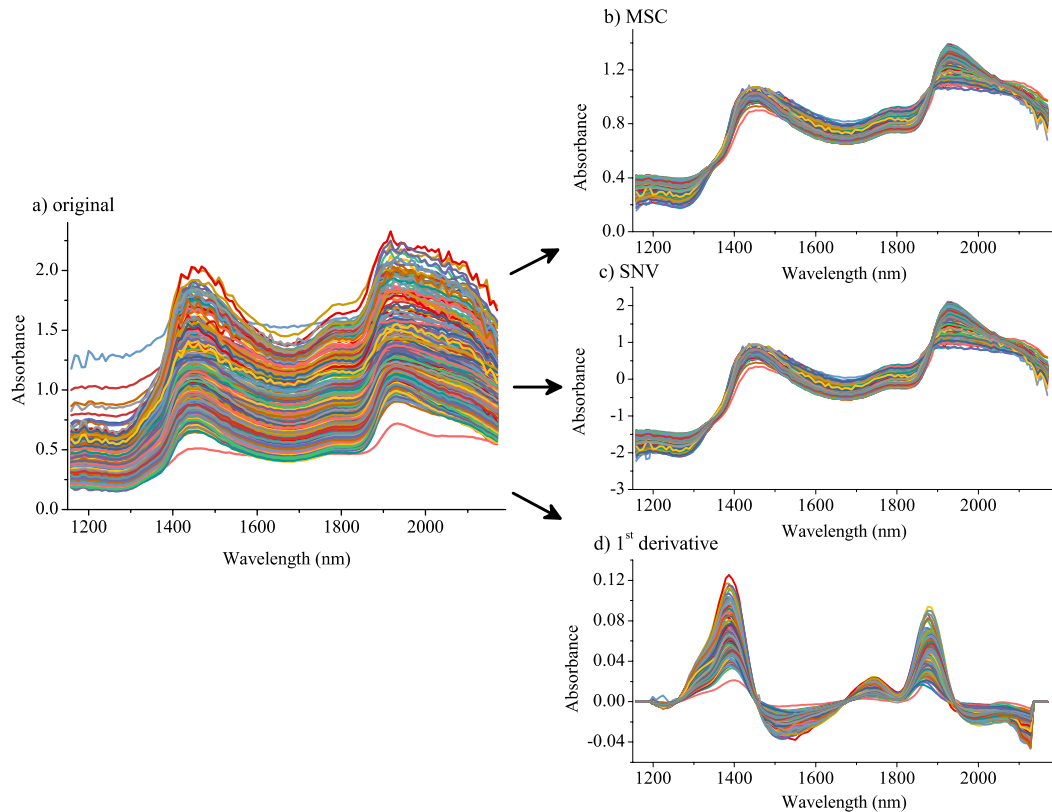


Figure 1 - a) Raw absorbance spectra from 135 cashew apple obtained using the portable NIR spectrometer, and the same spectra after the follow treatment: b) MSC; c) SNV; and d) second derivative using Savitzky-Golay filter with a second order polynomial for five points.

The MSC and SNV algorithms often present similar results (as shown in Fig. 1b and 1c) and are regarded as exchangeable. After the application of the MSC algorithm, the spectra are adjusted to present the same scatter level estimated by a mean spectrum. On the other hand, the SNV algorithm treats each spectrum separately by autoscaling the absorbance values [13]. The derivative algorithm emphasizes steep peaks and improves the peaks overlapping as well as reduces the measurement variations, which improve the differentiation of the bands. The first and second derivatives are more commonly applied, and the second derivative NIR spectrum results in sharp signals [14-17].

In order to comprehend the variability of the organic compounds in cashew apple according to their genotype and detect tendencies related to the cashew apple composition, a non-targeted multivariate analysis was developed on the numerical matrix from the spectral data. The unsupervised method based on Hierarchical

Clustering was applied to segregate the samples in groups according to their similarity. Figure 2 presents the 3-D dendrogram with the samples in columns, wavelength in rows, and the signals intensities are illustrated in color: deeper red color represents the higher relative intensity on the line (wavelength), the deeper blue color the lower relative intensity, and the intermediary intensity in white color. Important tendencies of main four clustering were observed in the dendrogram regarding the genotype of the cashew apple even considering the different fruits morphology (shape, size, and color). The results reflected the natural differences among the cashew apple groups, which were formed as a function of fruit composition dependent of the genotype. The cluster 4 presented the most distant samples with opposite behavior of their composition compared to other groups.

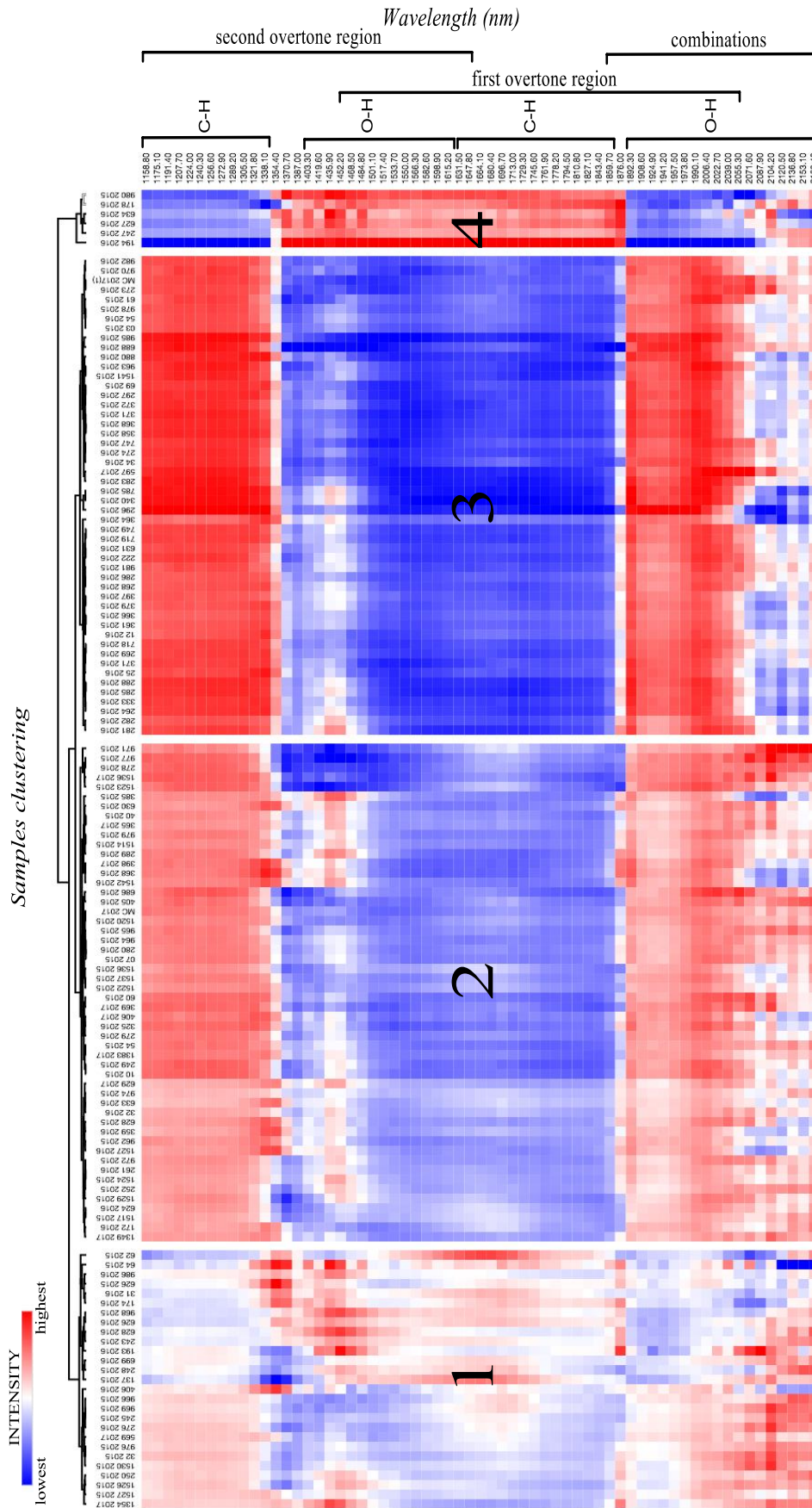


Figure 2 - 3-D dendrogram (sample x wavelength x intensity) representing chemical composition similarity relationships among the samples.

The main absorbance ranges related to the samples clustering are located between 1150 and 1340 nm, 1370 and 1850 nm, and 1900 and 2020 nm. The absorbance range between 2040 and 2170 nm related to combinations of functional groups was not relevant. The absorbance ranges between 1400-1620 nm and 1900-2020 nm are related to OH stretch second and first overtone respectively, and consequently, they may be related to water. The absorbance ranges between 1150-1340 nm and 1650-1850 nm are due to CH stretches related to the second and first overtone respectively, may be from carbohydrates and other organic species present in cashew skin and pulp. The band at 1340 nm is probably due to CH from cellulose content [18].

5.4.2 Classification analysis

Previously the Partial Least Squares (PLS) analysis, the concentrations of ascorbic acid (vitamin C), °Brix value, and acidity were determined in 40 cashew apple samples randomly chosen from the set of samples in order to create categorical variables (Y column) to regression analysis. Based on the non-targeted chemometric evaluation (Fig. 2) and considering the quantitative results presented in Table 1 (reference values), three classification models were created to each dependent variable (ascorbic acid, °Brix, and acidity) to improve the association of the chemical variability according to the cashew apple genotypes. The statistical parameters used to assess the quality of the calibration method (Table 2) indicated weak-adjusted models using 7 LVs for all the classifications, despite of the relative proximity (similarity) between the reference (real) and predicted values. In addition, despite of the proximity between the SEC (Standard Error of Calibration) and SEV (Standard Error of Cross Validation) values (SEC/SEV ratio above 0.50 indicates a well-adjusted model) [19-21], the high SEC and SEV values and the low correlation coefficients (R^2) expresses the rather weak classification performance of all the models estimated as a function of the global model error, samples leverage, and the sample residual X-variance. Through the linear regression analysis, it was possible to verify that the three models presented weak relation with the quality parameters of ascorbic acid ($R^2 = 0.46$), °Brix ($R^2 = 0.49$), and acidity ($R^2 = 0.32$), being these parameters normally used to define the points of harvest of cashew apple.

Therefore, the weak classification performances show that further parameters related to the fruit composition and environmental factors on the instrument performance, such as fruits morphology (shape, size, and color), nonhomogeneous distribution of particles in a samples, samples density variations, must be taken into account for improve the comprehension of the chemical and physical variability of the cashew apples from the germplasm bank of Embrapa. These small physical variations from sample to sample may lead to light scattering that influence the measured NIR spectra, which result in baseline shifts and scaling variations (intensity variations) and, consequently, disturbing future classifications and predictions evaluation.

Table 1 - Concentrations of ascorbic acid (in mg.100 g⁻¹ of cashew apple), °Brix values, and acidity in 40 randomly selected samples (real or reference values), and those predicted values by each regression model.

<i>Samples</i>	<i>ascorbic reference</i>	<i>pred. value</i>	<i>°Brix reference</i>	<i>pred. value</i>	<i>acidity reference</i>	<i>pred. value</i>
03 2015	380.66	304.90	13.70	12.96	1.01	0.59
12 2016	278.85	259.50	10.90	13.48	0.50	0.56
137 2015	212.69	252.63	16.65	13.94	0.93	0.81
1517 2015	200.36	189.86	11.70	13.99	0.37	0.45
1523 2015	287.76	279.06	10.90	14.12	0.41	0.53
1529 2015	203.82	252.66	13.15	13.49	0.43	0.56
1536 2015	530.68	275.92	17.30	13.92	0.68	0.60
1541 2015	192.18	244.20	11.90	12.50	0.76	0.43
222 2016	306.20	268.90	13.20	13.77	1.22	0.72
243 2015	225.28	215.17	14.90	14.79	0.83	0.53
248 2015	200.40	260.74	16.10	16.86	0.52	0.64
249 2015	235.96	248.94	15.10	14.24	0.59	0.46
264 2016	168.38	177.69	13.40	13.85	0.76	0.45
268 2016	180.94	200.85	12.30	13.44	0.24	0.49
276 2016	334.71	384.03	14.97	16.22	0.97	0.83
279 2016	301.77	269.55	14.30	15.10	0.42	0.58
281 2016	190.47	264.37	14.60	14.58	0.28	0.61
285 2016	201.60	259.67	12.90	13.12	0.41	0.58

286 2016	350.51	307.28	14.15	13.83	0.52	0.55
288 2016	275.69	251.35	11.30	13.19	0.42	0.56
289 2016	262.82	288.04	13.50	13.37	0.44	0.60
340 2015	338.39	257.27	10.90	11.30	0.46	0.51
358 2015	122.85	251.88	13.30	12.06	0.38	0.57
359 2016	216.59	111.27	17.10	17.02	0.96	0.63
366 2015	219.02	241.66	12.00	11.94	0.41	0.57
61 2015	278.59	280.40	15.05	13.34	0.53	0.58
62 2015	198.61	249.03	14.60	16.23	0.55	0.86
628 2016	193.84	202.45	17.50	17.61	0.31	0.28
686 2016	129.63	215.52	12.20	13.43	0.33	0.43
688 2016	329.12	265.61	15.65	16.25	0.46	0.33
962 2015	359.71	321.48	14.35	13.99	0.76	0.57
963 2015	134.47	157.68	11.60	12.87	0.26	0.35
971 2015	383.25	434.32	19.23	17.61	0.70	0.70
976 2015	206.84	242.33	13.85	14.22	0.43	0.57
978 2015	258.32	294.64	15.50	14.09	0.58	0.62
979 2015	212.43	231.57	14.85	13.49	0.35	0.53
981 2015	214.87	156.72	12.00	13.91	0.40	0.41
982 2016	174.16	215.07	12.50	13.72	0.25	0.38
MC 2017.1	283.36	235.59	14.90	14.54	0.30	0.32
MC 2017.2	428.45	380.76	14.45	15.65	0.71	0.61

Table 2 - Parameters obtained by multivariate calibration from each PLS model.

Model	7 LVs^a	r^2 cal^b	SEC^c	SEV^d	SEC/SEV^e
<i>ascorbic</i>	45.7 %	0.46	63.90	107.6	0.594
<i>°Brix</i>	49.2 %	0.49	1.44	2.30	0.626
<i>acidity</i>	32.3 %	0.32	0.19	0.37	0.513

^a Total variance percent in X matrix refer to seven Latent Variables (LVs).

^b Coefficient of correlation between the real concentration and the concentration predicted during the calibration.

^c Standard Error of Calibration.

^d Standard Error of Cross Validation.

^e Similarity criterion.

5.5 CONCLUSIONS

It was demonstrated that portable NIR spectroscopy provided a non-destructive, low-cost, and fast results of the cashew apple composition. Important composition tendencies were observed with four samples clustering according to the composition similarity and genotype, even considering the different fruits morphology (shape, size, and color). The weak classification performances of the models using the concentrations of ascorbic acid (vitamin C), °Brix value, and acidity show that further parameters related to the fruit composition and environmental factors on the instrument performance must be taken into account for a better comprehension of the chemical and physical variability of the cashew apples from the germplasm bank of Embrapa.

5.6 ACKNOWLEDGEMENTS

The authors thank CNPq and FUNCAP for the financial support and the award of scholarship (303791/2016-0).

5.7 REFERENCES

1. Alves Filho, E.G., et al., *Genotype evaluation of cowpea seeds (Vigna unguiculata) using ^1H qNMR combined with exploratory tools and solid-state NMR*. Food Research International, 2017. **91**: p. 140-147.
2. Ribeiro, L.P.D., et al., *Non-destructive determination of quality traits of cashew apples (Anacardium occidentale, L.) using a portable near infrared spectrophotometer*. Journal of Near Infrared Spectroscopy, 2016. **24**(1): p. 77-82.
3. Samamad, N.T.I., *Cashew apple quality by near infrared spectroscopy technique*. 2016.
4. Mathlouthi, M., *Water content, water activity, water structure and the stability of foodstuffs*. Food control, 2001. **12**(7): p. 409-417.
5. Nicolai, B.M., et al., *Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review*. Postharvest biology and technology, 2007. **46**(2): p. 99-118.
6. Connolly, C., *NIR spectroscopy for foodstuff monitoring*. Sensor Review, 2005. **25**(3): p. 192-194.
7. Dos Santos, C.A.T., et al., *A review on the applications of portable near-infrared spectrometers in the agro-food industry*. Applied spectroscopy, 2013. **67**(11): p. 1215-1233.
8. Alishahi, A., et al., *Identification of transgenic foods using NIR spectroscopy: a review*. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2010. **75**(1): p. 1-7.
9. Galtier, O., et al., *Geographic origins and compositions of virgin olive oils determined by chemometric analysis of NIR spectra*. Analytica chimica acta, 2007. **595**(1): p. 136-144.
10. LUTZ, I.A., *Normas analíticas, métodos químicos e físicos para análise de alimentos*. São Paulo: Instituto Adolfo Lutz, 1985. **1**.
11. Horwitz, W., P. Chichilo, and H. Reynolds, *Official methods of analysis of the Association of Official Analytical Chemists*. Official methods of analysis of the Association of Official Analytical Chemists., 1970: p. 1115.
12. Strohecker, R. and H.M. Henning, *Análisis de vitaminas, métodos comprobados*. 1967.
13. Varmuza, K. and P. Filzmoser, *Introduction to multivariate statistical analysis in chemometrics*. 2016: CRC press.
14. Press, W.H., et al., *Numerical recipes 3rd edition: The art of scientific computing*. Vol. 3. 2007: Cambridge university press Cambridge.

15. Gorry, P.A., *General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method*. Analytical Chemistry, 1990. **62**(6): p. 570-573.
16. Ni, Z. and F. FENG, *Progress and effect of spectral data pretreatment in NIR analytical technique*. Chinese Journal of Pharmaceutical Analysis, 2008. **28**(5): p. 824-829.
17. Lu, Y., et al., *Research on the method of choosing optimum wavelengths combination by using multiple scattering correction technique*. Guang pu xue yu guang pu fen xi= Guang pu, 2007. **27**(1): p. 58-61.
18. Shenk, J.S., J.J. Workman, and M.O. Westerhaus, *Application of NIR spectroscopy to agricultural products*. Practical Spectroscopy Series, 2001. **27**: p. 419-474.
19. Beebe, K.R., R.J. Pell, and M.B. Seasholtz, *Chemometrics: a practical guide*. Vol. 4. 1998: Wiley-Interscience.
20. Wold, S., M. Sjöström, and L. Eriksson, *PLS-regression: a basic tool of chemometrics*. Chemometrics and intelligent laboratory systems, 2001. **58**(2): p. 109-130.
21. Ferrão, M.F., et al., *Simultaneous determination of ash content and protein in wheat flour using infrared reflection techniques and partial least-squares regression (PLS)*. Food Science and Technology (Campinas), 2004. **24**(3): p. 333-340.

6 CONSIDERAÇÕES FINAIS

A hifenação das técnicas LC e MS foi eficiente na análise dos genótipos de caju, sendo possível, por meio de ferramentas quimiométricas, separar grupamentos por afinidade na composição química.

A identificação das moléculas responsáveis pela separação dos grupos foi valiosa, pois adicionou informações sobre o conhecimento do metaboloma do caju. Essa identificação é importante pois pode ajudar o Programa de Melhoramento Genético a selecionar genótipos que atendam a necessidades nutricionais dos consumidores para serem lançados no mercado.

Os parâmetros vitamina C, sólidos solúveis e acidez total não foram capazes de explicar a separação dos grupamentos dos genótipos. Este resultado sugere que a novas pesquisas na área precisam ser feitas a fim de determinar os compostos responsáveis pela variabilidade encontrada nos genótipos do Banco de Germoplasma da Embrapa.