

PAULA CAROLINA PIRES BUENO

**Variabilidade química infraespecífica e dinâmica de metabólitos secundários
fenólicos e diterpênicos em *Casearia sylvestris* Sw. (Salicaceae): correlação
metabolômica e proteômica utilizando cromatografia, espectrometria de massas,
ressonância magnética nuclear e quimiometria**

Tese de doutoramento apresentada ao Instituto
de Química, Universidade Estadual Paulista
como parte dos requisitos para a obtenção do
título de DOUTORA EM QUÍMICA.

Orientador: Prof. Dr. Alberto José Cavalheiro

Araraquara

2015

FICHA CATALOGRÁFICA

B928v Bueno, Paula Carolina Pires
Variabilidade química infraespecífica e dinâmica de metabólitos secundários fenólicos e diterpênicos em *Casearia sylvestris* Sw. (Salicaceae): correlação metabolômica e proteômica utilizando cromatografia, espectrometria de massas, ressonância magnética nuclear e quimiometria / Paula Carolina Pires Bueno – Araraquara : [s.n], 2015
255 f. : il.

Tese (doutorado) – Universidade Estadual Paulista, Instituto de Química
Orientador: Alberto José Cavalheiro

1. Produtos naturais. 2. Metabolômica. 3. Proteômica.
4. Cromatografia líquida de alta eficiência. 5. Metabólitos.
I. Título.

Elaboração: Diretoria Técnica de Biblioteca e Documentação do Instituto de Química
Seção Técnica de Aquisição e Tratamento da Informação

PAULA CAROLINA PIRES BUENO

Tese de Doutorado apresentada ao
Instituto de Química, Universidade
Estadual Paulista, como parte dos
requisitos para obtenção do título de
Doutora em Química.

Araraquara, 18 de setembro de 2015.

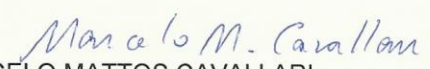
BANCA EXAMINADORA

Prof. Dr. ALBERTO JOSÉ CAVALHEIRO (Orientador)
Instituto de Química / UNESP / Araraquara – SP.


Profª. Drª. MAYSA FURLAN
Instituto de Química / UNESP / Araraquara – SP.


Prof. Dr. LEONARDO GOBBO NETO
Faculdade Ciências Farmacêuticas / USP / Ribeirão Preto – SP.


Profª. Drª. MARINA FRANCO MAGGI TAVARES
Instituto de Química / USP / São Paulo – SP.


Dr. MARCELO MATTOS CAVALLARI
Empresa Brasileira de Pesquisa Agropecuária / EMBRAPA / São Luís – MA.

DADOS CURRICULARES

Paula Carolina Pires Bueno

Departamento de Química Orgânica, Instituto de Química de Araraquara - UNESP
Endereço: Rua Dr. Francisco Degni 55, CEP 14800-900, Araraquara/SP, Brasil
E-mail: paulabueno@yahoo.com

Formação Acadêmica.....

2011 – 2015 **Doutorado em Química**

- Área de Concentração: Química Orgânica
- Instituto de Química de Araraquara (IQ-UNESP)
- Projeto: Variabilidade química infra-específica e dinâmica de metabólitos secundários fenólicos e diterpênicos em *Casearia sylvestris* Sw. (Salicaceae) provenientes de diferentes ecossistemas brasileiros: correlação metabolômica e proteômica
- Agência financiadora: FAPESP (2011/21440-3)
- Supervisor: Prof. Dr. Alberto José Cavalheiro (albjcava@iq.unesp.br)

2012 – 2013 **Estágio de Doutorado BEPE (FAPESP)**

- Institut für Biologie und Biotechnologie der Pflanzen – Westfälische Wilhelms, Universität Münster, Münster – Germany
- Projeto: Differential proteomics of two varieties of *Casearia sylvestris* Sw. (Salicaceae)
- Agência financiadora: BEPE/FAPESP (2012/21921-4)
- Supervisor: Prof. Dr. Michael Hippler (mhippler@uni-muenster.de)

2005 – 2007 **Mestrado em Ciências Farmacêuticas**

- Área de Concentração: Produtos Naturais e Sintéticos
- Instituição: Faculdade de Ciências Farmacêuticas de Ribeirão Preto (FCFRP-USP)
- Dissertação: Desenvolvimento e validação analítica por cromatografia gasosa para o controle de qualidade de *Eucalyptus globulus* e seus produtos: planta, extratos, óleo essencial e xarope de eucalipto
- Supervisor: Prof. Dr. Jairo Kenupp Bastos (jkbastos@fcrfp.usp.br)

2005 – 2005 **Especialização em Ciências Farmacêuticas**

- Programa de Atualização em Ciências Farmacêuticas
- Área de Concentração: Produtos Naturais e Sintéticos
- Instituição: Faculdade de Ciências Farmacêuticas de Ribeirão Preto (FCFRP-USP)
- Projeto: Quantificação de cumarina em *Milkenia glomerata* planta seca e extrato de guaco
- Supervisor: Prof. Dr. Jairo Kenupp Bastos

2000 – 2004 **Iniciação Científica**

- Instituição: Faculdade de Ciências Farmacêuticas de Ribeirão Preto
- Projeto: Estudo fitoquímico e da atividade tripanocida dos metabólitos secundários das raízes de *Baccharis dracunculifolia* Dc (Asteraceae).
- Agência financiadora: FAPESP (2002/03493-3)
- Supervisor: Prof. Dr. Jairo Kenupp Bastos

2000 – 2004 **Graduação em Ciências Farmacêuticas**

- Instituição: Faculdade de Ciências Farmacêuticas de Ribeirão Preto (FCFRP-USP)

Experiências Profissionais e Didáticas.....

2010 – 2011 **SENAC – Serviço Nacional de Aprendizagem Comercial**

- Docente do Curso Técnico em Farmácia

2007 – 2007 **Programa de Aperfeiçoamento de Ensino - PAE**

- Bolsista PAE na disciplina de Princípios Ativos Naturais - FCFRP-USP

2005 – 2011 **Apis Flora Indl. Com. Ltda**

- Supervisora Controle de Qualidade, Desenvolvimento e Validação de Métodos Analíticos

Orientações e Supervisões Científicas.....

2013 – 2015 **Supervisão Científica**

- Supervisão científica de Livia L.Thomaz (FAPESP # 2012/14111-6)
- Supervisão científica Naira B. Anhesine (FAPESP # 2014/02738-0)

2006 – 2011 **Orientação de Alunos em Estágio Curricular**

Supervisão e orientação de alunos de graduação em estágio curricular na empresa Apis Flora Indl. Com. Ltda.

Títulos e Prêmios.....

2015 **SciFinder Future Leaders in Chemistry**

- Distinção outorgada pela American Chemical Society. Columbus/EUA, de 10/08/2015 a 15/08/2015.

2014 **Um dos melhores quatro trabalhos apresentados no XV Colacro**

- Título do trabalho: Intraspecific chemical variability of *Casearia sylvestris* Sw. (Salicaceae) based on UHPLC-DAD combined with chemometrics tools. XV COLACRO. Cartagena/Colombia, de 29/10/2014 a 03/10/2014.

Artigos Publicados.....

Bueno, PCP; Pereira, FMV; Torres, RB; Cavalheiro, AJ. Development of a comprehensive method for analysing clerodane-type diterpenes and phenolic compounds from *Casearia sylvestris* Swartz (Salicaceae) based on ultra-high performance liquid chromatography combined with chemometric tools. Journal of Separation Science, v. 38, p. 1649-1656, 2015.

Souza, MC; Bueno, PCP; Morellato, LPC; Habermann, G. Ecological strategies of Al-accumulating and non-accumulating functional groups from the cerrado sensu stricto. Anais da Academia Brasileira de Ciências. Anais da Academia Brasileira de Ciências, v. 7 (2), p. 813-823, 2015.

Bueno, PCP; Groppo Junior, M; Bastos, JK. A GC-FID validated method for the quality control of *Eucalyptus globulus* raw material and its pharmaceutical products, and GC-MS fingerprinting of 12 *Eucalyptus* species. *Natural Product Communications*, v. 9, p. 1787-1790, 2014.

Braga, FTMM; Santos, ACF; Bueno, PCP; Silveira, RCCP; Santos, CB; Bastos, JK; Carvalho, EC. Use of *Chamomilla recutita* in the prevention and treatment of oral mucositis in patients undergoing hematopoietic stem cell transplantation. *Cancer Nursing*, v. 38 (4), p. 322-329, 2014.

Rocha, BA; Bueno, PCP; Vaz, MMOLL; Nascimento, AP; Ferreira, NU; Moreno, GP; Rodrigues, MR; Costa-Machado, ARM; Barizon, EA; Campos, JCL; De Oliveira, PF; Acésio, NO; Martins, SPL; Tavares, DC; Berretta, AA. Evaluation of a propolis water extract using a reliable RP-HPLC methodology and in vitro and in vivo efficacy and safety characterisation. *Evidence-Based Complementary and Alternative Medicine*, v. 2013, p. 1-11, 2013.

Figueiredo-Rinhel, AGSG; Kabeya, LM; Bueno, PCP; Jorge-Tiossi, RF; Azzolini, AECS; Bastos, JK; Lucisano-Valim, YM. Inhibition of the human neutrophil oxidative metabolism by *Baccharis dracunculifolia* DC (Asteraceae) is influenced by seasonality and the ratio of caffeic acid to other phenolic compounds. *Journal of Ethnopharmacology*, v. 150, p. 655-664, 2013.

Guimarães, NSS; Mello, JC; Paiva, JS; Bueno, PCP; Berretta, AA; Torquato, RJ; Nantes, IL; Rodrigues, T. *Baccharis dracunculifolia*, the main source of green propolis, exhibits potent antioxidant activity and prevents oxidative mitochondrial damage. *Food and Chemical Toxicology*, v. 50, p. 1091-1097, 2012.

Berretta, AA; Nascimento, AP; Bueno, PCP; Vaz, MMOLL; Marchetti, JM. Propolis standardized extract (EPP-AF), an innovative chemically and biologically reproducible pharmaceutical compound for treating wounds. *International Journal of Biological Sciences*, v. 8, p. 512-521, 2012.

Rocha, BA; Rodrigues, MR; Bueno, PCP; Vaz, MMOLL; Costa, ARM; Barud, H; Berretta, AA. Preparation and thermal characterization of inclusion complex of Brazilian green propolis and hidroxypropyl-B-cyclodextrin. *Journal of Thermal Analysis and Calorimetry*, v. 108 (1), p. 87-94, 2012.

Reis, PE; Carvalho, EC; Bueno, PCP; Bastos, J K. Aplicação clínica da *Chamomilla recutita* em flebites: estudo de curva dose-resposta. *Revista Latino-Americana de Enfermagem*, v. 19 (1), p. 3-10, 2011.

Costa, ARM; Marquiasfável, FS; Vaz, MMOLL; Rocha, BA; Bueno, PCP; Amaral, PLM; Barud, H; Berreta-Silva, AA. Quercetin-PVP K25 solid dispersions. *Journal of Thermal Analysis and Calorimetry*, v. 104 (1), p. 273-278, 2011.

Sousa, JPB; Bueno, PCP; Gregório, LE; Silva-Filho, AA; Furtado, NAJC; Jorge, R; Bastos, JK. A validated reverse phase HPLC analytical method for the quantification of phenolic compounds in *Baccharis dracunculifolia*. *Phytochemical Analysis*, v. 20, p. 24-32, 2009.

Bueno, PCP; Bastos, JK. A validated capillary gas chromatography method for guaco (*Mikania glomerata* S.) quality control and rastreability: from plant biomass to phytomedicines. *Revista Brasileira de Farmacognosia*, v. 19, p. 218-223, 2009.

Missima, F.; Silva-Filho, AA; Nunes, GA; Bueno, PCP; Sousa, JPB; Bastos, JK; Sforcin, JM. Effect of *Baccharis dracunculifolia* D.C. (Asteraceae) extracts and its isolated compounds on macrophage activation. Journal of Pharmacy and Pharmacology, v. 59, p. 01-06, 2007.

Sousa, JPB; Bueno, PCP; Gregório, LE; Silva-Filho, AA; Furtado, NAJC; Sousa, ML; Bastos, JK. A reliable quantitative method for the analysis of phenolic compounds in Brazilian propolis by reverse phase high performance liquid chromatography. Journal of Separation Science v. 30, p. 2656-2665, 2007.

Silva-Filho, AA; Bueno, PCP; Gregório, LE; Silva, MLA; Albuquerque, S; Bastos, JK. In vitro trypanocidal activity evaluation of crude extract and isolated compounds from *Baccharis dracunculifolia* D. C. (Asteraceae). Journal of Pharmacy and Pharmacology, England, v. 56, n. 9, p. 1195-1199, 2004.

Participação em Eventos Científicos.....

- | | |
|------|--|
| 2015 | Ouvinte: 250th American Chemical Society National Meeting Exposition. Boston/MA, EUA. |
| 2015 | Poster: 1st Brazilian Workshop on Bioinformatics/Chemometrics for Metabolomics. Ribeirão Preto/SP, Brasil. |
| 2015 | Palestra: Semana de Integração do Curso de Farmácia do Centro Universitário Barão de Mauá. Ribeirão Preto/SP, Brasil. |
| 2014 | Poster: 62nd International Congress & Annual Meeting of the Society for Medicinal Plant and Natural Product Research. Guimarães, Portugal. |
| 2014 | Apresentação oral: XV COLACRO - Congresso Latino-Americano de Cromatografia e Técnicas Relacionadas. Cartagena, Colômbia. |
| 2014 | Palestra: XIV Jornada de Integração Farmacêutica do Centro Universitário Barão de Mauá. Ribeirão Preto/SP, Brasil. |
| 2014 | Poster: VIII Reunião de Avaliação do programa BIOTA-FAPESP. São Pedro/SP, Brasil. |
| 2013 | Poster: V BrMass. Campinas/SP, Brasil. |
| 2013 | Ouvinte: São Paulo Advanced School on Bioorganic Chemistry. Araraquara/SP, Brasil. |
| 2012 | Ouvinte: Workshop ANVISA, Indústrias Farmacêuticas e Ministério da Saúde: Padrões de Referência e Metodologias Analíticas em Insumos Vegetais e Medicamentos Fitoterápicos. Brasília/DF, Brasil. |
| 2011 | Poster: 3rd Brazilian Conference on Natural Products - 3rd BCNP / XXXIX Annual Meeting on Micromolecular Evolution, Systematics and Ecology (RESEM). Ouro Preto/MG, Brasil. |
| 2011 | Ouvinte: Phytopharmactec o Workshop on Phytopharmaceutical Technology. Ribeirão Preto/SP, Brasil |
| 2011 | Poster: 8th Congress of Pharmaceutical Sciences (8th CIFARP). Ribeirão Preto/SP, Brasil. |
| 2011 | Ouvinte: Workshop Inovação no Complexo Industrial de Fitoterápicos. São Paulo/SP, Brasil |

- 2010 Ouvinte: Segundo Simpósio Internacional da Farmacopéia Brasileira / 4 ° Encontro anual da Farmacopéia Brasileira / Lançamento da Farmacopéia Brasileira 5 ed. Brasília/DF, Brasil.
- 2009 Ouvinte: 2nd Brazilian Conference on Natural Products - 2nd BCNP / XXXVIII Annual Meeting on Micromolecular Evolution, Systematics and Ecology (RESEM). São Pedro/SP, Brasil.
- 2008 Ouvinte: 1°. Simpósio Internacional de Revisão da Farmacopéia Brasileira. Brasília/DF, Brasil.
- 2008 Poster: XII Congresso Latino-Americano de Cromatografia e Técnicas Relacionadas. Florianópolis/SC, Brasil.
- 2007 Ouvinte: 1st Brazilian Conference on Natural Products - 1st BCNP / XXVII Annual Meeting on Micromolecular Evolution, Systematics and Ecology (RESEM) – São Pedro, SP, Brasil.
- 2007 Ouvinte: VII Jornada Paulista de Plantas Medicinais (VII JPPM) – São Paulo/SP, Brasil.
- 2006 Ouvinte: 2nd German Conference on Chemoinformatics – 20.CIC Workshop. Goslar, Alemanha.
- 2006 Organização: 2nd International Symposium of Post Graduation and Research (II SINPOSPq). Ribeirão Preto/SP, Brasil.
- 2003 Poster: 4th Congress of Pharmaceutical Sciences (CIFARP). Ribeirão Preto/SP, Brasil.
- 2002 Organização: III Encontro Farmacêutico de Ribeirão Preto (III ENFARP). Ribeirão Preto/SP, Brasil.

Idiomas.....

Inglês (avançado), alemão (básico)

AGRADECIMENTOS

Ao Prof. Dr. Alberto José Cavalheiro pela oportunidade, aprendizado, orientação, amizade e confiança. É uma honra ter tido estes quatro anos de convivência pacífica, prazerosa e extremamente produtiva.

Aos professores do NuBBE, ao Prof. Dr. Michael Hippler, à Dra. Roseli Buzanelli Torres e à Prof. Dra. Fabiola Manhas Verbi Pereira, pelas valiosas contribuições e interesse em compartilhar conhecimento.

À aluna de iniciação científica Livia de Lima Thomaz pela preciosa e competente ajuda.

Aos meus amados familiares, amigos próximos e pessoas extremamente queridas por estarem sempre presentes, pelo suporte e companheirismo. Fizemos tudo isto juntos.

Ao Fernando por ser incomparavelmente exemplar. Pelo constante apoio, generosidade, pelas inerentes sugestões e contribuições.

À todos os funcionários da Seção de Pós-Graduação, da Biblioteca, do Departamento de Química Orgânica, do escritório FAPESP, aos motoristas e funcionários da Seção de Transportes e à Diretoria do Instituto de Química de Araraquara, pela atenção, profissionalismo e prestatividade.

À pós-graduanda Rosana Casoti e aos técnicos Jacqueline Nakau Mendonça Galiete Silva e José Carlos Tomáz da Faculdade de Ciências Farmacêuticas de Ribeirão Preto (FCFRP-USP) pela preciosa ajuda na obtenção dos espectros de massas.

Aos técnicos e especialistas João Luis Bronzel, Juliana Rodrigues e Nivaldo Boralle pelas inúmeros auxílios e dedicação impecável.

À todos os colegas do Instituto de Química, do NuBBE e do Institut für Biologie und Biotechnologie der Pflanzen em Münster pela amizade, aprendizado, ajuda, troca de experiências e apoio.

E à todos que, de que alguma forma, contribuíram para a realização deste trabalho ou que viabilizaram a oportunidade de aprender.

Agradeço também às agências de fomento à pesquisa FAPESP, CAPES e CNPQ pelo apoio financeiro concedido.

RESUMO

A espécie *Casearia sylvestris* Sw. (Salicaceae) é um importante representante do gênero *Casearia* Jacq. destacando-se nos pontos de vista químico, farmacológico, econômico, biotecnológico e ecológico. Por possuir alta capacidade adaptativa, encontra-se amplamente disseminada nas Américas Central e do Sul, sendo que, no Brasil, ocorre em praticamente todos os biomas. Os dados morfo-anatômicos, químicos e genéticos disponíveis sobre *C. sylvestris*, indicam relação próxima entre variedade *lingua* e bioma Cerrado e entre variedade *sylvestris* e Mata Atlântica, havendo também dados descrevendo diferenças genéticas significativas entre estas duas variedades. Em ecótonos Cerrado/Mata Atlântica, estas duas variedades co-existem, havendo principalmente nessas regiões indivíduos com características intermediárias, tanto do ponto de vista morfo-anatômico quanto genético. Há também indícios de que em *C. sylvestris* var. *lingua*/Cerrado ocorram predominantemente compostos fenólicos, enquanto em *C. sylvestris* var. *sylvestris*/Mata Atlântica predominam os diterpenos clerodânicos. Assim, este trabalho tem como objetivo principal avaliar e confirmar indícios de que a composição química, no que concerne a metabólitos secundários, está relacionada e/ou condicionada pelos biomas, e principalmente associada aos respectivos morfotipos predominantes e, portanto, sob forte controle genético. Para tanto, desenvolveu-se e validou-se um método analítico por UHPLC-DAD para a análise simultânea de compostos fenólicos e diterpenos clerodânicos produzidos pelas duas variedades de *C. sylvestris*. Através do estudo metabolômico destas duas variedades utilizando foram identificados 15 diterpenos clerodânicos em *C. sylvestris* var. *sylvestris* e 14 flavonoides glicosilados e uma catequina em *C. sylvestris* var. *lingua*. O estudo dos ciclos circadiano e sazonal de ambas as variedades evidenciou um decréscimo na produção de ambas as classes de metabólitos no período reprodutivo das duas variedades. Nos estudos de variabilidade química infraespecífica, 134 indivíduos foram coletados em diversos ecossistemas/biomas brasileiros. Foi possível confirmar a produção de prioritariamente flavonoides glicosilados pela variedade *lingua*, bem como correlacionar sua ocorrência em áreas de alta incidência de luz solar como em áreas de Cerrado; para a variedade *sylvestris*, observou-se a produção característica de diterpenos clerodânicos, correlacionando sua ocorrência em áreas de menor incidência de luz solar, a exemplo das áreas remanescentes de Mata Atlântica. Finalmente, o estudo proteômico permitiu a identificação de 547 proteínas expressas diferencialmente nas duas variedades em estudo. Além da identificação de proteínas do metabolismo secundário, a análise das proteínas fotossintéticas evidenciou uma maior expressão de RuBisCo juntamente com menor expressão das proteínas do Fotossistema II na var. *lingua* quando comparada com a var. *sylvestris*, sugerindo um possível mecanismo de adaptação do aparato fotossintético destas variedades dependendo do ambiente ao qual estão submetidas.

Palavras chave: *Casearia sylvestris*, metabolômica, proteômica, produtos naturais de origem vegetal

ABSTRACT

The species *Casearia sylvestris* Sw (Salicaceae) is an important representative of the *Casearia* Jacq. genus, due to its chemical, pharmacological, economic, environmental and biotechnological aspects. Because of its high adaptive capacity, this species is widespread in Central and South America and, in Brazil, it occurs in practically all biomes. The morpho-anatomical, chemical and genetic data available on *C. sylvestris*, indicate close relationship between the variety “*lingua*” and the Cerrado biome, and between the variety “*sylvestris*” and Atlantic Forest. Also, there are data describing significant genetic differences between these two varieties. In ecotones Cerrado/Atlantic Forest, it can be found these two varieties, including individuals with intermediate characteristics, considering both morpho-anatomical and genetic ones. There is evidence that in *C. sylvestris* var. “*lingua*”/Cerrado occur predominantly the phenolic compounds, while in *C. sylvestris* var. “*sylvestris*”/Atlantic Forest the clerodane diterpenes predominate. So, the main objective of this work is to evaluate if the chemical composition, considering the secondary metabolites, is related to and/or conditioned by biomes, and mainly associated to their predominant morphotypes and, therefore, under strong genetic control. To evaluate this, an analytical method using UHPLC-DAD was developed and validated for the simultaneous analysis of phenolic compounds and clerodane-type diterpenes produced by the two *C. sylvestris* varieties. The metabolomic study allowed the identification of 15 clerodane-type diterpenes in *C. sylvestris* var. *sylvestris* and 14 glycosylated flavonoids and one catechin in *C. sylvestris* var. *lingua*. The study of the circadian rhythm and seasonal cycle of both varieties showed a decrease in the production of the two secondary metabolites classes in the reproductive period. In the infraspecific chemical variability study, 134 *C. sylvestris* individuals were collected in different ecosystems/biomes. It was possible to confirm the production of glycosylated flavonoids by var. *lingua* individuals as well as correlate their occurrence in areas of high sunlight incidence, like in Cerrado areas; for var. *sylvestris* individuals, it was confirmed the production clerodane-type diterpenes, correlating their occurrence in areas of lower sunlight incidence, like in the Atlantic Forest remnant areas. Finally, the proteomic study allowed the identification of 547 proteins differentially expressed in the two varieties. Besides the identification of secondary metabolism proteins, the analysis of the photosynthetic proteins showed an up-regulation of RuBisCo along with a down-regulation of Photosystem II proteins in var. *lingua* when compared to var. *sylvestris*, suggesting a possible adaptation mechanism of the photosynthetic apparatus of these two varieties depending on the environment to which they are subjected.

Key words: *Casearia sylvestris*, metabolomics, proteomics, plant natural compounds

LISTA DE ILUSTRAÇÕES

Figura 1- Integração das ciências “ômicas” no estudo da função gênica. Traduzido de Sumner et al., 2003.	4
Figura 2- Representação esquemática das abordagens utilizadas no estudo proteômico. Traduzido e modificado a partir de Matros et al., 2011.	6
Figura 3- Distribuição geográfica de <i>C. sylvestris</i> . Fonte: Marquete, 2010.	11
Figura 4- Esqueleto básico dos diterpenos clerodânicos e suas formas isoméricas, característicos do gênero <i>Casearia</i>	13
Figura 5- Gradiente exploratório utilizado no processo de composição da amostra representativa de <i>C. sylvestris</i> (ARCS).....	22
Figura 6- Método provisório por UHPLC-DAD para análise de <i>C. sylvestris</i>	23
Figura 7- Gradiente simplificado para a otimização do método cromatográfico por UHPLC-DAD.....	25
Figura 8- Gradiente de 65-100% de B utilizado para monitoramento das frações ricas em casearinas.	32
Figura 9- Gradiente rápido usado na análise por UHPLC-DAD das frações selecionadas obtidas no fracionamento do extrato bruto liofilizado da var. <i>lingua</i> (EBLL).....	35
Figura 10- Amostra de folhas de <i>C. sylvestris</i> var. <i>lingua</i> (à esquerda) e <i>C. sylvestris</i> var. <i>sylvestris</i> (à direita).....	50
Figura 11- Perfil cromatográfico da amostra representativa de <i>C. sylvestris</i> (ARCS), utilizada durante todo o desenvolvimento analítico. Cromatograma obtido por HPLC-DAD (235 nm).....	51
Figura 12- Cromatogramas da avaliação da influência da adição de ácido fórmico à fase móvel. (A) método provisório acrescido de 0,1% de ácido fórmico. (B) método provisório sem adição do ácido. Em azul, cromatograma em 254 nm; em preto, cromatograma em 235 nm.	52
Figura 13- Gráficos do planejamento experimental Simplex Centróide. (A) planejamento com os pontos e proporções dos solventes utilizados; (B) resultados do planejamento em 235 nm; (C) resultados do planejamento em 254 nm; (D) resultados do planejamento com a média dos resultados.	54
Figura 14- Cromatograma da amostra representativa extraída na proporção ótima de solventes definida pelo planejamento Simplex Centróide. Em azul cromatograma em 254 nm; em preto, em 235 nm.....	55
Figura 15- Cromatogramas obtidos no teste da influência de diferentes ácidos no procedimento de extração. 1- água, 2- tampão pH 4,5, 3- TFA, 4- ácido acético, 5- ácido fórmico, A = 235 nm, B = 254 nm.	56
Figura 16- Perfis cromatográficos btidos por UHPLC-DAD. (A) amostra representativa de <i>C. sylvestris</i> (ARCS); (B) <i>C. sylvestris</i> var. <i>lingua</i> (CSL); (C) <i>C. sylvestris</i> var. <i>sylvestris</i>	

(CSS). Os cromatogramas foram plotados em 254 nm (de 0 a 25 min) e 235 nm (de 25 a 45 min). Os sinais *, i and PI representam os picos característicos dos compostos fenólicos (mais precisamente flavonoides glicosilados), diterpenos clerodânicos e padrão interno, respectivamente. As letras X e Y representam os picos selecionados para a validação do método. Condições cromatográficas definidas pelo planejamento: coluna cromatográfica Kinetex 150 mm × 2,1; 2,6 μ m, 100 Å. Fase móvel: água (A) e acetonitrila (B). Gradiente de eluição: 10-25% de B de 0 a 15 min, 25-90% de B até 35 min, mantendo-se a 90% de B até 40 min, retorno à condição inicial em 2 min e tempo de equilíbrio de 3 min. Vazão: 0,4 mL/min. Temperatura do forno: 35 °C. Volume de injeção: 2 μ L.59

Figura 17- Cromatogramas em 254 nm, do extrato de *C. sylvestris* variedade *lingua* (CSL, em vermelho), do extrato de *C. sylvestris* variedade *sylvestris* (CSS, em azul) e da mistura de galatos (picos de 1 a 13, respectivamente, ácido gálico, galatos de metila, etila, propila, butila, pentila, hexila, heptila, octila, nonila, decila, undecila e dodecila), nas condições cromatográficas otimizadas. Destacado, está o galato de butila.....60

Figura 18- Efeito das matrizes das duas variedades de *C. sylvestris* sobre o padrão interno, galato de butila, em cinco diferentes concentrações.....61

Figura 19- Curva analítica da rutina apresentando a regressão linear.....63

Figura 20- Curva analítica da caseargrewiina F apresentando a regressão linear.63

Figura 21- Padrões de casearinas isoladas neste estudo.....67

Figura 22- Cromatograma de *C. sylvestris* var. *sylvestris* (235 nm), destacando (em rosa) a região rica em diterpenos clerodânicos70

Figura 23- Ampliação da região do cromatograma onde ocorre a separação e a coleta das frações referentes aos diterpenos clerodânicos de *C. sylvestris* var. *sylvestris*.71

Figura 24- Estruturas químicas da casearina D e casearina M.....73

Figura 25- Espectro de massas (MS/MS) da casearina D (pico # 20).73

Figura 26- Estrutura química da casearina S.....74

Figura 27- Espectro de massas (MS/MS) da casearina S (pico #31).....74

Figura 28- Estruturas químicas da casearina K e casearina Q75

Figura 29- Estrutura química da casearina J.75

Figura 30- Cromatograma de *C. sylvestris* var. *sylvestris* obtido por UHPLC-DAD mostrando os picos correspondentes às substâncias identificadas. O cromatograma foi plotado em 254 nm (de 0 a 25 min) e 235 nm (de 25 a 45 min). Em vermelho, substâncias identificadas por desreplicação (UHPLC-DAD-ESI-Orbitrap). Em azul, substâncias também identificadas por desreplicação e confirmadas através do isolamento em pequena escala (UHPLC-DAD) e identificadas por inserção direta em espectrômetro de massas (microTOF Q II - ESI-TOF). Para a rotina, casearina x, casearina J e caseargrewiina F, os tempos de retenção também coincidem com os tempos de retenção de padrões autênticos.76

Figura 31- Perfil cromatográfico obtido em 254 nm do extrato bruto de <i>C. sylvestris</i> var. <i>lingua</i> (EBLL, em vermelho) e da fração enriquecida em compostos fenólicos (FRCF, em preto).....	77
Figura 32- Cromatogramas das subfrações obtidas por EFS com poliamida da fração FRCF (254 nm). Em preto a fração tal e qual; em azul a fração hidro-metanólica; em rosa a fração metanólica rica em flavonoides e em vermelho a fração metanólica basificada com hidróxido de amônio.....	78
Figura 33- Cromatograma da fração metanólica em condição isocrática (20% de acetonitrila em água acidificada com 0,1% de ácido fórmico, 254 nm) durante o isolamento dos compostos fenólicos. As letras de A a F correspondem às substâncias identificadas por desreplicação (MS e MS/MS); os números de 1 a 5 correspondem substâncias que foram identificadas por elucidação estrutural (NMR).....	79
Figura 34- Numeração da estrutura básica dos flavonoides	82
Figura 35- Estrutura química da (+)-catequina (substância 1).	82
Figura 36- Estrutura química da rutina (substância 2)	84
Figura 37- Estrutura química da isoramnetina-3-O-neohesperosideo (substância 3)	86
Figura 38- Estrutura química da narcisina (substância 4).	88
Figura 39- Estrutura química da isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- α -L-arabinopiranosideo (substância 5).	90
Figura 40- Perfil cromatográfico de <i>C. sylvestris</i> var. <i>lingua</i> , mostrando os flavonoides isolados e identificados neste estudo). O cromatograma foi plotado em 254 nm (de 0 a 25 min) e 235 nm (de 25 a 45 min).	92
Figura 41- Teor de água nas folhas de <i>C. sylvestris</i>	94
Figura 42- Umidade relativa do ar (acima) e temperatura média (abaixo) durante o ano, no período de 24 horas.....	95
Figura 43- Gráficos de radar exemplificando o ciclo circadiano dos flavonoides e diterpenos nas duas variedades de <i>C. sylvestris</i> . Em azul: var. <i>lingua</i> ; Em vermelho: var. <i>sylvestris</i> . A escala circular indica hora do dia e o eixo vertical principal indica o teor em mg/g. A área sombreada em azul corresponde ao período noturno e a área sombreada em amarelo corresponde ao período diurno.	97
Figura 44-Gráficos box plot obtidos para a análise do ciclo sazonal dos flavonoides glicosilados na variedade <i>lingua</i>	98
Figura 45- Gráfico box plot obtido para a análise do ciclo sazonal da rutina na variedade <i>sylvestris</i>	99
Figura 46- Gráfico box plot obtido para a análise do ciclo sazonal das casearinas na variedade <i>sylvestris</i>	100
Figura 47- Avaliação da conformidade das injeções comparando-se as 12 amostras reinjetadas (análise 2) com as injeções iniciais (análise 1) focando-se na razão de áreas	

dos diterpenos pelo padrão interno. OBS: as letras L e S que aparecem logo em seguida do mês correspondem às variedades *lingua* e *sylvestris*, respectivamente. 102

Figura 48- Avaliação da conformidade das injeções comparando-se as 12 amostras reinjetadas (análise 2) com as injeções iniciais (análise 1) focando-se na razão de áreas dos compostos fenólicos pelo padrão interno. OBS: as letras L e S correspondem às variedades *lingua* e *sylvestris*, respectivamente. 102

Figura 49- Cromatogramas representativos das coletas de variabilidade infraespecífica. Mapa e legenda adaptados do IBGE (<http://atlasescolar.ibge.gov.br>). 104

Figura 50- (A) Gráfico de scores (PC1 x PC2) das amostras de *C. sylvestris* provenientes do Estado de São Paulo, mostrando o agrupamento de acordo com as populações de origem. (B) Gráfico de scores (PC1 x PC2) das amostras de *C. sylvestris* classificadas de acordo com as variedades (intermediária, *lingua* e *sylvestris*). (C) Gráfico de *loadings* (PC1 x PC2) para discriminação de variáveis. 106

Figura 51- A) Gráfico de scores (PC1 x PC2) das amostras de *C. sylvestris* provenientes do território brasileiro, mostrando o agrupamento de acordo com as populações de origem. (B) Gráfico de scores (PC1 x PC2) das amostras de *C. sylvestris* classificadas de acordo com as variedades (intermediária, *lingua* e *sylvestris*). (C) Gráfico de *loadings* (PC1 x PC2) para discriminação de variáveis. As elipses concentram amostras de características comuns: azul (clima quente e úmido, vegetação menos exposta ao sol): Floresta Estacional Semi-Decidual; verde (clima quente e super úmido, vegetação menos exposta ao sol): Floresta Ombrófila Densa e Mata ciliar; preto (clima quente e seco, exposição excessiva ao sol): Cerrado e Carrasco; amarelo (clima seco porém com áreas alagadas e vegetação excessiva ao sol): Pampa e Pantanal. 109

Figura 52- Curva analítica de BSA para quantificação de proteínas pelo método do BCA. 112

Figura 53- Curva analítica de BSA para quantificação de proteínas pelo método do amido black. 114

Figura 54- Gráfico do teor médio de proteínas das amostras de *C. sylvestris* pelos métodos do BCA e amido black. 115

Figura 55- SDS-PAGEs dos extratos proteicos normalizados por dois métodos de quantificação. (A) amostra normalizada pelo método do BCA (aplicou-se em cada poço 20 µg de proteínas); (B) amostra normalizada pelo método do amido black (aplicou-se em cada poço 10 µg de proteínas). Coloração feita com *comassie blue*. As região com proteínas mais abundantes (faixas I e II) correspondem às proteínas RuBisCo (55 KDa) e proteínas coletoras de luz (*light-harvesting complex proteins* ou LHCs, 25 KDa). 116

Figura 56- *Western blotting* das amostras PL1 (variedade *lingua*) e PS1 (variedade *sylvestris*) utilizando os anticorpos anti-rcbl, P25K e IsPE. Imagem obtida após revelação manual. 117

Figura 57- Análise semi-quantitativa do teor de RuBisCo (anti-rcbl) em três concentrações. 118

Figura 58- Análise semi-quantitativa do teor de LHC (anti-P25K) em três concentrações. 118

Figura 59- Análise semi-quantitativa do teor de ISPE (anti-IsPE) em três concentrações.	118
Figura 60- Classificação de todas proteínas identificadas nas duas variedades em estudo de acordo com suas funções biológicas.....	122
Figura 61- Análise comparativa entre <i>C. sylvestris</i> var. <i>sylvestris</i> e <i>C. sylvestris</i> var. <i>lingua</i> em função dos <i>scan counts</i>	123
Figura 62- Análise comparativa quantitativa das proteínas fotossintéticas expressas por <i>C. sylvestris</i> var. <i>lingua</i> e <i>C. sylvestris</i> var. <i>sylvestris</i>	124
Figura 63- Análise comparativa quantitativa das proteínas do metabolismo secundário expressas por <i>C. sylvestris</i> var. <i>lingua</i> e <i>C. sylvestris</i> var. <i>sylvestris</i>	125

LISTA DE TABELAS

Tabela 1- Proporções de cada solvente nas misturas extratoras testadas no planejamento Simplex Centróide.	24
Tabela 2- Planejamento e parâmetros do Planejamento Fatorial Fracionário 2^{4-1} com ponto central.	26
Tabela 3 - Planejamento Composto Central e variáveis selecionadas para a definição da melhor condição cromatográfica.	26
Tabela 4- Distribuição F – 5% de probabilidade.	28
Tabela 5- Fracionamento da fração EFS2 por cromatografia em coluna de fase normal.	30
Tabela 6- Massas das subfrações obtidas pelo fracionamento da fração EFS2.	31
Tabela 7- Amostras coletadas no território brasileiro para o estudo de variabilidade infraespecífica de <i>C. sylvestris</i>	40
Tabela 8- Descrição dos anticorpos e detalhes da análise de <i>western blotting</i>	45
Tabela 9- Resultados do planejamento Simplex Centróide para a definição do melhor solvente extrator.	53
Tabela 10- Resultados do Planejamento Fatorial Fracionário 2^{4-1} com ponto central para triagem de variáveis.	57
Tabela 11- Resultados do Planejamento Composto Central definição do método cromatográfico.	58
Tabela 12- Resultados obtidos para determinação da precisão do método analítico para análise simultânea de compostos fenólicos e diterpenos clerodânicos por UHPLC-DAD.	62
Tabela 13- Resultados das curvas analíticas para a rutina e caseargrewiina F, em função do padrão interno galato de butila (n=3).	64
Tabela 14- Dados de NMR de ^{13}C obtidos para as três casearinas em piridina- d_5 isoladas neste trabalho.	66
Tabela 15- Dados de desreplicação de <i>C. sylvestris</i> var. <i>sylvestris</i> obtidos por UHPLC-DAD-ESI-Orbitrap.	69
Tabela 16- Tabela de identificação das casearinas coletadas de <i>C. sylvestris</i> var. <i>sylvestris</i> por UHPLC-DAD e analisadas por espectrometria de massas com inserção direta (microTOF Q II - ESI-TOF).	72
Tabela 17- Picos coletados e quantidades das substâncias obtidas durante o isolamento dos compostos fenólicos de <i>C. sylvestris</i> var. <i>lingua</i>	79

Tabela 18- Dados de desreplicação de <i>C. sylvestris</i> var. <i>lingua</i> obtidos por HRMS (inserção direta, micrOTOF Q II - ESI-TOF).	81
Tabela 19- Dados espectrométricos de NMR (^1H) obtidos para a Substância 1 a 600 MHz, em DMSO- d_6 : (+)-catequina.	83
Tabela 20- Dados espectrométricos de NMR (^1H) obtidos para a substância 2 e para o padrão de rutina a 600 MHz, em DMSO- d_6	85
Tabela 21- Dados espectrométricos de NMR (^1H) obtidos para a substância 3 a 600 MHz em DMSO- d_6	87
Tabela 22- Dados espectrométricos de NMR (^1H) obtidos para a substância 4 a 600 MHz em DMSO- d_6	89
Tabela 23- Dados espectrométricos de NMR (^1H) obtidos para a substância 5 a 600 MHz em DMSO- d_6	91
Tabela 24- Cálculos de recuperação da fração dos extratos brutos das duas variedades de <i>C. sylvestris</i> (EBLL e EBLs) e da fração rica em compostos fenólicos (FRCF) retidas em poliamida.	96
Tabela 25- Resultados de absorbância dos pontos da curva analítica de BSA para quantificação de proteínas pelo método do BCA.	112
Tabela 26- Resultados obtidos na quantificação de proteínas pelo método do BCA.	113
Tabela 27- Resultados de absorbância dos pontos da curva analítica de BSA para quantificação de proteínas pelo método do amido black.	114
Tabela 28- Resultados obtidos na quantificação de proteínas pelo método do amido black.	114
Tabela 29- Resultados do rendimento dos extratos proteicos e do teor médio de proteínas das amostras de <i>C. sylvestris</i> pelos métodos do BCA e amido black.	115
Tabela 30- Número de peptídeos identificados de acordo com duas diferentes estratégias.	120

LISTA DE QUADROS

Quadro 1- Protocolo para obtenção dos extratos proteicos	42
Quadro 2- Protocolo FASP	46

LISTA DE ABREVIATURAS E SIGLAS

2D-PAGE	Eletroforese em gel de poliacrilamida bidimensional (<i>Two dimensional polyacrylamide gel electrophoresis</i>)
3-PGA	3-fosfoglicerato
ACN	Acetonitrila
AcOEt	Acetato de Etila
ADP	Adenosina-Di-Fosfato
ATP	Adenosina-Tri-Fosfato
ax	Axial
BCA	Ácido bicinchinínico (<i>Bicinchoninic acid</i>)
CID	Dissociação induzida por colisão (<i>Collision induced dissociation</i>)
<i>d</i>	Duplete
DAD	Detector de arranjo de diodos (<i>Diode array detector</i>)
DCM	Diclorometano
<i>dd</i>	Duplo duplete
<i>DIGE</i>	Eletroforese em gel diferencial (<i>Difference gel electrophoresis</i>)
DMSO-d ₆	Dimetil sulfóxido deuterado
DPR	Desvio padrão relativo
<i>eq</i>	Equatorial
ESI	Ionização por eletrospray (<i>Electrospray ionization</i>)
FASP	Preparo de amostra assistida por filtro (<i>Filter aided sample preparation</i>)
GAP	Gliceraldeído fosfato
GC	Cromatografia gasosa (<i>Gas chromatography</i>)
gli	Glicose
HEX	<i>n</i> -Hexano
HPLC	Cromatografia líquida de alta eficiência (<i>High performance liquid chromatography</i>)
HRMS	Espectrometria de massas de alta resolução (<i>High resolution mass spectrometry</i>)
IPA	Isopropanol
IV	Infravermelho
<i>J</i>	Constante de acoplamento
LC	Cromatografia líquida (<i>Liquid chromatography</i>)
LHC	Complexo antena coletor de luz (<i>Light-harvesting antenna complex</i>)
<i>m/z</i>	Massa/carga

MALDI	Dessorção/ionização a laser assistida por matriz (<i>Matrix-assisted laser desorption/ionization</i>)
MeOH	Metanol
MS/MS	Espectrometria de massas sequencial
NADP ⁺	Nicotinamida adenina dinucleotídeo fosfato
NADPH	Nicotinamida adenina dinucleotídeo fosfato reduzido
NMR	Ressonância magnética nuclear (<i>Nuclear magnetic resonance</i>)
PSI	Fotossistema I (<i>Photosystem I</i>)
PSII	Fotossistema II (<i>Photosystem II</i>)
ram	Ramnose
RuBisCo	Ribulose-1,5-bisfosfato carboxilase oxigenase
RuBP	Ribulose-1,5-bifosfato
s	Singleto
SDS-PAGE	<i>Sodium dodecyl sulfate polyacrylamide gel electrophoresis</i>
TLC	Cromatografia em camada delgada (<i>Thin layer chromatography</i>)
t_R	Tempo de retenção
UHPLC	Ultra high performance liquid chromatography
UV/Vis	Ultravioleta/Visível
Var.	Variedade
δ	Deslocamento químico

SUMÁRIO

DADOS CURRICULARES	i
AGRADECIMENTOS	vi
RESUMO	vii
ABSTRACT	viii
LISTA DE ILUSTRAÇÕES	ix
LISTA DE TABELAS	xiv
LISTA DE QUADROS	xvi
LISTA DE ABREVIATURAS E SIGLAS	xvii
SUMÁRIO	xix
1 INTRODUÇÃO	1
1.1 O metabolismo vegetal e as ciências ômicas	1
1.2 Ferramentas do estudo proteômico	5
1.3 Estudo metabolômico de produtos de origem vegetal	8
1.4 <i>Casearia sylvestris</i> Sw. (Salicaceae)	10
2 OBJETIVOS	15
3 MATERIAL	16
3.1 Equipamentos utilizados e suas especificações	16
3.2 Reagentes, solventes, adsorventes e padrões	18
3.3 Softwares	19
3.4 Material vegetal	19
4 MÉTODOS	21
4.1 PARTE 1- Desenvolvimento de método cromatográfico por UHPLC-DAD	22
4.1.1 Composição da amostra representativa de <i>C. sylvestris</i> (ARCS)	22
4.1.2 Definição de um método provisório por UHPLC-DAD	23
4.1.3 Otimização das condições de extração das amostras	23
4.1.3.1 Planejamento Simplex Centróide para definição do solvente extrator	23
4.1.3.2 Estudo da influência do pH na etapa de extração	24
4.1.4 Otimização da condição de eluição cromatográfica	24
4.1.5 Escolha do padrão interno	27
4.1.6 Avaliação da precisão	27
4.1.7 Avaliação da linearidade e obtenção de curvas analíticas	29
4.2 PARTE 2- Estudo metabolômico de <i>C. sylvestris</i>	30
4.2.1 Obtenção de padrões de diterpenos clerodânicos (casearininas)	30
4.2.2 Identificação de casearininas em <i>C. sylvestris</i> var. <i>sylvestris</i>	32
4.2.2.1 Análise qualitativa utilizando UHPLC-DAD-ESI-Orbitrap	32
4.2.2.2 Validação dos dados obtidos	33
4.2.3 Identificação dos compostos fenólicos em <i>C. sylvestris</i> var. <i>lingua</i>	34
4.2.3.1 Desreplicação por HRMS utilizando substâncias isoladas	34
4.2.3.2 Validação dos dados obtidos	35

4.3	PARTE 3 - Estudos de dinâmica metabólica	36
4.3.1	Coletas, processamento e caracterização das amostras	36
4.3.2	Avaliação do teor de taninos por extração em fase sólida com poliamida	37
4.3.3	Análise das amostras e avaliação da conformidade das análises	37
4.3.3.1	Estudo de dinâmica metabólica: ciclos circadiano e sazonal	37
4.3.3.2	Avaliação da conformidade das análises	38
4.4	PARTE 4 – Estudos de variabilidade química infraespecífica	39
4.4.1	Coletas e classificação das amostras	39
4.4.2	Análise das amostras e processamento dos dados	39
4.5	PARTE 5- Estudo proteômico diferencial de <i>C. sylvestris</i>	41
4.5.1	Coleta, obtenção e avaliação dos extratos proteicos	41
4.5.1.2	Análise quantitativa pelo método do BCA	42
4.5.1.3	Análise quantitativa pelo método do amido black	43
4.5.1.4	Análise qualitativa por SDS-PAGE	44
4.5.2	<i>Western blotting</i>	44
4.5.3	Identificação das proteínas do estudo proteômico diferencial	45
4.5.3.2	Preparo das amostras para LC-MS, etapa de digestão: FASP	45
4.5.3.3	Preparo das amostras para LC-MS, etapa de fracionamento: Pós-FASP	47
4.5.3.4	Análise por LC-MS/MS	47
4.5.3.5	Análise dos dados por bioinformática	48
5	RESULTADOS E DISCUSSÃO	50
5.1	PARTE 1- Desenvolvimento de método cromatográfico por UHPLC-DAD	50
5.1.1	Composição da amostra representativa de <i>C. sylvestris</i> (ARCS)	50
5.1.2	Definição de um método provisório por UHPLC-DAD	51
5.1.3	Otimização das condições de extração das amostras	52
5.1.3.1	Planejamento Simplex Centróide para definição do solvente extrator	52
5.1.3.2	Estudo da influência do pH na etapa de extração	55
5.1.4	Otimização da condição de eluição cromatográfica	56
5.1.5	Escolha do padrão interno	60
5.1.6	Avaliação da precisão	61
5.1.7	Avaliação linearidade e obtenção de curvas analíticas	63
5.2	PARTE 2- Estudo metabolômico de <i>C. sylvestris</i>	65
5.2.1	Obtenção de padrões de diterpenos clerodânicos (casearininas)	65
5.2.2	Identificação de casearininas em <i>C. sylvestris</i> var. <i>sylvestris</i>	67
5.2.2.1	Análise qualitativa utilizando UHPLC-DAD-ESI-Orbitrap	67
5.2.2.2	Validação dos dados obtidos	70
5.2.3	Identificação dos compostos fenólicos em <i>C. sylvestris</i> var. <i>lingua</i>	76
5.2.3.2	Desreplicação por HRMS utilizando substâncias isoladas	76
5.2.3.3	Validação dos dados obtidos	82
5.2.3.2.1	Substância 1	82
5.2.3.2.2	Substância 2	84
5.2.3.2.3	Substância 3	86
5.2.3.2.4	Substância 4	88
5.2.3.2.5	Substância 5	90
5.3	PARTE 3 - Estudos de dinâmica metabólica	94
5.3.1	Coletas, processamento e caracterização das amostras	94
5.3.2	Avaliação do teor de taninos por extração em fase sólida com poliamida	95
5.3.3	Análise das amostras e avaliação da conformidade das análises	96
5.3.3.1	Estudo de dinâmica metabólica: ciclos circadiano e sazonal	96

5.3.3.2 Avaliação da conformidade das análises	101
5.4 PARTE 4 - Estudos de variabilidade química infraespecífica	103
5.5 PARTE 5- Estudo proteômico diferencial de <i>C. sylvestris</i>	111
5.5.1 Coleta do material vegetal, obtenção e avaliação dos extratos proteicos	111
5.5.2 <i>Western blotting</i>	116
5.5.3 Identificação das proteínas do estudo proteômico diferencial	119
<u>6 CONCLUSÕES E PERSPECTIVAS</u>	<u>127</u>
<u>REFERÊNCIAS</u>	<u>129</u>
<u>APÊNDICES</u>	<u>138</u>

1 INTRODUÇÃO

1.1 O metabolismo vegetal e as ciências ômicas

Grande parte do genoma de uma planta codifica enzimas responsáveis pela síntese de substâncias que compõem os metabolismos primário e secundário. Através da ação destas enzimas, inúmeras transformações bioquímicas envolvendo pequenas e grandes moléculas culminarão na captação, utilização da luz solar e fixação de carbono, como é o caso da fotossíntese, na produção de energia, como é o caso da glicólise do ciclo do ácido cítrico, na síntese de moléculas estruturais, a exemplo dos aminoácidos, ácidos graxos e macromoléculas, tais como ácidos polinucleicos, proteínas e parede celular (FRIDMAN; PICHERSKY, 2005), ou mesmo nos processos específicos do metabolismo secundário, que culminam na biossíntese de diversas classes de metabólitos (HARBORNE et al., 1993).

Processo comum a todos os organismos fotossintéticos, a transformação da energia luminosa em energia química é o ponto de partida para que todas as reações bioquímicas necessárias à sobrevivência e desenvolvimento da planta possam ocorrer. No processo fotossintético a captação de luz solar é eficientemente realizada por um conjunto de proteínas conhecidas como complexo antena, ou proteínas coletoras de luz (LHCs). Tratam-se de proteínas transmembrana presentes nos tilacóides dos cloroplastos e que são fundamentais no processo de fotossíntese, por se complexarem às moléculas de clorofila e carotenóides na função de absorver e transferir a energia luminosa do fotossistema II (PSII) para o fotossistema I (PSI) através de diversas reações de oxiredução que, por sua vez, gerarão elétrons e prótons. Neste processo, o fluxo de elétrons do PSII para o PSI resultará na redução de NADP^+ à NADPH no sistema ferredoxina/tioredoxina, enquanto que os prótons serão acumulados no interior do lúmen dos tilacóides e comporão uma diferença de pH entre o lúmen e o estroma, favorável para a síntese de ATP a partir do ADP pela enzima ATP sintase (DAMKJAER et al., 2009).

Já no Ciclo de Calvin-Benson, a ribulose-1,5-bifosfato carboxilase oxigenase (RuBisCo), outra enzima bastante abundante nos sistemas vegetais, catalisa a carboxilação da ribulose-1,5-bifosfato (RuBP), produzindo duas moléculas de 3-fosfoglicerato (3-PGA), que serão fosforiladas pela glicerato-3-fosfato quinase, gerando glicerato-1,3-bifosfato que será, por sua vez, reduzido à gliceraldeído fosfato (GAP) pela NADP-gliceraldeído fosfato desidrogenase, tendo NADPH como doador de elétrons. O resultado deste processo é a fixação do carbono, que será utilizado pela planta em todos os demais processos fisiológicos. Sendo assim, observa-se que o processo de

fotossíntese para a fixação do CO₂ atmosférico pela RuBisCo depende da ação das LHCs no processo de captação da energia luminosa e geração de ATP e NADPH (NELSON; YOCUM, 2006; NELSON; COX, 2008).

Da mesma forma, outras enzimas estão envolvidas nos mais variados processos do metabolismo celular especializado que é responsável pela produção de metabólitos secundários que são destinados às mais diversas finalidades (HARBORNE, 1993; HEY et al., 2010), muitas das quais têm sido elucidadas em crescente frequência, enquanto que outras ainda permanecem desconhecidas.

Sabe-se, por exemplo, que existem metabólitos ou classes de metabólitos que possuem propriedades tóxicas contra predadores, tais como os alcalóides, taninos e iridóides, ou atrativas para agentes polinizadores, a exemplo dos pigmentos (flavonoides, antocianinas e betalaínas) e óleos essenciais, estes constituídos majoritariamente de monoterpenos, sesquiterpenos e fenilpropanóides (HARBORNE, 1993; OLIVEIRA et al., 2003). Outras substâncias, tais como as fitoalexinas, compostos fenólicos, furanoterpenos e poliacetilenos, estão envolvidas nos mecanismos de sinalização e defesa quando do ataque da planta por fungos e parasitas. Ainda, outros metabólitos, como é o caso do hormônio ácido abscísico, estão envolvidos nos processos de sinalização e resposta a estresses abióticos, tais como o frio, o calor, os estresses osmóticos ou hídricos, e até mesmo sinalizando condições de herbivoria e ataque por patógenos (HARBORNE, 1993; HEY et al., 2010).

Uma vez que a diferenciação dos vários tipos de metabólitos primários e secundários só é possível graças à ação de tais enzimas codificadas nos genes nucleares, cloroplastidiais ou mitocondriais que catalisam os passos em cada rota metabólica (HARBORNE, 1993), é extremamente importante entender como, onde, quando e porque as enzimas específicas das rotas biossintéticas dos metabolismos secundário, primário e energético são transcritas e reguladas a partir das informações codificadas nos genes, como elas interagem entre si e com outras moléculas, como elas trabalham para promover o crescimento e desenvolvimento programado, bem como para interagir com os fatores bióticos e abióticos (ABRIL et al., 2011).

Além da expressão genética de cada planta e a ação das proteínas envolvidas na biossíntese dos precursores e dos compostos secundários, o metabolismo vegetal é um sistema altamente dinâmico, modulando e afetando os níveis de metabólitos ao responder instantaneamente a estímulos tais como luz, temperatura, disponibilidade de nutrientes ou concentrações enzimáticas, dentre outros (HARBORNE, 1993). A interação das plantas com seu meio ambiente, assim como sua adaptação anatômica e

morfológica a diferentes fatores bióticos e abióticos, ocorrem através da diversificada produção de tais metabólitos, os quais possuem importante variabilidade em sua abundância, mesmo quando determinadas a partir de grupos geneticamente idênticos sob as mesmas condições experimentais (HARBORNE, 1993; MORGENTHAL et al., 2006).

A crescente necessidade de se entender tais mecanismos, bem como o advento de técnicas modernas para o estudo de sistemas biológicos, impulsionaram o desenvolvimento de três grandes eventos que revolucionaram a biologia. O primeiro deles foi o sequenciamento genético que, além de fornecer um imenso banco de dados acerca da identidade genética de um organismo, gerou a necessidade de se elucidar o fenótipo molecular a partir da interpretação funcional do genótipo. A primeira espécie vegetal a ter o genoma totalmente mapeado foi *Arabidopsis thaliana* (Brassicaceae) e, desde então, muitas outras plantas, algas, cianobactérias e outros organismos fotossinteticamente ativos tiveram seu genoma elucidado, impulsionando também o desenvolvimento das “ciências ômicas” e dos estudos *in silico* (JORRIN-NOVO et al., 2009; WECKWERTH, 2011).

O segundo evento, aqui generalizado como “ciências ômicas”, inclui a transcriptômica, a proteômica e a metabolômica, e coincide com a demanda gerada pela genômica de se analisar os dados moleculares oriundos de sistemas vivos na mesma escala do genoma, no contexto de genômica funcional. Até os anos 90, a pesquisa biológica era principalmente focada nos estudos *in vitro* de componentes individuais, nos quais genes e proteínas eram investigadas um por vez. Mais tarde esta estratégia migrou para estudos *in vivo* e para a pesquisa molecular em larga escala. O advento da era genômica e a disponibilidade de genomas inteiramente sequenciados mudou a forma de como a pesquisa é desenvolvida, passando de uma visão reducionista para uma visão holística e com a análise de alto rendimento acerca da função gênica através do emprego das novas tecnologias “ômicas”. Desta forma, juntamente com a genômica, a transcriptômica, a proteômica e a metabolômica constituem as bases da biologia de sistemas moderna, visto que nenhuma abordagem sozinha é capaz de elucidar a complexidade dos organismos vivos, permitindo a análise desde o gene até o fenótipo, incluindo proteínas e metabólitos (ABRIL et al., 2011).

A **Figura 1**, a seguir, ilustra a integração das ciências “ômicas” no contexto da genômica funcional. Os efeitos das perturbações nos genes são avaliados em múltiplos níveis, incluindo o transcriptoma, o proteoma e o metaboloma. Mudanças no metaboloma ocorrem como consequências destas mudanças no transcriptoma que

resultam em mudanças nos níveis ou atividades catalíticas enzimáticas. Portanto, a análise metabolômica é uma ferramenta valiosa para se inferir a função gênica (SUMNER et al., 2003).

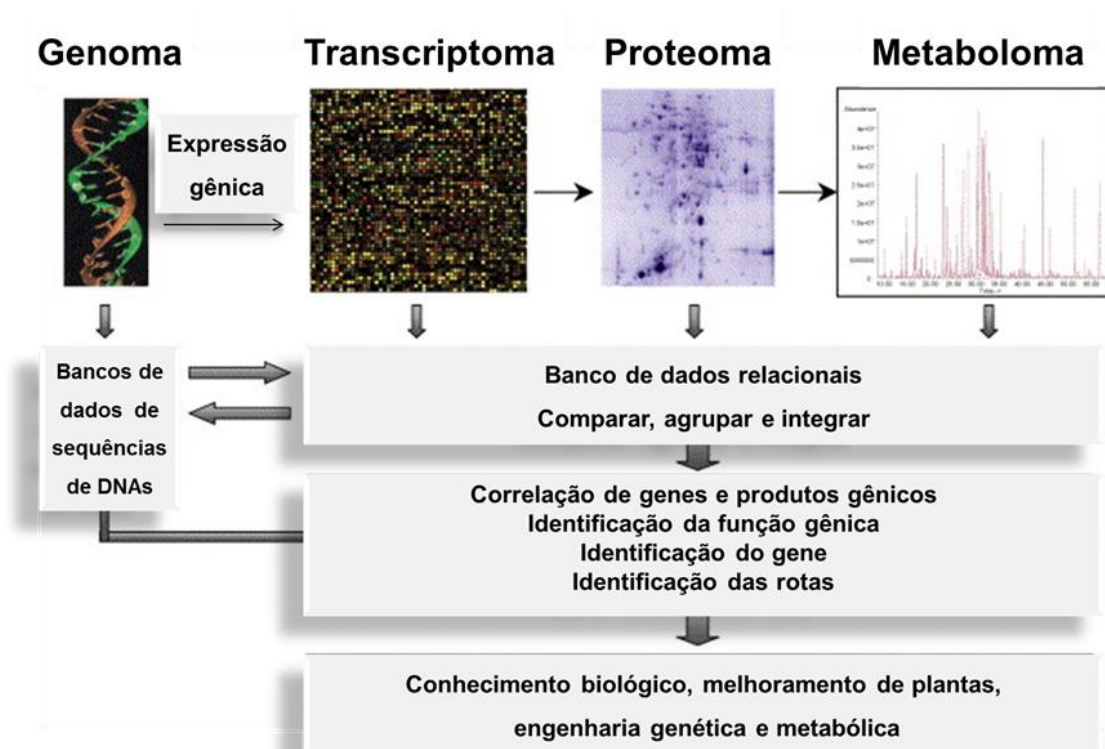


Figura 1- Integração das ciências “ômicas” no estudo da função gênica. Traduzido de Sumner et al., 2003.

Finalmente, o terceiro evento vem de encontro com a necessidade de integração destes dados, numa escala genômica e de dinâmica molecular, com o objetivo de prever o fenótipo e melhor entender os processos fisiológicos dos organismos em estudo. Trata-se da biologia teórica e modelagem assistida por computadores, incluindo também a criação e disponibilização de grandes bancos de dados científicos (WECKWERTH, 2011). Além disso, os dados gerados a partir de experimentos tais como os transcriptômicos, proteômicos e metabolômicos têm se tornado cada vez mais numerosos e complexos. Por conta disso a mineração destes dados é um grande desafio, principalmente quando se deseja integrá-los. Neste sentido, ferramentas em bioinformática e métodos estatísticos multivariados, incluindo a análise de componentes principais (*principal component analysis*, PCA), análise discriminante com método de mínimos quadrados parciais (*partial least squares discriminatory analysis*, PLS-DA), análise hierárquica de clusters (*hierarchical clustering analysis*, HCA), redes neurais e

coeficiente de correlação de Pearson têm sido comumente utilizadas (YANG et al., 2014).

Por conta da enorme diversidade de micromoléculas, a metabolômica, a proteômica e a genômica de plantas estão em rápido desenvolvimento e integração e têm se tornado importantes ferramentas tecnológicas na abordagem da biologia de sistemas, nos estudos de interações entre as plantas e o meio ambiente e nos estudos de atividades biológicas de plantas medicinais (VERPOORTE et al., 2005). Grande parte dos princípios ativos utilizados nos medicamentos foi inspirada em produtos de origem natural, principalmente os metabólitos secundários. A razão pela qual tais produtos são promissores deve-se à alta diversidade química, que por sua vez está relacionada aos efeitos da pressão evolutiva para a biossíntese de moléculas biologicamente ativas, na similaridade estrutural com os alvos proteicos nas diferentes espécies, além da grande variabilidade em suas propriedades físico-químicas intrínsecas, o que, adicionalmente, faz de sua detecção e caracterização um desafio (WOLFENDER et al., 2010).

1.2 Ferramentas do estudo proteômico

Por definição, proteômica é o termo dado ao estudo em larga escala de todas as proteínas expressas num sistema biológico, seja ele uma organela, uma célula, um tecido ou órgão, um indivíduo, uma espécie e até mesmo um ecossistema, sob determinado estágio de desenvolvimento ou condições ambientais específicas (ABRIL et al., 2011; CONN, 2003), podendo indicar, inclusive, o estado fisiológico de um organismo (SARASWATHY; RAMALINGAM, 2011).

Para tanto, as técnicas utilizadas para a análise de extratos proteicos complexos podem ser divididas em dois grupos principais. O primeiro e mais tradicional, compreende as técnicas que empregam separações em gel, como é o caso das eletroforeses em gel de poliacrilamida unidimensional (SDS-PAGE) e bidimensional (2D-PAGE, 2D-DIGE) (SHEORAN et al., 2009). Dentre as vantagens da utilização da SDS-PAGE estão relativo baixo custo e alta disponibilidade dos materiais necessários. No entanto, tais técnicas consomem muito tempo, demandam maiores quantidades de amostra e fornecem informações qualitativas limitadas.

Já o segundo grupo refere-se às técnicas livres de gel, as quais podem ou não ser realizadas com a utilização de isótopos marcados, e fornecem dados extremamente precisos e sensíveis para a análise de perfis proteicos de tecidos, células ou frações subcelulares (MATROS et al., 2011). Dentre eles estão a espectrometria de massas de

dessorção/ionização a laser assistida por matriz (MALDI) e a cromatografia líquida acoplada à espectrometria de massas sequencial (LC-MS/MS) (NATARAJAN et al., 2009), técnicas que têm conquistado cada vez mais espaço, complementando ou substituindo a abordagem de separação em gel (MATROS et al., 2011).

A **Figura 2** a seguir evidencia as diferentes abordagens para o estudo proteômico.

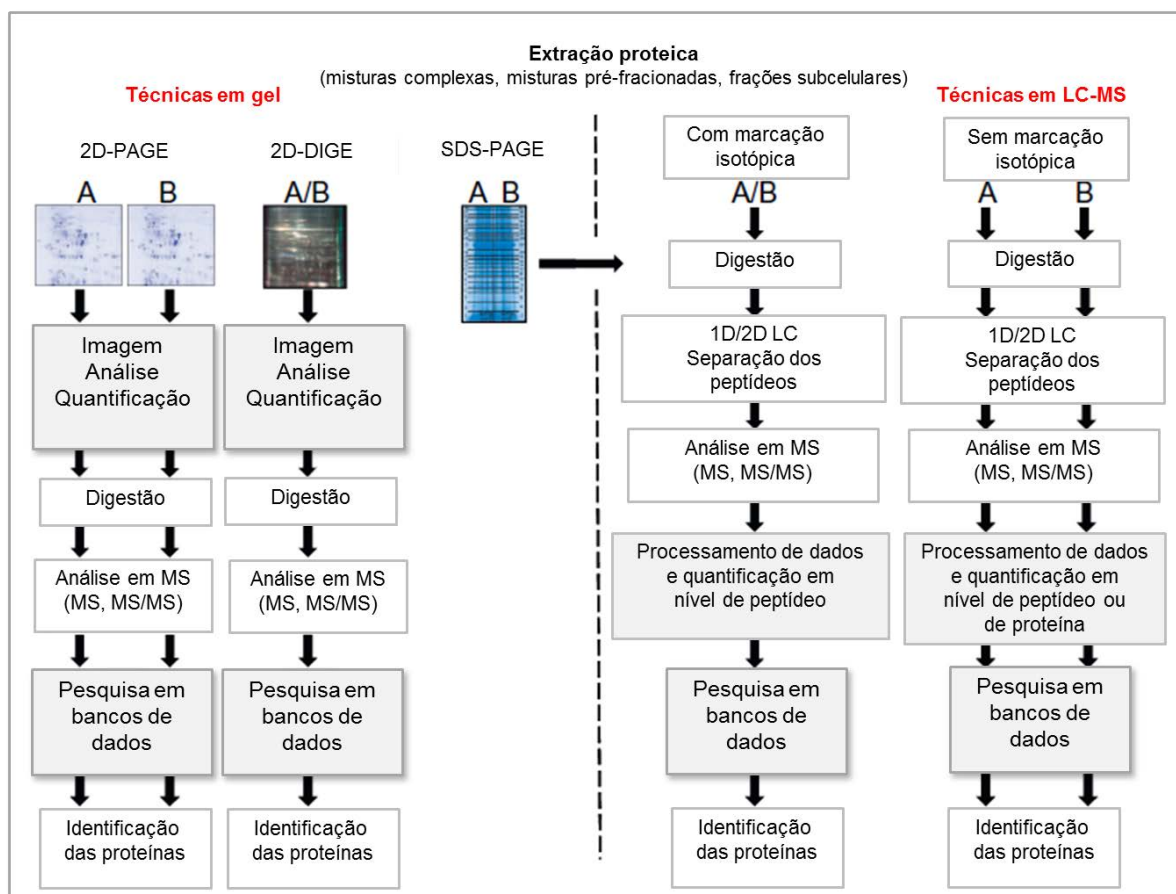


Figura 2- Representação esquemática das abordagens utilizadas no estudo proteômico. Traduzido e modificado a partir de Matros et al., 2011.

Dentre as estratégias do estudo proteômico estão a identificação e a quantificação de tantas proteínas quanto forem possíveis, numa abordagem não alvo, também chamada de *shotgun proteomics*. Para que isto seja possível, primeiramente promove-se a digestão proteolítica das proteínas em peptídeos, visando promover a solubilidade da amostra bem como favorecer a ionização. Numa segunda etapa, tais peptídeos são analisados empregando-se LC-MS/MS (WECKWERTH, 2011).

Nesta técnica, os íons dos peptídeos são primeiramente analisados como fragmentos, depois são selecionados em função da relação massa/carga (m/z) e finalmente submetidos à dissociação induzida por colisão (CID). Sob tais condições de baixa energia de dissociação, os fragmentos de íons de peptídeos apresentam um padrão previsível, fato que permite a construção e a utilização de bancos de dados de espectros teóricos de sequência de proteínas (ou de nucleotídeos) para a identificação *in silico*. Sendo assim, uma das vantagens em se usar a espectrometria de massas sequencial para a identificação de proteínas é a habilidade inerente de se sequenciar peptídeos diretamente de misturas (DELAHUNT; YATES, 2005; WECKWERTH, 2011).

Uma vez obtidos dados espectrométricos, o próximo passo é a identificação em larga escala dos peptídeos ou proteínas presentes nas amostras, através do emprego de técnicas computacionais que permitem organizar e entender as informações associadas a elas. Neste processo, conhecido como *high throughput proteomics*, os espectros podem ser identificados de acordo com diferentes tipos de abordagens, tais como (i) o sequenciamento *De Novo* a partir de dados obtidos diretamente dos espectros de massas, (ii) o uso de sequências marcadas de peptídeos (ou *peptide sequence tags*), derivadas de espectros utilizados na busca de sequências conhecidas, (iii) métodos de correlação cruzada que comparam espectros experimentais com espectros teóricos, ou (iv) comparação baseada em probabilidade, que calcula um *score* baseado na significância estatística de uma identificação feita com base na comparação de um dado candidato de peptídeo com uma sequência obtida de uma biblioteca de dados (GEER et al., 2004).

De forma geral, a identificação das proteínas se baseia na análise comparativa das sequências de peptídeos fragmentadas por MS/MS com sequências produzidas pela digestão *in silico* de proteínas provenientes de grandes bancos de dados. No entanto, para proteínas ainda desconhecidas, ou para aquelas presentes em amostras selvagens e que não possuem bancos de dados, tampouco o genoma sequenciado, esta abordagem é falha. O que se faz, neste caso é realizar a identificação com base na similaridade, e não na identidade, de sequências de peptídeos fragmentados e sequências de proteínas homólogas conhecidas de espécies filogeneticamente relacionadas (JUNQUEIRA et al., 2008). Neste contexto, diversos algoritmos e softwares computacionais têm sido desenvolvidos visando a melhoria das estratégias de identificação, uma vez que é crescente a demanda por técnicas que permitam a análise de grandes lotes de dados oriundos de amostras complexas.

O uso de tais técnicas extremamente poderosas, tanto para a identificação de enzimas como também de proteínas regulatórias, fornece, por conseguinte, uma ferramenta adicional para o estudo do metabolismo secundário em plantas. No entanto, devido à dificuldades no preparo de amostras e na identificação das proteínas, a aquisição de dados proteômicos bem como sua integração com outros dados “ômicos” acerca do metabolismo vegetal são raros (YANG et al., 2014).

1.3 Estudo metabolômico de produtos de origem vegetal

Ao passo que o metabolismo primário sintetiza, degrada e interconverte substâncias tais como açúcares, lipídeos, aminoácidos e nucleotídeos, encontradas em todas as espécies e que participam dos processos metabólicos essenciais para a sobrevivência da planta, o metabolismo secundário biossintetiza substâncias que são diferencialmente distribuídas em grupos taxonômicos do reino vegetal, constituindo uma expressão da individualidade das espécies (CROTEAU et al., 2000; DEWICK, 2009).

De forma geral, os metabólitos secundários são pequenas moléculas que são quimicamente transformadas durante o metabolismo e que, como tal, proporcionam um indicativo do estado fisiológico de um organismo. Ao contrário dos genes e proteínas, cujas funções estão sujeitas à regulação epigenética e modificações pós-translacionais, respectivamente, tais metabólitos são evidências diretas da atividade bioquímica e que, por conseguinte, podem ser mais facilmente correlacionados com o fenótipo (PATTI et al., 2012).

Ao estudo em larga escala dos metabólitos presentes em determinados organismos em diferentes estados fisiológicos dá-se o nome de metabolômica . De fato, a aquisição e análise de dados do perfil metabolômico pode efetivamente fornecer informações funcionais, permitindo uma visão ampla da situação bioquímica de um organismo além de poder ser utilizado para monitorar variações significativas no perfil dos metabólitos gerados, ou seja, os produtos finais de processos de regulação celular (WOLFENDER et al., 2010).

Devido à alta complexidade química dessas amostras, à variabilidade biológica inerente à maioria dos organismos e às limitações dinâmicas das abordagens experimentais, a análise quantitativa e qualitativa de todo o metaboloma de um sistema biológico é um objetivo muito ambicioso e ainda está longe de ser alcançado. Não há uma única técnica analítica capaz de analisar um metaboloma como um todo, obstáculo que tem sido vencido através da extração seletiva e análises paralelas usando

combinação de tecnologias para se obter uma visão mais completa possível do metaboloma. Dentre elas estão os métodos baseados em espectroscopia de infravermelho (IV), a ressonância magnética nuclear (NMR), espectrometria de massas sequencial (MS/MS), a cromatografia em camada delgada (TLC), a cromatografia líquida de alta eficiência (HPLC) acoplada aos mais variados tipos de detectores (UV, DAD, MS, NMR), a eletroforese capilar (CE) acoplada à detecção no UV, fluorescência ou MS e cromatografia gasosa com detecção por espectrometria de massas (GC-MS) (SUMNER et al., 2003; GLAUSER et al., 2013).

Atualmente, a ressonância magnética nuclear e a espectrometria de massas têm sido as duas principais técnicas utilizadas no estudo em larga escala de metabólitos secundários. Enquanto que a primeira é extremamente reprodutível e permite a quantificação absoluta dos sinais detectados, ela perde em detectabilidade e sua capacidade de identificação de substâncias em misturas complexas é limitada. Ao contrário, a habilidade da espectrometria de massas em detectar múltiplas substâncias em concentrações extremamente baixas faz dela uma ferramenta poderosa e versátil para o estudo metabolômico, especialmente quando combinada com técnicas de separação, sejam elas cromatográficas ou eletroforéticas (GLAUSER et al. 2013).

Neste contexto, o acoplamento entre a cromatografia gasosa ou a cromatografia líquida, especialmente a cromatografia líquida de ultra eficiência, com a espectrometria de massas (GC-MS e UHPLC-MS, respectivamente) permite uma detecção bidimensional, em que cada metabólito é caracterizado tanto na dimensão cromatográfica em termos de tempo de retenção, quando na espectrométrica, em termos de massa/carga (m/z). Tal hifenação tem sido uma ferramenta fundamental e extremamente poderosa em termos de detecção, quantificação e identificação de uma vasta gama de produtos de origem natural (WOLFENDER et al., 2010).

Ao se utilizar um detector de massas, o melhor ponto de partida para a identificação destas micromoléculas é a determinação inequívoca de sua composição elementar e, neste contexto, obter espectros com alta resolução de massas e alta resolução espectral é um pré-requisito. Isto torna imprescindível a utilização de espectrômetros de massas de altíssima resolução, tais como as tecnologias de tempo de voo (TOF, com resolução entre 30.000 a 40.000), a ressonância ciclotrônica de íons por transformada de Fourier (FT-ICR, com poder de resolução acima de 1.000.000) e Orbitrap, com resolução de 70.000 – m/z 200 (KIND; FIEHN, 2007; GLAUSER et al., 2013).

Apesar da utilização de técnicas de separação cromatográficas hifenadas a modernos detectores, sejam elas associadas *on line* ou *off line*, serem imprescindíveis para a realização de estudos metabolômicos e, apesar dos consideráveis avanços em termos de separação e detecção, a identificação precisa dos produtos de origem natural é ainda um desafio, uma vez que não existe uma estratégia universal para a interpretação dos espectros de massas (MS ou MS/MS), e a disponibilidade de bancos de dados contendo tais informações para estas substâncias é ainda escassa (GLAUSER et al., 2013).

1.4 *Casearia sylvestris* Sw. (Salicaceae)

A espécie *Casearia sylvestris* Sw. (Salicaceae) é um exemplo notável de como variações morfológicas ocorrem de acordo com diferentes ecossistemas de origem, sendo que tais eventos parecem estar relacionados com as adaptações ao ambiente e com a plasticidade fenotípica. Trata-se de um importante representante do gênero *Casearia* Jacq. que contém cerca de 180 espécies de distribuição neotropical (SLEUMER, 1980; MARQUETE, 2010) e com 50 delas largamente distribuídas no Brasil (MAISTRO et al., 2004). É importante destacar que, tal gênero, inicialmente classificado como pertencente à antiga família Flacourtiaceae, foi recentemente reclassificado (de acordo com método baseado em seqüências de DNA - plastídeo *rbcL*), como pertencente à família Salicaceae, tribo Samydeae (CHASE et al., 2002; JUDD, 2002).

Devido à sua alta capacidade adaptativa *C. sylvestris* encontra-se bastante disseminada na América Central e do Sul, ocorrendo desde o México até a Argentina e Uruguai (SLEUMER, 1980; LORENZI, 2002), apresentando, portanto, importante variação morfológica. No Brasil, esta espécie ocorre nos biomas Mata Atlântica (Floresta Estacional Decidual, Semi-Decidual, Floresta Ombrófila Densa, Mista e Restinga), Amazônia, Cerrado (*latu sensu*, Cerradão e Campo Cerrado), Caatinga e Pantanal (MARQUETE, 2010), dentre outras formações e ecótonos.

Morfológicamente *C. sylvestris* se caracteriza por grandes variações relacionadas ao tamanho, forma e textura das folhas, pilosidade dos ramos jovens e inflorescências, número de flores por inflorescências e comprimento dos pedicelos (TORRES; YAMAMOTO, 1986). A ocorrência de variedades morfo-anatômicas nesta espécie é relatada por Corrêa (1984), apontando o valor da adaptação para a criação de formas. Diferenças fenológicas também fornecem um importante contraponto da adaptação desta planta ao meio ambiente. Dados de Lorenzi (1992) indicam o florescimento de *C.*

sylvestris nos meses de junho a agosto, enquanto que Klein e Sleumer (1984) relatam o florescimento de julho a novembro.

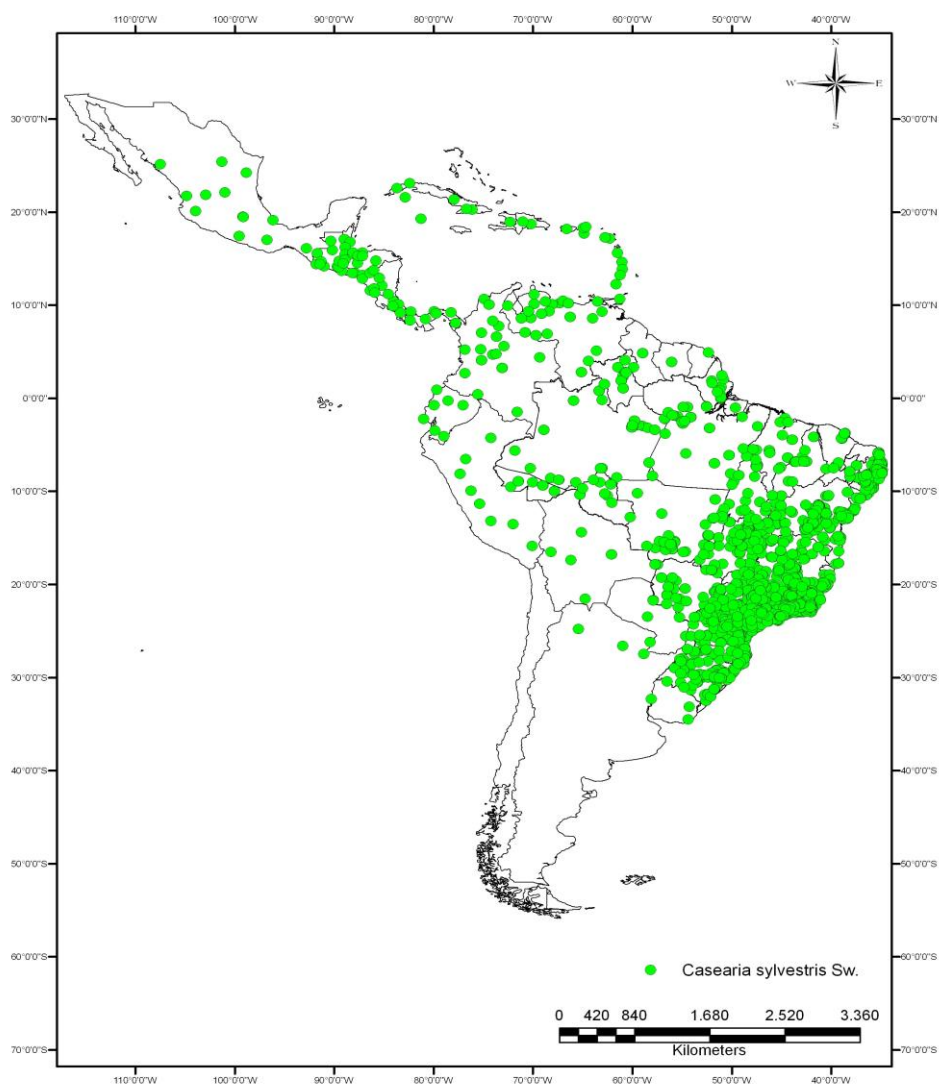


Figura 3- Distribuição geográfica de *C. sylvestris*. Fonte: Marquete, 2010.

Devido à enorme diversidade morfológica, Sleumer (1980) admite duas variedades distintas para esta espécie: *C. sylvestris* var. *sylvestris* e *C. sylvestris* var. *lingua*. Há relatos de diferenças genéticas importantes entre essas suas variedades, obtidas através de estudos de 6 populações via RAPD (DA SILVA, 2003).

Recentemente, um estudo mais extenso utilizando 9 populações e marcadores microsatélites, comparou os padrões de fluxo gênico e diversidade genética entre e dentro de tais variedades e ecossistemas, corroborando a existência de diferenças genéticas entre as duas variedades morfo-anatômicas de *C. sylvestris* características dos biomas Mata Atlântica e Cerrado, revelando a co-existência das duas variedades

em zonas ecótonas do Estado de São Paulo, onde híbridos destas duas variedades podem ocorrer. Os dados morfo-anatômicos, químicos e genéticos disponíveis sobre *C. sylvestris* indicam relação próxima entre variedade *lingua* e o bioma Cerrado e entre variedade *sylvestris* e Mata Atlântica, havendo também dados descrevendo diferenças genéticas significativas entre estas duas variedades. Em ecótonos Cerrado/Mata Atlântica essas duas variedades co-existem, havendo principalmente nessas regiões indivíduos com características intermediárias, tanto do ponto de vista morfo-anatômico quanto genético. Há indícios de que em *C. sylvestris* var. *lingua*/Cerrado ocorram predominantemente compostos fenólicos, enquanto em *C. sylvestris* var. *sylvestris*/Mata Atlântica predominam os diterpenos clerodânicos (CAVALLARI, 2008; CAVALLARI et al., 2010).

Além disso, maiores teores de rutina foram determinados em árvores coletadas em áreas de Cerrado, em detrimento daquelas coletas em áreas de Mata Atlântica (DA SILVA, 2003), enquanto o oposto parece ocorrer com os diterpenos clerodânicos (CAVALLARI, 2008).

Além da grande importância química e ecológica, esta espécie possui também notável destaque econômico e farmacológico devido aos usos populares e comerciais de *C. sylvestris*, que incluem o tratamento de úlceras e gastrites, utilização como cicatrizante e antiofídico (FERREIRA et al., 2011). Propriedades biológicas de extratos, frações e/ou substâncias isoladas desta espécie têm sido estudadas, tais como as atividades antiulcerogênica (BASILE et al., 1990; ESTEVES et al., 2005), antiofídica (BORGES et al., 2000; BORGES et al., 2001; CAVALCANTE et al., 2007;), citotóxica (DA SILVA et al., 2008; MAISTRO et al., 2004; OLIVEIRA et al., 2009; FERREIRA et al., 2010; SANTOS et al., 2010) e antiinflamatória (RASLAN et al., 2002), dentre outras.

Por este motivo, *C. sylvestris* é uma das espécies consideradas como prioritárias para conservação e manejo sustentável, uma vez que devido ao disseminado uso popular, torna-se alvo fácil para a exploração comercial e medicinal. No entanto, conforme bem destacado por Facanali (2004) em seu estudo acerca da ecologia de populações de *C. sylvestris*, são escassos os dados sobre a forma de distribuição dos indivíduos nas populações, bem como sobre os fatores que a afetam, como é a dinâmica, a regeneração e a demografia, dados essenciais para o conhecimento da estrutura de comunidades biológicas naturais e imprescindíveis para o estabelecimento de programas de manejo e conservação de espécies.

Quimicamente, tanto *C. sylvestris* quanto outras espécies do gênero *Casearia*, caracterizam-se pela presença de diterpenos clerodânicos altamente oxigenados,

marcados pela presença de insaturações e geralmente com hidroxilas esterificadas. Tais compostos, também denominados casearinas, constituem marcadores taxonômicos para *Casearia* e são amplamente encontrados nas folhas, caule, casca do caule, raiz e sementes de indivíduos do gênero (CARVALHO et al., 1998; CARVALHO et al., 2009; SANTOS et al., 2007; WANG et al., 2009).

Do gênero *Casearia* foram descritos cerca de 287 metabólitos secundários, incluindo terpenóides, fenilpropanóides, flavonoides e derivados do ácido elágico, dentre outros. No entanto a presença dos diterpenos clerodânicos é predominante, o que os tornam os constituintes representativos das espécies do gênero. Estudos fitoquímicos de espécies do gênero, principalmente *C. corymbosa*, *C. grewiifolia*, *C. membranacea* e *C. sylvestris* permitiram o isolamento e caracterização estrutural de 152 diterpenos clerodânicos, sendo que destes, 41 foram isolados de *C. sylvestris* (XIA et al., 2015).

O esqueleto básico da maioria dos diterpenos clerodânicos encontrados em espécies de *Casearia* é formado pelos anéis A e B do sistema decalínico com duas metilas, uma em C-8 (Me-17) e outra em C-9 (Me-20); uma cadeia lateral em C-9 com 6 carbonos (C-11 a C-16) que contém um dieno conjugado (o qual pode assumir três formas isoméricas); um terceiro anel C diidrofurânico/diacetálico; além de substituintes oxigenados (R1 a R5, os quais podem ser ésteres, éteres, hidroxilas ou cetonas) e que podem estar presentes nos carbonos 2, 6, 7, 18 e 19 (SANTOS, 2008), conforme pode ser visualizado na **Figura 4**, a seguir:

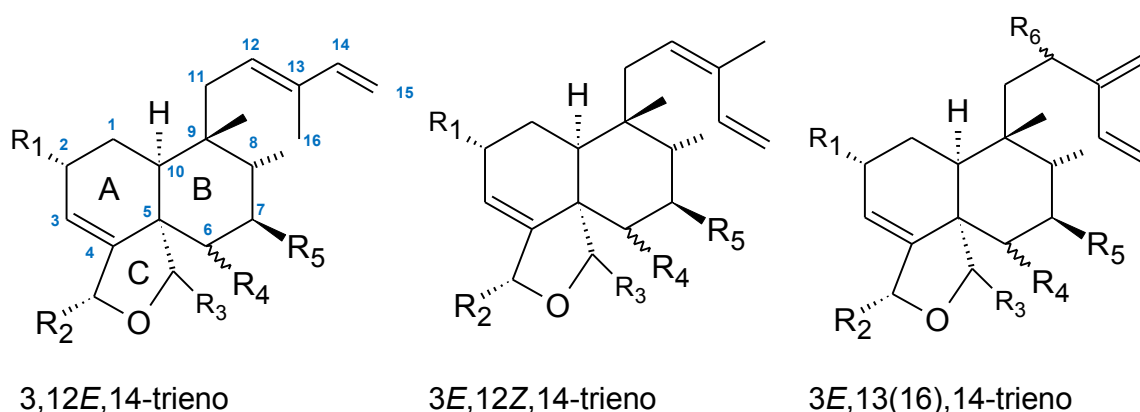


Figura 4- Esqueleto básico dos diterpenos clerodânicos e suas formas isoméricas, característicos do gênero *Casearia*.

A presença do dieno conjugado nos diterpenos clerodânicos isolados de espécies de *Casearia* permitem a detecção e diferenciação de tais substâncias por espectrofotometria na região do ultravioleta. Dentre as possíveis formas isoméricas, estão o dieno na cadeia lateral com uma ligação dupla terminal em C-14 e outra que pode diferir quanto a posição e a estereoquímica: em C-12 E ou Z ou, ainda, em C-13(16) formando uma segunda ligação dupla terminal (**Figura 4**). Dados da literatura evidenciam um padrão de absorção em que valores de $\lambda_{\text{máx}}$ de 232-235 nm ocorrem para o dieno em C-12/C-14 e de 223-229 nm (deslocamento hipsocrômico) para o dieno em C-13(16)/C-14, coerente com seu menor grau de substituição (CARVALHO et al., 2009).

Além da ocorrência dos diterpenos clerodânicos no gênero *Casearia*, merece destaque a ocorrência de compostos fenólicos, com 18 substâncias já descritas, sendo 7 delas obtidas de *C. sylvestris*, incluindo ácido elágico, ácidos 3'-O-metil elágico, ácido 3,3'-di-O-metil elágico, ácido 3-O-metil-3',4'-metilenodioxil elágico, 3,5-dimetil éter do galato de isobutílica, 3,5-dimetil éter do galato de metila e tirosol. Apenas onze flavonoides foram descritos para esse gênero, sendo eles kaempferol, kaempferol-3-O- β -glicopiranosídeo, quercetina-3-O- β -glicopiranosídeo, kaempferol-3-O- β -glicopiranosil-7-O- α -ramnopiranosídeo, quercetina-3-O- β -glicopiranosil-7-O- α -ramnopiranosídeo, kaempferol-3,7-di-O- α -ramnopiranosídeo, quercetina-3,7-di-O- α -ramnopiranosídeo, hiperosídeo, catequina, amentoflavona e 3'-O-metilquercetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 6)- β -D-glicopiranosídeo, porém nenhum foi obtido da espécie *C. sylvestris* (XIA et al., 2015).

Considerando o que já se sabe sobre a composição química de *C. sylvestris* var. *sylvestris*, na qual predominam os diterpenos clerodânicos, bem como a falta de estudos acerca do metabolismo secundário de *C. sylvestris* var. *lingua*, bem como as variações que podem ocorrer na ocorrência destas variedades dependendo dos biomas de origem, este trabalho objetiva avaliar e confirmar, do ponto de vista da constituição química micromolecular, a associação existente entre morfotipo predominante, bioma de origem e perfil metabólico em *C. sylvestris*.

2 OBJETIVOS

O objetivo principal deste trabalho é avaliar e confirmar a associação entre a produção diferencial de metabólitos secundários diterpênicos e fenólicos em *C. sylvestris* var. *sylvestris* e *C. sylvestris* var. *lingua* em função de seus biomas de origem.

Os objetivos específicos são:

- (I) Desenvolver e validar um método analítico por UHPLC-DAD-MS visando a detecção e quantificação simultânea de diterpenos clerodânicos característicos de *C. sylvestris*, e de compostos fenólicos, ainda muito pouco estudados nesta espécie;
- (II) Identificar os diterpenos clerodânicos nas amostras de *C. sylvestris* var. *sylvestris*, bem como identificar os compostos fenólicos presentes em *C. sylvestris* var. *lingua* ainda muito pouco conhecidos;
- (III) Realizar o estudo de dinâmica metabólica (ciclos circadiano e sazonal) das duas variedades em estudo;
- (IV) Realizar o estudo de variabilidade química infraespecífica com amostras de *C. sylvestris* provenientes dos biomas brasileiros Caatinga, Cerrado, Mata Atlântica, Pantanal e Pampa, mapear e correlacionar, do ponto de vista químico, a ocorrência desta espécie e suas variedades no território brasileiro;
- (V) Realizar estudos de proteômica comparada visando avançar no conhecimento acerca das enzimas diferencialmente expressas nas duas variedades em estudo.

3 MATERIAL

3.1 Equipamentos utilizados e suas especificações

- Agitador magnético com aquecimento RTC Basic, Ika Labortechnik
- Agitador tipo vortex modelo G-560E (Genie 2), Scientific Industries
- Balança analítica modelo AG204 (máx. 210 g; d = 0,1 mg), Mettler Toledo
- Balança analítica modelo BG2000 (máx. 2020 g; d = 0,1 g), Gehaka
- Balança analítica, modelo AG245 (máx. 41/210 g; d = 0,01/0,1 mg), Mettler Toledo
- Balança modelo SBA 52 (2200g/0,01g), Scaltec
- Balança modelo SBC 31 (200 g/0,0001g), Scaltec
- Banho de ultrassom modelo USC 2800, UNIQUE
- Cartucho de extração em fase sólida em sílica de fase reversa C18 (2x1 cm x 40-63 μm), Merck
- Cartucho de extração em fase sólida em poliamida, Macherey-Nagel
- Centrífuga refrigerada modelo Avanti J-25, Beckman Coulter
- Centrífuga refrigerada modelo Microfuge 22R, Beckman Coulter
- Coluna cromatográfica C18 PepMap100, 5 μm , 100 Å, 300 μm x 5 mm, Dionex
- Coluna cromatográfica capilar C18 Atlantis, 3 μm , 100 Å, 75 μm x 10 mm, Waters
- Coluna cromatográfica Kinetex C18, 2,6 μm , 100 Å, 150 x 2,1 mm, Phenomenex
- Coluna cromatográfica Luna C18, 5 μm , 100 Å, 250 x 4,6 mm, Phenomenex
- Coluna cromatográfica Luna preparativa C18, 10 μm , 150 x 21,2 mm, Phenomenex
- Coluna cromatográfica Luna preparativa C18, 10 μm , 250 x 21,2 mm, Phenomenex
- Cromatógrafo Líquido de Alta Eficiência (HPLC) equipado com duas bombas LC-20AT, amostrador automático modelo 20A, detector de arranjo de diodos SPD-M20A, controlador SCL-10A vp, forno para coluna CTO-20AC, operando o software LC Solution 1.23 SP1, Shimadzu (cromatógrafo 1)
- Cromatógrafo Líquido de Ultra Eficiência (UHPLC) modelo Ultimate 3000 RS acoplado a detector de arranjo de diodos, compartimento para coluna com aquecimento, bomba quaternária e amostrador automático, todos modelo 3000RS, Dionex (cromatógrafo 2)
- Cromatógrafo Líquido De Ultra Eficiência (UHPLC) modelo ACELLA (Thermo Fisher Scientific) acoplado a um espectrômetro de massas com fonte electrospray (ESI) e analisador OrbiTrap (Exactive Plus Orbitrap, Thermo Scientific) (cromatógrafo 3)
- Cromatógrafo Líquido de Alta Eficiência (HPLC) modos analítico e preparativo, equipado com duas bombas LC-8A, amostrador automático e manual, sistema de

controle SCL-10Avp, detector de arranjo de diodos SPD-M10Avp, operando software LC Solution 1.23 SP1, Shimadzu (cromatógrafo 4)

- Cromatógrafo Líquido De Ultra Eficiência (UHPLC) modelo Ultimate 3000 Nanoflow, acoplado a detector de massas híbrido iontrap/orbitrap, modelo MS-LTQ Orbitrap XL, Thermo Scientific / Dionex (cromatógrafo 5)
- Cuba para eletroforese modelo 45-1010-i, Peqlab Biotechnologie
- Disco de extração em fase sólida, troca aniônica (Solid Phase Extraction Disc, Empore 2252 Anion), Empore
- Equipamento para blotting Eletroblotter, Peqlab Biotechnologie
- Espectrofotômetro Ultrospec 3100 pro UV-Vis, Amersham Biosciences
- Espectrômetro de massas de alta resolução modelo micrOTOF Q II - ESI-TOF operando com uma bomba de infusão modelo Cole Parmer, vazão de 300 µl/h, Bruker Daltonics
- Espectrômetro/Espectropolarímetro de Dicroísmo Circular modelo j-815, JASCO
- Estufa de secagem de material de laboratório modelo MA 033/3, Marconi
- Estufa para esterilização e secagem modelo G4023D, Gehaka
- Filme para autoradiografia de alta resolução Hyperfilm MP (18 x 24 cm), Amersham GE Healthcare
- Filtro 30K, Amicon Ultra Cel 0,5 mL (membrana de celulose regenerada 30.000 MWCO), Millipore
- Filtro de nylon 0,22 µm x 13 mm de diâmetro, LAC
- Fonte de energia Pac 200 ou Pac 300, BioRad
- Fonte de energia PEQPower 250V, Peqlab Biotechnologie
- Freezer - 86°C modelo 92, Forma Scientific
- Liofilizador modulyo, Savant
- Microcentrífuga 6000 rpm, HT
- Microcentrífuga para tubos Eppendorfs Mini Spin Plus, Eppendorf
- Microplacas de 96 poços para leitor de ELISA
- Moinho analítico modelo A11 Basic, Ika Labortechnik
- pHmetro modelo Ultra Basic UB10, Denver Instrument
- Pipetas automáticas de 10 – 100 – 200 – 1000 µL, LabMate
- Pipetas automáticas de 10, 100, 200 e 1000 µL, LabMate
- Pipetas automáticas de 2,5 – 20 – 200 – 1000 µL, Eppendorf
- NMR 600 MHz Bruker Avance III HD 600 equipado com as probes BBO 600Mz S3 5mm com Z-gradiente (utilizada neste estudo) e uma Crioprobe com 4 canais com

controle de temperatura. O equipamento possui trocador automático de amostras, automação quanto aos procedimentos de lock, sintonia dos núcleos e shimming; Software TopSpin 3.2, Bruker

- Sistema de detecção por fluorescência/quimioluminescência Fusion SL, Peqlab Biotechnologie
- Sistema de purificação de água Milli-Q®, Millipore
- Sistema Flash-Prep-LC system Puriflash 4100®, Interchim
- Speed Vac Concentration Plus, Eppendorf
- Speed Vac modelo SPD SpeedVac, Thermo Scientific
- Termomixer Confort 1,5 mL, Eppendorf

3.2 Reagentes, solventes, adsorventes e padrões

Para as análises cromatográficas foram utilizados os solventes de grau cromatográfico metanol, etanol, hexano, acetato de etila, isopropanol e acetonitrila, todos da marca J.T.Baker, grau HPLC (Center Valley, PA, USA). Os ácidos fórmico (Êxodo Científica, Hortolândia, SP, Brasil), acético (Synth, São Paulo, SP, Brasil), trifluoroacético (Sigma Aldrich, Saint Louis, MO, USA), bem como formiato de amônio e hidróxido de amônio (Synth, São Paulo, SP, Brasil) foram de grau analítico. Já a série de derivados de ácido gálico, os galatos de metila, etila, propila, butila, pentila, hexila, heptila, octila, nonila, decila, undecila e dodecila foram sintetizados e gentilmente cedidos pelo Prof. Dr. Luis Octávio Regasini do Instituto de Biociências, Letras e Ciências Exatas da UNESP de São José do Rio Preto (SILVA et al., 2013). Também foi utilizado galato de butila, Sigma Aldrich® (Saint Louis, MO, USA). Água purificada foi obtida por um sistema Milli-Q® (Millipore).

Para as análises de NMR utilizou-se metanol- d_4 ; 99,8% + 0,05 v/v TMS (Aldrich, São Paulo, Brasil), piridina- d_5 ; 99,5% (Cambridge Isotope Laboratories, Andover, USA) e dimetil sulfóxido- d_6 ; 99,9% + 0,03 v/v TMS (Aldrich, São Paulo, Brasil).

Para o isolamento de substâncias foram utilizados os adsorventes sílica gel de 40-63 μm , 60 Å (Merck, Darmstadt, Germany), sílica C18 de 40-63 μm , 60 Å (Merck, Darmstadt, Germany), Celite (Synth, São Paulo, SP, Brasil), sílica C18 de 50 μm (Puriflash IR 60, Interchrom) e poliamida (Sigma Aldrich, Saint Louis, MO, USA).

Para os estudos proteômicos, utilizou-se um kit para doseamento de proteínas com ácido bicinchonínico Pierce BCA Protein Assay Kit (Thermo Scientific), mistura de padrões de pesos moleculares para SDS-PAGE, PageRuler *prestained protein ladder*

10-170 KDa, Pierce Protein Biology Products (Thermo Scientific), membrana de nitrocelulose para ECL *Western blotting*, Hybond ECL 0,45 μ m, 30 x 3 m (Amersham GE Healthcare), revelador para autorradiografia Ultrafin liquid B/W negative Developer, (Tetenal) e fixador para autorradiografia Superfix Plus – Rapid Fixer (Tetenal). Ainda para o estudo proteômico, todos os demais reagentes e soluções utilizadas nos protocolos FASP e Pós-FASP, extração de proteínas, SDS-PAGE e *Western blotting*, e quantificação por BCA e Amido Black foram de grau analítico e estão descritos no Apêndice I.

3.3 Softwares

Para o tratamento e aquisição dos dados cromatográficos utilizou-se o software Chromeleon (versão 6.80SR12, Dionex). Para a manipulação e tratamento dos dados para as análises quimiométricas utilizaram-se OriginPro 8 (OriginLab, USA), Microsoft Excel 2010 (Microsoft, USA), Matlab® v. 7.5 (MathWorks Inc., Natick, USA) com um algoritmo implementado (www.models.kvl.dk/source/DTW_COW/index.asp) e Pirouette (Versão 4.0 rev. 2, Infometrix, Bothell, USA).

Para a interpretação dos espectros de NMR utilizou-se o software MestReNova 6.0.2-5475 (Mestrelab Research, Santiago de Compostela, Espanha). Já para a análise dos dados espectrométricos utilizou-se os softwares ESI Compass 1.3 for micrOTOF DataAnalysis version 4.0 (Bruker Daltonic, USA) e Xcalibur versão 2.2 (Thermo Fischer Scientific). Para a desreplicação dos dados utilizou-se os softwares Protowizard (KESSNER et al., 2008) e MZmine versão 2.10 (PLUSKAL et al., 2010).

Para a etapa de bioinformática do estudo proteômico foram utilizados os softwares Proteomic® (SPECHT et al., 2011), OMSSA (versão 2.1.9, GERR et al., 2004), X! Tandem (versão 2012.10.01.1, CRAIG et al., 2004), PEAKS (versão 2.4, MA et al., 2003) e BLAST (algoritmo BLAST-P, disponível em <http://blast.ncbi.nlm.nih.gov>).

3.4 Material vegetal

Para o estudo de dinâmica metabólica (ciclos circadiano e sazonal) e estudo proteômico, as amostras das duas variedades de *C. sylvestris* foram coletadas na cidade de Araraquara, SP, Brasil. Para os estudos de variabilidade infraespecífica, as amostras foram coletadas em diversas cidades brasileiras. De forma geral, foram

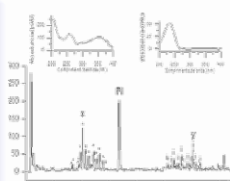
coletadas apenas as folhas das árvores, as quais foram imediatamente processadas conforme a finalidade do estudo.

Todas as amostras coletadas tiveram a identificação confirmada pela Dra. Roseli B. Torres, botânica especialista na antiga família Flacourtiaceae e curadora do Herbário do Instituto Agronômico de Campinas (IAC). Amostras testemunho da var. *lingua* e da var. *sylvestris* foram depositadas no referido herbário sob os códigos IAC 55839 (ARA 13, P.C.P. Bueno et al. 56) e IAC 55840 (ARA 14, P.C.P. Bueno 57), respectivamente.

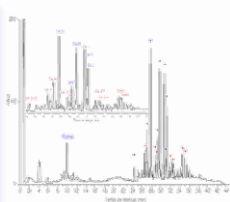
Para cada estudo, o procedimento de coleta e tratamento de tais amostras serão detalhadamente descritos na seção correspondente em Métodos.

4 MÉTODOS

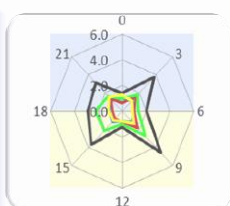
Para melhor exposição e entendimento das atividades de pesquisa desenvolvidas neste trabalho, a apresentação desta seção, bem como a de Resultados e Discussão, será dividida em 5 partes, conforme estratégia experimental descrita abaixo:



Parte 1 - Desenvolvimento de método cromatográfico utilizando UHPLC-DAD para análise simultânea de compostos fenólicos e diterpenos clerodânicos em *C. sylvestris* var. *sylvestris* e *C. sylvestris* var. *lingua*.



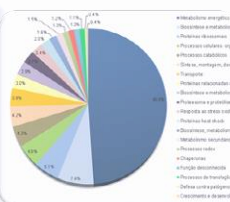
Parte 2 - Estudo metabolômico de *C. sylvestris* objetivando análise qualitativa das duas variedades em estudo, bem como a identificação dos compostos fenólicos presentes na variedade *lingua*, ainda muito pouco estudada quimicamente.



Parte 3 - Estudos de dinâmica metabólica e variação sazonal das duas variedades para a avaliação da oscilação temporal dos produtos do metabolismo secundário em estudo.



Parte 4 - Estudos de variabilidade química infraespecífica de *C. sylvestris* com o objetivo de se correlacionar o perfil químico predominante das variedades *lingua* e *sylvestris* e seus locais de origem.



Parte 5 - Estudo proteômico diferencial de *C. sylvestris* visando avançar no conhecimento acerca das enzimas expressas nas duas variedades em estudo e correlacioná-las com o perfil metabólico observado.

4.1 PARTE 1- Desenvolvimento de método cromatográfico por UHPLC-DAD

4.1.1 Composição da amostra representativa de *C. sylvestris* (ARCS)

Uma amostra representativa da diversidade micromolecular de *C. sylvestris* foi elaborada amostrando-se quatro indivíduos da variedade *lingua*, todos provenientes da Estação Experimental de Araraquara (Rua Vereador Mário Ananinas s/n, Chácara Flora, Araraquara), e quatro da variedade *sylvestris*, três provenientes do Horto de Plantas Medicinais e Tóxicas da Faculdade de Ciências Farmacêuticas de Araraquara-UNESP e uma cultivada no Departamento de Estradas de Rodagem - DER (Rua Castro Alves, 1253, Jardim Santa Lucia, Araraquara).

A amostra foi constituída por folhas dos oito acessos indicados acima, as quais foram desidratadas a 40 °C por 36 horas, trituradas em moinho analítico e extraídas em banho de ultrassom numa proporção de 50 mg de pó para 1 mL de uma mistura de metanol/água 80:20 v/v. Em seguida o extrato foi submetido à uma extração em fase sólida em sílica de fase reversa (cartucho C18 Merck, 2x1 cm x 40-63 µm), utilizando uma mistura de metanol/água 98:02 (v/v) como eluente (SANTOS, 2008). O eluato foi então filtrado em membrana de nylon de 0,22 µm e as amostras analisadas por HPLC-DAD (cromatógrafo 1), em gradiente exploratório utilizando como fase móvel água (A) e metanol (B) conforme indicado na **Figura 5**. Utilizou-se uma coluna Luna C18 (250 x 4,6 mm, 5 µm, 100 Å, Phenomenex), vazão de 0,800 mL/min, volume de injeção de 10 µL e faixa de detecção de 200 a 800 nm.

A partir dos resultados dessas análises foi estimada a proporção a usar de cada uma dessas oito amostras para se obter uma amostra representativa da diversidade química de *C. sylvestris* (ARCS).

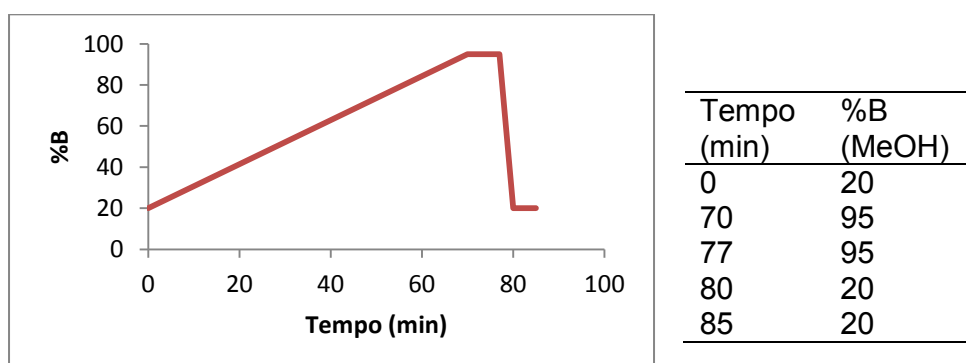


Figura 5- Gradiente exploratório utilizado no processo de composição da amostra representativa de *C. sylvestris* (ARCS).

4.1.2 Definição de um método provisório por UHPLC-DAD

Uma vez constituída a amostra representativa de *C. sylvestris* (ARCS), um método provisório por UHPLC-DAD foi estabelecido (cromatógrafo 2) utilizando-se uma coluna de fase reversa com núcleo sólido (coluna Core-Shell Kinetex de 150 mm x 2,1 mm, 2,6 μ m, Phenomenex), com pré-coluna de mesma fase estacionária. Com base nos primeiros resultados obtidos por HPLC (cromatógrafo 1) compôs-se uma fase móvel ternária de água (A), acetonitrila (B) e metanol (C). O gradiente multilinear de 37 minutos está indicado na **Figura 6** a seguir. Utilizou-se uma vazão de 0,250 mL/min, volume de injeção de 2 μ L e faixa de detecção de 200-800 nm.

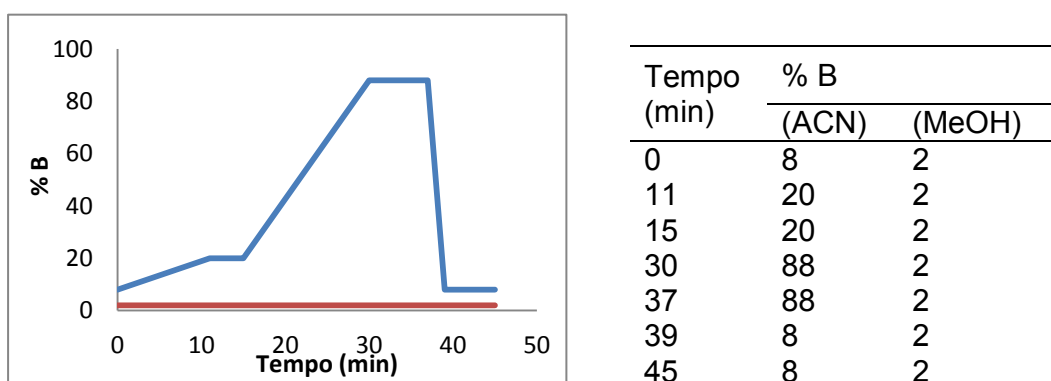


Figura 6- Método provisório por UHPLC-DAD para análise de *C. sylvestris*.

Ainda, utilizando esta mesma condição e a amostra padronizada, foi avaliada a influência do uso de apenas acetonitrila contendo 0,1% de ácido fórmico na fase B. A condição cromatográfica desenvolvida nessa etapa foi utilizada para analisar as amostras obtidas no desenvolvimento das condições de extração.

4.1.3 Otimização das condições de extração das amostras

4.1.3.1 Planejamento Simplex Centróide para definição do solvente extrator

A definição do melhor solvente extrator para a análise de *C. sylvestris* foi feita utilizando-se o modelo Simplex Centróide. Foram testadas 10 combinações de solventes extratores, com o ponto central em triplicata, conforme a **Tabela 1** a seguir. Cada amostra (ARCS) foi extraída na proporção de 50 mg de pó para um mililitro de cada proporção de solvente extrator e analisadas utilizando o método provisório por UHPLC-DAD descrito anteriormente. Como resposta, determinou-se a média do número de

picos observados em 235 e 254 nm, ou seja, nos comprimentos de onda ótimos para a detecção dos diterpenos clerodânicos e compostos fenólicos, respectivamente.

Tabela 1- Proporções de cada solvente nas misturas extratoras testadas no planejamento Simplex Centroides.

Experimento	Água	Etanol	Isopropanol
1	1	0	0
2	0	1	0
3	0	0	1
4	1/2	1/2	0
5	0	1/2	1/2
6	1/2	0	1/2
7	2/3	1/6	1/6
8	1/6	2/3	1/6
9	1/6	1/6	2/3
10 (n=3)	1/3	1/3	1/3

Com os resultados deste primeiro planejamento, definiu-se a composição de solventes capaz de extrair maior número de substâncias das folhas de *C. sylvestris*, isto é, que forneceria o maior número de picos cromatográficos. A seguir analisou-se a amostra (n=3) empregando-se a proporção de solventes correspondente ao ponto ótimo.

4.1.3.2 Estudo da influência do pH na etapa de extração

A influência de diferentes pHs na extração foi avaliada pela adição de 0,1% dos ácidos acético (pka 4,76), fórmico (pka 3,77) e trifluoroacético (pka 0,23) à mistura de solventes, sendo também avaliada a influência do tampão formiato de amônio 100 mM, pH 4,5. As amostras (ARCS) foram extraídas em triplicata conforme definido anteriormente, na proporção de 50 mg de pó para um mL da mistura solvente extrator acidificado ou tamponado.

4.1.4 Otimização da condição de eluição cromatográfica

A otimização da eluição cromatográfica iniciou-se através de um Planejamento Fatorial Fracionário 2^{4-1} com ponto central (nível de significância de 0,05) para seleção de variáveis. Para tanto, primeiramente simplificou-se o método preliminar, que passou a conter apenas água e acetonitrila (excluindo-se o metanol que estava presente no

método provisório) e ajustou-se a vazão da fase móvel para 0,400 mL/min. A **Figura 7** a seguir evidencia o gradiente simplificado utilizado.

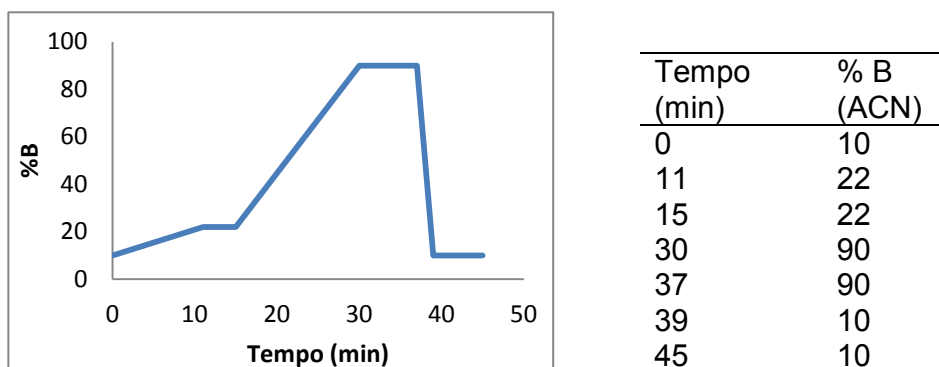


Figura 7- Gradiente simplificado para a otimização do método cromatográfico por UHPLC-DAD.

O planejamento, bem como as variáveis selecionadas para o estudo seguem descritas na **Tabela 2**. Como resposta, da mesma forma que no planejamento Simplex Centróide, contabilizou-se e calculou-se a média do número de picos observados nos cromatogramas em 235 e 254 nm.

Uma vez selecionadas as variáveis que mais influenciavam na separação cromatográfica, foi executado, em seguida, um Planejamento Composto Central para definição da melhor condição cromatográfica para a análise das duas variedades de *C. sylvestris*. O planejamento, bem como as variáveis selecionadas para o estudo, seguem descritas na **Tabela 3**. A temperatura foi mantida a 35°C. Como resposta, contabilizou-se e calculou-se a média do número de picos observados nos cromatogramas em 235 e 254 nm. Finalmente, para a seleção da melhor condição avaliou-se a condição geral dos cromatogramas e por fim pequenos ajustes foram feitos desta vez avaliando a resolução global dos picos presentes.

Tabela 2- Planejamento e parâmetros do Planejamento Fatorial Fracionário 2^{4-1} com ponto central.

Ensaio	Níveis dos fatores				Valores codificados			
	X ₁	X ₂	X ₃	X ₄	Vazão X ₁ , mL/min	Temperatura X ₂ , °C	Gradiente 1* X ₃ , min	Gradiente 2** X ₄ , min
1	-1	-1	-1	-1	0,300	30	6	10
2	1	-1	-1	1	0,500	30	6	20
3	-1	1	-1	1	0,300	40	6	20
4	1	1	-1	-1	0,500	40	6	10
5	-1	-1	1	1	0,300	30	16	20
6	1	-1	1	-1	0,500	30	16	10
7	-1	1	1	-1	0,300	40	16	10
8	1	1	1	1	0,500	40	16	20
pc1	0	0	0	0	0,400	35	11	15
pc2	0	0	0	0	0,400	35	11	15
pc3	0	0	0	0	0,400	35	11	15

* Tempo do primeiro segmento do gradiente, a partir da condição inicial de B.

** Tempo do segundo segmento do gradiente.

Tabela 3 - Planejamento Composto Central e variáveis selecionadas para a definição da melhor condição cromatográfica.

Ensaio	Níveis dos fatores		Valores codificados	
	X ₁	X ₂	Vazão X ₁ , mL/min	Gradiente 2 X ₂ , min
1	-1	-1	0,200	10
2	1	-1	0,600	10
3	-1	1	0,200	20
4	1	1	0,600	20
5	-1,41	0	0,120	15
6	1,41	0	0,600	15
7	0	-1,41	0,400	7,93
8	0	1,41	0,400	22,07
9	0	0	0,400	15
10	0	0	0,400	15

4.1.5 Escolha do padrão interno

A escolha do padrão interno foi feita injetando-se na condição cromatográfica final uma série homóloga de derivados esterificados do ácido gálico, conforme descrito na seção Materiais. Para cada derivado preparou-se uma solução a 2 mg/mL e transferiu-se 100 μ L de cada uma das soluções para um frasco de 1,5 mL, perfazendo uma única solução contendo a mistura de 150 μ g/mL de cada galato. Selecionou-se então, o galato de alquila cujo tempo de retenção não coincidiu com nenhum pico da amostra representativa de *C. sylvestris*.

Avaliou-se também o efeito das diferentes matrizes (folhas desidratadas da variedade *lingua* (CSL), *sylvestris* (CSS) e da amostra representativa - ARCS) sobre o padrão interno. Para tanto preparou-se 25 mL de soluções do padrão interno em cinco concentrações diferentes, ou seja, a 0,3, 0,4, 0,5, 0,6 e 0,7 mg/mL. Seguiu-se então o mesmo procedimento descrito para o preparo das amostras: transferiu-se, para tubos Eppendorf de 1,5 mL, 50 mg de cada matriz (CSL, CSS e ARCS) em quintuplicada e adicionou-se 1 mL de cada solução do padrão interno (nas cinco concentrações). Extraíu-se em banho de ultrasson por 1h, agitando-se manualmente a cada 20 minutos. Terminada a extração, centrifugou-se os tubos por 5 minutos a 5.000 x g e com o auxílio de uma pipeta automática, filtrou-se em filtro de nylon (0,22 μ m e 15 mm de diâmetro) uma alíquota de 0,7 mL do sobrenadante diretamente para o frasco de 1,5 mL. Avaliou-se o efeito de cada matriz através de um gráfico da concentração do padrão interno em função da respectiva área nas diferentes matrizes.

4.1.6 Avaliação da precisão

A precisão instrumental foi avaliada pela injeção em sextuplicata da solução de padrão interno (galato de butila 0,5 mg/mL) e detecção a 254 nm. Calculou-se, portanto, os desvios-padrão relativos (DPR) entre os tempos de retenção e também das áreas do pico do padrão interno. Já para a determinação da repetibilidade e precisão intermediária, injetou-se, em dois dias diferentes, seis replicatas independentes da amostra representativa de *C. sylvestris* (ARCS) contendo a solução de padrão interno de 0,5 mg/mL. A partir dos cromatogramas resultantes, foram seleccionados dois picos (**Figura 16A**), um representativo dos compostos fenólicos (flavonoides glicosilados, detectados a 254 nm) e outro dos diterpenos clerodânicos (detectados a 235 nm). Em seguida, calculou-se a relação entre as suas áreas contra a área do padrão interno

(detectada a 254 nm) e os valores obtidos foram usados para cálculos dos desvios-padrão relativos.

Ainda, para a avaliação da precisão intermediária, a comparação entre os dois dias de análise foi realizada pelo Teste F (LEITE, 2002; INSTITUTO NACIONAL DE METROLOGIA, 2010). Neste teste determinou-se a variância (S^2) dos diferentes conjuntos de medidas e aplicou-se a **Equação 1** a seguir:

Equação 1

$$F = \frac{\text{Maior variância}}{\text{Menor variância}}$$

Em seguida comparou-se o valor numérico obtido com o valor tabelado correspondente ao número de graus de liberdade do teste (n-1) de acordo com a **Tabela 4** a seguir, para verificar se:

- o resultado da equação foi menor que o valor tabelado, indicativo de que não houve diferença significativa entre as duas precisões no nível de 5%.
- o resultado foi maior que o valor tabelado, indicativo de que a diferença entre as duas precisões é significativa.

Tabela 4- Distribuição F – 5% de probabilidade.

Graus de liberdade menor variância	Graus de liberdade maior variância									
	1	2	3	4	5	6	7	8	9	10
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	3,79	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	4,21	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98

4.1.7 Avaliação da linearidade e obtenção de curvas analíticas

Para a determinação da linearidade foram elaboradas curvas analíticas a partir de diluições seriadas de uma solução-mãe contendo rutina (representando os flavonoides glicosilados), caseargrewiina F (representando os diterpenos clerodânicos) e o padrão interno na concentração de 0,5 mg/mL.

A curva analítica compreendeu uma faixa de concentrações de 1,25 a 400% e foi elaborada da seguinte forma: para um tubo Eppendorf 2,0 mL transferiram-se, respectivamente, 3,0 mg rutina e 1,5 mg de caseargrewiina F e homogeneizou-se com 1,5 mL da solução extratora (água/etanol/isopropanol 5:3:2) contendo 0,5 mg/mL do padrão interno galato de butila. Em seguida, desta solução-mãe sucederam uma série de diluições fornecendo as seguintes concentrações, em $\mu\text{g/mL}$, para cada padrão:

- rutina: 6,25; 62,5; 125; 250; 375; 500; 750; 1000; 1500 e 2000

- caseargrewiina F: 3,125; 31,25; 62,5; 125; 187,5; 250; 375; 500; 750; 1000

Finalmente, um volume de 2,0 μL de cada solução foi injetado em triplicata no cromatógrafo, fornecendo, desta forma, as áreas para a elaboração das curvas analíticas em função do padrão interno. A linearidade foi obtida através do coeficiente de correlação linear (R^2) da média das três curvas, para os dois padrões em estudo.

4.2 PARTE 2- Estudo metabolômico de *C. sylvestris*

4.2.1 Obtenção de padrões de diterpenos clerodânicos (casearinas)

Para o isolamento de casearinas partiu-se de uma fração rica em casearinas obtida a partir de um extrato bruto etanólico previamente preparado por Santos (2008).

Em síntese, cerca de 20 Kg de folhas de *C. sylvestris* foram maceradas exaustivamente em etanol, obtendo-se cerca de 200 L de extrato, o qual foi concentrado resultando em 1.540 g de extrato bruto. Em seguida, 462,6 g deste extrato foi submetido à uma extração em fase sólida utilizando uma mistura de sílica gel/carvão ativo na proporção de 1:1. Três frações foram obtidas: uma em hexano/acetato de etila 95:05 (EFS1), outra em acetato de etila (EFS2, rica em casearinas) e a última em metanol (SANTOS et al., 2008).

Para este trabalho, o procedimento experimental iniciou-se com o fracionamento desta fração EFS2 por cromatografia em coluna de fase normal. Para tanto, 28 g da fração foram solubilizados em uma mistura de hexano/acetato de etila/isopropanol 85:14:1 e aplicados em coluna de vidro contendo sílica gel (20 x 10 cm; 40-63 μ m; poro 60 Å; 832,0 g). A eluição foi realizada sob vácuo, utilizando diferentes misturas de hexano, acetato de etila e isopropanol, conforme pode ser visualizado na **Tabela 5** a seguir.

Tabela 5- Fracionamento da fração EFS2 por cromatografia em coluna de fase normal.

Fração	Solvente - Proporção	Volume (mL)	Subfrações
Vm	Hex/AcOEt/IPA 85:14:1	800	Vm
1	Hex/AcOEt/IPA 85:14:1	3 x 1000	1.1 a 1.3
2	Hex/AcOEt/IPA 78/20,5/1,5	5 x 600	2.1 a 2.5
3	Hex/AcOEt/IPA 76/22,4/1,6	5 x 600	3.1 a 3.5
4	Hex/AcOEt/IPA 73/25,2/1,8	5 x 600	4.1 a 4.5
5	AcOEt	1000	AcOEt
6	AcOEt/MeOH 1:1	1000	AcOEt/MeOH 1:1
7	MeOH	1000	MeOH

Vm: volume morto; Hex: hexano; AcOEt: acetato de etila; IPA: isopropanol; MeOH: metanol

As subfrações obtidas foram concentradas em rotaevaporador e secas com o auxílio de ar comprimido. A **Tabela 6** a seguir resume as massas obtidas.

Tabela 6- Massas das subfrações obtidas pelo fracionamento da fração EFS2.

Subfração	Massa (g)	% da fração	Subfração	Massa (g)	% da fração
Vm	0,442	1,58	3.3	1,612	5,76
1.1	...	...	3.4	0,908	3,24
1.2	...	...	3.5	0,857	3,06
1.3	...	...	4.1	1,040	3,71
2.1	0,960	3,43	4.2	1,448	5,17
2.2	0,541	1,93	4.3	1,168	4,17
2.3	0,582	2,08	4.4	0,628	2,24
2.4	0,679	2,43	4.5	0,530	1,89
2.5	1,219	4,35	sf. AcOEt	3,392	12,11
3.1	2,187	7,81	sf..AcOEt/MeOH 1:1	2,196	7,84
3.2	2,380	8,50	sf. MeOH	...	...

Total recuperado: 22,77 g (equivalente a 81,32% da fração EFS2 aplicada); sf. AcOEt: subfração em acetato de etila; sf. AcOEt/MeOH: subfração em acetato de etila e metanol 1:1; sf. MEON: subfração metanólica.

Em seguida as subfrações 2.1, 2.5, 3.1, 3.5, 4.1, 4.5 e metanólica foram analisadas por HPLC-DAD (cromatógrafo 1) utilizando como referência a fração 8FA do trabalho de Santos (2008), de onde havia sido isolada a casearina X, de interesse para este trabalho. Tais amostras, na concentração aproximada de 1 mg/mL em metanol foram analisadas em gradiente água/metanol conforme indicado na **Figura 8**. Utilizou-se uma coluna Luna C18 (Phenomenex, 5 μ m, 100 Å, 250 x 4,6 mm), vazão de 1,0 mL/min, volume de injeção de 10 μ L e faixa de detecção de 200-400 nm e monitoramento em 235 nm.

A partir de então, a fração 3.1, de menor complexidade e com perfil semelhante à fração 8FA e também a fração 3.4, foram submetidas à cromatografia líquida de alta eficiência preparativa para o isolamento de casearinas. O pré-tratamento dos 2,187 g da fração 3.1 e 0,908 da fração 3.4 incluiu uma extração em fase sólida em fase reversa (3 x 5 cm, sílica C-18 Merck®; 40-63 μ m, 20 g,) usando como eluente metanol/água (90:10, v/v, 200 mL). Para a fração 3.1, após evaporação do solvente, o resíduo seco (2,0 g) foi ressuspendido em 100 mL de metanol na proporção de 20 mg/mL e para que a amostra estivesse com uma concentração inicial de solvente o mais próximo possível da concentração inicial da fase móvel, gotejou-se água até que houvesse turvação da solução (cerca de 5%). Foram realizadas sucessivas injeções de 5 mL em uma coluna preparativa Luna C18 (250 x 21,2 mm, 10 μ m, Phenomenex), em condição isocrática de 73% de metanol durante 50 minutos, a uma vazão de 15 mL/min. O mesmo procedimento foi realizado para a fração 3.4, porém com a amostra preparada na concentração de 65 mg/mL (934 mg da fração 3.4 em 15 mL de metanol) e volume de

injeção de 2 mL. Para a identificação das casearinas, utilizou-se dados dos espectros de absorção no UV e de ressonância magnética nuclear de ^{13}C e ^1H , bem como comparação com os dados da literatura.

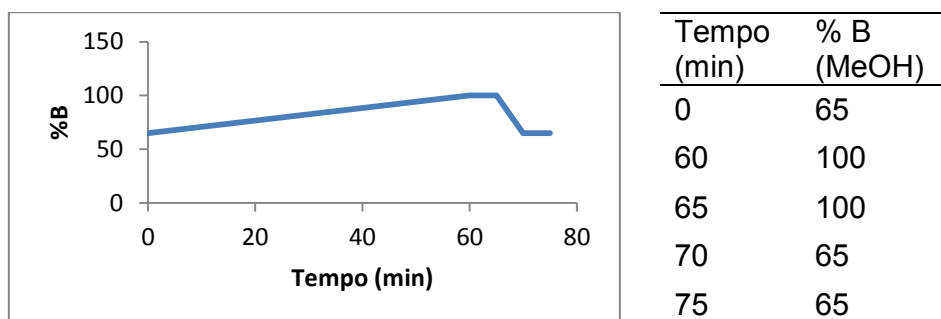


Figura 8- Gradiente de 65-100% de B utilizado para monitoramento das frações ricas em casearinas.

4.2.2 Identificação de casearinas em *C. sylvestris* var. *sylvestris*

4.2.2.1 Análise qualitativa utilizando UHPLC-DAD-ESI-Orbitrap

A identificação dos diterpenos clerodânicos conhecidos (desreplicação) presentes na amostra de *C. sylvestris* foi feita partindo-se do extrato bruto liofilizado obtido de folhas da variedade *sylvestris* (EBLS), coletadas durante um ano nos experimentos de dinâmica metabólica na cidade de Araraquara (conforme descrito no item 4.3.3). O extrato foi preparado na concentração de 15 mg/mL e analisado nas condições analíticas já definidas para este trabalho (Apêndice II), porém utilizando UHPLC-DAD-ESI-Orbitrap (cromatógrafo 3), gentilmente cedido pelo grupo de pesquisa AsterBiochem da Faculdade de Ciências Farmacêuticas de Ribeirão Preto (FCFRP-USP). Para as análises no espectrômetro de massas selecionou-se o modo alternado de ionização, ou seja, modos positivo e negativo com fragmentação em ambas polaridades. As demais condições foram estabelecidas como se segue: temperatura do capilar e da probe em 300 °C; voltagem do spray de 3,6 kV (modo positivo) e 3,2 kV (modo negativo); faixa de detecção do scan de 150 a 2000 Da; resolução full scan: 70.000 (200 m/z – 35.000 para cada modo de ionização); resolução fragmentação: 35.000 (200 m/z); HCD: 35 unid.; gás: nitrogênio, 300 °C.

A análise qualitativa (desreplicação) foi conduzida de acordo com o trabalho de Chagas-Paula et al. (2015) com pequenas modificações. Os dados espectrais brutos foram exportados para o software Protowizard (KESSNER et al., 2008), onde foram

divididos em dois conjuntos contemplando os modos positivo e negativo. E em seguida os dados foram convertidos em formato .mzML e exportados para o software MZmine 2.10 (PLUSKAL et al., 2010). Realizou-se então a desconvolução dos dados (*threshold* de 5%; faixa de tempo de retenção de 12 s; altura mínima relativa de 15%), eliminação dos isótopos (tolerância de m/z de 5 ppm; tolerância de tempo de retenção de 12s), alinhamento (tolerância de m/z de 5 ppm; tolerância de tempo de retenção de 30 s; peso para m/z de 15 e para tempo de retenção de 10), preenchimento dos espaços (intensidade, m/z e tempo de retenção de 1%, 5 ppm e 12s, respectivamente). Finalmente, utilizando os arquivos gerados em .csv, realizou-se a desreplicação dos dados utilizando um banco de dados de casearinas construído para este trabalho (Apêndice III).

4.2.2.2 Validação dos dados obtidos

Para a validação dos dados obtidos, isolaram-se padrões de casearinas, porém em pequena quantidade, apenas para identificação por espectrometria de massas, utilizando-se a mesma amostra de *C. sylvestris* var. *sylvestris*. O extrato foi preparado na concentração de 16 mg/mL e 5 μ L desta amostra foram injetados no UHPLC-DAD nas condições analíticas já definidas para este trabalho (Apêndice II). Foram realizadas 25 injeções sequenciais e para cada uma delas foram coletadas 40 frações, num intervalo de 0,2 minutos cada uma. As frações coletadas sucessivamente contendo as casearinas foram então imediatamente secas em *speed vac* e solubilizadas em 50 μ L de metanol/água 1:1 para obtenção dos espectros de massas de alta resolução por inserção direta. Para tanto, utilizou-se um espectrômetro de massas modelo micrOTOF Q II - ESI-TOF, gentilmente cedido pelo Núcleo de Pesquisas de Produtos Naturais e Sintéticos – NPPNS, da Faculdade de Ciências Farmacêuticas de Ribeirão Preto (FCFRP-USP). Os parâmetros utilizados foram os seguintes: modo de ionização positivo, calibração interna Na-TFA a 10 mg/ml (TOF); end plate: -500 Volts; capilar: 3500 Volts; temperatura e vazão do gás de secagem (nitrogênio) 180° C e 4L/min, respectivamente; pressão do gás de nebulização (nitrogênio): 0,4 bar.

4.2.3 Identificação dos compostos fenólicos em *C. sylvestris* var. *lingua*

4.2.3.1 Desreplificação por HRMS utilizando substâncias isoladas

Para o isolamento de compostos fenólicos, 3 g do extrato bruto liofilizado da amostra de *C. sylvestris* var. *lingua* (EBLL, obtido de acordo com o item 4.3.3) foi disperso em 7 g de C18 (Puriflash, 50 μ m) e 2 g de celite (Synth). O pó obtido foi então transferido para um cartucho de extração e este acoplado à mais dois cartuchos de separação de C18 (Puriflash C18HQ 15 μ m, 35g, Flashscolumn). Foi utilizado um sistema de cromatografia de media pressão Puriflash 4100 (Interchim), operando sob o seguinte gradiente de água (A) e metanol (B): de 0 a 8 minutos a 5% de B; de 8 a 35 minutos de 5 a 25% de B; de 35 a 65 minutos de 25 a 100% de B; de 65 a 90 minutos a 100% de B e finalmente de 90 a 110 minutos 100% de B. Utilizou-se vazão de 8 mL/min, e detecção no UV/Vis, na faixa de 250 – 600 nm. Ao fim do procedimento foram coletadas 113 frações cujos volumes variaram de 3 a 15 mL, dependendo da detecção dos picos. Após redução do solvente, apenas as frações que tiveram massa considerável foram avaliadas por UHPLC sob um gradiente rápido de 22 minutos composto de água (A) e acetonitrila (B), conforme indicado na **Figura 9**. Utilizou-se vazão de 0,400 mL/min, a coluna Kinetex C18 (2,6 μ m, 100 Å, 150 x 2,1 mm) e detecção no UV/Vis na faixa de 200 a 800 nm, monitorando em 254 nm.

Apenas as frações 2, 3 e 4 apresentaram os compostos fenólicos de interesse e logo após serem liofilizadas foram reunidas sob o código FRCF (fração rica em compostos fenólicos). Em seguida a FRCF foi submetida à extração em fase sólida (EFS) com poliamida para a remoção da grande quantidade de taninos presente. Para tanto, utilizou-se uma coluna com 4 g de poliamida (3,5 cm de altura 2,7 cm de diâmetro) e três extrações sucessivas utilizando 200, 400 e 600 mg de amostra, obtendo-se três subfrações eluídas sucessivamente com 75 mL de metanol/água 1:1, metanol e metanol contendo 5% de hidróxido de amônio. As subfrações foram reunidas e analisadas no gradiente de 22 minutos (**Figura 9**).

O isolamento dos compostos fenólicos foi feito utilizando-se a subfração metanólica. Para tanto, 113 mg desta subfração foram solubilizados em 6 mL de metanol/água 2:1 e em seguida foi submetida à cromatografia líquida preparativa, em condição isocrática (água/acetonitrila 80:20 + 0,1% de ácido fórmico, 45 min). Utilizou-se uma coluna Luna C18 (15 cm x 21,5, Phenomenex), vazão de 10 mL/min e volumes de injeção de 0,4 mL. Foram coletados 11 picos, os quais foram imediatamente analisados por HPLC (cromatógrafo 1), utilizando uma coluna Luna C18 (25 cm x 4,6 mm, Phenomenex) e mesma condição isocrática, por 30 minutos.

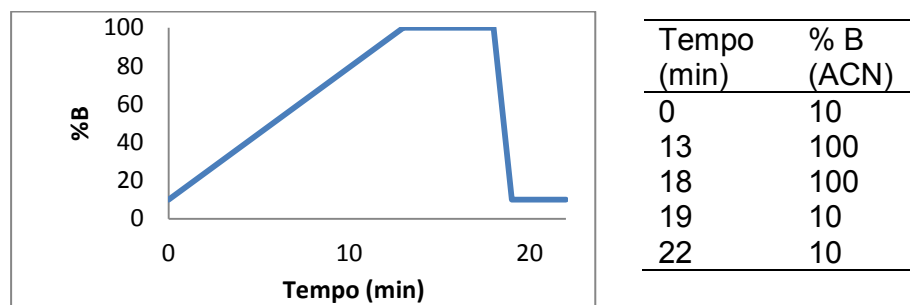


Figura 9- Gradiente rápido usado na análise por UHPLC-DAD das frações selecionadas obtidas no fracionamento do extrato bruto liofilizado da var. *lingua* (EBLL).

Finalmente as soluções contendo as substâncias referentes aos picos coletados tiveram a acetonitrila retirada por rotaevaporação e foram finalmente liofilizadas. Para todas as substâncias isoladas foram obtidos espectros de massas de alta resolução por inserção direta (micrOTOF Q II - ESI-TOF) e também por MS/MS. Os parâmetros utilizados foram os seguintes: modo de ionização negativo, calibração interna NA-TFA a 10mg/ml (TOF); end plate: -500 Volts; capilar: 3500 Volts; temperatura e vazão do gás de secagem 180° C e 4L/min, respectivamente; pressão do gás de nebulização (nitrogênio): 0,4 bar.

A identificação foi feita utilizando-se os bancos de dados disponíveis *on line*, a citar o *Dictionary of Natural Products* (www.dnp.chemnetbase.com) e MassBank (<http://www.massbank.jp>).

4.2.3.2 Validação dos dados obtidos

Para as substâncias que apresentaram massa e pureza adequadas, foram obtidos os dados de NMR de ^{13}C e ^1H , uni e bidimensionais. Os dados dos espectros de massas e de NMR foram também comparados com dados da literatura e, quando disponível, utilizaram-se também padrões autênticos das substâncias previstas.

4.3 PARTE 3 - Estudos de dinâmica metabólica

4.3.1 Coletas, processamento e caracterização das amostras

O estudo de dinâmica metabólica contemplou a coleta das folhas de dois indivíduos de *C. sylvestris*, um da variedade *sylvestris* e outro da variedade *lingua*. Oportunamente, ambos localizam-se na mesma área (Clube ASCAR, Estrada Abílio Augusto Correia n. 200, Bairro dos Machados, Araraquara/SP) a cerca de 50 m de distância um do outro. Antes do início da rodada de coletas, as árvores foram amostradas e analisadas por HPLC-DAD (cromatógrafo 1), confirmando-se o perfil químico distinto. As atividades de coleta estenderam-se do mês de agosto de 2012 ao mês de julho de 2013, com coletas mensais ao longo de 48h, realizadas a cada 3 horas. Em cada amostragem foram coletadas cerca de 20 folhas frescas, que foram imediatamente limpas com água destilada, secas, acondicionadas em tubos Falcon de 15 mL e congeladas em nitrogênio líquido. As amostras foram estocadas em freezer – 80°C, onde permaneceram até o final das coletas.

Para o processamento prévio das amostras para as análises cromatográficas, cerca de 5 folhas coletadas e congeladas foram transferidas para placas de Petri, desidratadas em estufa a 55°C por 24 horas e em seguida trituradas utilizando moinho analítico e nitrogênio líquido. Para todas as amostras, foi também determinado o teor de água em % m/m de acordo com a **Equação 2** a seguir:

Equação 2

$$\% \text{ Água (m/m)} = 100 - (((M3 - M1) / (M2 - M1)) \times 100)$$

Em que:

M1 = massa das placas de Petri

M2 = massa das placas de Petri + folhas frescas e congeladas

M3 = massa das placas de Petri + folhas desidratadas

A avaliação do teor de água foi feita com a média mensal dos resultados e os dados organizados em gráficos de coluna.

Os dados de temperatura e umidade dos dias e horários de todas as coletas foram monitorados com o uso de um termo higrômetro e dados mais detalhados foram obtidos em parceria com Michele Souza, aluna de doutorado do professor Dr. Arnaldo

Alves Cardoso do Departamento de Química Analítica do Instituto de Química da Unesp Araraquara.

4.3.2 Avaliação do teor de taninos por extração em fase sólida com poliamida

Uma vez que o perfil químico preliminar das duas variedades apresentaram consideráveis diferenças principalmente com relação à presença de compostos fenólicos e por ter sido possível inferir a presença de taninos, principalmente nas amostras da variedade *lingua*, a partir dos perfis cromatográficos obtidos, realizou-se um estudo comparativo das frações obtidas por extração em fase sólida (EFS) com poliamida utilizando os extratos brutos liofilizados das duas variedades de *C. sylvestris* (EBLL e EBLS). Primeiramente fez-se o condicionamento do adsorvente (cartucho contendo 0,5 g de poliamida) com 6 mL metanol/água 1:1 e em seguida aplicou-se cerca de 15 mg de cada amostra (uma de cada vez, extrato bruto liofilizado da variedade *lingua* - EBLL, extrato bruto liofilizado da variedade *sylvestris* - EBLS e fração rica em compostos fenólicos - FRCF, conforme obtenção descrita no item 4.2.3.1). Em seguida coletou-se 3 frações de 6 mL eluidas com água destilada, seguida de metanol/água 1:1 e finalmente metanol contendo 5% de hidróxido de amônio. Para determinação da quantidade de solutos extraídos, alíquotas de 1,5 mL de cada fração foram transferidas para frascos de vidro de 1,5 mL previamente tarados, sendo o solvente evaporado a seguir em *speed vac*. Calculou-se, finalmente, o teor percentual em % m/m de resíduo para cada amostra.

4.3.3 Análise das amostras e avaliação da conformidade das análises

4.3.3.1 Estudo de dinâmica metabólica: ciclos circadiano e sazonal

As 384 amostras provenientes das variedades *lingua* e *sylvestris*, coletadas para as análises de dinâmica metabólica foram analisadas por UHPLC-DAD em sequência aleatória, de acordo com o método cromatográfico desenvolvido (Apêndice II).

Para a análise dos dados, foram selecionados três flavonoides glicosilados, a citar rutina, narcisina e isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- α -L-arabinopiranosideo (cuja área foi registrada em 254 nm) e cinco diterpenos clerodânicos, a citar casearina D, caseargrewiina F, casearina S, casearina I e casearina J (com áreas registradas em 235 nm). Calculou-se então o teor de cada substância, em mg/g utilizando-se as curvas analíticas obtidas na etapa de validação do método, sendo

que para tanto, considerou-se a razão da área de cada marcador pela área do padrão interno (registrada em 254 nm).

A organização dos dados do ciclo circadiano foi feita utilizando-se gráficos do tipo radar, enquanto que os dados da variação sazonal estão apresentados em gráficos do tipo *box plot*.

Ao fim das análises de dinâmica metabólica, as 192 amostras de cada variedade coletadas ao longo de um ano foram reunidas para a elaboração de um extrato bruto liofilizado de *C. sylvestris* var. *sylvestris* (EBLS) e de um extrato bruto liofilizado de *C. sylvestris* var. *lingua* (EBLL). Para tanto, 27,0 g das folhas secas e trituradas da variedade *lingua* e 31,0 g das folhas secas e trituradas da var. *sylvestris* foram extraídas exaustivamente com três porções de 175 mL do solvente extrator (água/etanol/isopropanol 5:3:2) em banho de ultrasson conforme o procedimento de extração descrito no Apêndice II, sem a adição do padrão interno. Os extratos tiveram então o volume reduzido por rotaevaporação e foram finalmente liofilizados. O rendimento foi calculado e os extratos armazenados em tubo falcon de 50 mL. Além disso, os extratos brutos obtidos foram utilizados nos estudos qualitativos.

4.3.3.2 Avaliação da conformidade das análises

Para a certificação da qualidade e conformidade das análises, uma amostra de cada mês (seis meses contemplando a variedade *lingua* e seis meses contemplando a variedade *sylvestris*) foi selecionada aleatoriamente e analisada novamente. Desta forma, os dados da razão da soma das áreas dos picos dos compostos de interesse pela área do padrão interno da nova análise (análise 2) foram comparadas com os dados da primeira análise (análise 1).

4.4 PARTE 4 – Estudos de variabilidade química infraespecífica

4.4.1 Coletas e classificação das amostras

Foram coletadas as folhas de 134 indivíduos de *C. sylvestris* oriundos de dezesseis populações distintas, compreendendo diferentes formações dos biomas Mata Atlântica, Cerrado, Caatinga, Pampa e Pantanal no período entre junho e setembro de 2012 e 2013. Cada local de coleta foi classificado de acordo com as regiões fitoecológicas e áreas de vegetação conforme indicado no Mapa de Vegetação do Brasil, confeccionado pelo IBGE (disponível em <http://mapas.ibge.gov.br/tematicos/vegetacao>).

Dependendo da disponibilidade de material vegetal, coletou-se de 10 a 20 folhas, as quais foram desidratadas em estufa de ar circulante a 40 °C e armazenadas inteiras em sacos plásticos hermeticamente fechados à temperatura ambiente (aproximadamente 10 g de cada amostra). Para todas as amostras foram feitas exsiccatas, as quais foram analisadas e tiveram as identidades confirmadas pela Dra. Rozeli Buzzanelli Torres. A **Tabela 7** evidencia as coletas realizadas.

4.4.2 Análise das amostras e processamento dos dados

Todas as amostras foram preparadas e analisadas aleatoriamente de acordo com a metodologia analítica já estabelecida. Para a análise dos dados, todos os cromatogramas, adquiridos em 235 e 254 nm, foram exportados em formato ASCII e compilados em uma planilha do OriginPro 8 (OriginLab, USA). As matrizes resultantes foram então exportadas para o Microsoft Excel 2010 (Microsoft, USA). A seguir os cromatogramas registrados em 235 nm (de 0 a 25 min) e 254 nm (de 25 a 45 min) foram normalizados em função do padrão interno galato de butila. Na sequência foi realizada uma redução de variáveis (4 vezes) e em seguida realizado o alinhamento dos picos através de algoritmo (http://www.models.kvl.dk/source/DTW_COW/index.asp) implementado no Matlab® v.7.5 (MathWorks Inc., Natick, USA). Os dados foram normalizados e centrados na média (como etapa de pré-tratamento) e finalmente foi realizada a análise de componentes principais (PCA) utilizando o software Pirouette 4.0 rev.2 (Infometrix, Bothell, USA).

Tabela 7- Amostras coletadas no território brasileiro para o estudo de variabilidade infraespecífica de *C. sylvestris*.

Estado	População	Código	N*	Morfologia**			Bioma	Cobertura Vegetal***
				L	S	I		
SP	Presidente Venceslau	PRE/SP	10	-	10	-	Mata Atlântica	Floresta Estacional Semi-Decidual
	Angatuba	ANG/SP	02	-	02	-	Cerrado	Cerrado
	Araraquara	ARA/SP	16	04	02	10	Cerrado	Cerrado
	Campinas	CAM/SP	14	02	06	06	Cerrado	Cerrado / Mata ciliar
	Mogi-Guaçu	MOG/SP	05	-	-	05	Ecótono	Cerrado / Floresta Est. Semi-Decidual
	Luis Antônio	LUI/SP	07	05	-	02	Ecótono	Cerrado / Floresta Est. Semi-Decidual
MT	Cáceres	CAC/MT	08	03	-	05	Pantanal	Campo-sujo-de-Cerrado
	Porto Estrela	POR/MT	05	-	-	05	Pantanal	Campo-sujo-de-Cerrado e Floresta Est. Semi-Decidual
	Cariri	CAR/CE	06	06	-	-	Ecótono	Cerrado / Savana Estéptica (Carrasco)
CE	Guaramiranga	GUA/CE	07	-	07	-	Mata Atlântica	Floresta Estacional Semi-Decidual
	Pacoti	PAC/CE	05	-	05	-	Mata Atlântica	Floresta Estacional Semi-Decidual
SC	Florianópolis	FLO/SC	14	-	14	-	Mata Atlântica	Floresta Ombrófila Densa
MG	Pocinhos do Rio Verde	PRV/MG	11	01	-	10	Ecótono	Cerrado / Floresta Est. Semi-Decidual
	São Roque de Minas	SRM/MG	08	08	-	-	Cerrado	Cerrado
RS	Rio Grande	RIO/RS	07	-	01	06	Pampa	Campo alagado / Estepe
SE	Sergipe	SER/SE	09	-	09	-	Mata Atlântica	Floresta Estacional Semi-Decidual

* Número de árvores amostradas; ** Morfologia baseada na classificação botânica; *** Cobertura vegetal de acordo com o mapa de regiões fitoecológicas do IBGE; L: var. *lingua*; S: var. *sylvestris*; I: morfologia intermediária; Ecótono: zona de tensão ecológica

4.5 PARTE 5- Estudo proteômico diferencial de *C. sylvestris*

4.5.1 Coleta, obtenção e avaliação dos extratos proteicos

As amostras das duas variedades de *C. sylvestris* foram coletadas no período da manhã, entre 7h00 e 09h00, na cidade de Araraquara/SP - Brasil. Foram amostradas cinco árvores de cada variedade e, de cada árvore, cerca de 10 g de folhas foram coletadas, lavadas com água destilada e secas com papel toalha. Em seguida as folhas foram colocadas em tubos Falcon de 50 mL e imediatamente congeladas em nitrogênio líquido. As amostras foram mantidas em freezer a -80°C até o momento da extração proteica.

O procedimento de extração foi baseado nos trabalhos de Wang et al. (2003) e Wang et al. (2006) com modificações apenas no dimensionamento da amostra. O procedimento que consiste basicamente em extração das proteínas com uma mistura 1:1 de fenol/tampão SDS e precipitação com acetato de amônio em metanol segue detalhadamente descrito no **Quadro 1** a seguir.

A quantificação dos extratos proteicos obtidos foi feita pelos métodos do ácido bicinchonínico (BCA) e amido black. Sendo assim, para tubos Epenforff de 1,5 mL, transferiu-se 5 mg de cada extrato proteico seco e solubilizou-se com 0,5 mL de tampão H6, perfazendo-se soluções proteicas brutas a 10 $\mu\text{g}/\mu\text{L}$. Tais soluções foram mantidas em freezer a -20 °C e utilizadas durante todo o estudo proteômico.

Quadro 1- Protocolo para obtenção dos extratos proteicos

Preparo do tecido vegetal: cerca de 12,0 g de folhas frescas congeladas foram trituradas em moinho analítico com o auxílio de nitrogênio líquido. Obteve-se um pó bem fino.



Lavagem com TCA/acetona: 5,0g do pó foram transferidos para um tubo de centrifuga de 35 mL. Adicionou-se 30 mL de uma solução de TCA 10% em acetona, agitou-se bem em vortex e centrifugou-se a 16.000 x g, 4°C, por 3 minutos. Removeu-se o sobrenadante por decantação.



Lavagem com metanol: adicionou-se 30 mL de uma solução de metanol 80% em água contendo 0,1M de acetato de amônio, agitou-se em vortex e centrifugou-se a 16.000 x g, 4°C, por 3 minutos. Removeu-se o sobrenadante por decantação.



Lavagem com acetona: adicionou-se 30 mL de uma solução de acetona 80% em água, agitou-se em vortex até dispersar-se o precipitado e centrifugou-se a 16.000 x g, 4°C, por 3 minutos. Removeu-se o sobrenadante por decantação.



Secagem: o precipitado lavado foi seco em ar comprimido à temperatura ambiente por cerca de 15 minutos.



Extração e precipitação de proteínas: adicionou-se 30 mL de uma solução de fenol/tampão SDS 1:1, agitou-se e incubou-se por 5 minutos. Em seguida, centrifugou-se a 16.000 x g por 3 minutos. Transferiu-se a camada fenólica (porção intermediária entre o precipitado e a camada superior correspondente ao tampão SDS) para um novo tubo de 30 mL. Adicionou-se 25 mL de metanol 80% em água contendo 0,1M de acetato de amônio e deixou-se à -20 °C por 4 horas. Descartou-se o sobrenadante e verificou-se a formação de um precipitado branco.



Lavagem e secagem do precipitado: primeiro lavou-se o precipitado com metanol 100% e posteriormente com acetona 80% em água, descartando o sobrenadante. Em cada lavagem, agitou-se bem em vortex e centrifugou-se a 16.000 x g, 4 °C, por 3 minutos. Finalizado o processo, o extrato proteico foi seco em ar comprimido e estocou-se em freezer -80 °C até o momento da análise por SDS-PAGE.

*O preparo de todas as soluções e reagentes utilizados neste protocolo encontram-se descritos no Apêndice I

4.5.1.2 Análise quantitativa pelo método do BCA

Para a quantificação pelo método do ácido bicinchonínico utilizou-se o kit Pierce BCA Protein Assay (Thermo Scientific). Para tanto, construiu-se uma curva analítica (em duplicata) a partir de diluições seriadas de uma solução padrão de albumina de soro bovino (BSA) a 2,0 $\mu\text{g}/\mu\text{L}$ nas concentrações de 0,0625, 0,125, 0,250, 0,500, 1,0 e 2,0

$\mu\text{g}/\mu\text{L}$. Para as amostras, que foram analisadas em triplicata, estabeleceram-se três diluições a partir da solução a $10 \mu\text{g}/\mu\text{L}$, a citar: 1:20, 1:100 e 1:200. Para todas as soluções amostras e padrões, fixou-se um volume de $20 \mu\text{L}$, em tubos Eppendorf de 0,5 mL. Também preparou-se um branco contendo apenas água bidestilada. Em seguida adicionaram-se $100 \mu\text{L}$ da solução de ácido bicinchonínico (BCA/sulfato de cobre 50:1), homogeneizou-se e incubou-se à 55°C por 30 minutos. Finalmente transferiu-se $100 \mu\text{L}$ para uma microplaca de 96 poços e fez-se a leitura em espectrofotômetro a 562 nm. Para os cálculos, a curva analítica foi plotada com uma regressão polinomial e expressou-se os resultados em proteínas totais em $\mu\text{g}/\mu\text{L}$, referente à solução inicial de proteína bruta de $10 \mu\text{g}/\mu\text{L}$.

4.5.1.3 Análise quantitativa pelo método do amido black

Para a quantificação pelo método do amido black, baseada no método descrito por Schaffner & Weissmann (1973), primeiramente construiu-se uma curva analítica (em triplicata) a partir de diluições seriadas de uma solução padrão de albumina de soro bovino (BSA) a $0,2 \mu\text{g}/\mu\text{L}$ nas concentrações de 0,0125, 0,025, 0,05, 0,1 e $0,2 \mu\text{g}/\mu\text{L}$. Para as amostras, que também foram analisadas em triplicata, estabeleceu-se uma diluição 1:50 a partir da solução proteica bruta a $10 \mu\text{g}/\mu\text{L}$ (tomando-se $2 \mu\text{L}$ da amostra e adicionando-se $98 \mu\text{L}$ de água bidestilada). Também preparou-se um branco contendo apenas água bidestilada. Para todas as soluções amostras e padrões, fixou-se um volume de $100 \mu\text{L}$, num tubo Eppendorf de 1,5 mL. Em seguida adicionaram-se $500 \mu\text{L}$ da solução de amido black a $260 \mu\text{g}/\text{mL}$, homogeneizou-se em agitador tipo vortex e incubou-se à temperatura ambiente por 5 minutos. Em seguida centrifugou-se a $15.000 \times g$ por 5 minutos e descartou-se o sobrenadante. Com o auxílio de vortex, ressuspendeu-se o precipitado com 1 mL de uma solução de ácido acético/etanol 1:10, centrifugou-se a $15.000 \times g$ por 5 minutos e descartou-se o sobrenadante. Repetiu-se este processo de lavagem por mais duas vezes. Finalmente ressuspendeu-se o precipitado com $300 \mu\text{L}$ de NaOH 0,1N, transferiu-se $100 \mu\text{L}$ para uma microplaca de 96 poços e fez-se a leitura em espectrofotômetro a 624 nm. Para os cálculos, a curva analítica foi plotada com uma regressão linear e expressou-se os resultados em proteínas totais em $\mu\text{g}/\mu\text{L}$, referente à solução inicial de proteína bruta de $10 \mu\text{g}/\mu\text{L}$.

4.5.1.4 Análise qualitativa por SDS-PAGE

Após a obtenção e quantificação dos extratos proteicos tanto pelo método BCA quanto pelo método do amido black, as amostras foram analisadas por SDS-PAGE, utilizando-se géis de poliacrilamida de 9 x 7 cm, com 0,5 mm de espessura. Para o preparo das amostras, o equivalente a 10 ou 20 μg de proteínas (quantidade que dependeu do método de quantificação e também da resposta do anticorpo a ser usado na etapa posterior de *western blotting*) foram transferidas para um tubo Eppendorf de 1,5 mL, juntamente com 1,5 mL de tampão-amostra SDS-PAGE 10X e uma quantidade de água bidestilada necessária para um volume final de 15 μL . Em seguida as amostras foram incubadas a 65 °C por 15 minutos e finalmente aplicadas no gel. Em um dos poços aplicou-se também 5 μL de um mix de padrões de pesos moleculares de 10, 15, 25, 35, 40, 55, 70, 110, 130 e 170 KDa. A eletroforese foi realizada à tensão de 10 mA por cerca de 60 minutos (gel de empilhamento) e posteriormente a 20 mA por cerca de 90 minutos (gel de separação), utilizando o tampão de corrida SDS-PAGE. Para avaliação qualitativa, a revelação foi feita com *comassie blue* por cerca de 18 horas (*overnight*) à uma temperatura de 2 - 4 °C e posteriores lavagens com água destilada.

4.5.2 Western blotting

O procedimento de *western blotting* iniciou-se com a separação das proteínas nas amostras por SDS-PAGE, de acordo com o que foi descrito anteriormente, porém sem a etapa de revelação com *comassie blue*. Com os géis prontos, fez-se a transferência das proteínas do gel para as membranas de nitrocelulose utilizando tampão-corrida e cuba para *western blotting* operando a uma tensão de 80 mA *overnight*. Finalizada a transferência as membranas foram coradas com vermelho de ponceau por 5 minutos e posteriormente descoradas repetidas lavagens com água bidestilada.

O procedimento de incubação com os anticorpos iniciou-se com o bloqueio da membrana com leite reconstituído em tampão PBS, pH 7,2-7,4 (leite/PBS) em 3 lavagens de 15 minutos cada. Em seguida incubou-se cada membrana com seu respectivo anticorpo diluído no leite/PBS. Para a análise qualitativa, os anticorpos utilizados, a diluição dos mesmos, a quantidade de proteína aplicada no gel, bem como o tempo de incubação seguem resumidos na **Tabela 8** a seguir. Já para a análise semi-quantitativa e comparação entre as variedades, aplicou-se no gel uma quantidades de amostra equivalentes a 5, 10 e 20 μg .

Após o tempo de incubação, as membranas foram lavadas com leite/PBS (3 x 15 minutos) e reincubadas com o anticorpo secundário (*anti-rabbit IgG*) numa diluição de 1:10.000 (1,4 μ L em 14 mL) por 2 horas à temperatura ambiente. Finalmente lavou-se novamente as membranas com o leite/PBS (2 x 10 min) e incubou-se com o reagente ECL (luminol) por 1 minuto para detecção por quimioluminescência. A forma de revelação variou entre a utilização de filmes para autorradiografia e prioritariamente a detecção direta utilizando um sistema de detecção e aquisição de imagem por fluorescência/quimioluminescência.

Tabela 8- Descrição dos anticorpos e detalhes da análise de *western blotting*.

Anticorpo*	Diluição	Quantidade/ incubação
anti-rbcl (<i>RuBisCo large subunit</i>) ¹	1:5000 (5 μ L/ 25 mL)	10 μ g/96h
anti-P25 (LHCII - <i>light harvesting complex</i>) ²	1:3000 (5 μ L/15 mL)	10 μ g/96h
anti-IspE (<i>4-diphosphocytidyl-2C-methyl-D-erythritol kinase</i>) ³	1:1000 (15 μ L/15 mL)	20 μ g/72h

¹ RuBisCo subunidade grande/metabolismo energético; ² proteínas coletoras de luz ou fotossistema II/metabolismo energético; ³ 4-difosfocitidil-2C-metil-D-eritritol kinase / via do metileritritol.

4.5.3 Identificação das proteínas do estudo proteômico diferencial

4.5.3.2 Preparo das amostras para LC-MS, etapa de digestão: FASP

O fracionamento do extrato proteico e análise por LC-MS/MS foram precedidos de uma etapa de purificação e digestão pelo procedimento conhecido como FASP, do inglês *filter aided sample preparation*, que envolve a eliminação de detergentes, sais e pigmentos presentes nas amostras e uma etapa de digestão com tripsina (WISNIEWSKI; ZOUGMAN; MANN, 2009). Por conter diversos passos, a descrição do procedimento segue no **Quadro 2** a seguir.

Quadro 2- Protocolo FASP

Tomada da amostra: para um tubo Eppendorf de 1,5 mL, pipetou-se uma quantidade de extrato proteico a $10 \mu\text{g}/\mu\text{L}$ de forma que se obtivesse $100 \mu\text{g}$ de proteína (considerando o teor obtido pela análise pelo método do amido black).



Incubação com DTT: Adicionou-se um volume de DTT 0,1 M de forma que o volume final fosse de 250 μL . Homogeneizou-se em vortex e incubou-se por 30 min à temperatura ambiente. Centrifugou-se a 14.000 x g por 10 minutos, à 20 °C.



Transferência para o filtro 30K: Centrifugou-se as amostras a 14.000 x g por 5 minutos à temperatura ambiente e transferiu-se o sobrenadante para um filtro molecular 30K. Centrifugou-se a 14.000 x g por 20 minutos, à 20 °C.



Lavagem com UA: Adicionou-se 200 μL de tampão UA, lavando-se as paredes do filtro e centrifugou-se a 14.000 x g por 15 minutos, à 20 °C. Repetiu-se uma vez.



Incubação com IAA: Adicionou-se 100 μL de tampão IAA, lavando as paredes do filtro e homogeneizou-se no termomixer por 1 minuto, a 600 rpm. Incubou-se, ao abrigo da luz por 20 minutos. Centrifugou-se a 14.000 x g por 10 minutos, à 20 °C.



Lavagem com UA: Adicionou-se 100 μL de tampão UA e centrifugou-se a 14.000 x g por 15 minutos, à 20 °C. Repetiu-se este passo mais uma vez.



Lavagem com ABC: Adicionou-se 100 μL de solução ABC 50 mM e centrifugou-se a 14.000 x g por 15 minutos, à 20 °C. Repetiu-se este passo mais uma vez.



Digestão com tripsina: Trocou-se o tubo Eppendorf onde estava acoplado o filtro e adicionou-se 40 μL de solução de tripsina $0,05 \mu\text{g}/\mu\text{L}$. Acondicionaram-se as amostras em uma caixa de plástico contendo um papel embebido em água e incubou-se por 16 horas à 37 °C



Obtenção do extrato peptídico: Ao filtro contendo a amostra digerida, adicionou-se 50 μL de solução ABC 50 mM e centrifugou-se a 14.000 x g por 10 minutos, à 20 °C. As amostras foram mantidas em banho de gelo e procedeu-se o próximo passo que foi o fracionamento por Pós-FASP.

O preparo de todas as soluções e reagentes utilizados neste protocolo encontram-se descritos no Apêndice I

4.5.3.3 Preparo das amostras para LC-MS, etapa de fracionamento: Pós-FASP

Uma vez digeridas, as amostras, agora constituídas de uma solução de peptídeos, foram fracionadas utilizando mini-colunas cromatográficas empacotadas em laboratório com resina aniônica, utilizando ponteiras descartáveis de 200 μL acopladas à tubos Eppendorfs de 1,5 mL, como descrito nos procedimentos relatados por Rappsilber et al. (2007) e Wisniewski et al. (2009). Para o preparo de tais colunas, compactaram-se 6 camadas de resina (troca aniônica, Empore) com o auxílio de uma micro-seringa adaptada. Em seguida as mini-colunas foram condicionadas com 100 μL de metanol, centrifugando-se a 7000 x g por cerca 15 minutos. Em seguida, mais um passo de condicionamento foi feito adicionando-se 100 μL de NaOH 1M, centrifugando-se da mesma forma. Finalmente o último passo de condicionamento foi feito, repedindo-se duas vezes, com 100 μL de tampão Britton Robinson pH 11 (BB pH 11, solução de trabalho) e centrifugando-se a 7.000 x g por 15 minutos. Uma vez as colunas condicionadas, trocou-se o tubo Eppendorf por um novo, aplicou-se a amostra, adicionou-se mais 100 μL de BB pH 11 e centrifugou-se a 7.000 x g por 15 minutos, iniciando-se o fracionamento. Coletou-se a amostra no tubo Eppendorf e reservou-se. Assim por diante, com tubos novos a cada pH, as amostras foram eluídas com os pHs 8, 6, 5, 4 e 3. Secou-se as amostras em *speed vac* a 30 °C e armazenou-se as frações secas em freezer a -80 °C.

4.5.3.4 Análise por LC-MS/MS

As frações secas e armazenadas em freezer -80 °C oriundas do procedimento Pos-FASP foram solubilizadas em 5 μL de fase móvel A imediatamente antes da análise por LC-MS/MS (cromatógrafo 5), que utilizou como eluentes água contendo 5% de acetonitrila e 0,1% de ácido fórmico (A) e acetonitrila 80% contendo 0,1% de ácido fórmico em água deionizada. Primeiramente, 1 μL da amostra foi injetada numa pré-coluna *trap* (PepMap100 C18, 300 μm x 5mm, 5 μm , 100Å) e dessaltificada com o eluente A a 25 $\mu\text{L}/\text{mL}$ por 4 minutos. Subsequentemente, esta coluna *trap* foi acoplada em série com uma coluna capilar (Atlantis C18, 75 μm x 10mm, 3 μm , 100Å) e o seguinte gradiente foi aplicado para a separação dos peptídeos: 0 a 15% de B por 10 minutos, 15-60% de B por 45 minutos, 60-100% de B por 2 minutos e 100% de B por 5 minutos. Finalmente a coluna foi reequilibrada com 100% de A por 10 minutos. A temperatura da coluna foi mantida em 35°C. Os peptídeos eluíram direto para a fonte *nanospray* do espectrômetro de massas, que operou no modo positivo. Os scans foram obtidos por

FT-MS (m/z 400-2000) com resolução de 60.000 FWHM usando calibração interna de m/z 445.120025. Os 12 íons mais intensos foram sequencialmente submetidos à fragmentação por dissociação induzida por colisão (CID), com energia de colisão normalizada a 35% no *ion trap* linear (MS^2). A exclusão dinâmica foi feita com uma duração de repetição de 45 s, contagem repetida de 1 e intervalo de exclusão de massa de ± 5 ppm.

4.5.3.5 Análise dos dados por bioinformática

Os dados obtidos na análise por LC-MS/MS de cada amostra (PL1 e PS1) foram analisados individualmente utilizando o software Proteomatic® (SPECHT et al., 2011), que automatiza e processa a análise de grandes conjuntos de informações provenientes de espectrometria de massas *tandem*, utilizando diversas ferramentas de conversão, tratamento, identificação, quantificação, sanitização e comparação de dados proteômicos, dentre outras. Utilizando esta plataforma, a identificação das proteínas presentes nas amostras foi feita através dos programas OMSSA (versão 2.1.9, GERR et al., 2004) e X! Tandem (versão 2012.10.01.1, CRAIG et al., 2004).

Para que estes programas pudessem ser executados e como parte da estratégia para identificação do maior número possível de proteínas, trabalhou-se em duas vertentes para a seleção de banco de dados (Estratégia 1 e Estratégia 2, Apêndice IV). A primeira utilizou um mistura de bancos de dados de proteínas de plantas que já possuem genoma sequenciado disponíveis na plataforma NCBI (<http://www.ncbi.nlm.nih.gov/>). A escolha das espécies foi feita com base em dois critérios: primeiro, selecionou-se aquelas que fossem filogeneticamente próximas à espécie *C. sylvestris* e, segundo, que tivessem ocorrência de diferentes classes de metabólitos, tais como terpenos, diterpenos, flavonoides e outros compostos fenólicos. A outra vertente utilizou apenas o banco de dados de proteínas de *Populus trichocarpa* (Salicaceae). As sequências de peptídeos assim obtidas (ou PSM, do inglês *peptide sequence match*) foram então filtradas considerando aquelas obtidas com melhor score. Para a avaliação estatística da qualidade das sequências de peptídeos foi utilizada a estratégia *target/decoy* (ou sequência alvo/sequência falsa). Para tanto, sequências peptídicas artificiais inexistentes foram construídas a partir do banco de dados de proteínas das espécies de plantas utilizadas na identificação por OMSSA e X! Tandem, embaralhando-se randomicamente os peptídeos tripticos (script disponível no Proteomatic®). Em seguida removeu-se as sequências que coincidiram com aquelas

fictícias. Finalmente, para a validação estatística da qualidade das proteínas, utilizou-se a ferramenta *Qvalue* (versão 2,02, KÄLL et al., 2009), com um erro de probabilidade inferior a 0,05. Como um passo adicional de filtragem, todos os PSMs, com um desvio de massa precursor de mais de 5 ppm foram descartados.

Além disso, outra abordagem também foi utilizada para análise dos dados de MS/MS, que foram submetidos ao programa PEAKS (versão 2.4, MA et al., 2003) fornecendo sequências de aminoácidos num procedimento conhecido como *De Novo Sequencing* ou síntese *De Novo* de proteínas. Finalmente, a anotação dos dados foi feita com o programa BLAST (disponível em <http://blast.ncbi.nlm.nih.gov>), algoritmo BLAST P, utilizando-se todo o banco de dados de todas as espécies já descritas disponível na plataforma NCBI (ncbi-nr).

Em seguida comparou-se a variedade *sylvestris* com a variedade *lingua* no nível de peptídeos (identificados com base apenas na estratégia que utilizou apenas o banco de dados de *P. trichocarpa*), calculando-se o número de scans (*scan counts*) de cada uma. Nesta etapa, considerou-se segura uma razão maior que 4 para 1 como etapa de filtro adicional. No Apêndice IV encontram-se de forma resumida e esquemática, as estratégias utilizadas em bioinformática, enquanto que o Apêndice V traz a interface gráfica do Proteomatic®.

5 RESULTADOS E DISCUSSÃO

5.1 PARTE 1- Desenvolvimento de método cromatográfico por UHPLC-DAD

5.1.1 Composição da amostra representativa de *C. sylvestris* (ARCS)

Para o desenvolvimento do método cromatográfico foi necessário ter em mãos uma amostra de perfil químico altamente complexo e representativo da diversidade molecular a ser encontrada nas futuras análises de variabilidade química infra-específica, contemplando as duas principais classes de metabólitos secundários em questão: a dos compostos fenólicos, reconhecidamente característico da variedade *lingua*, e ainda pouco estudada e a dos diterpenos, característica da variedade *sylvestris* e já bem estabelecida. A seleção das amostras foi realizada a priori com base nas características morfológicas das folhas e posteriormente a diferença no perfil metabólico de cada uma foi confirmada por cromatografia. Para tanto, quatro amostras de cada variedade foram coletadas na cidade de Araraquara, analisadas individualmente e um *blend* foi composto de forma que as duas classes de metabólitos estivessem presentes em quantidade adequada para a análise por HPLC-DAD. A **Figura 10** a seguir mostra com detalhes as diferenças entre as folhas de ambas as variedades.



Figura 10- Amostra de folhas de *C. sylvestris* var. *lingua* (à esquerda) e *C. sylvestris* var. *sylvestris* (à direita).

Uma vez analisadas, todas as amostras já secas e trituradas foram combinadas e compôs-se uma amostra representativa cujo perfil cromatográfico segue apresentado na **Figura 11**. Obteve-se 9,1 g desta mistura, quantidade suficiente para a execução de todos os experimentos de desenvolvimento de método.

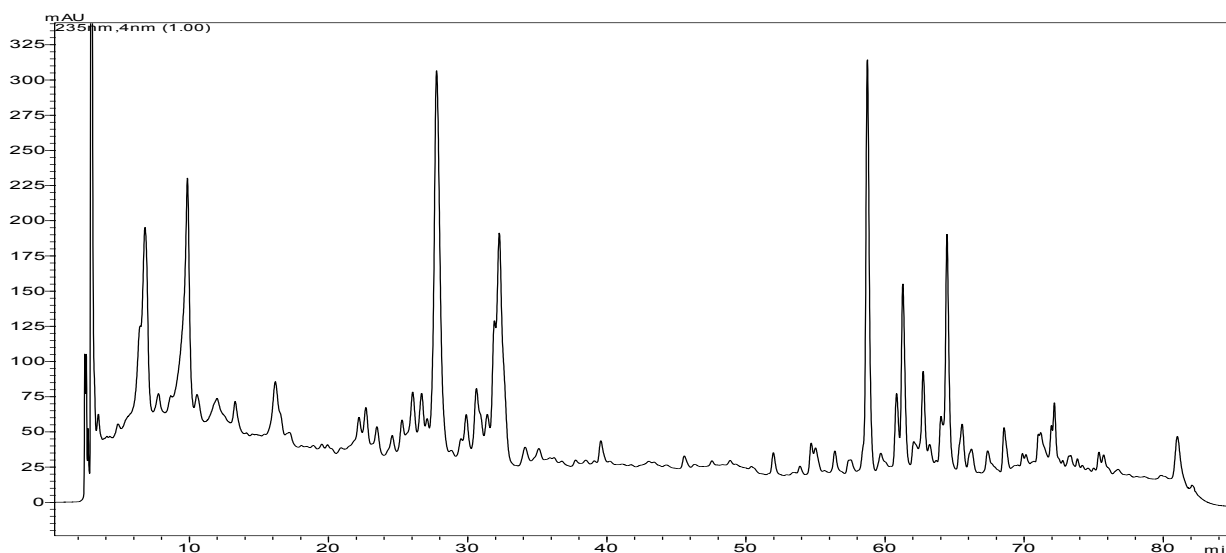


Figura 11- Perfil cromatográfico da amostra representativa de *C. sylvestris* (ARCS), utilizada durante todo o desenvolvimento analítico. Cromatograma obtido por HPLC-DAD (235 nm).

5.1.2 Definição de um método provisório por UHPLC-DAD

Até a composição da amostra representativa de *C. sylvestris* (ARCS) os perfis cromatográficos foram obtidos em modo HPLC, trabalhando com uma coluna C18 convencional de 25 cm, com tamanho de partícula de 5 μm . No entanto, para os próximos passos, visando uma maior capacidade de picos e maior resolução, decidiu-se trabalhar no modo UHPLC, com uma coluna de núcleo sólido (*Core-Shell*) de 15 cm e tamanho de partícula de 2,6 μm , objetivando inclusive, diminuir o tempo de análise.

Sendo assim, um primeiro método provisório, contendo apenas água e acetonitrila foi estabelecido e utilizado para a execução da próxima fase que foi a definição das melhores condições de preparo das amostras. Nesta etapa também avaliou-se a influência do uso de 0,1% de ácido fórmico na fase móvel, considerando-se a presença dos compostos fenólicos principalmente na variedade *lingua*.

A adição de ácido fórmico não incrementou a separação e resolução das bandas de interesse ($t_R > 10$ min). Por este motivo, e para viabilizar um recondicionamento mais rápido da coluna, decidiu-se não adicionar ácido à fase móvel. Os cromatogramas estão apresentados nas **Figuras 12A e 12B**.

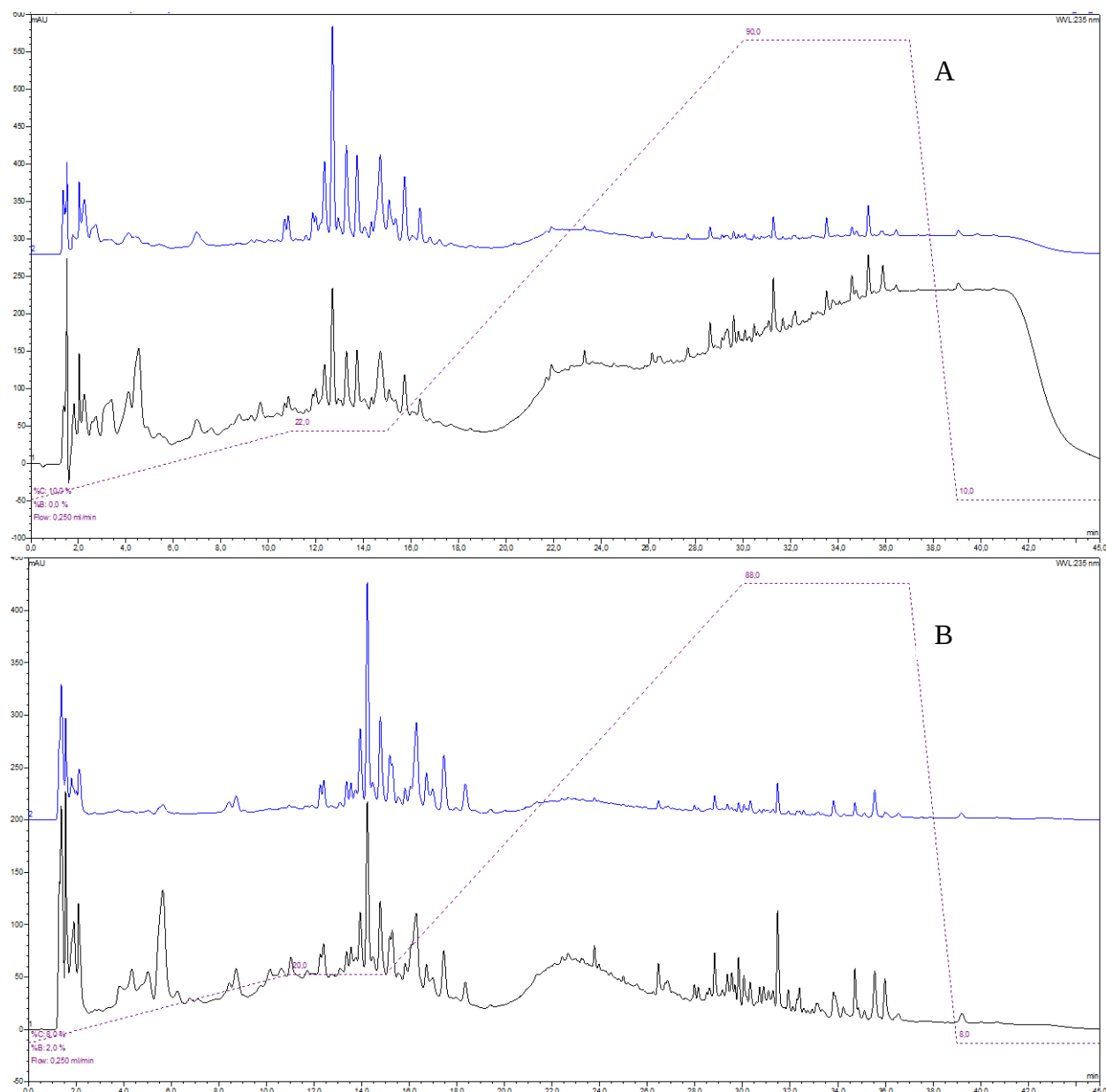


Figura 12- Cromatogramas da avaliação da influência da adição de ácido fórmico à fase móvel. (A) método provisório acrescido de 0,1% de ácido fórmico. (B) método provisório sem adição do ácido. Em azul, cromatograma em 254 nm; em preto, cromatograma em 235 nm.

5.1.3 Otimização das condições de extração das amostras

5.1.3.1 Planejamento Simplex Centróide para definição do solvente extrator

Neste trabalho priorizou-se a extração simultânea abrangente tanto dos compostos fenólicos de maior polaridade, quanto dos diterpenos de menor polaridade, os quais poderiam estar presentes nas diferentes amostras e variedades de *C. sylvestris*. Portanto, o uso de uma amostra representativa desta espécie (ARCS), contendo tanto

as folhas da variedade *lingua* (rica em compostos fenólicos) quanto as folhas da variedade *sylvestris* (rica em diterpenos clerodânicos) garantiu que as duas classes de metabólitos estivessem presentes. No trabalho realizado por Bandeira et al. (2006), o qual priorizou a análise dos diterpenos clerodânicos presentes em *C. sylvestris*, a extração destes metabólitos foi realizada durante uma hora em banho de ultrasson e utilizando diclorometano saturado com hidróxido de amônio. No entanto, incluir os compostos fenólicos no mesmo processo extrativo requeria o aumento da polaridade do solvente de extração.

Portanto, para se alcançar tal objetivo, utilizou-se o modelo Simplex Centróide para a definição do melhor solvente extrator. Avaliou-se que a utilização de misturas de água e isopropanol poderia contemplar boa faixa de polaridade, mas que a viscosidade desses solventes poderia ser um fator negativo. Por isso foi incluído o etanol como terceiro solvente, na expectativa de se obter uma mistura ternária que permitisse o ajuste adequado de polaridade para a mistura extratora, com menor interferência da viscosidade, considerando-se também sua menor toxicidade e menor impacto ambiental. Uma análise visual dos cromatogramas também foi considerada de forma que se pudesse monitorar a confiabilidade dos resultados. No Apêndice VI estão dispostos todos os cromatogramas obtidos. A **Tabela 9** e a **Figura 13** a seguir resumem os resultados.

Tabela 9- Resultados do planejamento Simplex Centróide para a definição do melhor solvente extrator.

Ensaio	Água	Etanol	Isopropanol	Número de picos detectados		Média
				235 nm	254 nm	
1	1	0	0	69	49	59
2	0	1	0	79	53	66
3	0	0	1	64	41	52
4	1/2	1/2	0	110	82	96
5	0	1/2	1/2	83	57	70
6	1/2	0	1/2	115	84	99
7	2/3	1/6	1/6	101	75	88
8	1/6	2/3	1/6	106	69	87
9	1/6	1/6	2/3	79	54	66
10	1/3	1/3	1/3	102	80	91

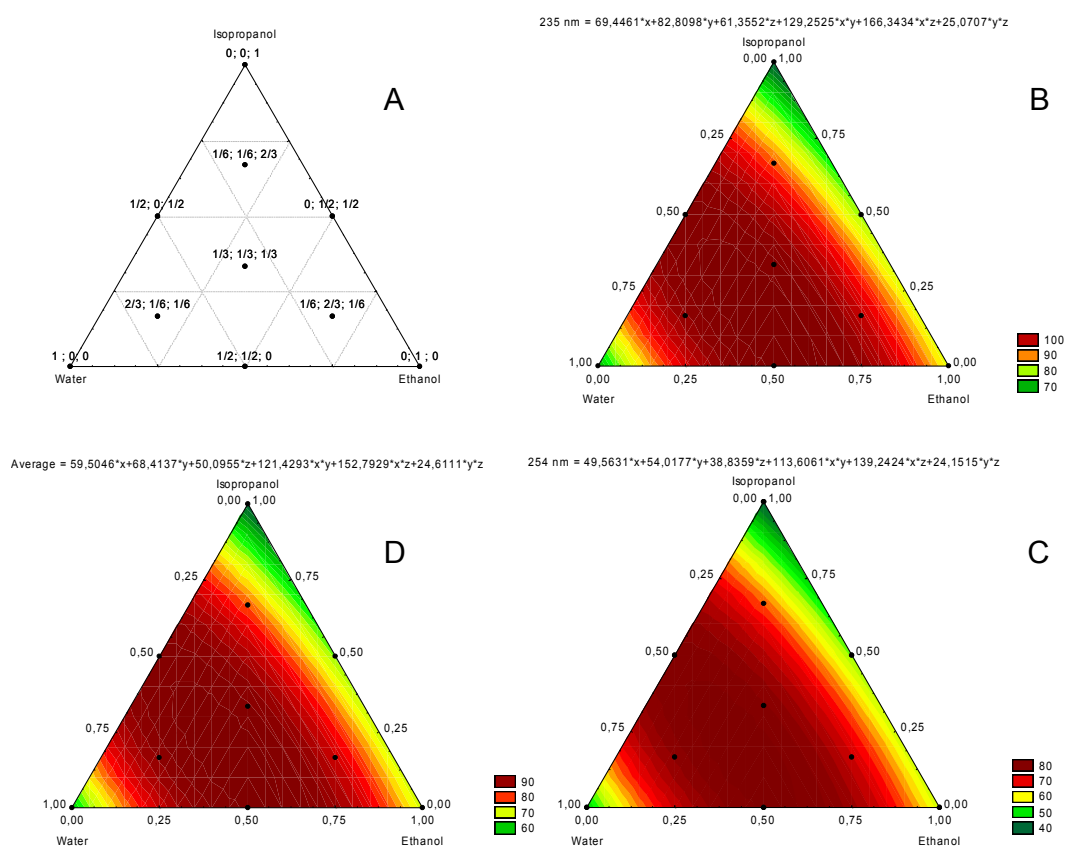


Figura 13- Gráficos do planejamento experimental Simplex Centroide. (A) planejamento com os pontos e proporções dos solventes utilizados; (B) resultados do planejamento em 235 nm; (C) resultados do planejamento em 254 nm; (D) resultados do planejamento com a média dos resultados.

Observa-se que maior número de picos é detectado no comprimento de onda de 235 nm, uma vez que em 254 a detecção das casearinas é comprometida devido à baixa absorção dessas substâncias nesse comprimento de onda. No entanto, em 235 nm há um comprometimento da linha de base, tornando-o adequado somente para a detecção das casearinas. Sendo assim, resolveu-se analisar os gráficos de mistura de extração ternária considerando a média do número de picos obtidos nos dois registros. A partir desta estratégia, o modelo quadrático (**Equação 3**) mostrou-se ser o mais adequado no nível de confiança de 95% ($R^2 = 0,910$).

Equação 3

$$N = 59,5x + 68,4y + 50,1z + 121xy + 153xz + 24,6yz$$

$\pm 23,0 \quad \pm 23,0 \quad \pm 23,0 \quad \pm 106 \quad \pm 106 \quad \pm 106$

Através da **Equação 3** é possível identificar a proporção ótima de solventes para o processo extrativo, a qual foi definida na proporção de 5:3:2 (v/v) de água (x), etanol (y) e isopropanol (z), respectivamente (**Fig. 13D**). De acordo com a equação obtida, a utilização da melhor proporção de solventes leva a um número de picos (N) igual a 94, como segue: $N = (59,5 \times 0,5) + (68,4 \times 0,3) + (50,1 \times 0,2) + (121 \times 0,5 \times 0,3) + (153 \times 0,5 \times 0,2) = 94$. Para validar este resultado, analisou-se uma amostra em triplicata utilizando a proporção de solventes definida pelo ponto ótimo. Desta forma, obteve-se uma média de 111 picos detectados em 235 nm (113, 111 e 112 picos) e uma média de 79 picos detectados em 254 nm (74, 81 e 81 picos), perfazendo-se uma média final de 95 picos. Tal resultado reflete uma excelente concordância entre o número de picos previstos pela equação ($N=94$) e o número real de picos obtidos ($N = 95$).

A **Figura 14** a seguir evidencia a melhora significativa na qualidade do cromatograma obtido, no qual foi possível observar 81 picos em 254 nm e 112 picos em 235 nm.

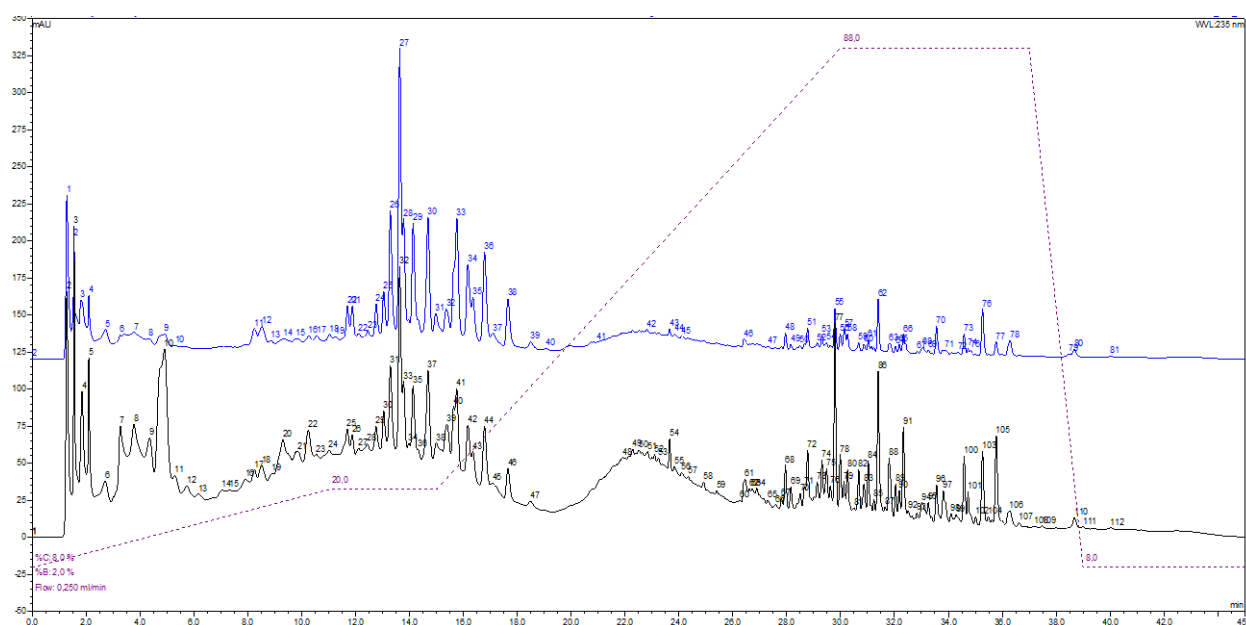


Figura 14- Cromatograma da amostra representativa extraída na proporção ótima de solventes definida pelo planejamento Simplex Centróide. Em azul cromatograma em 254 nm; em preto, em 235 nm.

5.1.3.2 Estudo da influência do pH na etapa de extração

Devido à presença dos compostos fenólicos em folhas de *C. sylvestris*, foi testada a influência do pH na etapa de extração das amostras, modulando-se este fator com a adição de tampão e de ácidos de diferentes pKa's à mistura de solventes definida pelo

planejamento Simplex Centróide. Não se observou melhoria quando se adiciona qualquer ácido ao solvente extrator. A **Figura 15** evidencia os resultados.

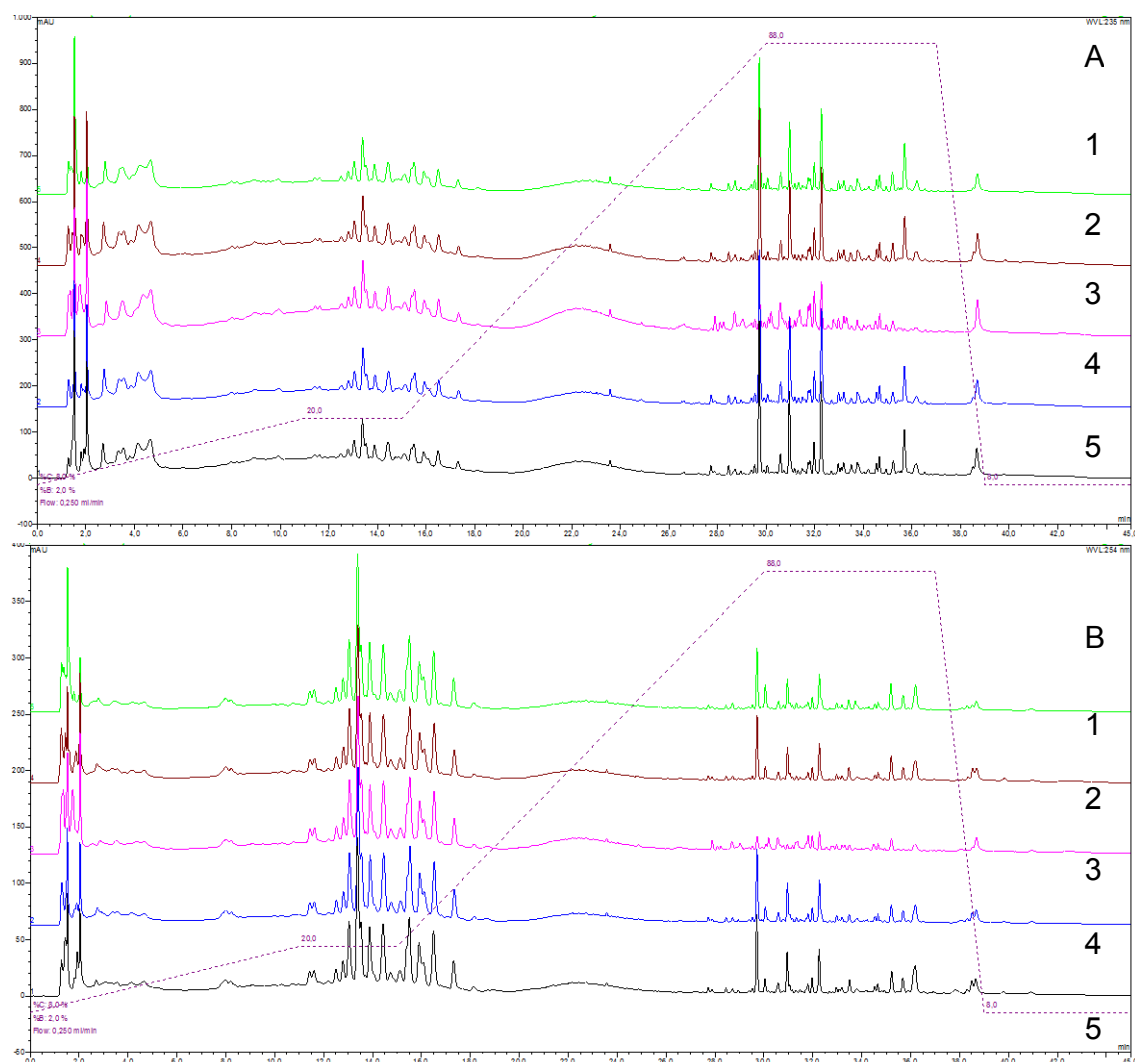


Figura 15- Cromatogramas obtidos no teste da influência de diferentes ácidos no procedimento de extração. 1- água, 2- tampão pH 4,5, 3- TFA, 4- ácido acético, 5- ácido fórmico, A = 235 nm, B = 254 nm.

5.1.4 Otimização da condição de eluição cromatográfica

Tendo por princípio que o modo cromatográfico, a coluna e o procedimento de extração já estavam definidos, um planejamento fatorial fracionário foi utilizado para a triagem das variáveis que mais exercem influência no método de separação. Sendo assim, considerou-se a vazão da fase móvel (X_1), a temperatura da coluna (X_2) e os dois segmentos do gradiente multilinear (X_3 e X_4 , respectivamente). Ajustou-se a vazão ótima para 0,400 mL/min, por ser a vazão nominal ótima para o tipo de coluna cromatográfica em uso. A **Tabela 10** a seguir evidencia o planejamento e os resultados encontrados,

que se deu em função do número de picos observados. Da mesma forma, no Apêndice VII encontram-se os cromatogramas obtidos.

Tabela 10- Resultados do Planejamento Fatorial Fracionário 2^{4-1} com ponto central para triagem de variáveis.

Ensaio	Níveis dos fatores				N*	Cálculo do efeito			
	X ₁	X ₂	X ₃	X ₄		Contraste	Quadrado	%	
1	-1	-1	-1	-1	94	C2	-0,25	0,0625	0,0
2	1	-1	-1	1	96	C3	0,75	0,5625	0,2
3	-1	1	-1	1	109	C13	-0,75	0,5625	0,2
4	1	1	-1	-1	87	C24	-0,75	0,5625	0,2
5	-1	-1	1	1	107	C14	-3,25	10,5625	4,5
6	1	-1	1	-1	91	C23	-3,25	10,5625	4,5
7	-1	1	1	-1	99	C12	-3,75	14,0625	6,0
8	1	1	1	1	92	C34	-3,75	14,0625	6,0
pc1	0	0	0	0	102	C4	8,25	68,0625	29,0
pc2	0	0	0	0	98	C1	-10,75	115,5625	49,3
pc3	0	0	0	0	99	Soma		234,625	100

*número de picos detectados; ** nível de significância = 0,05

Num nível de significância de 0,05 observa-se que as variáveis que mais influenciaram o número de picos cromatográficos foram a vazão da fase móvel (X₁) e o segundo segmento do gradiente (X₄), pois mostraram ser estatisticamente significativos. Juntos, estes dois parâmetros foram responsáveis por 78,3% do efeito. Por este motivo, o próximo passo foi um planejamento composto central contendo estes dois parâmetros, requerendo, inclusive, um menor número de experimentos. No Apêndice VIII estão compilados os cromatogramas obtidos. A **Tabela 11** evidencia os resultados obtidos para este último planejamento, os quais forneceram o modelo matemático num nível de confiança de 95%, (**Equação 4**), com R² de 0,990.

Equação 4

$$N = 89,9 - 1,21x_1 + 4,50x_2 + 0,44x_1 x_1 + 3,71x_2 x_2 - 1,75x_1 x_2$$

$\pm 46,9$ $\pm 23,5$ $\pm 23,5$ ± 31 $\pm 31,1$ $\pm 33,2$

Analisando-se os resultados obtidos, verifica-se que o método mostrou-se robusto, visto que apenas o intercepto ($b_0 = 89,9$) foi estatisticamente significativo na faixa estudada. As outras variáveis não foram significativas uma vez que seus intervalos de confiança (valores denotados com $\pm$) tiveram magnitude muito maior que os coeficientes calculados. Neste sentido, confirmou-se a vazão ótima de 0,400 mL/min, a segunda porção do gradiente foi estabelecida em 20 minutos e a temperatura mantida em 35 °C.

Tabela 11- Resultados do Planejamento Composto Central definição do método cromatográfico.

Ensaio	Níveis dos fatores		N*	Coeficientes		Erro
	X ₁	X ₂				
1	-1	-1	96	b0	89,9	46,9
2	1	-1	101	b1	-1,2	23,5
3	-1	1	109	b2	4,5	23,5
4	1	1	107	b11	0,4	31,1
5	-1,41	0	86	b22	3,7	31,1
6	1,41	0	77	b12	-1,8	33,2
7	0	-1,41	82			
8	0	1,41	94			
9	0	0	89			
10	0	0	91			

*número de picos detectados; ** nível de significância = 0,05

Ao final dos experimentos, alguns ajustes fora do planejamento ainda foram feitos considerando-se a resolução global do cromatograma. Desta forma, o tempo do primeiro e segundo segmento do gradiente foram ajustados e o método definido conforme pode ser observado na **Figura 16** a seguir, que mostra um cromatograma da amostra representativa de *C. sylvestris* (ARCS) obtido de acordo com as condições estabelecidas após a execução de todos planejamentos experimentais para a definição do método cromatográfico final. Para comparação, são apresentados também os perfis cromatográficos de uma amostra de *C. sylvestris* var. *lingua* (CSL) e *C. sylvestris* var. *sylvestris* (CSS) nas novas condições de eluição.

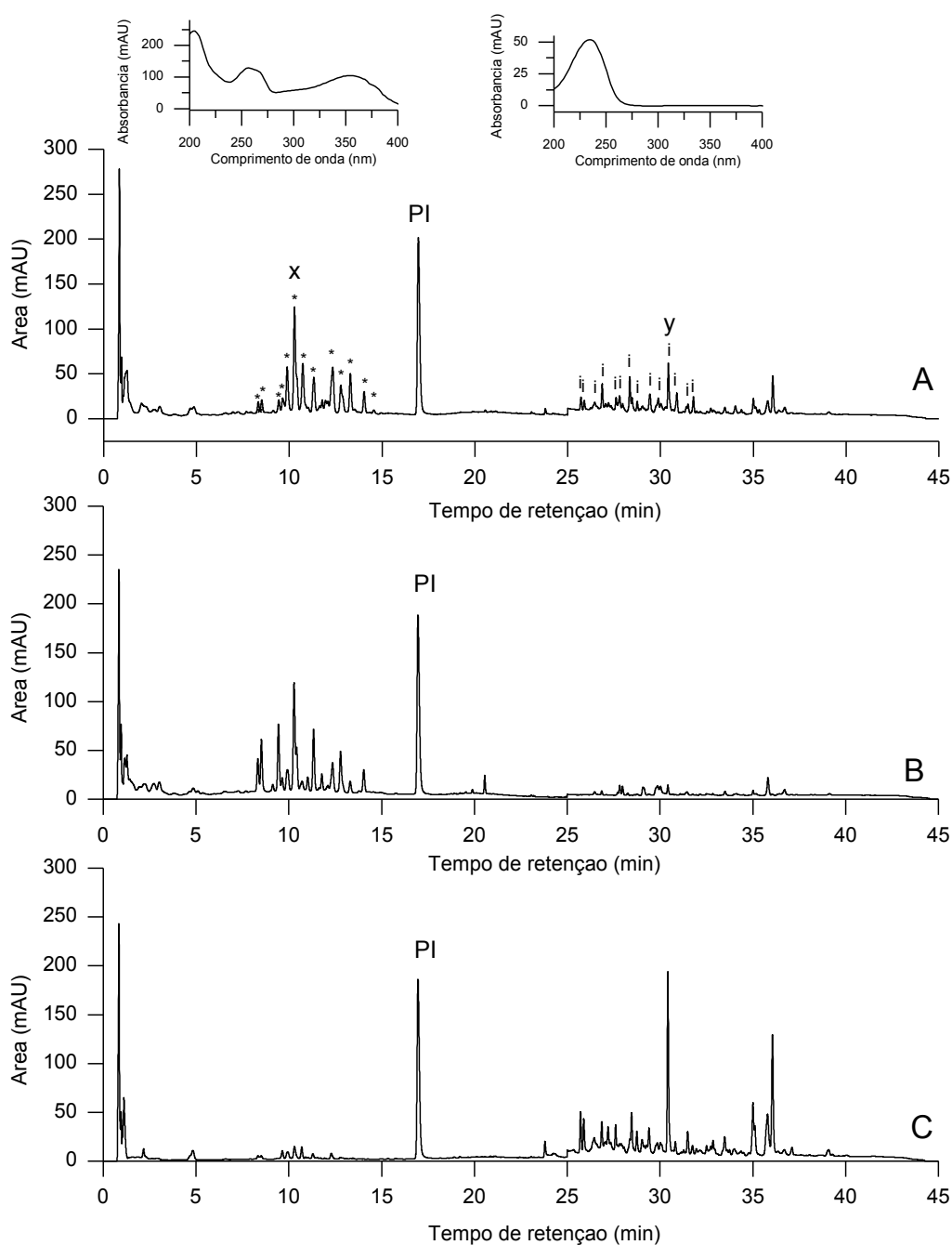


Figura 16- Perfis cromatográficos btidos por UHPLC-DAD. (A) amostra representativa de *C. sylvestris* (ARCS); (B) *C. sylvestris* var. *lingua* (CSL); (C) *C. sylvestris* var. *sylvestris* (CSS). Os cromatogramas foram plotados em 254 nm (de 0 a 25 min) e 235 nm (de 25 a 45 min). Os sinais *, i and PI representam os picos característicos dos compostos fenólicos (mais precisamente flavonoides glicosilados), diterpenos clerodânicos e padrão interno, respectivamente. As letras X e Y representam os picos selecionados para a validação do método. Condições cromatográficas definidas pelo planejamento: coluna cromatográfica Kinetex 150 mm × 2,1; 2,6 μm, 100 Å. Fase móvel: água (A) e acetonitrila (B). Gradiente de eluição: 10-25% de B de 0 a 15 min, 25-90% de B até 35 min, mantendo-se a 90% de B até 40 min, retorno à condição inicial em 2 min e tempo de equilíbrio de 3 min. Vazão: 0,4 mL/min. Temperatura do forno: 35 °C. Volume de injeção: 2 μL.

É possível concluir que o uso dos planejamentos experimentais para a otimização do método cromatográfico permitiu, com um número reduzido de experimentos, chegar o mais próximo possível do método ótimo. O Apêndice II resume o protocolo para a análise de amostras de *C. sylvestris*, de acordo com as condições definidas.

5.1.5 Escolha do padrão interno

A escolha do padrão interno foi feita analisando-se uma mistura de uma série homóloga de galatos de alquila nas condições cromatográficas definidas acima (**Figura 17**). Tais ésteres possuem grupos fenólicos e cadeias alquílicas presentes nos flavonoides e diterpenos, respectivamente, porém não estão presentes na matriz natural, possuem ótima absorção em 235 e em 254 nm, além de serem estáveis nas temperaturas e solventes utilizados no método de análise. O galato de butila foi escolhido como padrão interno por ter eluído em região do cromatograma livre de metabólitos de *C. sylvestris* e apresenta todas as características indicadas a cima, além de ser disponível comercialmente. Cabe destacar que a região de eluição deste derivado encontra-se exatamente após a eluição dos compostos fenólicos e antes da eluição das casearinas.

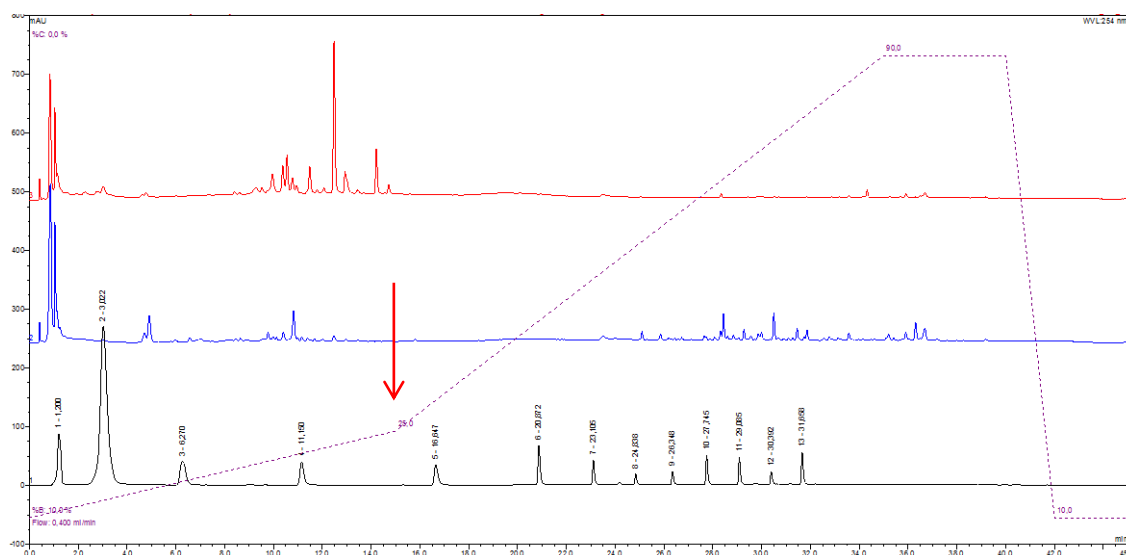


Figura 17- Cromatogramas em 254 nm, do extrato de *C. sylvestris* variedade *lingua* (CSL, em vermelho), do extrato de *C. sylvestris* variedade *sylvestris* (CSS, em azul) e da mistura de galatos (picos de 1 a 13, respectivamente, ácido gálico, galatos de metila, etila, propila, butila, pentila, hexila, heptila, octila, nonila, decila, undecila e dodecila), nas condições cromatográficas otimizadas. Destacado, está o galato de butila.

Após a seleção do galato de butila como padrão interno, foi definida a concentração considerando-se a altura média dos picos dos cromatogramas. A solução a 0,5 mg/mL mostrou-se ideal para uma boa detecção e para os cálculos futuros de razão de áreas. No entanto, um ensaio adicional foi necessário pois percebeu-se que para *C. sylvestris* existe um importante efeito de matriz. Sendo assim, soluções em cinco concentrações de galato de butila (0,3, 0,4, 0,5 0,6 e 0,7 mg/mL) foram preparadas e adicionadas de 50 mg do pó das folhas das duas variedades. Foi possível observar claramente o efeito da matriz sobre a área do padrão interno (**Figura 18**). Nota-se que a matriz da variedade *lingua* exerce maior efeito que a matriz da variedade *sylvestris*, principalmente por conta da presença de, provavelmente, taninos. Também é importante destacar que as inclinações das retas não diferem entre si (coeficiente de variação do coeficiente angular entre as três retas = 6,1%), mostrando que mesmo em concentrações crescentes nas diferentes matrizes, a resposta é a mesma. Isto mostra, mais uma vez, a adequabilidade do galato de butila como padrão interno.

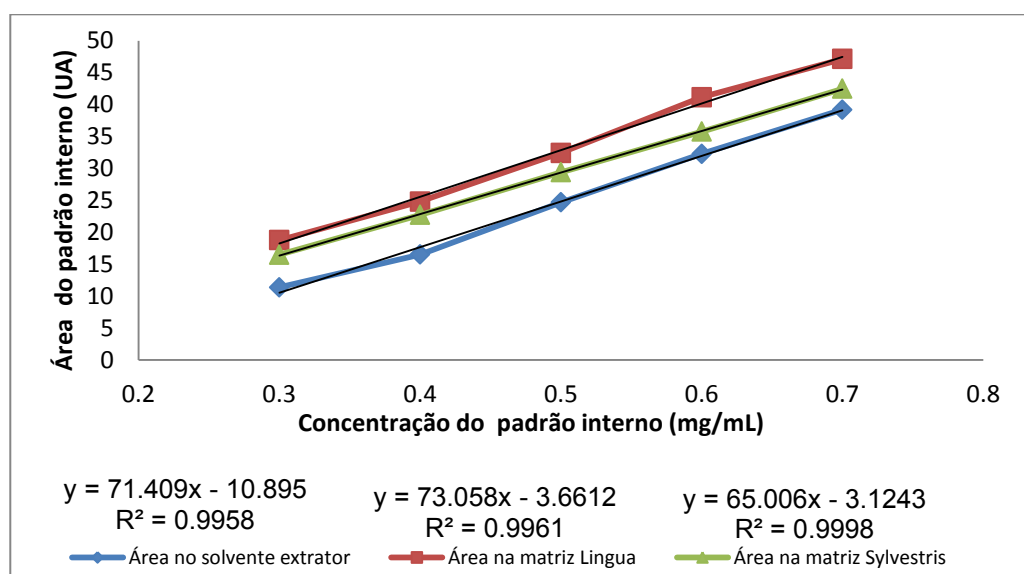


Figura 18- Efeito das matrizes das duas variedades de *C. sylvestris* sobre o padrão interno, galato de butila, em cinco diferentes concentrações.

5.1.6 Avaliação da precisão

A precisão de um método analítico representa a dispersão dos resultados obtidos a partir de análises múltiplas de uma mesma amostra, amostras semelhantes ou padrões, sob as mesmas condições definidas para o método em questão (GREEN, 1996; RIBANI, 2004).

Para a avaliação do desempenho do sistema cromatográfico, ou seja, a precisão instrumental, calculou-se o desvio padrão relativo do tempo de retenção e da área do pico do padrão interno, injetado em sextuplicata. Os desvios-padrão relativos obtidos foram de 0,08% e 1,12%, respectivamente. Tais valores foram considerados satisfatórios visto que para tal parâmetro, espera-se um valor de desvio padrão relativo da ordem de 1% (GREEN, 1996).

A repetibilidade e precisão intermediária foram determinadas em dois dias diferentes, pelo preparo e injeção de seis replicatas da amostra representativa de *C. sylvestris* (ARCS) contendo a solução padrão interno a 0,5 mg/mL. Dois picos foram selecionados a partir dos cromatogramas resultantes (**Fig. 16A**) e incluiu tanto um flavonoide glicosilado (narcisina, detectada a 254 nm) quanto um diterpeno clerodânico (casearina D, detectada a 235 nm). A razão entre as áreas destes picos selecionados pela área do padrão interno (detectada a 254 nm) foi usada para os cálculos dos desvios-padrão relativos, cujos resultados podem ser visualizados na **Tabela 12**.

Tabela 12- Resultados obtidos para determinação da precisão do método analítico para análise simultânea de compostos fenólicos e diterpenos clerodânicos por UHPLC-DAD.

% intervalo	Amostras	Razão: área flavonoide/PI		Razão: área diterpeno/PI	
		Dia 1	Dia 2	Dia 1	Dia 2
100%	1	0,604	0,566	0,199	0,127
	2	0,611	0,586	0,193	0,136
	3	0,664	0,635	0,200	0,145
	4	0,710	0,701	0,199	0,140
	5	0,734	0,708	0,205	0,143
	6	0,627	0,640	0,196	0,131
	Média	0,659	0,639	0,199	0,137
	Desvio Padrão	0,054	0,058	0,004	0,007
	DPR (%)	8,20	9,07	2,02	5,05
	Variância (S^2)	$2,92 \times 10^{-3}$	$3,36 \times 10^{-3}$	$1,61 \times 10^{-5}$	$4,78 \times 10^{-5}$

No primeiro e no segundo dia, o DPR das razões entre a área do pico do flavonoide glicosilado e a área do pico do padrão interno foi de 8,20 e 9,07%, respectivamente. Já para a razão entre a área do pico do diterpeno clerodânico e a área do pico do padrão interno foi de 2,02% no primeiro dia e 5,04% no segundo dia. Considerando a elevada complexidade da matriz da amostra, os valores obtidos (DPR inferiores a 10%) demonstram boa precisão durante a preparação e injeção das mesmas.

Finalmente, a avaliação da precisão intermediária foi feita em função do Teste F. Para os compostos fenólicos, comparando-se os resultados obtidos nos dois diferentes dias, verifica-se que o quociente das variâncias obtidas $F=1,15$ é menor que o valor tabelado 5,05 (para $n=6$). O mesmo foi observado para os diterpenos clerodânicos, cujo quociente das variâncias obtidas foi de $F=2,97$ também é menor que o valor tabelado. Sendo assim, verifica-se que não há diferença significativa entre os resultados das duas precisões, em um nível de 5% de confiança.

5.1.7 Avaliação linearidade e obtenção de curvas analíticas

A obtenção das curvas analíticas a partir de diluições seriadas da solução-mãe permitiu a obtenção de coeficientes de correlação lineares bastante satisfatórios para os dois padrões utilizados: rutina, $R^2 = 0,9990$ e caseargrewiina F, $R^2 = 0,9994$. A **Tabela 13** e as **Figuras 19** e **20** a seguir apresentam os dados de cada curva de analítica obtida.

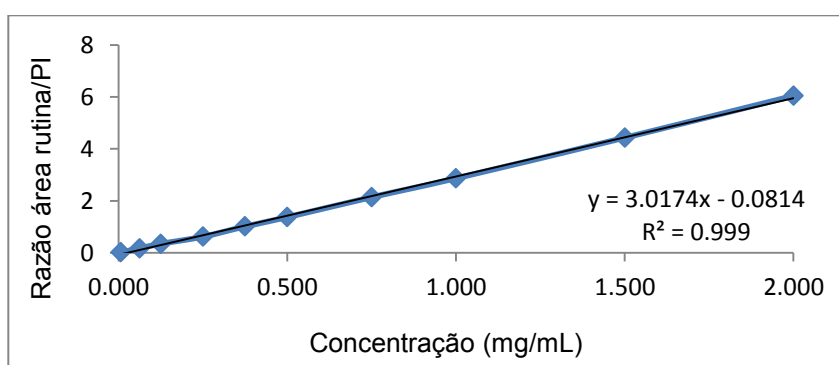


Figura 19- Curva analítica da rutina apresentando a regressão linear.

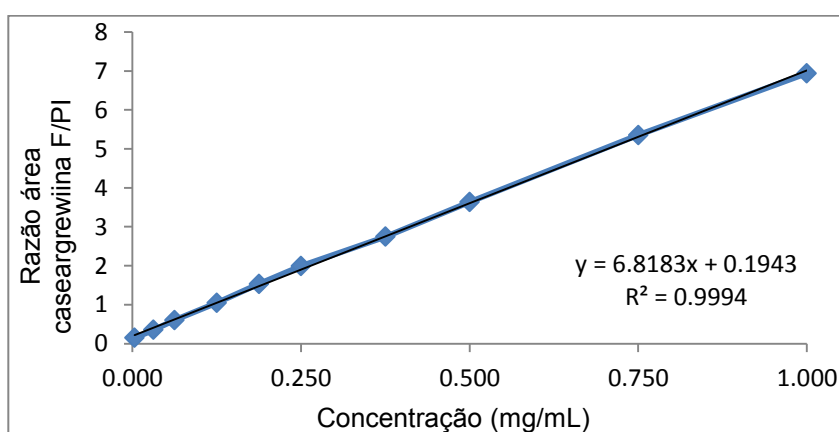


Figura 20- Curva analítica da caseargrewiina F apresentando a regressão linear.

Tabela 13- Dados experimentais das curvas analíticas (rutina e caseargrewiina F), em função do padrão interno galato de butila (n=3).

Conc. (%)	Área			Razão de áreas		Média		Desvio padrão		DPR (%)	
	Rutina	Cas. F	PI	Rutina/PI	Cas. F/PI	Rutina/PI	Cas. F/PI	Rutina/PI	Cas. F/PI	Rutina/PI	Cas. F/PI
1,25	0,224	2,277	14,691	0,015	0,155	0,014	0,147	0,001	0,007	6,09	5,02
	0,224	2,285	15,904	0,014	0,144						
	0,219	2,279	16,152	0,014	0,141						
	2,781	5,849	15,731	0,177	0,372						
12,5	2,885	5,871	16,881	0,171	0,348	0,173	0,353	0,004	0,017	2,2	4,69
	2,928	5,867	17,253	0,170	0,340						
	6,079	10,779	17,595	0,346	0,613						
	6,186	10,809	17,910	0,345	0,604						
25	6,226	10,818	18,090	0,344	0,598	0,345	0,605	0,001	0,007	0,21	1,22
	11,831	20,326	19,155	0,618	1,061						
	12,171	20,312	19,675	0,619	1,032						
	12,034	20,262	19,350	0,622	1,047						
50	18,523	28,303	18,226	1,016	1,553	0,619	1,047	0,002	0,014	0,36	1,37
	18,728	28,254	18,525	1,011	1,525						
	18,878	28,292	18,707	1,009	1,512						
	26,576	38,829	19,346	1,374	2,007						
75	26,829	38,877	19,559	1,372	1,988	1,012	1,530	0,004	0,021	0,36	1,35
	26,815	38,779	19,614	1,367	1,977						
	42,971	54,509	19,710	2,180	2,766						
	42,250	54,568	19,993	2,113	2,729						
100	42,373	54,524	19,969	2,122	2,731	2,139	2,742	0,036	0,021	1,70	0,75
	57,664	73,245	20,081	2,872	3,648						
	57,784	73,265	20,117	2,872	3,642						
	57,850	73,272	20,266	2,855	3,616						
150	88,829	107,526	19,990	4,444	5,379	2,866	3,635	0,010	0,017	0,35	0,47
	88,835	107,567	20,086	4,423	5,355						
	89,225	107,433	20,204	4,416	5,318						
	122,238	141,007	20,363	6,003	6,925						
200	88,829	107,526	19,990	4,444	5,379	4,428	5,351	0,014	0,031	0,32	0,58
	88,835	107,567	20,086	4,423	5,355						
300	89,225	107,433	20,204	4,416	5,318	6,040	6,938	0,037	0,013	0,62	0,18
	122,238	141,007	20,363	6,003	6,925						
	122,944	140,582	20,229	6,078	6,950						
	122,420	140,666	20,272	6,039	6,939						

5.2 PARTE 2- Estudo metabolômico de *C. sylvestris*

5.2.1 Obtenção de padrões de diterpenos clerodânicos (casearinas)

Para as futuras análises quantitativas do teor de diterpenos nas amostras de casearia, fez-se necessário o isolamento de padrões de casearinas. Sendo assim, tendo-se por base estudos anteriores do nosso grupo de pesquisa, uma fração rica em diterpenos proveniente de *C. sylvestris* var. *sylvestris* foi fracionada e três casearinas foram isoladas.

A casearina X (243,2 mg, oriunda da fração 3.1) foi obtida na forma de um sólido branco-amarelado, de aspecto levemente oleoso. O espectro de massas de alta resolução no modo positivo (Apêndice IX-A) apresentou o sinal característico do íon $[M+Na]^+$ com o respectivo valor de m/z de 555,2959, correspondente à massa teórica calculada para tal molécula cationizada de 555,2928. A banda de absorção no UV em 235,0 nm é característica da região do dieno conjugado das casearinas.

A Casearina J (60,7 mg, também oriunda da fração 3.1) foi obtida na forma de um pó branco. O espectro de massas de alta resolução no modo positivo (Apêndice IX-B) apresentou os sinais característicos dos íons $[M+Na]^+$ e $[M+K]^+$ com os valores de m/z de 585,3027 e 601,2767, correspondentes às massas teóricas calculadas para as respectivas moléculas cationizadas de 585,3034 e 601,2773. A banda de absorção no UV em 235,0 nm é também característica da região do dieno conjugado das casearinas.

Por fim, a caseargrewiina F (320 mg, oriunda da fração 3.4) foi isolada na forma de um pó branco. O espectro de massas de alta resolução no modo positivo (Apêndice IX-C) apresentou os sinais característicos dos íons $[M+Na]^+$ e $[M+K]^+$ com os respectivos valores de m/z de 527,2615 e 543,2357, correspondentes às massas teóricas calculadas para as moléculas cationizadas de 527,2615 e 543,2355, respectivamente. A banda de absorção no UV em 235,0 nm é característica da região do dieno conjugado das casearinas.

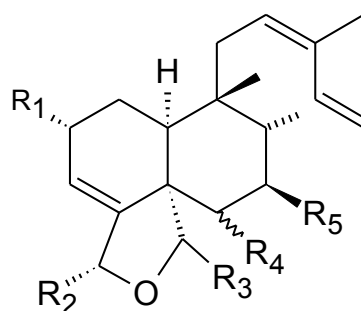
Por serem já bastante conhecidas, a identificação inequívoca foi feita com base nos dados de NMR de ^{13}C das casearinas solubilizadas em piridina comparados com os dados da literatura, conforme trabalho de Passareli (2010). Na **Tabela 14** estão compilados os dados experimentais com os dados da literatura para as três casearinas isoladas, enquanto que os espectros de ^{13}C encontram-se dispostos no Apêndice X.

Tabela 14- Dados de NMR de ^{13}C (NMR 600 MHz Bruker Avance III HD 600) obtidos para as três casearinas em piridina- d_5 isoladas neste trabalho.

C	Casearina X	Ref.*	Casearina J	Ref.*	Caseargrewiina F	Ref.*
1	27,8	27,6	26,8	26,5	28,0	27,7
2	67,6	67,4	73,4	73,3	67,4	67,3
3	121,8	121,6	124,0	123,8	121,8	121,6
4	148,0	147,8	144,7	144,5	147,7	147,7
5	55,0	54,8	55,2	55,0	54,9	54,8
6	73,0	72,8	74,9	74,4	73,0	72,8
7	38,2	38,1	76,6	76,4	38,2	38,1
8	37,2	37,0	42,0	41,7	37,8	37,7
9	38,8	38,6	40,3	40,1	38,7	38,6
10	37,9	37,7	37,1	36,9	37,1	37,0
11	30,1	30,0	31,2	30,9	30,1	30,0
12	128,6	128,4	127,8	127,5	128,6	128,4
13	134,2	134,0	134,3	134,1	134,1	134,0
14	134,6	134,4	134,5	134,3	134,6	134,4
15	114,8	114,6	115,0	114,7	114,7	114,6
16	20,9	20,7	20,6	20,4	20,8	20,7
17	16,3	16,1	12,0	11,7	16,2	16,1
18	97,5	97,3	97,4	97,1	97,5	97,4
19	99,4	99,2	99,3	99,1	99,3	99,1
20	25,7	25,5	26,2	26,0	25,7	25,5
OAc 1.1	170,3	170,0	170,0	169,7	170,3	170,1
OAc 1.2	22,2	22,0	22,0	21,8	21,6	21,4
OAc 2.1	-	-	-	-	171,0	170,8
OAc 2.2	-	-	-	-	22,1	22,0
OBu 1.1	173,6	173,4	173,9	173,6	173,5	173,3
OBu 1.2	36,9	36,7	37,0	36,8	36,8	36,7
OBu 1.3	19,4	19,2	19,3	19,1	19,4	19,2
OBu 1.4	14,2	14,0	14,3	14,1	14,1	14,0
OBu 2.1	173,4	173,3	173,7	173,4	-	-
OBu 2.2	36,9	36,7	36,9	36,7	-	-
OBu 2.3	19,1	18,9	19,1	18,9	-	-
OBu 2.4	14,1	13,8	14,2	13,8	-	-
OMe			57,1	56,9	-	-

* Passareli (2010)

De forma geral, todos os valores de deslocamentos químicos das casearinas isoladas são condizentes com os valores descritos na literatura. Além disso, para todas elas os valores dos deslocamentos químicos do sistema decalínico são bastante parecidos, fazendo que a diferenciação seja feita em função dos valores de deslocamentos químicos correspondentes às cadeias laterais, conforme pode ser visualizado na **Figura 21** a seguir.



Substância	R1	R2	R3	R4	R5
Caseargrewiina F	OBu	OAc	OAc	OH	H
Casearina X	OBu	OBu	OAc	OH	H
Casearina J	OMe	OBu	OAc	OH	OBu

Figura 21- Padrões de casearinas isoladas neste estudo.

5.2.2 Identificação de casearinas em *C. sylvestris* var. *sylvestris*

5.2.2.1 Análise qualitativa utilizando UHPLC-DAD-ESI-Orbitrap

Desreplicação é o processo de identificar substâncias já conhecidas presentes em matrizes complexas dentre aqueles ainda inéditos e de possível interesse, poupando tempo e recursos materiais (WOLFENDER et al., 2010). Trata-se de uma análise qualitativa em misturas complexas, na qual se utiliza técnicas hífenadas entre um elemento de separação de alta eficiência com um detector com capacidade de varredura espectrométrica ou espectroscópica (FUNARI et al., 2013).

Dentre os métodos comumente utilizados nos estudos de desreplicação, o acoplamento entre a cromatografia líquida de alta eficiência com a espectrometria de massas de alta resolução (HPLC-HRMS) permite a aquisição de informações estruturais, tais como fórmula e massa molecular de alta acurácia, combinadas com o tempo de retenção ou dados de fragmentação (MS/MS). Outra vantagem é a excelente detectabilidade da técnica, que permite observar analitos presentes na escala de nanogramas, substâncias minoritárias. Além disso, a disponibilidade de banco de dados comerciais abrangentes ou elaborados *in-house* permitem uma rápida e eficiente identificação a partir dos dados obtidos. No entanto, é importante pontuar que a co-injeção de padrões de referência, estudos de fragmentação e a identificação de substâncias isoladas por NMR devem ser também utilizadas no processo de

confirmação da identidade das substâncias desreplicadas (CHAGAS-PAULA et al., 2015).

A estratégia para a desreplicação por UHPLC-DAD-ESI-Orbitrap contemplou a aquisição dos dados de massas em alta resolução, o tratamento dos dados no software MzMine e a utilização de um banco de dados de casearinas especialmente construído para este estudo. Desta forma foi possível a identificação de 15 casearinas, além do flavonóide glicosilado rutina, no extrato bruto de *C. sylvestris* var. *sylvestris*, sendo que todas as substâncias propostas já foram previamente descritas para a espécie. Para cada substância identificada, os tempos de retenção obtidos no UHPLC-DAD-ESI-Orbitrap e as massas (m/z) encontradas foram confirmados através da busca destes dados diretamente no cromatograma de íon extraído, no software Xcalibur. Os cromatogramas de íon extraído contendo cada substância identificada estão dispostos no Apêndice XI.

Para que a identificação de substâncias desconhecidas por desreplicação seja bem sucedida, é imprescindível assinalar corretamente a composição elementar e, portanto, tanto a exatidão das massas e a acurácia espectral são os dois fatores críticos. É consenso que, uma exatidão de massas de 3 ppm aliada a 2% de erro isotópico remove pelo menos 95% de falsos candidatos para uma dada composição elementar (GLAUSER et al., 2013). Portanto, os erros de massa encontrados foram considerados excelentes, visto que variaram 0 a 1,966 ppm. A **Tabela15** a seguir evidencia os resultados encontrados.

Tabela 15- Dados de desreplicação de *C. sylvestris* var. *sylvestris* obtidos por UHPLC-DAD-ESI-Orbitrap.

t_R (min) ^a	m/z (HRMS) encontrada	Aduto	Massa exata correspondente	Erro (ppm)*	UV max (nm)	Desreplicação ^b	Massa exata dos candidatos ^c
Modo positivo							
23,37	435,2382	[M+H] ⁺	434,2309	0,921	232,3	Dit 24** Dit 25**	434,2305
25,35	571,2888	[M+Na] ⁺	548,2996	1,966	225,0	Casearina M	548,2985
26,27	557,2719	[M+Na] ⁺	534,2827	0,416	233,0	Casearina L ou A	534,2829
26,79	571,2870	[M+Na] ⁺	548,2978	1,317	232,7	Casearina D**	548,2985
27,39	557,2730	[M+Na] ⁺	534,2838	1,643	232,0	Casearina L ou A	534,2829
27,66	527,2615	[M+Na] ⁺	504,2723	0,044		Casearina H	504,2723
27,76	527,2623	[M+Na] ⁺	504,2731	1,543	230,0	Caseargrewiina F	504,2723
28,83	497,2509	[M+Na] ⁺	474,2617	0,257	233,0	Casearina S**	474,2618
29,39	555,2934	[M+Na] ⁺	532,3042	1,086	225,2	Casearina I Caseargrewiina A	532,3036
29,81	555,2924	[M+Na] ⁺	532,3032	0,793	228,4	Casearina X	532,3036
29,95	585,3030	[M+Na] ⁺	562,3138	0,750	228,8	Casearina J**	562,3142
30,91	541,2771	[M+Na] ⁺	518,2879	0,193	232,0	Casearvestrina B	518,2880
31,38	627,3134	[M+Na] ⁺	604,3242	1,029	232,0	Casearina O ou N	604,3248
33,51	669,3603	[M+Na] ⁺	646,3711	0,962	232,0	Casearina C	646,3717
34,11	613,3710	[M+Na] ⁺	590,3818	0,207	234,0	Casearina E	590,3819
Modo negativo				0,921			
8,653	609,1467	[M-H] ⁻	610,1540	1,966	255,0/355,0	Rutina	610,1534

^acorrespondente ao equipamento utilizado (UHPLC-DAD-ESI-Orbitrap; para referência: tempo de retenção do padrão interno (235 nm) = 15,14 min); ^bcandidatos já identificados na espécie *C. sylvestris* (banco de dados de casearinas, Apêndice III); ^cmassa exata calculada (H= 1,00783, C = 12,00000, O = 15,99492, Na = 22,98977, carga = 0,00055); *Erro (exatidão) em ppm = $(m_{\text{exata}} - m_{\text{medida}})/(m_{\text{exata}}) \times 10^6$; ** substâncias que tiveram a desreplicação confirmada a partir da análise da substância isolada.

5.2.2.2 Validação dos dados obtidos

A validação dos dados obtidos por desreplicação foi feita através do fracionamento e coleta em pequena escala dos picos majoritários presentes no extrato bruto liofilizado de *C. sylvestris* var. *sylvestris* e posterior análise por espectrometria de massas de alta resolução (HRMS).

A identificação considerou a massa (m/z) em alta resolução do íon molecular obtida experimentalmente e comparação com a massa (m/z) do íon molecular calculada com base na fórmula molecular de casearinas já descritas para o gênero *Casearia*, o histórico da presença do candidato na espécie *C. sylvestris*, a comparação dos tempos de retenção da substância isolada e padrões de casearinas e o perfil de fragmentação por espectrometria de massas. Os cromatogramas apresentados nas **Figuras 22 e 23** mostram a região do extrato bruto da variedade *sylvestris* rica em casearinas e o fracionamento realizado (frações de #1 a #40).

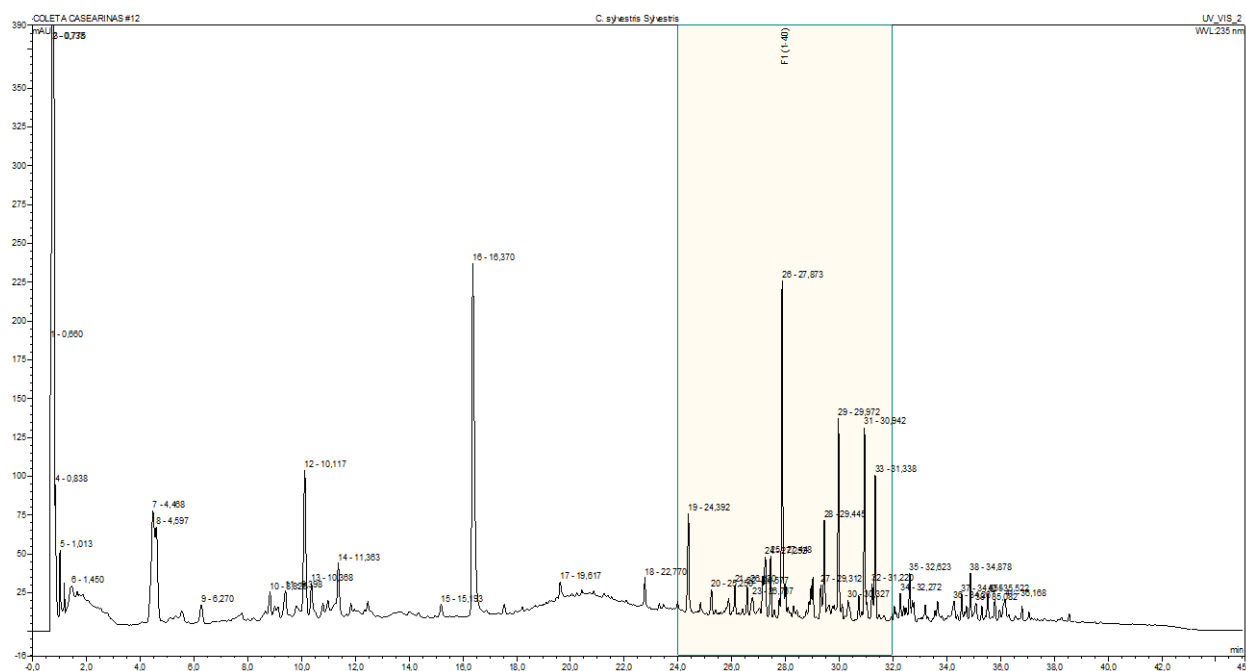


Figura 22- Cromatograma de *C. sylvestris* var. *sylvestris* (235 nm), destacando (em rosa) a região rica em diterpenos clerodânicos

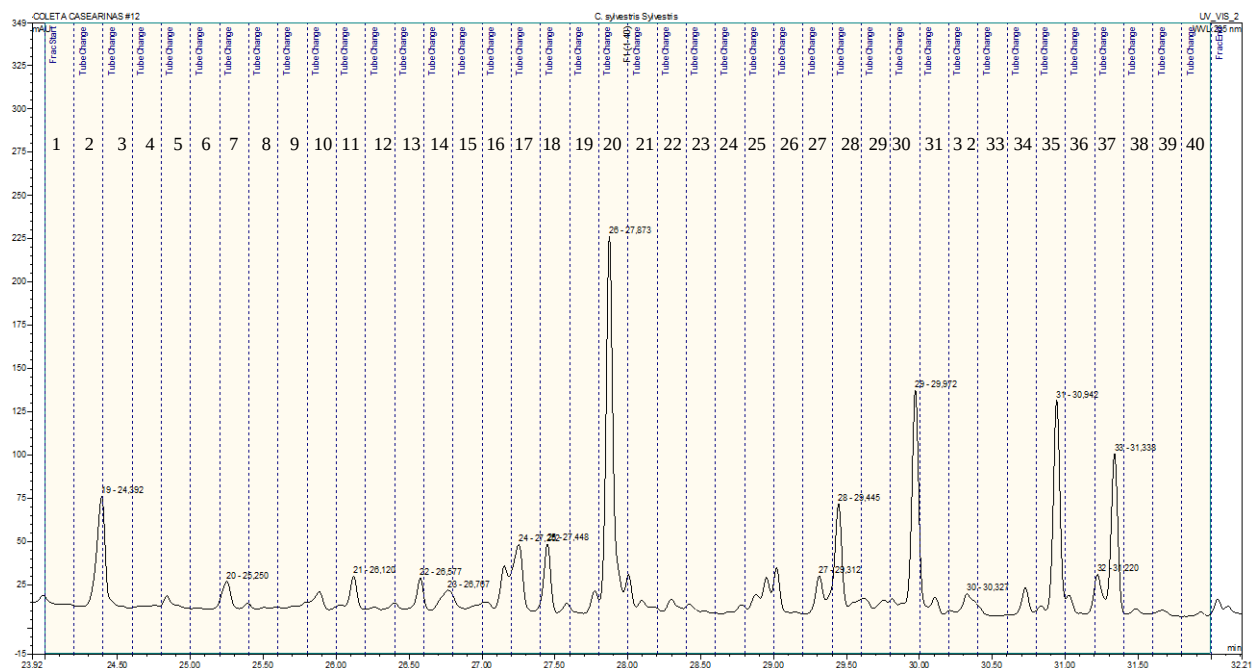


Figura 23- Ampliação da região do cromatograma onde ocorre a separação e a coleta das frações referentes aos diterpenos clerodânicos de *C. sylvestris* var. *sylvestris*.

Considerando a análise dos picos majoritários foi possível identificar, utilizando esta estratégia, três casearinas (picos #20, #31 e # 37). A **Tabela 16** a seguir evidencia os resultados encontrados.

A primeira substância identificada trata-se do pico majoritário em 27,873 minutos (pico #20), cuja massa do íon molecular em alta resolução foi de $[M]^+ = 548,2951$. Dentre as opções oriundas do banco de dados de massas de alta resolução de casearinas existem vários isômeros, como a casearina D, casearina M, casearina P e caseamembrina P. A caseamembrina P foi descartada por não ter sido isolada da espécie *C. sylvestris*. Já a casearina P, por conseguinte, também foi excluída por não apresentar na molécula um grupo butanoato, já que tal perda foi verificada no espectro de massas. Restaram os isômeros casearina D e casearina M de fórmula molecular $C_{30}H_{44}O_9$ e massa MM = 548,2985 (**Figura 24**).

Tabela 16- Tabela de identificação das casearinas coletadas de *C. sylvestris* var. *sylvestris* por UHPLC-DAD e analisadas por espectrometria de massas com inserção direta (microTOF Q II - ESI-TOF).

Pico (#)	t_R (min)	Ionização modo positivo			UV max (nm)	Candidatos ^a	Massa exata [M] ⁺ dos candidatos ^c
		Alta resolução (m/z)		MS/MS (m/z)			
		[M + Na] ⁺	[M] ⁺				
20	27,873	5eV: 571,2849	548,2951	30 eV: 571 → 481, 423	236,2	Casearina P ^b	548,2616
						Caseamembrina P	548,2980
						Casearina D ^b	548,2980
						Casearina M ^b	548,2980
31	29,972	0eV: 497,2292	474,2394	35 eV: 497 → 437, 409, 349	233,7	Casearina S ^b	474,2612
						Caseamembrina N	562,2773
						Caseanigrescena A	562,2773
						Caseanigrescena C	562,2773
37	31,330	0eV: 585,3006	562,3108	0 eV: 585, 409, 392, 374, 360 40 eV: 392 → 297, 262	235,7	Casearina K ^b	562,2773
						Casearina Q ^b	562,2773
						Caseamembrina C	562,3136
						Caseamembrina L	562,3136
						Casearina J ^b	562,3136
Casearina J		0eV: 585,3027	562,3129				

^acandidatos descritos no banco de dados de casearinas (Apêndice III); ^bcandidatos já identificados na espécie *C. sylvestris*; ^cmassa exata calculada (H= 1,00783, C = 12,00000, O = 15,99492, Na = 22,98977).

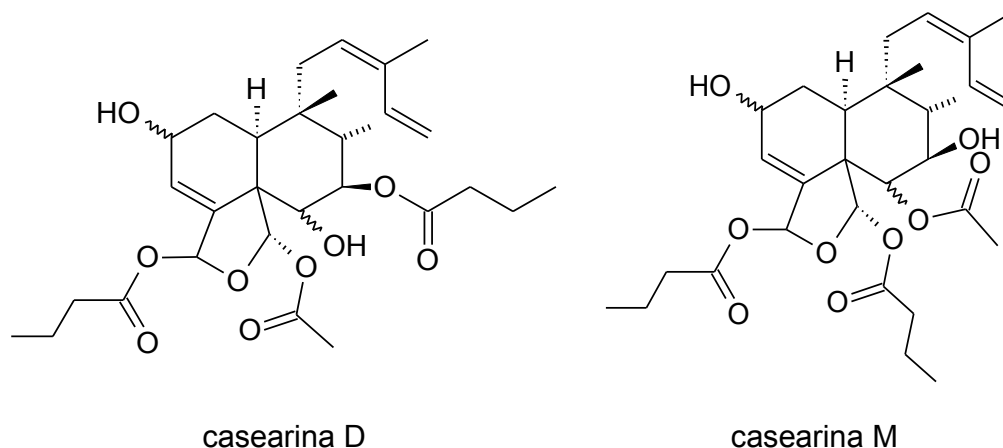


Figura 24- Estruturas químicas da casearina D e casearina M.

Através da análise do espectro de massas (MS/MS), as perdas características de um grupo butanoato na posição R2 (em m/z 483) seguida de um grupo acetato em R3 (m/z 423) levaram à identificação deste pico como sendo casearina D (**Figura 25**). Tais propostas de identificação através da análise do espectro de fragmentação estão em consonância com estudos de fragmentação de casearinas utilizando ionização por eletrospray (ESI-MS/MS), em que primeiro ocorre a perda neutra do substituinte em C-18 na forma de ácido carboxílico (R2) seguida da eliminação neutra de ceteno e água através de um mecanismo concertado que leva à formação de um íon aromático.

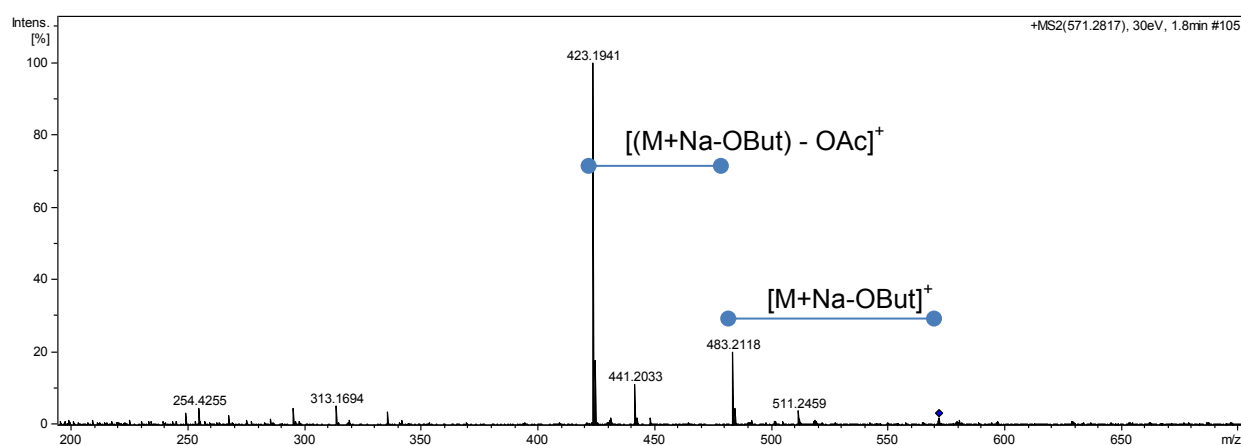


Figura 25- Espectro de massas (MS/MS) da casearina D (pico # 20).

A segunda substância identificada foi a casearina S (**Figura 26**), de fórmula molecular $C_{27}H_{38}O_7$ (pico #31, em 29,972 minutos, MM = 474,2618), por ser o único

candidato possível cuja massa do íon molecular em alta resolução ($[M]^+$ 474,2612) é correspondente à massa obtida experimentalmente ($[M]^+$ 474,2394). Além disso, o espectro de massas (MS/MS) mostra as perdas características de um grupo acetato em m/z 437 e de um grupo butanoato em m/z 409, conforme pode ser visualizado na **Figura 27**.

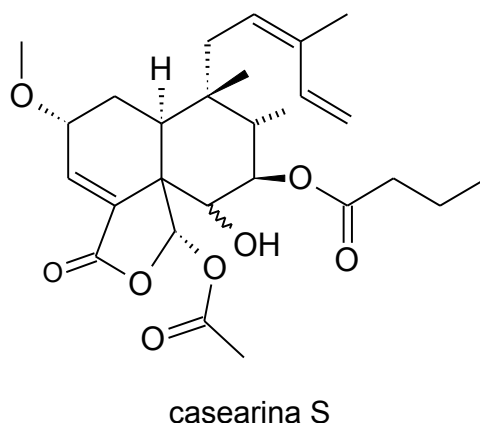


Figura 26- Estrutura química da casearina S.

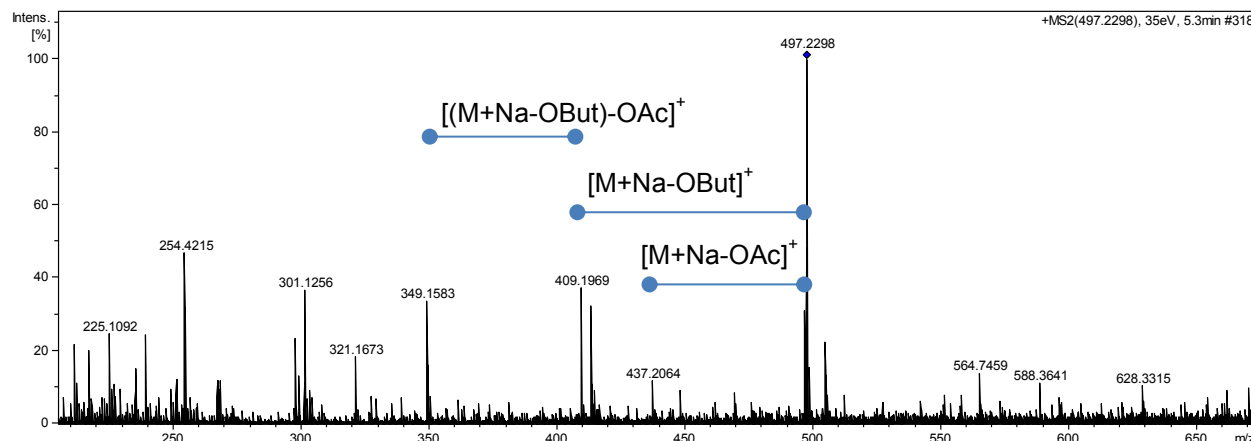


Figura 27- Espectro de massas (MS/MS) da casearina S (pico #31)

Finalmente, a última substância que pode ser identificada por esta metodologia correspondeu ao pico #37 em 31,330 minutos, com massa do íon molecular de $[M]^+$ = 562,3108. De acordo com o banco de dados, esta massa poderia corresponder às seguintes substâncias isoladas de *C. sylvestris*: casearina J ($[M]^+$ = 562,3136), casearina K ($[M]^+$ = 562,2773) ou casearina Q ($[M]^+$ = 562,2773), cujas estruturas das duas últimas seguem dispostas na **Figura 28**.

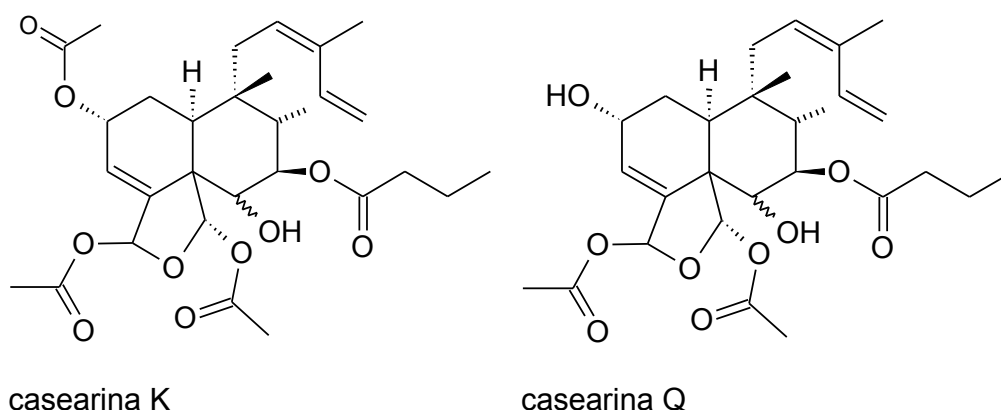


Figura 28- Estruturas químicas da casearina K e casearina Q

As casearinas Q e K foram excluídas uma vez que a massa para o íon molecular obtida experimentalmente difere bastante da massa exata calculada para este íon. Além disso, o tempo de retenção de um padrão autêntico de casearina J coincidiu com o tempo de retenção da substância em questão, sugerindo que a substância eluída em 31,330 minutos ser a casearina J (**Figura 29**).

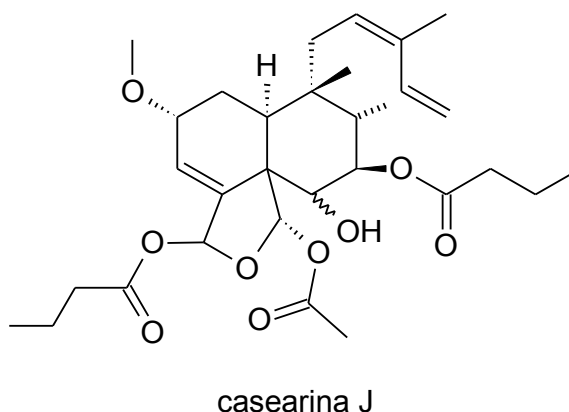


Figura 29- Estrutura química da casearina J.

Além disso, os tempos de retenção dos padrões de casearina X, de caseargrewiina F e de rutina, também corroboram com a proposta de identificação feita por desreplicação.

A partir dos dados obtidos foi possível, finalmente, atribuir as substâncias identificadas aos picos do cromatograma de *C. sylvestris* var. *sylvestris*, conforme pode ser visualizado na **Figura 30** a seguir. É importante pontuar que o cromatograma

apresentado foi obtido por UHPLC-DAD e, por este motivo, os tempos de retenção correspondentes às substâncias identificadas por UHPLC-DAD-ESI-Orbitrap (**Tabela 15**) variam cerca de 1 minuto.

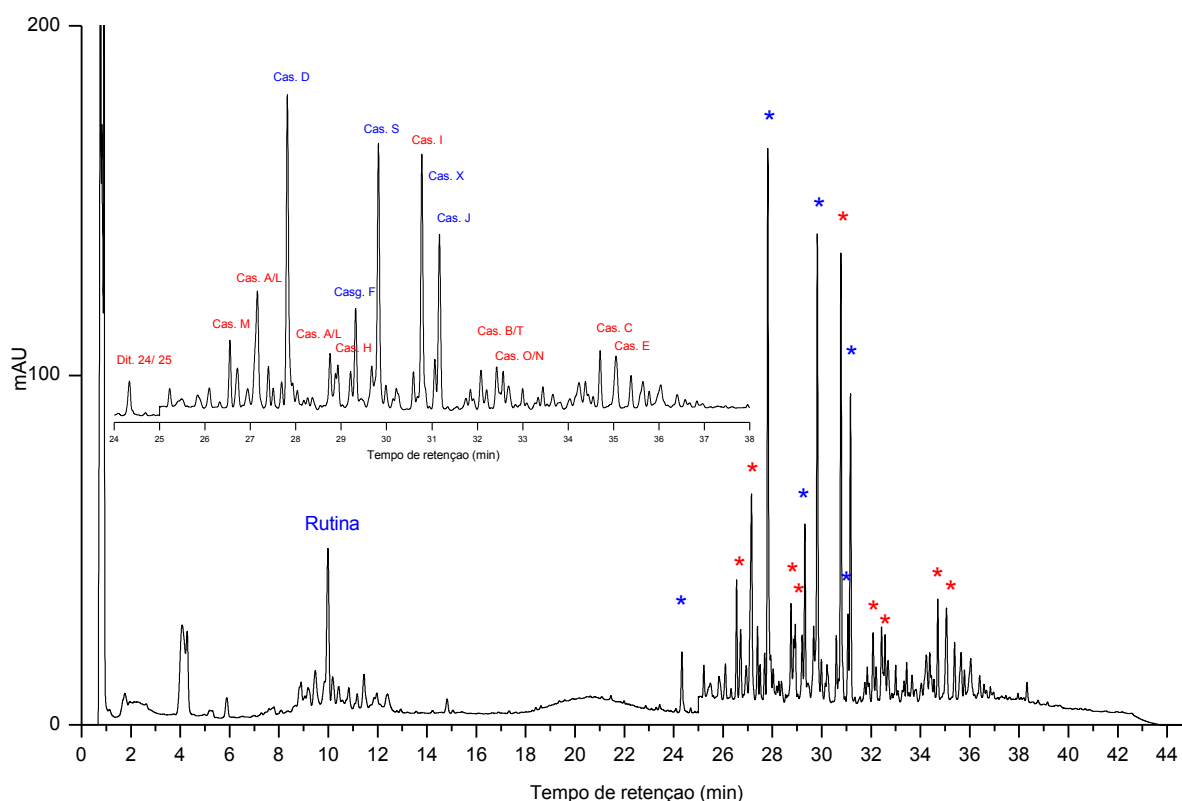


Figura 30- Cromatograma de *C. sylvestris* var. *sylvestris* obtido por UHPLC-DAD mostrando os picos correspondentes às substâncias identificadas. O cromatograma foi plotado em 254 nm (de 0 a 25 min) e 235 nm (de 25 a 45 min). Em vermelho, substâncias identificadas por desreplicação (UHPLC-DAD-ESI-Orbitrap). Em azul, substâncias também identificadas por desreplicação e confirmadas através do isolamento em pequena escala (UHPLC-DAD) e identificadas por inserção direta em espectrômetro de massas (microTOF Q II - ESI-TOF). Para a rutina, casearina x, casearina J e caseargrewiina F, os tempos de retenção também coincidem com os tempos de retenção de padrões autênticos.

5.2.3 Identificação dos compostos fenólicos em *C. sylvestris* var. *lingua*

5.2.3.2 Desreplicação por HRMS utilizando substâncias isoladas

Existem pouquíssimos dados acerca da presença e identidade dos compostos fenólicos em *C. sylvestris*. Embora a presença de glicosídeos fenólicos seja comum em espécies da família Salicaceae, muito pouco se sabe acerca da ocorrência destes compostos no gênero, principalmente em *C. sylvestris* (MOSADDIK et al., 2007). Neste sentido, existem relatos apenas da ocorrência de quatro derivados do ácido elágico em

extratos aquosos de *C. sylvestris* (DA SILVA et al., 2008), dois derivados do ácido gálico em extratos alcoólicos e tirosol (DA SILVA et al., 2009; WANG et al., 2009). No entanto, em estudos com amostras de *C. sylvestris* provenientes de áreas de cerrado do Estado de São Paulo, foram encontrados altos teores de rutina (SILVA et al.; 2003).

Um fracionamento inicial do extrato bruto liofilizado de *C. sylvestris* var. *lingua* (EBLL) por cromatografia de média pressão (sistema Puriflash 4100) resultou na obtenção de 2,33 g de uma fração rica em flavonoides e de outros compostos fenólicos (fração FRCF). A **Figura 31** a seguir demonstra a grande similaridade entre o extrato bruto liofilizado da variedade *lingua* (EBLL) e da fração FRCF obtida a partir deste extrato.

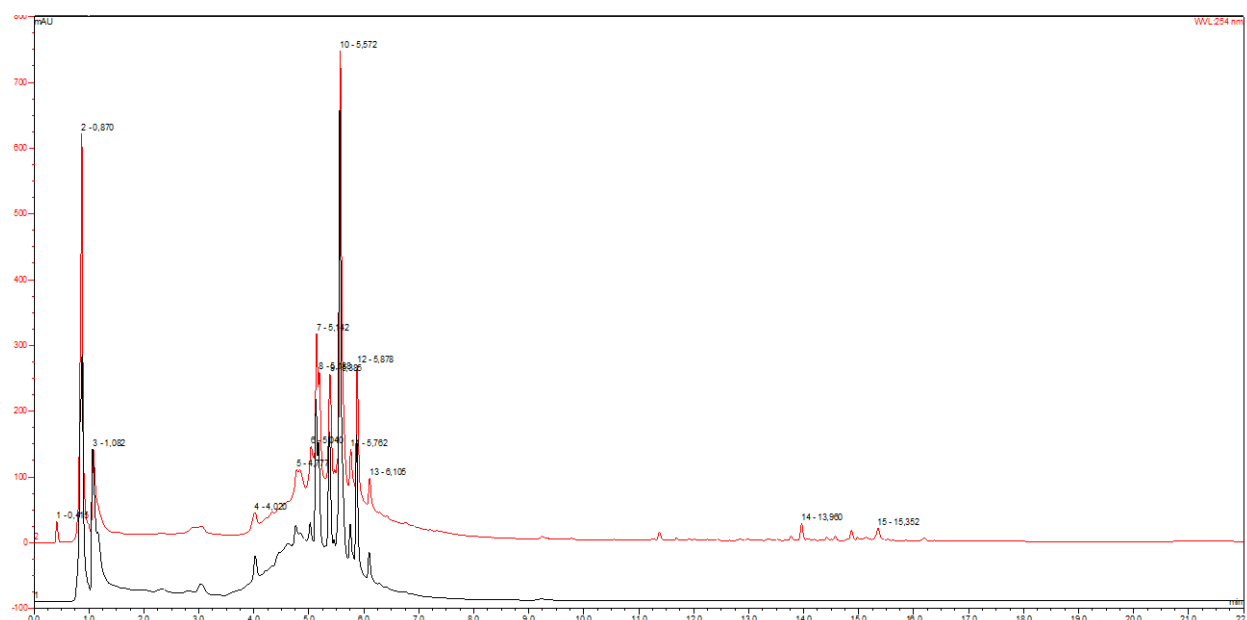


Figura 31- Perfil cromatográfico obtido em 254 nm do extrato bruto de *C. sylvestris* var. *lingua* (EBLL, em vermelho) e da fração enriquecida em compostos fenólicos (FRCF, em preto).

Observa-se que nesta fração estão presentes a maioria dos compostos fenólicos observados no extrato bruto. No entanto, pode-se perceber também que co-eluem com os picos referentes aos flavonoides (classe química confirmada pelos espectros no ultravioleta dos picos majoritários), provavelmente taninos (banda larga entre 4-6 min). Sendo assim, para que a fração FRCF pudesse ser fracionada, fez-se necessário uma etapa de extração em fase sólida com poliamida. Foram obtidas três subfrações, conforme pode ser observado na **Figura 32**.

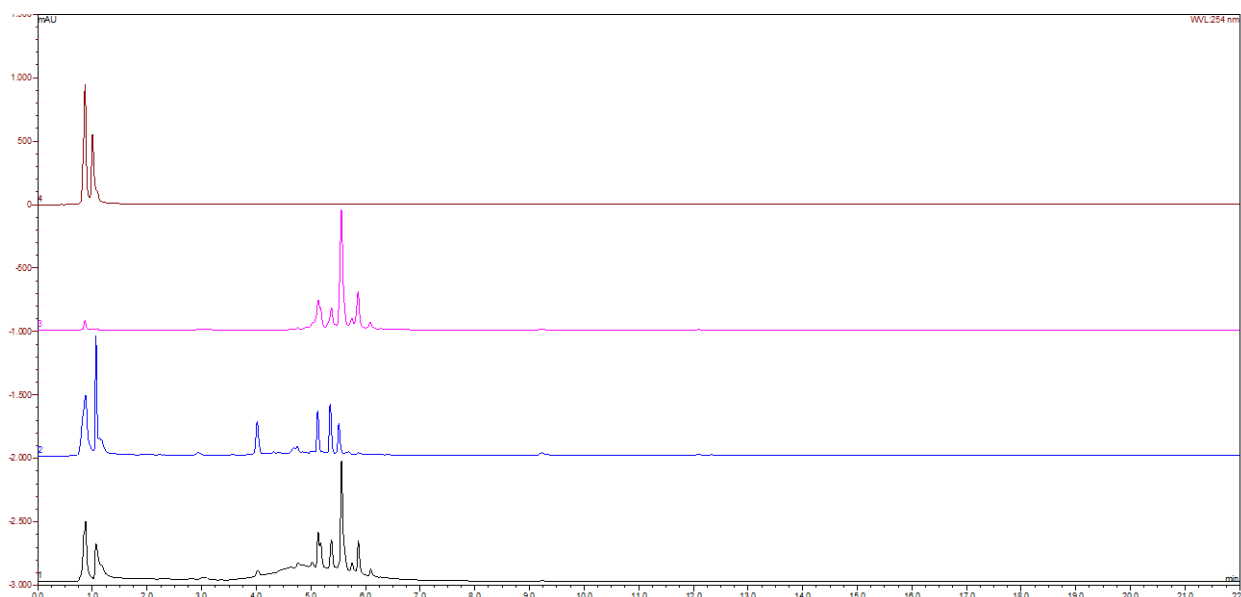


Figura 32- Cromatogramas das subfrações obtidas por EFS com poliamida da fração FRCF (254 nm). Em preto a fração tal e qual; em azul a fração hidro-metanólica; em rosa a fração metanólica rica em flavonoides e em vermelho a fração metanólica basificada com hidróxido de amônio.

Das subfrações obtidas, a fração metanólica (113 mg) foi a que apresentou-se rica em flavonoides glicosilados e foi a escolhida para o fracionamento por cromatografia líquida preparativa. A **Figura 33** a seguir mostra o cromatograma obtido em modo preparativo da fração em questão. Foram coletados 11 picos, os quais foram analisados na sequência por HPLC em modo analítico. Espectros de massas de alta resolução também foram obtidos e os resultados seguem dispostos na **Tabela 17** a seguir.

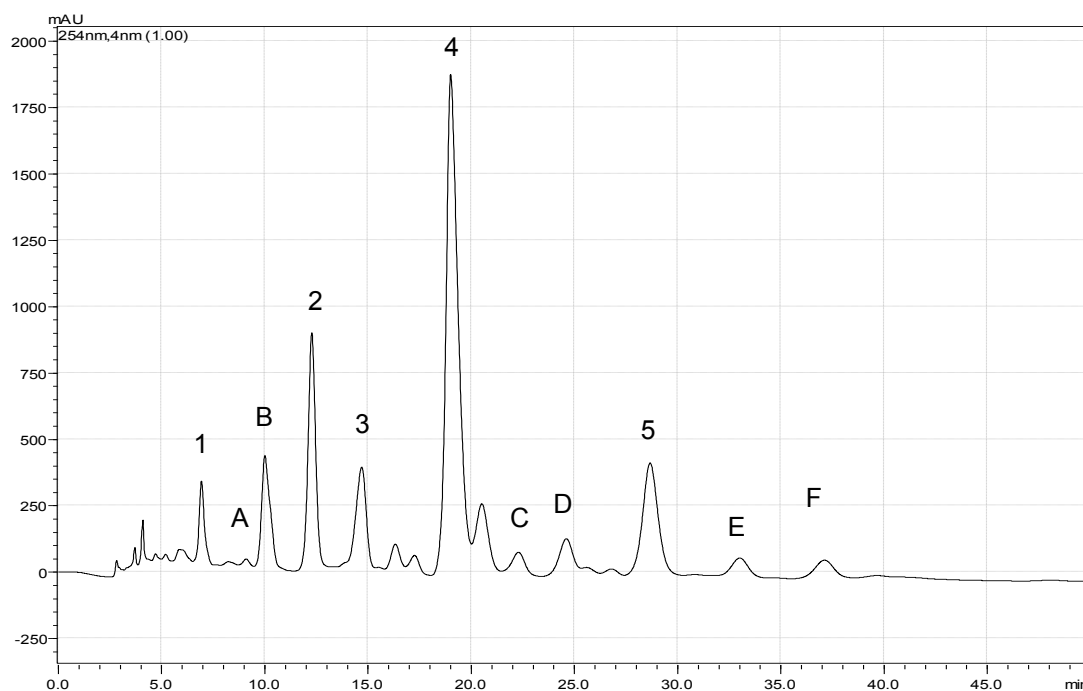


Figura 33- Cromatograma da fração metanólica em condição isocrática (20% de acetonitrila em água acidificada com 0,1% de ácido fórmico, 254 nm) durante o isolamento dos compostos fenólicos. As letras de A a F correspondem às substâncias identificadas por desreplicação (MS e MS/MS); os números de 1 a 5 correspondem substâncias que foram identificadas por elucidação estrutural (NMR).

Tabela 17- Picos coletados e quantidades das substâncias obtidas durante o isolamento dos compostos fenólicos de *C. sylvestris* var. *lingua*.

Pico (#)	t_R -HPLCprep (min)	Método de identificação		Massa (mg)
		Desreplicação (MS e MS/MS)	Elucidação estrutural NMR	
Pico 1	7	-	1	11,3
Pico 2	8	A	-	2,4
Pico 3	10	B	-	8,7
Pico 4	13	-	2	5,3
Pico 5	15	-	3	3,2
Pico 6	17	-	4	14,9
Pico 7	22	C	-	3,9
Pico 8	25	D	-	5,3
Pico 9	28	-	5	5,1
Pico 10	33	E	-	5,1
Pico 11	37	F	-	5,5

A estratégia para a desreplicação baseou-se, para cada substância isolada, na aquisição dos dados de massas em alta resolução e pesquisa em bancos de dados públicos. Desta forma foi possível a identificação de 15 flavonoides, sendo que quatro deles foram identificados em mistura.

De forma geral, todos os flavonoides identificados apresentaram uma, duas ou três unidades de açúcares, exceto para a (+)-catequina (**Tabela 18**). Além disso, de acordo com os espectros de massas obtidos (MS/MS), propôs-se que tais flavonoides são todos 3-O-glicosilados, devido ao padrão de fragmentação característico destes compostos. Para todos os casos observou-se a formação do radical aglicona $[Y_0-H]^- \bullet$ por clivagem homolítica entre a ligação glicosídica entre a aglicona e o açúcar (CUYCKENS; CLAEYS, 2004). Já a posição da glicosilação foi distinguida pela abundância relativa dos íons $[Y_0-H]^- \bullet$ e $[Y_0]^-$. Nos flavonoides 3-O-glicosilados, a abundância relativa dos íons $[Y_0-H]^- \bullet$ é maior que a abundância relativa dos íons $[Y_0]^-$ (ABLAJAN et al., 2006; ZHOU et al., 2014), o que ocorreu em todos os casos.

Além disso, para todos os casos em que foi proposta a presença de ramnose na composição do açúcar, mesmo quando em mistura, confirmou-se a existência dessa unidade glicosídica através da observação do duplete característico na faixa de 1,0 a 1,3 ppm ($J \sim 6.0$) através da análise de espectros de NMR de 1H .

Tabela 18- Dados de desreplicação de *C. sylvestris* var. *lingua* obtidos por HRMS (inserção direta, micrOTOF Q II - ESI-TOF).

Pico	m/z (HRMS) [M-H] ⁻	m/z (MS/MS)	Massa exata	Erro (ppm)	UV max (nm)	Substância	Massa exata
1	289,0729	30 eV: 289 → 245; 221; 205; 203; 187; 179; 165	290,0802	4,137	279	(+)-Catequina	290,0790
A	769,2203	55 eV: 769 → 314 (315)	770,2276	0,779	254 (265); 353	Isoramnetina-3-O-triglicosídeo (ramnose +ramnose+ glicose ou galactose)	770,2270
	579,1389	40 eV: 579 → 300 (301)	580,1462	5,861		Quercetina-3-O-diglicosídeo (ramnose+ arabinose ou xilose)	580,1428
B	593,1537	50 eV: 593 → 284 (285)	594,1610	4,376	256 (265); 353	Kaempferol-3-O-diglicosídeo (ramnose+glicose ou galactose)	594,1584
	463,0907	30 eV: 463 → 300 (301)	464,0980	5,387		Quercetina-3-O-glicosídeo (glicose ou galactose)	464,0955
2	609,1486	45 eV: 609 → 300 (301)	610,1559	4,097	256(265); 353	Quercetina-3-O-rutinosídeo (Rutina)	610,1534
3	623,1646	45eV: 623 → 314 (315)	624,1719	4,646	254 (265); 354	Isoramnetina-3-O-neohesperidosídeo	624,1690
4	623,1665	50 eV: 623 → 314 (315)	624,1738	7,690	255 (267); 354	Isoramnetina-3-O-rutinosídeo (Narcisina)	624,1690
C	433,0811	35 eV: 433 → 300 (301)	434,0884	8,063	256 (267); 354	Quercetina-3-O-glicosídeo (arabinose ou xilose)	434,0849
	477,1076	45 eV: 477 → 314 (315)	478,1149	7,948		Isoramnetina-3-O-glicosídeo (glicose ou galactose)	478,1111
D	767,2065	30 eV: 767 → 623; 314 (315)	768,2138	1,953	254(267); 352	Isoramnetina-3-O-[Ramnopiranosil-3- hidroxi-3-metilglutaroil-glucopiranosídeo	768,2113
	563,1448	35 eV: 563 → 284 (285)	564,1521	7,445		Kaempferol-3-O-diglicosídeo (ramnose+arabinose ou xilose)	564,1479
5	593,1553	40 eV: 593 → 314 (315)	594,1626	6,901	254 (265); 353	Isoramnetina-3-O- α -L-ramnopiranosil- (1→2)- α -L-arabinopiranosídeo	594,1585
E	621,1502	55 eV: 621 → 477, 314 (315)	622,1575	6,590	254 (267); 351	Isoramnetina-3-O-[3-hidroxi-3- metilglutaroil-galactopiranosídeo	622,1534
F	447,0975	40 eV: 447 → 314 (315)	448,1048	9,373	255 (265); 351	Isoramnetina-3-O-pentosídeo (arabinose ou xilose)	448,1006

5.2.3.3 Validação dos dados obtidos

A validação dos dados obtidos por desreplicação foi feita através da elucidação estrutural completa de cinco flavonoides utilizando NMR. Para efeitos didáticos, a numeração básica do núcleo aglicona de um flavonóide segue representado na **Figura 34** a seguir e será utilizada em toda a discussão referente à elucidação estrutural dessas substâncias. Todos os espectros referentes às elucidações estruturais discutidas a seguir estão organizados no Apêndice XII.

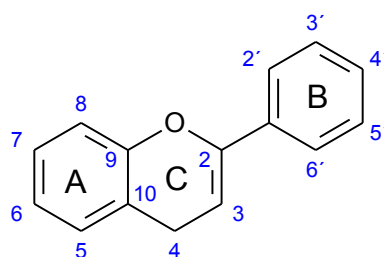


Figura 34- Numeração da estrutura básica dos flavonoides

5.2.3.2.1 Substância 1

A substância 1, uma flavan-3-ol, foi identificada como (2R,3S)-2,3-trans-(+)-catequina, ou seja (+)-catequina, cuja estrutura química segue representada na **Figura 35** a seguir.

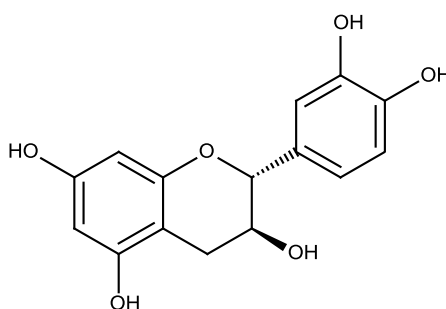


Figura 35- Estrutura química da (+)-catequina (substância 1).

A (+)-catequina foi obtida como um pó levemente amarelado (11,3 mg), com banda de absorção na região do ultravioleta em 279 nm. O espectro de massas de alta resolução (micrOTOF Q II - ESI-TOF) no modo negativo apresentou o sinal relativo ao íon $[M-H]^-$ com valor de m/z de 289,0729, consistente com o valor teórico de m/z

289,0718 correspondente ao íon desprotonado referente à fórmula empírica $C_{15}H_{14}O_6$. Foram também observados os sinais característicos em m/z 271 e 245.

O espectro de NMR de 1H dessa substância (**Tabela 19**) evidencia dois dupletos em δ 5,88 e 5,68, com $J = 2,1$ Hz referentes aos hidrogênios em relação meta H-6 e H-8 do anel A, um duplete em δ 4,47 com $J = 7,5$ Hz e um duplo duplete em δ 3,81 ($J = 2,1$ e 13,3 Hz) referentes à H-2 e H-3 do anel C, respectivamente, bem como dois duplo dupletos com deslocamentos e constantes de acoplamento condizentes com dois hidrogênios em H-4 nas posições axial e equatorial. Os demais dados de deslocamentos químicos dos hidrogênios do anel B mostram um duplete em δ 6,71 ($J = 2,0$ Hz), outro duplete em δ 6,68 ($J = 8,1$ Hz) e um duplo duplete em δ 6,58 (dd , $J = 8,2$ e 2,0 Hz), referentes à H2', H5' e H6', respectivamente. Já o espectro de NMR de ^{13}C apresentou 15 sinais similares àqueles já relatados para catequinas (SHEN et al., 1993), conforme pode ser visualizado na **Tabela 19** a seguir.

Tabela 19- Dados espectrométricos de NMR (1H) obtidos para a Substância 1 a 600 MHz, em DMSO- d_6 : (+)-catequina.

Posição	Substância 1		Referência (+)-catequina*	
	δH	δC	δH	δC
2	4,47 (d ; 7,5)	81,05	4,51 (d ; 7,3)	81,00
3	3,81 (dd ; 7,5; 13,3)	66,36	3,84 (m)	66,40
4	2,34 (dd ; 16,0; 8,1)	27,93	2,38 (ax ; dd ; 16,0; 7,9)	27,70
	2,65 (dd ; 16,0; 5,3)		2,68 (eq ; dd ; 16,0; 5,3)	
5		156,52		156,10
6	5,88 (d ; 2,1)	95,15	5,90 (d ; 2,2)	95,30
7		156,52		156,40
8	5,68 (d ; 2,1)	93,88	5,72 (d ; 2,2)	94,00
9		155,41		155,30
10		99,10		99,20
1'		130,62		130,70
2'	6,71 (d ; 2,0)	114,59	6,74 (d ; 1,9)	114,50
3'		144,93		144,80
4'		144,93		144,80
5'	6,68 (d ; 8,1)	115,14	6,70 (d ; 8,0)	115,10
6'	6,58 (dd ; 8,1; 2,0)	118,49	6,61 (dd ; 8,0; 1,9)	118,40
3'-OH	--- (pressat)		4,75 (d ; 4,7)	

5.2.3.2.2 Substância 2

A substância 2 é um flavonóide glicosilado e foi identificada como rutina (quercetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 6)- β -D-glucopiranosideo), cuja estrutura química segue na **Figura 36**.

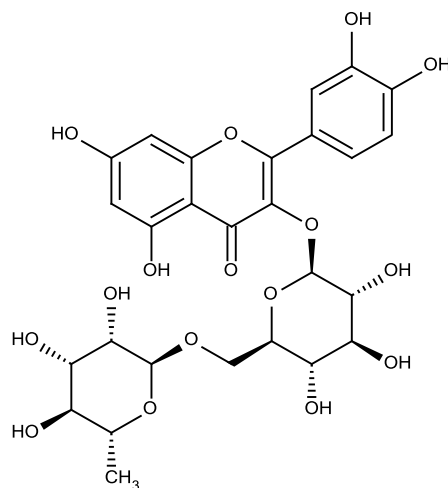


Figura 36- Estrutura química da rutina (substância 2)

A rutina foi obtida como um pó amarelo (4,0 mg), com bandas de absorção na região do ultravioleta em 256 (ombro em 265 nm) e 353 nm, características de flavonol. O espectro de massas de alta resolução (microTOF Q II - ESI-TOF) no modo negativo apresentou o sinal relativo ao íon $[M-H]^-$ com valor de m/z de 609,1486, consistente com o valor de 609,1461 calculado para o íon desprotonado correspondente à fórmula empírica $C_{27}H_{30}O_{16}$. O espectro de fragmentação do íon em m/z 609 mostrou a presença de um íon de m/z 300, referente à perda dos açúcares $[M-gli-ram-H]^-$. Por ser uma substância muito comum, foram obtidos sob as mesmas condições os espectros de massas e de NMR de um padrão autêntico de rutina e os dados encontrados foram idênticos ao da substância 2 em questão. A **Tabela 20** a seguir reúne os dados de NMR obtidos experimentalmente.

Tabela 20- Dados espectrométricos de NMR (^1H) obtidos para a substância 2 e para o padrão de rutina a 600 MHz, em DMSO- d_6 .

Estrutura	Posição	Substância 2	
		δH	δH (rutina; padrão)
Flavonol	2		
	3		
	4		
	5		
	6	6,19 (<i>d</i> ; 2,0)	6,14 (<i>s</i> ; broad)
	7		
	8	6,38 (<i>d</i> ; 2,0)	6,34 (<i>s</i> ; broad)
	9		
	10		
	1'		
	2'	7,54 (<i>dd</i> ; 2,2; 8,3)	7,54 (<i>dd</i> ; 1,7; 8,2)
	3'	6,84 (<i>d</i> ; 8,2)	6,82 (<i>d</i> ; 8,2)
	4'		
	5'		
	6'	7,52 (<i>d</i> ; 2,2)	7,51 (<i>s</i> ; broad)
Açúcar 1	3'	3,71 (<i>s</i>)	3,71 (<i>s</i>)
	1'' - 6''	5,34 (<i>d</i> ; 7,3)	5,31 (<i>d</i> ; 7,2)
	2''	3,00 – 3,75 (<i>m</i>)	3,00 – 3,75 (<i>m</i>)
Açúcar 2	1'''	4,38 (<i>s</i>)	4,38 (<i>s</i>)
	2'''-5'''	3,00 – 3,75 (<i>m</i>)	3,00 – 3,75 (<i>m</i>)
	6'''	0,99 (<i>d</i> ; 6,2)	0,99 (<i>d</i> ; 6,2)

5.2.3.2.3 Substância 3

A substância 3 foi identificada como isoramnetina-3-O-neohesperosideo ou isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- β -D-glicopiranosídeo (**Figura 37**).

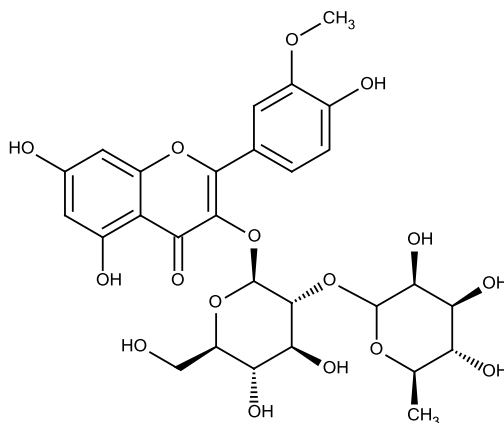


Figura 37- Estrutura química da isoramnetina-3-O-neohesperosideo (substância 3)

A isoramnetina-3-O-neohesperosideo foi obtida como um pó amarelo (5,3 mg), com bandas de absorção na região do ultravioleta em 255, 265 (ombro) e 354 nm, característicos de flavonol. O espectro de massas de alta resolução (microTOF Q II - ESI-TOF) no modo negativo apresentou o sinal relativo ao íon $[M-H]^-$ com valor de m/z de 623,1917, consistente com o valor de 623,1618 calculado para o íon desprotonado da fórmula $C_{28}H_{32}O_{16}$. O espectro de fragmentação do íon em m/z 623 mostrou a presença de um íon de m/z 315, referente à perda dos açúcares $[M-gli-ram-H]^-$ e indicando presença de metoxila ligada à aglicona. O espectro de NMR de 1H (**Tabela 21**) confirma a existência de um $O-CH_3$ ligado à aglicona, cuja posição foi atribuída ao C-3' (δ 3,86, s). Ainda para o anel B, o espectro evidencia um duplete com δ 8,08 ($J = 1,8$ Hz) e um duplete com δ 6,89 ($J = 8,6$ Hz) referentes à H-2' em *meta* e H-5' em *orto* em relação à H-6' (δ 7,49, *dd*, $J = 8,6$ e 1,8 Hz). Os dupletos em δ 6,44 ($J = 1,7$) e δ 6,19 ($J = 1,7$) foram atribuídos aos hidrogênios em *meta* H-8 e H-6 do anel A. A presença de sinal de CH_3 em δ 0,62 (*d*, $J = 6,2$ Hz) juntamente com um H-anomérico em δ 4,94 (s) indicam uma ramnose dentre os açúcares presentes. Outro H-anomérico em δ 5,70 (*d*, $J = 7,8$ Hz) foi atribuído à uma unidade de glicose. A constante de acoplamento de $J=7,8$ Hz observada no sinal do H anomérico indica se tratar de unidade de β -D-glicose. Os deslocamentos químicos de ambos os hidrogênios anoméricos foram confirmados por Tocsy1D. Pelo espectro de HMBC, o hidrogênio anomérico da β -D-glicose em δ 5,70 (H-1'') correlacionou com o sinal de carbono em δ

132,75 (C-3), enquanto que pelo espectro de HSQC, a correlação direta se deu em δ 100,96, sugerindo um flavonóide 3-diglicosilado. O tipo de ligação entre os dois açúcares, foi definida em 1 $\rightarrow$ 2, uma vez que o hidrogênio anomérico da ramnose em δ 4,94 (H-1''') correlacionou com o sinal de carbono em δ 75,48 (C-2''), enquanto que pelo espectro de HSQC, a correlação direta foi observada em δ 103,91.

É interessante pontuar que os rutinosídeos (ramnosil-(1 $\rightarrow$ 6)-glicosídeo) podem ser distinguidos dos neohesperosídeos (ramnosil-(1 $\rightarrow$ 2)-glicosídeo) em soluções de DMSO-d₆, pelo deslocamento químico do H-anomérico da ramnose. Por exemplo: os H-anoméricos de glicose e ramnose nos 3-rutinosídeos aparecem em 5,3 e 4,4 ppm, respectivamente, enquanto que nos 3-neohesperosídeos, eles são encontrados em 5,7 e 5,0 ppm, respectivamente (HARBORNE, 1996).

Tabela 21- Dados espectrométricos de NMR (¹H) obtidos para a substância 3 a 600 MHz em DMSO-d₆.

Estrutura	Posição	Substância 3		
		δ H	δ C	HMBC
Flavonol	2		156,43	
	3		132,75	
	4		177,37	
	5		161,33	
	6	6,19 (<i>d</i> ; 1,7)	98,98	
	7		165,07	
	8	6,44 (<i>d</i> ; 1,7)	93,89	
	9		155,87	
	10		105,94	
	1'		121,15	
	2'	8,08 (<i>d</i> ; 1,8)	113,53	
	3'		149,53	
	4'		147,69	
	5'	6,89 (<i>d</i> ; 8,6)	115,26	
	6'	7,49 (<i>dd</i> ; 1,8; 8,6)	121,74	
	3'-OCH ₃	3,86 (<i>s</i>) - pressat	56,09	C-3'
Açúcar 1	1''	5,70 (<i>d</i> ; 7,8)	100,96	C-3
	2''	3,06 – 3,43 (<i>m</i>)	75,48	
	3''		77,87	
	4''		68,47	
	5''		75,90	
	6''		60,45	
Açúcar 2	1'''	4,94 (<i>s</i>)	103,91	C-2''
	2'''	3,06 – 3,43 (<i>m</i>)	70,77	
	3'''		68,67	
	4'''		70,82	
	5'''		68,35	
	6'''	0,62 (<i>d</i> ; 6,2)	17,16	

5.2.3.2.4 Substância 4

A substância 4 foi identificada como narcisina (isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 6)- β -D-glucopiranosídeo) cuja estrutura química segue na **Figura 38**.

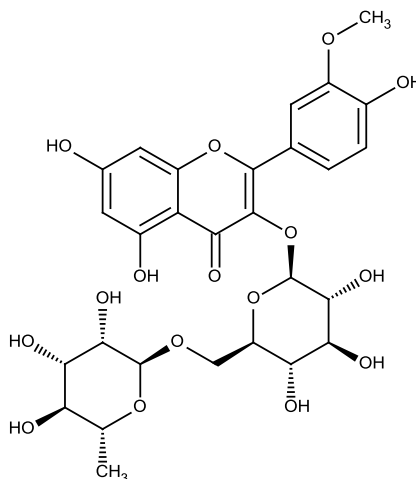


Figura 38- Estrutura química da narcisina (substância 4).

A narcisina, substância majoritária do extrato bruto liofilizando de *C. sylvestris* var. *lingua*, foi obtida como um pó amarelo (14,9 mg), com bandas de absorção na região do ultravioleta em 255, 265 e 354 nm, característicos de flavonol. O espectro de massas de alta resolução (microTOF Q II - ESI-TOF) no modo negativo apresentou o sinal relativo ao íon $[M-H]^-$ com valor de m/z de 623,1665, consistente com o valor de 623,1618 calculado para o íon desprotonado da fórmula $C_{28}H_{32}O_{16}$. O espectro de fragmentação do íon em m/z 623 mostrou a presença de um íon de m/z 315, referente à perda dos açúcares $[M-gli-ram-H]^-$ e indicando presença de metoxila ligada à aglicona.

O espectro de NMR de 1H (**Tabela 22**) confirma a existência de um $O-CH_3$ ligado à aglicona, cuja posição foi atribuída ao C-3' (δ 3,84, s). Ainda para o anel B, o espectro evidencia um duplete com δ 7,98 ($J = 1,8$ Hz) e um duplete com δ 6,90 ($J = 8,4$ Hz) referentes à H-2' em *meta* e H-5' em *orto* em relação à H-6' (δ 7,51, *dd*, $J = 8,4$ e 1,8 Hz). Os dupletos em δ 6,42 ($J = 1,3$) e δ 6,19 ($J = 1,8$) foram atribuídos aos hidrogênios em *meta* H-8 e H-6 do anel A. A presença de um CH_3 em δ 1,03 (*d*, $J = 6,2$ Hz) juntamente com um H-anomérico em δ 4,41 (*s*) sugerem uma ramnose dentre os açúcares presentes. Outro H-anomérico em δ 5,42 (*d*, $J = 7,8$ Hz) foi atribuído à uma unidade de β -D-glicose. Os deslocamentos químicos de ambos os hidrogênios anoméricos foram confirmados por Tocsy1D.

Pelo espectro de HMBC, o hidrogênio anomérico da β -D-glucose em δ 5,42 (H-1'') correlacionou com o sinal de carbono em δ 133,14 (C-3), enquanto que pelo espectro de HSQC, a correlação direta se deu em δ 101,91, sugerindo um flavonóide 3-diglicosilado.

Pelo espectro de HMBC, também foi possível observar o tipo de ligação entre os dois açúcares, que foi definida em 1 $\rightarrow$ 6, uma vez que o hidrogênio anomérico da ramnose em δ 4,41 (H-1''') correlacionou com o sinal de carbono em δ 65,27 (C-6''), enquanto que pelo espectro de HSQC, a correlação direta foi observada em δ 100,14.

Tabela 22- Dados espectrométricos de NMR (^1H) obtidos para a substância 4 a 600 MHz em DMSO- d_6 .

Estrutura	Posição	Substância 4		
		δH	δC (DMSO)	HMBC
Flavonol	2		156,56	
	3		133,14	
	4		177,38	
	5		161,26	
	6	6,19 (<i>d</i> ; 1,8)	98,96	
	7		164,77	
	8	6,42 (<i>d</i> ; 1,3)	93,94	
	9		156,43	
	10		103,90	
	1'		121,12	
	2'	7,98 (<i>d</i> ; 1,8)	113,48	
	3'		149,54	
	4'		147,07	
	5'	6,90 (<i>d</i> ; 8,4)	115,25	
	6'	7,51 (<i>dd</i> ; 1,8; 8,4)	122,05	
	3'-OCH ₃	3,84 (<i>s</i>)	56,01	C-3'
	1''	5,42 (<i>d</i> ; 7,8)	101,91	C-3
Açúcar 1	2''		71,97	
	3''		73,64	
	4''	3,07 – 3,38 (<i>m</i>)	68,38	
	5''		73,02	
	6''		65,27	
	1'''	4,41 (<i>s</i>)	100,14	C-6''
Açúcar 2	2'''		70,69	
	3'''		70,52	
	4'''	3,07 – 3,38 (<i>m</i>)	71,21	
	5'''		68,06	
	6'''- CH ₃	1,03 (<i>d</i> ; 6,2)	17,98	

5.2.3.2.5 Substância 5

A substância 5, também um flavonol glicosilado, foi identificada como isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- α -L-arabinopiranosideo, cuja estrutura química segue na **Figura 39**.

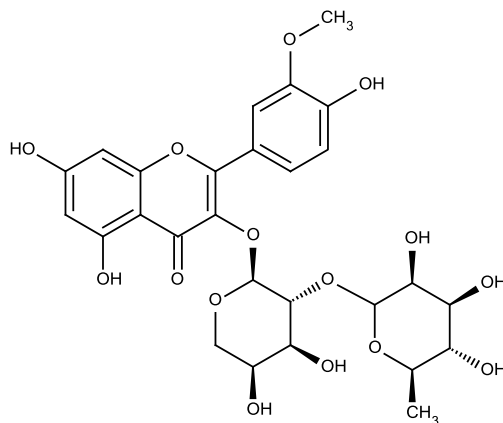


Figura 39- Estrutura química da isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- α -L-arabinopiranosideo (substância 5).

A isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- α -L-arabinopiranosideo foi obtida como um pó amarelo (5,1 mg), com bandas de absorção na região do ultravioleta em 255, 265 e 353 nm, característicos de flavonol. O espectro de massas de alta resolução (microTOF Q II - ESI-TOF) no modo negativo apresentou o sinal relativo ao íon $[M-H]^-$ com valor de m/z de 593,1553, consistente com o valor de 593,1512 calculado para o íon desprotonado correspondente à fórmula empírica $C_{27}H_{30}O_{15}$. O espectro de fragmentação do íon em m/z 593 mostrou a presença de um íon de m/z 314/315, referente à perda dos açúcares $[M-ara-ram-H]^-$ e indicando a possível presença de uma metoxila ligada à aglicona.

O espectro de NMR de 1H (**Tabela 23**) confirma a existência de um $O-CH_3$ ligado à aglicona, cuja posição foi atribuída ao C-3' (δ 3,88, s). Ainda para o anel B, o espectro evidencia um duplete com δ 7,90 ($J = 1,8$ Hz) e um duplete com δ 6,91 ($J = 8,4$ Hz) referentes à H-2' em meta e H-5' em orto em relação à H-6' (δ 7,57, dd , $J = 8,4$ e $1,5$ Hz). Os dupletos em δ 6,37 e δ 6,13 foram atribuídos aos hidrogênios em meta H-8 e H-6 do anel A. A presença de um CH_3 em δ 0,87 (d , $J = 6,2$ Hz) juntamente com um H-anomérico em δ 4,91 (s) sugerem uma raminose dentre os açúcares presentes. Outro H-anomérico em δ 5,50 (d , $J = 5,7$ Hz) foi atribuído à uma unidade de α -L-arabinose. Os deslocamentos químicos de ambos os hidrogênios anoméricos foram confirmados por

Tocsy1D. Pelo espectro de HMBC, o hidrogênio anomérico da α -L-arabinose em δ 5,50 (H-1'') correlacionou com o sinal de carbono em δ 133,09 (C-3), enquanto que pelo espectro de HSQC, a correlação direta se deu em δ 99,41.

Pelo espectro de HMBC, também foi possível observar o tipo de ligação entre os dois açúcares, que foi definida em 1 $\rightarrow$ 2, uma vez que o hidrogênio anomérico da ramnose em δ 4,91 (H-1''') correlacionou com o sinal de carbono em δ 71,38 (C-2''), enquanto que pelo espectro de HSQC, a correlação direta foi observada em δ 100,38.

Tabela 23- Dados espectrométricos de NMR (^1H) obtidos para a substância 5 a 600 MHz em DMSO- d_6 .

Estrutura	Posição	Substância 5		
		δH	δC	HMBC
Flavonol	2		156,35	
	3		133,09	
	4		177,05	
	5		161,01	
	6	6,13 (s)	94,05	
	7		166,49	
	8	6,37 (s)	93,15	
	9		155,40	
	10		109,27	
	1'		120,82	
	2'	7,90 (d; 1,8)	113,07	
	3'		150,75	
	4'		146,99	
	5'	6,91 (d; 8,4)	115,31	
	6'	7,51 (dd; 1,8; 8,4)	122,15	
	3'-OCH ₃	3,88 (s)	55,77	C-3'
Açúcar 1	1''	5,50 (d; 5,7)	99,41	C-3
	2''	3,20 – 3,90 (m)	71,38	
	3''		75,07	
	4''		70,7	
	5''		71,85	
Açúcar 2	1'''	4,91 (s)	100,38	C-2''
	2'''	3,20 – 3,90 (m)	68,49	
	3'''		78,64	
	4'''		70,62	
	5'''		66,96	
	6'''	0,87 (d; 6,2)	17,38	

A importância em se descrever tais compostos em *C. sylvestris* baseia-se na variabilidade química infraespecífica entre *C. sylvestris* var. *sylvestris* e *C. sylvestris* var. *lingua*. Neste trabalho demonstrou-se, do ponto de vista químico micromolecular, que os diterpenos clerodânicos prevalecem na variedade *sylvestris*, enquanto que os flavonoides predominam na variedade *lingua*.

Comparando-se os perfis cromatográficos destas duas variedades, os quais foram coletados no mesmo campo e, por conseguinte, sob as mesmas condições ambientais, é possível observar importantes diferenças na expressão do metabolismo secundário. Com efeito, os flavonoides glicosilados são notáveis na variedade *lingua* e apenas rutina foi encontrada em quantidades consideráveis nas duas variedades. Por outro lado, apesar de ter sido possível observar casearinas na variedade *lingua* (mesmo em nível de traço), tais substâncias são fortemente expressas apenas na variedade *sylvestris*. Num ponto de vista abrangente, a expressão diferencial destas duas classes de metabólitos secundários em ambas as variedades confirma a relação quimiotaxonômica entre eles.

Tais conclusões corroboram com o estudo de variabilidade genética realizado por Cavallari et al. (2010), no qual ele observou indivíduos de *C. sylvestris* ocorrendo simpatricamente e alopatricamente em diferentes ecossistemas brasileiros, o que sugere que as duas variedades desta espécie são unidades biológicas relativamente independentes em vez de uma única espécie respondendo a diferentes condições ambientais.

Conclui-se, portanto, que além dos diterpenos tipo clerodânicos já bastante conhecidos na espécie em estudo, os flavonoides glicosilados podem também desempenhar um importante papel de marcador quimiotaxonômico para se distinguir as duas variedades de *C. sylvestris*.

5.3 PARTE 3 - Estudos de dinâmica metabólica

5.3.1 Coletas, processamento e caracterização das amostras

Após as coletas, as folhas das duas variedades foram desidratadas e determinou-se o teor de água individualmente. Para a variedade *sylvestris* o teor de água variou entre 47,5 e 60,0% m/m (com desvios padrão relativos mínimo e máximo de 2,4 e 12,6, respectivamente), enquanto que para a variedade *lingua* de 47,5 a 60,7 % m/m (com desvios padrão relativos mínimo e máximo de 2,5 e 13,7 %, respectivamente). Analisando-se a média mensal dos resultados obtidos (**Figura 41**), verifica-se que a variedade *sylvestris*, quando comparada com a variedade *lingua*, apresenta um teor de água maior durante todos os meses do ano, exceto nos meses de janeiro, setembro e outubro.

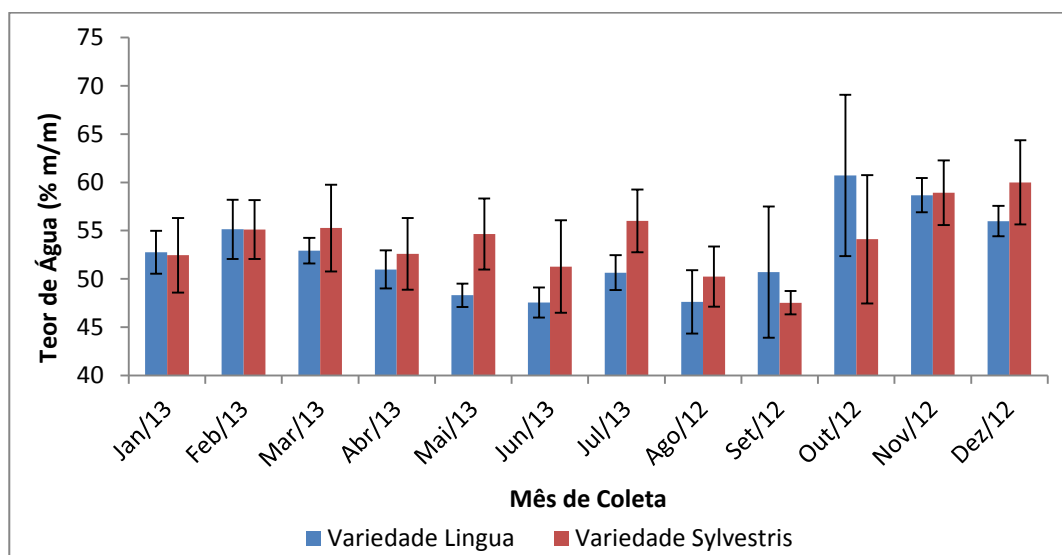


Figura 41- Teor de água nas folhas de *C. sylvestris*

Para efeito de documentação e monitoramento, os dados de temperatura e umidade relativa do ar também foram compilados durante as 48h de coleta de cada mês (**Figura 42**). Documentar tais dados é importante visto que mudanças climáticas globais podem alterar significativamente a fenologia vegetal, uma vez que variações na temperatura, tanto sozinhas quanto em conjunto com outros fatores, tais como o fotoperíodo, influenciam a dinâmica de desenvolvimento do vegetal. No entanto, sabe-se que em ecossistemas tropicais, a fenologia pode ser menos sensível à temperatura e fotoperíodo, e mais influenciada por mudanças sazonais de precipitação. A expectativa é que tais mudanças ocorram em conjunto com o aumento das temperaturas globais, mas tanto a direção quanto a magnitude de tais eventos podem variar de região para região (CLELAND et al., 2007).

Os registros obtidos neste trabalho mostram que a umidade do ar apresenta relação inversa à temperatura, conforme esperado. Para todos os meses, às 15h00 ocorre os picos de temperatura máxima com menores valores de umidade relativa. Comparando os resultados dessas análises com o teor de água nas folhas verifica-se que nos meses de menores teores de umidade relativa do ar e maiores temperaturas (Outubro a Fevereiro), os teores de água foliar nas duas variedades são maiores.

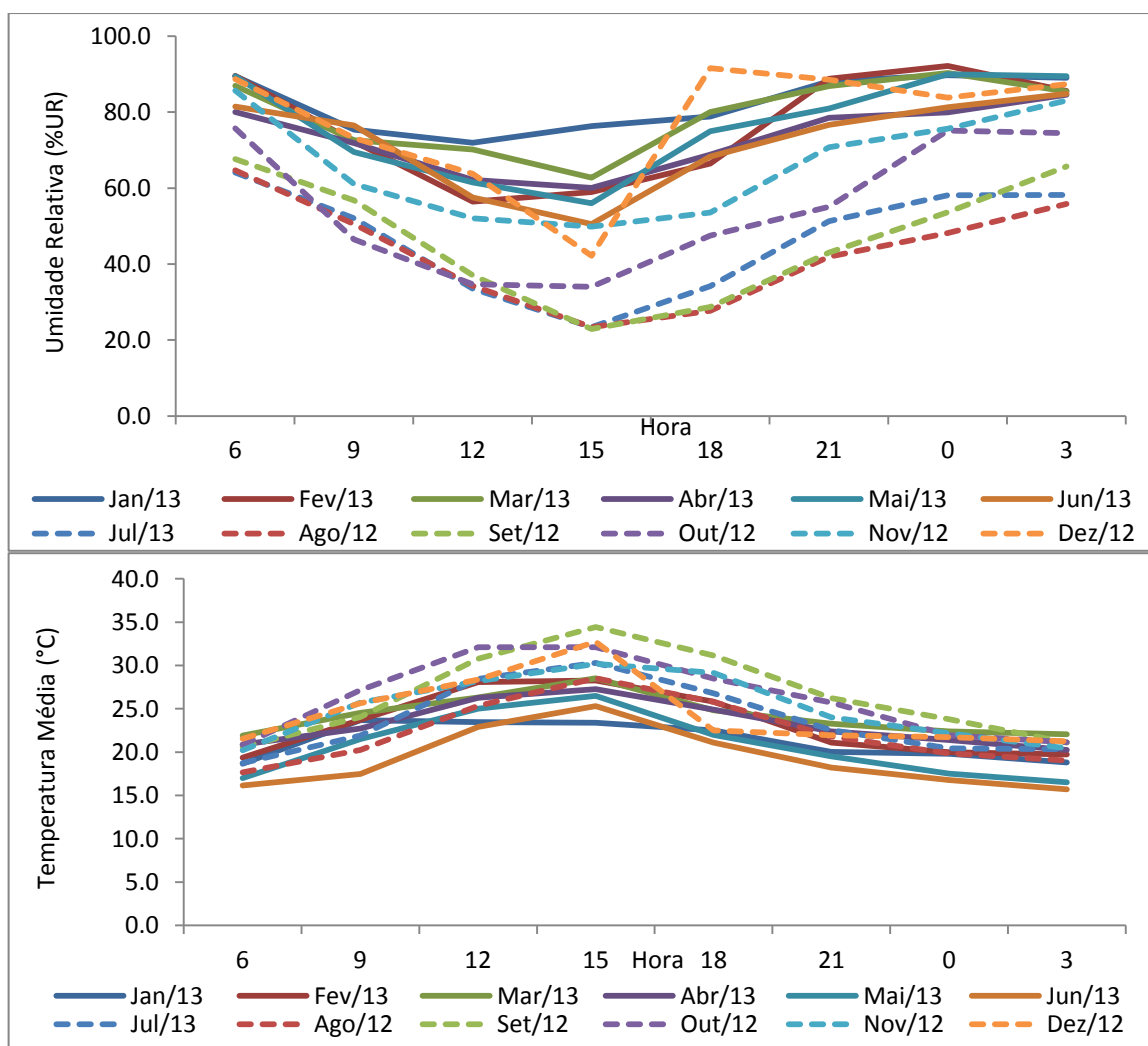


Figura 42- Umidade relativa do ar (acima) e temperatura média (abaixo) durante o ano, no período de 24 horas.

5.3.2 Avaliação do teor de taninos por extração em fase sólida com poliamida

Para que se pudesse conhecer melhor os extratos brutos das variedades *lingua* e *sylvestris*, fez-se o estudo da porcentagem do material retido em poliamida, da mesma forma que para a fração rica em compostos fenólicos (FRCF) oriunda do extrato bruto da variedade *lingua*, conforme descrito na seção 4.2.3. Verifica-se que tanto o extrato bruto liofilizado da variedade *lingua* quanto a fração enriquecida em compostos fenólicos

(FRCF) deixam cerca de 30% de material retido na poliamida. Ao contrário, o extrato bruto liofilizado da variedade *sylvestris* deixa retido pouco menos que 1%, mostrando uma grande diferença no teor de compostos fenólicos (incluindo os flavonoides glicosilados) entre essas variedades (**Tabela 24**).

Tabela 24- Cálculos de recuperação da fração dos extratos brutos das duas variedades de *C. sylvestris* (EBLL e EBLS) e da fração rica em compostos fenólicos (FRCF) retidas em poliamida.

Amostra*	Massa amostra de partida (mg)	Massa fração obtida (mg)	% m/m	Recuperação (%)
L - água		4,4	33,1	
L – água/MeOH 1:1	13,3	2,4	18,0	69,2
L - MeOH/NH ₄		2,4	18,0	
S - água		10,4	71,7	
S - água/MeOH 1:1	14,5	2,4	16,6	99,3
S - MeOH/NH ₄		1,6	11,0	
FRCF – água		7,6	40,6	
FRCF - água/MeOH 1:1	18,7	2,8	15,0	72,7
FRCF - MeOH/NH ₄		3,2	17,1	

* L = variedade *lingua*; S = variedade *sylvestris*; FRCF = fração rica em compostos fenólicos; MeOH/NH₄ = metanol adicionado de 5% de hidróxido de amônio.

5.3.3 Análise das amostras e avaliação da conformidade das análises

5.3.3.1 Estudo de dinâmica metabólica: ciclos circadiano e sazonal

Muitas das variações qualitativas e quantitativas que ocorrem na produção dos metabólitos secundários pelas plantas, ou seja, variações químicas fenotípicas, relacionam-se com o espaço (como resultado de adaptações ao meio ambiente), com variações genéticas e com o tempo, a exemplo dos estágios ontogênicos e fenológicos bem como com os ciclos circadianos e sazonais (MOORE et al., 2013). Neste contexto, para que se pudesse caracterizar e entender a dinâmica da produção dos metabólitos secundários nas duas variedades de *C. sylvestris*, optou-se pela realização de coletas mensais das folhas das duas variedades durante um ano, sendo que tal amostragem estendeu-se por dois dias consecutivos com tomadas de amostras a cada três horas. Uma vez processados, os cromatogramas resultantes das análises em UHPLC-DAD foram registrados em dois comprimentos de onda, 254 nm e 235 nm, a fim de monitorar os compostos fenólicos e os diterpenos, respectivamente. A adição do padrão interno foi essencial na prevenção e correção de problemas oriundos de variações na preparação e injeção das amostras, além de permitir a normalização dos dados. A **Figura 43**

evidencia os gráficos obtidos para a narcisina e para a casearina D. Para as demais substâncias, os gráficos de radar seguem dispostos nos Apêndices XIII e XIV.

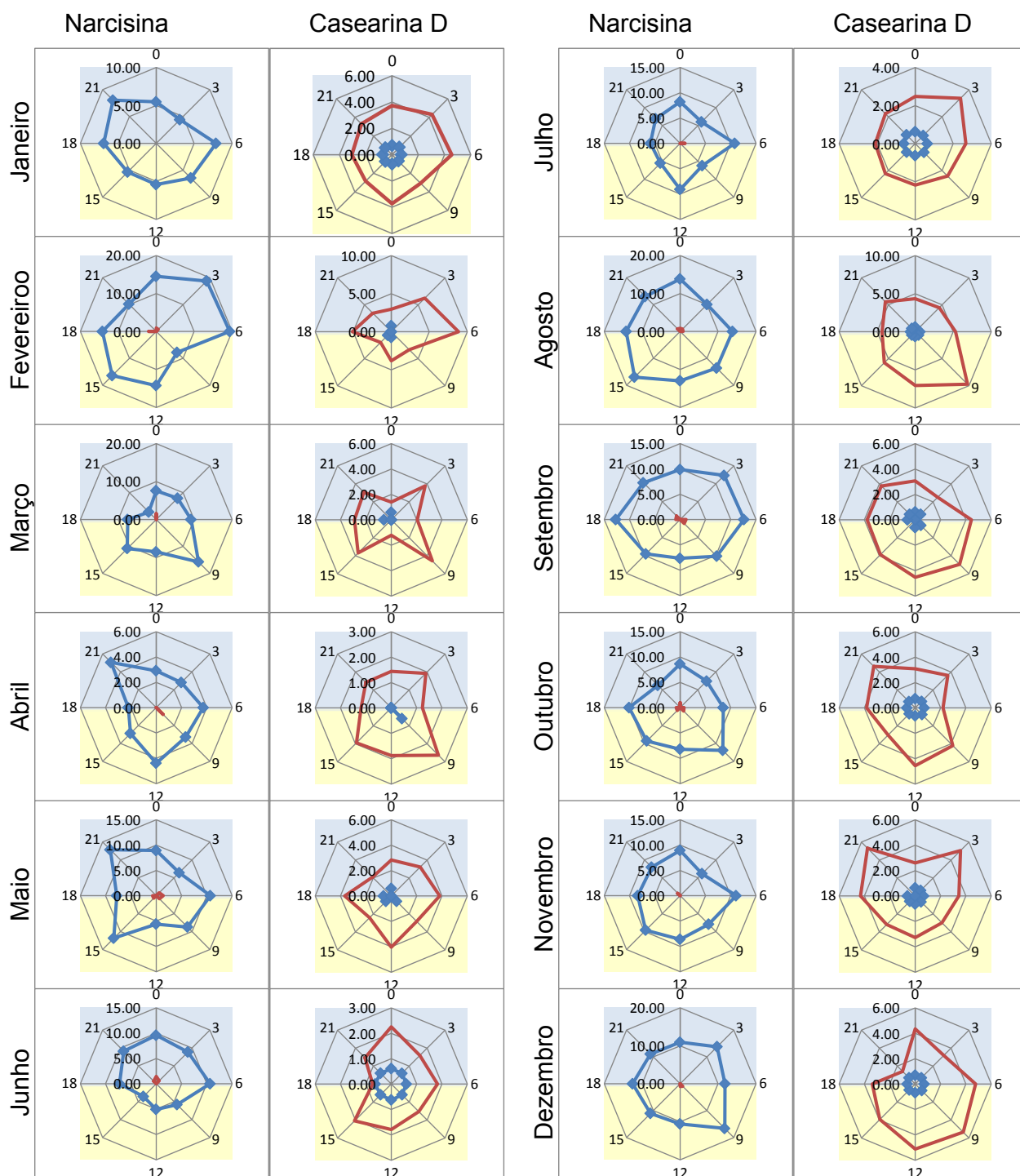


Figura 43- Gráficos de radar exemplificando o ciclo circadiano dos flavonoides e diterpenos nas duas variedades de *C. sylvestris*. Em azul: var. *lingua*; Em vermelho: var. *sylvestris*. A escala circular indica hora do dia e o eixo vertical principal indica o teor em mg/g. A área sombreada em azul corresponde ao período noturno e a área sombreada em amarelo corresponde ao período diurno.

De forma constante, em todas as horas do dia e em todos os dias do ano, os resultados obtidos indicam que na variedade *sylvestris* os diterpenos clerodânicos, representados pelas casearinas, estão com teor maior que os flavonoides, conforme já era esperado. Situação inversa ocorre com a variedade *lingua*, onde os flavonoides prevalecem sobre os diterpenos. Tanto para os três flavonoides quanto para os cinco diterpenos selecionados não foram observados picos de produção em qualquer horário amostrado, indicando teores relativamente constantes durante os dias. Já a análise da variação sazonal foi feita através de gráficos do tipo box plot, que permite contemplar todos os valores encontrados nas 16 pontos amostrados em cada mês. A **Figura 44** a seguir evidencia os resultados.

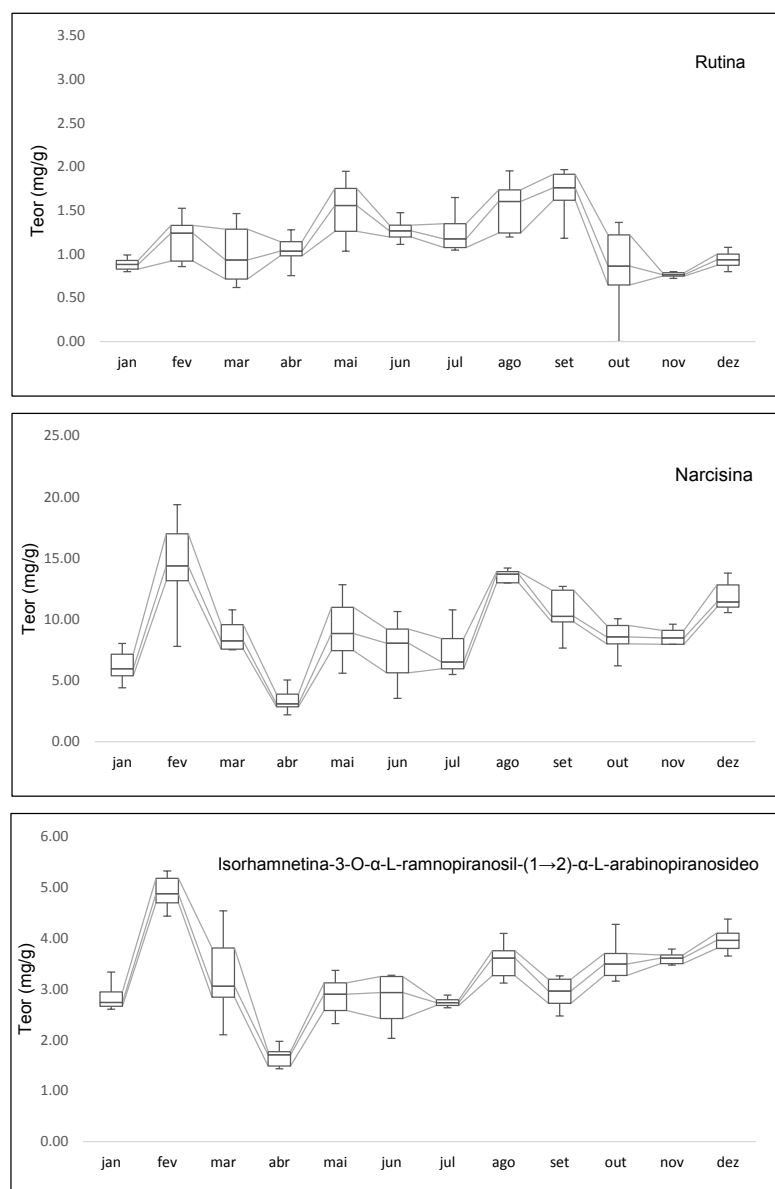


Figura 44-Gráficos box plot obtidos para a análise do ciclo sazonal dos flavonoides glicosilados na variedade *lingua*.

Uma vez que cada classe de metabólitos é diferencialmente expressa em cada variedade, a análise sazonal foi feita em função dos flavonoides para a variedade *lingua* e dos diterpenos para a variedade *sylvestris*.

Iniciando-se pela variedade *lingua*, é possível observar, claramente, o mesmo padrão de produção para os três flavonoides em estudo. É interessante notar que há um decréscimo acentuado no teor dos três flavonoides no mês de Abril, que volta a aumentar no mês de Agosto (**Figura 44**). Esta observação é pertinente visto que as observações em campo permitiram verificar que o período de floração se inicia em maio, com pico em junho, sendo que até meados de agosto ainda era possível observar flores coexistindo com frutos. Ou seja, os menores teores de flavonoides foram verificados exatamente durante a fase reprodutiva da planta.

Para a variedade *sylvestris*, ocorre o mesmo padrão na produção dos flavonoides glicosilados durante o ano, mais precisamente e unicamente para a Rutina, sendo que no mês de Abril esta substância esteve presente apenas em nível de traços (**Figura 45**).

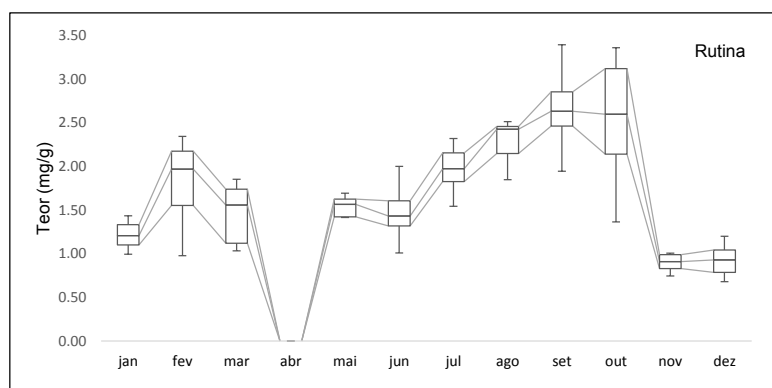


Figura 45- Gráfico box plot obtido para a análise do ciclo sazonal da rutina na variedade *sylvestris*.

Finalmente, focando na análise do ciclo sazonal dos diterpenos clerodânicos produzidos pela variedade *sylvestris* é possível observar a mesma dinâmica observada nos flavonoides: ocorre um decréscimo no teor das cinco casearinas avaliadas nos meses de floração, principalmente no mês de Abril, sendo que o teor máximo só é retomado no mês de Agosto, ou seja, no fim do período reprodutivo (**Figura 46**).

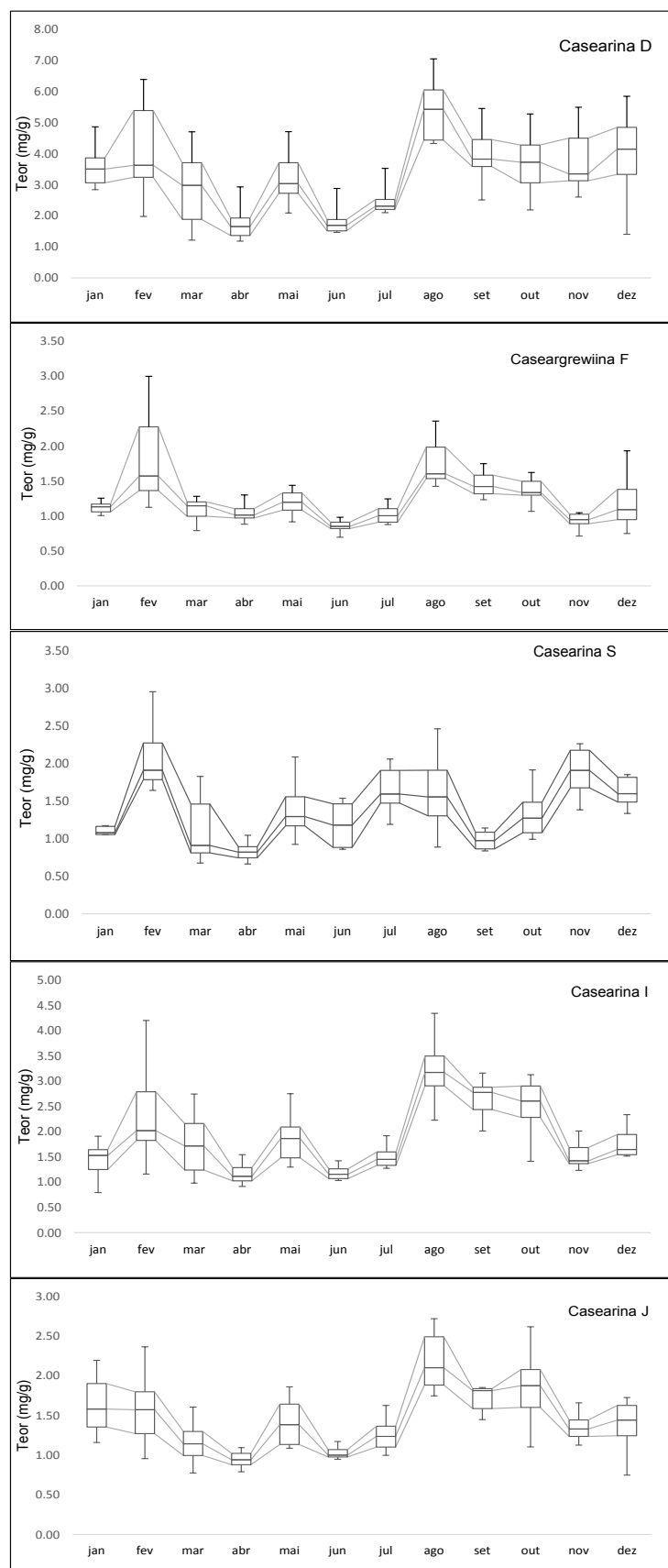


Figura 46- Gráfico box plot obtido para a análise do ciclo sazonal das casearinas na variedade *sylvestris*.

De fato, eventos fenológicos, ou seja, os eventos periódicos do ciclo de vida de animais e plantas que são influenciados pelo meio ambiente (especialmente variações sazonais de temperatura e precipitação) são aspectos dominantes na ecologia de plantas, tanto em escala individual quanto para ecossistemas inteiros. O momento da mudança entre a fase vegetativa e a reprodutiva que ocorre de forma concertada com o período de floração é crucial para o desenvolvimento das sementes, tanto para os indivíduos quanto para as populações. Variações fenológicas entre espécies é um mecanismo importante para a manutenção da coexistência das espécies em diversas comunidades vegetais, reduzindo a concorrência por polinizadores e outros recursos (CLELAND et al., 2007).

5.3.3.2 Avaliação da conformidade das análises

O desenvolvimento do método analítico por UHPLC-DAD priorizou, em cada etapa, minimizar quaisquer erros que pudessem interferir nos resultados finais, uma vez que um grande número de amostras foi analisado para posterior comparação. Para tanto foram tomados alguns cuidados: a utilização da padronização interna, do estabelecimento de um protocolo de extração das amostras com poucas etapas, da utilização de uma fase móvel simples composta apenas por acetonitrila e água (sem a adição de ácido, o que permitiu um rápido acondicionamento da coluna) e finalmente a manutenção de um controle rigoroso da qualidade das injeções e resolução da coluna cromatográfica (através da injeção diária da solução do padrão interno 0,5 mg/mL e de uma mistura de benzeno e tolueno). Adicionalmente, para certificação da qualidade e reprodutibilidade das análises, 6 amostras da variedade *lingua* e 6 amostras da variedade *sylvestris*, uma de cada mês foram reinjetadas e a conformidade com os resultados anteriores foi avaliada. Observa-se que para todos os casos, existe a mesma tendência dos resultados anteriores. As **Figuras 47 e 48** a seguir mostram os resultados obtidos.

Para o grupo dos diterpenos, observa-se uma tendência bastante uniforme entre os dois momentos de análise. Nos meses de fevereiro e agosto (ambos correspondentes à var. *sylvestris*) ocorrem as maiores diferenças. Ainda assim os valores de desvio padrão relativo entre os dois dias de análise foi de 34,8% e 33,6%, respectivamente. Já para o grupo dos compostos fenólicos, esta variação foi maior, principalmente nos meses de março e setembro (ambos correspondentes à var. *lingua*), com valores de desvio padrão relativo entre os dois dias de análise de 40,3%

e 29,7%, respectivamente. . Estas porcentagens foram consideradas adequadas visto que nesta etapa desejava-se observar a tendência dos resultados; para as análises quantitativas, os dados foram normalizados e os teores calculados em função de curvas analíticas.

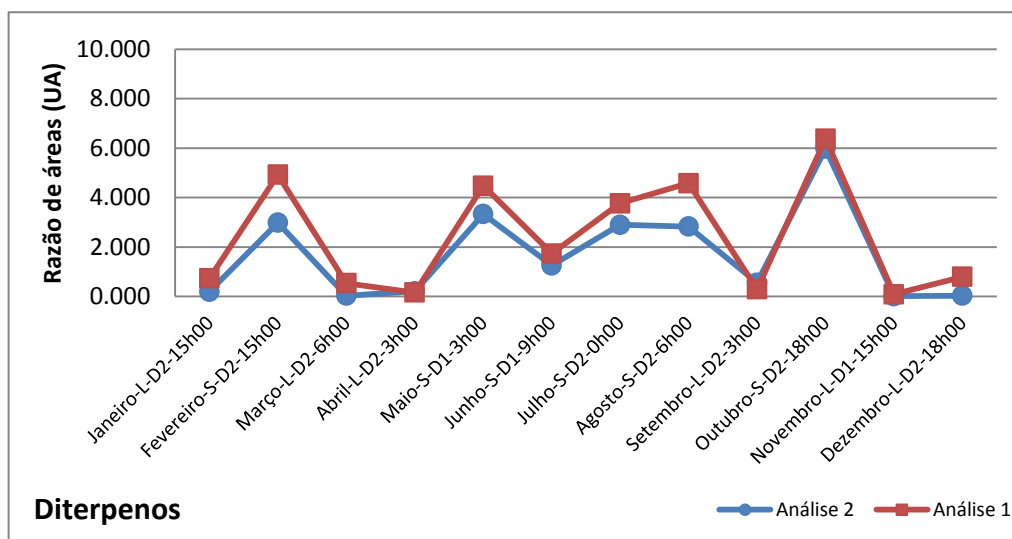


Figura 47- Avaliação da conformidade das injeções comparando-se as 12 amostras reinjetadas (análise 2) com as injeções iniciais (análise 1) focando-se na razão de áreas dos diterpenos pelo padrão interno. OBS: as letras L e S que aparecem logo em seguida do mês correspondem às variedades *lingua* e *sylvestris*, respectivamente.

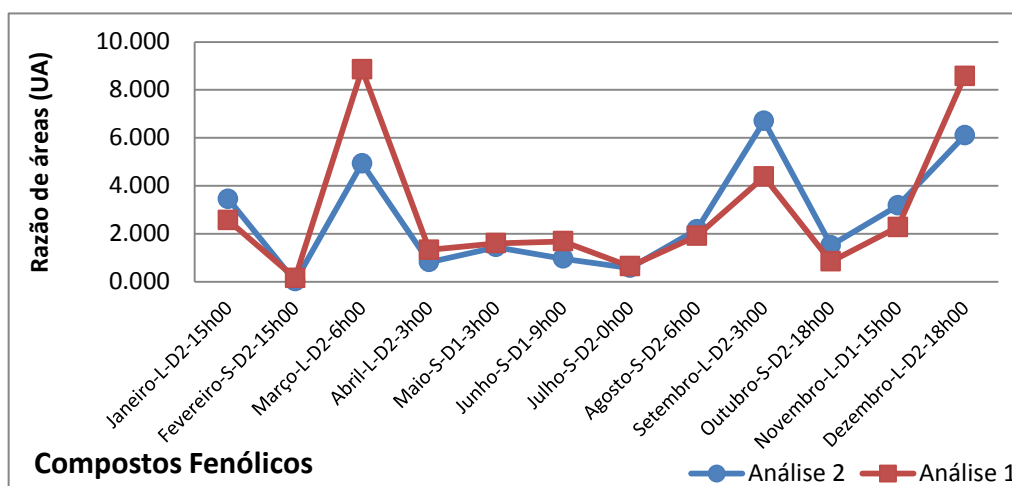


Figura 48- Avaliação da conformidade das injeções comparando-se as 12 amostras reinjetadas (análise 2) com as injeções iniciais (análise 1) focando-se na razão de áreas dos compostos fenólicos pelo padrão interno. OBS: as letras L e S correspondem às variedades *lingua* e *sylvestris*, respectivamente.

5.4 PARTE 4 - Estudos de variabilidade química infraespecífica

Biodiversidade, ou diversidade biológica, significa o alcance total de vida na Terra, incluindo todas as diferentes espécies em conjunto com a variação genética entre populações e indivíduos e a variedade de ecossistemas, comunidades e habitats. Sua determinação envolve tanto o número de espécies (riqueza) quanto os padrões de alocação do número ou de biomassa entre as diferentes espécies, no sentido de uniformidade (COX; MOORE, 2009). Neste contexto, a Biodiversidade pode ser considerada em três níveis: a diversidade intraespecífica (dentro da mesma espécie, portanto diversidade genética), entre espécies (organismal) e entre comunidades (diversidade ecológica) (RAMBALDI et al., 2003). No presente trabalho objetivou-se o estudo da biodiversidade no nível infraespecífico visando entender a variabilidade química apresentada pela espécie *C. sylvestris*.

Portanto, para o estudo de variabilidade química infraespecífica, foram coletadas amostras de *C. sylvestris* abrangendo os biomas Mata Atlântica (clima subquente e úmido ou superúmido), Cerrado (clima quente e semi-úmido, com estação seca e chuvosa), Caatinga (semi-árido, quente e seco), Pantanal (clima quente com planícies alagadas), Pampa (clima mesotérmico brando e úmido) e ecótonos, visando contemplar as duas variedades em estudo. A **Figura 49** a seguir ilustra as populações amostradas.

Todas as amostras foram preparadas e analisadas de acordo com o método analítico definido para este trabalho fornecendo os dados cromatográficos necessários para a análise quimiométrica. Cada cromatograma foi normalizado em função do padrão interno objetivando eliminar inconsistências no preparo das amostras e no procedimento de injeção. Adicionalmente, foi necessária uma redução do número de variáveis originais (sem perda de resolução) para que os cálculos computacionais pudessem ser viabilizados. Portanto, as 27.004 variáveis obtidas em cada cromatograma foram reduzidas para 6.751, resultando numa matriz de dados de 134×6751. Finalmente, os dados foram centrados na média, normalizados e submetidos à análise de componentes principais (PCA). No Apêndice XV encontram-se os cromatogramas organizados por local de coleta.

Tendo em vista a ocorrência das duas variedades em estudo em cada bioma, a análise dos dados obtidos a partir de todas as amostras coletadas no território brasileiro revelou um complexo padrão de distribuição. No entanto, para uma análise detalhada tais biomas foram subdivididos em domínios de vegetação. Para que esta análise pudesse ser viabilizada, em função destes domínios, resolveu-se então utilizar

primeiramente apenas amostras do Estado de São Paulo, as quais correspondem a aproximadamente 50% das amostras coletadas.

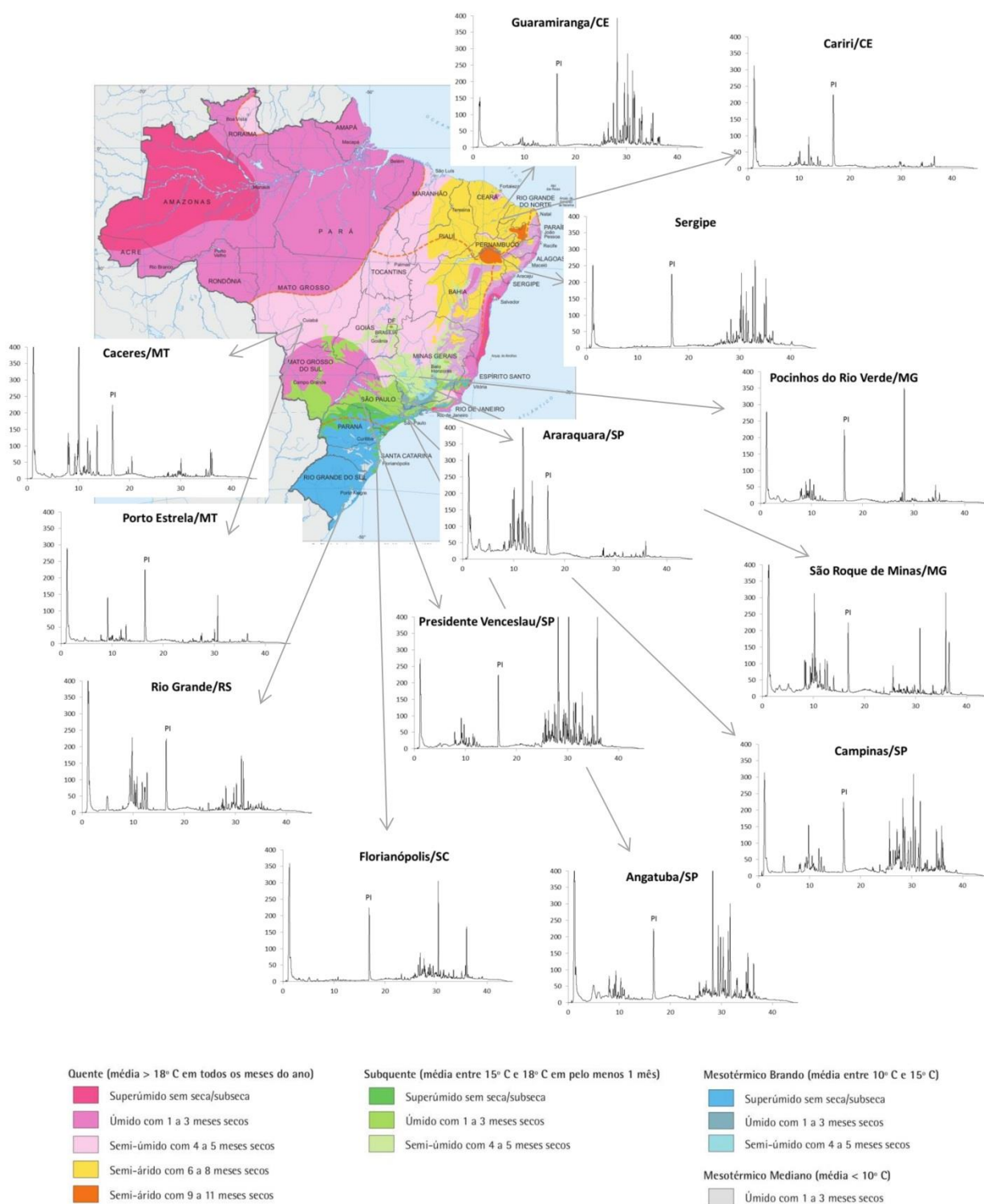


Figura 49- Cromatogramas representativos das coletas de variabilidade infraespecífica. Mapa e legenda adaptados do IBGE (<http://atlasescolar.ibge.gov.br>).

Sendo assim, a análise quimiométrica das amostras coletadas no Estado de São Paulo contemplou os dados cromatográficos de 54 indivíduos oriundos de seis populações, contemplando resquícios de Mata Atlântica (Floresta Estacional Semi-Decidual), Cerrado e ecótonos.

A **Figura 50** a seguir evidencia os resultados encontrados, sendo que nos gráficos de *scores* (Fig. **50A** e **Fig. B**, respectivamente) as amostras estão classificadas de acordo com as populações de origem e a variedade assinalada tendo em vista as características morfológicas. Embora esta classificação possa ser considerada subjetiva, é possível observar uma clara distinção entre as variedades *lingua* e *sylvestris*, enquanto que as amostras com morfologia intermediária sobrepõem a ambos os grupos. O gráfico de *loadings* (**Figura 50C**) permite a distinção das principais variáveis responsáveis pelo agrupamento observado no gráfico de *scores*. Essencialmente, a ocorrência de compostos fenólicos pode ser correlacionada com a var. *lingua* enquanto que a ocorrência de diterpenos clerodânicos pode ser correlacionada com var. *sylvestris*, como esperado. Além disso, é possível observar diferenças na composição de cada grupo de metabólitos.

Uma vez que as amostras são melhor separadas pela PC1, dentre as variáveis que mais influenciam no agrupamento observado estão o flavonóide glicosilado narcisina (ou isoramnetina-3-O-rutinosídeo, em aproximadamente 10,5 min, na parte negativa da PC1) e os diterpenos clerodânicos casearina D (aproximadamente 28 min), casearina S (em aproximadamente 30,5 min) e casearina I (32 min), na porção positiva da PC1. No entanto, tais casearinas são também responsáveis pela separação observada pela PC2, com a adição da isoramnetina-3-O-neohesperidosídeo (em 9,5 min), na parte positiva da PC2. Ainda, na parte negativa da PC2, é interessante notar a presença de um pico em aproximadamente 36 min, o qual ainda não foi identificado, porém, seu espectro no UV-Vis evidencia as bandas de absorção em 265 e 435 nm.

Uma análise cuidadosa por populações destaca a ocorrência de indivíduos ricos em diterpenos clerodânicos não só nas populações de Presidente Venceslau/SP e Angatuba/SP, as quais situam-se em áreas de Floresta Estacional Semi-Decidual, mas também alguns indivíduos com este perfil podem ser encontrados nas populações de Mogi-Guaçu/SP (ecótono) e Campinas/SP (mata ciliar em área de Cerrado). Embora Campinas/SP pertença a uma área de Floresta Estacional Semi-Decidual as amostras foram coletadas em uma região de mata ciliar localizada em área de Cerrado, onde a

vegetação é mais úmida e densa, o que pode explicar a ocorrência de indivíduos com morfologias típicas de *var. sylvestris*, *var. lingua* e intermediária.

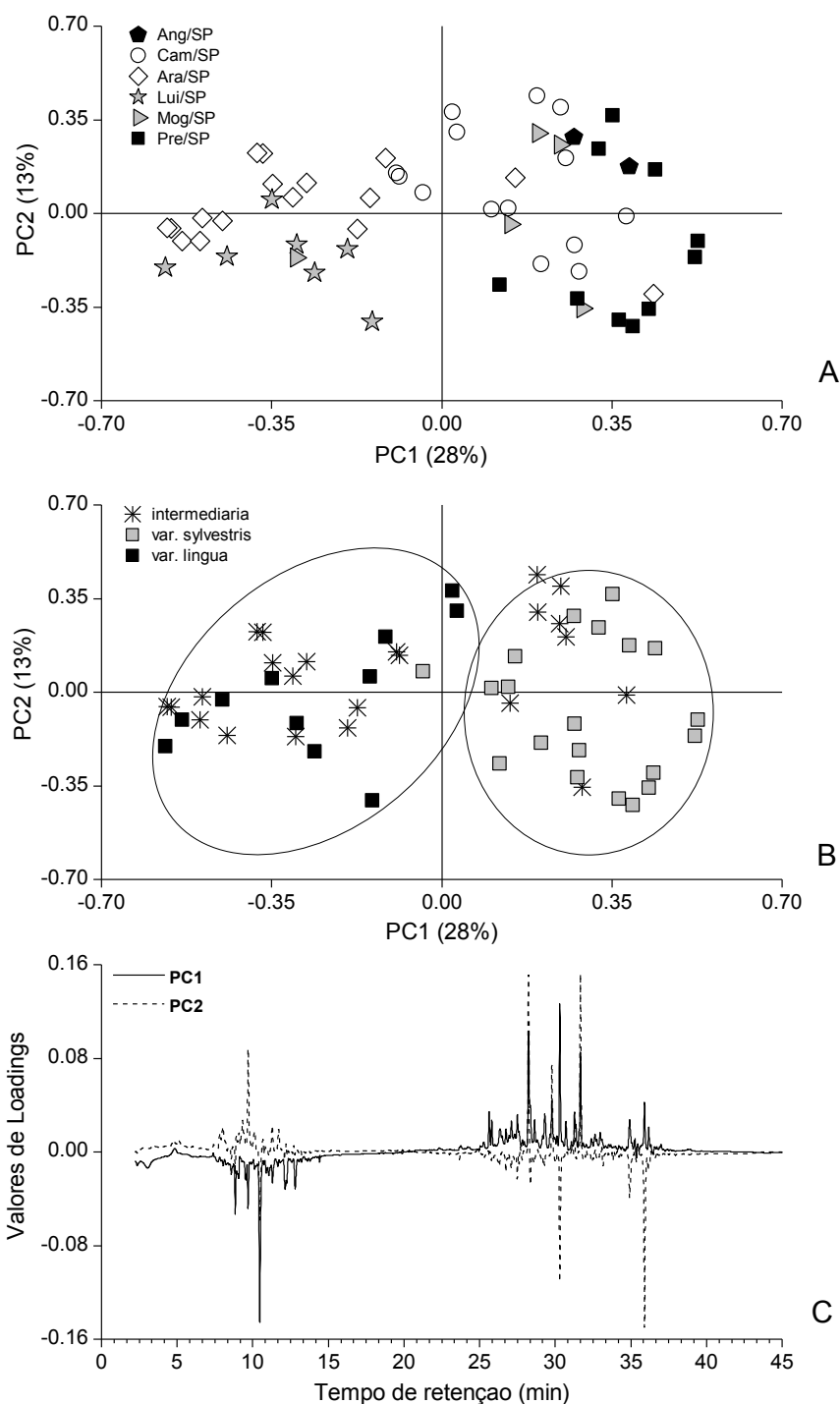


Figura 50- (A) Gráfico de scores (PC1 x PC2) das amostras de *C. sylvestris* provenientes do Estado de São Paulo, mostrando o agrupamento de acordo com as populações de origem. (B) Gráfico de scores (PC1 x PC2) das amostras de *C. sylvestris* classificadas de acordo com as variedades (intermediária, *lingua* e *sylvestris*). (C) Gráfico de *loadings* (PC1 x PC2) para discriminação de variáveis.

Por outro lado, a grande parte das amostras ricas em compostos fenólicos foi encontrada nas populações de Luis Antônio/SP (ecótono) e Araraquara/SP (área típica de Cerrado). No entanto, no trabalho de campo em Araraquara/SP, foi possível também observar e coletar alguns indivíduos com morfologia típica de var. *sylvestris*, embora a maioria dos indivíduos amostrados se caracterizavam como var. *lingua* ou intermediária. Finalmente, em ecótonos (Mogi-Guaçu/SP e Luis Antônio/SP) só foi possível encontrar indivíduos classificados como var. *lingua* e intermediária. Embora não tenham sido encontrados indivíduos da var. *sylvestris*, algumas amostras provenientes dessas zonas de ecótonos, evidenciaram a produção de diterpenos clerodânicos.

Expandindo-se tal análise e contemplando, desta vez, todas as amostras coletadas no território brasileiro, através do gráfico de scores (**Figura 51A**) é possível observar que amostras coletadas em biomas mais restritos e homogêneos sob o ponto de vista edáfico e de fauna e flora como o Pampa e o Pantanal apresentam-se melhor agrupados na PCA, enquanto que as amostras oriundas do Cerrado, por ser considerado um bioma contínuo e portanto sujeito à maior fluxo gênico (maior variabilidade), apresentam-se mais espalhadas. Quanto às amostras provenientes de zonas de transição, ou ecótonos, incluindo o carrasco, verifica-se que tanto do ponto de vista químico, quanto morfológico, o grau de variabilidade é maior, sendo que pouco ou nenhum agrupamento pode ser visualizado.

Com relação ao padrão de distribuição das amostras oriundas das áreas remanescentes de Mata Atlântica (bioma altamente fragmentado), os agrupamentos podem ser analisados sob o ponto de vista biogeográfico, visto que tais resultados vêm de encontro com padrões de fragmentação de habitats.

A fragmentação de um habitat original é o processo no qual um habitat contínuo é dividido em manchas, ou fragmentos mais ou menos isolados, sendo que a distância entre tais fragmentos e o isolamento entre estes, são responsáveis pelo grau de conectividade entre os fragmentos e o habitat contínuo (RAMBALDI et al., 2003). Habitats fragmentados têm o tamanho de suas populações reduzido e o isolamento espacial entre elas aumentado. Consequentemente, ocorre a erosão da variabilidade genética e o aumento da divergência interpopulacional devido à fatores tais como (i) o aumento da deriva genética, (ii) aumento das taxas de endogamia intrapopulacional, (iii) redução do fluxo gênico entre as populações (YOUNG et al., 1996). Populações de plantas e animais em fragmentos isolados têm menores taxas de migração e dispersão

e, em geral, com o tempo sofrem problemas de troca gênica e declínio populacional (RAMBALDI et al., 2003).

Cinco diferentes populações de áreas remanescentes de Mata Atlântica foram amostradas a citar, uma da área de Floresta Ombrófila Densa (FLO/SC) e quatro correspondentes à Floresta Estacional Semi-Decidual, uma do Estado de São Paulo (PRE/SP) e três do Nordeste (CAR/CE, PAC/CE e SER/SE). Analisando-se os resultados encontrados (**Figura 51A**), observa-se que as diferentes populações oriundas do bioma Mata Atlântica, se situam principalmente na parte positiva da PC1. No entanto, ao se analisar cada população separadamente, observa-se que elas apresentam-se agrupadas entre si, ou seja, cada qual caracterizando um habitat, separadamente.

Clara distinção pode ser feita quando se considera a classificação morfológica (**Figura 51B**): amostras classificadas como variedade *sylvestris* correlacionam-se majoritariamente com as áreas de Floresta Estacional Semi-Decidual e Floresta Ombrófila Densa, ou seja, áreas caracterizadas por um clima úmido e menor exposição à luz solar (na **Figura 51B**, elipses azul e verde). Amostras provenientes da área de mata ciliar (amostras coletadas em Campinas) situam-se na mesma região das amostras classificadas como variedade *sylvestris*. Já as amostras classificadas como variedade *lingua* correlacionam-se com áreas onde a vegetação é típica de Cerrado ou em áreas de ecótono contemplando a vegetação de cerrado, regiões caracterizadas por um clima quente e seco, e cuja vegetação está sujeita à forte exposição solar (na **Figura 51B**, elipse de cor preta). Finalmente, em ambientes de clima quente cuja vegetação se encontra sob forte exposição solar, porém em áreas sujeitas à alagamentos ou a um clima mais úmido, a exemplo do Pantanal e Pampa (elipse amarela) concentram-se amostras de características morfológicas intermediárias.

Da mesma forma que para a análise feita apenas com as amostras provenientes do Estado de São Paulo, o gráfico de loadings desta PCA (**Figura 51C**) evidencia a predominância dos compostos fenólicos na porção negativa da PC1 e na positiva da PC2, onde justamente agrupam-se a maioria das amostras do Cerrado, Pampa, Pantanal e Caatinga. Já os diterpenos, ocorrem principalmente na porção positiva da PC1. Dentre as variáveis mais significativas, desta vez estão os flavonoides glicosilados rutina e narcisina (em aproximadamente 8,5 e 10,5 min, respectivamente) e os diterpenos clerodânicos casearina D (28 min), casearina S em 30,5 min e casearina I (32 min).

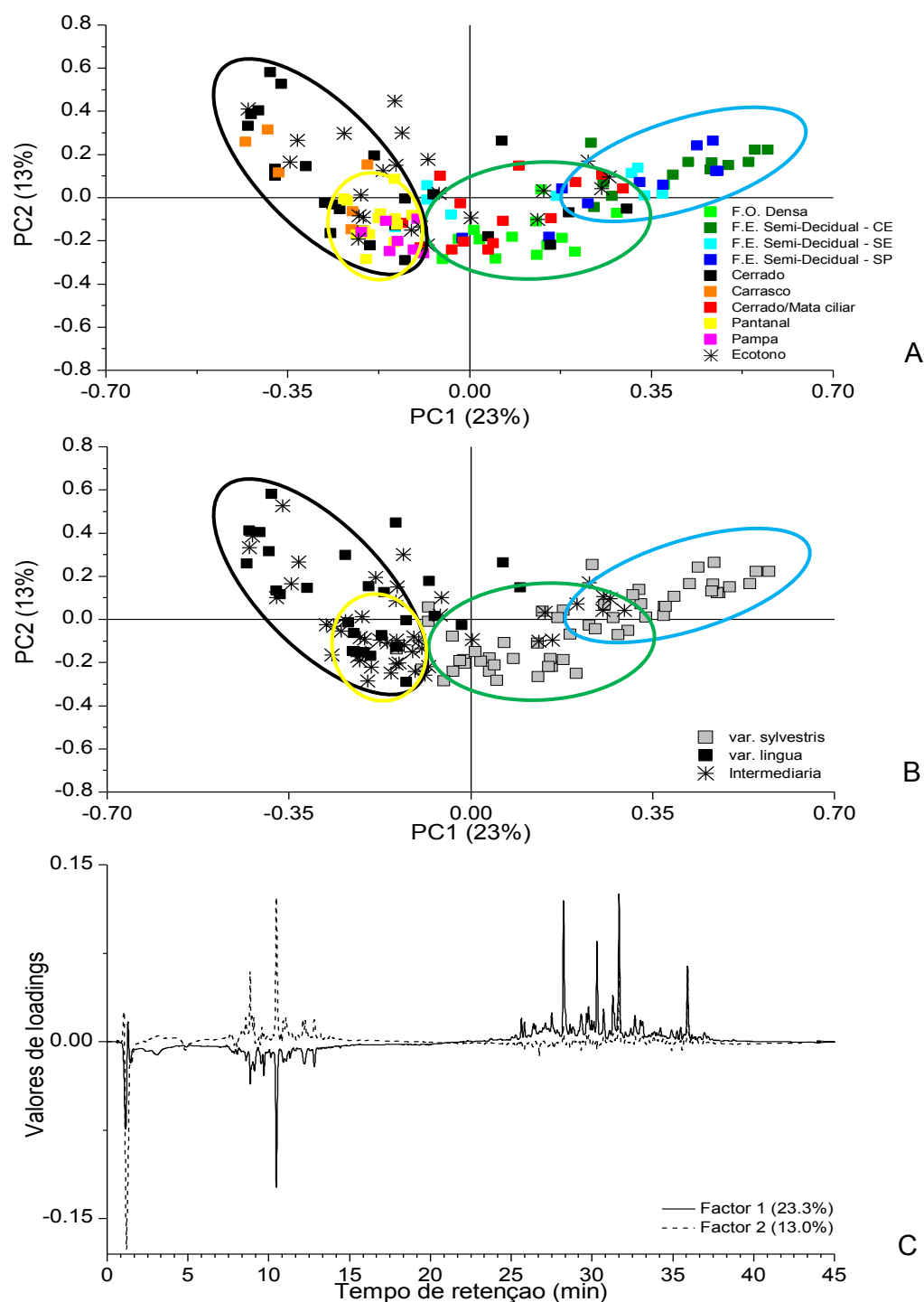


Figura 51- A) Gráfico de scores (PC1 x PC2) das amostras de *C. sylvestris* provenientes do território brasileiro, mostrando o agrupamento de acordo com as populações de origem. (B) Gráfico de scores (PC1 x PC2) das amostras de *C. sylvestris* classificadas de acordo com as variedades (intermediária, *lingua* e *sylvestris*). (C) Gráfico de *loadings* (PC1 x PC2) para discriminação de variáveis. As elipses concentram amostras de características comuns: azul (clima quente e úmido, vegetação menos exposta ao sol): Floresta Estacional Semi-Decidual; verde (clima quente e super úmido, vegetação menos exposta ao sol): Floresta Ombrófila Densa e Mata ciliar; preto (clima quente e seco, exposição excessiva ao sol): Cerrado e Carrasco; amarelo (clima seco porém com áreas alagadas e vegetação excessiva ao sol): Pampa e Pantanal.

Tal produção diferenciada de compostos fenólicos e diterpenos clerodânicos em *C. sylvestris* e sua correspondência com as condições ambientais de origem podem ser analisadas à luz do que já se sabe acerca da diferenciação dos metabólitos secundários de origem vegetal. Tanto o tipo quanto a quantidade de metabólitos secundários diferem significativamente de acordo com os ambientes físicos e bióticos, como resultado da adaptação local ou triagem genotípica ou seleção através habitats. Além disso, grande parte desta variação é atribuível a plasticidade fenotípica dos genótipos em resposta à variações dos recursos disponíveis e necessários para o crescimento, tais como nutrientes, luz e água e a presença ou ausência de inimigos, concorrentes ou mutualistas (MOORE et al., 2013).

5.5 PARTE 5- Estudo proteômico diferencial de *C. sylvestris*

5.5.1 Coleta do material vegetal, obtenção e avaliação dos extratos proteicos

Na análise proteômica a extração das proteínas e a qualidade das amostras é um fator crítico tanto para as separações em gel quanto nos procedimentos livres de gel (AGRAWAL et al., 2005). Considerando as propriedades bioquímicas das proteínas tais como carga, tamanho e hidrofobicidade, bem como características intrínsecas das diferentes espécies, estágio de desenvolvimento, tipo de célula e tecido e condições de cultivo, diversos protocolos de extração têm sido desenvolvidos para o estudo proteômico (SARASWATHY; RAMALINGAM, 2011; SHEORAN et al., 2009). Os mais comuns são os métodos que utilizam homogeneização do tecido vegetal com tampões específicos, e também aqueles que utilizam solventes orgânicos tais como ácido tricloroacético-acetona e fenol, fornecendo extratos proteicos com bom rendimento e com a menor quantidade de interferentes possível (JORRIN-NOVO et al., 2009; SHEORAN et al., 2009).

No presente trabalho, a extração das proteínas foi feita de acordo com o método otimizado por WANG et al. (2006) visando eliminar interferentes tais como terpenóides, lipídeos, polifenóis e pigmentos, que dificultam as separações eletroforéticas e cromatográficas de extratos proteicos (SHEORAN et al., 2009; WANG et al., 2003). O procedimento combina uma sequência de breves lavagens com ácido tricloroacético, metanol básico e acetona visando remover estes contaminantes sem expor a amostra por muito tempo a pHs baixos, ao mesmo tempo em que promove a precipitação das proteínas. Em seguida, a extração com a mistura de fenol e tampão SDS denso permite uma eficiente solubilização e concentração das proteínas do precipitado, incluindo proteínas de membrana, deixando na fase aquosa substâncias hidrofílicas tais como carboidratos e ácidos nucleicos (WANG et al., 2003; WANG et al., 2006). Como amostras utilizou-se a mistura de folhas dos 5 indivíduos coletados de cada variedade e também as folhas de apenas um indivíduo das variedades *lingua* e *sylvestris* (árvores utilizadas no estudo de dinâmica metabólica), para fins de controle. As extrações foram feitas em duplicata e o rendimento obtido pode ser visualizado na **Tabela 29**.

Após a ressuspensão dos extratos secos em tampão H6, na proporção de 5 mg de extrato seco para 0,5 mL de volume final (extrato proteico bruto equivalente a 10,0 $\mu\text{g}/\mu\text{L}$) a quantificação do teor de proteína foi feita por dois diferentes métodos objetivando normalizar a quantidade de extrato a ser utilizada nas etapas posteriores de SDS-PAGE, *Western blotting* e LC-MS/MS.

Primeiramente, analisou-se as amostras pelo método do ácido bicinchonínico (BCA), sendo que foram feitas três diluições visando aumentar a confiabilidade do resultado. No entanto, apenas as diluições 1:20 e 1:100 foram consideradas para os cálculos. A **Tabela 25** a **Figura 52** a seguir detalham a curva analítica do BSA elaborada em duplicata para os cálculos pelo método do BCA. Já a **Tabela 26** resume os resultados encontrados na quantificação das proteínas por este método.

Tabela 25- Resultados de absorvância dos pontos da curva analítica de BSA para quantificação de proteínas pelo método do BCA.

Y Conc. ($\mu\text{g}/\mu\text{L}$)	Xa (ABS)	Xb (ABS)	X (media)
2,0	2,2231	2,2653	2,2442
1,0	1,3131	1,4163	1,3647
0,5	0,7730	0,8645	0,8188
0,25	0,4770	0,4850	0,4810
0,125	0,2848	0,2630	0,2739
0,0625	0,1557	0,1472	0,1515

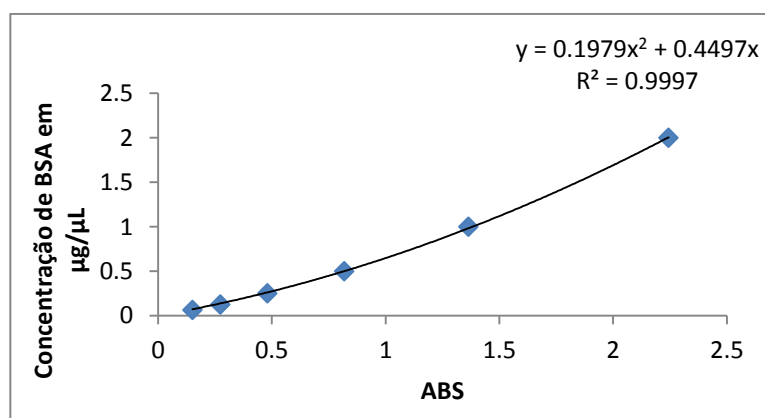


Figura 52- Curva analítica de BSA para quantificação de proteínas pelo método do BCA.

Tabela 26- Resultados obtidos na quantificação de proteínas pelo método do BCA.

Amostra/Diluição 1:20	Absorbância das replicatas			Média	Conc. $\mu\text{g}/\mu\text{L}$
L	0,4174	0,4145	0,4179	0,4166	4,43
PL1	0,3619	0,3828	0,4000	0,3816	4,01
PL2	0,3598	0,3776	0,3961	0,3778	3,96
S	0,3912	0,4183	0,4501	0,4199	4,47
PS1	0,5526	0,5748	0,6011	0,5762	6,50
PS2	0,4781	0,4742	0,5083	0,4869	5,32
Amostra/Diluição 1:100	Absorbância das replicatas			Média	Conc. $\mu\text{g}/\mu\text{L}$
L	0,1099	0,1172	0,1267	0,1179	5,58
PL1	0,1072	0,1031	0,0950	0,1018	4,78
PL2	0,0878	0,1048	0,0981	0,0969	4,54
S	0,0939	0,0882	0,0928	0,0916	4,29
PS1	0,1651	0,1598	0,1752	0,1667	8,05
PS2	0,1467	0,147	0,1591	0,1509	7,24

O método do BCA foi utilizado para que se pudesse fazer uma primeira avaliação quantitativa dos extratos obtidos. Este método, originalmente desenvolvido por Smith et al. (1985), tem como princípio a redução dos íons Cu^{2+} do sulfato cúprico pelas ligações peptídicas das proteínas em meio alcalino e subsequente complexação com o ácido bicinchonínico, levando à formação de uma cor lilás que absorve fortemente na região de 560 nm. Apesar desta reação ser similar aos métodos de Lowry e biureto e devido à presença de íons cúpricos e ser extensivamente utilizada por ser mais sensível e estável quando comparada aos demais métodos, ainda assim uma vasta gama de compostos tais como EDTA, lipídios, 2-mercaptoetanol, açúcares (ZAIA et al., 1998) Tween 20, DTT, glutatona (SAPAN et al., 1999), polifenóis (WEISS and BISSON, 2001) e SDS (GATES, 1991), dentre outros, podem causar resultados falso-positivos. Por este motivo, após a avaliação da SDS-PAGE obtida pela normalização das amostras através da quantificação pelo método de BCA, decidiu-se utilizar também a quantificação de proteínas totais pelo método do amido black, por ser menos susceptível a interferentes, principalmente os compostos fenólicos, SDS e EDTA, que provavelmente estavam presentes nas amostras.

Sendo assim e, da mesma forma que para o método do BCA, a **Tabela 27** e a **Figura 53** evidenciam os resultados a curva analítica do BSA elaborada em triplicata para os cálculos. Já a **Tabela 28** resume os resultados encontrados na quantificação das proteínas pelo método do amido black. Desta vez, foi feita apenas uma diluição em triplicata visto que uma análise quantitativa já havia sido feita pelo método do BCA.

Tabela 27- Resultados de absorvância dos pontos da curva analítica de BSA para quantificação de proteínas pelo método do amido black.

Y = Massa (μg)	Xa (ABS)	Xb (ABS)	Xc (ABS)	X (média)
0	0,0000	0,0000	0,0000	0,0000
1,25	0,0158	0,0232	0,0223	0,0204
2,50	0,0327	0,0382	0,0281	0,0330
5,00	0,0625	0,0961	0,0985	0,0857
10,00	0,1604	0,1279	0,1677	0,1520
20,00	0,3154	0,3233	0,1891	0,2759

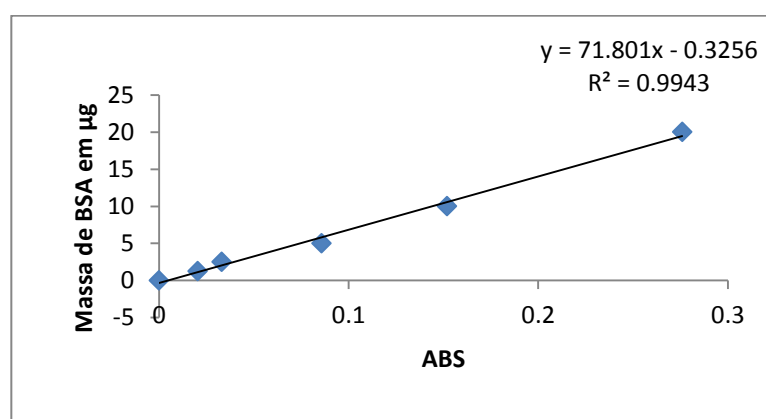


Figura 53- Curva analítica de BSA para quantificação de proteínas pelo método do amido black.

Tabela 28- Resultados obtidos na quantificação de proteínas pelo método do amido black.

Amostra	Replicatas			Média	Conc. ($\mu\text{g}/\mu\text{L}$)
L	0,0676	0,0875	0,1149	0,0900	3,07
PL1	0,0837	0,1040	0,1094	0,0990	3,39
PL2	0,0614	0,0621	0,0684	0,0639	2,13
S	0,0842	0,0945	0,1004	0,0930	3,18
PS1	0,1010	0,1260	0,1118	0,1129	3,89
PS2	0,0835	0,0943	0,0686	0,0821	2,78

Observando-se a **Tabela 29** e a **Figura 54** a seguir, nota-se que pela quantificação por BCA, os resultados obtidos são maiores que pelo método do amido black, provavelmente por conta de uma superestimação causada pela presença de possíveis interferentes no primeiro método.

Tabela 29- Resultados do rendimento dos extratos proteicos e do teor médio de proteínas das amostras de *C. sylvestris* pelos métodos do BCA e amido black.

Amostra	Código	Rendimento	*Teor (BCA)	*Teor (amido black)
Indivíduo var. <i>lingua</i>	L	24 mg	5,01	3,07
Mix var. <i>lingua</i> 1	PL1	54 mg	4,39	3,39
Mix var. <i>lingua</i> 2	PL2	40 mg	4,25	2,13
Indivíduo var. <i>sylvestris</i>	S	12 mg	4,38	3,18
Mix var. <i>sylvestris</i> 1	PS1	33 mg	7,27	3,89
Mix var. <i>sylvestris</i> 2	PS2	10 mg	6,28	2,78

* Teor de proteína no extrato proteico bruto a 10 $\mu\text{g}/\mu\text{L}$ (em $\mu\text{g}/\mu\text{L}$); media de três replicatas.

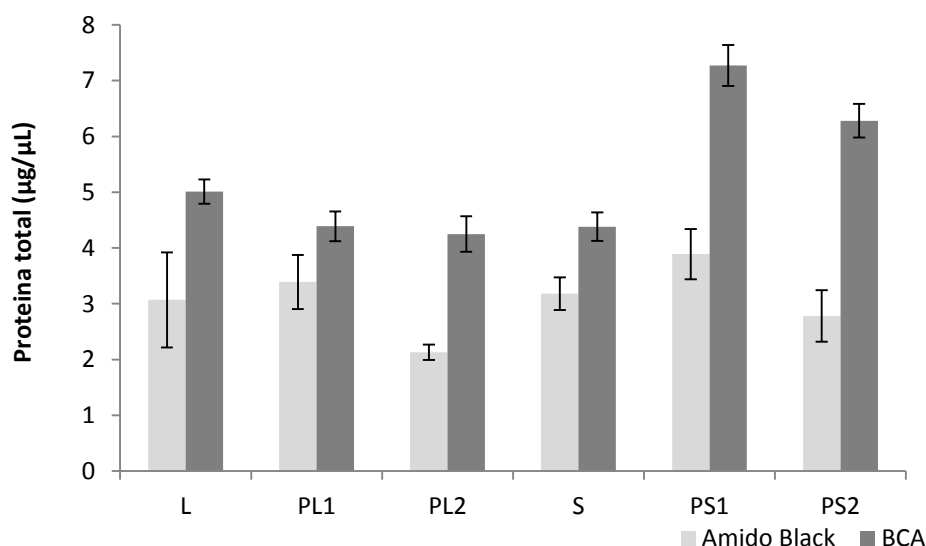


Figura 54- Gráfico do teor médio de proteínas das amostras de *C. sylvestris* pelos métodos do BCA e amido black.

Tal fato também pode ser facilmente observado nas diferentes eletroforeses obtidas a partir das normalizações obtidas com as diferentes alíquotas de amostras dependendo do método de quantificação. Mesmo normalizando-se a quantidade aplicada em cada duplicata, o perfil qualitativo é diferente, como pode ser visto na **Figura 55**, que mostra as eletroforeses obtidas pela normalização das amostras de acordo com os métodos do BCA e amido black, respectivamente.

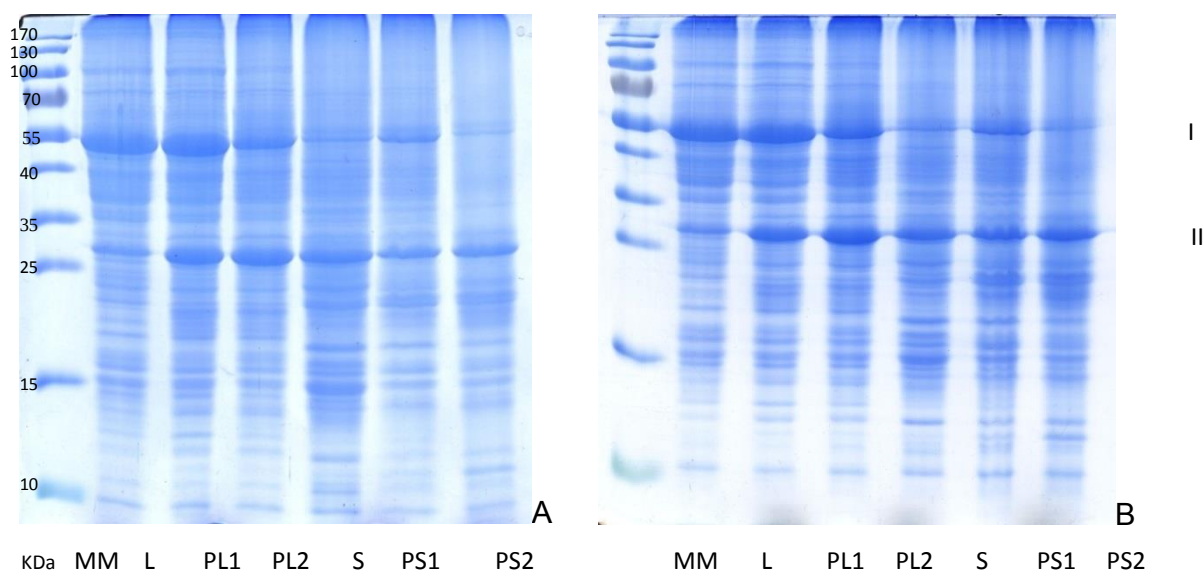


Figura 55- SDS-PAGEs dos extratos proteicos normalizados por dois métodos de quantificação. (A) amostra normalizada pelo método do BCA (aplicou-se em cada poço 20 μ g de proteínas); (B) amostra normalizada pelo método do amido black (aplicou-se em cada poço 10 μ g de proteínas). Coloração feita com *comassie blue*. As regiões com proteínas mais abundantes (faixas I e II) correspondem às proteínas RuBisCo (55 KDa) e proteínas coletoras de luz (*light-harvesting complex proteins* ou LHCs, 25 KDa).

Por este motivo e, considerando-se a menor sensibilidade do método do amido black à interferentes, principalmente no caso dos compostos fenólicos como já demonstrado por WEISS & BISSON (2001) estudando o teor proteico de sucos de uva e vinhos (reconhecidamente ricos em compostos fenólicos), todos os passos subsequentes tiveram as amostras normalizadas de acordo com este método.

5.5.2 Western blotting

O processo de normalização das amostras foi feito com base nas análises espectrofotométricas quantitativas e também pela análise qualitativa das eletroforeses obtidas. Este procedimento foi necessário para que se pudesse calcular a tomada de amostra para as etapas de digestão com tripsina e fracionamento (FASP e Pós-FASP) que viabilizariam as análises por LC-MS/MS. No entanto, durante este processo verificou-se um perfil qualitativo muito interessante considerando-se duas enzimas bastante abundantes em organismos fotossintéticos: a RuBisCo (**Figura 55**, região I, 55 KDa) e as proteínas coletoras de luz – LHCs (**Figura 55**, região II, 25 KDa)

Além disso, outro fator que estimulou um estudo mais aprofundado da expressão destas proteínas nas amostras das duas variedades de *C. sylvestris* foi o fato de que

dependendo da condição luminosa onde as plantas estão expostas, diferentes estratégias moleculares de adaptação irão ocorrer em sincronia para se modular a taxa da luz absorvida (GIVNISH, 1988; RUBAN, 2009).

À primeira vista, observando-se o gel corado com *comassie blue*, verificou-se que a variedade *lingua* apresentava maior expressão de proteínas RuBisCo quando comparada à variedade *sylvestris*. Já a expressão das LHCs, pareceu ser similar nas duas variedades. Sendo assim, para que se pudesse confirmar a presença diferencial destas proteínas nas duas variedades, foram feitos *Western blottings* utilizando anticorpos específicos para RuBisCo (anti-*rbcl*, *large subunit*) e LHC (anti-P25K).

Também, testou-se mais um anticorpo específico da via biossintética do metileritritol (IsPE, correspondente à 4-diphosphocytidyl-2C-methyl-D-erythritol kinase, Apêndice XVI), responsável pela biossíntese dos terpenóides para avaliar se desta forma seria possível diferenciar uma variedade de outra considerando a maior ocorrência de diterpenos na variedade *lingua*.

A **Figura 56**, a seguir, evidencia os resultados encontrados para *rbcl* e P25K. Comparando-se as duas amostras, verifica-se que há uma superexpressão de RuBisCo (*rbcl*) na variedade *lingua*, enquanto que as proteínas coletoras de luz permanecem (P25K), aparentemente semelhantes. Já IsPE apresentou-se mais abundante na variedade *lingua*.

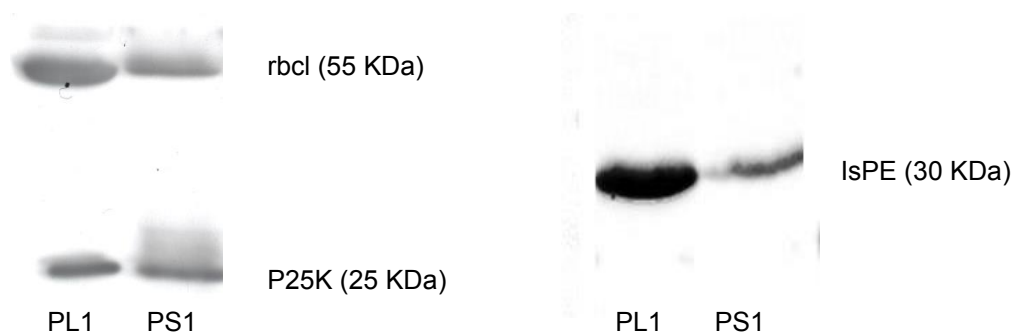


Figura 56- *Western blotting* das amostras PL1 (variedade *lingua*) e PS1 (variedade *sylvestris*) utilizando os anticorpos anti-*rbcl*, P25K e IsPE. Imagem obtida após revelação manual.

Para que todos estes resultados pudessem ser confirmados e para que se pudesse investigar a magnitude dessas diferenças, decidiu-se realizar uma quantificação relativa,

desta vez em três concentrações. As **Figuras 57, 58 e 59** mostram os resultados encontrados utilizando-se densitometria.

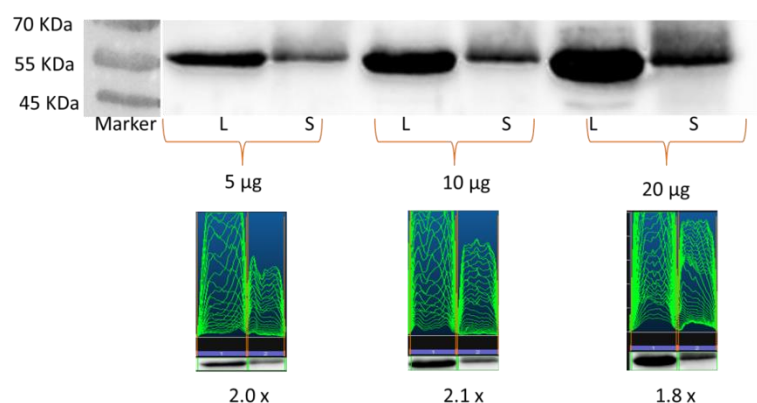


Figura 57- Análise semi-quantitativa do teor de RuBisCo (anti-rcbl) em três concentrações.

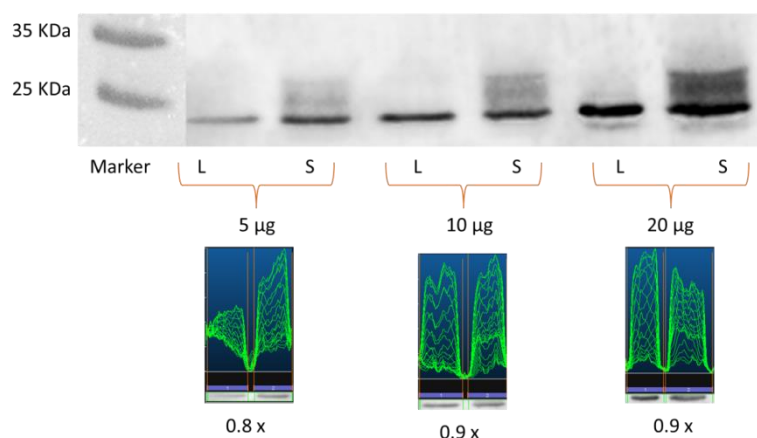


Figura 58- Análise semi-quantitativa do teor de LHC (anti-P25K) em três concentrações

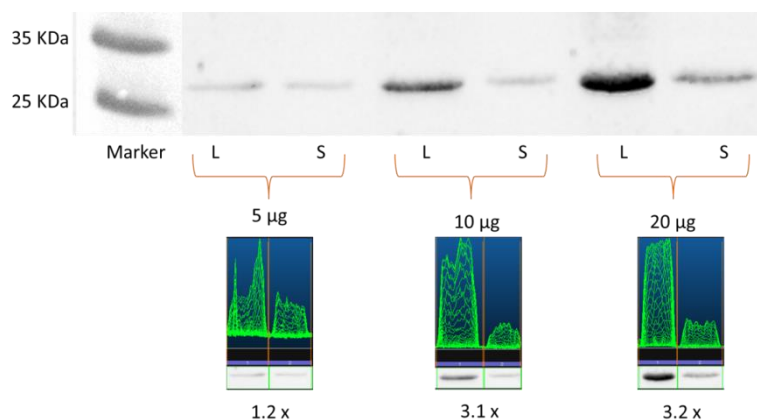


Figura 59- Análise semi-quantitativa do teor de ISPE (anti-IsPE) em três concentrações.

Analisando-se os resultados encontrados, confirmam-se algumas diferenças entre as variedades *lingua* e *sylvestris*: RuBisCo está sendo expressa na primeira cerca de 2 vezes mais que na *sylvestris*, enquanto que as LHCs cerca de 10% menos. O perfil destas enzimas fotossintéticas sugerem uma interessante relação destas variedades com a exposição à luz solar.

Já para IsPE confirma-se um resultado não esperado para a variedade *lingua*, ou seja, esperava-se encontrar menor expressão das enzimas responsáveis pela biossíntese de isoprenóides visto que o perfil metabolômico mostra que esta variedade é caracterizada pela presença de compostos fenólicos em detrimento dos diterpenos característicos da variedade *sylvestris*. No entanto, uma análise cuidadosa levanta a hipótese de que tais isoprenóides estariam sendo destinados à biossíntese não de diterpenos, mas sim de carotenoides, visto que existe uma relação bastante importante entre o teor de pigmentos nas folhas e os fotossistemas I e II nos mecanismos de adaptação do aparato fotossintético à luminosidade. Sabe-se que folhas de plantas mais expostas à altas taxas de luminosidade possuem maiores taxas de β -caroteno, quando comparadas com plantas de ambiente de sombra (LICHTENTHALER, 2009). De fato, observações de campo deste trabalho revelaram que os indivíduos da variedade *lingua* ocorrem geralmente em ambientes mais abertos sendo, portanto, expostos à maior incidência de luz solar.

5.5.3 Identificação das proteínas do estudo proteômico diferencial

Estudos proteômicos em plantas superiores é uma área bastante recente na pesquisa de produtos naturais, e vem de encontro com a biologia sistêmica e genômica funcional. Neste contexto, um dos desafios do presente trabalho foi iniciar o estudo proteômico de *C. sylvestris*, visando conhecer melhor a espécie e obter dados que permitam ou auxiliem a diferenciação de duas variedades, *lingua* e *sylvestris*, (SLEUMER, 1980) que, do ponto de vista metabolômico, produzem classes de metabólitos pontualmente diferentes. Uma vez que nesta espécie existem importantes variações morfológicas e fenotípicas e que estas ocorrem de acordo com os diferentes ecossistemas de origem (TORRES; YAMAMOTO, 1986), sugere-se que tais eventos podem estar relacionados com as adaptações ao ambiente e com a plasticidade fenotípica.

Na etapa de identificação de proteínas utilizando ferramentas em bioinformática, duas estratégias foram utilizadas para a identificação do maior número possível de

proteínas nas duas variedades visto que para a espécie em estudo não existem bancos de dados disponíveis. Para a primeira estratégia (Estratégia 1, Apêndice IV), construiu-se um banco de dados misto de diversas espécies vegetais que estivessem filogeneticamente próximas à *C. sylvestris* e/ou que tivessem um histórico da ocorrência de classes de metabólitos secundários que também são descritas para esta espécie. Tomando-se por base que a espécie com estudos proteômicos registrados mais próxima de *C. sylvestris* é a espécie *Populus trichocarpa*, ambas pertencentes à família Salicaceae, emergiu-se os bancos de dados de espécies próximas à esta última. Desta forma foram utilizados os bancos de dados de *Ricinus communis*, *Populus trichocarpa*, *Linum usitatissimum*, *Medicago truncatula*, *Glycine max* e *Oriza sativa* (esta última pelo histórico de ocorrência de diterpenos). A segunda estratégia (Estratégia 2, Apêndice IV) considerou apenas o banco de dados de *P. trichocarpa*. A **Tabela 30** a seguir mostra o número de peptídeos identificados para cada variedade nas diferentes estratégias.

Tabela 30- Número de peptídeos identificados de acordo com duas diferentes estratégias.

Amostra	Número de peptídeos identificados de acordo com os diferentes bancos de dados	
	Banco composto por 6 espécies vegetais	Banco de dados de <i>P. trichocarpa</i>
PL1 (<i>lingua</i>)	3456	2953
PS1 (<i>sylvestris</i>)	3270	2925

Verifica-se que o número de peptídeos identificados é maior na estratégia que utiliza o banco de dados composto por seis diferentes espécies, o que era realmente esperado. Nesta abordagem foi possível identificar, preliminarmente, cerca de 140 enzimas pertencentes ao metabolismo secundário, mais especificamente das vias do metileritritol e do ácido chiquímico mesmo depois das diversas etapas de filtros de qualidade aplicados no Proteomatic®, diferentemente da abordagem quando se utilizou apenas banco de dados de *P. trichocarpa*, visto que pouquíssimas enzimas das rotas biossintéticas do ácido chiquímico e do metileritritol foram encontradas. O Apêndice XVII resume as proteínas mais importantes, bem como os peptídeos identificados para cada uma delas com base no banco de dados misto. Apenas para citar, dentre as principais enzimas identificadas estão diversas 4-coumarato-CoA ligases, antocianidinas

glucosiltransferases e redutases, caroteno oxigenases, hidroxilases e isomerases, chalcona sintases e isomerases, diidroflavonol redutases, flavonol sintases, isoflavonas sintases, caureno sintases, protocatecuato 3,4-dioxigenase, fenilalanina e histidina amônia-liases, chiquimato desidrogenases, licopeno ciclase, geranilgeranil difosfato sintases, geranilgeranil transferases e farnesil sintases, dentre outras.

No entanto, ao se analisar os dados mais detalhadamente e ao tentar agrupar os peptídeos em nível de proteínas, percebeu-se que realizar esta tarefa automaticamente utilizando softwares não era exequível, impossibilitando as análises comparativas subsequentes. Isto porque ao se misturar diferentes bancos de dados, muitos peptídeos idênticos recebiam diferentes códigos ou notações (que variam de acordo com o próprio banco de dados). Assim, a única alternativa foi utilizar apenas um banco de dados para o grupamento e análise comparativa, neste caso o banco de dados de *P. trichocarpa*.

Nesta etapa, mais uma estratégia de filtro foi aplicada, em nível de peptídeos, visto que comparar duas variedades diferentes, que não sofreram nenhum tipo de marcação em laboratório (*label-free* quantitation), exigia maior critério de qualidade. Neste caso, considerou-se apenas peptídeos que tivessem sido detectados em pelo menos 4 *scans counts*, ou seja, em 4 espectros de massas. Para aqueles peptídeos que atenderam a este critério, calculou-se a razão, o número de *scan counts* para a variedade *lingua* e para a variedade *sylvestris*. Finalmente, os peptídeos que atenderam à este critério foram identificados através de um *script* de anotação utilizando-se os arquivos anteriores com maioria dos peptídeos anotados via BLASTP, ou utilizando-se os arquivos de cálculos com as anotações provenientes dos bancos de dados mistos ou ainda manualmente via NCBI web site.

Desta forma foram identificadas 820 sequências de peptídeos (Apêndice XVIII), referentes a 547 diferentes proteínas, as quais finalmente foram agrupadas de acordo com sua função biológica. A **Figura 60** a seguir resume o resultado encontrado considerando-se todos os grupos de proteínas identificadas nas duas variedades. Nota-se que aproximadamente 60% das proteínas identificadas pertencem à função do metabolismo energético, em parte por serem mais abundantes no extrato proteico total (como já evidenciado na SDS-PAGE) e também por pertencerem às regiões mais conservadas entre as espécies permitindo assim a identificação com maior sucesso.



Figura 60- Classificação de todas proteínas identificadas nas duas variedades em estudo de acordo com suas funções biológicas.

Já na análise comparativa entre as duas variedades em estudo, ainda considerando-se as proteínas identificadas e agrupadas de acordo com a função biológica, os dados foram agrupados em função da porcentagem dos *scan counts*, conforme apresentado na **Figura 61** a seguir.

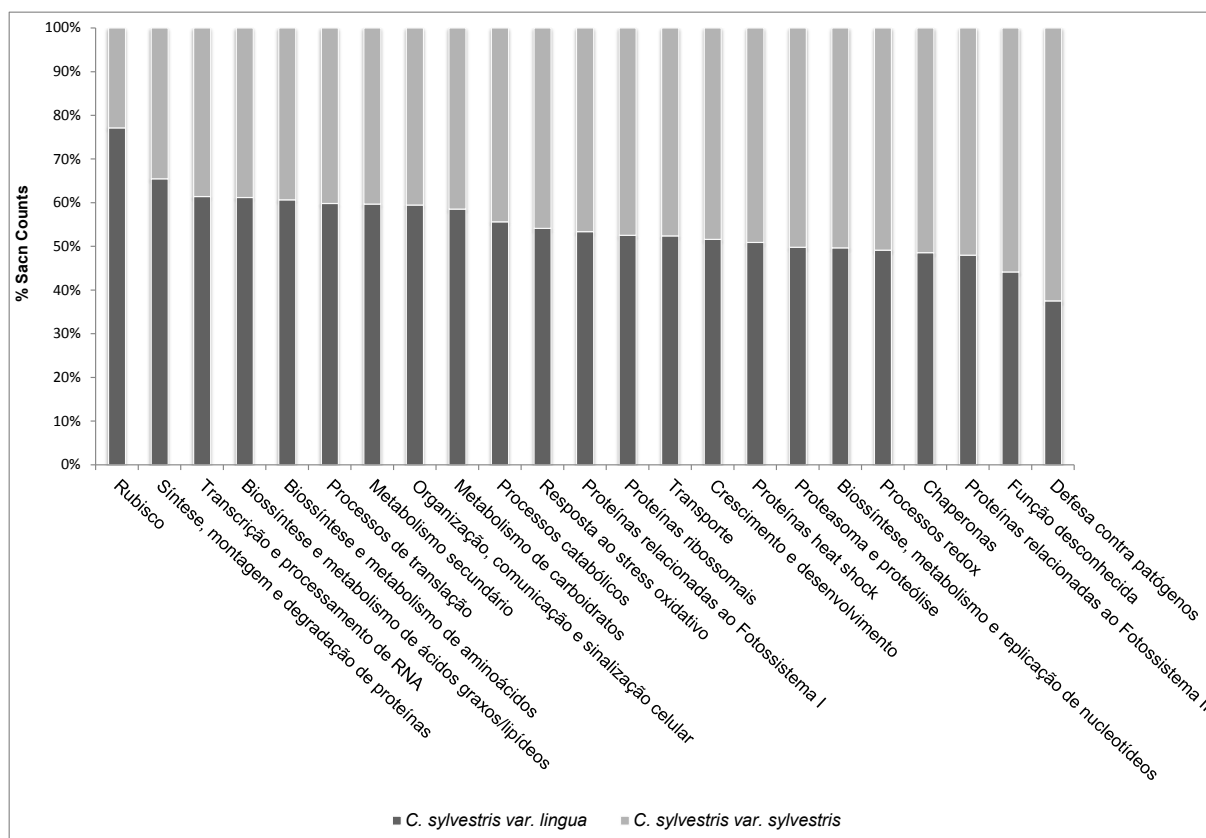


Figura 61- Análise comparativa entre *C. sylvestris* var. *sylvestris* e *C. sylvestris* var. *lingua* em função dos scan counts.

Dentre as classes funcionais, a mais abundante é o grupo de “Metabolismo energético e Fotossíntese”, com 48,9% de todas as proteínas identificadas. Este grupo pode ser dividido em proteínas relacionadas com o metabolismo de carboidratos (61,5%), proteínas relacionadas ao fotossistema I (4,2%), proteínas relacionadas ao fotossistema II (12,9%) e Rubisco (21,4%). Um olhar mais atento evidencia dados interessantes relacionados com o aparato fotossintético, corroborando com os resultados obtidos por *Western blotting*. Os dados evidenciam que as proteínas do fotossistema II estão 10% mais expressas na variedade *sylvestris*, quando comparada com variedade *lingua* (281 x 259 scan counts, respectivamente), enquanto que as proteínas do fotossistema I são 10% menos expressas (83 x 95 scan counts, respectivamente). Da mesma forma, corroborando com os dados obtidos por *Western blotting*, a RuBisCo está cerca de três vezes mais expressa na variedade *lingua*, em comparação com a variedade *sylvestris* (692 x 205 scan counts, respectivamente). Os resultados seguem sumarizados na **Figura 62**, que mostra uma análise mais detalhada da expressão das proteínas RuBisCo, fotossistema II e fotossistema I (LHC II e LHC I, respectivamente).

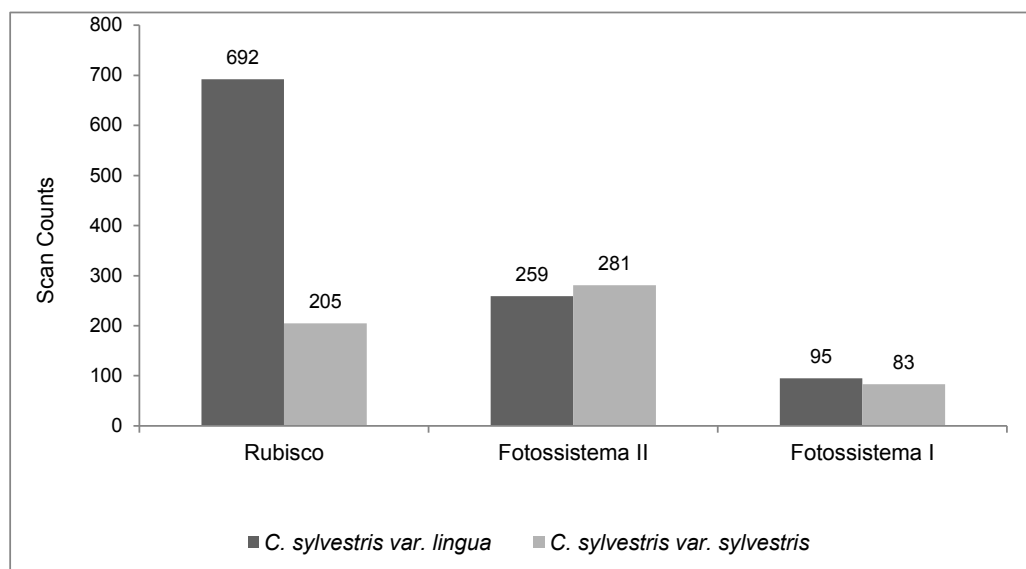


Figura 62- Análise comparativa quantitativa das proteínas fotossintéticas expressas por *C. sylvestris* var. *lingua* e *C. sylvestris* var. *sylvestris*.

A análise conjunta destes resultados permite uma diferenciação preliminar das duas variedades, sugerindo uma possível relação com o ambiente ao qual estão submetidas. Dependendo da intensidade da luz em que as plantas se encontram, diferentes estratégias de adaptação molecular poderão ocorrer em sincronia para modular a taxa de luz absorvida (GIVNISH, 1988; RUBAN, 2009). As plantas expostas a baixas intensidades de luz tendem a expressar níveis mais baixos de RuBisCo, uma vez que a regeneração de substrato RuBP é dependente da produção de ATP e NADP, que são menores em ambientes de pouca luz (SEEMAN, 1989). De fato, a variedade *sylvestris*, encontrada principalmente em áreas de baixa intensidade de luz, como áreas de Mata Atlântica, mostrou esse padrão de menor expressão de RuBisCo. Em comparação, as plantas encontradas sob alta intensidade de luz solar (como as áreas de Cerrado do Brasil, bem como em micro-localidades onde o efeito do dossel é reduzido (onde variedade *lingua* é comumente encontrada) tendem a expressar níveis mais elevados de RuBisCo. Da mesma forma, o complexo antena (onde estão localizadas as proteínas coletoras de luz - LHC, especialmente o Fotossistema II) de plantas altamente expostas ao sol é relativamente menor quando comparado com as plantas que crescem em ambientes de sombra. Sabe-se que este complexo é reduzido pelas enzimas proteolíticas do complexo antena (também dependentes de ATP), que degradam a porção exterior do complexo LHCII (RUBAN, 2009).

Embora as proteínas do metabolismo secundário não tenha sido o grupo mais abundante em número de proteínas identificadas (representando apenas 1,5% do total de proteínas identificadas neste estudo), delas foi possível obter informações importantes para a análise diferencial das duas variedades de *C. sylvestris* (**Figura 63**).

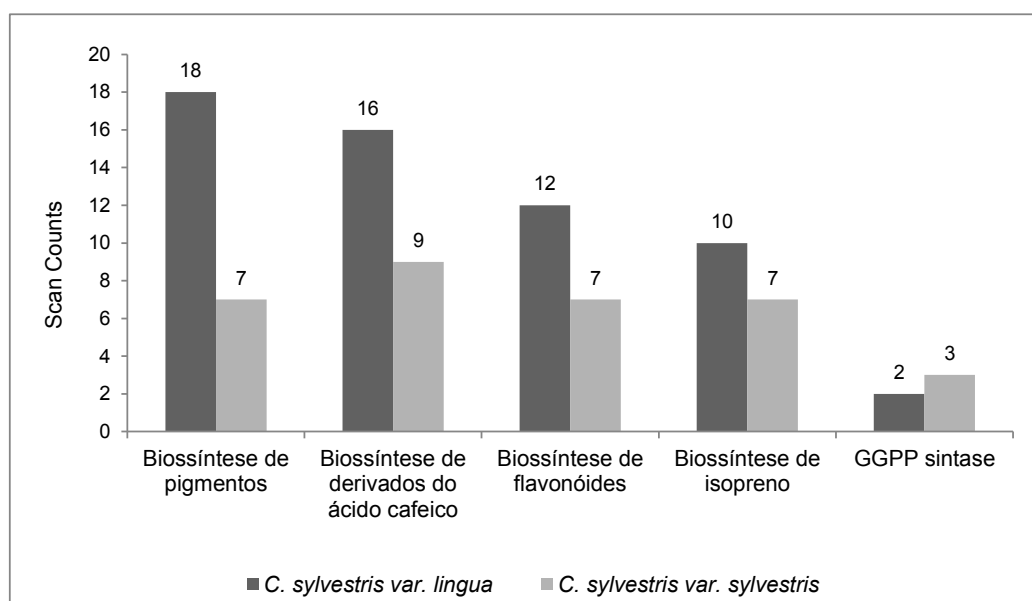


Figura 63- Análise comparativa quantitativa das proteínas do metabolismo secundário expressas por *C. sylvestris* var. *lingua* e *C. sylvestris* var. *sylvestris*.

Um olhar atento neste grupo revelou uma maior expressão de geranylgeranil difosfato sintase (ou GGPP sintase, enzima chave envolvida na biossíntese diterpenóides) na variedade *sylvestris*, como esperado, uma vez que esta variedade é caracterizada pela presença maciça de diterpenos clerodânicos. Para a variedade *lingua*, observou-se um aumento da expressão de enzimas responsáveis pela biossíntese de compostos fenólicos, mais precisamente com a biossíntese de flavonoides (antocianidina 3-O-glucosiltransferase, flavanona 3-hidroxilase e oxidoredutase) e de derivados do ácido cafeico (ácido cafeico 3-O-metiltransferase e ácido cafeico/ácido 5-hidroxiferúlico-3,5-O-metiltransferase). Embora as análises cromatográficas realizadas neste estudo tenham evidenciado a ocorrência de flavonoides glicosilados na variedade *lingua*, até o momento, não há indícios da presença de derivados do ácido cafeico nesta espécie.

Finalmente, as enzimas 2-C-metil-D-eritritol 2,4-cyclodiphosphate sintase (ISPF) e 2-C-metil-D-eritritol 4-fosfato cytidyltransferase (PDSI) relacionadas com a biossíntese de isoprenos (precursor de todas as vias biossintéticas dos terpenóides,

incluindo monoterpênos, diterpênos e carotenos) estão super-expressas na variedade *lingua*. Este resultado sugere uma relação interessante com a biossíntese de pigmentos, uma vez que as enzimas carotenóide 9,10(9',10)-clivagem dioxigenase, carotenóide-clivagem dioxigenase, licopeno epsilon ciclase e prolicopeno isomerase também são super-expressas na variedade *lingua*. Do ponto de vista fotossintético, os pigmentos são essenciais para os processos de captação de luz, bem como para a fotoproteção. Sabe-se que o excesso de energia luminosa é dissipada na forma de calor dentro dos complexos antena, protegendo o aparelho fotossintético contra o dano oxidativo (ASHRAF; HARRIS, 2013). Além disso, dependendo das condições de crescimento da folha e as intensidades de luz a que estão expostas, a razão entre a quantidade de clorofilas e carotenóides será diferente. Folhas muito expostas à luz solar possuem níveis de β -caroteno que são pelo menos duas vezes maior que em folhas de sombra, enquanto que o teor de clorofila total em folhas altamente expostas ao sol pode ser cerca de 1,3 a 1,5 vezes maior em comparação com folhas de sombra (LICHTENTHALER, 2009).

Os resultados relativos à expressão das enzimas do aparato fotossintético e das enzimas do metabolismo secundário nas duas variedades de *C. sylvestris*, corroboram com os perfis químicos apresentados pela espécie.

6 CONCLUSÕES E PERSPECTIVAS

A estratégia experimental multidisciplinar utilizada no decorrer deste trabalho bem como os resultados aqui descritos e discutidos permitem fundamentar a hipótese de que a produção diferencial de metabólitos secundários nas duas variedades de *C. sylvestris* está relacionada com os biomas de origem e, portanto, também associada aos respectivos morfotipos predominantes.

Tal proposição deve-se à comprovação da produção diferencial de flavonoides e diterpenos clerodânicos pelas variedades *lingua* e *sylvestris*, respectivamente, constantes durante todo o ciclo circadiano e com modulações apenas no período reprodutivo.

Da mesma forma, o estudo da variabilidade química infraespecífica confirma a produção prioritária de flavonoides glicosilados pela variedade *lingua*, bem como correlaciona a ocorrência destes compostos em áreas de alta incidência de luz solar (como é o caso das áreas de Cerrado e Carrasco); para a variedade *sylvestris*, foi possível observar a produção característica de diterpenos clerodânicos, correlacionando sua ocorrência em áreas de menor incidência de luz solar, onde o efeito do dossel da vegetação é pronunciado (condição característica das áreas de Floresta Ombrófila Densa e Floresta Estacional Semi-Decidua). Pela análise de componentes principais, indivíduos de classificação morfológica intermediária concentraram-se principalmente em áreas de clima quente, com alta incidência de luz solar, porém em regiões alagadiças, a exemplo do Pampa e Pantanal.

Os estudos proteômicos complementam os resultados químicos obtidos e evidenciam os mecanismos de adaptação do aparato fotossintético de plantas oriundas de ambientes de maior ou menor exposição à luz solar. Os resultados mostraram que a RuBisCo está cerca de três vezes mais expressa na variedade *lingua*, em comparação com a variedade *sylvestris*, ao contrário das proteínas do fotossistema II, que estão sendo mais expressas na variedade *sylvestris*. Além disso, os estudos proteômicos levantam indícios de que a rota biossintética dos isoprenóides está sendo desviada para a produção de pigmentos na variedade *lingua*, uma vez que foi detectada uma maior expressão de enzimas relacionadas a estes metabólitos nesta variedade.

Entretanto, vale a pena ressaltar que os resultados obtidos neste trabalho lançam ainda uma questão prioritária para estudos vindouros: qual a importância, função ou

relação dos diterpenos clerodânicos e flavonoides glicosilados em ambientes sombreados e úmidos ou ensolarados e secos, respectivamente, onde tanto a variedade *sylvestris* quanto a variedade *lingua* possuem vantagens competitivas?

Ainda, dentre os resultados alcançados neste trabalho também destacam-se:

- a) O desenvolvimento e validação de um método analítico por UHPLC-DAD para a análise simultânea de diterpenos clerodânicos e flavonoides em *C. sylvestris*, o qual poderá ser utilizado de forma sistemática em trabalhos futuros;
- b) A identificação de 15 diterpenos clerodânicos em *C. sylvestris* var. *sylvestris* utilizando técnicas de desreplicação e a utilização de um banco de dados construído especialmente para este estudo;
- c) A identificação inédita de 15 flavonoides em *C. sylvestris* var. *lingua*, sendo 14 deles com o mesmo padrão de glicosilação (flavonoides 3-O-glicosilados), e que até então eram ainda muito pouco conhecidos para a espécie;
- d) A identificação de 547 proteínas diferencialmente expressas nas duas variedades em estudo, resultados também inéditos para tal espécie selvagem.

Finalmente, conclui-se que a integração de diferentes técnicas cromatográficas com as abordagens metabolômica e proteômica, bem como quimiometria permitiram acessar e analisar importantes informações acerca da variabilidade química apresentada por *C. sylvestris*. Tais resultados poderão ser utilizados na fundamentação de futuros trabalhos acerca da escolha de matrizes para cultivo desta espécie para fins medicinais e biotecnológicos, no estudo das rotas biossintéticas para a produção dos metabólitos secundários de interesse, na condução de ensaios de atividades biológicas relacionadas com ambas as variedades, em estudos de adaptação de plantas à diferentes condições ambientais, bem como no estudo de ecologia de populações e, finalmente, no melhor conhecimento da nossa biodiversidade.

REFERÊNCIAS

- ABLAJAN, K.; ABLIZ, Z.; SHANG, X.; HE, J.; ZHANG, R.; SHI, J. Structural characterization of flavonol 3,7-di-O-glycosides and determination of the glycosylation position by using negative ion electrospray ionization tandem mass spectrometry. **Journal of Mass Spectrometry**, v. 41, n. 3, p. 352-360, 2006.
- ABRIL, N.; GION, J.; KERNER, R.; MÜLLER-STARCK, G.; CERRILLO, R. M. N.; PLOMION, C.; RENAUT, J.; VALLEDOR, L.; JORRIN-NOVO, J. V. Proteomics research on forest trees, the most recalcitrant and orphan plant species. **Phytochemistry**, v. 72, n. 10, p. 1219-242, 2011.
- AGRAWAL, G. K.; YONEKURA, M.; IWAHASHI, Y.; IWAHASHI, H.; RAKWAL, R. System, trends and perspectives of proteomics in dicot plants. Part I: technologies in proteome establishment. **Journal of Chromatography B**, v. 815, n. 1/2, p. 109-123, 2005.
- ASHRAF, M.; HARRIS, P. J. C. Photosynthesis under stressful environments: an overview. **Photosynthetica**, v. 51, n. 2, p. 163-190, 2013.
- BANDEIRA, K. F.; TININIS, A. G.; BOLZANI, V. S.; CAVALHEIRO, A. J. Optimisation of conditions for the extraction of casearins from *Casearia sylvestris* using response surface methodology. **Phytochemical Analysis**, v. 17, n. 3, p. 168-175, 2006.
- BASILE, A. C.; SERTIE, J. A. A.; PANNIZZA, S.; OSHIRO, T. T.; AZZOLINI, C. A. Pharmacological assay of *Casearia sylvestris* I: preventive anti-ulcer activity and toxicity of the leaf crude extract. **Journal of Ethnopharmacology**, v. 30, n. 2, p. 185-197, 1990.
- BORGES, M. H.; SOARES, A. M.; RODRIGUES, V. M.; ANDRIÃO-ESCARSO, S. H.; DINIZ, H.; HAMAGUCHI, A.; QUINTERO, A.; LIZANO, S.; GUTIÉRREZ, J. M.; GIGLIO, J. R.; HOMSI-BRANDEBURGO, M. I. Effects of aqueous extract of *Casearia sylvestris* (Flacourtiaceae) on actions of snake and bee venoms and on activity of phospholipases A2. **Comparative Biochemistry and Physiology**, v. 127, n. 1, p. 21-30, 2000.
- BORGES, M. H.; SOARES, A. M.; RODRIGUES, V. M.; OLIVEIRA, F.; FRANSCHESCHI, A. M.; RUCAVADO, A.; GIGLIO, J. R.; HOMSI-BRANDEBURGO, M. I. Neutralization of proteases from *Bothrops* snake venoms by the aqueous extract from *Casearia sylvestris* (Flacourtiaceae). **Toxicon**, v. 39, n. 12, p. 1863-1869, 2001.
- CARVALHO, E. S.; SANTOS, A. G.; CAVALHEIRO, A. J. Identificação de diterpenos clerodânicos em diferentes órgãos de *Casearia sylvestris* Sw. **Revista de Ciências Farmacêuticas Básica e Aplicada**, v. 30, n. 3, p. 277-284, 2009.
- CARVALHO, P. R. F.; FURLAN, M.; YOUNG, M. C. M.; KINGSTON, D. G. I.; BOLZANI, V. S. Acetylated DNA-damage clerodane diterpenes from *Casearia sylvestris*. **Phytochemistry**, v. 49, n. 6, p. 1659-1662, 1998.
- CAVALCANTE, L. G. W.; CAMPOS, T. O.; SILVA, M. D. P.; PEREIRA, P. S.; OLIVEIRA, C. Z.; SOARES, A. M.; GALLACCI, M. Neutralization of snake venom phospholipase A2 toxins by aqueous extract of *Casearia sylvestris* (Flacourtiaceae) in mouse neuromuscular preparation. **Journal of Ethnopharmacology**, v. 112, n. 3, p. 490-497, 2007.

- CAVALLARI, M. M. **Variabilidade genética e química entre e dentro de populações de *Casearia sylvestris* Sw. (Salicaceae)**. 2008. 127 f. Tese (Doutorado em Ciências Biológicas) - Instituto de Biociências, Universidade Estadual Paulista, Botucatu, 2008.
- CAVALLARI, M. M.; GIMENES, M. A.; BILLOT, C.; TORRES, R. B.; CAVALHEIRO, A. J.; BOUVET, J. M. Population genetic relationships between *Casearia sylvestris* (Salicaceae) varieties occurring sympatrically and allopatrically in different ecosystems in south-east Brazil. **Annals of Botany**, v. 106, n. 4, p. 627-636, 2010.
- CHAGAS-PAULA, D.; ZHANG, T.; COSTA, F. B. da; EDRADA-EBEH, R. A metabolomic approach to target compounds from the Asteraceae family for dual COX and LOX inhibition. **Metabolites**, v. 5, n. 3, p. 404-430, 2015.
- CHASE, M. W.; ZMARTZTY, S.; LLEDÓ, M. D.; WURDACK, K. J.; SWENSEN, S. M.; FAY, M. F. When in doubt, put it in Flacourtiaceae: a molecular phylogenetic analysis based on plastid rbcL DNA sequences. **Kew Bulletin**, v. 57, p. 141-181, 2002.
- CLELAND, E. E.; CHUINE, I.; MENZEL, A.; MOONEY, H. A.; SCHWARTZ, M. D. Shifting plant phenology in response to global change. **Trends in Ecology and Evolution**, v. 22, n. 7, 2007. doi:10.1016/j.tree.2007.04.003.
- COON, M. P. (Ed.). **Handbook of proteomics methods**. New Jersey: Humana Press, 2003. 510 p.
- CORRÊA, M. P. **Dicionário das plantas úteis do Brasil e das exóticas cultivadas**. Rio de Janeiro: Imprensa Nacional, 1984. v. 2, 747 p.
- COX, C. B.; MOORE, P. D. **Biogeografia**: uma abordagem ecológica e evolucionária. 7. ed. Rio de Janeiro: LTC, 2009. 412 p.
- CRAIG, R.; BEAVIS, R. C. TANDEM: matching proteins with tandem mass spectra. **Bioinformatics**, v. 20, n. 9, p. 1466-1467, 2004.
- CROTEAU, R.; KUTCHAN, T. M.; LEWIS, N. G. Natural products (secondary metabolites). In: BUCHANAN, B.; GRUISSEM, W.; JONES, R. (Ed.). **Biochemistry & molecular biology of plants**. Rockville: American Society of Plant Physiologists, 2000. Chap. 24, p. 1250-1318.
- CUYCKENS, F.; CLAEYS, M. Mass spectrometry in the structural analysis of flavonoids. **Journal of Mass Spectrometry**, v. 39, n. 1, p. 1-15, 2004.
- DAMKJAER, J. T.; KEREICHE, S.; JOHNSON, M. P.; KOVACS, L.; KISS, A. Z.; BOEKEM, E. J.; RUBAN, A. V.; HORTON, P.; JANSSEN, S. A. The photosystem II light-harvesting protein Lhcb3 affects the macrostructure of photosystem II and the rate of state transitions in *Arabidopsis*. **The Plant Cell**, v. 21, n. 10, p. 3245-3256, 2009.
- DELAHUNTY, C.; YATES III, J. R. Protein identification using 2D-LC-MS/MS. **Methods**, v. 35, n. 3, p. 248-255, 2005.
- DEWICK, P. M. **Medicinal natural products**: a biosynthetic approach. 3rd ed. Chichester: Wiley, 2009. 539 p.

- ESTEVES, I.; SOUZA, I. R.; RODRIGUES, M.; CARDOSO, L. G. V.; SANTOS, L. S.; SERTIE, J. A. A.; PERAZZO, F. F.; LIMA, L. M.; SCHNEEDORF, M.; BASTOS, J. K.; CARVALHO, J. C. Gastric antiulcer and anti-inflammatory activities of the essential oil from *Casearia sylvestris* Sw. **Journal of Ethnopharmacology**, v. 101, n. 1/3, p. 191-196, 2005.
- FACANALI, R. **Ecologia de populações de espécies prioritárias para conservação e uso**: um estudo de caso, usando como modelo a *Casearia sylvestris* Sw. 2004. Monografia apresentada para a Disciplina NT238, Ecologia de Populações de Plantas, Instituto de Biologia, Universidade de Campinas, Campinas, 2004.
- FERREIRA, P. M. P.; COSTA-LOTUFO, L. V.; MORAES, M.; BARROS, F. W. A.; MARTINS, A. M. A.; CAVALHEIRO, A. J.; BOLZANI, V. S.; SANTOS, A. G.; PESSOA, C. Folk uses and pharmacological properties of *Casearia sylvestris*: a medicinal review. **Anais da Academia Brasileira de Ciências**, v. 83, n. 4, p. 1-12, 2011.
- FERREIRA, P. M. P.; SANTOS, A. G.; TININIS, A. G.; COSTA, P. M.; CAVALHEIRO, A. J.; BOLZANI, V. S.; MORAES, M. O.; COSTA-LOTUFO, L. V.; MONTENEGRO, R. C.; PESSOA, C. Casearin X exhibits cytotoxic effects in leukemia cells triggered by apoptosis. **Chemical-Biological Interactions**, v. 188, n. 3, p. 497-504, 2010.
- FRIDMAN, E.; PICHERSKY, E. Metabolomics genomics, proteomics, and the identification of enzymes and their substrates and products. **Current Opinion in Plant Biology**, v. 8, n. 3, p. 242-248, 2005.
- FUNARI, C. S.; CASTRO-GAMBOA, I.; CAVALHEIRO, A. J.; BOLZANI, V. S. Metabolômica, uma abordagem otimizada para exploração da biodiversidade brasileira: estado da arte, perspectivas e desafios. **Química Nova**, v. 36, n. 10, p. 1605-1609, 2013.
- GATES, R. R. Elimination of interfering substances in the presence of detergent in the bicinchoninic acid protein assay. **Analytical Biochemistry**, v. 196, n. 12, p. 290-295, 1991.
- GEER, L. Y.; MARKEY, S. P.; KOWALAK, J. A.; WAGNER, L.; XU, M.; MAYNARD, D. M.; YANG, X.; SHI, W.; BRYANT, S. H. Open mass spectrometry search algorithm. **Journal of Proteome Research**, v. 3, n. 5, p. 958-964, 2004.
- GIVNISH, T. J. Adaptation to sun and shade: a whole-plant perspective. **Australian Journal of Plant Physiology**, v. 15, n. 2, p. 63-92, 1988.
- GLAUSER, G.; VEYRAT, N.; ROCHAT, B.; WOLFENDER, J.; TURLINGS, T. C. Ultra-high pressure liquid chromatography-mass spectrometry for plant metabolomics: a systematic comparison of high-resolution quadrupole-time-of-flight and single stage Orbitrap mass spectrometers. **Journal of Chromatography A**, v. 1292, p. 151-159, 2013.
- GREEN, J. M. A practical guide to analytical method validation. **Analytical Chemistry**, v. 68, n. 9, p. 305-309, 1996.
- HARBORNE, J. B. **Introduction to ecological biochemistry**. 4th ed. London: Academic Press, 1993. 318 p.

HARBORNE, J. B. **The flavonoids**: advances in research since 1986. London: Chapman & Hall, 1996. 675 p.

HEY, S. J.; BYRNE, E.; HALFORD, N. G. The interface between metabolic and stress signalling. **Annals of Botany**, v. 105, n. 2, p. 197-203, 2010.

INSTITUTO NACIONAL DE METROLOGIA. **DOQ-CGCRE-008 Revisão 3**: orientações sobre validação de métodos de ensaios químicos. Brasília/DF, 2010. 20 p.

JORRIN-NOVO, J. V.; MALDONADO, A. M.; ECHEVARRÍA-ZOMEÑO, S.; VALLEDOR, L.; CASTILLEJO, M. A.; CURTO, M.; VALERO, J.; SGHAIER, B.; DONOSO, G.; REDONDO, I. Plant proteomics update (2007–2008): second-generation proteomic techniques, an appropriate experimental design, and data analysis to fulfill MIAPE standards, increase plant proteome coverage and expand biological knowledge. **Journal of Proteomics**, v. 72, n. 3, p. 285-312, 2009.

JUDD, W. S. **Plant systematics**: a phylogenetic approach. 2nd ed. Sunderland: Sinaues Associates, 2002. 576 p.

JUNQUEIRA, M.; SPIRIN, V.; BALBUENA, T. S.; THOMAS, H.; ADZHUBEI A.; SUNYAEV S.; SHEVCHENKO, A. Protein identification pipeline for the homology-driven proteomics. **Journal of Proteomics**, v. 71, n. 3, p. 346-356, 2008.

KÄLL, L.; STOREY, J. D.; NOBLE, W. S. Qvalue: non-parametric estimation of q-values and posterior error probabilities. **Bioinformatics**, v. 25, n. 7, p. 964-966, 2009.

KESSNER, D.; CHAMBERS, M.; BURKE, R.; AGUS, D.; MALLICK, P. ProteoWizard: open source software for rapid proteomics tools development. **Bioinformatics**, v. 24, n. 21, p. 2534-2536, 2008.

KIND, T.; FIEHN, O. Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. **BMC Bioinformatics**, v. 8, 2007. doi: 10.1186/1471-2105-8-105.

KLEIN, R. M.; SLEUMER, H. O. Flacourtiáceas. In: REITZ, R. (Ed.). **Flora ilustrada catarinense**. Itajaí: Herbário Barbosa Rodrigues, 1984. p. 1-96.

LAEMMLI, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. **Nature**, v. 227, n. 5259, p. 680-685, 1970.

LEEPER, F. J.; VEDERAS, J. C. (Ed.). **Biosynthesis**: aromatic polyketides, isoprenoids, alkaloids. Heidelberg: Springer, 2010. 247 p.

LEITE, F. **Validação em análise química**. 4. ed. Campinas: Átomo, 2002. 278 p.

LICHTENTHALER, H. K. Biosynthesis and accumulation of isoprenoid carotenoids and chlorophylls and emission of isoprene by leaf chloroplasts. **Bulletin of the Georgian National Academy of Sciences**, v. 3, n. 3, p. 81-94, 2009.

LORENZI, H. **Árvores brasileiras**: manual de identificação e cultivo de plantas arbóreas nativas do Brasil. Nova Odessa: Plantarum, 1992. v. 1, 532 p.

- LORENZI, H. **Árvores brasileiras**: manual de identificação e cultivo de plantas arbóreas do Brasil. 4. ed. Nova Odessa: Instituto Plantarum, 2002. v. 1, 368 p.
- MA, B.; ZHANG K.; HENDRIE, C.; LIANG, C.; LI, M. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. **Rapid Communications in Mass Spectrometry**, v. 17, n. 20, p. 2337-2342, 2003.
- MAISTRO, E. L.; CARVALHO J. C.; MANTOVANI, M. S. Evaluation of the genotoxic potential of the *Casearia sylvestris* extract on HTC and V79 cells by the comet assay. **Toxicology in Vitro**, v. 18, n. 3, p. 337-342, 2004.
- MARKHAM, K. R. **Techniques of flavonoid identification**. Londres: Academic Press, 1982. 113 p.
- MARQUETE, R. **O gênero Casearia Jacq. no Brasil**. 2010. 438 f. Tese (Doutorado em Botânica) - Escola Nacional de Botânica Tropical, Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Rio de Janeiro, 2010.
- MATROS, A.; KASPAR, R.; WITZEL, K.; MOCK, H. P. Recent progress in liquid chromatography-based separation and label-free quantitative plant proteomics. **Phytochemistry**, v. 72, n. 10, p. 963-974, 2011.
- MOORE, B. D.; ANDREW, R. L.; KÜLHEIM, C.; FOLEY, W. Explaining intraspecific diversity in plant secondary metabolites in an ecological context. **New Phytologist**, v. 201, p. 733-750, 2014.
- MORGENTHAL, K.; WECKWERTH, W.; STEUER, R. Metabolomic networks in plants: transitions from pattern recognition to biological interpretation. **Biosystems**, v. 83, n. 2/3, p. 108-117, 2006.
- MOSADDIK, M. A.; FORSTER, P. I.; BOOTH, R.; WATERMANN, P. G. New clerodane and halimane diterpenes from the leaves and woody stems of *Casearia grayi* (Flacourtiaceae/Salicaceae). **Natural Products Communications**, v. 1, n. 6, p. 441-448, 2006.
- MOSADDIK, M. A.; FORSTER, P. I.; BOOTH, R.; WATERMANN, P. G. Phenolic glycosides from some australian species of Flacourtiaceae (Salicaceae *sensu lato*). **Biochemical Systematics and Ecology**, v. 35, n. 3, p. 166-168, 2007.
- NATARAJAN, S. S.; XU, C.; CREGAN, P.; CAPERNA, T. J.; GARRETT, W. M.; LUTHRIA, D. Utility of proteomics techniques for assessing protein expression. **Regulatory Toxicology and Pharmacology**, v. 54, n. 3, p. S32-S36, 2009.
- NELSON, D. L.; COX, M. M. **Lehninger**: principles of biochemistry. 5th ed. New York: Freeman, 2008. 1100 p.
- NELSON, N.; YOCUM, C. F. Structure and function of photosystems I and II. **Annual Review of Plant Biology**, v. 57, p. 521-565, 2006.

OLIVEIRA, A. M.; SANTOS, A. G.; SANTOS, R. A.; CSIPAK, A. R.; OLIVATO, C.; SILVA, I. C.; FREITAS, M. B.; BASSI, C. L.; CAVALHEIRO, A. J.; BOLZANI, V. S.; SILVA, D. H. S.; SAKAMOTO-HOJO, E. T.; TAKAHASHI, C. S.; SOARES, C. P. Ethanolic extract of *Casearia sylvestris* and its clerodane diterpen (caseargrewiin f) protect against dna damage at low concentrations and cause dna damage at high concentrations in mice's blood cells. **Mutagenesis**, v. 24, n. 6, p. 501-506, 2009.

OLIVEIRA, R. B.; GODOY, S. A. P.; COSTA, F. B. da. **Plantas tóxicas**: conhecimento e prevenção de acidentes. Ribeirão Preto: Holos, 2003. 64 p.

PASSARELI, F. **Preparação e caracterização de casearinas de *Casearia sylvestris* como padrão fitoquímico**. 2010. 183 f. Dissertação (Mestrado em Química) - Instituto de Química, Universidade Estadual Paulista, Araraquara, 2008.

PATTI, G. J.; YANES, O.; SIUZDAK, G. Metabolomics: the apogee of the omic trilogy. **Nature Reviews Molecular Cell Biology**, v. 13, n. 4, p. 263-269, 2012.

PLUSKAL, T.; CASTILLO, S.; VILLAR-BRIONES, A.; ORESIC, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. **BMC Bioinformatics**, v. 11, 2010.
doi:10.1186/1471-2105-11-395.

RAMBALDI, D. M.; OLIVEIRA, D. A. S. (Org.). **Fragmentação de ecossistemas**: causas, efeitos sobre a biodiversidade e recomendações de políticas públicas. Brasília/DF: MMA/SBF, 2003. 510 p.

RAPPSILBER, J.; MANN, M.; ISHIHAMA, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. **Nature Protocols**, v. 2, n. 8, p. 1896-1906, 2007.

RASLAN, D. S.; JAMAL, C. M.; DUARTE, D. S.; BORGES, M. H.; LIMA, M. E. de. Anti-PLA2 action test of *Casearia sylvestris* Sw. **Bollettino Chimico Farmacêutico**, v. 141, n. 6, p. 457-460, 2002.

RIBANI, M.; BOTTOLI, C. B. G.; COLLINS, C. H.; JARDIM, I. C. S. F.; MELO, L. F. C. Validação em métodos cromatográficos e eletroforéticos. **Química Nova**, v. 27, n. 5, p. 771-780, 2004.

RUBAN, A. V. Plants in light. **Communicative & Integrative Biology**, v. 2, n. 1, p. 50-55, 2009.

SANTOS, A. G. A. **Identificação dos princípios ativos antiulcerogênicos das folhas de *Casearia sylvestris***: contribuição para o desenvolvimento de um fitoterápico. 2008. 361 f. Tese (Doutorado em Química) - Instituto de Química, Universidade Estadual Paulista, Araraquara, 2008.

SANTOS, A. G.; PEREZ, C. C.; TININIS, A. G.; BOLZANI, V. S.; CAVALHEIRO, A. J. Clerodane diterpenes from leaves of *Casearia sylvestris* SWARTZ. **Química Nova**, v. 30, n. 5, p. 1100-1103, 2007.

- SANTOS, A. G.; FERREIRA, P. M. P.; VIEIRA JÚNIOR, G. M.; PEREZ, C. C.; TININIS, A. G.; SILVA, G. H.; BOLZANI, V. S.; COSTA-LOTUFO, L. V.; PESSOA, C.; CAVALHEIRO, A. J. Casearin X, its degradation product and other clerodane diterpenes from leaves of *Casearia sylvestris*: evaluation of cytotoxicity against normal and tumour human cells. **Chemistry and Biodiversity**, v. 7, n. 1, p. 205-215, 2010.
- SAPAN, C. V.; LUNDBLAD, R. L.; PRICE, N. C. Colorimetric protein assay techniques. **Biotechnology and Applied Biochemistry**, v. 29, n. 2, p. 99-108, 1999.
- SARASWATHY, N.; RAMALINGAM, P. **Concepts and techniques in genomics and proteomics**. New York: Biohealthcare Publishing, 2011. 268 p.
- SCHAFFNER, W.; WEISSMANN, C. A rapid, sensitive, and specific method for the determination of protein in dilute solution. **Analytical Biochemistry**, v. 56, n. 2, p. 502-514, 1973.
- SEEMAN, J. R. Light adaptation/acclimation of photosynthesis and the regulation of ribulose-1,5-bisphosphate carboxylase activity in sun and shade plants. **Plant Physiology**, v. 91, n. 1, p. 379-386, 1989.
- SHEN, C.; CHANG, Y.; HOTS, L. Nuclear magnetic resonance studies of 5,7-dihydroxyflavonoids. **Phytochemistry**, v. 37, n. 3, p. 843-845, 1993.
- SHEORAN, I. S.; ROSS, A. R. S.; OLSON, D. J. H.; SAWHNEY, V. K. Compatibility of plant protein extraction methods with mass spectrometry for proteome analysis. **Plant Science**, v. 176, n. 1, p. 99-104, 2009.
- SILVA, I. C.; REGASINI, L. O.; PETRONIO, M. S.; SILVA, D. H. S.; BOLZANI, V. S.; BELASQUE, J. R. J.; SACRAMENTO, L. V. S.; FERREIRA, H. Antibacterial activity of alkyl gallates against *Xanthomonas citri* subsp. **Journal of Bacteriology**, v. 195, n. 1, p. 85-94, 2013.
- SILVA, M. A. S. da. **Variabilidade genética e fitoquímica de *Casearia sylvestris* Sw em populações do cerrado e mata atlântica do Estado de São Paulo**. 2003. 110 f. Tese (Doutorado em Agronomia) - Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista, Botucatu, 2003.
- SILVA, S. L. da; CALGAROTTO, A. K.; CHAAR, J.; MARANGONI, S. Isolation and characterization of ellagic acid derivatives isolated from *Casearia sylvestris* SW aqueous extract with anti-PLA2 activity. **Toxicon**, v. 52, n. 6, p. 655-666, 2008.
- SILVA, S. L. da; CHAAR, J. S.; YANO, T. Chemotherapeutic potential of two gallic acid derivative compounds from leaves of *Casearia sylvestris* Sw. (Flacourtiaceae). **European Journal of Pharmacology**, v. 608, n. 1/3, p. 76-83, 2009.
- SLEUMER, H. O. **Flacourtiaceae**. New York: The New York Botanic Garden, 1980. 499 p. (Flora neotropica monograph, n. 22).
- SMITH, P. K.; KROHN, R. I.; HERMANSON, G. T.; MALLIA, A. K.; GARTNER, F. H.; FROVENZANO, M. D.; FUJIMOTO, E. K.; GOEKE, N. M.; OLSON, B. J.; KLENK, D. C. Measurement of protein using bicinchoninic acid. **Analytical Biochemistry**, v. 150, n. 1, p. 76-85, 1985.

- SPECHT, M.; KUHLGERT, S.; FUFEZAN, C.; HIPPLER, M. Proteomics to go: proteomatic enables the user-friendly creation of versatile MS/MS data evaluation workflows. **Bioinformatics**, v. 27, n. 8, p. 1183-1184, 2011.
- SUMNER, L. W.; MENDES, P.; DIXON, R. A. Plant metabolomics: large-scale phytochemistry in the functional genomics era. **Phytochemistry**, v. 62, n. 6, p. 817-836, 2003.
- TORRES, R. B.; YAMAMOTO, K. Taxonomia das espécies de *Casearia* Jacq. (Flacourtiaceae) do Estado de São Paulo. **Revista Brasileira de Botânica**, v. 9, n. 2, p. 239-258, 1986.
- VERPOORTE, R.; CHOI, Y. H.; KIM, H. K. Ethnopharmacology and systems biology: a perfect holistic match. **Journal of Ethnopharmacology**, v. 100, n. 1/2, p. 53-56, 2005.
- WANG, W.; SCALI, M.; VIGNANI, R.; SPADAFORA, A.; SENSI, E.; MAZZUCA, S.; CRESTI, M. Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. **Electrophoresis**, v. 24, n. 14, p. 2369-2375, 2003.
- WANG, W.; VIGNANI, R.; SCALI, M.; CRESTI, M. A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. **Electrophoresis**, v. 27, n. 13, p. 2782-2786, 2006.
- WANG, W.; ALI, Z.; LI, X.; SMILLIE, T. A.; GUO, D.; KHAN, I. A. New clerodane diterpenoids from *Casearia sylvestris*. **Fitoterapia**, v. 80, n. 7, p. 404-407, 2009.
- WECKWERTH, W. Green systems biology - from single genomes, proteomes and metabolomes to ecosystems research and biotechnology. **Journal of Proteomics**, v. 75, n. 1, p. 285-301, 2011.
- WEISS, K. C.; BISSON, L. F. Optimisation of the amido black assay for determination of the protein content of grape juices and wines. **Journal of the Science of Food and Agriculture**, v. 81, n. 6, p. 583-589, 2001.
- WISNIEWSKI, J. R.; ZOUGMAN, A.; MANN, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. **Journal of Proteome Research**, v. 8, n. 12, p. 5674-5678, 2009.
- WISNIEWSKI, J. R.; ZOUGMAN, A.; NAGARAJ, N.; MANN, M. Universal sample preparation method for proteome analysis. **Nature Methods**, v. 6, n. 5, p. 359-363, 2009.
- WOLFENDER, J.; MARTI, G.; QUEIROZ, E. F. Advances in techniques for profiling crude extracts and for the rapid identification of natural products: dereplication, quality control and metabolomics. **Current Organic Chemistry**, v. 14, n. 16, p. 1808-1832, 2010.
- XIA, L.; GUO, Q.; TU, P.; CHAI, X. The genus *Casearia*: a phytochemical and pharmacological overview. **Phytochemistry Reviews**, v. 14, p. 99-135, 2015.

YANG, D.; DU, X.; YANG, Z.; LIANG, Z.; GUO, Z.; LIU, Y. Transcriptomics, proteomics, and metabolomics to reveal mechanisms underlying plant secondary metabolism. **Engineering in Life Sciences**, v. 14, n. 5, p. 456-466, 2014.

YOUNG, A.; BOYLE, T.; BROWN, T. The population genetic consequences of habitat fragmentation for plants. **Trends in Ecology and Evolution**, v. 11, n. 10, p. 413-418, 1996.

ZAIA, D. A. M.; ZAIA, C. T. B. V.; LICHTIG, J. Determinação de proteínas totais via espectrofotometria: vantagens e desvantagens dos métodos existentes. **Química Nova**, v. 21, n. 6, p. 786-793, 1998.

ZHOU, H.; TANG, W.; ZENG, J.; TANG, C. Screening of terpene lactones and flavonoid glycosides in *Ginkgo biloba* capsule by UPLC-Orbitrap high resolution MS, with emphasis on isomer differentiation. **Journal of Food and Nutrition Research**, v. 2, n. 7, p. 369-376, 2014.

APÊNDICES

APÊNDICE I

Procedimentos de preparo de soluções para o estudo proteômico

A) Extração de proteínas

Acetona 80% em água: para um balão volumétrico de 250 mL, transferiu-se 200 mL de acetona e completou-se o volume com água bidestilada.

Fenol/tampão SDS 1:1: Para um frasco com capacidade de 250 mL, transferiu-se 100 mL do tampão SDS e 100 mL de fenol pH 8,0 Sigma. Homogeneizou-se imediatamente antes do uso.

Metanol 80% em água contendo 0,1M de acetato de amônio: para um balão volumétrico de 500 mL contendo cerca de 50 mL de água bidestilada transferiu-se e dissolveu-se 3,75 g de acetato de amônio. Em seguida adicionou-se 400 mL de metanol e completou-se o volume com água bidestilada.

Tampão SDS: para um balão volumétrico de 100 mL contendo cerca de 30 mL de tampão Tris-HCl 0,1M, pH 8,0, transferiu-se 30 g de sucrose, 2 g de dodecilsulfato de sódio (SDS, MM = 288,37 g/mol) e 5 mL de 2-mercaptoetanol (MM = 78,13 g/mol). Homogeneizou-se cuidadosamente para evitar formação de espuma de completou-se o volume com o tampão Tris-HCl.

Tampão Tris-HCl 0,1 M, pH 8,0: para um balão volumétrico de 100 mL contendo cerca de 50 mL de água bidestilada, transferiu-se e homogeneizou-se 1,21 g de 2-Amino-2-hydroxymethyl-propane-1,3-diol (Tris, MM = 121,14 g/mol) e corrigiu-se o pH para 8,0 com cerca de 0,4 mL de ácido clorídrico concentrado. Completou-se o volume com água bidestilada.

TCA 10% em acetona: para um balão volumétrico de 250 mL, transferiram-se 200 mL de acetona e 25 mL de ácido tricloroacético (TCA, MM = 163.4 g/mol). Homogeneizou-se e completou-se o volume com acetona.

B) FASP e Pós-FASP

DTT 0,1M: para um tubo Eppendorf transferiu-se 0,015 g de ditioneitol (DTT, 154,25 g/mol) e dissolveu-se em 1 mL de tampão UA.

Hepes 100 mM, pH 6,5: Para um tubo falcon de 50 mL, transferiu-se 0,477 g de ácido 2-[4-(2-hidroxietil)1-piperazinil]-etanossulfônico (HEPES, MM = 238,31 g/mol) e dissolveu-se com 10 mL de água bidestilada. Ajustou-se o pH com NaOH 1M e completou-se o volume para 20 mL com água bidestilada. Esta solução tem que ser preparada no dia do uso.

NaOH 1M: Para um tubo falcon de 15 mL, transferiu-se 0,400 g de hidróxido de sódio (NaOH, MM= 40,01 g/mol) e dissolveu-se com 10 mL de água bidestilada.

Solução ABC 50 mM: Para um tubo falcon de 15 mL, transferiu-se 0,0198 g de bicarbonato de amônio (ABC, MM=79,06 g/mol) e completou-se 5 mL com água bidestilada.

Solução estoque de tripsina 0,5 µg/µL: para o vial do Kit Pierce Trypsin Protease MS Grade (Thermo Scientific), contendo 20 µg de tripsina, adicionou-se 40 µL de tampão de ressuspensão e homogeneizou-se em vortex. O procedimento foi feito em banho de gelo.

Tampão Britton & Robinson – solução de trabalho: para cada solução com determinado pH, foi feita uma diluição de 1:5.

Tampão Britton & Robinson, diversos pHs – solução estoque 5X: Primeiramente preparou-se 600 mL de uma solução contendo 20 mM ácido acético (MM= 60,05 g/mol; 6,86 mL), 20 mM ácido fosfórico (MM=98 g/mol; 8,21 mL), 20 mM ácido bórico (MM=61,83 g/mol; 7,416 g) e cerca de 585 mL de água bidestilada. Em seguida, fracionou-se esta solução transferindo-se 100 mL para 6 diferentes frascos. Em seguida, cada frasco teve a solução titulada com NaOH 1 M até os pHs 3, 4, 5, 6, 8 e 11. Para o pH 3, adicionou-se 10 mL de uma solução de NaCl 5M (MM= 58,44 g/mol; perfazendo-se no final 0,25M). Finalmente, completou-se o volume para 200 mL com água destilada. Todas as soluções foram alíquotadas em tubos Eppendorf de 2 mL e congeladas em freezer -20 °C.

Tampão IAA: Para um tubo Eppendorf de 1,5 mL, transferiu-se 0,0094 g de 2-iodoacetamida (IAA, MM=184,96 g/mol) e adicionou-se 1 mL de tampão UA. Esta solução tem que ser preparada no dia do uso.

Tampão UA: Para um tubo falcon de 15 mL, transferiu-se 1 mL da solução HEPES 100 mM pH 6,5, 4,8 g de uréia (UA, MM=60,06 g/mol) e 1,52 g de tiouréia (MM=76,12 g/mol) e dissolveu-se com água bidestilada completando um volume de 10 mL. Esta solução tem que ser preparada no dia do uso.

Tripsina 0,05 µg/µL: Para um tubo Eppendorf de 0,5 mL, transferiu-se 24 µL da solução estoque de tripsina e 216 µL de solução ABC 50 mM. Homogeneizou-se. Esta quantidade é suficiente para a digestão de 6 amostras.

C) Quantificação por BCA e Amido Black

Ácido acético/etanol 1:10: para um balão volumétrico de 10 mL contendo já 5 mL de etanol, transferiu-se 1 mL de ácido acético e completou-se o volume com etanol.

Amido black a 260 µg/mL: para um tubo Eppendorf de 1,5 mL, transferiu-se 0,260 mg de amido black e dissolveu-se com 1 mL de água bidestilada.

BSA a 2,0 µg/µL: para um tubo Eppendorf de 1,5 mL, transferiu-se 2 mg de albumina de soro bovino (BSA) e dissolveu-se com 1 mL de água bidestilada.

NaOH 0,1N: Para um tubo falcon de 15 mL, transferiu-se 0,040 g de hidróxido de sódio (NaOH, MM= 40,01 g/mol) e dissolveu-se com 10 mL de água bidestilada.

D) Soluções para SDS-PAGE e *Western blotting*

3-Aminoftahidrazina (luminol) 250 mM: para um tubo Eppendorf de 1,5 mL, transferiu-se 44 mg de 3-aminoftahidrazina e dissolveu-se com 1 mL de DMSO.

Ácido p-cumárico 90 mM: para um tubo Eppendorf de 1,5 mL, transferiu-se 14 mg de ácido p-cumárico e dissolveu-se com 1 mL de DMSO.

APS 50%: Para um balão volumétrico de 10 mL, transferiu-se 5 g de persulfato de amônio (APS, MM = 228,18 g/mol) e dissolveu-se, completando o volume com água bidestilada.

Comassie Blue: Para um balão volumétrico de 250 mL, transferiu-se 0,25 g de Coomassie Brilliant Blue, 100 mL de etanol e 25 mL de ácido acético e completou-se o volume com água bidestilada.

EDTA 0,5 M pH 8,0: para um balão volumétrico de 100 mL, transferiu-se 14,61 g de ácido etilenodiamino tetra-acético (EDTA, MM = 292.2 g/mol) e dissolveu-se com cerca de 50 mL de água bidestilada. Corrigiu-se o pH com NaOH 1M e completou-se o volume com água bidestilada.

Gel de poliacrilamida: Cada gel de separação (13% de poliacrilamida) de 6 cm de altura foi elaborado com 2,15 mL de solução de acrilamida 30% (29:1 acrilamida/bisacrilamida), 1,25 mL de tampão tris 1,5 M contendo 0,4% de SDS, pH 8,8, 0,815 mL de água bidestilada, 20,2 µL de EDTA 0,5 M pH 8,0, 0,72 mL de sacarose 2M, 1,3 µL de tetrametiletenodiamina (TEMED) e 12,25 µL de persulfato de amônio 50% (APS). Já o gel de empilhamento de 1 cm conteve 0,3 mL de acrilamida 30%, 0,44 mL de tampão tris 0,5 M contendo 0,4% de SDS, pH 6,8, 0,975 mL de água bidestilada, 7 µL de EDTA 0,5 M pH 8,8, 1,75 µL de TEMED e 3,52 µL de APS 50% (Laemmli, 1970).

Reagente ECL – Solução A: em um tubo Falcon de 15 mL transferiu-se e homogeneizou-se 13,3 µL da solução de ácido p-cumárico 90 mM, 3 mL de tampão Tris-HCl 0,1 M, pH 8,5 e 1,66 µL de peróxido de hidrogênio (30%).

Reagente ECL – Solução B: em um tubo Falcon de 15 mL transferiu-se e homogeneizou-se 30 µL da solução de 3-Aminoftahidrazina (luminol) 250 mM e 3 mL de tampão Tris-HCl 0,1 M, pH 8,5.

Reagente ECL: imediatamente antes do uso, misturou-se as soluções A e B.

Sacarose 2M: Para um balão volumétrico de 500 mL, transferiu-se 342,30 g de sacarose (MM = 342,30 g/mol) e dissolveu-se, completando o volume com água bidestilada.

Solução Leite/PBS: para um frasco com capacidade de 1000 mL, transferiu-se e homogeneizou-se 20 g de leite reconstituído em pó, 100 mL de tampão PBS 10X e completou-se o volume com água bidestilada.

Tampão de corrida SDS-PAGE: a partir do tampão-corrida estoque SDS-PAGE 10X, fez-se uma diluição 1:10, transferindo-se 100 mL da solução 10X concentrada para uma proveta de 1000 mL e completou-se o volume com água bidestilada.

Tampão H6: Para um balão volumétrico de 100 mL, transferiu-se 0,119 g de ácido 2-[4-(2-hidroxietil)1-piperazinil]-etanosulfônico (HEPES, MM = 238,31 g/mol) e dissolveu-se com 20 mL de água bidestilada. Ajustou-se o pH para 7,5 com NaOH 1M. Adicionou-se 0,2922 g de EDTA (MM = 292.2 g/mol) e completou-se o volume para 100 mL com água bidestilada.

Tampão PBS 10X: Para um balão volumétrico de 1000 mL transferiu-se 80,1 g de cloreto de sódio (NaCl), 2 g de fosfato de potássio monobásico (KH₂PO₄, MM = 136,086 g/mol), 11,2 g de fosfato de sódio bibásico heptahidratado (Na₂HPO₄.7H₂O, MM = 268.07 g/mol), 2 g de cloreto de potássio (KCl, MM = 74,551 g/mol) e completou-se o volume com água destilada.

Tampão Tris-HCl 0,1 M, pH 8,5: para um balão volumétrico de 100 mL contendo cerca de 50 mL de água bidestilada, transferiu-se e homogeneizou-se 1,21 g de 2-Amino-2-hydroxymethyl-propane-1,3-diol (Tris, MM = 121,14 g/mol) e corrigiu-se o pH para 8,5 com ácido clorídrico concentrado. Completou-se o volume com água bidestilada.

Tampão Tris-HCl 0,5M, pH 6,8: para um balão volumétrico de 500 mL contendo cerca de 100 mL de água bidestilada, transferiu-se e homogeneizou-se 30,29 g de 2-Amino-2-hydroxymethyl-propane-1,3-diol (Tris, MM = 121,14 g/mol) e 2 g de dodecilsulfato de sódio (SDS, MM = 288,37 g/mol). Corrigiu-se o pH para 6,8 com de ácido clorídrico diluído. Completou-se o volume com água bidestilada.

Tampão Tris-HCl 1,5M, pH 8,8: para um balão volumétrico de 500 mL contendo cerca de 100 mL de água bidestilada, transferiu-se e homogeneizou-se 90,86 g de 2-Amino-2-hydroxymethyl-propane-1,3-diol (Tris, MM = 121,14 g/mol) e 2 g de dodecilsulfato de sódio (SDS, MM = 288,37 g/mol). Corrigiu-se o pH para 8,8 com de ácido clorídrico diluído. Completou-se o volume com água bidestilada.

Tampão-amostra SDS-PAGE 10X: para um balão volumétrico de 250 mL, transferiu-se 25 mL de glicerina, 12,5 mL de β -mercaptoetanol, 7,75 g de dodecilfulfato de sódio (SDS), 31,25 mL de tris-HCl 0,5M, pH 6,8 e 0,0025 g de azul de bromofenol. Homogeneizou-se.

Tampão-corrida estoque SDS-PAGE 10X (Laemmli, 1973): para um balão volumétrico de 1 L transferiu-se e dissolveu-se 10 g de dodecilsulfato de sódio (SDS) 30 g de 2-Amino-2-hydroxymethyl-propane-1,3-diol (Tris), 144 g de glicina e completou-se o volume com água bidestilada.

Tampão-corrida *Western blotting*: Para um balão volumétrico de 1000 mL, transferiu-se 100 mL de tampão-corrida estoque SDS-PAGE 10X, 200 mL de metanol e completou-se o volume com água bidestilada.

Vermelho de Ponceau: para um balão volumétrico de 500 mL, contendo cerca de 100 mL de água bidestilada, transferiu-se 0,5 g de *Ponceau* e 25 mL de ácido acético. Completou-se o volume com água bidestilada.

APÊNDICE II

Protocolo de rotina para preparo e análise das amostras por UHPLC-DAD

A) Preparo do solvente extrator contendo o padrão interno galato de butila a 0,5 mg/mL: O solvente extrator, definido através do triângulo de seletividade, foi composto de etanol, isopropanol e água na proporção de 3:2:5, respectivamente, sendo que o padrão interno é adicionado nesta mistura. Para tanto, primeiro preparou-se esta mistura em frasco separado (pois há retração de volume na mistura) e com esta preparou-se a solução final contendo o padrão interno (adicionando-se 5 mg de galato de butila para cada 10 mL de solvente extrator). Homogeneizou-se muito bem e manteve-se o balão volumétrico fechado.

B) Extração das amostras: para um tubo Eppendorf de 1,5 mL, transferiu-se 50 mg das folhas secas finamente trituradas e adicionou-se 1 mL do solvente extrator. Tampou-se o tubo e extraiu-se em banho de ultrasson por 1h, agitando-se manualmente a cada 20 minutos. Terminada a extração, centrifugou-se os tubos por 5 minutos a 5.000 x g e com o auxílio de uma pipeta automática, filtrou-se uma alíquota de 0,7 mL do sobrenadante em filtro de nylon (0,22 μm x 15 mm de diâmetro) diretamente para o frasco de 1,5 mL.

C) Análise cromatográfica: As amostras foram analisadas por cromatografia líquida acoplada à detecção por arranjo de diodos (UHPLC-DAD) utilizando como eluentes água (A) e acetonitrila (B). Um volume de 2 μL de amostra foi injetado numa coluna C18 de núcleo sólido (Kinetex C18, 150 x 2,1 mm, 2,6 μm , 100 Å) utilizando o seguinte gradiente: 10 a 25% de B por 15 minutos e em seguida de 25 a 90% de B por 20 minutos, permanecendo nesta condição por 5 minutos. Um método para recondicionamento foi incorporado de forma que se voltasse à condição inicial em 2 minutos e permanecesse nesta por mais 3 minutos, perfazendo um tempo total de 45 minutos. A vazão foi de 0,400 mL/min, temperatura da coluna de 35°C, faixa de detecção de 200-800 nm e plotagem dos cromatogramas em 235 nm/254 nm.

APÊNDICE III

Tabela usada para desreplicação de diterpenos clerodânicos em *Casearia*, por espectrometria de massas de alta resolução

Formula	MM (alta)	[M+H] ⁺	[M] ⁺	[M+Na] ⁺	Nome
C20H30O2	302.22458	303.23186	302.22403	325.21380	2-OXO-18-OH-Cleroda-3,13(16),14-trieno
C20H34O2	306.25588	307.26316	306.25533	329.24510	15-OH-3-Cleroden-2-ona
C20H28O3	316.20385	317.21112	316.20330	339.19307	Ác. Hardwickiico
C24H34O5	402.24063	403.24790	402.24008	425.22985	2-OXO-18,19-Diacetoxi-cleroda-3,13(16),14-trieno
C24H32O6	416.21989	417.22717	416.21934	439.20911	Esculentina A
C25H36O5	416.25628	417.26355	416.25573	439.24550	Caseamembrina F
C25H36O5	416.25628	417.26355	416.25573	439.24550	Caseamembrina J
C24H34O6	418.23554	419.24282	418.23499	441.22476	Caseargrewiina D
C24H36O6	420.25119	421.25847	420.25064	443.24041	Dt 13
C24H36O6	420.25119	421.25847	420.25064	443.24041	Dt 23
C24H36O6	420.25119	421.25847	420.25064	443.24041	Dt 26
C26H38O5	430.27193	431.27920	430.27138	453.26115	Dt 22
C24H32O7	432.21481	433.22208	432.21426	455.20403	Caseamembrina K
C24H34O7	434.23046	435.23773	434.22991	457.21968	Dt 24
C24H34O7	434.23046	435.23773	434.22991	457.21968	Dt 25
C26H38O7	462.26176	463.26903	462.26121	485.25098	Dt 14
C26H38O7	462.26176	463.26903	462.26121	485.25098	Dt 17
C27H38O7	474.26176	475.26903	474.26121	497.25098	Casearina S
C26H36O8	476.24102	477.24830	476.24047	499.23024	Casearinona A
C26H36O8	476.24102	477.24830	476.24047	499.23024	Casearinona B
C26H36O8	476.24102	477.24830	476.24047	499.23024	Corimbotina D
C26H36O8	476.24102	477.24830	476.24047	499.23024	Dt 16
C27H40O7	476.27741	477.28468	476.27686	499.26663	Caseargrewiina G
C25H34O9	478.22029	479.22756	478.21974	501.20951	Corimbotina G
C28H40O7	488.27741	489.28468	488.27686	511.26663	Caseargrewiina E
C27H38O8	490.25667	491.26395	490.25612	513.24589	Casearlucina B
C27H38O8	490.25667	491.26395	490.25612	513.24589	Corimbotina A
C28H42O7	490.29306	491.30033	490.29251	513.28228	Caseamembrina B
C28H42O7	490.29306	491.30033	490.29251	513.28228	Caseamembrina I
C28H42O7	490.29306	491.30033	490.29251	513.28228	Dt 19
C28H42O7	490.29306	491.30033	490.29251	513.28228	Dt 21
C28H38O8	502.25667	503.26395	502.25612	525.24589	Corimbotina F
C29H42O7	502.29306	503.30033	502.29251	525.28228	Casearlucina F
C29H42O7	502.29306	503.30033	502.29251	525.28228	Casearlucina G
C29H42O7	502.29306	503.30033	502.29251	525.28228	Dt 3
C27H36O9	504.23594	505.24321	504.23539	527.22516	Corimbotina I
C28H40O8	504.27232	505.27960	504.27177	527.26154	Caseamembrina G
C28H40O8	504.27232	505.27960	504.27177	527.26154	Caseanigrescena D
C28H40O8	504.27232	505.27960	504.27177	527.26154	Caseargrewiina F (Cas V

C28H40O8	504.27232	505.27960	504.27177	527.26154	Casearina H
C28H40O8	504.27232	505.27960	504.27177	527.26154	Casearvestrina A
C28H40O8	504.27232	505.27960	504.27177	527.26154	Corimbotina E
C28H40O8	504.27232	505.27960	504.27177	527.26154	Dt 9
C29H44O7	504.30871	505.31598	504.30816	527.29793	Caseargrewiina J
C29H44O7	504.30871	505.31598	504.30816	527.29793	Dt 15
C29H44O7	504.30871	505.31598	504.30816	527.29793	Dt 20
C27H38O9	506.25159	507.25886	506.25104	529.24081	Corimbotina H
C28H42O8	506.28797	507.29525	506.28742	529.27719	Casearina F
C29H40O8	516.27232	517.27960	516.27177	539.26154	Corimbotina C
C30H44O7	516.30871	517.31598	516.30816	539.29793	Caseargrewiina H
C30H44O7	516.30871	517.31598	516.30816	539.29793	Dt 12
C28H38O9	518.25159	519.25886	518.25104	541.24081	Casearina R
C29H42O8	518.28797	519.29525	518.28742	541.27719	Caseamembrina E
C29H42O8	518.28797	519.29525	518.28742	541.27719	Caseamembrol A
C29H42O8	518.28797	519.29525	518.28742	541.27719	Casearina G
C29H42O8	518.28797	519.29525	518.28742	541.27719	Casearinol A
C29H42O8	518.28797	519.29525	518.28742	541.27719	Casearlucina A
C29H42O8	518.28797	519.29525	518.28742	541.27719	Casearvestrina B
C29H42O8	518.28797	519.29525	518.28742	541.27719	Corimbotina B
C29H42O8	518.28797	519.29525	518.28742	541.27719	Dt 11
C29H42O8	518.28797	519.29525	518.28742	541.27719	Dt 18
C29H42O8	518.28797	519.29525	518.28742	541.27719	Dt 2
C29H42O8	518.28797	519.29525	518.28742	541.27719	Dt 8
C28H40O9	520.26724	521.27451	520.26669	543.25646	Caseanigrescena B
C29H44O8	520.30362	521.31090	520.30307	543.29284	Caseargrewiina K
C29H44O8	520.30362	521.31090	520.30307	543.29284	Casearinol B
C30H44O8	532.30362	533.31090	532.30307	555.29284	Caseamembrina H
C30H44O8	532.30362	533.31090	532.30307	555.29284	Caseargrewiina A
C30H44O8	532.30362	533.31090	532.30307	555.29284	Caseargrewiina I
C30H44O8	532.30362	533.31090	532.30307	555.29284	Casearina I
C30H44O8	532.30362	533.31090	532.30307	555.29284	Casearina U
C30H44O8	532.30362	533.31090	532.30307	555.29284	Casearlucina E
C30H44O8	532.30362	533.31090	532.30307	555.29284	Casearvestrina C
C30H44O8	532.30362	533.31090	532.30307	555.29284	Dt 1
C30H44O8	532.30362	533.31090	532.30307	555.29284	Dt 10
C29H42O9	534.28289	535.29016	534.28234	557.27211	Caseamembrol B
C29H42O9	534.28289	535.29016	534.28234	557.27211	Caseargrewiina L
C29H42O9	534.28289	535.29016	534.28234	557.27211	Casearina A
C29H42O9	534.28289	535.29016	534.28234	557.27211	Casearina L
C29H42O9	534.28289	535.29016	534.28234	557.27211	Casearlucina H
C29H42O9	534.28289	535.29016	534.28234	557.27211	Casearlucina I
C29H42O9	534.28289	535.29016	534.28234	557.27211	Casearlucina J
C29H42O9	534.28289	535.29016	534.28234	557.27211	Casearlucina K
C29H42O9	534.28289	535.29016	534.28234	557.27211	Esculentina B
C31H38O8	538.25667	539.26395	538.25612	561.24589	Casearborina A

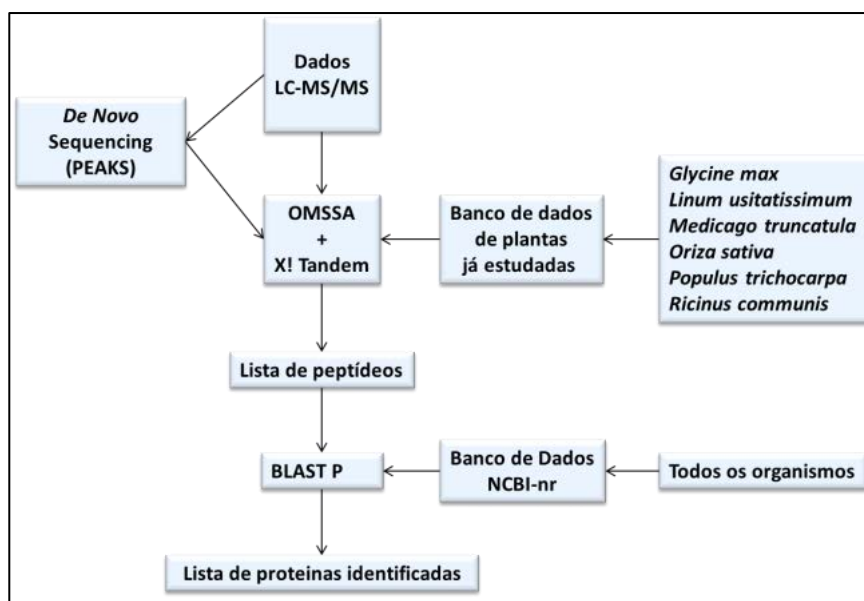
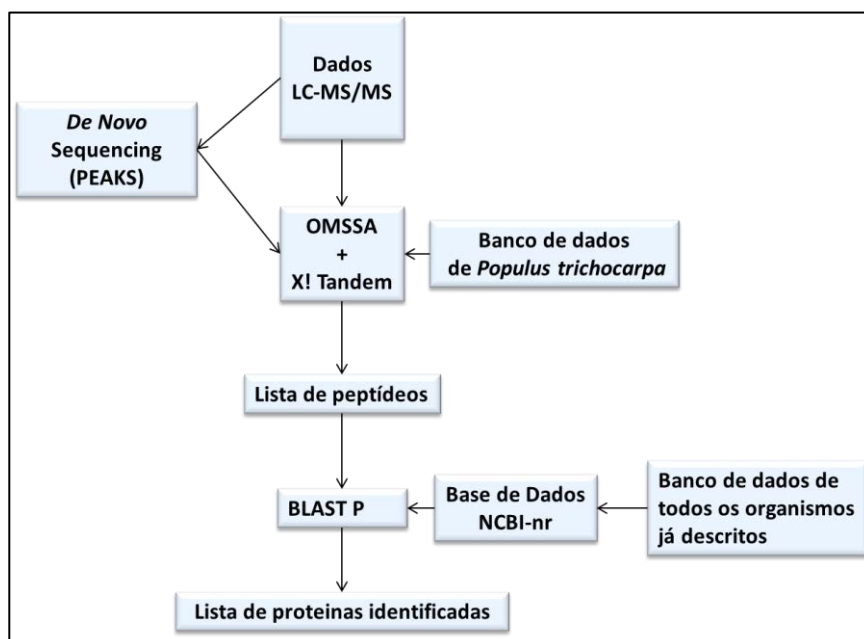
C31H46O8	546.31927	547.32655	546.31872	569.30849	Caseamembrina A
C31H46O8	546.31927	547.32655	546.31872	569.30849	Casearina Y
C31H46O8	546.31927	547.32655	546.31872	569.30849	Dt 7
C29H40O10	548.26215	549.26943	548.26160	571.25137	Casearina P
C30H44O9	548.29854	549.30581	548.29799	571.28776	Caseamembrina P
C30H44O9	548.29854	549.30581	548.29799	571.28776	Casearina D
C30H44O9	548.29854	549.30581	548.29799	571.28776	Casearina M
C31H38O9	554.25159	555.25886	554.25104	577.24081	Casearborina C
C31H38O9	554.25159	555.25886	554.25104	577.24081	Casearborina D
C31H44O9	560.29854	561.30581	560.29799	583.28776	Casearlucina C
C31H44O9	560.29854	561.30581	560.29799	583.28776	Casearlucina D
C32H48O8	560.33492	561.34220	560.33437	583.32414	Dt 4
C30H42O10	562.27780	563.28508	562.27725	585.26702	Caseamembrina N
C30H42O10	562.27780	563.28508	562.27725	585.26702	Caseanigrescena A
C30H42O10	562.27780	563.28508	562.27725	585.26702	Caseanigrescena C
C30H42O10	562.27780	563.28508	562.27725	585.26702	Casearina K
C30H42O10	562.27780	563.28508	562.27725	585.26702	Casearina Q
C31H46O9	562.31419	563.32146	562.31364	585.30341	Caseamembrina C
C31H46O9	562.31419	563.32146	562.31364	585.30341	Caseamembrina L
C31H46O9	562.31419	563.32146	562.31364	585.30341	Casearina J
C32H40O9	568.26724	569.27451	568.26669	591.25646	Casearborina B
C32H46O9	574.31419	575.32146	574.31364	597.30341	Casearina X
C31H44O10	576.29345	577.30073	576.29290	599.28267	Caseamembrina O
C31H44O10	576.29345	577.30073	576.29290	599.28267	Caseargrewiina C
C31H44O10	576.29345	577.30073	576.29290	599.28267	Casearina B
C31H44O10	576.29345	577.30073	576.29290	599.28267	Casearina T
C32H48O9	576.32984	577.33711	576.32929	599.31906	Dt 6
C34H48O8	584.33492	585.34220	584.33437	607.32414	Pitumbina
C33H48O9	588.32984	589.33711	588.32929	611.31906	Caseamembrina M
C32H46O10	590.30910	591.31638	590.30855	613.29832	Caseamembrina Q
C34H54O8	590.38187	591.38915	590.38132	613.37109	Casearina E
C33H40O10	596.26215	597.26943	596.26160	619.25137	Casearborina E
C35H54O8	602.38187	603.38915	602.38132	625.37109	Dt 5
C33H48O10	604.32475	605.33203	604.32420	627.31397	Caseamembrina D
C33H48O10	604.32475	605.33203	604.32420	627.31397	Caseamembrina R
C33H48O10	604.32475	605.33203	604.32420	627.31397	Casearina N
C33H48O10	604.32475	605.33203	604.32420	627.31397	Casearina O
C36H54O10	646.37170	647.37898	646.37115	669.36092	Casearina C
C37H46O10	650.30910	651.31638	650.30855	673.29832	Caseargrewiina B

Diterpeno	Nome
1	rel-18S,19R-diacetóxi-18,19-epóxi-2S-(2-metilbutanoiloxi)-6R-metoxi-5R,8S,9S,10R-cleroda-3,13(16),14-trieno
2	rel-18S,19R-diacetóxi-18,19-epóxi-2S-(2-metilbutanoiloxi)-6R-hidróxi-5R,8S,9S,10R-cleroda-3,13(16),14-trieno
3	rel-18S,19R-diacetóxi-18,19-epóxi-2S-(2-metilbutanoiloxi)-5R,8S,9S,10R-cleroda-3,13(16),14-trieno

4	rel-18S,19R-diacetóxi-18,19-epóxi-6R-hidróxi-2S-octanoiloxi-5R,8S,9S,10R-cleroda-3,13(16),14-trieno
5	rel-18S,19R-diacetóxi-18,19-epóxi-6R-hidróxi-2S-undecanoiloxi-5R,8S,9S,10R-cleroda-3,13(16),14-trieno
6	rel-18S,19R-diacetóxi-18,19-epóxi-6R-hidróxi-2S-(3-hidroxi-octanoiloxi)-5R,8S,9S,10R-cleroda-3,13(16),14-trieno
7	rel-19R-acetóxi-18S-butanoiloxi-18,19-epóxi-6R-hidróxi-2S-(2-metilbutanoilóxi)-5R,8S,9S,10R-cleroda-3,13(16),14-trieno
8	ent-6 β -hidroxii-sozuelanina-2b-(2-metil)butanoato
9	ent-6 β -hidroxii-sozuelanina-2b-(2-metil)propanoato
10	ent-6 β -metoxii-sozuelanina-2b-(2-metil)butanoato
11	ent-6 β -metoxii-sozuelanina-2b-(2-metil)propanoato
12	rel-18S,19R-diacetóxi-18,19-epóxi-2R-hexanoilóxi-5R,8S,9S,10R-cleroda-3,13(16),14-trieno
13	ent-2 β -hidróxi-3,4-di-hidro-4a-isozuelanina
14	ent-2 β -acetoxi-3,4-di-hidro-4a-isozuelanina
15	ent-2 β -(2-metilbutanoiloxi)-3,4-di-hidro-4a-isozuelanina
16	rel-2R,18S,19R-triacetóxi-18,19-epóxi-4S,5S,8S,9S,10R-clerodan-13(16),14-dien-6-ona
17	rel-2R,18S,19R-triacetóxi-18,19-epóxi-4S,5R,8S,9S,10R-clerodan-13(16),14-dieno
18	rel-18S,19R-diacetóxi-18,19-epóxi-2R-(2-metilbutanoilóxi)-4S,5S,8S,9S,10R-clerodan-13(16),14-dien-6-ona
19	rel-18S,19R-diacetóxi-18,19-epóxi-2R-(isobutanoilóxi)-4S,5R,8S,9S,10R-clerodan-13(16),14-dieno
20	rel-18S,19R-diacetóxi-18,19-epóxi-2R-(2-metilbutanoilóxi)-4S,5R,8S,9S,10R-clerodan-13(16),14-dieno
21	rel-18R-acetóxi-19S-metóxi-18,19-epóxi-6S-hidróxi-2S-(2-metilbutanoilóxi)-5S,8R,9R,10S-clerodan-3,13(16),14-trieno
22	rel-(2R,5R,6R,8S,9S,10R)-6-metóxi-2-(2-metilbutanoilóxi)-cleroda-3,13(16),14-trieno-4,5-dial
23	rel-2R-hidróxi-18S,19R-diacetóxi-18,19-epóxi-4S,5R,8S,9S,10R-clerodan-13(16),14-dieno
24	rel-2R-hidróxi-18S,19R-diacetóxi-18,19-epóxi-4S,5S,8S,9S,10R-clerodan-13(16),14-dien-6-ona
25	rel-2R-hidróxi-18S,19R-diacetóxi-18,19-epóxi-4S,5S,8S,9S,10R-clerodan-12E,14-dien-6-ona
26	rel-2R-hidróxi-18S,19R-diacetóxi-18,19-epóxi-4S,5R,8S,9S,10R-clerodan-12E,14-dieno

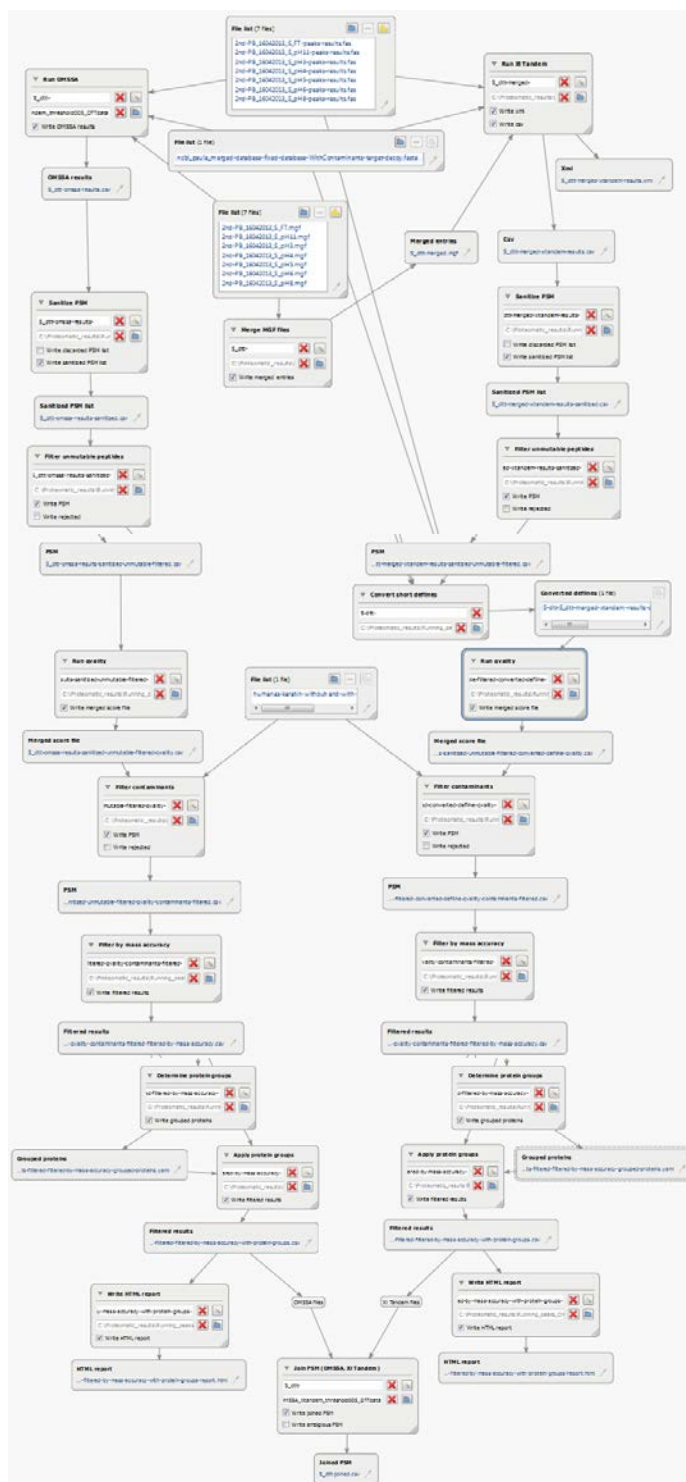
*Referência dos dados: SANTOS, 2008.

Apêndice IV

Fluxograma resumindo as estratégias utilizadas em bioinformática para a identificação das proteínas nas duas variedades de *C. sylvestris*A) **Estratégia 1:** Utilização de banco de dados misto para a identificação de proteínasB) **Estratégia 2:** Utilização de banco de dados de *Populus trichocarpa* para a identificação de proteínas

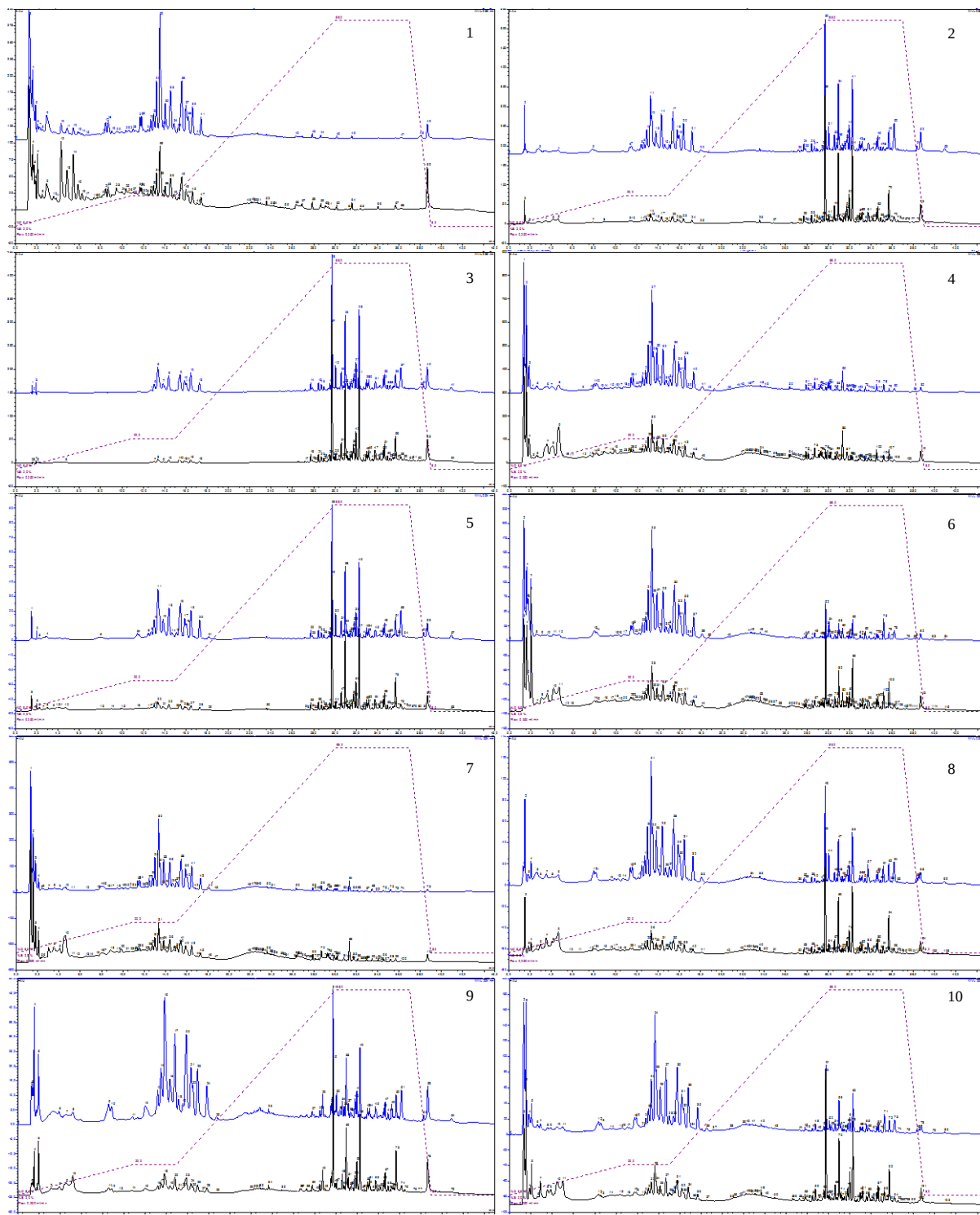
Apêndice V

Interface gráfica do software Proteomatic® mostrando a estratégia final para a identificação das proteínas

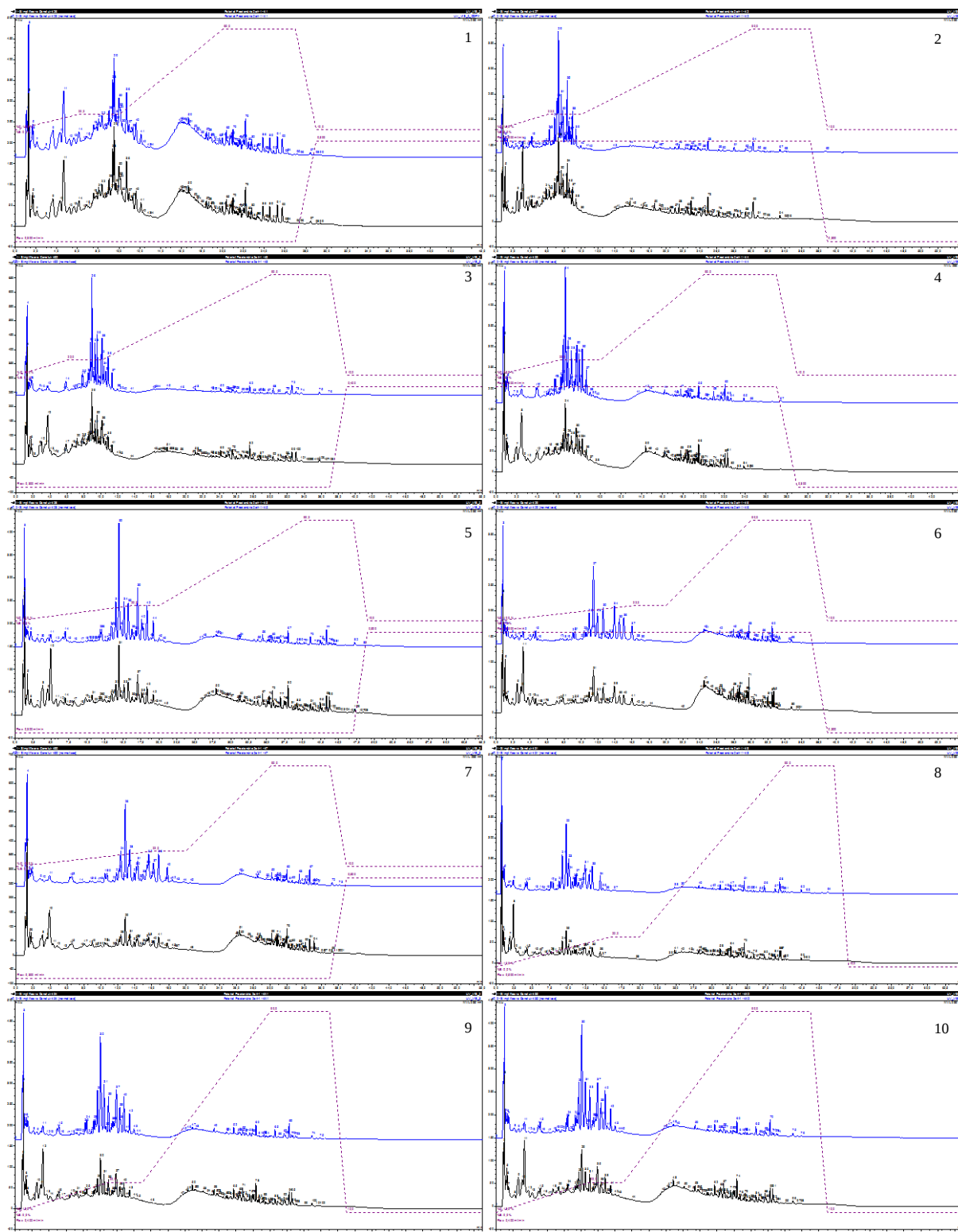


APÊNDICE VI

Cromatogramas do planejamento Simplex Centróide

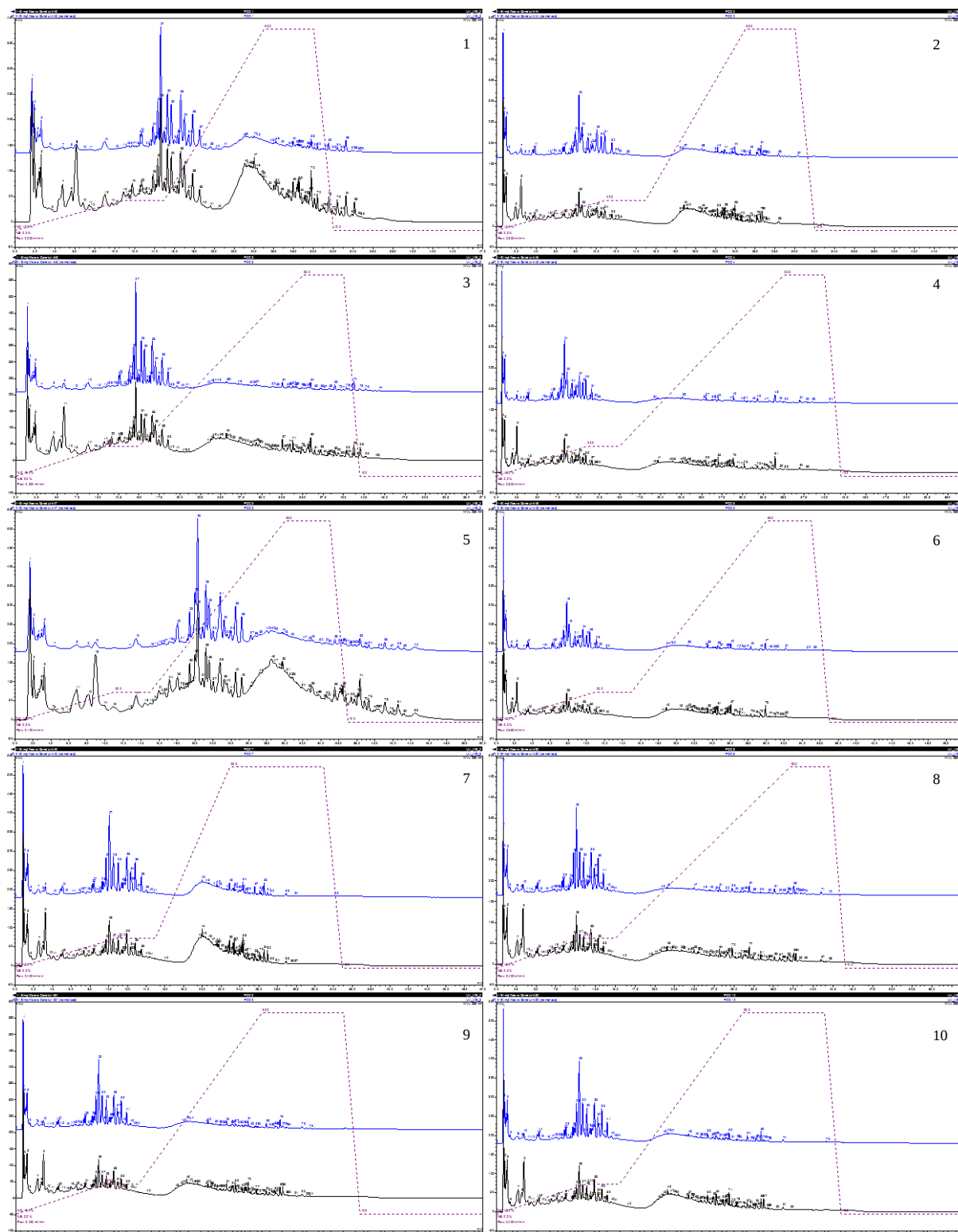


APÊNDICE VII

Cromatogramas do Planejamento Fatorial Fracionário 2^{4-1} : triagem de variáveis

APÊNDICE VIII

Cromatogramas do Planejamento Composto Central



APÊNDICE IX

Espectros de massas dos padrões de casearinas isolados

A) Casearina X

Generic Display Report

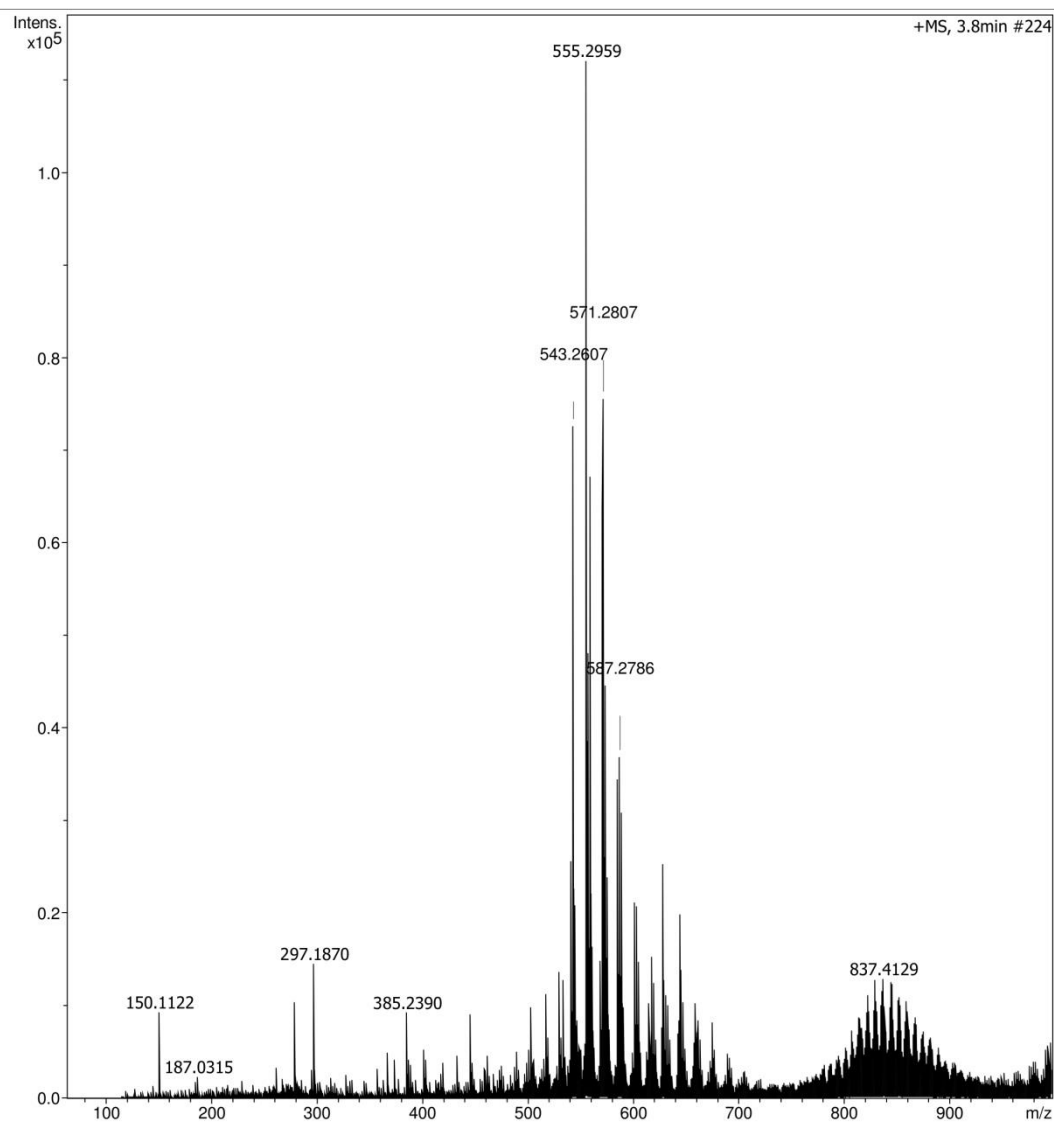
Faculdade de Ciencias Farmaceuticas de Ribeirao Preto-USP

Analysis Info

Analysis Name D:\Data\Paulinha_UNESP\CASEARINA_X_POS.d
Method Tune_Low_Tomaz_Pos_1300u_NOVO.m
Sample Name CASEARINA_X_POS
Comment

Acquisition Date 9/11/2014 3:14:19 PM

Operator BDAL@DE
Instrument micrOTOF-Q II



B) Casearina J

Generic Display Report

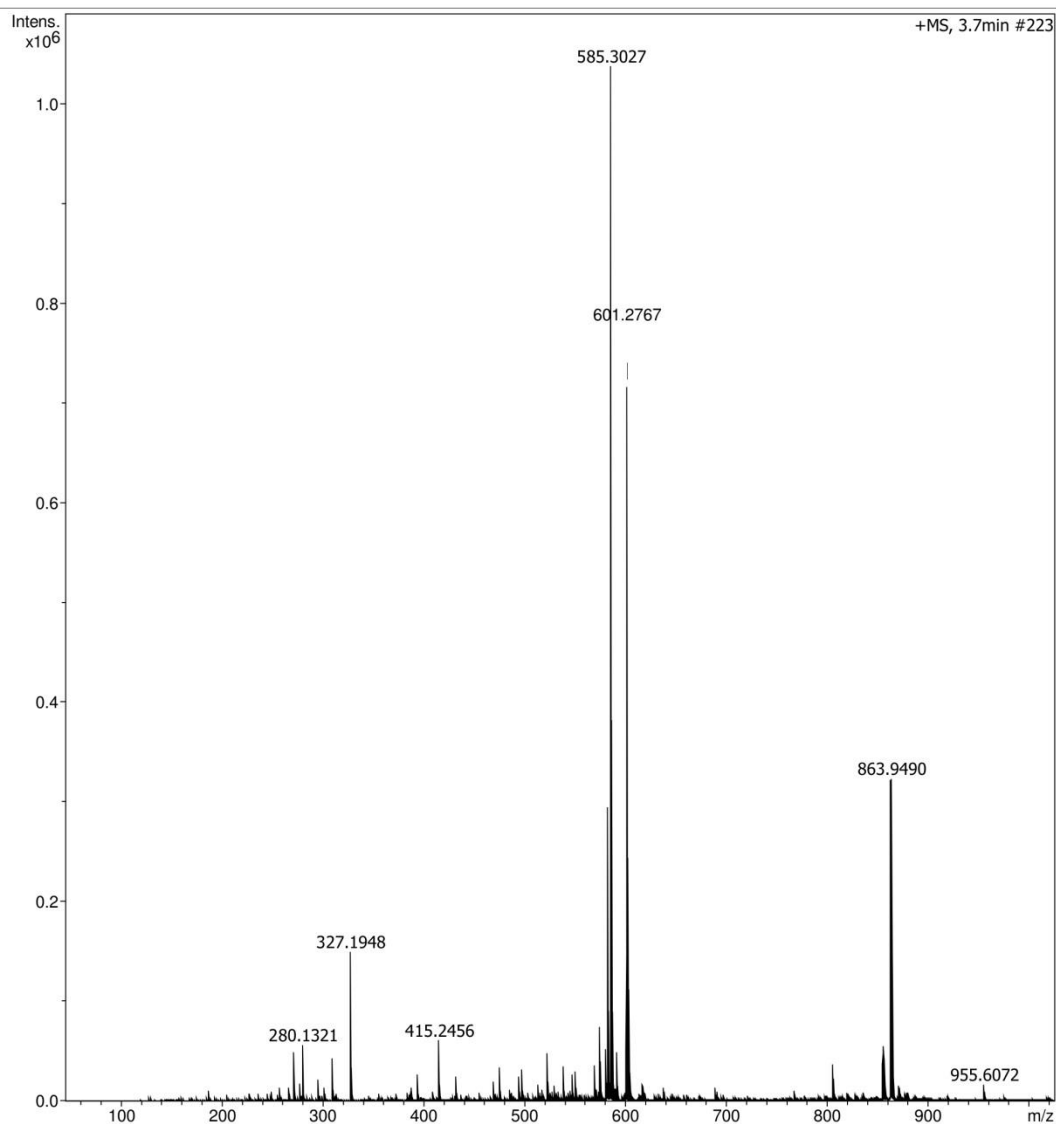
Faculdade de Ciencias Farmaceuticas de Ribeirao Preto-USP

Analysis Info

Analysis Name D:\Data\Paulinha_UNESP\CASEARINA_J_POS.d
Method Tune_Low_Tomaz_Pos_1300u_NOVO.m
Sample Name CASEARINA_J_POS
Comment

Acquisition Date 9/11/2014 4:06:20 PM

Operator BDAL@DE
Instrument micrOTOF-Q II



C) Caseargrewiina F

Generic Display Report

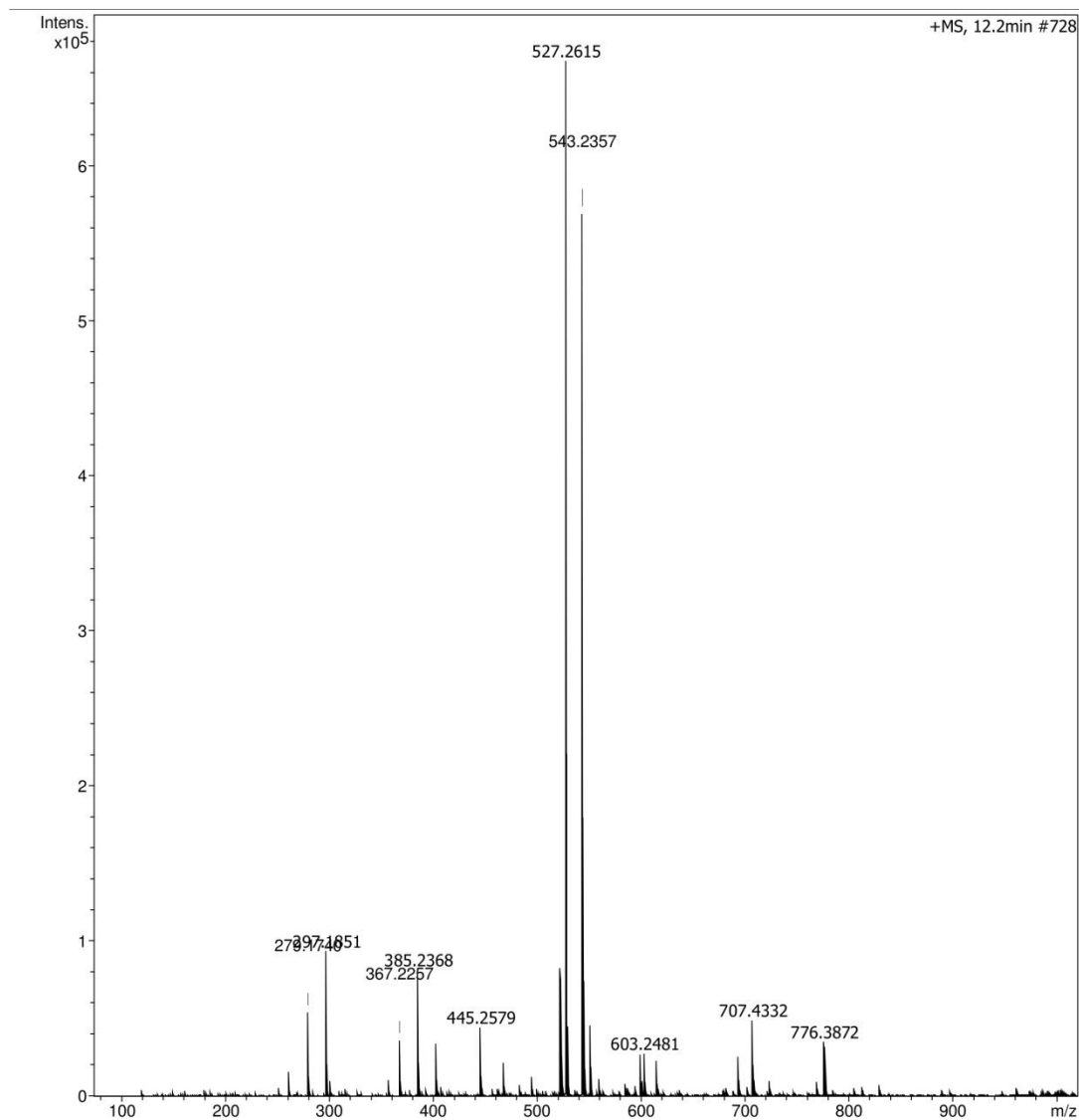
Faculdade de Ciencias Farmaceuticas de Ribeirao Preto-USP

Analysis Info

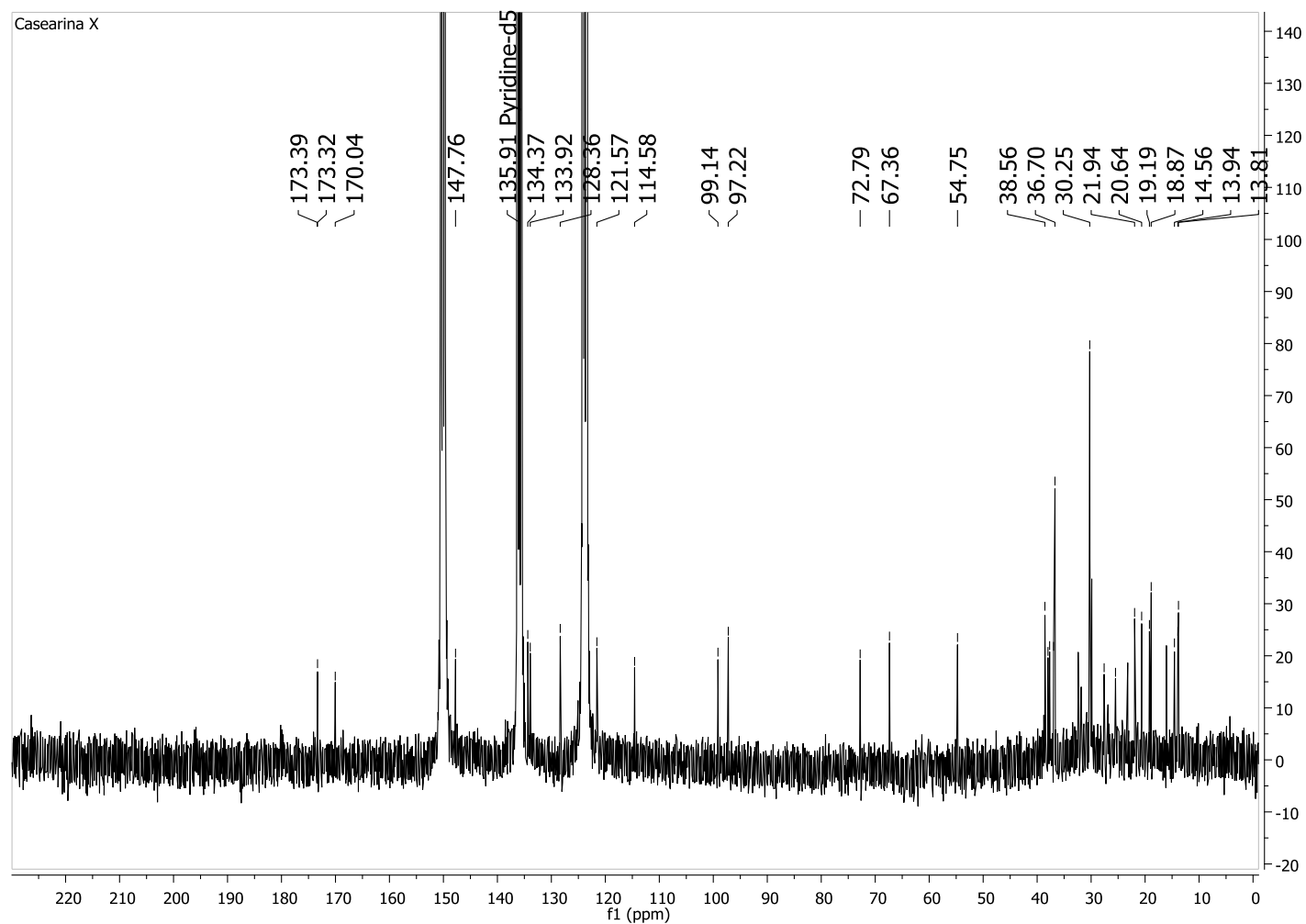
Analysis Name D:\Data\Paulinha_UNESP\CASEARINA_F_POS.d
Method Tune_Low_Tomaz_Pos_1300u_NOVO.m
Sample Name CASEARINA_F_POS
Comment

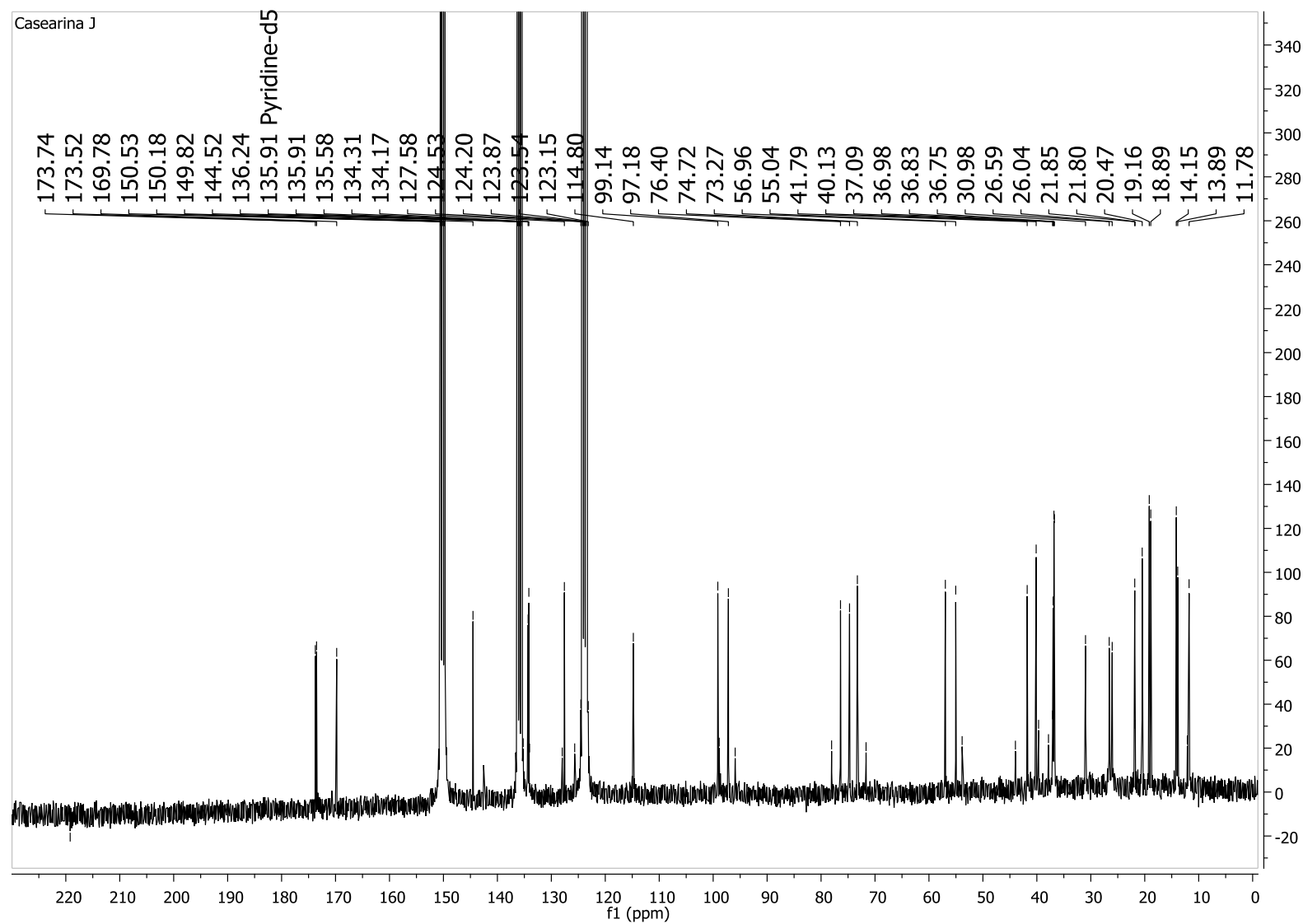
Acquisition Date 9/11/2014 3:42:52 PM

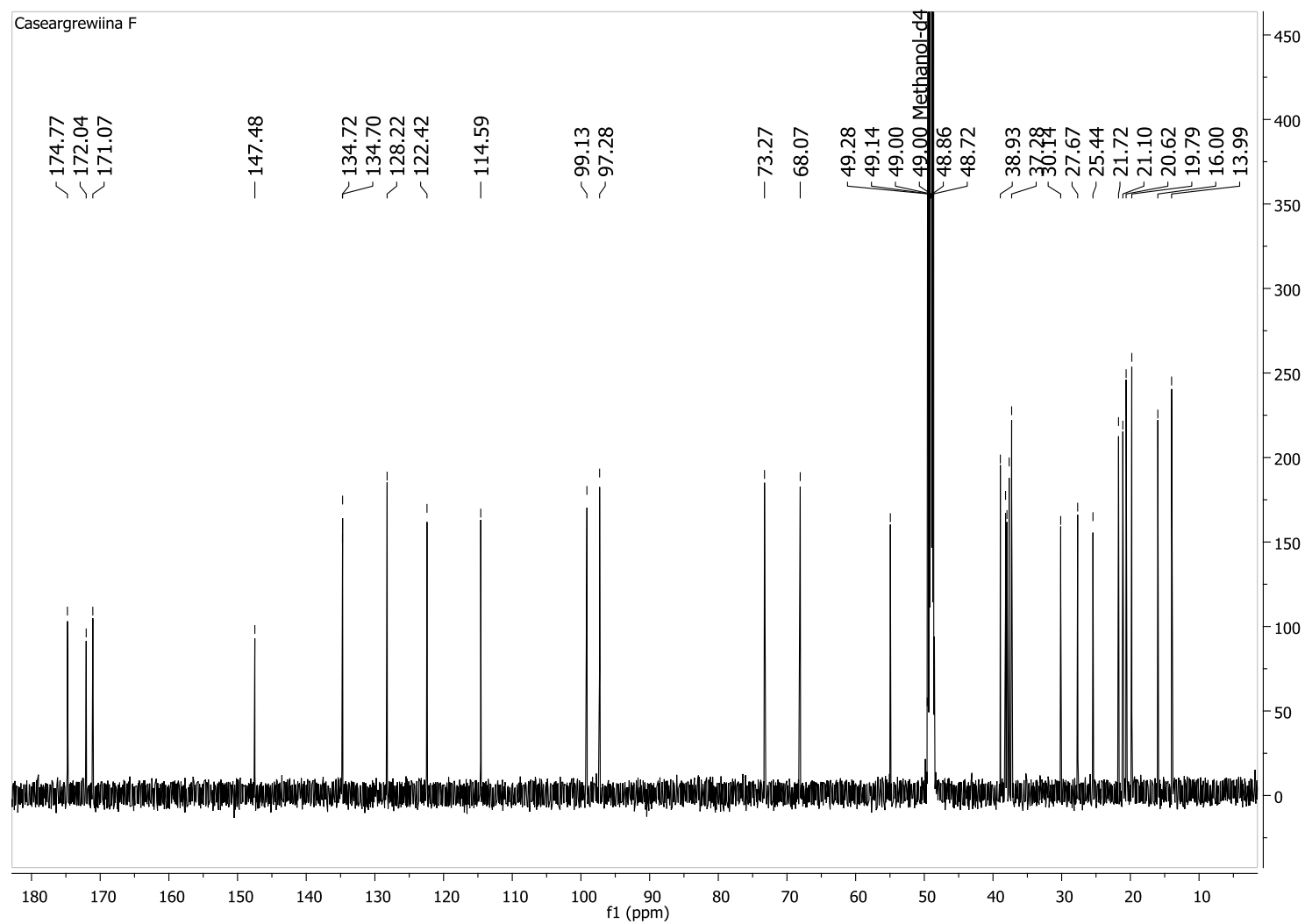
Operator BDAL@DE
Instrument micrOTOF-Q II



Apêndice X

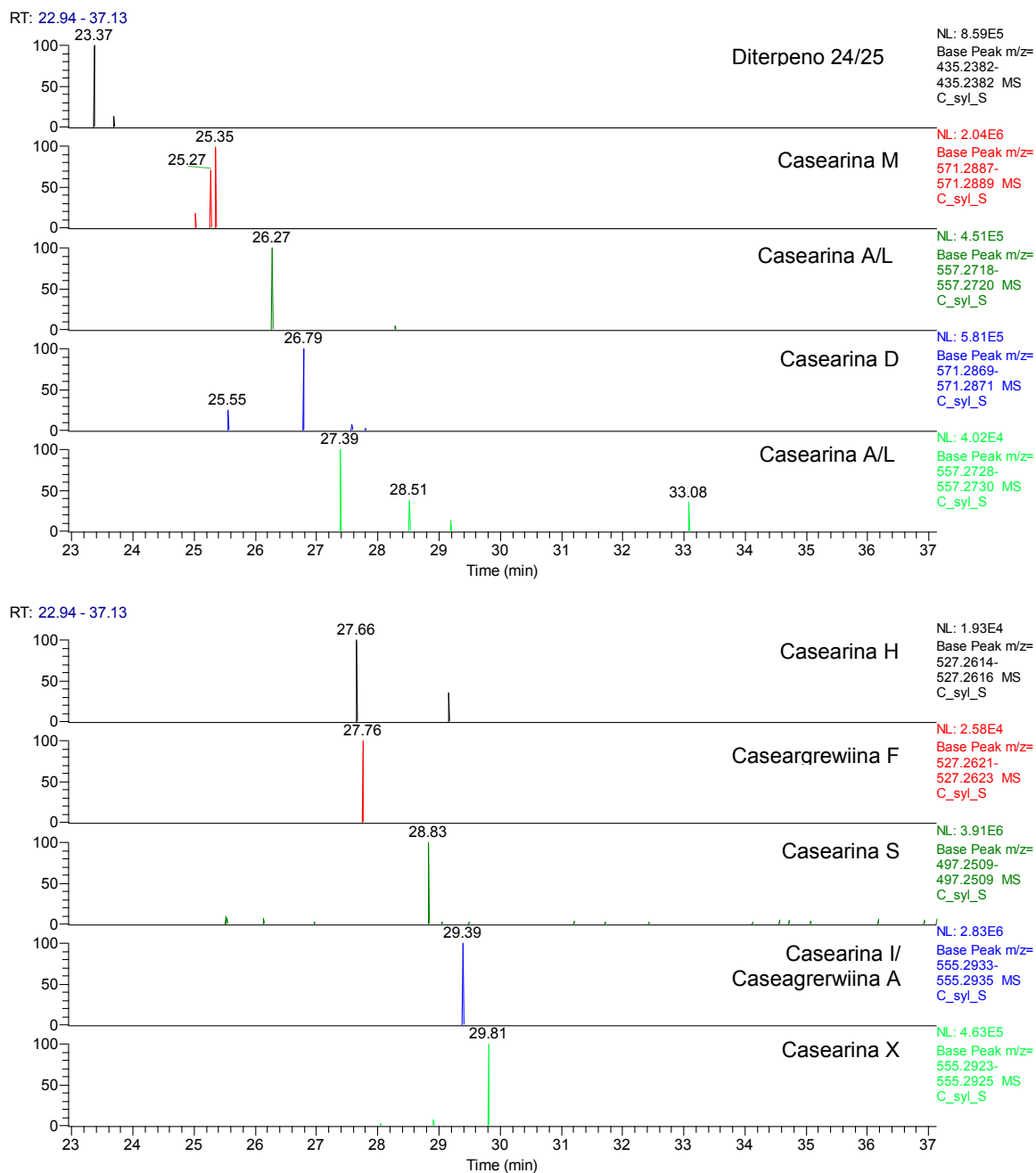
Espectros de NMR de ^{13}C dos padrões de casearinas isoladosA) Casearina X em piridina- d_5 

B) Casearina J em piridina- d_5 

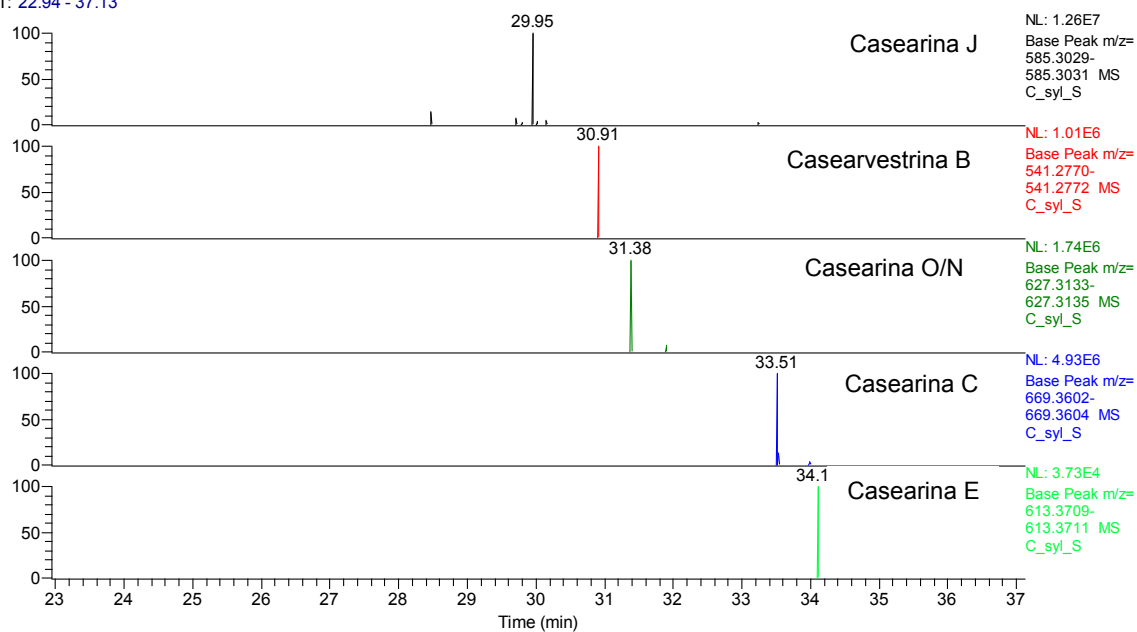
C) Caseagrerwiina F em $\text{MEOH-}d_4$ 

Apêndice XI

Cromatogramas de íon extraído (obtidos por LC-DAD-ESI-Orbitrap) de cada casearina identificada por desreplicação



RT: 22.94 - 37.13



Apêndice XII

Espectros obtidos para a elucidação estrutural dos flavonoides isolados de *C. sylvestris* var. *lingua*

A) Espectros obtidos para a elucidação estrutural da (+)-catequina

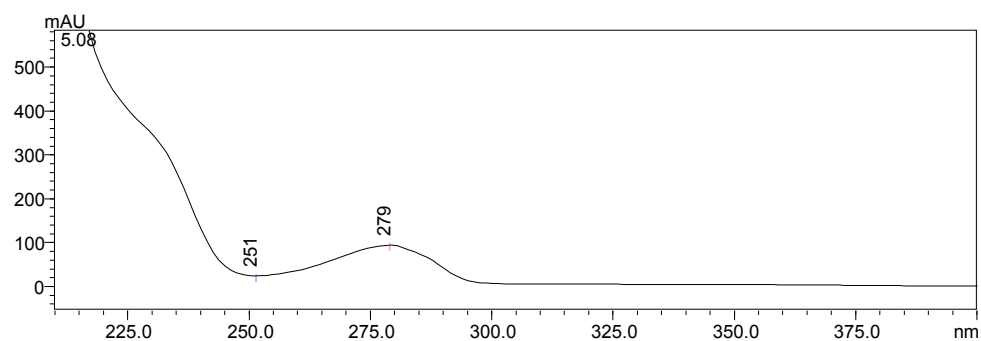
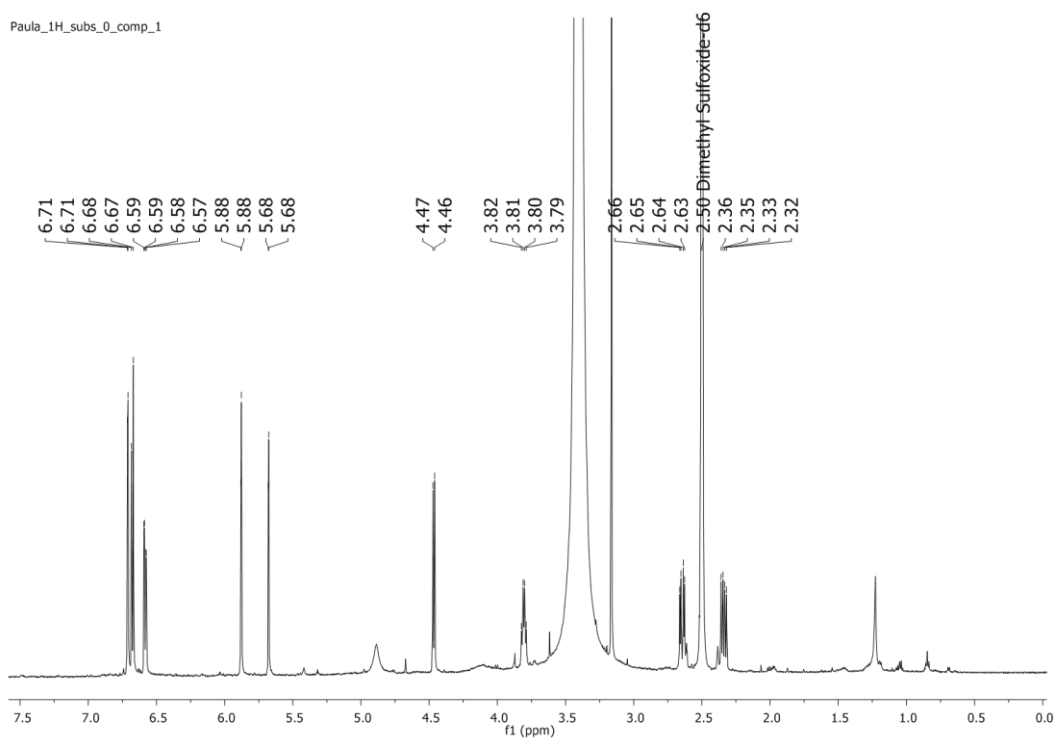


Figura A.1 - Espectro no UV da (+)-catequina

Figura A.2 – Espectro de NMR de ¹H (600 MHz, (CD₃)₂SO) da (+)-catequina

Paula_13C_subs_0_comp_1

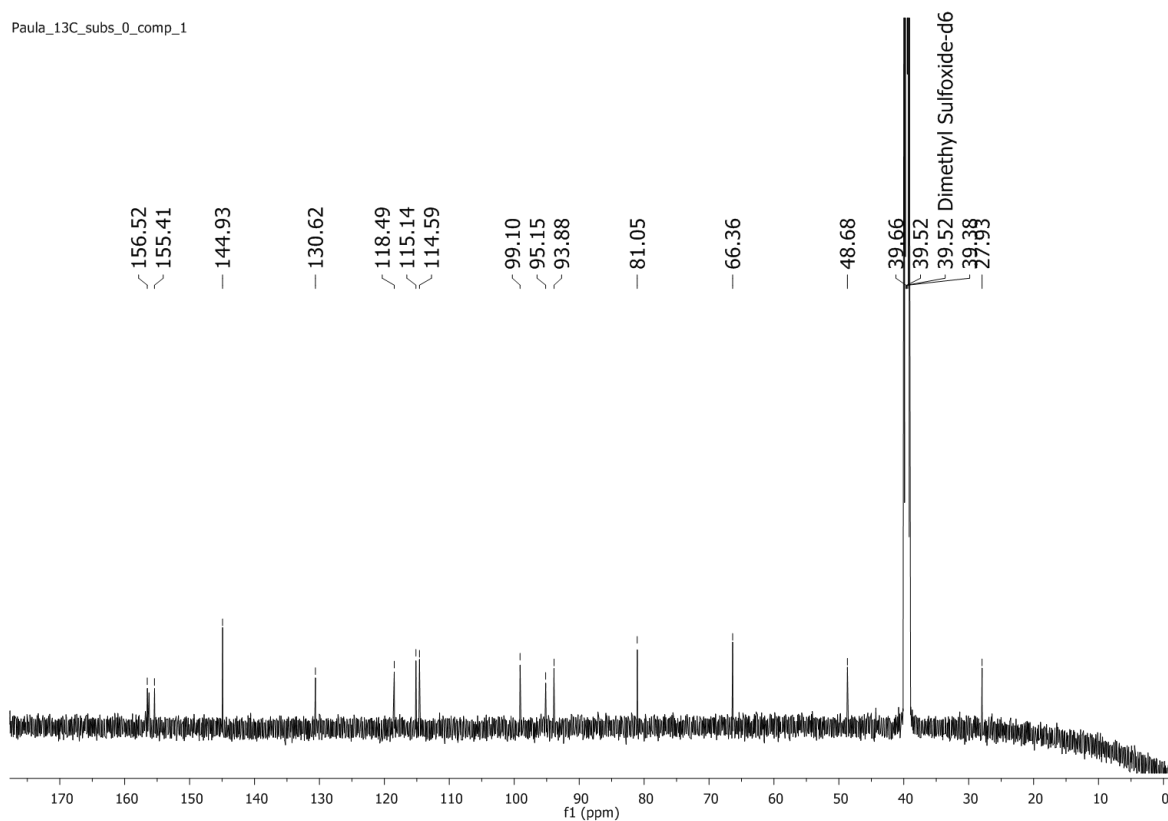


Figura A.3 – Espectro de NMR de ^{13}C (150 MHz, $(\text{CD}_3)_2\text{SO}$) da (+)-catequina

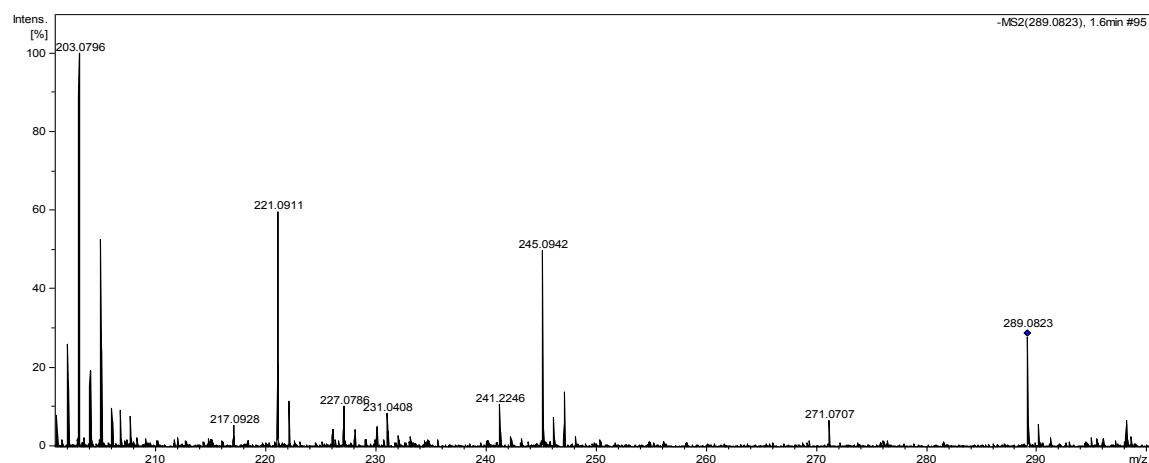


Figura A.4 – Espectro de massas (MS/Ms) da (+)-catequina

B) Espectros obtidos para a elucidação estrutural da quercetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 6)- β -D-glucopiranosideo (rutina)

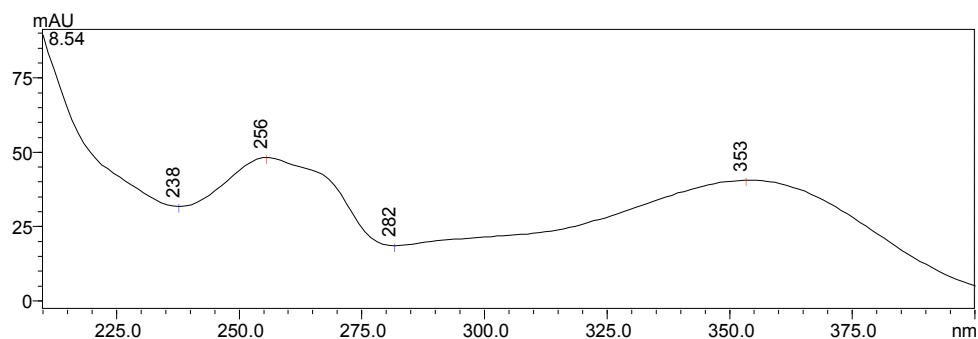


Figura B.1 - Espectro no UV da rutina

Paula_1H_Subst_2_Comp_2

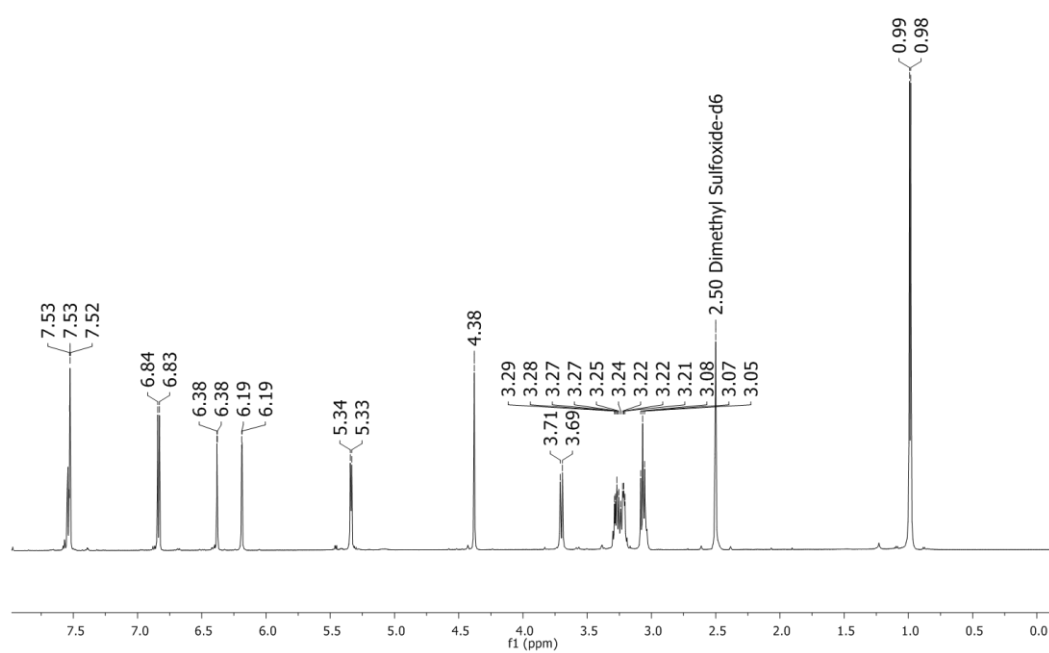


Figura B.2 – Espectro de NMR de ^1H (600 MHz, $(\text{CD}_3)_2\text{SO}$) da rutina

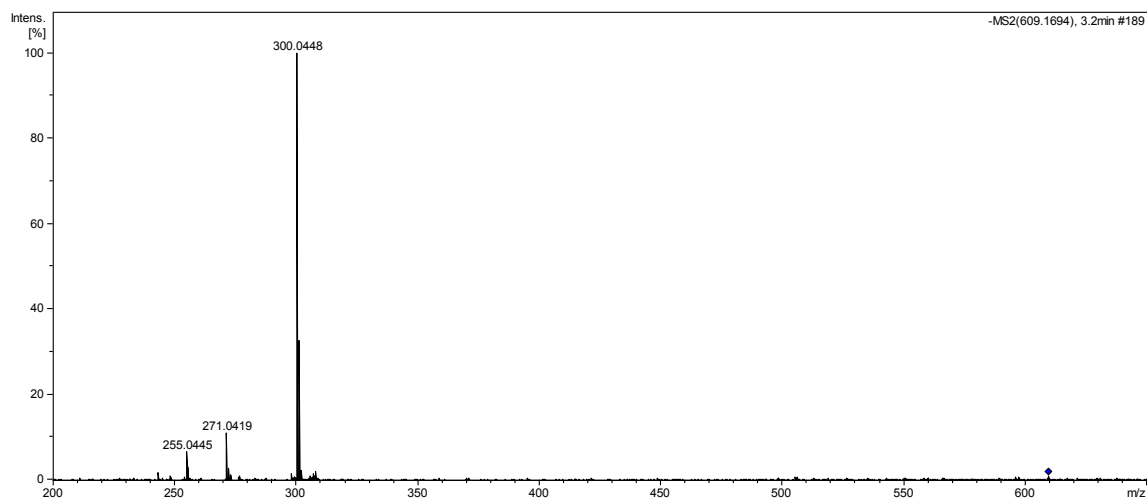


Figura B.3 – Espectro de massas (MS/Ms) da rutina

C) Espectros obtidos para a elucidação estrutural da isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- β -D-glucopiranosideo (isoramnetina-3-O-neohesperidosideo)

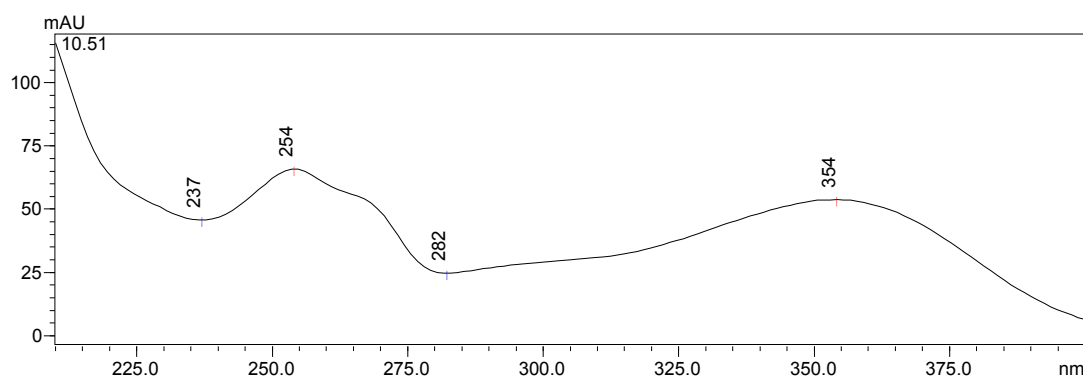


Figura C.1 – Espectro no UV da isoramnetina-3-O-neohesperidosideo

Paula_1H_Subst_4_Comp_3

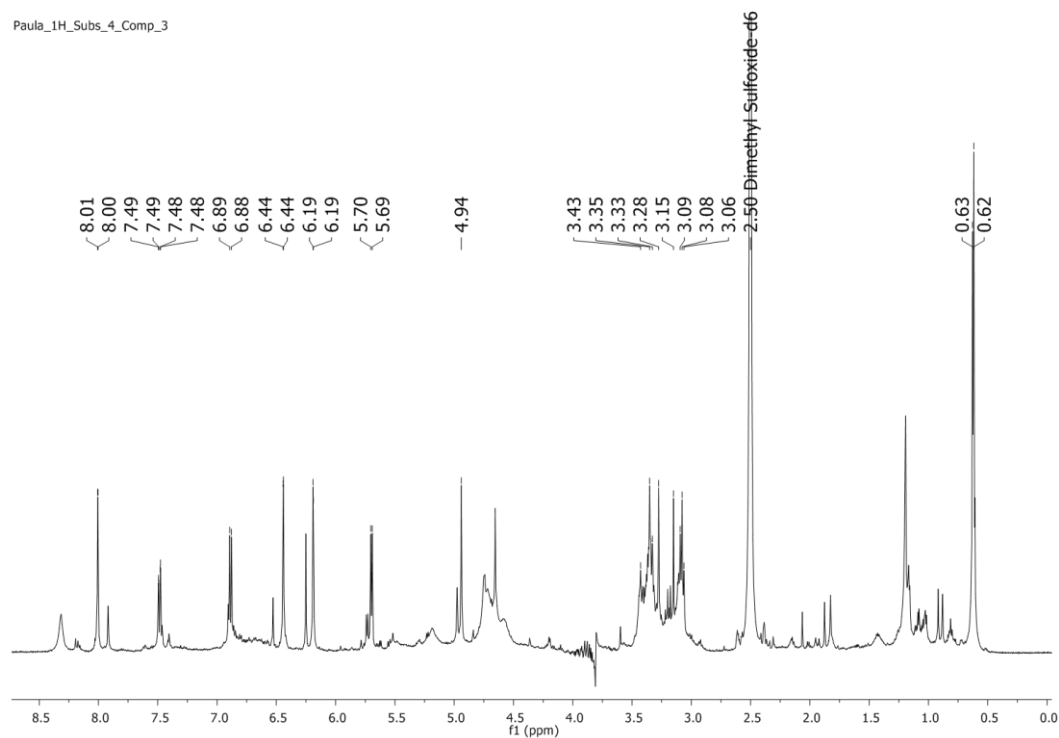


Figura C.2 – Espectro de NMR de ^1H (600 MHz, $(\text{CD}_3)_2\text{SO}$) da isoramnetina-3-O-neohesperidosideo

Paula_13C_Subst_4_Comp_3

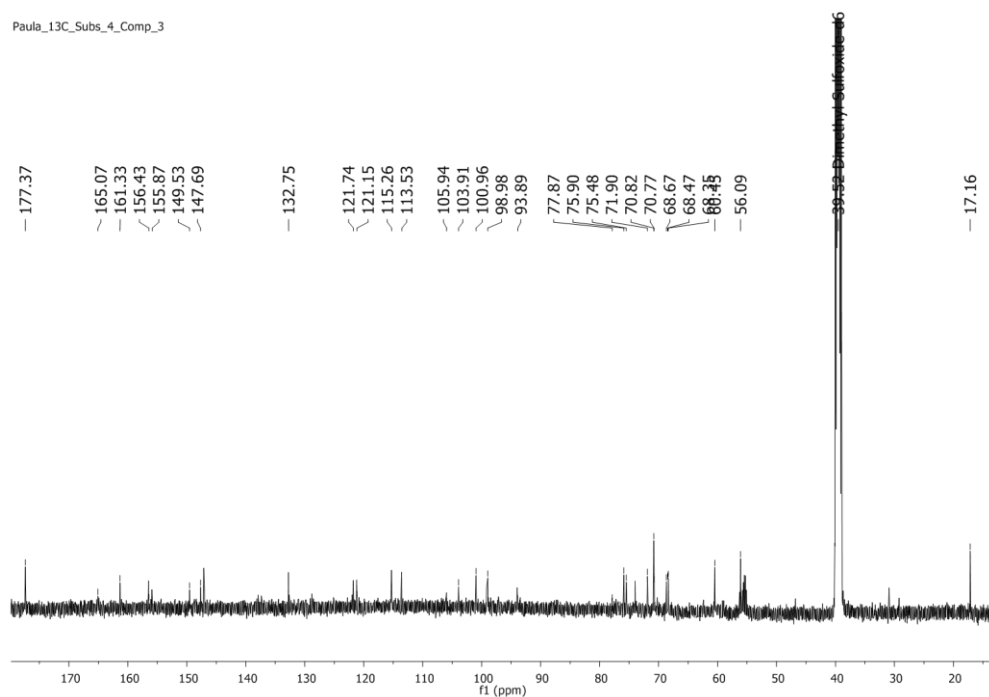


Figura C.3 – Espectro de NMR de ^{13}C (150 MHz, $(\text{CD}_3)_2\text{SO}$) da isoramnetina-3-O-neohesperidosideo

Paula_Tocsy1D_5.70_Sub3_4_Comp_3

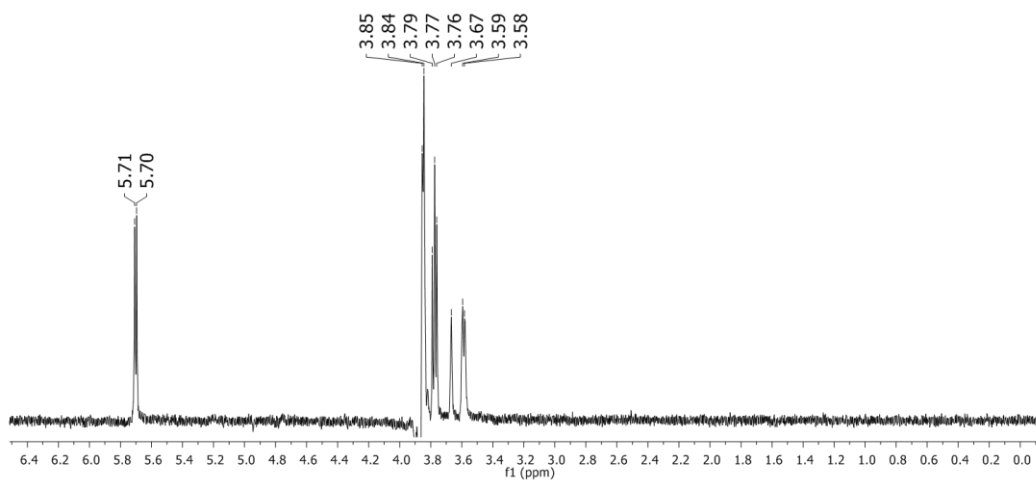


Figura C.4 - TOCSY 1D (irradiando o sinal em 5,70 ppm, $(\text{CD}_3)_2\text{SO}$) da isoramnetina-3-O-neohesperidosideo

Paula_Tocsy1D_4.95_Sub3_4_Comp_3

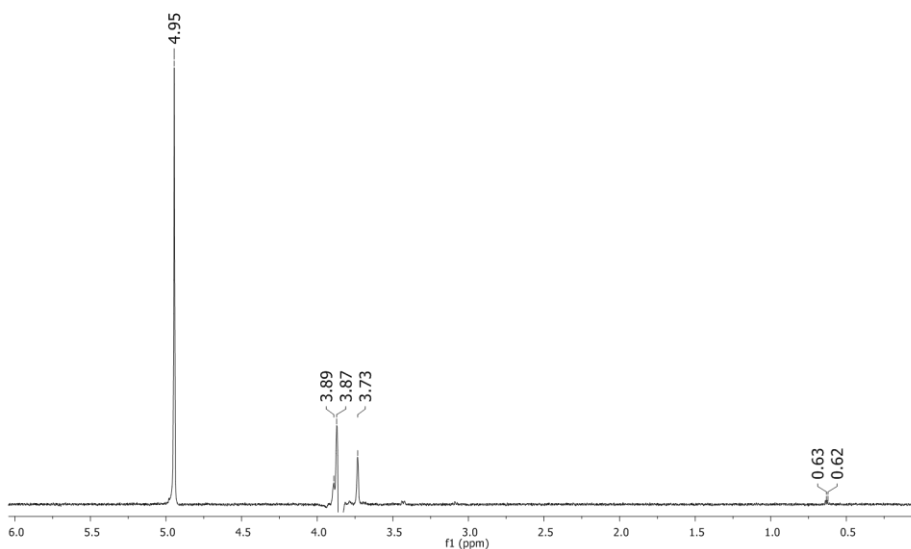


Figura C.5 - TOCSY 1D (irradiando o sinal em 4,95 ppm, $(\text{CD}_3)_2\text{SO}$) da isoramnetina-3-O-neohesperidosideo

Paula_Tocsy1D_0.63_Subst_4_Comp_3

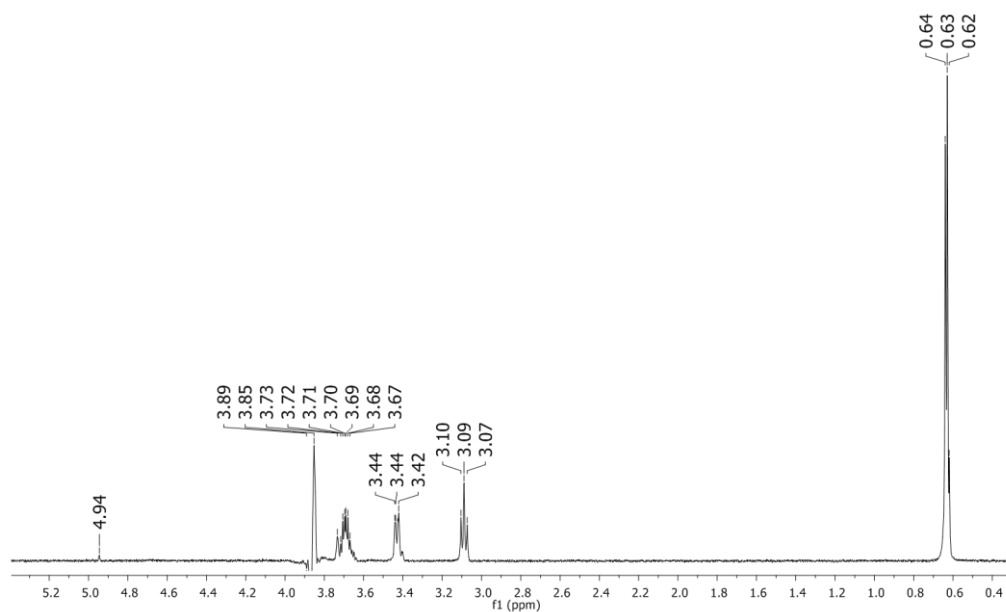


Figura C.6 - TOCSY 1D (irradiando o sinal em 0,63 ppm, $(\text{CD}_3)_2\text{SO}$) da isoramnetina-3-O-neohesperidosideo

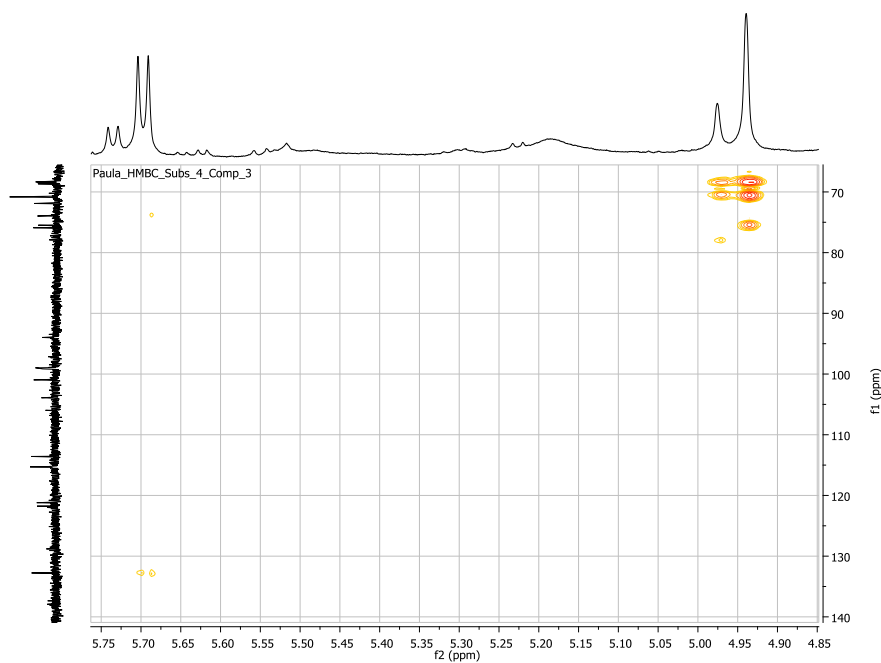


Figura C.7 – Espectro de HMBC (expansão da região dos sinais correspondentes ao H anomérico, $(\text{CD}_3)_2\text{SO}$) da isoramnetina-3-O-neohesperidosideo

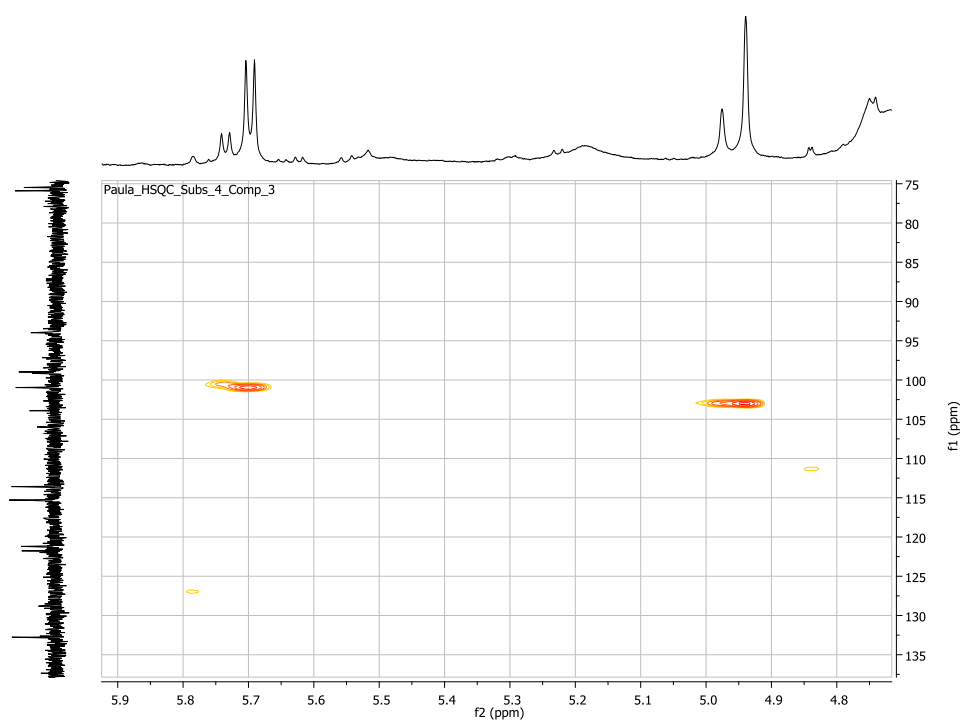


Figura C.8 – Espectro de HSQC (expansão da região dos sinais correspondentes ao H anomérico, $\text{CD}_3)_2\text{SO}$) da isoramnetina-3-O-neohesperidosideo

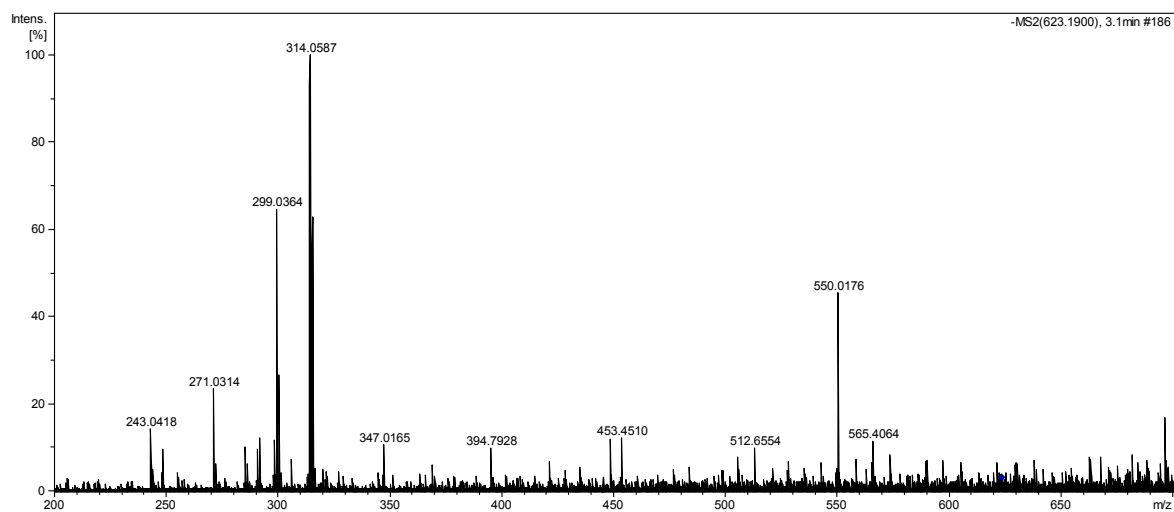


Figura C.9 – Espectro de massas (MS/MS) da isoramnetina-3-O-neohesperidosideo

D) Espectros obtidos para a elucidação estrutural da isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 6)- β -D-glucopiranosideo (narcisina)

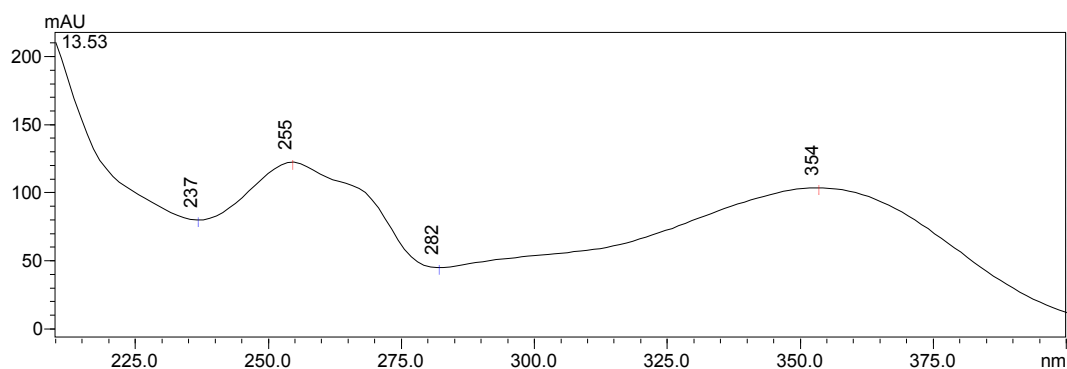


Figura D.1 – Espectro no UV da narcisina

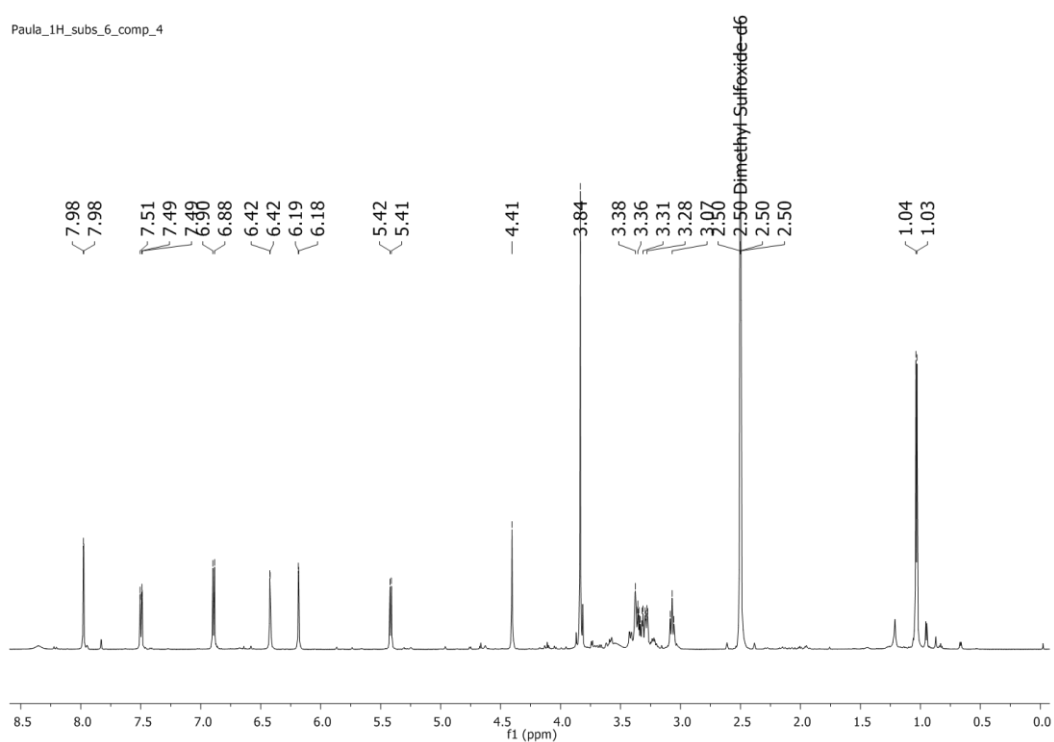


Figura D.2 – Espectro de NMR de ^1H (600 MHz, $(\text{CD}_3)_2\text{SO}$) da narcisina

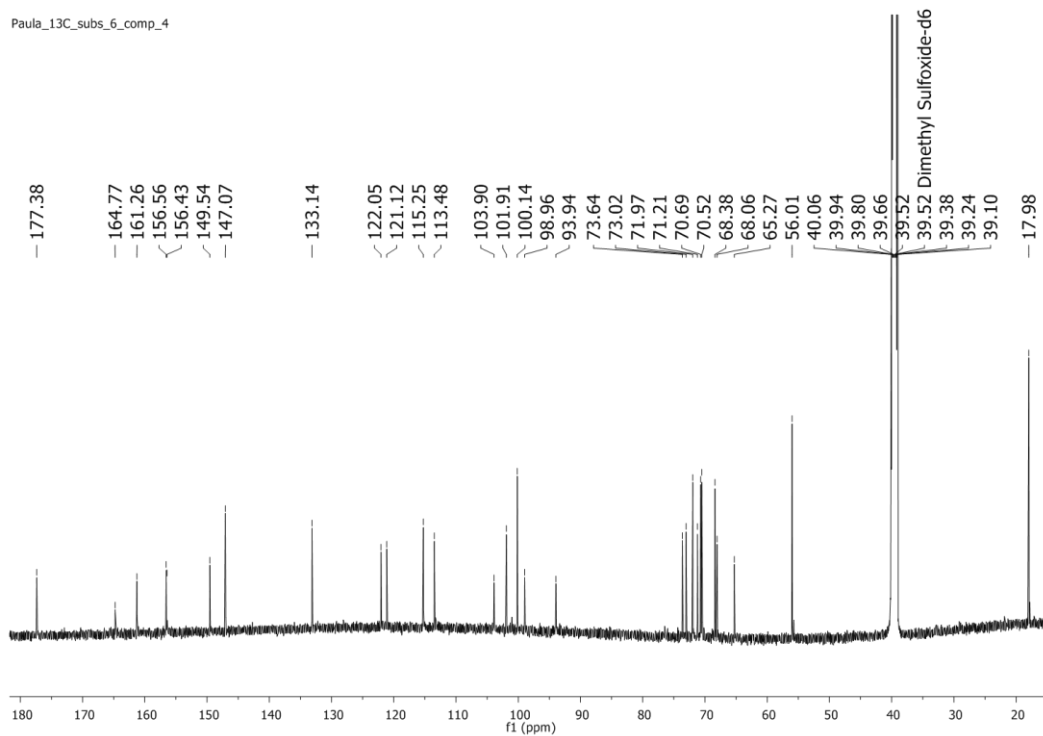


Figura D.3 – Espectro de NMR de ^{13}C (150 MHz, $(\text{CD}_3)_2\text{SO}$) da narcisina

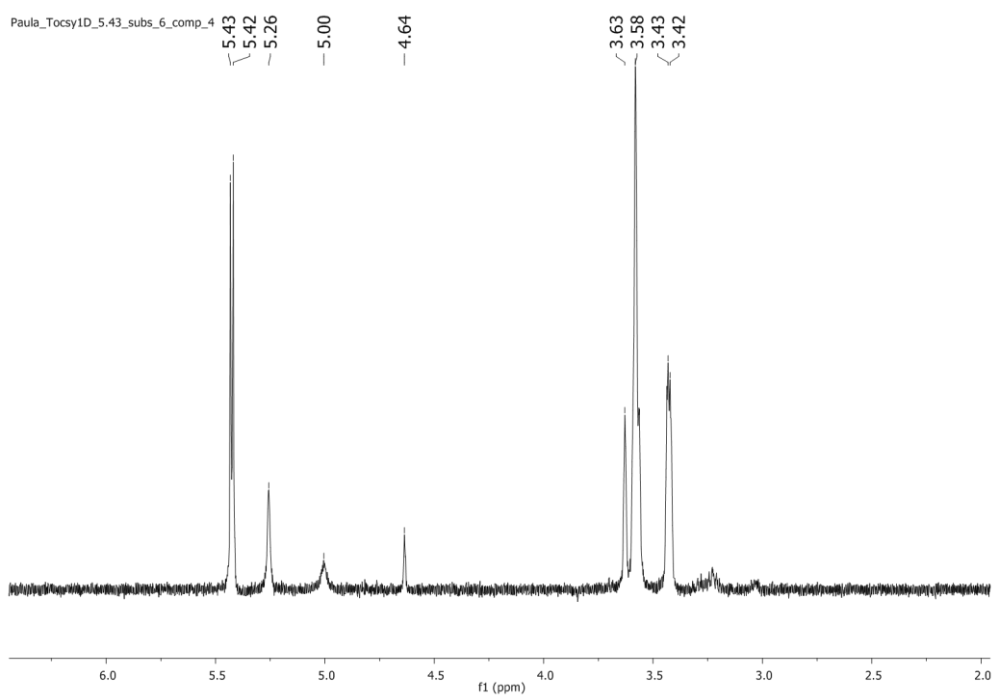
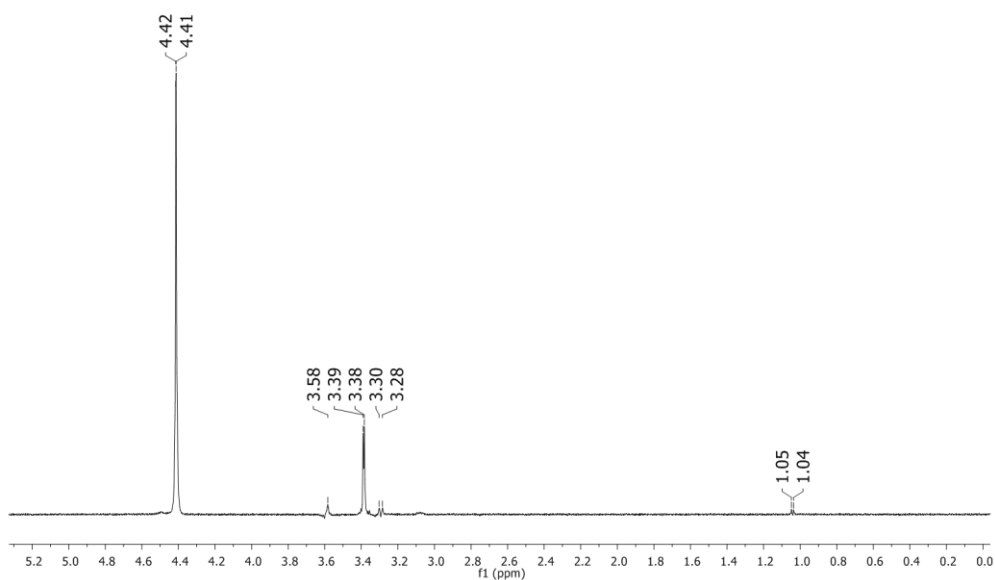
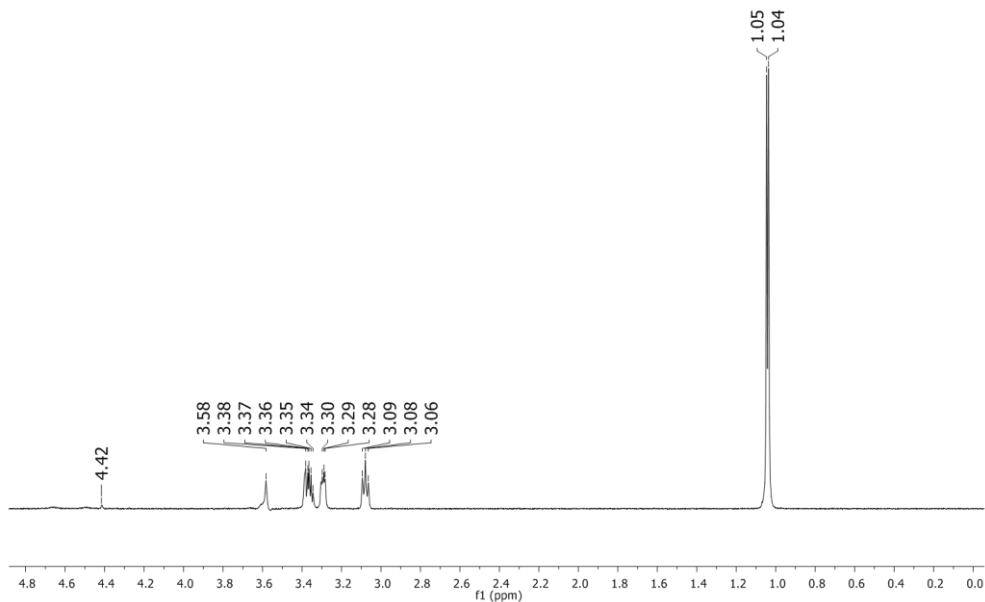


Figura D.4 - TOCSY 1D (irradiando o sinal em 5,43 ppm, $(\text{CD}_3)_2\text{SO}$) da narcisina

Paula_Tocsy1D_4.42_subs_6_comp_4

Figura D.5 - TOCSY 1D (irradiando o sinal em 4,42 ppm, $(\text{CD}_3)_2\text{SO}$) da narcisina

Paula_Tocsy1D_1.04_subs_6_comp_4

Figura D.6 - TOCSY 1D (irradiando o sinal em 1,04 ppm, $(\text{CD}_3)_2\text{SO}$) da narcisina

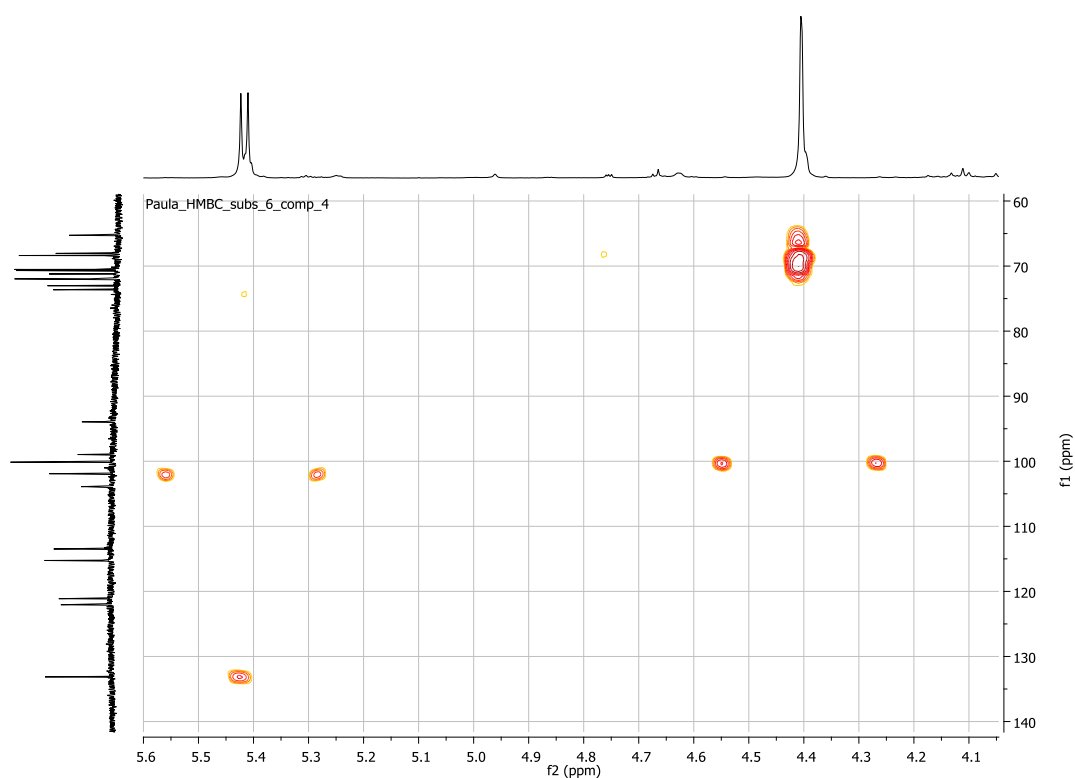


Figura D.7 – Espectro de HMBC (expansão da região dos sinais correspondentes ao H anomérico, $(\text{CD}_3)_2\text{SO}$) da narcisina

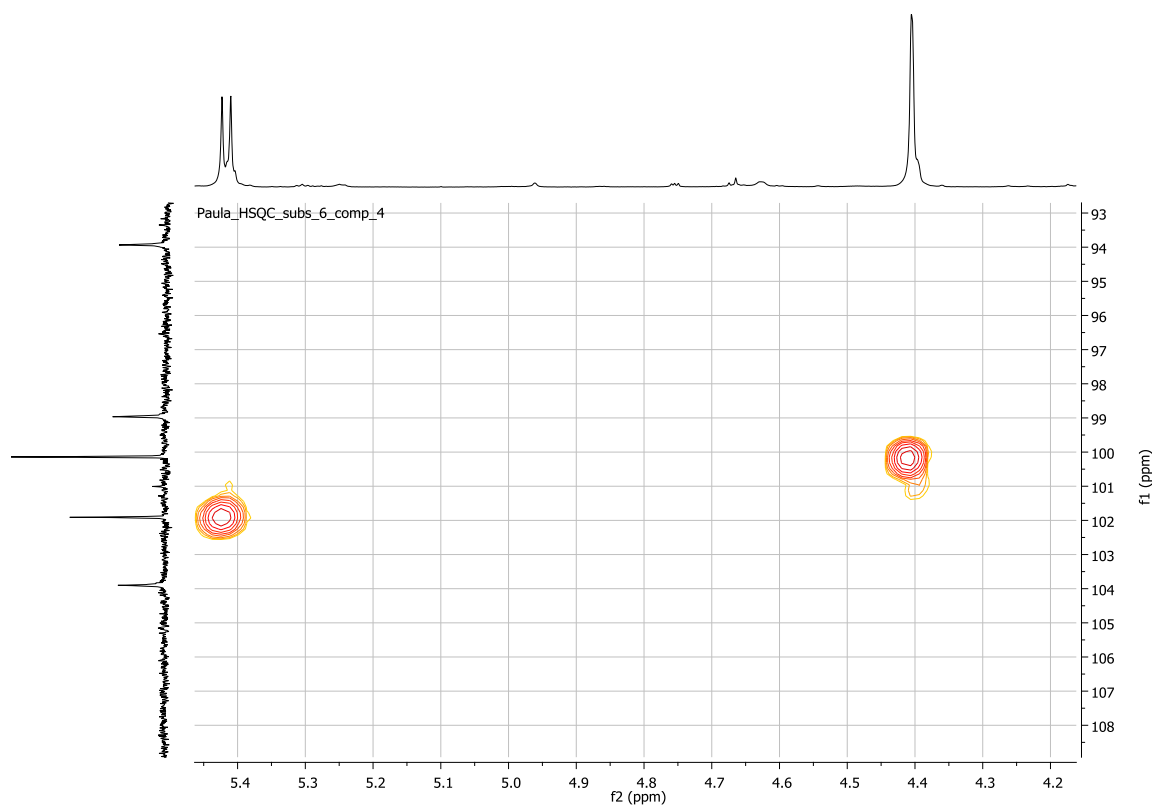


Figura D.8 – Espectro de HSQC (expansão da região dos sinais correspondentes ao H anomérico, $(\text{CD}_3)_2\text{SO}$) da narcisina

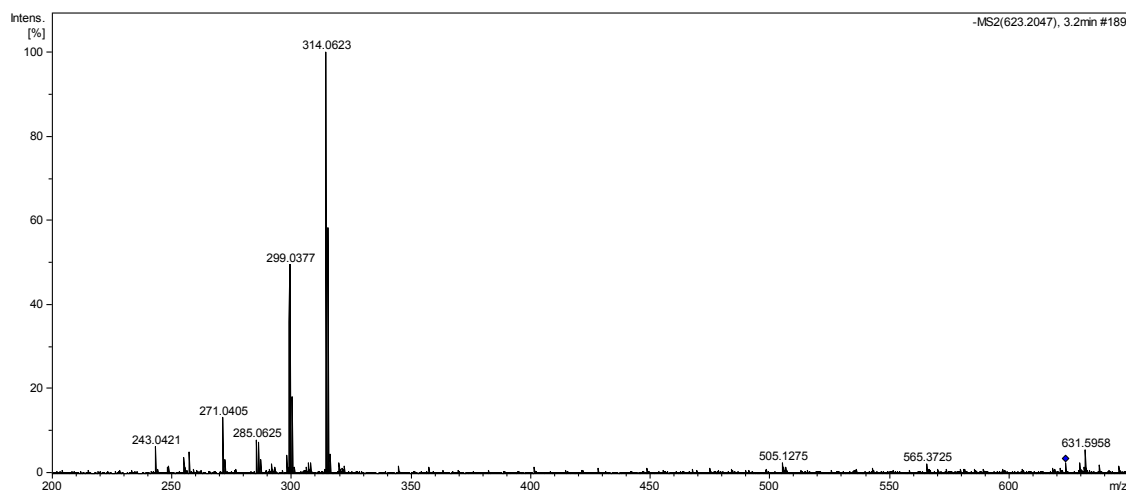


Figura D.9 – Espectro de massas (MS/MS) da narcisina

E) Espectros obtidos para a elucidação estrutural da isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- β -D-arabinopiranosideo

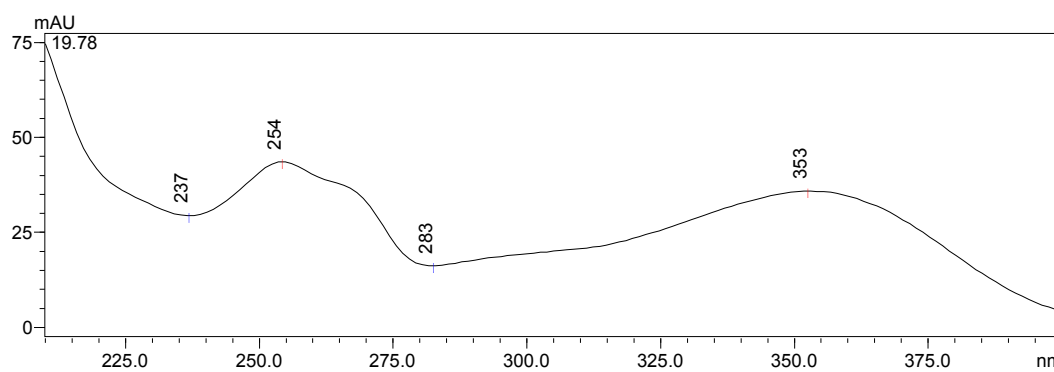


Figura E.1 – Espectro no UV da isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- β -D-arabinopiranosideo

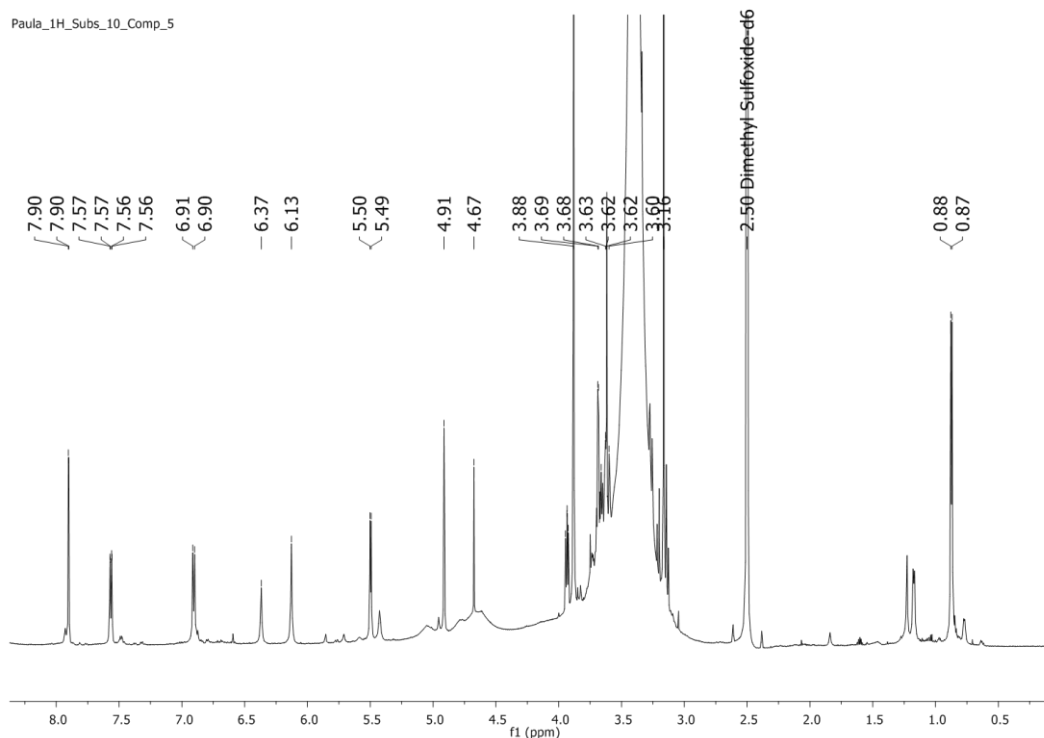


Figura E.2 – Espectro de NMR de ^1H (600 MHz, $(\text{CD}_3)_2\text{SO}$) da isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- β -D-arabinopiranosideo

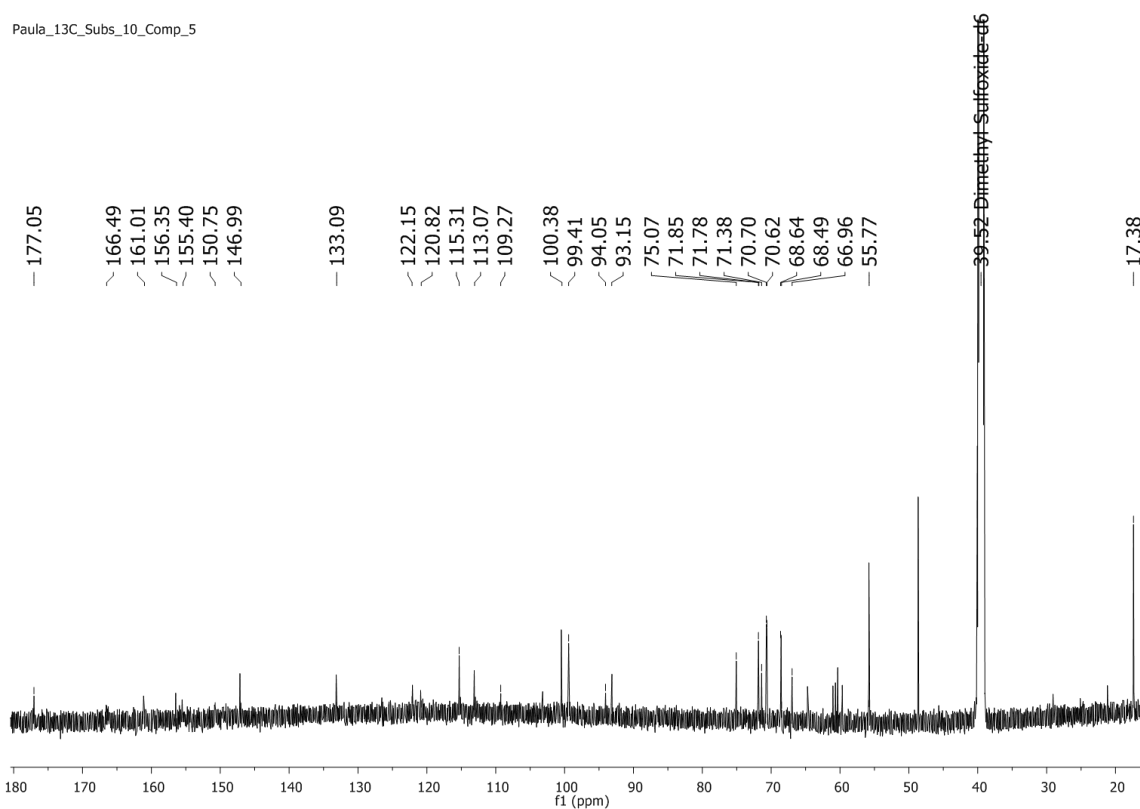


Figura E.3 – Espectro de NMR de ^{13}C (150 MHz, $(\text{CD}_3)_2\text{SO}$) da isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- β -D-arabinopiranosideo

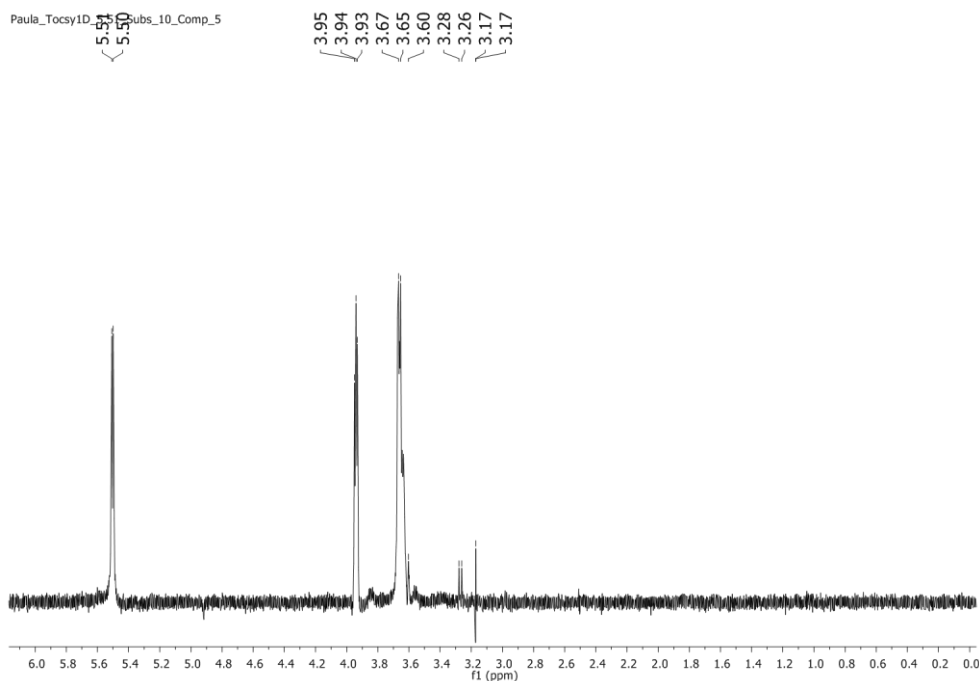


Figura E.4 - TOCSY 1D (irradiando o sinal em 5,51 ppm, $(\text{CD}_3)_2\text{SO}$) da isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- β -D-arabinopiranosideo

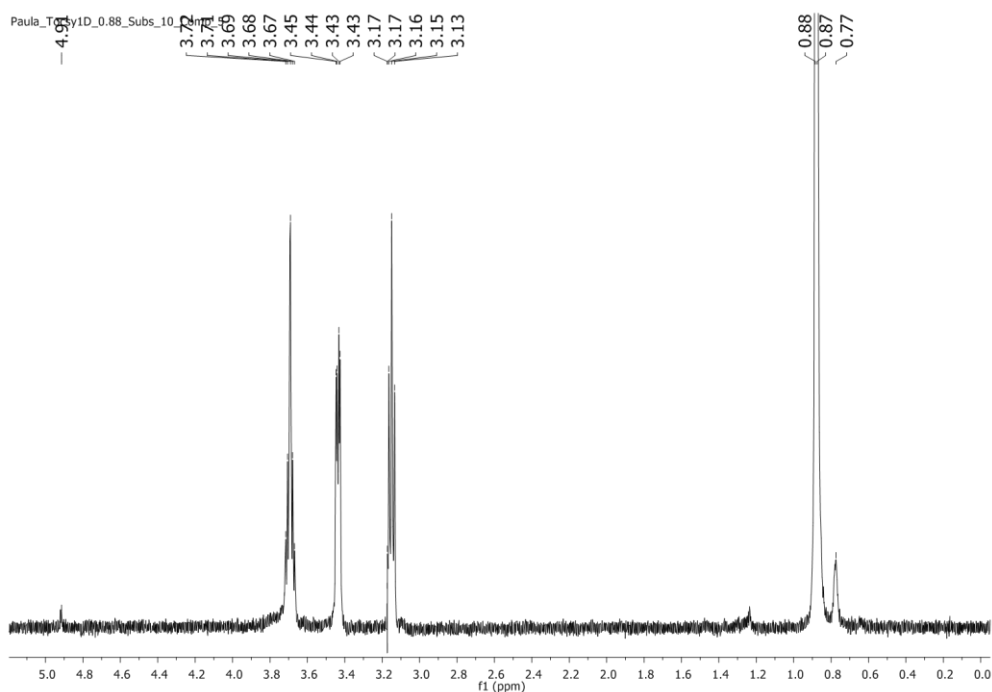


Figura E.5 - TOCSY 1D (irradiando o sinal em 0,88 ppm, $(\text{CD}_3)_2\text{SO}$) da isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- β -D-arabinopiranosideo

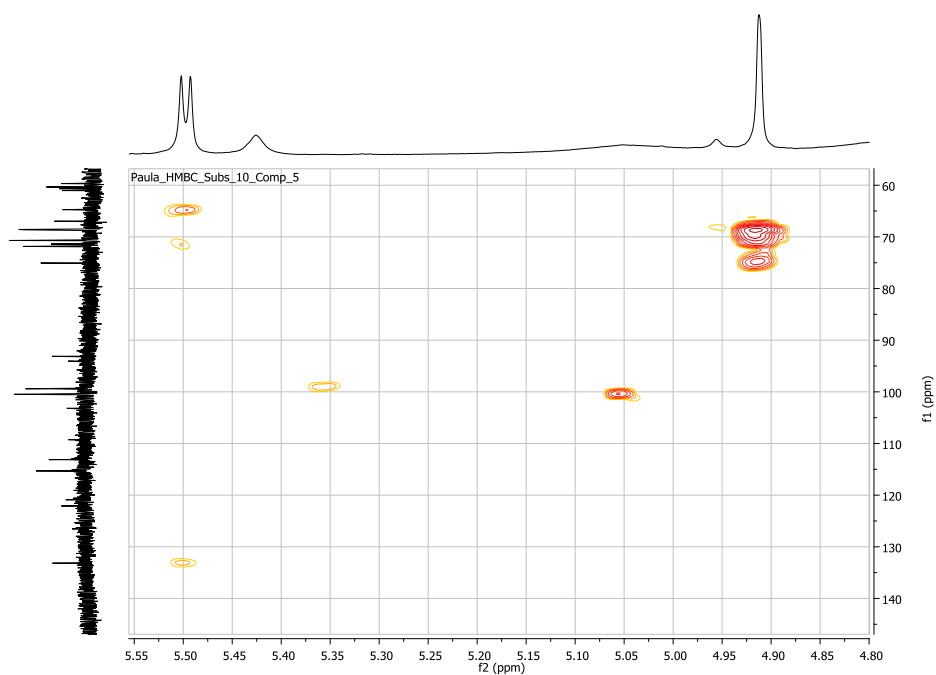


Figura E.6 – Espectro de HMBC (expansão da região dos sinais correspondentes ao H anomérico, $(\text{CD}_3)_2\text{SO}$) da isoramnetina-3-O- α -L-ramnopyranosil-(1 $\rightarrow$ 2)- β -D-arabinopyranosideo

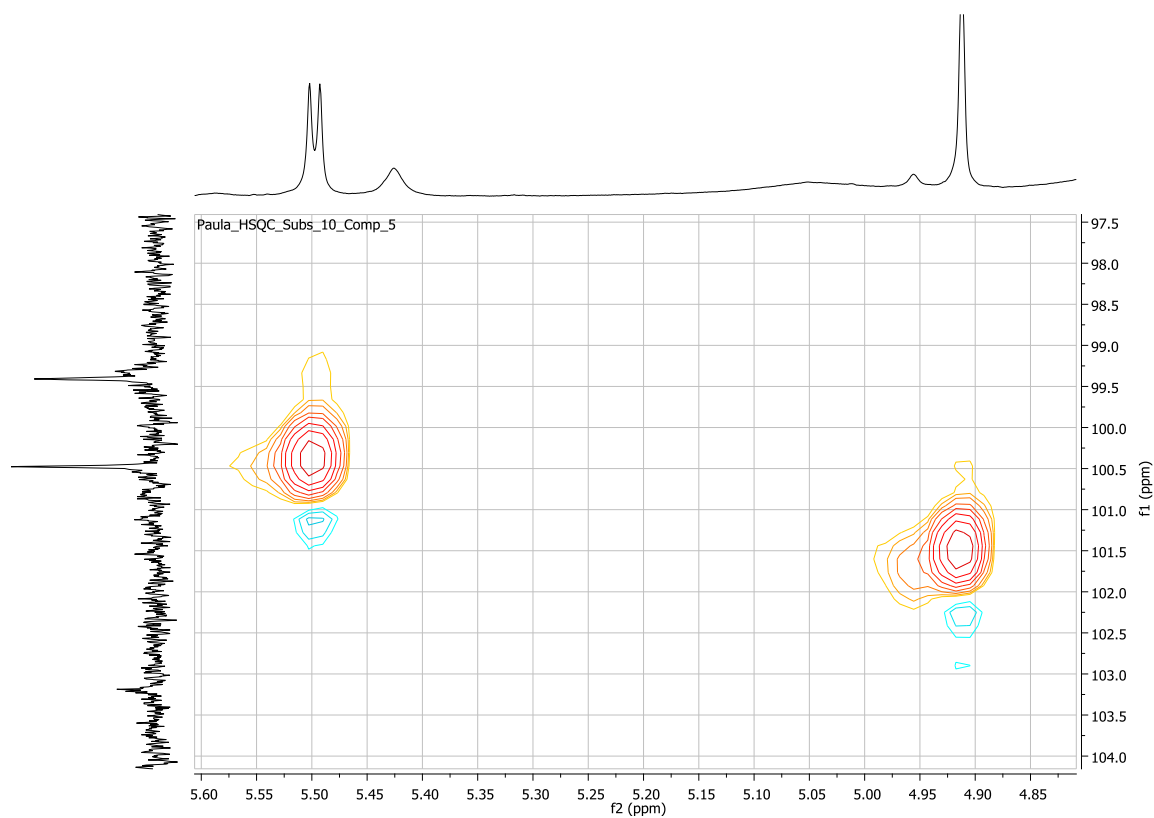


Figura E.7 – Espectro de HSQC (expansão da região dos sinais correspondentes ao H anomérico, $(\text{CD}_3)_2\text{SO}$) da isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- β -D-arabinopiranosideo

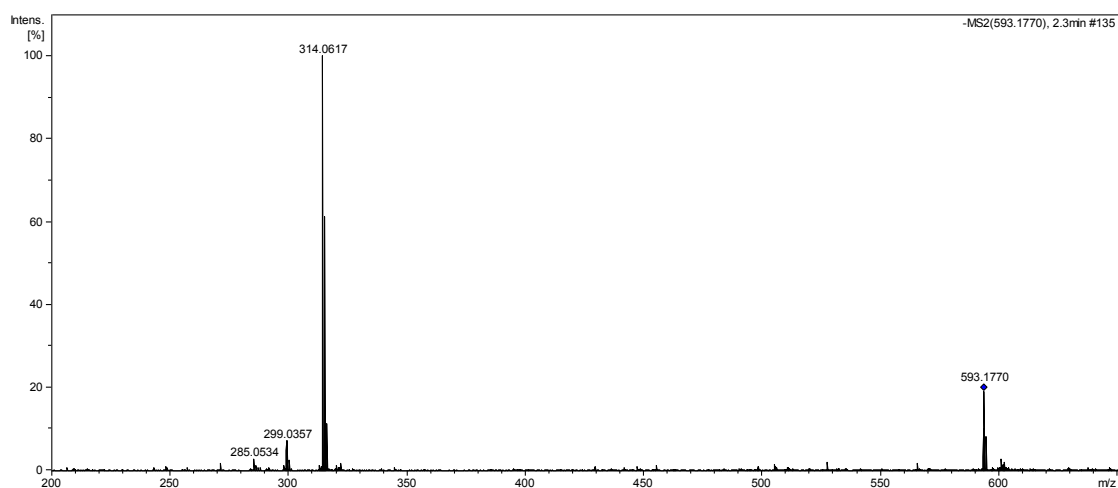
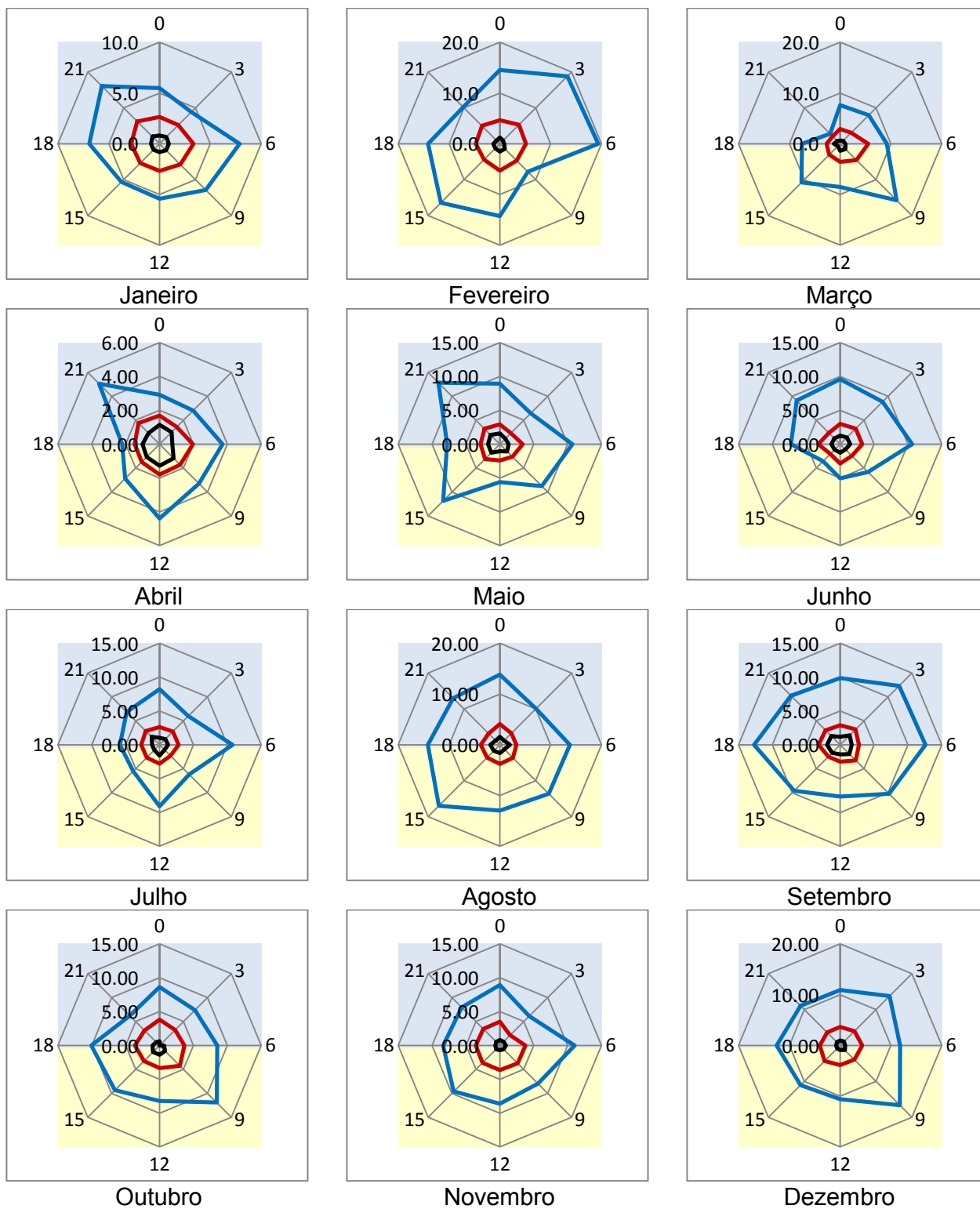
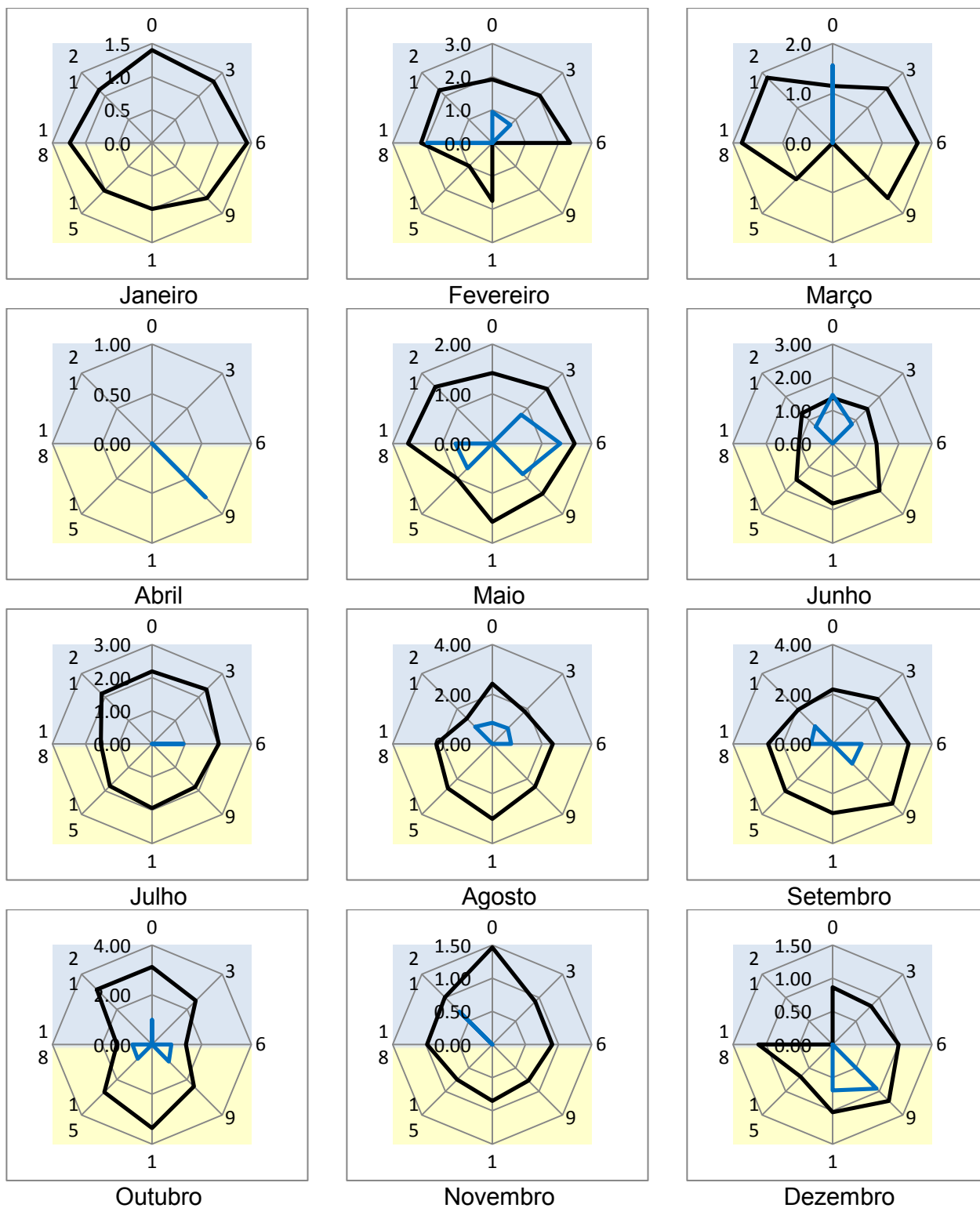


Figura E.8 – Espectro de massas (MS/MS) da isoramnetina-3-O- α -L-ramnopiranosil-(1 $\rightarrow$ 2)- β -D-arabinopiranosideo

Apêndice XIII

Gráficos do ciclo circadiano - Flavonoides: *C. sylvestris* var. *lingua*

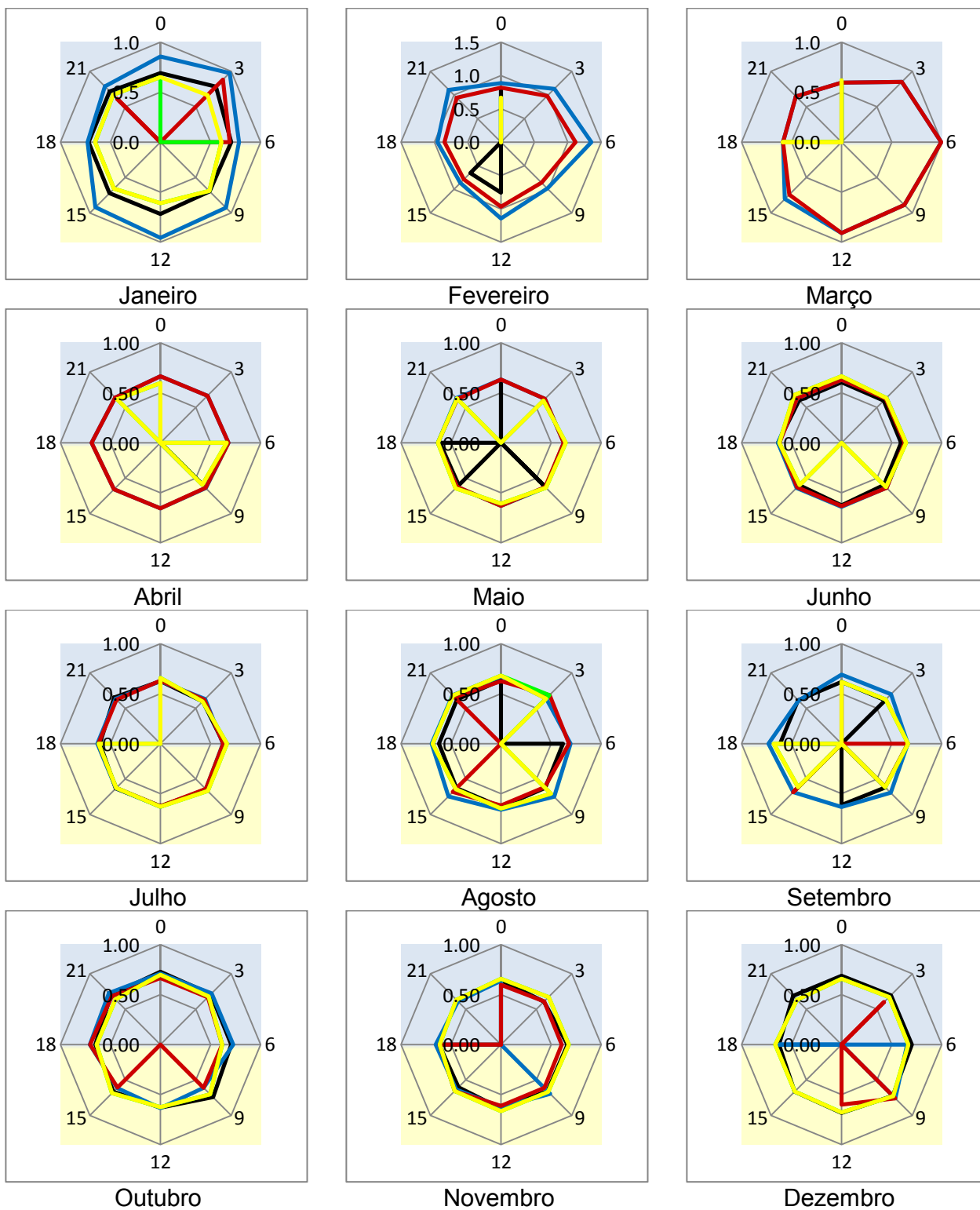
Gráficos do ciclo circadiano - Flavonoides: *C. sylvestris* var. *sylvestris*



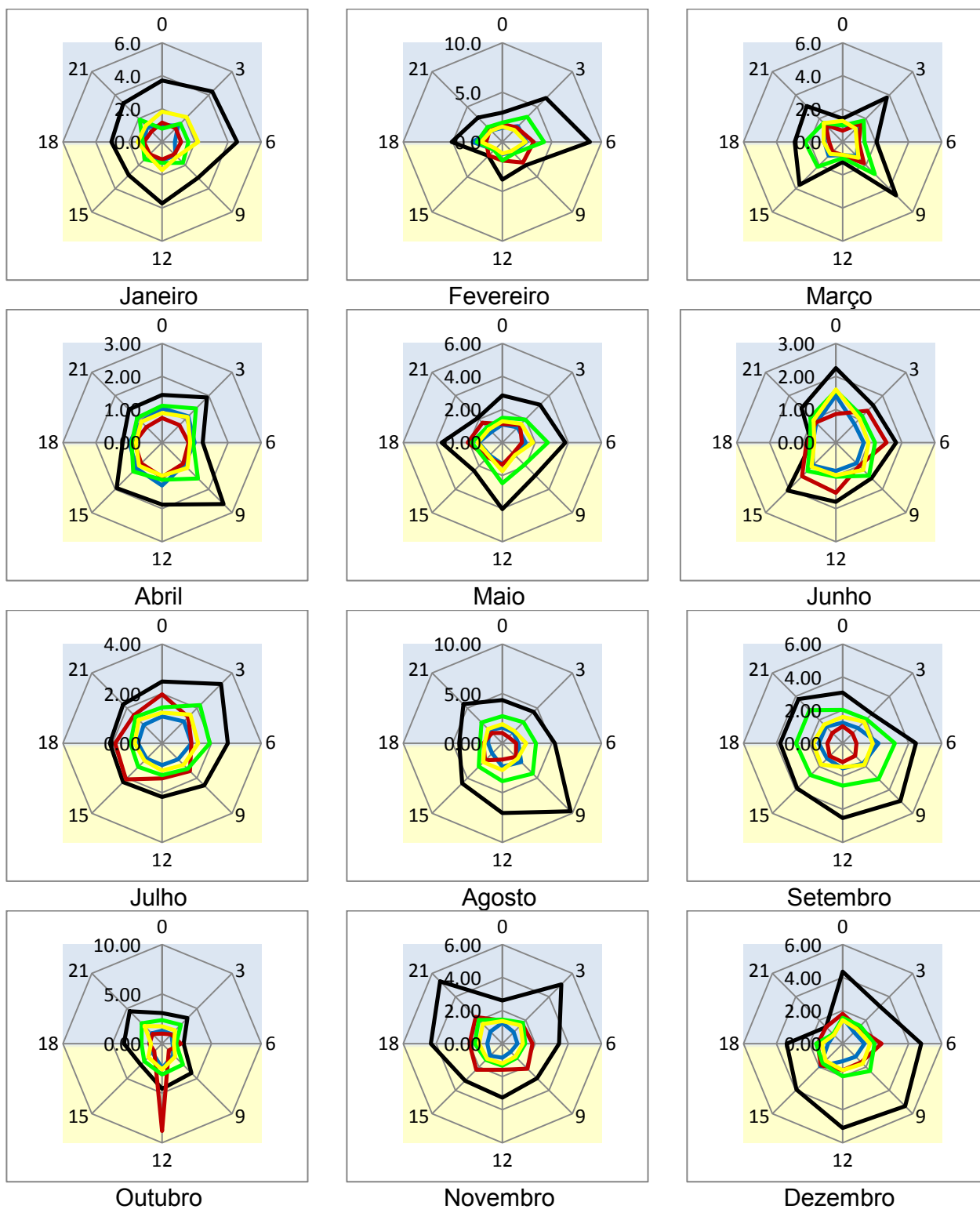
Legenda: preto: rutina, azul: narcisina

Apêndice XIV

Gráficos do ciclo circadiano - Casearinas: *C. sylvestris* var. *lingua*



Gráficos do ciclo circadiano - Casearinas: *C. sylvestris* var. *sylvestris*

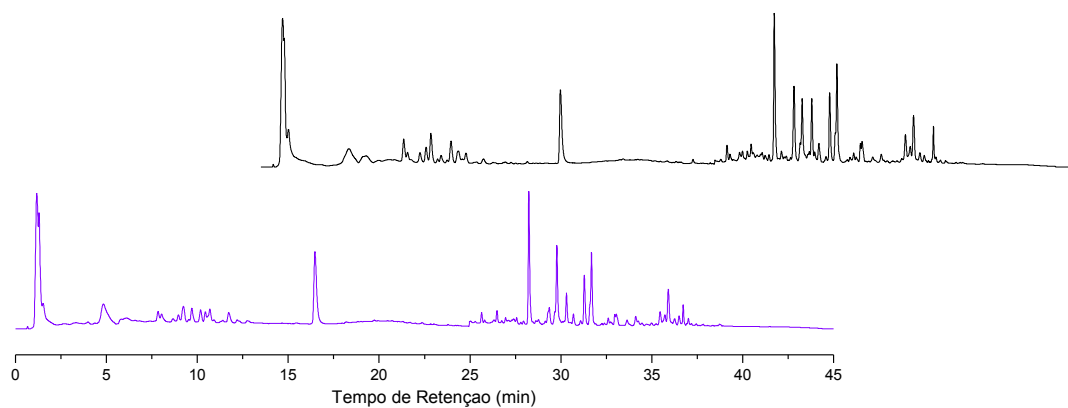


Legenda: preto: casearina D, azul: caseargreuiina F, vermelho: casearina S, em verde: casearina I, em amarelo casearina J

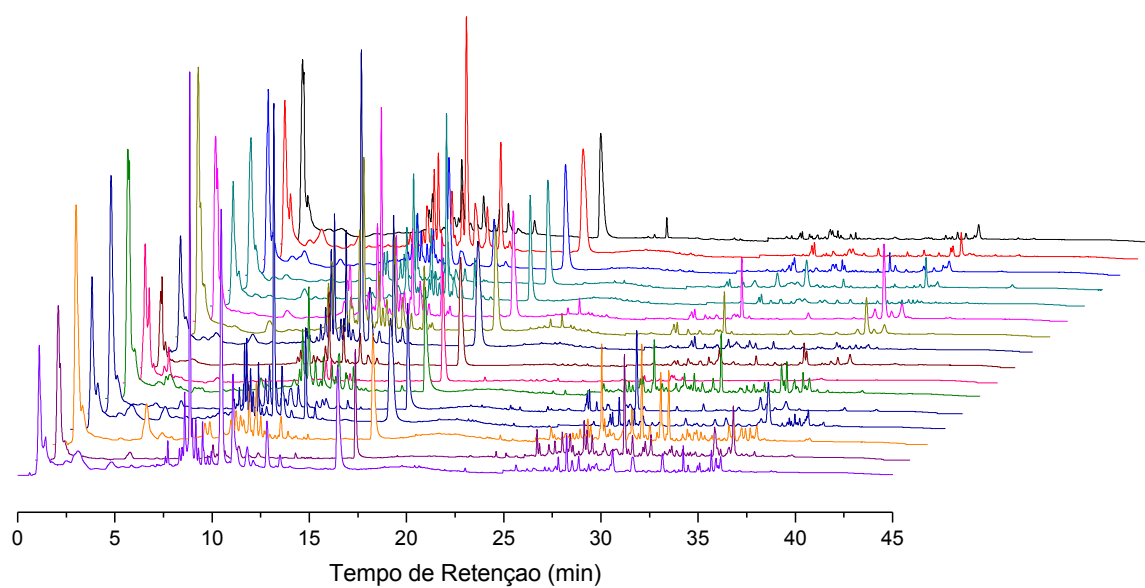
APÊNDICE XV

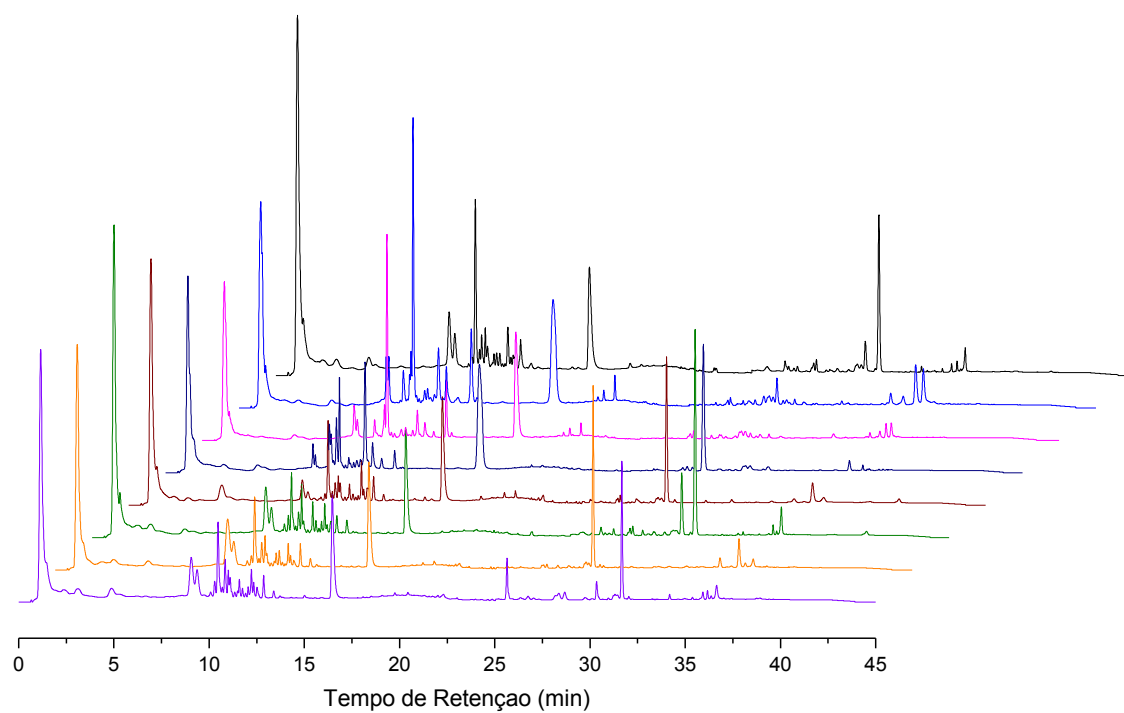
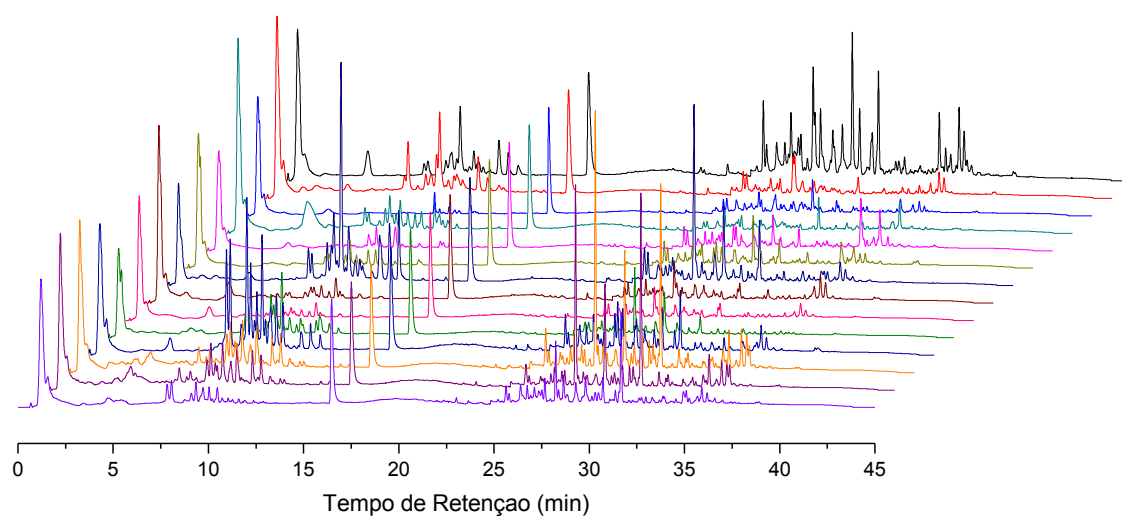
Cromatogramas das amostras coletadas no Brasil organizados por população

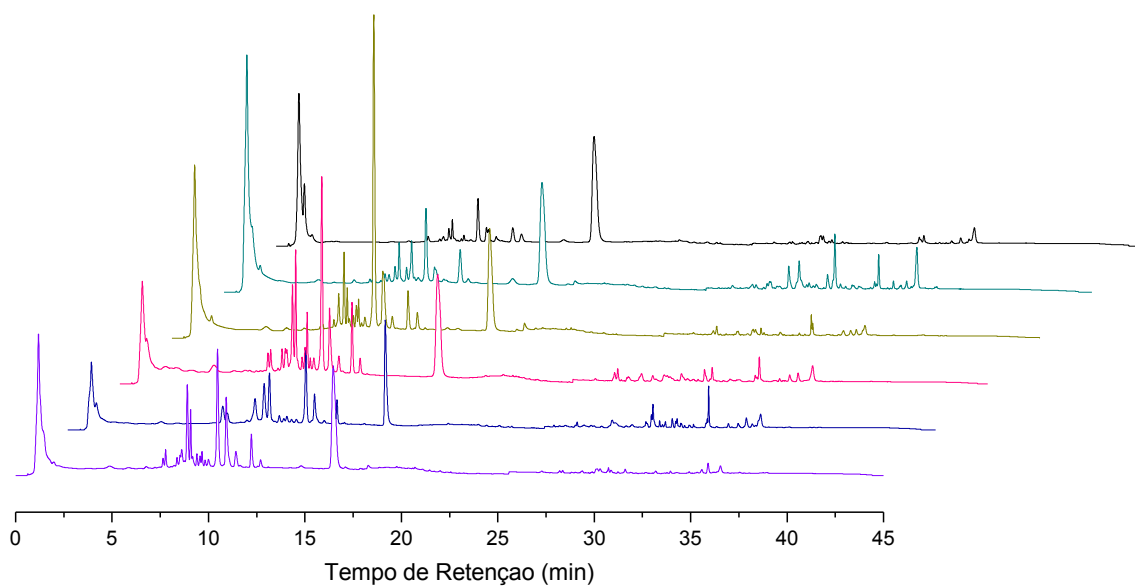
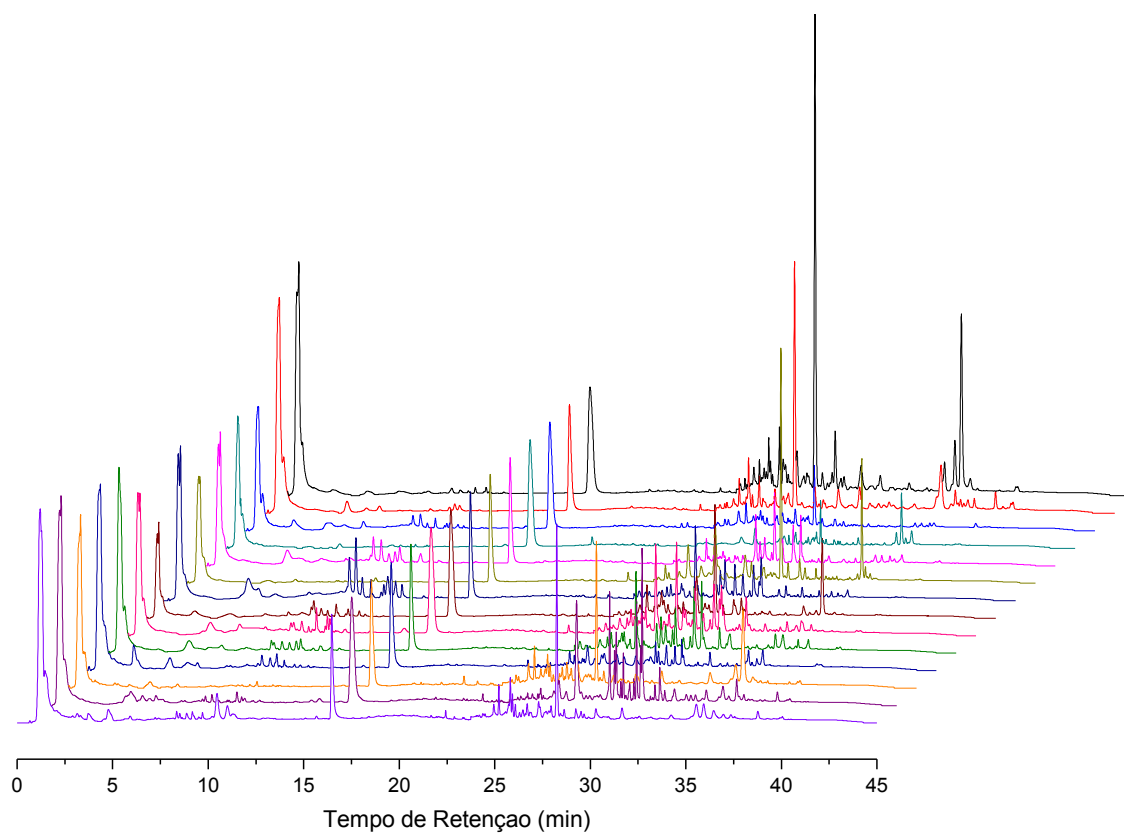
A) Angatuba/SP

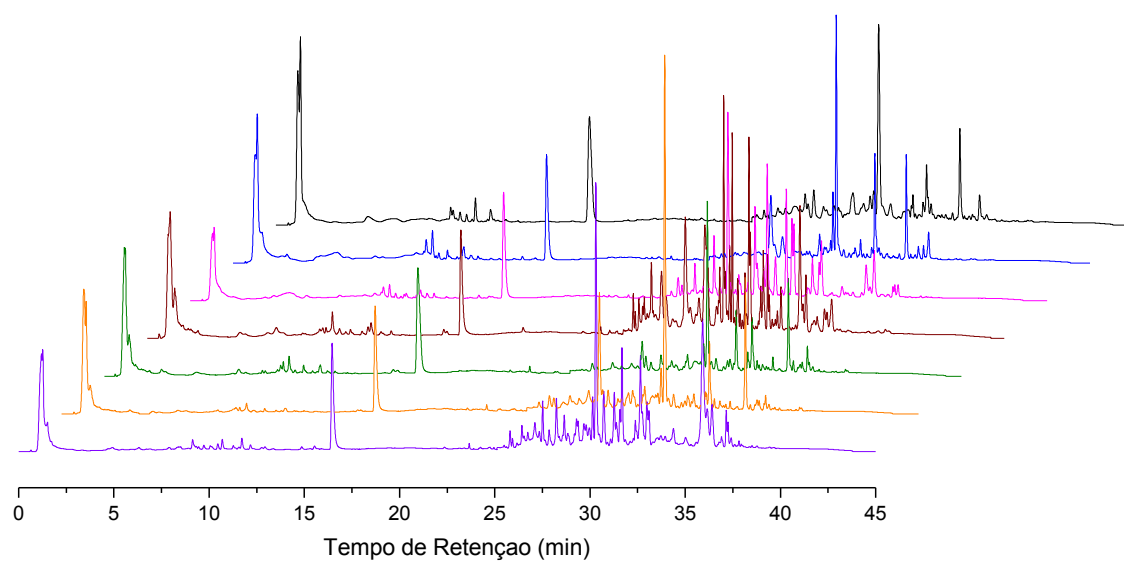
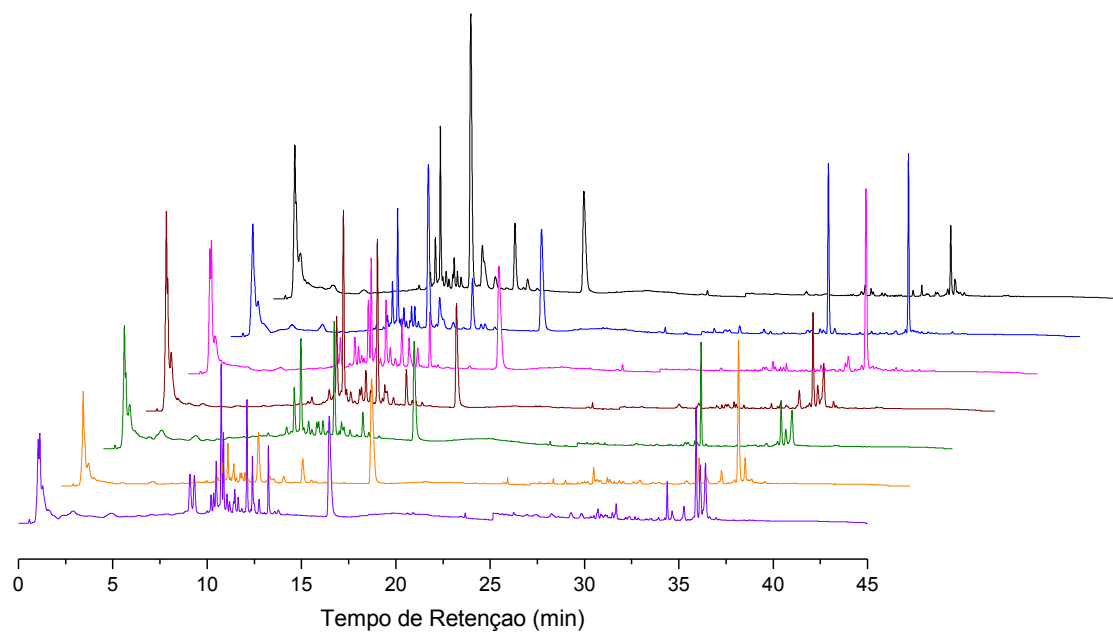


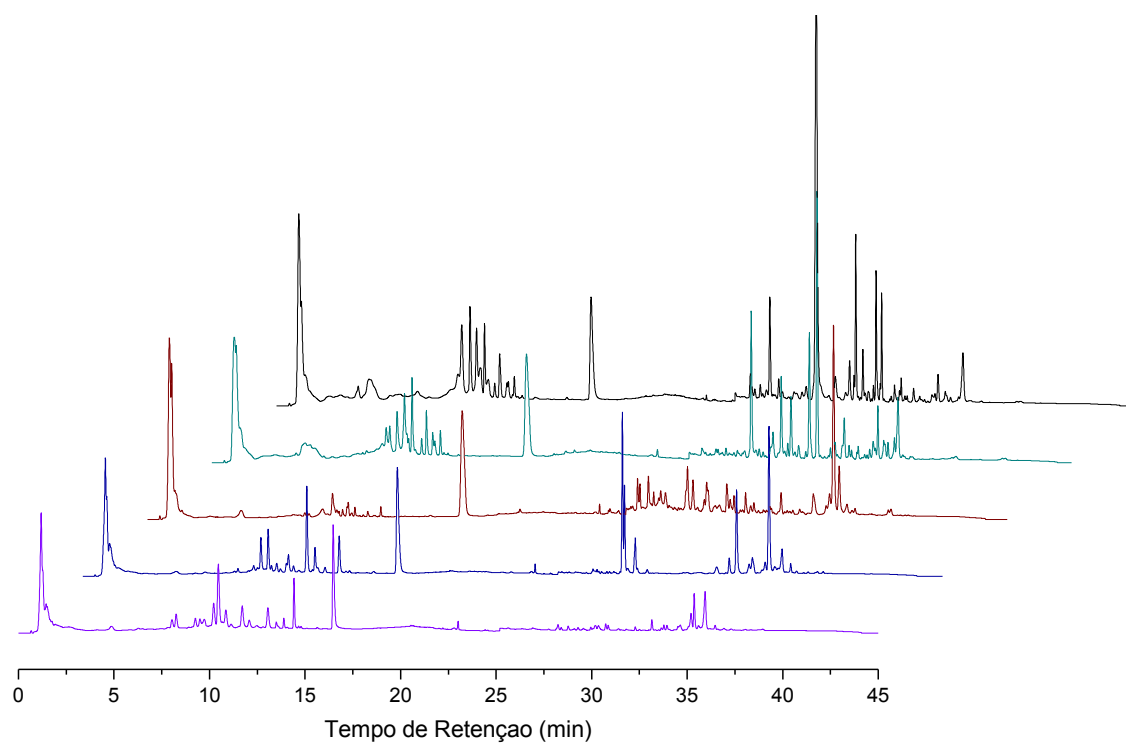
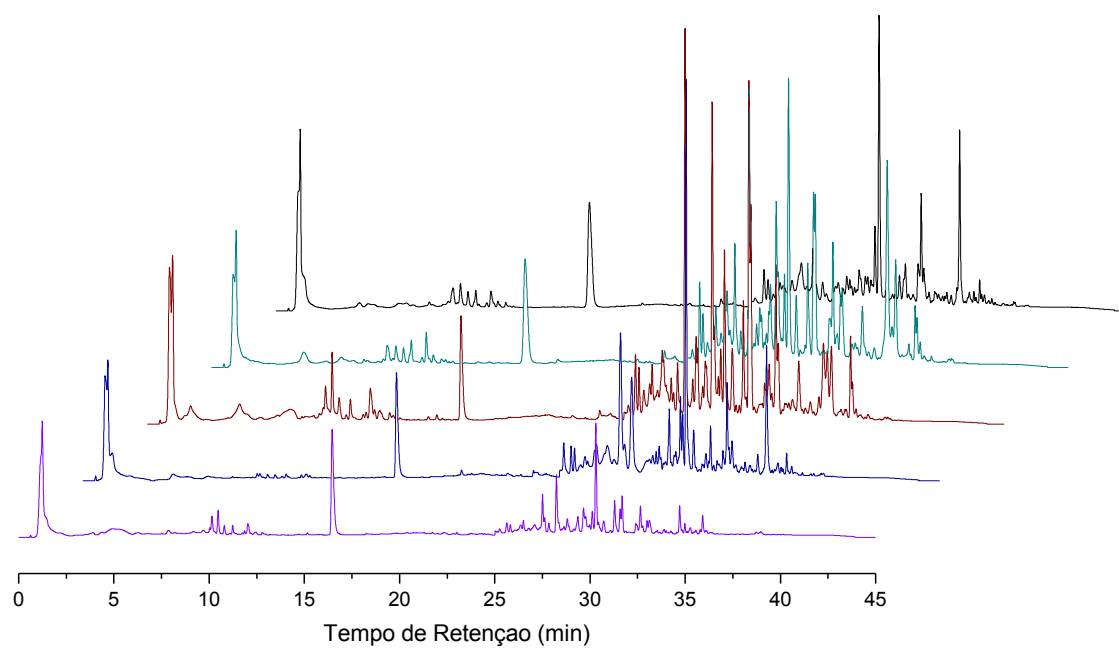
B) Araraquara/SP

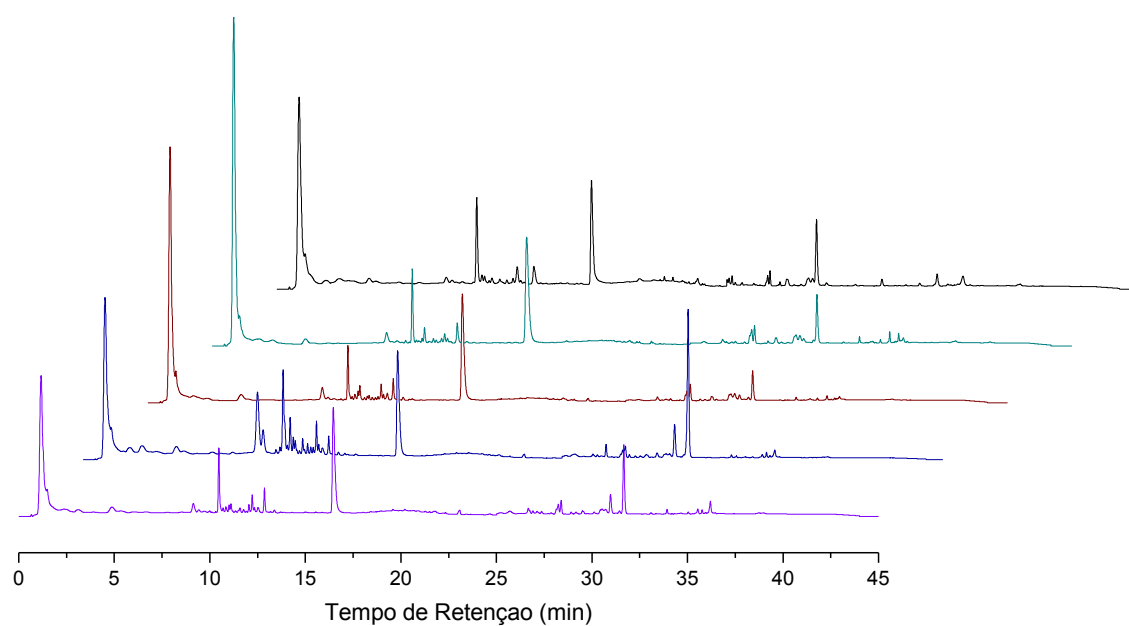
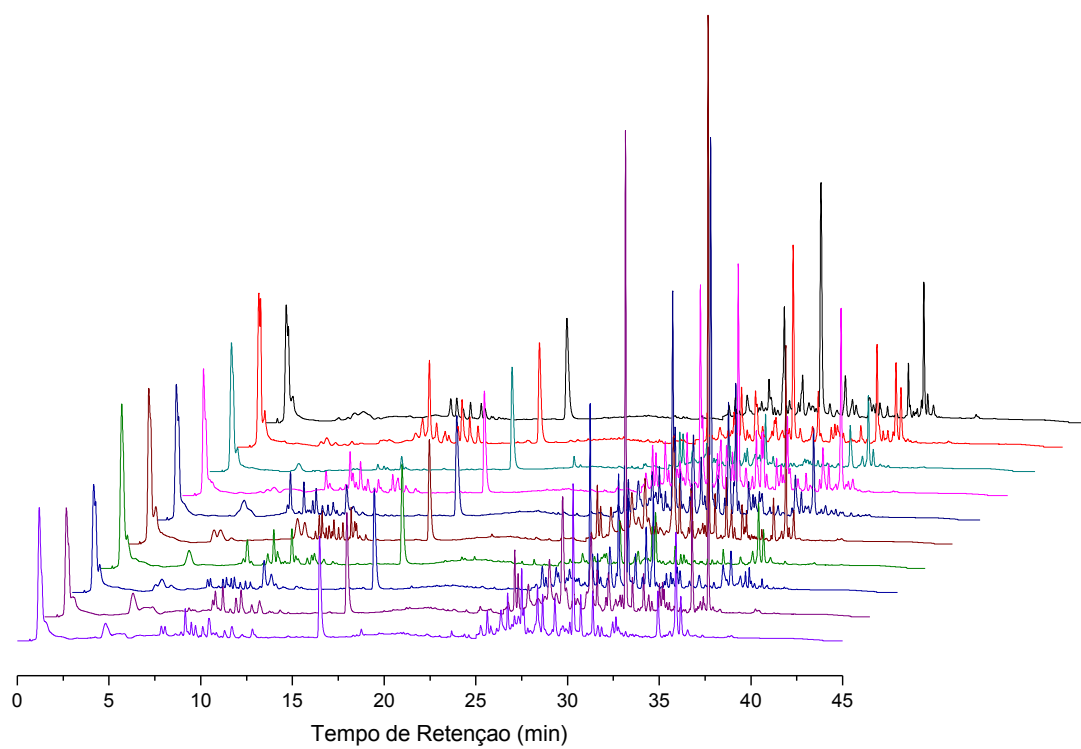


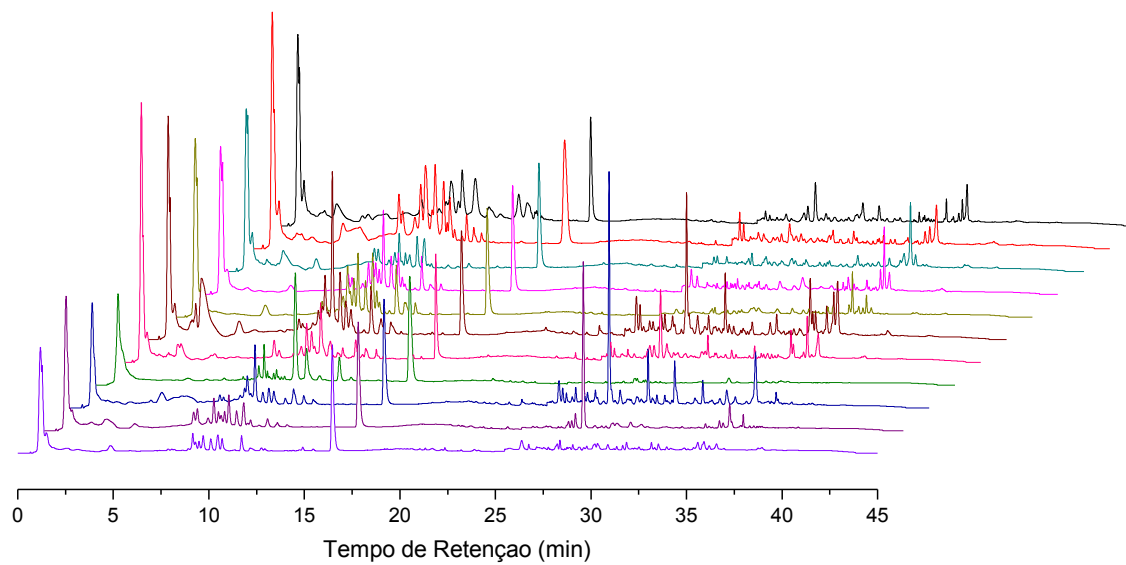
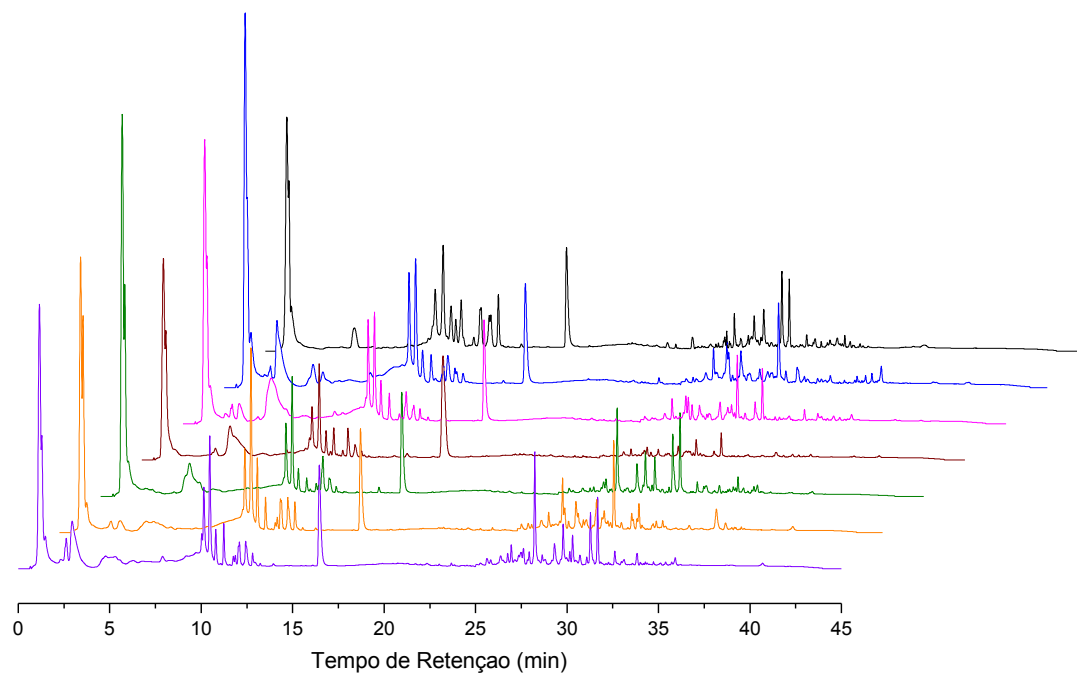
C) Cáceres/MT**D) Campinas/SP**

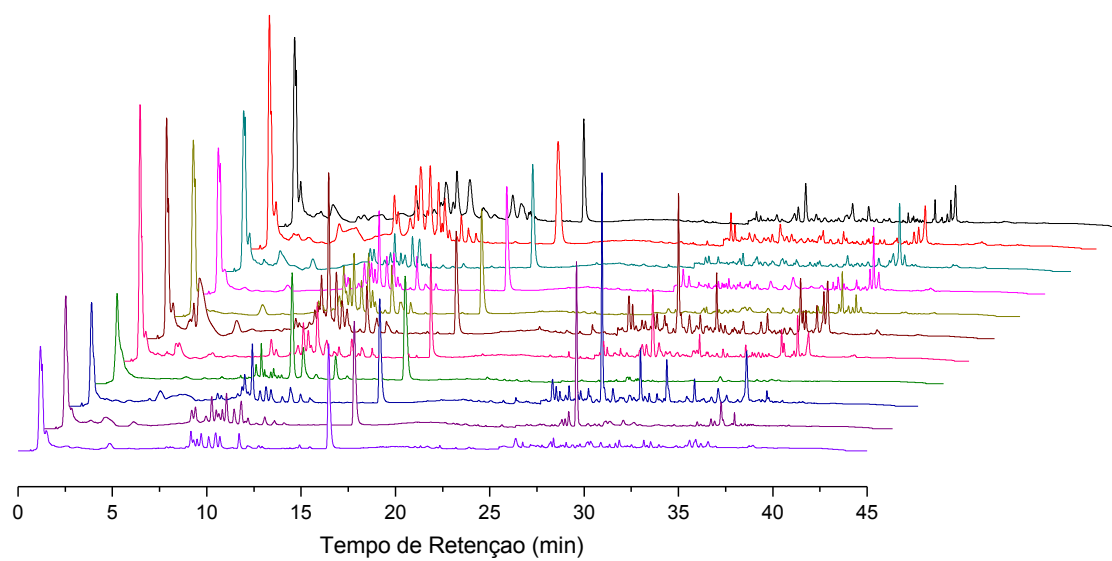
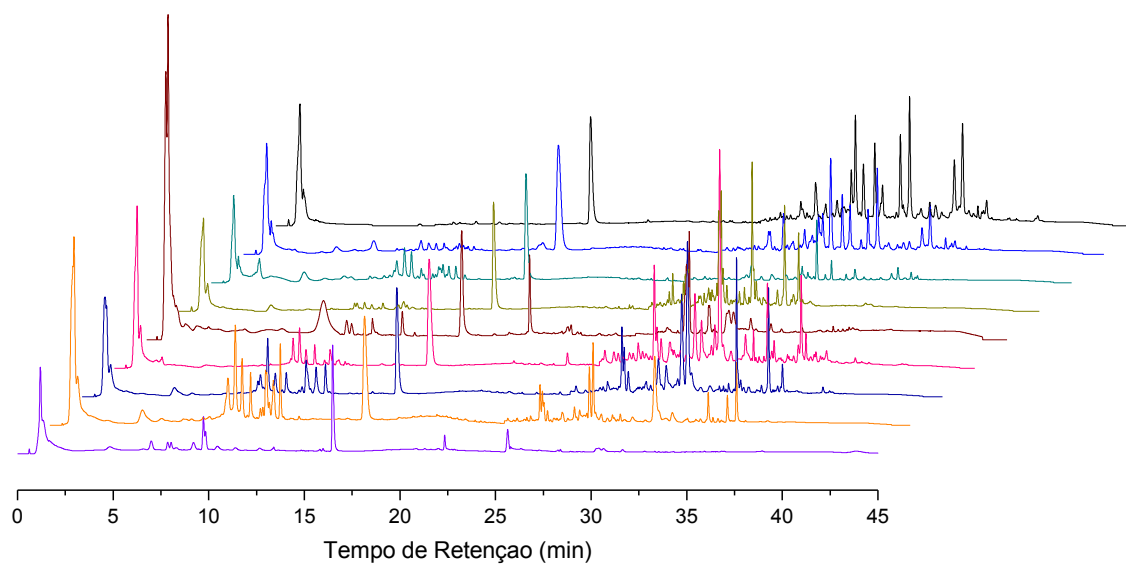
E) Cariri/CE**F) Florianópolis/SC**

G) Guaramiranga/CE**H) Luis Antônio/SP**

I) Mogi-Guaçu/SP**J) Pacoti/CE**

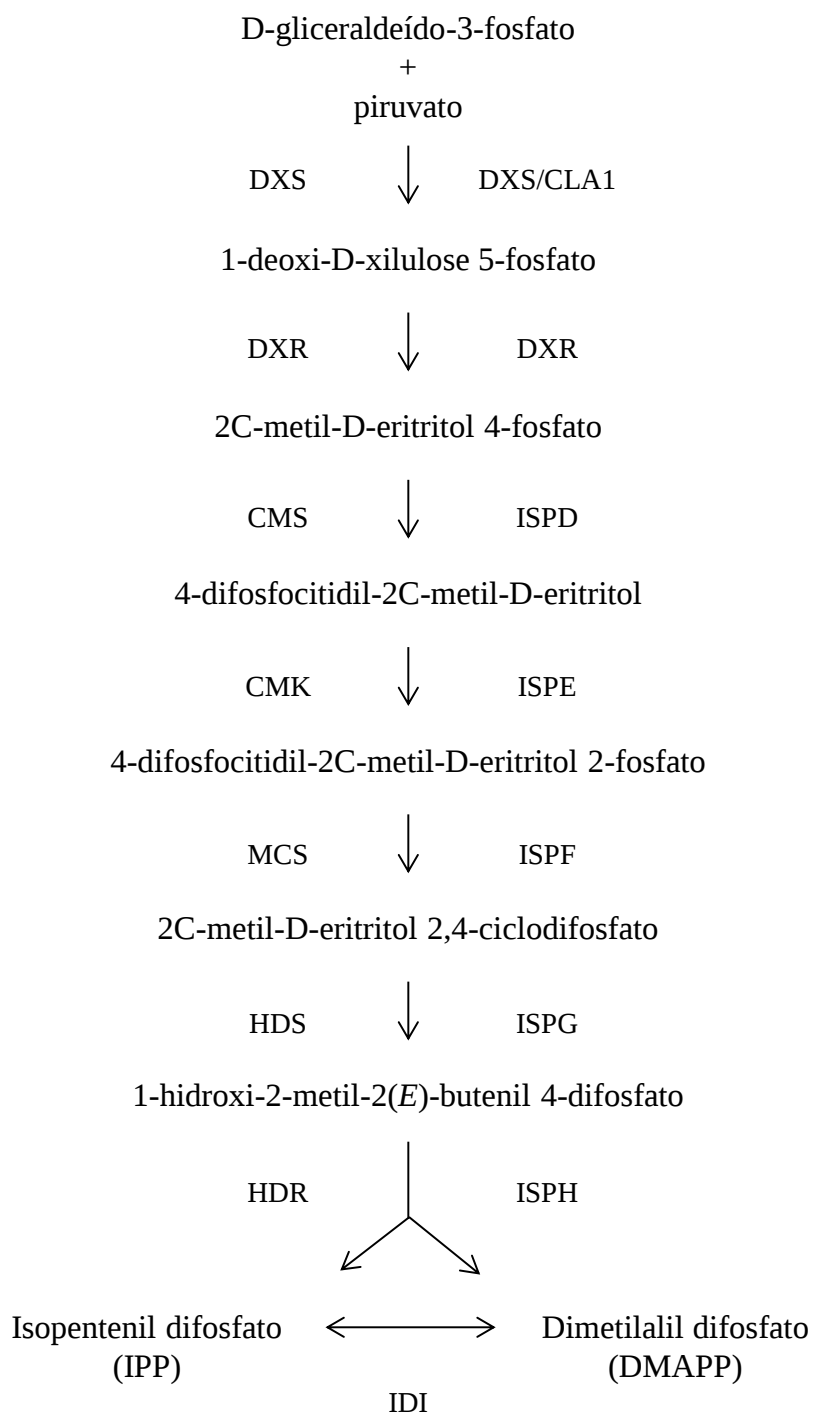
K) Porto Estrela/MT**L) Presidente Venceslau/SP**

M) Pocinhos do Rio Verde/MG**N) Rio Grande/RS**

O) São Roque de Minas/MG**P) Sergipe**

Apêndice XVI

Via biossintética do metileritritol



Apêndice XVII

Lista de proteínas/peptídeos do metabolismo secundário identificadas no estudo proteômico

Protein name	Peptide sequence
(R)-limonene synthase	SVVVTEPACPHIICLTNTQPNmLR
12-oxophytodienoate reductase 3	LFISNPDLVLR
12-oxophytodienoic acid reductase	GIPSSIELIETYRTWGRGNIGAIVTGNVmIDPNHIEVEGNPIIPPSAPFSGER
	LFLANPDLPR
	EQVEAWKPIVDAVHAK
	MSGHNLF SRLVTPALVR
	IDPALFNPLKRPAFLANNK
	GSYIFMQLWALGRVANPANLQK
	SDLMYEPTPNIVESSRKANGGYGFTAVDMFR
	VSNSSFQPNGQEPISSDKSLTSNDFEKFSPPR
	LSPFADYmDCHDSPEVLALRVIGLMNNLGVLYCHmIEPRMCVGAGEDGSK
	CRTAAIEALYYGVVQNPMRWEGGGLLLPIEVINIASPFIQGATATIYDPGTTVSEK
	WGFFQVINHGVPLELR
	IDRYSLALFAFNKGILQVPEELVDEEHPLMYK
	SSYHRVVVSSEEDRYALGLFTFLDGVVQPPEELVDDEHPLQYK
	CISMSLLVTLPSESTVSPLLPKVITDGINANEDEIREDEEQDNIHSVCK
3-phosphoshikimate 1-	TFPDYFEVLR
	LNVTEIDTYDDHR
	EANVDVELFLGNAGTAMRPLTAAVTAAGGNSSYILDGVPRmR
4-coumarate-CoA ligase	KFDIDEVIR
	IGTNAFDFVFLKYKIPVIPVTAAPVALIK
	YKISGLAAEAPVHELVEPAAAGLQMLVDKGmDVDFEAAPVEVK
	EAAMGLAmCGAGYTLFPVKFDIEPK
	RLGASANKFPQSIALVGILAMQEGSHLSFLGPVDILR
	QVNDQYADKHIVSDADLCVmDNDALPDFEVMTPSCVDQQR

	VFFIEAIPK
	ATAETLDSEGLWK
	AKYTQLAAVIEIK
	MLKSGGPSGLKLTDEEVR
	TDmDLGSQAWLCQDLFNPMLIAVGNLEVSILAVENDTPCmSPK
	FDLNIGSVWPVSFALKLLLPLLmNVmIGIGHKVAVTTTVLHEDR
	LIITQSVYLNKINDFAK
	L LSLNPVENFLT SVIDFQSG LIGYLVIGLVEPSVVHGM L L P K
	ARPAFVDASTGAALSFAGLRALSRL
4-coumarate-CoA ligase 1	AmYWLGN EYGLIIRYLCmVGIVAGADVENPTSAERWVSLQQWK
	AKLWETENVTDNGSPVMPYDTLVQLNGQVPSLGENKTDHWTVDGA FEER
	YKHAGFPmEEVNSLAVQSPA AVDIVAALMK
	IGVLIDPEAEASVS SAHVSPAFLAIKVGDELVAEIFAPVRSNGSK
	KGHVVIVVLPNVVEYAIVALGImAAGGVFSGANPTSHVSEIKK
	PDPVHASLAAA AVSEEGILLAESVFIEGEQVLDSVVELPPEGAAPSKmEHSAEAYNDIASYEVNK
4-coumarate-CoA ligase 2	MRMLIAVIAGGHSGFTLKFQHR
	VVYIIINHSAFAVSNVCTTIHNSR
	ITIFFTAPPIYLMITKSPDVKDHFASLVR
	KLYVMEPTETGALVVSLVGLTISGNHMMWR
	AELHGLEAmPLVVGSQLRNR
4-coumarate-CoA ligase 5	TLTAVYCNDLSGNTAAGSPATSLNEVTIFVLVIVLSTPIVAKHSVSAQDK
	IPmSTKETGSLGYAIADESLTNGLWR
	LLFWQVDVSGSQDDVFFAFIPmFHIYGMIFFGLLCIGITTVLMQKYDFQAMLVAIQK
	FRATYLPLVPPILVAMLNNA AAIK
	NIMENASALNTANLAANDNMPGPLNNDVTNIDVIAK
	GLSAAVVSSGmVDIERKCLIAGLPINSSLSMVTTVAITLVTGVVHPNLITPNFR
	PGIGAGLADSVLLVSVREVSQR
	NVYLAVLPMFHVYGLSLFAVGLLSLGSTVVV mRKFDIDEVVR
4-coumarate--CoA ligase 7	NDVVLFLAPNDIR
4-coumarate--CoA ligase 9	DHVTDGYDLTSLEAICGGSRLR
	IYFTHGLILGFNRSPFFFADHVFTDLPLVFHMPGLNGLVRIQATEFAGLmVVR

5-enol-pyruvylshikimate-3-phosphate	AGIVDEAMPLTPAEVALAYGFEIPEmALNARVK
9-cis-epoxycarotenoid dioxygenase	NEGAYCAASRTAETFRLGQER
	FGSEAVFVPR
	GPYHLFDGDGMLHSIR
	FGNLGADVGNVATRmSADQVPLLDYER
	FTRPRAAAAATNSVLSAPSSVPPAYVPPPPPPPTK
	SRDLDTLHQPTmIHDFAVTRNFVIPDHQVVK
	RHEVISYCAHPPSQLFSPPLNHKPTVFSVITPDTDTK
	VTQYYKYLVRAADPmVVGFSAGFLHGAVGNPR
	PVFPKHGGEAHALSIIGRFALYmLR
	LAAAmFmTPMFDIPDTADRILGYmAGHCGGNLPGDGHTRLDR
abietadienol/abietadienal oxidase	TGHLPSLCYNVGNQIPMEMQIAISNQIVLAMSVDGVLLSFHSSVAYFEKmgLKSVMQMLK
adenylate isopentenyltransferase	FGSmINKIDSTEVFMAILR
	EAIGKTKHNTCILVENQLLK
	YKSLPYSWCIDLFCFVmLELK
adenylate isopentenyltransferase 3,	VmCRQLTPPKNPSSVQSNIIQNISGR
	EEQRGVPHLLGTIKPNmDFSVNDFCDTSSEIDSITR
adenylate isopentenyltransferase 5	IHKWYTFmNGGmEK
allene oxide synthase	DPRVVALDAASFPVLFDTSLVDKTDLFTGTfPmPSTDLTGGYR
	STVFRPIPPSVNFVVNAFVLPFVTNVNDKDSDFSDVLLMHK
	LFYTFYLK
	YDSFEIEVGK
	VQELFVYEINER
	EGEMLFQFPFATK
	LDYFYNQGRDEYFK
	SNFGALLDTVQSQLASGGGK
	EYFGPEGQLMMATKDPR
	DLFTGTYPSTELTGGYR
	TVVKQEGGVTIQSLEKmgPLVK
	RTDLFVTSSYIIDSGGAETSSLLKR
	IQKYQSTVFRTNmPPGPFLAPNPmVVLLDAK

allene-oxide cyclase	RIILCGTIIGETIHK
alpha-humulene/(-)-(E)-beta-caryophyllene	VVEGTRITEINVYSMK
anthocyanidin 3-O-glucosyltransferase	IAHNLGIKSLQCWIVSPATVSYNASPSRMCEGTNLTEFDLMK
	PYEmmRmKSLFK
	FFHLYHGPMLLATAMALNKLAKK
	LLHEIQKQYWEQFLPVLNFPFILTNDQKLGK
	LFGPGAQPEPTNVQSPFMDKDIISKLFQQGTIFLQQIEAYLR
	HRWSIGVLQAELEAPMTPLNVWTHALDAVCSVLVFGGVAVDLK
	PTHVALLSSPGLGHLIPTIELGKR
	LASELEIKSVRFVSFSAISDSCINVPSR
anthocyanidin 5,3-O-glucosyltransferase	EDKEEEVVEGVVLLDLGRMK
	GMVmKPNQAWVKTAImSPENGWVYVHLHVLVGSQWFDVGAEPWGCSSVR
	FLWVVR
	VILVEEMK
	MPANIDFFNLVDK
	QECIAYVMFADNTGIR
	IPGLPPLSPEDMPTSLDDR
	EKGVVVKVETHVQTGHPVEEVDLVIFGSGGWNSGWSFLK
	KDDNECLSWLDSQPSQSVLFLSFRSMGR
anthocyanidin reductase	VILTSSAAAVSINK
	FAEENNIDLITVIPSLmTGPSLTLDIPSSVHLSMSLITGNEFLKNALK
	EWDVYVHDSETGLTNMATKHGPGYAPSWK
aristolochene synthase, 3'-partial	HAWWLYGYDAINSSSPDEAIGSSRFCmMEEPGR
beta-carotene 15,15'-dioxygenase	ENGDTFQVSHWFDGFSETHRFQLVAPDATHSSLR
beta-carotene 3-hydroxylase	FTYLVAAMSSFGITSMAVFAVYYRFSWQMEGGDVPWSEMLGTFS
beta-carotene hydroxylase	FDLPmFGLYVLGCTAAMVLGLGHHFGGIGLRRNPVGIEVFYFPPR
	RCSGTVSmTFLEDVK
	FSWQmEGGEFPALFMFGTFALSVGAAVGmEFWARWAHR
	DGPFELNDVFAITNAVPAMSLLAYGFFTRGLVPGLCFGAGLGITLFGMAYMFVHDGLVHR
	CSGTVSMTFLEDVKEIKEEATQVQPAIavgqvvesqek
beta-carotene hydroxylase 2	EEFPIWAVFGWGSASEmGGFAVMVFALQLGAETRWAHRELAHHHAHSSLmWWHR

bifunctional 3-dehydroquinat	LLLSTVTQAQVLGDPTTGHRmDVIVILGYGGVKS SVVGMSASSVmSSLPGVK
	AFQGLDEDVEISDR
	LFVVIGAGGAGKALAYGAKAK
bifunctional dihydroflavonol 4-	LFLHYFLPEVFQFIRGFFTIVGFImVDGVK
	VLHSHIDGGECYGEAILmVDVQYNRDVCNLDLLEYTLAMDSASAR
9,10-9',10' carotenoid cleavage dioxygenase	TMSADPVAVVVELPHR
	IGDLKGFYGLFMVQmQQLR
carotenoid cleavage dioxygenase 4a	GPYHLFDGDGmLHALR
	YKIYGFEAALLPFVQSIGASNNILNGERVAVASGR
carotenoid isomerase	LLKGEKQLEFEQEPNK
carotenoid oxygenase	PNRIEADIFELETSGEIPGTIDGVFFR
	AIFNLLDLIDADKGADII EAHDPGWGPICCANFLLGHmVGHLRSAR
casbene synthase, chloroplast precursor	ACANWIKKEENmDCMK
	DRMAEIFFWAVAmYFEPDYAHTRMIIAK
chalcone isomerase Chalcone--flavonone	LLSEAVLESMIGK
chalcone isomerase 4B	VIEFFQSK
	YEEEEEEALEK
chalcone isomerase protein	IVEFFQSK
	IKVENANVVEMIK
	IKVENANVVEMIKK
chalcone synthase	RICHKSGIEK
	MQATRYLLSSSVIPEVmGELmHFDSNR
	PVPHFPLAGEGmLCTMGTEMIGSFVWKEQ
	LNDAANDATTEDHGEHARPPPLSPIALACFARCCR
	ALVDAFQPLGVSDYNSIFWIAHPGGPAILDQVEQKLALKPEK
	ALLMNILIAAGGEEAALDDSRQVLRLVEESIVTGNYGTFLmNR
	EVIVMASRQAQRQYVFTCPGPTYGAADPNYLTINADSAPER
	AIAVNSGTSMGKAmpsAQPNGSKTGQQGDLMYVMVAPFLLFAVHDAIQR
	LSMCVFEmAIAVVRVPGHHASFDSAALGDEGLASLADASPDEGVEIVDGR
	ELGLSLYIYSAAmYNGVENVMDET IKIRK
	LLTEVIEAGRVDQTSFLK

	ALFGDGAAAVIVGSDPLPEVEKPLFELVWTAQTIIPDSEGAIK
	EVGLTFHLLK
	MSDFLAYAIK
	YMYLTEEILK
	EQWGPKSKMDHVDAGILYPSFQGTTVLTCTK
	IEILGCHTLYEMTKADCNPSLAEmYAPmNR
	KDALGSQKEEHISLLETGIGLVGVTFVGPLGVGFTWTA
	ALEEWGGDAATEITHTVAVTCTNQGNPGYDLMVAQRLGLGAGVDRTLLSGVGCAGGLAILR
	LPGGDLYLANELGLKSDVNRVmLYFLGCYGGVSGLR
	LGmGMAAAWLGTGIVAVWLFGDGSGAVVVGTDPRER
chalcone synthase Naringenin-chalcone	mVSVAEIR
chalcone synthase 1A	NNIGEHEMAPSLDAR
chalcone synthase 5	EDVVQmPVMKLGKEAATK
chalcone synthase 6	YLMYNILEEKMYAPAEAVSDLPNSGRVEVvmPQVDK
chalcone synthase 8	EKDEDESQQHEWGVmLAFGPGITETIVmR
chalcone-flavanone isomerase family protein	GSQYGVQLESSVR
chalcone-flavanone isomerase 2-B	ILPESAlmPVTLEEVGPA SLKTSLEARFLYFEK
cinnamate 4-hydroxylase, partial	MDLLLLLEK
	GQDMVFTVYGEHWR
copalyl diphosphate synthase, partial	TLYRMALALNNENNYNVK
	VGLVNKIKSILNSMEDGEITISAYDTAWVALVEDVHNSGTPQFPSSLEWISK
coumarate 3-hydroxylase, partial	FLEEDVDMKGHDFR
	HQGLVSMPHIWAVGLmLGT LHENFLTCCPPGKVSEEGTDMIEYNMLPPRPAAQVPLTTR
cytokinin biosynthetic isopentenyltransferase	LSAYLSKQVQPRNFQVMNLGGK
dehydroquinate dehydratase/ shikimate	LVDVLEKPGCLAITQTLVATLHVLDfSmIIAPR
	PGCLAITQTLVATLHVLDfSmIIAPR
	MTNLACPLGSAIAmVARmQQEVIVVES AQMDVAEGSDAARTIIFPQLES DLNSQDIR
diaminobutyrate--2-oxoglutarate	TATGRSTVLSFQGAYHGmSQGALSLmGSLGPK
	AIGGSLPLAVMVYRDWLDTWAPGAHAGTFRGNQmAmAAGSAVMR
dihydroflavonol 4-reductase flavanone 4-	SAWALAMDRGVNMVSINAGLLmAHDL SVKHPYLR
dihydroflavonol-4-reductase	IVFTSSAGTVDVEEHK

	LQIFNADLSNPESFNAAIEGCIGVFHTATPVDFELKEPEEIVIK
	KDCEFDAPVGRYLHPLR
	LNIFNADLSNPKSFNAAIEGCIGVFHTATPIDFELEEPEEIVTKR
	YLLTGENASFNQVFDmAAVITNTSKPmVSIPLCVIEAYGWLLVLISRITGK
	GVNMVSINAGLLmTHDLSIKHPYLRGAEmYEDGVFVTVDLGFLVDTHICVYEDVSSYGR
	FDLLMLQR
	KLEDLGFSYKHGLEDIYQTLmCFLDYGYLPPV
	IAVQAASEGVPIVLLYPGVYGPVKVTAGNVLARmIVER
	GTMNVLTAKEAGVR
	GSGMSAKITVVYVWISGCAASGVKFLLLR
	VFSHVCDHQAAGVGPLPCQIKAELLDIPKmntTAGVALK
dihydroflavonol-4-reductase 2	MTGWMYFVSK
ent-copalyl diphosphate synthase	AAYLYVMPmELNDKYELYCWQHNADNMDK
	LQNENVEHMPIGFEVAFPSLLDmARGLDIEVPNNSPILNKIFAMR
ent-copalyl diphosphate synthase kaurene	DGEFFCFVGQSTQAVTGMYNLNRASQISFPGEDILQRAR
ent-isokaurene C2-hydroxylase	RICPGINLGVANLELPLASLLYHFDWK
	ALSEKHGPIQLWmGEVPAVVVSSPAVAEEVLKHQDLR
ent-kaurene oxidase	MLGGGRVATRTPGGSPFAPLALPLVTLYIAPEPTLDPTSHDYNTR
	RLNTVTGPLADEASFAINK
	NIEAAGSCVPDPTKmLLFK
	FFLLSVMQHQRLLYLLGHALCTAIVMGR
	QEASAALAAHDGAFTAPALHTmKKMDSVMK
	LmPDTGEEIAVLTRAFFDLWNLHmASINR
	ELFAGYSGFEPIAAIADDANIIQTVARLNITTSK
	SALTYISILLIPIVITALSSVAALRGEKVPNVGPVPR
	LSGWATLRYSPNGQMTGSGPIVANLTSGNFFIDAPLAGGEVVASKFADR
	VEPVMPLTLALmSFNSTGANTGPFLSLLmKHTNIWAEEGWGGPmSmSTLALVNPAMSVDAK
	SLVNDNWDLISHIPVKAPK
	QYECNmDKNQWENPQQWMPK
	MVILPGAIPQPGFmITDPLGGDSFENRR
	DLLLAQWKFSAIMTVVPQGAGLVLAALQSIFQETYVPSAVILGRK

ent-kaurene synthase	GLRFIELNSASVTDNEQHKPIGFDIIFPGMIEYAIIDLNLPLK
	TVAFVFPFGPEVDVLHTMYLIPLAATSGPYLSKLPVYELGESELLNAHK
	LLDEEGISFQFEGKATLMEINAAKmSR
	VLLLYCYYPNTALLVAPELADARLSLAKNFVLATVIDDFFDVGDSQEELENLIQLIER
ent-kaurene synthase 2	QRLMATCEGDEQVDFIMRMLR
ent-kaurene synthase B, chloroplast	LINDRVATQSDGAQ GK
	TSYRFCNVDERYLLTFACQDFNICQSMHQK
ent-cassa-12,15-diene synthase/	VEAEWQRSQHLPTFEEYmESGMVSLGQGATVmSALFLIGEKLPEGVVELEEYDEMFR
	YGDFAPDNNSVYTCVESmEFPmLLVQSIIGILYLQAKSLSYTmSGWDLIKVEDTIAmMFTDCAELR
	VHTGSMSLDGMGVSASVSAELNK
ent-kaurene synthase like 5/	LTVQGPYCDmAEPQPFVLSNGYVmRYmGSDELSAISIEFITmDCR
	RYFGNDmVVQQNWLGESQAmAYLARK
	CYDALESHDSEYEVITGGIWSFFVNKQIAEK
	GMLTEAEWRmDKYVPTTVEEYMSAAVDSFAVGPIITSAALFVGPELSEEVFR
ent-kaurene synthase 8	KAMRSMANAVILEPLHLQE
ent-kaurene synthase like-2	LVHKRNIEHFAmSYISYAQNNIPDELER
	SMMTEVEWRISNYVPTPEEYmENAAMTFALGPVLPALYLVGPK
farnesyl diphosphate synthase	CSGAVHLMIEIYFQDDLDDWWLTRRWFGCQPR
farnesyl-diphosphate farnesyltransferase	TMADVYGAFYDFSCmLKSKVEMNDPNAEK
farnesyltransferase beta subunit	PPGDDACGTSSSTEAAYYAKFGFDFIEK
	APPHETVLAQmFVSKQTTINETPTIQEDELNLTLAVKPCFLLK
flavanone 3-hydroxylase	LLEVLSEAMGLEKEALTK
flavone synthase II	DLCLRLSLGPLYR
flavonol synthase/flavanone 3-hydroxylase	MLNVELGSQLLILNFYSPCPK
	PHYHGCEFWSEAVNTRDYLK
	YMTTDMRAAVRYGTSFSQTK
	EVVQEYNKEmLRVTDNILELLSEGLELESK
	ERYSVAFFmDPPSDCVVECFESCCSESSPPRFPPIR
	VPIVPDTLLINVGDMQMEImSNGIFQSPVHRAVVDSEKER
	TRISMPFVNPAPDAVIGTLPEVLENGEEPHYKQVVFSEYFNYFFSK
	LNQLPNLANDFPRETLK

	NLGIDPEDVATmFEDGVQGIR
	MAELMSHGILKALKEDMIEmK
	TQNYLLDNTGEVQFTQIIRDWIVPGNK
	NEFVAVMVNVDmMYSPNIGLGPRIHESRTTVASNK
	ILRAGIFELLVIAADLEEGWDSGGAEMARIGGPVYSIK
geranyl diphosphate synthase	GQLSDRAAFENVSQDDELQEPLLSERLR
geranylgeranyl diphosphate synthase	RFCAIVATAAASPVPAAAAAATGFDFNAYmGEKAAAVNR
geranylgeranyl hydrogenase geranylgeranyl	VALVGDAAGYVTK
geranylgeranyl pyrophosphate synthase	PmINHVVYGPNMELLTGDGIVPFGFELLARSDGGENSER
geranylgeranyl pyrophosphate synthase,	SVAVELFSETTLEAKPIDFFK
	EVLIFHHKLGGIVESGAMETEDGVIAEVAVGR
	AIGELAKSIGTEGLVAGQVVDIDSEGVANVGLETLEFIHVHK
	PVLCCLAAACSLVGGSEATAIPAACAVEmIHTMSLIHDDLPCmDNDDLRR
geranylgeranyl pyrophosphate synthase/	ISSTLAKSCRLGVVAAYETQAAGDLPAAYSR
	GFCDDLSEGKFSLILHTLNNSTADRIR
	ALLQTMHNDLLTELHRLEASFGLKNPILEGILTK
geranylgeranyl pyrophosphate synthetase	QQTKNQLVGR
	LSSISPNTFLQSPNVIPPTSSSGIHSNHIAPSK
	DANLSPSHLMVDNLDHIMRRVNSIGGTATSIGSHAFTQPAIYK
	IHRHDELADAGAICTGDWGRIMK
	FSYILVHLAHSRPDLMVQVEGIFK
	PTTLPIEAVRTIVRDANLSPSHLMVDNLDHIMR
geranylgeranyl transferase type I beta	EQAKYEVIRCAQGCIFYSASGSVHGGTLFGGmGDVSSER
geranylgeranyl transferase type II beta	LVKFILSCQDTENGGISDR
geranylgeranyl transferase type-1 subunit	NVLAEPQITLSSLDALFAAAGNAVLSGFENYLDACAGRVATLR
	GGSHFFNKSFNTEGIQPPRK
	FVYCAAVIYFmLDNWSGMDKEmTK
geranylgeranyltransferase	VVmGEDADPESYMKFADREGmGDEDNDm
homoprotocatechuate degradative operon	mQQLLALLDELK
hydroxycinnamoyl CoA shikimate/	PTPPPGYFGNVIFTATPIAVAGEIQSKPTWYAAGKIHDSLVR
	SWIVLCGNPTRWFAGLDHDGPIRLGmYTEImVESGPLmPQIIGDGAQASISFALPSNHFFLGPSEK

	QmGHNALSLVDMHCFDGTGFWHGSSAGAGVRPIILLDDGPTFRTLLR
	LRQLIPAVDYSAGIETYPLLVLQVTHFKCGGVSLGVGMQHHVADGASGLHFINSWSDVAR
	GREGIFVGNDNSNCFAGEFLVGEMAVITDEK
	SLWSSNLDLLVPMFHVQTVYFYK
	SNIFSYPLLLLQATLTKCGGVCLGVGLHHILGDGTSAIHFINSWSEIAR
	FFILQNASLTLLPYSK
	MIPPLPKNYFGNALAQAAAKGYIGEITSTLTILF
	ISTEGVNKAKMANWGDMAVSmFTSPTAYNGNYFIPGEK
	LLSDIDDQEGLR
	FTCGGFALGLCmNHCmFDGIGAmEFVNSWGETARGPLPCVPPFIDR
	QGEPTLVPPAEETEKGLYFLSNLDQNIIVRTIYCFK
	ITHFLNNQLmmPAGGmLLEGDARQWSGPIQAAPVLR
	PINTIDImPVMRDSLAKLYPLYHVGITIR
	IIPINmVDTMRDSLAKILVHYYPITGR
	CASKARELDEK
	PIFmGPGGIAYEGLSFIIPSSSTNDGSMVAIALPPEQMKVFQELFYDI
	TLSDVSVKPDCHFKEFK
	EAQKLGVLVACSPIDAALVATAGSDDR
	DADQRLEIAYSKAAYmVPPALFR
	SPADRDMPAVGVLSGFPFLQTQAIAALIFDLKFDYLYSNmYLLE
	PGSDSDNAAAAAAVSIFRLTREQLNTLK
	LPCTTNWISGRWFAGLDHDGPIRFQGGPPNIFILTILYSSSGGDSDImSMIVPHGAAPAEK
	LNLSDSGRmELDCNAKGVTLLAETTK
	METFSFHLTSSQLNYLQAQIWGTSLDQTPPFESLCAmIWRCmARIR
	PDFTDISSDSFLGAKAIVEPFHQVIEQTPLYSLIPLQELDR
isoflavone 2'-hydroxylase	DEIDNmVGQDRLVDES DIPKLPYLQNIYETLR
	GLGLLPFR
	QPLHHTFHTLSQKYGQIFSLWFGSRLVVVVSSLTIAQECFTK
	DNLIHSEQQYYNLQPTIITTSMDQQKNLWSLVNLEGSSEmVGAITLPATLmTMDVHmIK
	TLHALSQKHGPIFSLRIR
	GLALAMLFGGTDSSTGTLEWSLSNLLNHPEVLKKAK

isoflavone 3'-hydroxylase, partial	LVDEQDLSKLPYLQNISETLRLHPPAPLLLPHYSSDCTIGEFNVPK
isoflavone 7-O-methyltransferase isoform 2	TKTmSGINVmLSDLRER
isoflavone reductase family protein	PESADNPAHQELARAGATVETADFLDVQALAERLAGADVVISILPIK
isoflavone reductase homolog	ILIIGGTGYIGK
	NPTAFAVLGRTSSES DL PK
	FFPSEFGNDVDRVHAVEPAK
	RFFPSEFGNDVDRVHAVEPAK
	ENQIVLKmIAIERDAGVNK
	GPFS CNFILEFYMEIATVRLPPLPDGAQLLIPSAR
isoflavone reductase	SYGVTLIYGDIDHESLVKAIKQVEVVISTVGGHLHIAEQVK
	GVTGTDINFKTQIADESFDVDNVVmGFPARTLNK
	FFPSEFGNDVDR
	DLVYEQPILKSLENAmFILGHLPmPSSSQVAK
	KTIEDANIPFTYISANCFAAYFAGNLSQmGTLFPPR
	HINDVEDVLNGLVLGSLHKAIVVDIQVPQSHYSVDYLITK
	SVVVCIQPIDSKLQQVDRLLIR
	LVEAEGIHYYICCNFLMRYLLPSLVQPGLmTPPRDK
isoflavone synthase 1	RMCPGVNLVTSGMATLLASLIQCFDLQVLGPQGQILK
isoflavone synthase 2	LLLGLLLLAMLFEVLLHARPPATTK
isoflavone-7-O-methyltransferase	DENQLTQIKLLmNVTISCVNGK
	QNAIHWLDIINPYIGKSLVQLPDTSSSTVIPLK
	VVENLCGSNNLTyVGGDMFISVPKADAVLLKWILHDWTDR
	LNSmSELQAHmYTLSDGVAFEQQYLLLTkPmSLKGHINIEYALPQCIVNK
isopentenyl-diphosphate Delta-isomerase	ILYKAPSDGKWGEHELDYLLFIVR
	MSWTLNLEHICGLSLFWESPK
	AFSVFLFNSK
	TFNHPLSIPYLSLRPLLIFKSLSSFSPR
	LVVDNFLFK
	PMMDSAVETALGPLQVAASDAGASVRREIFmEDNDVLDLDCR
	NQLELLQRATEKCVNFSDSEGNISHPCTmWLGTIILPDGTM TGK
	GSSAAmSSATGFSSVALKGTQAAAADDQVEAmAGRFSQVIDHVCEDNDLIMDEGELK

kaurene synthase	ISTSVMPQEVECALQLPFYSSVLEPLEHKRNIEHFSTNGIQMR
kynurenine-oxoglutarate transaminase	IAGIVSFHKGVLIAPSLFVPGATR
	IGWAIAPPHLTWGV
	EVGVVAIPTSVFYLNPEEGK
leucoanthocyanidin reductase	PEmVYGSCLPGTVLEEQLAHHAIDQNGFIVYLLPYTLFNSPLK
	ISIATFYSPK
	DYAKIYQQYELGFKR
	ILKKPIAVNYmSQNQNGNVQCDEK
	DFLKLSDSPDFCTVWHPLLPHSMR
	YGVVPTASLISSRREVHYDDAVFYNDIPDTLLGLmTSLDNCAWYGK
limonene cyclase	HFMVSDGAPRRFDVFHMLHSIWL
lycopene beta cyclase, chloroplast precursor	mEDIQERmVAR
lycopene epsilon-cyclase	mLFCAPNVGEPDVTNVVVEDMQYSYEVYR
malonyl-CoA isoflavone 7-O-glucoside-6'-O-	VHGIPHPLRNRYNGFPAQLTCAPIGTER
	FIKAINLFGESLDDNASHLGSPTKNTLAK
myrcene synthase	DLLAMFLAFCLVLGVTRMGVAFVLFSMLAFARLVQACGLANNQIK
	WFNNGCVPNFKTYLDNGVISAGSCmALVHATFLIGDGLSKETmSIMMK
NAD(P)H-dependent 6'-deoxychalcone	NILNDWQIELDTKAQQAGHIK
	VGLIASRDELFITSKLWCCDNHPLVLPALQNSLR
Naringenin,2-oxoglutarate 3-dioxygenase	ELVTYFSYPIR
	KGGFLVSSHLQGEAVQDWR
	LRYSCSAAASEGYGSKmLATTSDQNDQVAVLDWR
naringenin,2-oxoglutarate 3-dioxygenase	LLEVELSEAMGLEK
naringenin-chalcone synthase	SELFIPDTMHGPLADQSAQLSIVGR
nine-cis-epoxycarotenoid dioxygenase4	YDFNGKLLmSMTAHPK
	WFDVPGFNImHTINAWDEEDGGTVTLVATNILSVQHVMER
	VNLWMEGTALNEASDPmLETMEEIGFGVNMNLIHLEHRARVDMVR
p-coumaroyl shikimate 3'-hydroxylase	LAFMLLLVYLVLQPLLILSAVYRLRFK
phenylalanine ammonia lyase	LAGIDSGFFELQPK
	EGLALVNGTAVGSGVASMVLFEANILALFSVVL SAVFTEVmQGK
	IGAFAEEELK

	TSPQWLGPLIEVIR
	HHPGQIEAAAlmEHILDGSSYVK
	INTLLQGYSGIRFEILEAITKLINTGVSPCLPLR
	DGLISLVAILALTTTSPYGGGRPNSKENEDAPVGALFTGVAK
	LTVGSGLASIVLFEANILAVLSEVLSAIFAEVmQGKPEFTDHLTHK
	ELIRFLNAGIFGNGTETCHTLPHSATRAAmLVR
phenylalanine ammonia-lyase 2	EGLALVNGTAVGSGLASVVLFEANLLVILSEVLSAIFAEVmQGK
	ASSDWVMESMDK
phenylalanine and histidine ammonia-lyase	EADARTLPPALALMQEELARDFPPLIEDR
	MIFSQCSEIINPNLSRGLPPNLSVDDPSLSFAFK
	DVVSMTLITAAQALALRDNLQIIAPEVAAFYHAVR
	YALRPPLWCQELHDGVLQILAKLLGTNVAQmSPAQEHGPNSETDFVETINDSVVIGDTLTK
p-hydroxybenzoate polyprenyltransferase	VSSSSYVGRGILLPATFSIYSPLSTSSSASPDPSPDESSPPHPSPAAAPPSSWVDR
prenyltransferase	HTILAQPFWEYADLFSWLTDDEQEQGATSmDLFIGRVPSLVRmGR
polycopene isomerase, chloroplastic	IAITNTSRGTDWFKLLK
	GLLGmPFNTTGIDGLYCVGDSCFPGQGVIAVAFSGVmCAHRVAADIGLEKK
protein farnesyltransferase subunit beta	LSSIINKQmEETSQIFAVSYVSEAKESLDGTSSHATCR
	SIGYNFINEWRAQEPLFHSIALQQYILLCAQEQAGGLRDK
protein prenyltransferase	ECMVSGLAQNKMIELAKK
	LPFLPLDPR QLLTELLR
protocatechuate 3,4-dioxygenase beta	LMTSLIRHVHDFARENEVTVDEWMAGVQLMNWAGQMSDDK
	FDPNFTDNVVNAMGPKTNPR
	DTDETFALDISGVHTWVNSSR
	VQLINPPIHISSALmPIAVYYDDPHDDIHTTHDVAFKDDVFVLDKPK
	TTNAEDSIFPYDTAGGFNPVVDIVPmDGKDVTk
protoheme IX farnesyltransferase	LAPGERAAALLWALVLGSALALYPFQTIVASPAILLTGAFGLLYFLANLLSLTTK
	LSMLVVATSGTGfVLGSGSAVDLSALSCTCLGTmmVAASANSNLNQVFEINNDAKmKR
	ITVPHAVGWASSVGLAGTALLATQVVNLLTSLITNmLAAGLAASNVLVYSFVYTPLK
quinate/shikimate dehydrogenase, putative	ASSVNTSANmYDGSVAAGAVDWLWPIPTYPMVKR
	TPVEEVKAGNAEYITLPPPQARQGGMQLQAPITSVK
rab geranylgeranyltransferase, beta subunit	LRAVEADEVIEWIMSCYHPESGGFGGNVGHDAHVLYTLsAVQVLCFLDR

sesquiterpene synthase	SmNETLLKFAKLDFNR
	MTVGAFQLACASLVGMGDFITEDTLDYLLTYPKLIKSYTTCVR
	EVARSYFAEAWEYNDGVLPTFDEFMQFGLVSSTFDYHASAFFLGVEDLGMKEIIWLR
shikimate 5-dehydrogenase	YLYRLIDLDPQLQDVNALPELLQAAERMAFTGLNITFPCK
shikimate dehydrogenase, putative	DLLDLYNFR EFTmAPERVIELAK
shikimate kinase	SVEETELKTLQSmGRGACANLGSDSLVVLTQLGR
	YMHKGVSVWLDVPVEALAQRIAAVGTNSR
syn-copalyl diphosphate synthase	LmAWAEVmPESLEGEHLHDILDLAFKEALDIYDPPDDLKEAAYIR
	FPGDDGVLGRAEVFCRSFLQDR
taxadiene 5-alpha hydroxylase	REELSKCVATMSLLLLDSLGLNSR VEIEKMLVK
	PECLQNYVKEMDELVNATLLR
	LLQEHVAIMSNKSRGENLTLEDIK
	LEILSNAMVHQKSREGEDNLTLIK
	LNILIFVHYVVTQYEWFLHPDEPVAMDPLPFPSLGMPIRISPKYT
terpene synthase family, metal binding	CSFAAPQDIVGLATGIFMFGSVNNRGLALmIGK
	NFSVAIDGQCAAPVGFNITFSGMLRLAIGmGLKFPVMETDIDSIFR
terpenoid cylases terpenoid synthase	TLKFILFVLYVLVLASFVKIADNGK
trans-cinnamate 4-monooxygenase	NVVFDFITGK RFESEEDPLFNK
	LVQNFELLPPPGQSK TLIGLFLAAVVAIAVSTLR
	GEISEDNTYIIVENINVAAIETTLWSMEWAIAELVNHPEVQSK
	ISEPENEQVENAISEVGLLIMVYVETAWNTAHDVAWIISTNEKIRDEISK
trichodiene oxygenase	LQDELRQVVDDPKNLPPWSVLEALPYLGAVIQEGLR
UDP rhamnose anthocyanidin-3-glucoside	AMNNNDHQLHVFmFPFLAFGHISPFVQLSNKLFSGGVHITFLSASSNIHK
	WFQPFLLDPYITQVQNLKLGKISPLYANSPLWLINIYQGSR
UDP-glucose:flavonoid 3-o-	VSCLVSDAFFWFAGEmAEEIGVWLPFWTAGPTSLSAHVYTDLIRDTFGVGGVAGHEDELLSLIPGMSK
zeta-carotene desaturase	PDQKTPVKNFFLSGSYTK
	VAKWSVIEQSGLSVTSSLQPVVFKGQYLSIR

Apêndice XVIII

Lista de todas as proteínas identificadas no estudo proteômico

Peptide sequence	Scan		Accession number	Reference	Protein name	Organism
	L	S				
YDDVFLK	2	4	gi 224146173	sp Q9XFH8.2	thioredoxin F1, chloroplastic	Populus trichocarpa
FLADGSATYTHALGLELDLQEK	1	4	gi 66360174	pdb 1TP9	peroxiredoxin, D Chain D	Populus Tremula
FALLVDDLK	4	4	gi 66360174	pdb 1TP9	peroxiredoxin, D Chain D	Populus Tremula
GINFDLPHVIEDAPSYPGVEHVGGDM	8	5	gi 223471421	prf 1906376A	caffeic acid O-methyltransferase	Magifera indica
GVNVVLYEGEVHAIMGK	1	4	gi 222850784	gb EEE88331.1	white-brown-complex ABC transporter family	Populus trichocarpa
ALGVDTPVPLVGPVSYLLLSKPAK	2	1	gi 222850043	gb EEE87590.1	vitamin-b12 independent methionine synthase, 5-	Populus trichocarpa
ALGVDTPVPLVGPVSYLLLSK	1	6	gi 222850043	gb EEE87590.1	vitamin-b12 independent methionine synthase, 5-	Populus trichocarpa
FALESFWDGK	5	5	gi 222850043	gb EEE87590.1	vitamin-b12 independent methionine synthase, 5-	Populus trichocarpa
YLFAGVVDGR	6	4	gi 222850043	gb EEE87590.1	vitamin-b12 independent methionine synthase, 5-	Populus trichocarpa
KLNLPIPTTTIGSFQTMDLR	7	1	gi 222837598	gb EEE75963.1	vitamin-b12 independent methionine synthase, 5-	Populus trichocarpa
LNLPIPTTTIGSFQTMDLR	3	4	gi 222837598	gb EEE75963.1	vitamin-b12 independent methionine synthase, 5-	Populus trichocarpa
WFDNTYHFIVPELGPDK	4	2	gi 222862731	gb EEF00238.1	vitamin-b12 independent methionine synthase, 5-	Populus trichocarpa
SEQAFYLDWAVHSFR	2	3	gi 222858086	gb EEE95633.1	vitamin-b12 independent methionine synthase, 5-	Populus trichocarpa
YSGDSDLQLER	3	5	gi 222846300	gb EEE83847.1	tubulin, beta chain	Populus trichocarpa
AYHEQLSVPEITNAVFEPAASMAK	2	3	gi 222841430	gb EEE78977.1	tubulin alpha-2 chain	Populus trichocarpa
MIPTKPMVVETFSAYPPLGR	1	7	gi 255740155	gb ACU31840.1	translation elongation factor 1 alpha, partial	Populus trichocarpa
VASAVTPVPGGVGPMTIAMLLSNTLSS	1	8	gi 222844297	gb EEE81844.1	tetrahydrofolate dehydrogenase/cyclohydrolase	Populus trichocarpa
WGVNVQPYSGSPANFAAYTAVLQPH	4	4	gi 403328048	gb AFR41397.1	serine hydroxymethyltransferase, partial	Populus nigra
GLELIPSENFTSVSVMQAVGSVMTNK	1	1	gi 403328215	gb AFR41476.1	serine hydroxymethyltransferase, partial	Populus alba
GLELIPSENFTSVSVMQAVGSVMTNK	6	0	gi 403328215	gb AFR41476.1	serine hydroxymethyltransferase, partial	Populus alba
QLNAPLEEIDPEIADIIIELEK	3	2	gi 403328215	gb AFR41476.1	serine hydroxymethyltransferase, partial	Populus alba
GLVEKDFEQIGEFLLHR	9	4	gi 403328176	gb AFR41461.1	serine hydroxymethyltransferase, partial	Populus alba
DIEALKADVEK	5	0	gi 403328176	gb AFR41461.1	serine hydroxymethyltransferase, partial	Populus alba
GFVEEDFAK	7	6	gi 222834153	gb EEE72630.1	serine hydroxymethyltransferase	Populus trichocarpa

VHTVLISTQHDETVTNDEIAADLK	8	9	gi 403327907	gb AFR41327.1	s-adenosylmethionine synthetase, partial	Populus nigra
VHTVLISTQHDETVTNDEIAADLKEHVI	6	3	gi 403327907	gb AFR41327.1	s-adenosylmethionine synthetase, partial	Populus nigra
IIIDTYGGWGAHGGGAFSGK	5	0	gi 403327907	gb AFR41327.1	s-adenosylmethionine synthetase, partial	Populus nigra
TAAYGHFGRDDPDFTWETVK	4	1	gi 222833395	gb EEE71872.1	s-adenosylmethionine synthetase 6	Populus trichocarpa
VLVNIEQQSPDIAQGVHGHFTK	5	2	gi 222843269	gb EEE80816.1	s-adenosylmethionine synthetase 5	Populus trichocarpa
RPEEIGAGDQGHMFGYATDEPELMP	7	3	gi 222857839	gb EEE95386.1	s-adenosylmethionine synthetase 2	Populus trichocarpa
VTPQPGVPPEEAGAAVAASSTGTWT	2	7	gi 134093207	ref YP_001109	ribulose-1,5-bisphosphate carboxylase/oxygenase	Populus trichocarpa
VTPQPGVPPEEAGAAVAASSTGTWT	1	2	gi 134093207	ref YP_001109	ribulose-1,5-bisphosphate carboxylase/oxygenase	Populus trichocarpa
VTPQPGVPPEEAGAAVAASSTGTWT	2	0	gi 134093207	ref YP_001109	ribulose-1,5-bisphosphate carboxylase/oxygenase	Populus trichocarpa
DDENVNSQPFMR	8	6	gi 134093207	ref YP_001109	ribulose-1,5-bisphosphate carboxylase/oxygenase	Populus trichocarpa
GGLDFTKDDENVNSQPFMR	8	2	gi 134093207	ref YP_001109	ribulose-1,5-bisphosphate carboxylase/oxygenase	Populus trichocarpa
FLFCAEALYK	8	1	gi 134093207	ref YP_001109	ribulose-1,5-bisphosphate carboxylase/oxygenase	Populus trichocarpa
GGLDFTK	5	3	gi 134093207	ref YP_001109	ribulose-1,5-bisphosphate carboxylase/oxygenase	Populus trichocarpa
AQAETGEIKGHYLNATAGTCEEMIKR	7	0	gi 134093207	ref YP_001109	ribulose-1,5-bisphosphate carboxylase/oxygenase	Populus trichocarpa
AQAETGEIK	3	3	gi 134093207	ref YP_001109	ribulose-1,5-bisphosphate carboxylase/oxygenase	Populus trichocarpa
DYKLTYTPDYETKDTDILAAFR	5	0	gi 134093207	ref YP_001109	ribulose-1,5-bisphosphate carboxylase/oxygenase	Populus trichocarpa
GHYLNATAGTCEEMIKR	0	5	gi 134093207	ref YP_001109	ribulose-1,5-bisphosphate carboxylase/oxygenase	Populus trichocarpa
MGNALPLTDMPLGTAIHNIEITLGR	1	2	gi 134093242	ref YP_001109	ribosomal protein L2	Populus trichocarpa
IMALDLPHGGHLSHGYQTDTK	1	8	gi 222864403	gb EEF01534.1	precursor of transferase serine	Populus trichocarpa
GYELVSGGTENHLVLNLK	9	1	gi 222864403	gb EEF01534.1	precursor of transferase serine	Populus trichocarpa
NTVPGDVSAMVPGGIR	3	5	gi 222864403	gb EEF01534.1	precursor of transferase serine	Populus trichocarpa
IMALDLPHGGHLSHGYQTDTKK	5	0	gi 222864403	gb EEF01534.1	precursor of transferase serine	Populus trichocarpa
AILLADMAHISGLVAAGVIPSPFEYADV	2	2	gi 222850736	gb EEE88283.1	precursor of transferase serine	Populus trichocarpa
LELQEVVDLKNPDKYTALGAK	1	7	gi 222843653	gb EEE81200.1	precursor of protein cell division protease ftsh-like	Populus trichocarpa
TPGFTGADLQNLNMNEAAILAAR	1	1	gi 222843653	gb EEE81200.1	precursor of protein cell division protease ftsh-like	Populus trichocarpa
RTPGFTGADLQNLNMNEAAILAAR	7	5	gi 222843653	gb EEE81200.1	precursor of protein cell division protease ftsh-like	Populus trichocarpa
SFDYDLIIIGAGVGGHGAALHAVEK	5	3	gi 222851215	gb EEE88762.1	precursor of dehydrogenase dihydrolipoamide	Populus trichocarpa
HGHVDYDKVPGVVYTHPEVASVGK	1	5	gi 118489203	gb EE88734.1	precursor of dehydrogenase dihydrolipoamide	Populus trichocarpa
AKGGDVSWHIHDER	7	2	gi 222832818	gb EEE71295.1	precursor of carboxylase t-protein 2, glycine	Populus trichocarpa
GGDVSWHIHDER	7	0	gi 222832818	gb EEE71295.1	precursor of carboxylase t-protein 2, glycine	Populus trichocarpa
NIMENPAWYTQYTPYQAEISQGR	1	5	gi 222854538	gb EEE92085.1	precursor of carboxylase p-protein 1, glycine	Populus trichocarpa

FGVPMGYGGPHAAFLATSQEYK	3	1	gi 222854538	gb EEE92085.1	precursor of carboxylase p-protein 1, glycine	Populus trichocarpa
ESPYLTHPIFNTYHTEHELLR	9	5	gi 222854538	gb EEE92085.1	precursor of carboxylase p-protein 1, glycine	Populus trichocarpa
FGVPMGYGGPHAAFLATSQEYKR	1	2	gi 222854538	gb EEE92085.1	precursor of carboxylase p-protein 1, glycine	Populus trichocarpa
LGTVEVQGLPFFDTVK	4	5	gi 222854538	gb EEE92085.1	precursor of carboxylase p-protein 1, glycine	Populus trichocarpa
NTAGIEPEDVAK	4	4	gi 222854538	gb EEE92085.1	precursor of carboxylase p-protein 1, glycine	Populus trichocarpa
GVNGTVAHEFIVDLR	4	3	gi 222854538	gb EEE92085.1	precursor of carboxylase p-protein 1, glycine	Populus trichocarpa
SYIGMGYYNTHVPPVILR	4	3	gi 222854538	gb EEE92085.1	precursor of carboxylase p-protein 1, glycine	Populus trichocarpa
RLENYYPILFR	5	2	gi 222854538	gb EEE92085.1	precursor of carboxylase p-protein 1, glycine	Populus trichocarpa
YATSHWVK	6	1	gi 222834032	gb EEE72509.1	precursor of carboxylase h-protein 4, glycine	Populus trichocarpa
VDTNSNLLTTITIDEPAPGLK	5	5	gi 222852528	gb EEE90075.1	porin/voltage-dependent anion-selective channel	Populus trichocarpa
AWMAAQDQPHENLIFPEEVLPR	2	2	gi 134093193	ref YP_001109	photosystem II protein D2	Populus trichocarpa
AAEDPEFETFYTK	4	1	gi 134093193	ref YP_001109	photosystem II protein D2	Populus trichocarpa
FGQEEETYNIVAAGHYFGR	8	9	gi 134093178	ref YP_001109	photosystem II protein D1	Populus trichocarpa
EESLWGR	6	8	gi 134093178	ref YP_001109	photosystem II protein D1	Populus trichocarpa
ANLGMEVMHER	2	5	gi 134093178	ref YP_001109	photosystem II protein D1	Populus trichocarpa
YSVEQVGVTVEFYGGELNGVSYSDPA	5	4	gi 134093227	ref YP_001109	photosystem II 47 kDa protein	Populus trichocarpa
DVFAGIDPDLDAQVEFGAFQK	1	2	gi 134093227	ref YP_001109	photosystem II 47 kDa protein	Populus trichocarpa
RMPTFFETFPVVLVDGDGIVR	9	9	gi 134093227	ref YP_001109	photosystem II 47 kDa protein	Populus trichocarpa
YQWDQGYFQQEIYRR	5	1	gi 134093227	ref YP_001109	photosystem II 47 kDa protein	Populus trichocarpa
SAEYMTTHAPLGSLNSVGGVATEINAV	1	1	gi 134093194	ref YP_001109	photosystem II 44 kDa protein	Populus trichocarpa
APWLEPLRGPNGLDLSR	1	6	gi 134093194	ref YP_001109	photosystem II 44 kDa protein	Populus trichocarpa
GIDRDFEPVLSMTPLN	7	7	gi 134093194	ref YP_001109	photosystem II 44 kDa protein	Populus trichocarpa
ITNLTLSPSVIFGYLLK	6	7	gi 134093194	ref YP_001109	photosystem II 44 kDa protein	Populus trichocarpa
DFEPVLSMTPLN	7	6	gi 134093194	ref YP_001109	photosystem II 44 kDa protein	Populus trichocarpa
FWDLRAPWLEPLRGPNGLDLSR	7	2	gi 134093194	ref YP_001109	photosystem II 44 kDa protein	Populus trichocarpa
VYLWHETTR	7	7	gi 134093257	ref YP_001109	photosystem I subunit VII	Populus trichocarpa
QILIEPIFAQWIQSAHGK	2	1	gi 134093198	ref YP_001109	photosystem I P700 chlorophyll a apoprotein A2	Populus trichocarpa
RIWFGIATAHDFESHDDITEER	6	1	gi 134093198	ref YP_001109	photosystem I P700 chlorophyll a apoprotein A2	Populus trichocarpa
DYNPEQNEDNVLAR	5	6	gi 134093198	ref YP_001109	photosystem I P700 chlorophyll a apoprotein A2	Populus trichocarpa
TSYGFDVLLSSTNSPAFNAGR	4	6	gi 134093198	ref YP_001109	photosystem I P700 chlorophyll a apoprotein A2	Populus trichocarpa
TNFGIGHSIKDLLEAHIPPGGR	4	4	gi 134093198	ref YP_001109	photosystem I P700 chlorophyll a apoprotein A2	Populus trichocarpa

IWFGIATAHDFESHDDITEER	2	4	gi 134093198	ref YP_001109	photosystem I P700 chlorophyll a apoprotein A2	Populus trichocarpa
PIAHAIWDPHFGQPAVEAFTR	2	3	gi 134093198	ref YP_001109	photosystem I P700 chlorophyll a apoprotein A2	Populus trichocarpa
YSEFLTFR	8	9	gi 134093199	ref YP_001109	photosystem I P700 chlorophyll a apoprotein A1	Populus trichocarpa
ILVDRDPIKTSFEEWAR	9	3	gi 134093199	ref YP_001109	photosystem I P700 chlorophyll a apoprotein A1	Populus trichocarpa
YNDLLDR	5	6	gi 134093199	ref YP_001109	photosystem I P700 chlorophyll a apoprotein A1	Populus trichocarpa
ILVDRDPIKTSFEEWARPGHFSR	8	0	gi 134093199	ref YP_001109	photosystem I P700 chlorophyll a apoprotein A1	Populus trichocarpa
TSFEEWARPGHFSR	6	1	gi 134093199	ref YP_001109	photosystem I P700 chlorophyll a apoprotein A1	Populus trichocarpa
EIPLPHEFILNR	3	3	gi 134093199	ref YP_001109	photosystem I P700 chlorophyll a apoprotein A1	Populus trichocarpa
FFPSEFGMDVDHVNAVEPAK	6	7	gi 222843751	gb EEE81298.1	phenylcoumaran benzylic ether reductase 1	Populus trichocarpa
NREWDFEYR	2	3	gi 134093262	ref YP_001109	NADH dehydrogenase subunit 7	Populus trichocarpa
FGFYEFFK	4	1	gi 222859438	gb EEE96985.1	mitochondrial phosphate carrier protein	Populus trichocarpa
KGSITSVQAIYVPADDLTDPAPATTFA	1	8	gi 222864942	gb EEF02073.1	mitochondrial beta subunit of F1 ATP synthase	Populus trichocarpa
QISELGIYPAVDPLDSTSR	0	6	gi 222864942	gb EEF02073.1	mitochondrial beta subunit of F1 ATP synthase	Populus trichocarpa
YDDLPEQSFYMGVGGIEEVIK	1	4	gi 222864942	gb EEF02073.1	mitochondrial beta subunit of F1 ATP synthase	Populus trichocarpa
AEYHQQYLEK	6	4	gi 222869209	gb EEF06340.1	methionine sulfoxide reductase type	Populus trichocarpa
YHELQGHAFGEK	5	2	gi 222845857	gb EEE83404.1	lysine/histidine transporter	Populus trichocarpa
LHPGGPFDPLGLAKDPDQAALLK	2	1	gi 222870973	gb EEF08104.1	light-harvesting complex II protein Lhcb5	Populus trichocarpa
HLSDPFGNNLLTVIAGSAERAPSL	9	0	gi 222870973	gb EEF08104.1	light-harvesting complex II protein Lhcb5	Populus trichocarpa
WAMLGAAGFIIPEAFNK	5	0	gi 222870973	gb EEF08104.1	light-harvesting complex II protein Lhcb5	Populus trichocarpa
TGALLLDGNTLNIFYGK	1	4	gi 222870973	gb EEF08104.1	light-harvesting complex II protein Lhcb5	Populus trichocarpa
FGEAVWFK	6	1	gi 222859802	gb EEE97349.1	light-harvesting complex II protein Lhcb1	Populus trichocarpa
IIPSFMIQGGDFTLGDGR	6	0	gi 222845570	gb EEE83117.1	isomerase peptidyl-prolyl cis-trans isomerase	Populus trichocarpa
IHENPELGFEETFSTK	2	4	gi 222840651	gb EEE78198.1	iaa-amino acid hydrolase 11	Populus trichocarpa
IFLENVIRDAVITYTEHAR	6	0	gi 222865907	gb EEF03038.1	histone H4	Populus trichocarpa
EIAQDFKTDLR	6	0	gi 222852864	gb EEE90411.1	histone H3	Populus trichocarpa
LLGQVTIANGGVLPNIHNTLLPK	5	1	gi 222848605	gb EEE86152.1	histone H2	Populus trichocarpa
TPSSHPPYEEMIKDAIVTLK	1	4	gi 222843795	gb EEE81342.1	histone H1	Populus trichocarpa
LLAGVTIAHGGVLPNINPVLLPK	7	1	gi 118485435	gb EEE91668.1	histone 2	Populus trichocarpa
LLQGVTIASGGVLPNINPVLLPK	1	1	gi 222857370	gb EEE94917.1	histone 2	Populus trichocarpa
GKGDIGSASQEF	7	8	gi 222838250	gb EEE76615.1	histone 2	Populus trichocarpa
LLGSVTIANGGVLPNIHQTLTPK	7	8	gi 222838250	gb EEE76615.1	histone 2	Populus trichocarpa

NDEELGKLLQGVTIASGGVLPNINPVL	9	1	gi 222857370	gb EEE94917.1	histone 2	Populus trichocarpa
HLQLAIRGDEELDTLIK	7	0	gi 222842536	gb EEE80083.1	histone 2	Populus trichocarpa
VDAVVYLVDAYDKER	5	1	gi 222857334	gb EEE94881.1	gtp-binding protein SAR1	Populus trichocarpa
VEAGSIVAIFGLGTVGLAVAEGAK	1	1	gi 222843562	gb EEE81109.1	glutathione-dependent formaldehyde dehydrogenase	Populus trichocarpa
IVDPHTVDVDGK	4	5	gi 222868511	gb EEF05642.1	glutathione reductase	Populus trichocarpa
GIEFHTEESPQAIK	1	5	gi 222868511	gb EEF05642.1	glutathione reductase	Populus trichocarpa
VLLTLEEK	6	6	gi 283135904	gb ADB11343.1	DHAR class glutathione transferase DHAR1	Populus trichocarpa
GALNVGAVLILPEGFELAPPDR	1	7	gi 134093213	ref YP_001109	cytochrome f	Populus trichocarpa
VQLSEMNF	5	6	gi 134093213	ref YP_001109	cytochrome f	Populus trichocarpa
KGALNVGAVLILPEGFELAPPDR	5	4	gi 134093213	ref YP_001109	cytochrome f	Populus trichocarpa
GALNVGAVLILPEGFELAPPDRISPEM	6	0	gi 134093213	ref YP_001109	cytochrome f	Populus trichocarpa
IVTGVPEAIPVIGSPLVELLR	1	1	gi 134093231	ref YP_001109	cytochrome b6	Populus trichocarpa
SKVYDWFEER	4	2	gi 134093231	ref YP_001109	cytochrome b6	Populus trichocarpa
SLVYLDITDCGSLRFMPAGMRQLICLR	4	2	gi 222844636	gb EEE82183.1	cc-nbs-lrr resistance protein	Populus trichocarpa
VGGANHSHATQDLYDSIAAGNYPEWK	3	5	gi 222856425	gb EEE93972.1	catalase	Populus trichocarpa
TWPEDILPLQPVGR	0	7	gi 222843646	gb EEE81193.1	catalase	Populus trichocarpa
WVEALSDPR	3	3	gi 222855876	gb EEE93423.1	catalase	Populus trichocarpa
AYGMTAFEYHGTDPR	6	1	gi 313104447	gb ADR31602.1	caffeic acid 5-hydroxyferulic acid 35-O-	Populus trichocarpa
VLMEWYHLK	2	3	gi 217037959	gb ACJ76442.1	caffeic acid 3-O-methyltransferase 2	Populus trichocarpa
QGEIEAVVIATGVHTFFGK	9	3	gi 222870522	gb EEF07653.1	autoinhibited H ⁺ ATPase	Populus trichocarpa
SASDIVLTEPGLSVIISAVLTSR	5	0	gi 222870522	gb EEF07653.1	autoinhibited H ⁺ ATPase	Populus trichocarpa
EGSITSIQAVVYPADDLTDPAATTFA	7	3	gi 134093206	ref YP_001109	ATP synthase CF1 beta subunit	Populus trichocarpa
TREGNDLYMEMK	9	7	gi 134093206	ref YP_001109	ATP synthase CF1 beta subunit	Populus trichocarpa
GMEVIDTGAPLSVPVGGATLGR	7	7	gi 134093206	ref YP_001109	ATP synthase CF1 beta subunit	Populus trichocarpa
AHGGVSVFGGVGER	7	6	gi 134093206	ref YP_001109	ATP synthase CF1 beta subunit	Populus trichocarpa
LILSGELDSLPEQAFYLVGNIDEATAK	6	6	gi 134093206	ref YP_001109	ATP synthase CF1 beta subunit	Populus trichocarpa
VALVYGQMNEPPGAR	6	5	gi 134093206	ref YP_001109	ATP synthase CF1 beta subunit	Populus trichocarpa
VVDLLAPYR	5	4	gi 134093206	ref YP_001109	ATP synthase CF1 beta subunit	Populus trichocarpa
TREGNDLYMEMKESGVINEENIAESK	7	0	gi 134093206	ref YP_001109	ATP synthase CF1 beta subunit	Populus trichocarpa
LSIFETGIKVVDLLAPYR	6	0	gi 134093206	ref YP_001109	ATP synthase CF1 beta subunit	Populus trichocarpa
VGLTALTMAEYFR	3	3	gi 134093206	ref YP_001109	ATP synthase CF1 beta subunit	Populus trichocarpa

LSIFETGIKVVDLLAPYRR	5	0	gi 134093206	ref YP_001109	ATP synthase CF1 beta subunit	Populus trichocarpa
RSVYEPLQTGLIAIDSMIPIGR	1	1	gi 134093182	ref YP_001109	ATP synthase CF1 alpha subunit	Populus trichocarpa
SVYEPLQTGLIAIDSMIPIGR	1	9	gi 134093182	ref YP_001109	ATP synthase CF1 alpha subunit	Populus trichocarpa
IVNTGTVLQVGDGIAR	8	8	gi 134093182	ref YP_001109	ATP synthase CF1 alpha subunit	Populus trichocarpa
LIESPAPGIISR	5	0	gi 134093182	ref YP_001109	ATP synthase CF1 alpha subunit	Populus trichocarpa
VTIRADEISNIIR	5	0	gi 134093182	ref YP_001109	ATP synthase CF1 alpha subunit	Populus trichocarpa
FLLQEQV	0	5	gi 134093182	ref YP_001109	ATP synthase CF1 alpha subunit	Populus trichocarpa
AIFVDLEPTVIDEVR	4	4	gi 403329340	gb AFR41980.1	alpha-tubulin, partial	Populus trichocarpa
DSAAVFAWK	5	0	gi 222847405	gb EEE84952.1	adenosylhomocysteinase/s-adenosyl-l-homocysteine	Populus trichocarpa
TTGIVLDSGDGVSHTVPIYEGYALPHAI	1	1	gi 222874629	gb EEF11760.1	actin 9	Populus trichocarpa
LAYVALDYEQELETAK	1	7	gi 222874629	gb EEF11760.1	actin 9	Populus trichocarpa
MPDVAMTLAVVALFADGPTAIR	9	0	gi 222843005	gb EEE80552.1	3-phosphoshikimate 1-carboxyvinyltransferase	Populus trichocarpa
GLFIIDKEGVIIQHSTINNLAIGR	1	1	gi 222855371	gb EEE92918.1	2-cys peroxiredoxin	Populus trichocarpa
EGVIQHSTINNLAIGR	2	3	gi 222855371	gb EEE92918.1	2-cys peroxiredoxin	Populus trichocarpa
DAINPTLMQTLEGTPVLVHAGPFANIA	4	2	gi 222844912	gb EEE82459.1	10-formyltetrahydrofolate synthetase	Populus trichocarpa
RVLTCPKYGLPALSGLWSHK	0	8	gi 110677919	ref YP_680926.	amine oxidase	Roseobacter
NKEFLEALMNPVATFVRR	0	5	gi 117921222	ref YP_870414.	phenylalanine 4-hydroxylase	Shewanella sp.
LPEPVYMPCGHLLR	0	5	gi 118374367	ref XP_001020	zinc finger domain protein	Tetrahymena
FGEELDLVFR	2	2	gi 152994037	ref YP_001351	replicase poliprotein	Scotophilus bat
CPTHTGTHLDTYQRPWK	0	5	gi 157363127	ref YP_001469	cyclase family protein isoform 1	Thermotoga
LLPALLAEDVTGEALATLVVNK	0	5	gi 134101	ref P08824.1	RuBisCO large subunit-binding protein subunit alpha	Ricinus communis
YPSVLTVSVDTLDRNDRR	0	7	gi 163858269	ref YP_001632	competence/damage inducible protein CinA	Bordetella petrii
WWMRWQPSDSFAQYLVVVLGWKK	0	5	gi 123243475	gb AMB74068.	beta-1,2-glucanase 3	Prunus avium
LWVFSGTDDSVLPVTSTR	0	5	gi 115466820	ref NP_001057	serine carboxylase II-2	Oriza sativa
VVDLAHLVASK	2	4	gi 211953671	gb ACJ13997.1	glyceraldehyde-3-phosphate dehydrogenase	Helianthus petiolaris
FVWYKPKDDSAVLPLGNTYTVVR	0	9	gi 218531476	ref YP_002422	Zn dependent metalloprotease - protein Mchl_3544	Methylobacterium
FLMDDLYR	2	3	gi 2499488	sp Q41140.1	pyrophosphate-dependent phosphofructokinase alpha	Ricinus communis
AHLDVTALLPK	3	2	gi 25150725	ref NP_498650.	telomere length regulation protein clk-2	Caenorhabditis
KFVFYDVTNVLLHFMKK	0	5	gi 281204347	gb EFA78543.1	protein kinase 3	Polysphondylium
MWFYPLLAGTTSGGVLTTLLQLR	0	8	gi 325106296	ref YP_004275	patatin	Pedobacter saltans
VEAVDASGTLMTK	0	5	gi 338210065	ref YP_004654	heme-binding protein	Runella slithyformis

LSDLFFFDK	3	3	gi 358333550	dbj GAA52039.	inactive phospholipase C-like protein 1	Clonorchis sinensis
NNVVTPEHLGDLVGLDTQK	5	4	gi 359497491	ref XP_003635	elongation factor G, chloroplastic-like	Vitis vinifera
WDWAWWCEFVNFATGKGALLCVP	0	1	gi 374300624	ref YP_005052	diguanylate cyclase/phosphodiesterase with	Desulfovibrio
SSLNPAKLVFGEAPVRK	0	8	gi 448634400	ref ZP_216747	UDP-glucose 6-dehydrogenase	Haloarcula
LPALSLLFK	0	6	gi 449503327	ref XP_004161	embryogenesis-associated protein EMB8-like	Cucumis sativus
VVHGDEFYKDVPTDDKK	0	9	gi 449521756	ref XP_004167	phosphoglucosyltransferase, cytoplasmic-like	Cucumis sativus
VPLWQNFGEYYLADNHDGFQK	0	8	gi 52426004	ref YP_089141.	bifunctional N-acetylglucosamine-1-phosphate	Mannheimia
EVTLGFVDLLR	1	6	gi 5821345	dbj BAA83779.	ribulose-1,5-bisphosphate carboxylase/oxygenase	Leptopteris
LDTLGFVDLLR	0	1	gi 61378666	gb AAX44980.1	ribulose-1,5-bisphosphate carboxylase/oxygenase	Psathyrostachys
MYSLLSEYK	2	3	gi 70982959	ref XP_747007.	C6 transcription factor	Aspergillus
RPMKENEGEYYLALGGHNMYGK	0	5	gi 82702686	ref YP_412252.	ABC transporter, periplasmic binding protein	Nitrosospora
FGEGLDLVFGK	1	0	gi 86747144	ref YP_483640.	cytochrome c oxidase, cbb3-type, subunit I	Rhodopseudomonas
YQWDKPVALLSVFEAVVR	0	6	gi 90412454	ref ZP_012204	phosphoenolpyruvate-protein phosphotransferase	Photobacterium
LPHFDEWGEKEVWHSQRR	0	1	gi 91077000	ref XP_975499.	vesamicol binding protein	Tribolium
LLPWVWRDLSAVLHEARNHPMK	8	0	gi 62290241	ref YP_222034.	oxidoreductase	Brucella abortus bv.
FTALWLTPVGLLGGGGSFSAGNPK	5	0	gi 443244391	ref YP_007377	glycosyltransferase	Nonlabens
VLEAHQNPFPQGMMEHK	9	0	gi 154686963	ref YP_001422	3-isopropylmalate dehydrogenase	Bacillus
LDYAPDGGGLGGPPHHHR	7	0	gi 356497689	ref XP_003517	UDP-glycosyltransferase 85A3-like	Glycine max
AFYSLQKVVLMVVFSGDTRR	1	0	gi 380007748	gb AFD29875.1	ATP synthase alpha subunit, partial (mitochondrion)	Androstaphylos
LWWWWWWRTPHTGALHVWTGDNY	6	0	gi 1741862	dbj BAA08235.	polyphenol oxidase	Phytolacca
LGAAMLDFVK	4	2	gi 366158720	ref ZP_094585	bifunctional glyoxylate/hydroxypyruvate reductase A	Escherichia sp.
WFRYYTWMSPRANSPNHAGFDLPW	2	0	gi 409402358	ref ZP_112519	pyruvate phosphate dikinase	Acidocella sp. MX-
SSLMPKHSAAATWRK	6	0	gi 146186256	ref XP_001033	phosphoenolpyruvate carboxykinase	Tetrahymena
CFAPYLPSPVSTNLPALLK	5	0	gi 134142800	gb ABO61734.1	mitochondrial glycine decarboxylase complex P-	Populus tremuloides
DLVSSLVTGLLTGPR	1	0	gi 449524892	ref XP_004169	ATP-citrate synthase beta chain protein 1-like	Cucumis sativus
WMFQHDVPVNGGHGVDGNLRR	5	0	gi 254500886	ref ZP_051130	endonuclease/exonuclease/phosphatase family	Labrenzia alexandrii
FCNLDFKLLKDESTVK	5	0	gi 385773235	ref YP_005645	ribonucleoside-diphosphate reductase	Sulfolobus
LHLQQLETMGGLLELDPK	1	0	gi 357515379	ref XP_003627	40S ribosomal protein S19-1	Medicago truncatula
LRDLYEHLTSSK	6	0	gi 420152366	ref ZP_146594	glycosyltransferase, group 1 family protein	Actinomyces
LLTALSTGTGVVEALDEFLK	7	0	gi 315126546	ref YP_004068	3-phosphoshikimate 1-carboxyvinyltransferase (5-	Pseudoalteromonas
QMNLVEALECMDDHK	5	0	gi 297294195	ref XP_001086	peptidyl-prolyl cis-trans isomerase A-like	Macaca mulatta

LVYPGLEDHPQHELLLSER	1	0	gi 357462339	ref XP_003601	Cystathionine gamma-lyase	Medicago truncatula
RYLDDQVAGNETQLQR	6	0	gi 255324301	ref ZP_053654	adenosylmethionine-8-amino-7-oxononanoate	Corynebacterium
WGVDALLNPAASETLARQLR	9	0	gi 34540961	ref NP_905440.	ribonuclease	Porphyromonas
YKAFRPHLTALKEEVVK	5	0	gi 225458828	ref XP_002285	calcium-transporting ATPase 1, chloroplastic-like	Vitis vinifera
YYPQQEPVVVLTCLAWK	5	0	gi 319405462	emb CBI79081.	3-methyl-2-oxobutanoate hydroxymethyltransferase	Bartonella sp. AR
WEYYRQESLEVSCLFVFK	6	0	gi 304384786	ref ZP_073671	acetyl-coA carboxylase carboxyl transferase subunit	Pediococcus
CGTLVATTTCKGVKPLVDLK	7	0	gi 312137337	ref YP_004004	magnesium chelatase	Methanothermus
YRHVGDRVTWQQGAPSYLAADNAFR	9	0	gi 322369433	ref ZP_080439	ABC-type transport system permease protein	Haladaptatus
HWWWRMYNLWFLAVANARQAHLR	1	0	gi 449480575	ref XP_004155	transmembrane 9 superfamily member 4-like	Cucumis sativus
EYYTMRKNMSLEVSDTFVFK	7	0	gi 326797330	ref YP_004315	type II secretion system protein E	Marinomonas
EPYYLWFWWFYTTVVDNKLPLSPYHA	7	0	gi 350536529	ref NP_001234	alkaline alpha-galactosidase seed imbibition protein	Solanum
RTMWYRGDEQLAEYLLSTLFPQK	7	0	gi 217977088	ref YP_002361	Enoyl-CoA hydratase/isomerase	Methylocella
HHWDLEWLVRNAAQSTVLPTVVPR	8	0	gi 209519986	ref ZP_032687	lysyl-tRNA synthetase	Burkholderia sp.
KMFYVQHCEEFMRR	6	0	gi 420256918	ref ZP_147597	Tannase and feruloyl esterase	Burkholderia sp.
LPASDAPRNVAVLPLFDHGR	8	0	gi 317034676	ref XP_001400	phosphotransferase enzyme family protein	Aspergillus niger
WWWWYQWRWRLYYTMGHGAGLAA	6	0	gi 269103889	ref ZP_061565	O-methyltransferase-related protein	Photobacterium
WHGDPNDKAVQTSSEYR	5	0	gi 30103014	gb AAP21427.1	putative calreticulin	Oryza sativa
LYLDLLK	2	7	gi 414877412	tpg DAA54543.	TPA: putative ulp1 protease family protein	Zea mays
LPPNLTTLLDLVELPLPK	6	0	gi 15239700	ref NP_199688.	TIR-NBS-LRR class disease resistance protein	Arabidopsis thaliana
WRYDAAQALLFFTADQHRR	5	0	gi 170090057	ref XP_001876	glycoside hydrolase family 15 protein	Laccaria bicolor
LLDELER	2	3	gi 126733243	ref ZP_017489	acetyl-CoA acetyltransferase	Roseobacter sp.
LLQWWYWWNLNLPDVVSFLLKTKSHD	0	1	gi 437161990	ref ZP_206958	putative glycosyl transferase, partial	Salmonella enterica
LFLPEGLDDR	0	7	gi 449462884	ref XP_004149	anthocyanidin 3-O-glucosyltransferase 5-like	Cucumis sativus
HIAAKLLVEPGHRVLDIGCGWGGLSR	4	1	gi 160895583	ref YP_001561	cyclopropane-fatty-acyl-phospholipid synthase	Delftia acidovorans
LLEPVYLVEIQAPEQALGGIYSVLNQK	2	1	gi 356556977	ref XP_003546	elongation factor 2-like	Glycine max
GPLENLADHLADPVNNNAWAYATNFV	1	2	gi 56809379	gb AAW31511.	light-harvesting chlorophyll-a/b binding protein Lhcb1	Pisum sativum
AVVVHADPDDLKGKGHELSK	2	1	gi 229464641	gb ACQ66643.	Cu/Zn superoxide dismutase, partial	Vigna radiata
NMITGAAQMDGAILVVSADGMPMPQT	1	1	gi 307931160	dbj BAJ21441.1	translation elongation factor Tu	Ulva arasaki
SLAAAGAEILVGTWVPALNIFETSLR	2	8	gi 317373789	gb ADV16377.1	enoyl-ACP reductase 2	Helianthus annuus
DGIAVEKPIYNHVTGLDPPELIKPPK	1	1	gi 356495988	ref XP_003516	phosphoribulokinase, chloroplastic-like	Glycine max
DVFGHMGSLSDKDIVALSGGHTLGR	1	1	gi 221327587	gb ACM17463.	ascorbate peroxidase	Citrus maxima

TYPEEMIQTGISTIDVMNSIAR	1	1	gi 134133942	gb ABO61030.1	vacuolar H ⁺ -ATPase B subunit	Medicago truncatula
GFLGQPQLESALTGMDLVIIPAGVPR	1	8	gi 356506619	ref XP_003522	malate dehydrogenase 1, glyoxysomal-like	Glycine max
EGNLIWPLLLDNVRPDMR	1	9	gi 449523177	ref XP_004168	NADP-dependent glyceraldehyde-3-phosphate	Cucumis sativus
GVMLGPDQPVLHMLDIPPAEALNGV	1	8	gi 449478013	ref XP_004155	malate dehydrogenase, cytoplasmic-like	Cucumis sativus
VDAIKDSLENDEEKVGADIVK	1	6	gi 356525839	ref XP_003531	ruBisCO large subunit-binding protein subunit beta,	Glycine max
QGVQGISDIITIPGLVNVDFADV K	1	1	gi 356537097	ref XP_003537	cell division protein ftsZ homolog 2-1, chloroplastic-	Glycine max
LASLADLYVNDAFGTAHR	1	1	gi 1022805	gb AAB60303.1	phosphoglycerate kinase, partial	Arabidopsis thaliana
EENPRVPIIVTGNDFSTLYAPLIR	1	1	gi 449450762	ref XP_004143	ribulose biphosphate carboxylase/oxygenase	Cucumis sativus
FGIVEGLMTTVHSITATQK	1	1	gi 380469908	gb AFD62268.1	glyceraldehyde-3-phosphate dehydrogenase	Prunus armeniaca
IGLAGLAVMGQNLALNIAEK	1	8	gi 307718128	ref YP_003873	6-phosphogluconate dehydrogenase	Spirochaeta
DILLLDVAPLTGIVTVGGVMTK	8	1	gi 406605278	emb CCH4330	Heat shock 70 kDa protein	Wickerhamomyces
VETGVLKPGMVVTFGPTGLTTEVK	1	8	gi 449508682	ref XP_004163	elongation factor 1-alpha-like	Cucumis sativus
YAGVGAAVEYAVLHLK	1	1	gi 339958979	gb AEK25173.1	chloroplast carbonic anhydrase	Dimocarpus longan
IQSVPLVIGQFMEMVDQNGIVGSTTG	1	9	gi 170037747	ref XP_001846	26S protease regulatory subunit 6B	Culex
ELDIQMMLAAEVHLGTK	1	4	gi 195622270	gb ACG32965.	40S ribosomal protein SA	Zea mays
FQSVPVQPGQTPPLLQYFGTLLTR	1	7	gi 356551424	ref XP_003544	clathrin heavy chain 1-like	Glycine max
KLASLADLYVNDAFGTAHR	1	4	gi 1022805	gb AAB60303.1	phosphoglycerate kinase, partial	Arabidopsis thaliana
GAALNAIQIAEMLL	1	8	gi 356524624	ref XP_003530	aspartate-semialdehyde dehydrogenase-like	Glycine max
LVAYHEAGHALVGALMPEYDPVAK	1	5	gi 18402995	ref NP_564563.	cell division protease ftsH-1	Arabidopsis thaliana
AVASISAGNDELIGTMIADAIDK	9	9	gi 357167236	ref XP_003581	ruBisCO large subunit-binding protein subunit alpha,	Brachypodium
LGGHLSEASSEEDFTGKPGQSLVLR	1	6	gi 225445867	ref XP_002276	leucine aminopeptidase 2, chloroplastic isoform 1	Vitis vinifera
ILLDGSPESPLDFIQVQIANVESK	1	7	gi 226498082	ref NP_001147	D-3-phosphoglycerate dehydrogenase	Zea mays
ESTLHLVLR	7	1	gi 449450454	ref XP_004142	ubiquitin-40S ribosomal protein S27a-like	Cucumis sativus
FQSVPVQAGQTPPLLQYFGTLLTR	1	6	gi 356498960	ref XP_003518	clathrin heavy chain 1-like	Glycine max
GLIKPGESVLEPTSGNTGIGLAFMAAA	8	9	gi 255542380	ref XP_002512	cysteine synthase, putative	Ricinus communis
MELSSLTQTNISLPFITATADGPK	1	7	gi 212375111	gb ACJ24805.1	heat shock protein 70B	Dunaliella salina
TGLISSVGNLGAPVAEWTGVTGALTSL	1	4	gi 449433998	ref XP_004134	pyrophosphate--fructose 6-phosphate 1-	Cucumis sativus
ATIANMSPEYGATMGFFPVDHVTQLQY	8	8	gi 356550020	ref XP_003543	aconitate hydratase, cytoplasmic-like	Glycine max
VLSVVGLLQDEVDPMVSVMK	8	7	gi 449462685	ref XP_004149	26S proteasome regulatory subunit 4 homolog A-like	Cucumis sativus
LLTAVSQSTPVTEAFNLADDILR	7	8	gi 356572556	ref XP_003554	cell division protein ftsZ homolog 2-1, chloroplastic-	Glycine max
ANPVALLSSAMMLR	1	3	gi 449444592	ref XP_004140	isocitrate dehydrogenase NAD regulatory subunit 1,	Cucumis sativus

LGHNFVGTEQILLGLIGEGTGIAAK	8	7	gi 449440498	ref XP_004138	chaperone protein ClpC, chloroplastic-like	Cucumis sativus
VAILGAAGGIGQPLALLMK	1	4	gi 145334349	ref NP_001078	malate dehydrogenase	Arabidopsis thaliana
NHYVEGGDAGNREDYINELIR	6	9	gi 307176212	gb EFN65855.1	60S ribosomal protein L7	Camponotus
KVVDEVPWFGEIEYETLLQTNVK	1	2	gi 255551511	ref XP_002516	glutamine synthetase plant, putative	Ricinus communis
GGTLLEVFSLELMGR	8	7	gi 356500948	ref XP_003519	pyruvate dehydrogenase E1 component subunit	Glycine max
MAPPLVTLLSAEPEIQYVALR	9	6	gi 356539170	ref XP_003538	beta-adaptin-like protein C-like	Glycine max
VLLQDFTGVPVVDLASMR	6	9	gi 42520011	ref NP_965926	aconitate hydratase	Wolbachia
VGADVMSHISTGGGASLELLEK	6	9	gi 420437801	ref ZP_149367	phosphoglycerate kinase	Helicobacter pylori
SINPDEAVAYGAAGVQAAILSGEGNEK	7	8	gi 48716124	dbj BAD22854	heat shock protein 70	Nicotiana
DIQAWYVPLGPFLGK	6	8	gi 225706644	gb ACO09168	Fumarylacetoacetase	Osmerus mordax
VVNIPNSMTILDELLPISIEMAK	1	3	gi 225426743	ref XP_002282	probable rhamnose biosynthetic enzyme 1	Vitis vinifera
GVQYLNEIKDSVAGFQWASK	1	2	gi 432915919	ref XP_004079	elongation factor 2-like	Oryzias latipes
GIDVQQVSLVINYLPTQPENYLHR	7	7	gi 47097038	ref ZP_002346	ATP-dependent RNA helicase DbpA	Listeria
HELGEHNGVNMSIEVSDTFVFK	9	5	gi 414870736	tpg DAA49293	TPA: O-succinylhomoserine sulphydrylase	Zea mays
GSFETIHVQDATGHEFATR	8	6	gi 356559288	ref XP_003547	40S ribosomal protein S4-1-like	Glycine max
IINIASVVGLVGNAGQANYSAK	8	6	gi 380853846	gb AFE88230.1	beta-keto-ACP reductase	Nicotiana tabacum
SNLSEFDDVQGHGPFDK	1	4	gi 62321494	dbj BAD94946	2-oxoglutarate dehydrogenase, E1 subunit - like	Arabidopsis thaliana
AVVVHADPDDLK	9	5	gi 3273755	gb AAC24833.1	copper/zinc superoxide dismutase	Arabidopsis thaliana
QLINSGPVIASQLLLVDEVIR	6	8	gi 255551408	ref XP_002516	chaperonin containing t-complex protein 1, zeta	Ricinus communis
IQPVYVVDVAAAIAALK	5	9	gi 255571168	ref XP_002526	NADH-ubiquinone oxidoreductase 39 kD subunit,	Ricinus communis
GLPGVLWVLPDSYIDVK	6	7	gi 225461197	ref XP_002283	DAG protein, chloroplastic isoform 1	Vitis vinifera
THKPLSVELGPGILGNIFDGIQRPLK	8	5	gi 15219234	ref NP_178011	V-type proton ATPase catalytic subunit A	Arabidopsis thaliana
EQVFSTYSDNQPGVLIQVYGER	6	7	gi 861170	emb CAA27330	heat shock protein 70	Zea mays
SHEVKDEGYEIEHDGK	8	5	gi 343488864	gb AEM45799	serine/threonine protein phosphatase 5	Solanum torvum
LLEPVYLVEIQAPEQALGGIYSVLNQK	1	0	gi 356556977	ref XP_003546	elongation factor 2-like	Glycine max
YGVSGYPTLQWFPK	9	4	gi 162461791	ref NP_001105	protein disulfide isomerase6 precursor	Zea mays
LRDGVATATDLVLTVTQMLR	9	4	gi 225468576	ref XP_002263	aconitate hydratase 1-like	Vitis vinifera
DATDDKVITIESAEATLK	8	5	gi 449483769	ref XP_004156	isocitrate dehydrogenase NADP, chloroplastic-like	Cucumis sativus
GQETPGEDPLLSSK	8	5	gi 356558612	ref XP_003547	beta-xylosidase/alpha-L-arabinofuranosidase 1-like	Glycine max
FYWAPTR	5	8	gi 383470439	gb AFH35543.1	chloroplast ribulose-1,5-bisphosphate	Brassica oleracea
LGGTTQYTVNNQMVNATLMNIADNPT	1	0	gi 300868066	ref ZP_071127	Ribulose biphosphate carboxylase/oxygenase	Oscillatoria sp. PCC

VESVPEMPLVISDSAESIEK	7	6	gi 413968492	gb AFW90583.	60S ribosomal protein L4/L1	Solanum tuberosum
DYPGTHEFLLLDEGKWQHVK	8	5	gi 356568654	ref XP_003552	sedoheptulose-1,7-bisphosphatase, chloroplastic-like	Glycine max
NFEGLDLGK	6	7	gi 3493460	gb AAC33509.1	glycolate oxidase, partial	Nicotiana tabacum
GIIEPISDSFINLVNADFTAK	4	9	gi 226498082	ref NP_001147	D-3-phosphoglycerate dehydrogenase	Zea mays
IAQIPVSEAYLGR	3	1	gi 115391887	ref YP_778475.	ATP synthase CF1 alpha subunit	Jasminum
GGYYDFVK	6	6	gi 387157286	dbj BAM15483.	carbonic anhydrase, partial	Solanum tuberosum
VSGVSLALFK	6	6	gi 195634715	gb ACG36826.	40S ribosomal protein S23	Zea mays
VVFVPWVETDFR	3	9	gi 356549980	ref XP_003543	cytochrome b6-f complex iron-sulfur subunit,	Glycine max
YAVVSAIAASAIPSLVMAR	7	5	gi 410908038	ref XP_003967	60S ribosomal protein L4-B-like	Takifugu rubripes
LDMSEYMER	6	6	gi 443309789	ref ZP_210394	ATPase with chaperone activity, ATP-binding subunit	Synechocystis sp.
IQPVYVVDVAAAIAALKDDGSSMGK	6	6	gi 356548101	ref XP_003542	NADH dehydrogenase ubiquinone 1 alpha	Glycine max
TLDTAHDDFHQPNFTGASIQK	1	1	gi 145974671	gb ABN13537.2	phospholipase D	Cucumis sativus
FLEYLDKDR	7	5	gi 449438119	ref XP_004136	ATP-dependent zinc metalloprotease FTSH 2,	Cucumis sativus
WGSTAFQNMLVPPGSGIVHQVNLEY	6	6	gi 350405429	ref XP_003487	cytoplasmic aconitate hydratase-like	Bombus impatiens
YAEDQEAFFKDYAEAHAK	8	4	gi 356515910	ref XP_003526	L-ascorbate peroxidase T, chloroplastic-like isoform	Glycine max
LDSTVISQPAIYVTSAAVEVLR	4	8	gi 357112661	ref XP_003558	malonyl-CoA-acyl carrier protein transacylase,	Brachypodium
VGAATETELED	6	6	gi 3790441	gb AAC68501.1	chaperonin 60 alpha subunit	Canavalia lineata
FNFSHGTHEYHQETLNNLR	7	5	gi 359484103	ref XP_003633	pyruvate kinase, cytosolic isozyme isoform 2	Vitis vinifera
HYETGESLPEEVYLK	7	5	gi 8978290	dbj BAA98181.	oligopeptidase A	Arabidopsis thaliana
VLYEVFPM SFLMEQAGGQAFTGK	5	7	gi 115336282	gb ABI94363.1	fructose-1,6-bisphosphatase 2	Oryza sativa
RQPEGPKPGFMVEGATLETVTPIPYD	6	6	gi 356516888	ref XP_003527	60S ribosomal protein L21-1-like	Glycine max
APAVDLQDGLISVAIGVAAQLSIEK	3	9	gi 449469377	ref XP_004152	inositol 2-dehydrogenase 2-like	Cucumis sativus
HKEHISAYGEGNER	1	0	gi 40457328	gb AAR86719.1	glutamine synthetase GS58	Nicotiana attenuata
LGDDEFGHMLAGILK	6	5	gi 51535181	dbj BAD38154.	putative fructokinase	Oryza sativa
DSPLDVIVVNDSSGGVK	6	5	gi 168523	gb AAA33484.1	glyceraldehyde-3-phosphate dehydrogenase	Zea mays
GVEYLVEASAAEAEISATLAAAK	7	4	gi 357471415	ref XP_003605	Serine/threonine protein phosphatase	Medicago truncatula
TQLGLPELTGLVIPGFGGTQR	7	4	gi 357141979	ref XP_003572	peroxisomal fatty acid beta-oxidation multifunctional	Brachypodium
EAPAFVEQATEQQILVTGIK	5	6	gi 212286514	dbj BAG82903.	ATP synthase beta subunit	Rhizobium sp. Cg-
QTDLPFLLAVEDVFSITGR	8	3	gi 218312	dbj BAA01975.	chloroplast elongation factor TuB (EF-TuB)	Nicotiana glauca
NLDFIQAAGLPLAIETAYEGLER	2	9	gi 449478569	ref XP_004155	quinone oxidoreductase-like protein At1g23740,	Cucumis sativus
VSGPVVVADGMAGAAMYELVR	7	4	gi 328874840	gb EGG23205.	vacuolar H ⁺ -ATPase A subunit	Dictyostelium

MDEWGV DVALTGSQK	4	7	gi 366984958	gb AEX09311.1	alanine glyoxylate aminotransferase	Peperomia andina
TSLWDL LDSTLYQHK	7	4	gi 106879583	emb CAJ38375	nucleoside-diphosphate-sugar dehydratase	Plantago major
GLLHSDQVLFNGGSTDSLVR	6	5	gi 356553152	ref XP_003544	peroxidase 4-like	Glycine max
RAQLGEIFELDR	4	7	gi 305671884	gb ADM63293.	photosystem II cp47 protein	Tetradoxa
LIYAEGAENGLEFTIVRPFNWIGPR	9	2	gi 347664493	gb AEP17007.1	UDP-D-apiose/UPD-D-xylose synthetase	Vitis vinifera
LYQVEYAMEAIGNAGSAIGILSK	4	7	gi 307186251	gb EFN71914.1	Proteasome subunit alpha type-4	Camponotus
YDYENV DAGAAK	3	8	gi 413933276	gb AFW67827.	phosphoglucosyltransferase, cytoplasmic 1	Zea mays
ISQEVSGDALGEEFKGYVFK	7	3	gi 356503097	ref XP_003520	40S ribosomal protein S6-like	Glycine max
ALVAPASLIAQNAGIEGEVVVEK	4	6	gi 358346943	ref XP_003637	RuBisCO large subunit-binding protein subunit alpha	Medicago truncatula
FGDAVHEDVGSR	5	5	gi 359481742	ref XP_003632	eukaryotic translation initiation factor 3 subunit G	Vitis vinifera
VIPDFPKPGIMFQDITLLLDTK	1	0	gi 356563876	ref XP_003550	adenine phosphoribosyltransferase 2-like isoform 1	Glycine max
SLTNDWEEHLAVK	5	5	gi 32967457	gb AAP51216.1	90-kDa heat-shock protein, partial	Haliclona rubens
QIQELVEAIVLPMTHK	6	4	gi 241236668	ref XP_002400	26S protease regulatory subunit 6A, putative	Ixodes scapularis
AVGDWIFYDR	5	5	gi 255557765	ref XP_002519	pectin acetylesterase, putative	Ricinus communis
TTQLRAPLLIAEDVTGEALATLVVNK	1	0	gi 449456032	ref XP_004145	ruBisCO large subunit-binding protein subunit alpha,	Cucumis sativus
VYISAASGAVGQLVGQFAK	7	3	gi 255577881	ref XP_002529	alcohol dehydrogenase, putative	Ricinus communis
LEGDNKYHAEHGLQDAK	8	2	gi 449501168	ref XP_004161	ubiquitin carboxyl-terminal hydrolase 12-like	Cucumis sativus
GVNPWIEVDGGVGPK	6	4	gi 255547345	ref XP_002514	ribulose-5-phosphate-3-epimerase, putative	Ricinus communis
RVEEVYYDVK	4	6	gi 255556962	ref XP_002519	translin	Ricinus communis
LTTFEDSEKESEYGYVR	5	5	gi 15982954	gb AAL11505.1	AF367445_1 V-ATPase catalytic subunit A	Prunus persica
LIFPFLELDIK	5	5	gi 15221788	ref NP_175836.	isocitrate dehydrogenase	Arabidopsis thaliana
GMPLEVITEFFAVGAK	2	8	gi 449475697	ref XP_004154	monosaccharide-sensing protein 2-like	Cucumis sativus
VVFVPDYNVSVAELLIPASELSQHISTA	7	3	gi 255570250	ref XP_002526	glycogen phosphorylase, putative	Ricinus communis
YGLDDNTIDFIGHALALHR	6	4	gi 82623395	gb ABB87112.1	GDP dissociation inhibitor 1-like	Solanum tuberosum
FTLPPFLTLK	5	5	gi 189418957	gb ACD93720.1	glycolate oxidase	Mikania micrantha
ETALLYDDQHSAANK	5	5	gi 356514669	ref XP_003526	puromycin-sensitive aminopeptidase-like	Glycine max
LIEGAGHMPQEDWPEK	5	5	gi 356526235	ref XP_003531	haloalkane dehalogenase 2-like	Glycine max
QFGYIVLTTSAGIMDHEEAR	5	5	gi 398407687	ref XP_003855	40S ribosomal protein S22	Zymoseptoria tritici
GIFWQEEIIPFFQAATLPK	8	2	gi 77416931	gb ABA81861.1	UDP-glucose:protein transglucosylase-like	Solanum tuberosum
VAVLGAAGGIGQPLALLMK	6	4	gi 350635444	gb EHA23805.1	malate dehydrogenase	Aspergillus niger
VSPSHVLDFQPGEAFMVR	4	6	gi 170102	gb AAA34026.1	carbonic anhydrase precursor	Spinacia oleracea

NEIRFEETLYGK	8	2	gi 255579134	ref XP_002530	sedoheptulose-1,7-bisphosphatase, chloroplast,	Ricinus communis
IETGHEDTVHVDVAMDYYGK	5	5	gi 356567182	ref XP_003551	protein SEC13 homolog	Glycine max
EVAFAAQFGSDLPATQALLNR	4	6	gi 357436047	ref XP_003588	ATP synthase subunit alpha	Medicago truncatula
LTHVYTAILKPDNELQILVDGEEK	2	8	gi 357520323	ref XP_003630	Calnexin-like protein	Medicago truncatula
LQEEEEAAEAEK	4	6	gi 255584185	ref XP_002532	50S ribosomal protein L29, chloroplast precursor,	Ricinus communis
GMAVEDPSAPHGLK	8	2	gi 71999169	gb AAZ57444.1	lipoxygenase LOX1	Populus deltoides
STNLDWYK	5	5	gi 449485915	ref XP_004157	elongation factor 1-alpha-like	Cucumis sativus
KFLVELR	4	5	gi 371926660	gb AEX58441.1	ATP synthase CF1 alpha subunit (chloroplast)	Theobroma
YHLPGETEEEFSTR	8	1	gi 449518935	ref XP_004166	gamma aminobutyrate transaminase 3, chloroplastic-	Cucumis sativus
YSEFLNAVK	4	5	gi 356508890	ref XP_003523	ATP-dependent zinc metalloprotease FTSH,	Glycine max
LSYGGAAQEYFGITPDLTTLGK	5	4	gi 307136451	gb ADN34256.1	glutamate-1-semialdehyde 21-aminomutase	Cucumis melo
LNPTSEYAPGLEDTLILTMK	6	3	gi 10303403	emb CAA65117	phosphoenolpyruvate carboxylase	Vanilla planifolia
VPTPNVSVVDLVVNVEKK	5	4	gi 7544132	dbj BAA94304.	NADP-glyceraldehyde-3-phosphate dehydrogenase	Chlamydomonas
IMLLPEDDPNATHIMIATGTGVAPFR	5	4	gi 195640470	gb ACG39703.	ferredoxin--NADP reductase, root isozyme	Zea mays
TFNQVEIKPEMIGHYLAEFSSISYKPVK	4	5	gi 395831618	ref XP_003788	40S ribosomal protein S15	Otolemur garnettii
VQDLLLLDVTPLSLGLETAGGVMTVLI	9	0	gi 1842232	gb AAC47456.1	heat shock protein 70	Babesia microti
KVDLFENGITIAIVEAVSPELGNR	3	6	gi 195623450	gb ACG33555.	FtsH6 - Zea mays FtsH protease	Zea mays
KELSETDSVYDSEK	4	5	gi 357167236	ref XP_003581	ruBisCO large subunit-binding protein subunit alpha,	Brachypodium
DVSPHEFVK	7	2	gi 449491502	ref XP_004158	40S ribosomal protein S19-3-like	Cucumis sativus
VTGDIYEGIAIGDVFPGSTLSDHVLRL	6	3	gi 15919087	emb CAC86995	ATP citrate lyase a-subunit	Lupinus albus
LAVEAWGLK	5	4	gi 357494811	ref XP_003617	Alanine glyoxylate aminotransferase	Medicago truncatula
MEVGFLGLGIMGK	3	6	gi 449458161	ref XP_004146	putative oxidoreductase GLYR1-like	Cucumis sativus
GHSYSEIINESLVDSLNPFMHAR	6	3	gi 195627844	gb ACG35752.	ketol-acid reductoisomerase	Zea mays
APLLIIAEDVTGEALATLVVNKLR	9	0	gi 356542646	ref XP_003539	ruBisCO large subunit-binding protein subunit alpha,	Glycine max
SNYNFEKPFLYLAR	4	5	gi 413950780	gb AFW83429.	ran GTP binding protein	Zea mays
FEETLYGK	4	5	gi 116582760	gb ABK00060.1	chloroplast sedoheptulose-1,7-bisphosphatase	Marchantia
FVHVSSAGVTRPDRPGLDSLK	6	3	gi 334186678	ref NP_001190	Rossmann-fold NAD(P)-binding domain-containing	Arabidopsis thaliana
VFEISLADLQGDHAYRK	5	4	gi 406870061	gb AFS65104.1	40S ribosomal protein S3a-like protein	Elaeis guineensis
LGSGISVQGNVDPAYLFSPLPALTDEI	5	4	gi 356549661	ref XP_003543	uroporphyrinogen decarboxylase 1, chloroplastic-like	Glycine max
QYETMAVLRPDMSEDER	5	4	gi 225426247	ref XP_002264	30S ribosomal protein S6 alpha, chloroplastic	Vitis vinifera
FGASELHAVAFIGGIASQEVK	5	4	gi 449452224	ref XP_004143	30S ribosomal protein S6 alpha, chloroplastic-like	Cucumis sativus

ENAPAIIFIDEVDAIATAR	1	8	gi 431907798	gb ELK11405.1	26S protease regulatory subunit 6B	Pteropus alecto
FQGGSPYIPLYGLGELPQAFAR	4	5	gi 40204891	emb CAF02075	GDP dissociation inhibitor	Medicago truncatula
EEVKVEVEDDR	3	6	gi 449454728	ref XP_004145	15.4 kDa class V heat shock protein-like	Cucumis sativus
ALNDHHVLLEGTLKPNMVTPGSDSP	1	8	gi 297803444	ref XP_002869	fructose-bisphosphate aldolase, cytoplasmic	Arabidopsis lyrata
KKPHVNIGTIGHVDHGK	7	1	gi 449529128	ref XP_004171	elongation factor Tu, chloroplastic-like	Cucumis sativus
SVPPRWDPALLSRFIHTYGTHIVIGMA	2	6	gi 413922466	gb AFW62398.	MAC/Perforin domain containing protein	Zea mays
IGVNAIPAIEEVNIFKDDLVIQFVNPK	5	3	gi 414873953	tpg DAA52510.	TPA: Transcription factor BTF3	Zea mays
YREAADIIK	8	0	gi 449459892	ref XP_004147	ribulose bisphosphate carboxylase/oxygenase	Cucumis sativus
ALLNAIGNLELTGAYAEALSK	5	3	gi 356573281	ref XP_003554	alpha-1,4 glucan phosphorylase L isozyme,	Glycine max
KGAALNAIQIAEMLL	4	4	gi 356524624	ref XP_003530	aspartate-semialdehyde dehydrogenase-like	Glycine max
QWYLQHYGVDIGR	6	2	gi 255587656	ref XP_002534	40S ribosomal protein S8, putative	Ricinus communis
YDSMLGTFK	3	5	gi 336390	gb AAD10210.1	glyceraldehyde 3-phosphate dehydrogenase B	Arabidopsis thaliana
LLTQFLEHLVSEGSQDVHVNALGK	5	3	gi 356495434	ref XP_003516	clathrin heavy chain 1-like	Glycine max
ESYDMLGISYDNHPR	4	4	gi 166092848	gb ABY82377.1	NADH dehydrogenase subunit J	Silene sorensenis
MGAEVYHHLK	6	2	gi 14423688	sp Q9LEJ0.1	Enolase 1 (2-phospho-D-glycerate hydro-lyase 1;	Hevea brasiliensis
ESPMPDPSELFNTNYYVK	5	3	gi 356500948	ref XP_003519	pyruvate dehydrogenase E1 component subunit	Glycine max
VLAMESEAEALGQLTTNELDGK	5	3	gi 359478255	ref XP_002273	membrane-associated 30 kDa protein, chloroplastic-	Vitis vinifera
LHGGTPANFLDVGGNASEGQVVEAFK	4	4	gi 146215968	gb ABQ10186.1	succinyl-CoA ligase beta subunit	Caragana jubata
LAEDVDLER	3	5	gi 308803516	ref XP_003079	Acyl-CoA synthetase (ISS)	Ostreococcus tauri
ALLFVDNSGADVVLGMLPLAR	6	2	gi 307212370	gb EFN88165.1	Pantothenate kinase 4	Harpegnathos
TEELQPYVLDVVK	5	3	gi 7548843	gb AAB26677.2	aspartate aminotransferase isozyme 5	Glycine max
FPFLLVDR	3	5	gi 313682962	ref YP_004060	3-hydroxyacyl-(acyl-carrier-protein) dehydratase	Sulfuricurvum
VFLMGEEVGEYQGAYK	3	5	gi 162464059	ref NP_001104	pyruvate dehydrogenase2	Zea mays
HYTAYDLDNWK	3	5	gi 356558612	ref XP_003547	beta-xylosidase/alpha-L-arabinofuranosidase 1-like	Glycine max
RPFAAIVGGSK	2	6	gi 83268386	gb ABB99705.1	3-phosphoglycerate kinase	Aegilops speltoides
LGFFDGDPSK	5	3	gi 356525896	ref XP_003531	beta-xylosidase/alpha-L-arabinofuranosidase 1-like	Glycine max
RDELTLEGIK	4	4	gi 443925317	gb ELU44176.1	ATP-dependent RNA helicase eIF4A	Rhizoctonia solani
EVAGFAPYEK	5	3	gi 357125815	ref XP_003564	60S ribosomal protein L36-2-like	Brachypodium
EAYPGDVFYLHSR	5	3	gi 449326777	gb AGE93355.1	ATP synthase CF1 alpha subunit	Alpinia zerumbet
FGDDAFNYKEEHDLDAAALKR	4	4	gi 21536967	gb AAM61308.	quinone oxidoreductase-like protein	Arabidopsis thaliana
LPGGMMIAVDLPGGQVYENTEVK	6	2	gi 356504704	ref XP_003521	ferredoxin-dependent glutamate synthase,	Glycine max

YGQDATNVGDEGGFAPNIQENKEGLE	6	2	gi 410925068	ref XP_003976	beta-enolase-like	Takifugu rubripes
QGNEYLINLIDSPGHVDFSSEVTAALR	6	2	gi 449503738	ref XP_004162	elongation factor 2-like, partial	Cucumis sativus
SAEVVNLVTQELGGK	6	2	gi 356531939	ref XP_003534	putative lactoylglutathione lyase-like	Glycine max
LSELLGVEVK	5	3	gi 413935730	gb AFW70281.	phosphoglycerate kinase	Zea mays
TTSCTPFIPTTGTGAWESALTNTLSP	2	6	gi 148524119	gb ABQ81922.1	aminotransferase 2	Cucumis sativus
RQPEGPKPGFMVEGATLETVTPIPYD	8	0	gi 225441539	ref XP_002280	60S ribosomal protein L21-1	Vitis vinifera
VLGGGVAGMLTAASPVQVIVGSFIADG	3	5	gi 255583444	ref XP_002532	DNA binding protein, putative	Ricinus communis
SDIIVSPSILSANFSK	4	4	gi 392412114	ref YP_006448	ribulose-5-phosphate 3-epimerase	Desulfomonile
GVSLLLPSDVIIADKFAPDANSK	3	5	gi 313585890	gb ADR71054.1	phosphoglycerate kinase	Nicotiana
YSITLLPGDGIGPEVISVAK	4	4	gi 318041639	ref ZP_079735	3-isopropylmalate dehydrogenase	Synechococcus sp.
ASAEVYVDVPDGWKER	4	4	gi 225450155	ref XP_002275	thylakoid lumenal 19 kDa protein, chloroplastic-like	Vitis vinifera
EVQKPLEDITDSLK	7	1	gi 15232123	ref NP_186797.	peptidyl-prolyl cis-trans isomerase CYP38	Arabidopsis thaliana
IGEVHEGTATMDWMEQEQR	3	5	gi 18407650	ref NP_564801.	elongation factor EF-G	Arabidopsis thaliana
QSAVIPLLDLAQQQHGGWLPVSAMNA	4	4	gi 449468942	ref XP_004152	NADH dehydrogenase	ubiquinone
SGEYEPSEKPEPDSYIR	4	4	gi 357111960	ref XP_003557	phospholipase D alpha 1-like	Brachypodium
SHGLSDSDFSAVIEALK	3	5	gi 255560761	ref XP_002521	3-hydroxyisobutyrate dehydrogenase, putative	Ricinus communis
ARFEELNMDLFR	4	4	gi 334855521	gb AEH16639.1	heat shock protein 70	Hymeniacidon sp.
ASSSYSSHEIDEQDNER	4	4	gi 154419981	ref XP_001583	Adenylate and Guanylate cyclase catalytic domain	Trichomonas
GDPTYLVVEK	1	7	gi 449438050	ref XP_004136	cytochrome b6-f complex iron-sulfur subunit,	Cucumis sativus
LDAQPELFIHIVPDK	4	4	gi 357114792	ref XP_003559	heat shock protein 81-1-like	Brachypodium
AQVSYAQPDDLIDFYHLK	6	1	gi 356526314	ref XP_003531	coatamer subunit beta-1-like	Glycine max
TLTHEETLEIPK	4	3	gi 449520511	ref XP_004167	gamma carbonic anhydrase-like 2, mitochondrial-like	Cucumis sativus
IIIDSNNNPEHFLTTNPYYDSR	7	0	gi 356540720	ref XP_003538	clathrin heavy chain 1-like	Glycine max
LTTFEDSEKESEYGYVRK	6	1	gi 15982954	gb AAL11505.1	AF367445_1 V-ATPase catalytic subunit A	Prunus persica
MFNDVQGTAGVAIAGLLGAVR	3	4	gi 444728483	gb ELW68940.	NADP-dependent malic enzyme, mitochondrial	Tupaia chinensis
WSPLTDVVDGYSPTLPHLK	7	0	gi 2773270	gb AAB96761.1	ferredoxin-dependent glutamate synthase	Glycine max
ALPSLAGLVVVDVVEGTAMASLIHMQK	6	1	gi 356538962	ref XP_003537	protein phosphatase methylesterase 1-like	Glycine max
LPLQDVYK	5	2	gi 449310747	gb AGE92582.1	elongation factor 1-alpha, partial	Triops granarius
TFQGPPHGIQVERDK	7	0	gi 384579082	gb AFI14485.1	ribulose-1,5-bisphosphate carboxylase/oxygenase	Leucopogon sp.
LQKPGSYEYLDK	3	4	gi 434386542	ref YP_007097	glutamate-1-semialdehyde-2,1-aminomutase	Chamaesiphon
TPGFSGADLANLLNEAAILAGRR	7	0	gi 350571030	ref ZP_089393	ATP-dependent metalloprotease FtsH	Neisseria

FVTAVVGFGK	3	4	gi 298712446	emb CBJ33221	calmodulin-like myosin-light chain	Ectocarpus
MVNHFVQEFK	5	2	gi 157382502	gb ABV48741.1	heat shock protein cognate 70	Bombus terrestris
LGEIVTTIPTIGFNVETVEYK	3	4	gi 351705750	gb EHB08669.1	ADP-ribosylation factor 5	Heterocephalus
KSGEYEPSEKPEPDSYIR	4	3	gi 449499356	ref XP_004160	phospholipase D alpha 2-like	Cucumis sativus
LALALGDENAVSMPVAAAANESFKK	4	3	gi 356520316	ref XP_003528	putative oxidoreductase GLYR1-like	Glycine max
LGDISNWLQK	4	3	gi 192910800	gb ACF06508.1	senescence-associated protein	Elaeis guineensis
GDLSPEELKER	4	3	gi 449460409	ref XP_004147	heat shock protein STI-like	Cucumis sativus
MPTFFETFPVVLVDGDGIVR	2	5	gi 305671916	gb ADM63316.	photosystem II cp47 protein	Moschopsis
HMYHDMYMK	6	1	gi 83284003	gb ABC01909.1	60S ribosomal protein L19-like protein	Solanum tuberosum
YVSGSYDTEDGFR	3	4	gi 356558777	ref XP_003547	glucose-6-phosphate 1-dehydrogenase, cytoplasmic	Glycine max
TMLGENLYPLVEQLEPEAAAK	5	2	gi 357145576	ref XP_003573	polyadenylate-binding protein 2-like	Brachypodium
IDIPYAASEDEL	3	4	gi 15228793	ref NP_191815.	GTP-binding protein SAR1	Arabidopsis thaliana
KRPVPIPGSGMQLTNIAHVR	7	0	gi 343172517	gb AEL98962.1	chloroplast stem-loop binding protein, partial	Silene latifolia
ITFPGEADEAPDVTGDIVFLQKQ	4	3	gi 413953636	gb AFW86285.	putative dnaJ chaperone family protein	Zea mays
SLQSVGQIIGEVLRPLDNER	4	3	gi 356551906	ref XP_003544	26S protease regulatory subunit 10B homolog A-like	Glycine max
TMREEGGYEAIK	6	1	gi 111162633	gb ABH07503.1	glutamine synthase, partial	Nicotiana attenuata
VGAATETELEDK	4	3	gi 1710807	sp P08926.2	RuBisCO large subunit-binding protein subunit alpha,	Pisum sativum
LLFAEDR	3	4	gi 18411523	ref NP_567209.	thylakoid rhodanese-like protein	Arabidopsis thaliana
LWGENFFDPATK	5	2	gi 356525774	ref XP_003531	elongation factor 2-like	Glycine max
FAPVAGYHWFDDGDMIHGMR	4	3	gi 82548252	gb ABB82946.1	carotenoid cleavage dioxygenase	Cucumis melo
YAEDQEAFFK	2	5	gi 356515910	ref XP_003526	L-ascorbate peroxidase T, chloroplastic-like isoform	Glycine max
LTYTTPDYETK	3	4	gi 7768578	dbj BAA95455.	ribulose 1,5-bisphosphate carboxylase/oxygenase	Gagea lutea
YVGIELWQVK	3	4	gi 255542876	ref XP_002512	calreticulin, putative	Ricinus communis
DGMKPHLVF	4	3	gi 297850244	ref XP_002893	60S ribosomal protein L6	Arabidopsis lyrata
GVSVPALDKDTLWEFQPK	2	5	gi 2506211	sp P13548.3	69 kDa - Vacuolar proton pump subunit alpha	Vigna radiata
QAELTEQWEHEK	3	4	gi 350534488	ref NP_001234	heat shock protein	Solanum
ELSETDSVYDSEKLAER	5	2	gi 145353092	ref XP_001420	chaperonin 60 alpha chain, chloroplast	Ostreococcus
LTEQVVFSDPYK	4	3	gi 255555259	ref XP_002518	Methylthioribose kinase, putative	Ricinus communis
QLPSGLLLVTGPFK	3	4	gi 449436263	ref XP_004135	60S ribosomal protein L6-like	Cucumis sativus
LNETEIIPLIPSVLQNLEHR	4	3	gi 297796593	ref XP_002866	coatamer subunit beta-1	Arabidopsis lyrata
MDFIPGIDGPSEGVPR	3	4	gi 347664493	gb AEP17007.1	UDP-D-apiose/UPD-D-xylose synthetase	Vitis vinifera

DLIGPAMYFGLMGDQGPIGR	3	4	gi 384248876	gb EIE22359.1	UDP-forming alpha-1,4-glucan-protein synthase	Coccomyxa
AVYECLR	1	6	gi 433804027	gb AGB51718.1	ribulose-1,5-bisphosphate carboxylase/oxygenase	uncultured
KTFAEENVNAEFR	3	3	gi 166702	gb AAD10209.1	glyceraldehyde 3-phosphate dehydrogenase A	Arabidopsis thaliana
MLLTLEK	3	3	gi 413968518	gb AFW90596.	putative 40S ribosomal protein S9	Solanum tuberosum
HNGNISLDDVIEIAK	6	0	gi 356537557	ref XP_003537	60S ribosomal protein L12-like	Glycine max
TQVAYGSKNEIRFEETLYGK	6	0	gi 99903817	gb ABF68592.1	chloroplast sedoheptulose-1,7-bisphosphatase	Marchantia
GWDAQVLGEAPYK	4	2	gi 357125604	ref XP_003564	2,3-bisphosphoglycerate-independent	Brachypodium
GAEAAALTAIEMASLFEHHLK	3	3	gi 195627252	gb ACG35456.	6,7-dimethyl-8-ribityllumazine synthase	Zea mays
SLAAAGAEILVGTWVPALNIFETSLRR	6	0	gi 7141083	gb AAF37208.1	AF207593_1 enoyl-ACP reductase	Arabidopsis thaliana
LSGTGSAGATVR	5	1	gi 357444163	ref XP_003592	Phosphoglucomutase	Medicago truncatula
KGGTSIEDLAEKFPDLIK	4	2	gi 83284007	gb ABC01911.1	succinyl CoA ligase beta subunit-like protein	Solanum tuberosum
AAEDANVYHDLVK	5	1	gi 356498960	ref XP_003518	clathrin heavy chain 1-like	Glycine max
EAIKDPGEFLLSDFSK	5	1	gi 1808656	emb CAA71762	Ubiquitin activating enzyme E1	Nicotiana tabacum
GYNELVLLSQTVGLQNYGAFLER	6	0	gi 34148077	gb AAQ62586.1	beta-galactosidase	Glycine max
LGDLFYR	4	2	gi 320043264	ref NP_001188	ubiquitin specific peptidase 54a	Danio rerio
VIYGADLAAFLQTLAK	4	2	gi 193290726	gb ACF17671.1	putative pyruvate dehydrogenase E2 subunit	Capsicum annuum
LDEIVVFDPLSHDQLR	1	5	gi 351725607	ref NP_001238	heat shock protein	Glycine max
VHINIVVIGHVDSGK	4	2	gi 345306703	ref XP_001514	elongation factor 1-alpha-like	Ornithorhynchus
YFEVELK	3	3	gi 59940382	gb AAX12729.1	glutamate decarboxylase	Oryza sativa
ESELTPSNANILDGR	2	4	gi 149390653	gb ABR25344.1	peptidyl-prolyl cis-trans isomerase	Oryza sativa Indica
MIPTKPMVVETFSEYPPLGR	3	3	gi 40287502	gb AAR83865.1	elongation factor 1-alpha	Capsicum annuum
YAEILASQGLLTTALEYLK	4	2	gi 449464024	ref XP_004149	protein transport protein Sec31A-like	Cucumis sativus
ELQETVQYPVEHPEKFEK	5	1	gi 303273578	ref XP_003056	cell division cycle protein 48	Micromonas pusilla
VFNQFTEQFSVNELASLVTK	4	2	gi 269854061	gb ACZ51244.1	UDP-sulfoquinovose synthase	Arachis hypogaea
VETGIKPGMVVTFGPTGLSTEVK	6	0	gi 126565779	gb ABO20800.1	translation elongation factor 1 alpha, partial	Plasmopara viticola
LADYFDVISGTSTGGLVTAMLAAPNEQ	6	0	gi 449461467	ref XP_004148	patatin group A-3-like	Cucumis sativus
DSGIGSQGITNSINMMTK	4	2	gi 164414441	ref NP_001105	NADP-dependent glyceraldehyde-3-phosphate	Zea mays
VVLVMESVADALVEK	5	1	gi 357437219	ref XP_003588	NADP-dependent glyceraldehyde-3-phosphate	Medicago truncatula
VTFLFLNLANDPTIER	3	3	gi 119874583	gb ABM05768.	vacuolar ATP synthase subunit B, partial	Phascolion
HLLHAYTPGTEPFELWEGLDRR	4	2	gi 449431858	ref XP_004133	prolycopene isomerase, chloroplastic-like	Cucumis sativus
FIFDAMHAVTGAYAKPIFVDK	2	4	gi 356535925	ref XP_003536	phosphoglucomutase, chloroplastic-like	Glycine max

VYGIPEVESAVLSGGK	4	2	gi 327493149	gb AEA86281.1	chloroplast cysteine synthase	Solanum nigrum
NIPLVFLQNITGFMVGSR	5	1	gi 87120143	ref ZP_010760	3-methylcrotonyl CoA carboxylase, beta subunit	Marinomonas sp.
LNAVESTFQDAESIAK	3	3	gi 383858535	ref XP_003704	tubulin alpha-1C chain-like	Megachile rotundata
KPADWDDREYIDDPNDVKPEGYDSIP	2	4	gi 225430904	ref XP_002276	calreticulin-3-like	Vitis vinifera
YTDDMELDDAVHTAILTLK	4	2	gi 195623644	gb ACG33652.	proteasome subunit alpha type 2	Zea mays
DYPGTHEFLLLDEGK	3	3	gi 357461143	ref XP_003600	Sedoheptulose-1,7-bisphosphatase	Medicago truncatula
EQLAIAEFAESLLIIPK	6	0	gi 356512570	ref XP_003524	T-complex protein 1 subunit alpha-like	Glycine max
VFVHVPSEIAAHEVEEIGVEHLLR	3	3	gi 340058004	emb CCC5235	26S proteasome regulatory subunit	Trypanosoma vivax
FESDETNGTAHFLEHMIFK	2	4	gi 410634	gb AAB28042.1	cytochrome c reductase-processing peptidase	potatoes, var.
ANNFDLMYEQVK	4	2	gi 349585018	dbj BAL03266.1	chloroplast phosphoribulokinase	Closterium
TPGFSGADLANLLNEAAILAGR	3	3	gi 257125522	ref YP_003163	ATP-dependent metalloprotease FtsH	Leptotrichia buccalis
FIIEAANHPTDPEADEILAK	3	3	gi 4150965	emb CAA09478	glutamate dehydrogenase	Asparagus
NKDTLEPLLDFLR	4	2	gi 151176306	gb ABR87939.1	sucrose synthase 1	Beta vulgaris
TFDPDDPEHMK	5	1	gi 255552279	ref XP_002517	ubiquitin-activating enzyme E1c, putative	Ricinus communis
MGLNDKEIVALSGAHTLGR	2	4	gi 449523509	ref XP_004168	L-ascorbate peroxidase T, chloroplastic-like, partial	Cucumis sativus
LLPLPELLQSIASIK	4	2	gi 399047049	ref ZP_107391	nucleoside-diphosphate-sugar epimerase	Brevibacillus sp.
LGINFFDTSPYYGGTLSEK	3	3	gi 299507808	gb ADJ21815.1	L-galactose dehydrogenase	Arabidopsis thaliana
AVQEEDELPEEK	2	4	gi 255538376	ref XP_002510	26S proteasome non-ATPase regulatory subunit,	Ricinus communis
VLDDYNVGDR	2	4	gi 365156331	ref ZP_093526	methionine synthase	Bacillus smithii
DLILQIIGEISVAGATYK	6	0	gi 357439825	ref XP_003590	3-isopropylmalate dehydratase	Medicago truncatula
MQGFLAASYHLYPK	4	2	gi 359478294	ref XP_002279	NADP-dependent alkenal double bond reductase P2	Vitis vinifera
LNIHEYQGAELMGK	5	1	gi 146215968	gb ABQ10186.1	succinyl-CoA ligase beta subunit	Caragana jubata
LELQEVVDFLK	3	3	gi 2077957	emb CAA73318	ATPase	Arabidopsis thaliana
YTGGMVPDVNQIIVK	4	2	gi 300681420	emb CBH32512	sedoheptulose-1,7-bisphosphatase, chloroplast	Triticum aestivum
QAVVNPENTFFSVK	3	3	gi 307107703	gb EFN55945.1	heat shock protein 70B	Chlorella variabilis
GVAMNPVEHPHGNGHGHASTV	2	4	gi 403352768	gb EJY75902.1	60S ribosomal protein L8	Oxytricha trifallax
VSSTETGVIFGNIVYETSNVLER	3	3	gi 255076481	ref XP_002501	coatomer subunit beta	Micromonas sp.
YADELIK	1	5	gi 432139325	gb AGB05600.1	fructose-bisphosphate aldolase 3	Camellia oleifera
VVGIDLGTNSAVAAMEGGKPTIVTNA	3	3	gi 1143427	emb CAA52149	heat shock protein 70	Cucumis sativus
RAQLGEIFELDRATLK	6	0	gi 305671884	gb ADM63293.	photosystem II cp47 protein	Tetradoxa
LAMQEFMILPVGASSFK	3	3	gi 289600010	gb ADD12953.1	2-phospho-D-glycerate hydrolase	Citrus trifoliata

AVIELENYGLPFSR	3	3	gi 343172675	gb AEL99041.1	succinate dehydrogenase	ubiquinone
DLVGVAYTEEETK	2	4	gi 7547401	gb AAB28813.2	cytochrome c1 precursor	Solanum tuberosum
SPSMGLLESTLFTGWMPILR	3	3	gi 18413285	ref NP_567349.	thioredoxin family protein	Arabidopsis thaliana
SSALLTVQTLLSELSSLHSR	3	3	gi 357492829	ref XP_003616	Sorting nexin-1	Medicago truncatula
YSGALSMGVPGEIAGLHEAWLQHGR	2	4	gi 356520836	ref XP_003529	gamma-glutamyltranspeptidase 1-like	Glycine max
VTTTPDDLILAER	4	2	gi 164604992	dbj BAF98292.	2-C-methyl-D-erythritol 4-phosphate	Hevea brasiliensis
VIVNTHAIEPQALVEFFGTLR	4	2	gi 410131067	gb AFV61748.1	clathrin heavy chain, partial	Leucaena
VFNQDLYFK	2	4	gi 449482370	ref XP_004156	pyruvate kinase, cytosolic isozyme-like	Cucumis sativus
LAEMPADSGYPAYLAAR	3	3	gi 269861253	ref XP_002650	vacuolar ATP synthase catalytic subunit A	Enterocytozoon
NGPGAPGGQSWTAEWLK	2	4	gi 2832920	dbj BAA24609.	thylakoid-bound ascorbate peroxidase	Spinacia oleracea
LMIVSOLDHTMVDHHPENMSLLR	3	3	gi 112383510	gb ABI17892.1	sucrose phosphatase	Coffea canephora
VIGTLLGSVLPDGTVDIR	4	2	gi 357492587	ref XP_003616	Eukaryotic translation initiation factor 3 subunit F	Medicago truncatula
GLDIPHVAHVNFDPNDIDDYVHR	3	3	gi 6735368	emb CAB68189	ATP-dependent RNA helicase-like protein	Arabidopsis thaliana
EALNSALDEEMSADPK	2	4	gi 162464059	ref NP_001104	pyruvate dehydrogenase2	Zea mays
LMHEAGYEIGNLDATLILQRPK	2	4	gi 374639357	gb AEZ55667.1	2-C-methyl-D-erythritol 2,4-cyclodiphosphate	Salvia miltiorrhiza
NLLAPSTPFFNTLYDPYR	1	5	gi 225456550	ref XP_002263	UDP-arabinopyranose mutase 1	Vitis vinifera
TVPPAVPGIVFLSGGQSEEEATLNLNA	1	5	gi 225708150	gb ACO09921.	Fructose-bisphosphate aldolase B	Osmerus mordax
HKEHIAAYGEGNER	5	0	gi 547508	emb CAA57346	glutamate--ammonia ligase	Glycine max
LELQEVVDLKNPDK	2	3	gi 356508890	ref XP_003523	ATP-dependent zinc metalloprotease FTSH,	Glycine max
FGFDEAFNYK	1	4	gi 342898879	gb AEL78825.1	ketone/zingerone synthase 1	Rubus idaeus
INQDTRDYFMSAK	2	3	gi 206901598	ref YP_002250	ATP-dependent Clp protease, proteolytic subunit	Dictyoglomus
MANYLASLNHEETR	3	2	gi 449531571	ref XP_004172	porphobilinogen deaminase, chloroplastic-like	Cucumis sativus
HAQGEINAGINVR	2	3	gi 255957396	emb CBA13552	cytosolic chaperonin, delta subunit	Glycine max
VLEALDDLKK	2	3	gi 255582939	ref XP_002532	30S ribosomal protein S20, putative	Ricinus communis
QFYVNVDKKEWK	3	2	gi 3776021	emb CAA09211	RNA helicase	Arabidopsis thaliana
FLPLGLGLLYLGK	2	3	gi 255728763	ref XP_002549	26S proteasome regulatory subunit rpn-1	Candida tropicalis
SVEMHHEALQEALPGDNVGFNVK	4	1	gi 413948814	gb AFW81463.	putative translation elongation factor Tu family protein	Zea mays
LGWHDAGTYNK	5	0	gi 3202024	gb AAC19393.1	thylakoid-bound L-ascorbate peroxidase precursor	Mesembryanthemu
GVSIDGKPLPFETGEIDFGPGTNGQH	3	2	gi 15239045	ref NP_199088.	glucose-6-phosphate isomerase	Arabidopsis thaliana
YGQDATNVGDEGGFAPNIQENK	3	2	gi 344242051	gb EGV98154.1	Alpha-enolase	Cricetulus griseus
KLQEEEEAAAEAEK	2	3	gi 255584185	ref XP_002532	50S ribosomal protein L29, chloroplast precursor,	Ricinus communis

GFIFGPPIALAIGAK	3	2	gi 356565339	ref XP_003550	adenine phosphoribosyltransferase 1, chloroplastic-	Glycine max
TGGLVWLAESFGHSVFK	5	0	gi 356504446	ref XP_003521	protein transport protein SEC23-like	Glycine max
VVDEVPWFGIEQEYTLQTNVK	2	3	gi 60686429	gb AAX35343.1	glutamine synthetase	Cucumis melo
YVYQPIEALYLLLVTNK	3	2	gi 359473067	ref XP_002276	coatomer subunit delta	Vitis vinifera
EATAFLLDVLKPNLPEHGFLQTK	3	2	gi 356551424	ref XP_003544	clathrin heavy chain 1-like	Glycine max
GVDLDALLDMSTDELVK	2	3	gi 356522093	ref XP_003529	40S ribosomal protein S15-like	Glycine max
GIYAYGF EKPSAIQQR	3	2	gi 386870481	gb AFJ42571.1	elongation initiation factor 4A	Sesamum indicum
IPYIDNPEFEDDPDLYVLKPIK	2	3	gi 296044570	gb ADG85705.	calreticulin	Triticum aestivum
FDSEVSGDREVQR	4	1	gi 283777738	gb ADB28898.1	26S protease regulatory subunit-like protein	Lolium perenne
AVIVGGGYIGLELSAALR	5	0	gi 302871151	ref YP_003839	FAD-dependent pyridine nucleotide-disulfide	Caldicellulosiruptor
GLTPYEFVEGMK	4	1	gi 357448925	ref XP_003594	ATP citrate lyase a-subunit	Medicago truncatula
DSFLDLHPEVSMLR	3	2	gi 356572556	ref XP_003554	cell division protein ftsZ homolog 2-1, chloroplastic-	Glycine max
TDHGEELLADVLFATGR	3	2	gi 50058094	dbj BAD27393.	glutathione reductase	Zinnia violacea
ANFLSSEDFEPPLIPAK	2	3	gi 255540003	ref XP_002511	calnexin, putative	Ricinus communis
EHALLAFTLGVK	2	3	gi 449463655	ref XP_004149	elongation factor 1-alpha-like	Cucumis sativus
EVLQREDDLNEIVQLVGK	3	2	gi 356600131	gb AET22418.1	vacuolar H ⁺ -ATPase	Citrus maxima
SVLGILGGGAGTR	4	1	gi 405132092	gb AFS17283.1	ADP glucose pyrophosphorylase, partial	Amaranthus
FGHEFLEFEFRPDGK	3	2	gi 195612480	gb ACG28070.	protein mago nashi	Zea mays
GSDHLRDVFGHMGSLDKDIALSGGH	5	0	gi 220898265	gb ACL81497.1	ascorbate peroxidase	Ginkgo biloba
VIATSAVFGTEIWPAAEYGR	3	2	gi 358636231	dbj BAL23528.1	N-acetylmuramoyl-L-alanine amidase	Azoarcus sp.
SMTQHNVGALVVVKPGEQK	3	2	gi 186489216	ref NP_001117	Cystathionine beta-synthase (CBS) family protein	Arabidopsis thaliana
VVNIPNSMTILDELLPISIEMAKR	5	0	gi 283488503	gb ADB24773.1	rhamnose synthase	Gossypium
IIIDSGDNPEHFLTTPPYDSR	3	2	gi 356540720	ref XP_003538	clathrin heavy chain 1-like	Glycine max
NLIEAAEQDYEKEKLNER	5	0	gi 449452644	ref XP_004144	ruBisCO large subunit-binding protein subunit beta,	Cucumis sativus
DKYSGVITYADLFQLASATAIEEAGGPK	3	2	gi 380863090	gb AFF18838.1	thylakoid ascorbate peroxidase precursor, partial	Dimocarpus longan
MYVTMDRYENDWNVVK	4	1	gi 71666782	ref XP_820347.	2,3-bisphosphoglycerate-independent	Trypanosoma cruzi
VVGFDQDKDVAVLR	3	2	gi 270342123	gb ACZ74706.1	serine-type peptidase	Phaseolus vulgaris
TTYVLALKDYPGTHEFLLLDGKWQH	5	0	gi 356539933	ref XP_003538	sedoheptulose-1,7-bisphosphatase, chloroplastic-like	Glycine max
LNLFPGYMER	1	4	gi 108709793	gb ABF97588.1	oxidoreductase, aldo/keto reductase family protein,	Oryza sativa
HIPVIDDK	2	3	gi 357491619	ref XP_003616	Cbs domain protein	Medicago truncatula
RDGYPSDPELIFLTDGASK	4	1	gi 225470587	ref XP_002273	alanine aminotransferase 2 isoform 1	Vitis vinifera

SGAAISMPGMMDTVNLGLNDQVVAG	5	0	gi 345303771	ref YP_004825	pyruvate, phosphate dikinase	Rhodothermus
TSEMSAVGLLPAALQGIDIR	4	1	gi 225459126	ref XP_002285	glucose-6-phosphate isomerase isoform 1	Vitis vinifera
VTSEDPDDGFKPTSGK	5	0	gi 1066857	gb AAA81578.1	acetyl-CoA carboxylase, partial	Glycine max
ATAGDTHLGGEDFDNR	3	2	gi 189380223	gb ACD93209.1	heat shock protein 70	Camellia sinensis
DMKPETFASLVSAILEYK	3	2	gi 21553663	gb AAM62756.	putative 20S proteasome beta subunit PBC2	Arabidopsis thaliana
KDDDLSEEDLALK	3	2	gi 322708848	gb EFZ00425.1	calcium-transporting ATPase type 2C member 1	Metarhizium
LLEDMAEK	5	0	gi 298707063	emb CBJ29865	ubiquitin protein ligase	Ectocarpus
EDFGGGHPDPNLTYAK	2	3	gi 238479033	ref NP_001154	phosphoglucomutase	Arabidopsis thaliana
LGTFWPYTPSIQLLYGLR	2	3	gi 366984898	gb AEX09281.1	alanine glyoxylate aminotransferase	Peperomia
VIVAIENPQDIIVQSAR	2	3	gi 366984898	gb AEX09281.1	alanine glyoxylate aminotransferase	Peperomia
VNLIGEHDYEGYSVLPMAIR	3	2	gi 126277534	ref XP_001369	n-acetylgalactosamine kinase	Monodelphis
YSGALESFYER	4	1	gi 350537279	ref NP_001234	vacuolar H ⁺ -ATPase A2 subunit isoform	Solanum
EEVAWYAGK	3	2	gi 413952851	gb AFW85500.	sesquiterpene cyclase1	Zea mays
FGFDEAFNYKEEPDLTAALK	3	2	gi 342898881	gb AEL78826.1	ketone/zingerone synthase 2	Rubus idaeus
LIGQGIPEDFLVSAVAEFK	3	2	gi 408717411	gb AFU52882.1	Cu/Zn-superoxide dismutase copper chaperone	Vitis vinifera
YTLDVDLK	3	2	gi 228991106	ref ZP_041510	4-hydroxy-2-oxovalerate aldolase	Bacillus
HSEFISYPISLWIEK	1	4	gi 2982291	gb AAC32131.1	heat shock protein	Picea mariana
EKFPGANDFGSEVIPGATSIGMR	4	1	gi 21401	emb CAA39181	ADP-glucose pyrophosphorylase	Solanum tuberosum
AYLHAEADVER	3	2	gi 225456483	ref XP_002280	magnesium-protoporphyrin O-methyltransferase	Vitis vinifera
ALPYLSGVILFEETLYQK	5	0	gi 218662307	ref ZP_035182	Fructose-bisphosphate aldolase	Rhizobium etli
KVEDPDMLER	3	2	gi 402566739	ref YP_006616	phospholipid/glycerol acyltransferase	Burkholderia
FGEGDEESMSRPFVSVLLIAGHDEK	4	1	gi 449455403	ref XP_004145	proteasome subunit alpha type-5-like	Cucumis sativus
QELAEMIFQSAL	3	2	gi 21537178	gb AAM61519.	xylose isomerase	Arabidopsis thaliana
MLDAVLLNMR	2	3	gi 255577881	ref XP_002529	alcohol dehydrogenase, putative	Ricinus communis
MEAEEGFDATR	3	2	gi 356540108	ref XP_003538	protein transport protein SEC23-like isoform 1	Glycine max
DGASGGVASVYYVGPDGWK	2	3	gi 356508794	ref XP_003523	proteasome subunit beta type-5-like	Glycine max
GVLGDAENLTEER	3	2	gi 357514625	ref XP_003627	(+)-neomenthol dehydrogenase	Medicago truncatula
TSVEGVFAAGDVQDHEWR	3	2	gi 110739775	dbj BAF01794.	thioredoxin reductase like protein	Arabidopsis thaliana
YLEDVMAHK	2	3	gi 284177758	gb ADB81472.1	60S ribosomal protein L17	Ignatius tetrasporus
VFGISMEPSILEAYVK	1	4	gi 32140557	dbj BAC78195.	prolyl-tRNA synthetase	Raphanus sativus
FVEFYGDGMGELSLADR	3	2	gi 262369900	ref ZP_060632	aconitate hydratase 1	Acinetobacter

VLQDGAVWEK	5	0	gi 356553641	ref XP_003545	coproporphyrinogen-III oxidase, chloroplastic-like	Glycine max
TEQQVSQLVAQLFQMR	2	3	gi 359481606	ref XP_002276	signal recognition particle 54 kDa protein,	Vitis vinifera
YDGVVGVGLGAIEAINVLK	4	1	gi 255551929	ref XP_002517	N-carbamoyl-L-amino acid hydrolase, putative	Ricinus communis
LWQVPETLPPEVLGK	3	2	gi 58500257	gb AAW78582.	quinone reductase 2	Triticum
VDFAYGFFEK	2	3	gi 313586457	gb ADR71239.1	60S ribosomal protein L3A	Hevea brasiliensis
IVANVVGDPALFNEWK	2	3	gi 225430398	ref XP_002285	aspartate aminotransferase, chloroplastic	Vitis vinifera
YQAHDPENQFK	4	1	gi 162463621	ref NP_001105	30S ribosomal protein S17, chloroplastic precursor	Zea mays
GHILNLGHGVLVGTPEEAVAHFFEVAR	3	2	gi 356549661	ref XP_003543	uroporphyrinogen decarboxylase 1, chloroplastic-like	Glycine max
WIEYDDDNPIPQREEDITK	3	2	gi 18403643	ref NP_564596.	ubiquitin carboxyl-terminal hydrolase 6	Arabidopsis thaliana
ASALDGMSDLSIVQMPR	3	2	gi 429123668	ref ZP_191842	phosphoribosylaminoimidazole carboxylase catalytic	Brachyspira
LWGENFFDPATKK	3	2	gi 449532873	ref XP_004173	elongation factor 2-like, partial	Cucumis sativus
APLEVFEFR	2	3	gi 428777022	ref YP_007168	DNA gyrase subunit A	Halothece sp. PCC
GDLGMEIPVEK	3	2	gi 52547714	gb AAU81892.1	pyruvate kinase	Phaeodactylum
ALIPDLYR	2	3	gi 15225693	ref NP_180811.	carboxymethylenebutenolidase	Arabidopsis thaliana
ILIDVSKIDMSTTVLGFK	5	0	gi 304368145	gb ADM26718.	glycolate oxidase	Nicotiana
MVDESDLR	2	3	gi 257440141	ref ZP_056158	bacteriocin	Faecalibacterium
GISDIITVPGLVNVDFADVR	3	2	gi 357131317	ref XP_003567	cell division protein ftsZ homolog 2-1, chloroplastic-	Brachypodium
GYEAAFGPIDETMESK	2	3	gi 356564135	ref XP_003550	subtilisin-like protease-like	Glycine max
GNEMAEVLMDFPQLTMTLPDGR	3	2	gi 149390851	gb ABR25443.1	vacuolar ATP synthase	Oryza sativa Indica
FVIQGFGNVGAWAAQLISEQGK	3	2	gi 255568914	ref XP_002525	glutamate dehydrogenase, putative	Ricinus communis
HYFQNTQGLIFVDSNDRDR	1	4	gi 38353642	gb AAR18698.1	ADP-ribosylation factor 1	Populus tomentosa
RKDPGTVFVAGATGQAGIR	2	3	gi 30686068	ref NP_565789.	NAD(P)-binding Rossmann-fold-containing protein	Arabidopsis thaliana
EYLQGNPLWLQYVK	3	2	gi 225458416	ref XP_002283	30S ribosomal protein S9, chloroplastic	Vitis vinifera
DAALNAIQSPLLDIGIER	1	4	gi 356546085	ref XP_003541	cell division protein ftsZ homolog 2-1, chloroplastic-	Glycine max
WIVSEKEDYK	1	4	gi 392374328	ref YP_003206	Histidine kinase	Candidatus
QNLAAATFVNAFVNAGFGQDK	2	3	gi 357474443	ref XP_003607	26S proteasome non-ATPase regulatory subunit	Medicago truncatula
GTEHEDEYMISEK	1	4	gi 357461137	ref XP_003600	Trafficking protein particle complex subunit	Medicago truncatula
VQQLQDFFNGK	2	3	gi 2642648	gb AAB97316.1	cytosolic heat shock 70 protein	Spinacia oleracea
FHAFLEDR	3	2	gi 326479546	gb EGE03556.1	chromatin remodeling complex subunit Arp8	Trichophyton
YEQIAELLHGIAER	2	3	gi 10130004	gb AAG13459.1	AF128453_1 gamma-glutamylcysteine synthetase	Glycine max
DVVATIVYELMPNPSDHK	1	4	gi 192910716	gb ACF06466.1	thioredoxin family Trp26-like protein	Elaeis guineensis

FLKPSVAGFLLQK	1	1	gi 1022805	gb AAB60303.1	phosphoglycerate kinase, partial	Arabidopsis thaliana
LIEEQGTNALAAALHLSGFSQPPSSR	7	0	gi 108711873	gb ABF99668.1	DEAD/DEAH box helicase family protein, expressed	Oryza sativa
AAVPSGASTGVYEALRLR	4	2	gi 110288668	gb ABG65935.1	Enolase, putative, expressed	Oryza sativa
MGINPIMMSAGELESGNAGEPAK	1	9	gi 12643757	sp Q40460.1	ribulose biphosphate carboxylase/oxygenase	Nicotiana tabacum
HGHVDYDKVPGVVYTHPEVASVGKTE	3	6	gi 134142802	gb ABO61735.1	mitochondrial lipoamide dehydrogenase	Populus tremuloides
AAEPFFLLAGPNVIESEEHILR	5	0	gi 13620976	dbj BAB41014.	2-dehydro-3-deoxyphosphooctonate aldolase	Arabidopsis thaliana
ADRDESSPYAAMLAAQDVSQR	3	2	gi 149390859	gb ABR25447.1	40S ribosomal protein S14	Oryza sativa Indica
LVASLPEGGVLLLENVR	7	2	gi 149392717	gb ABR26161.1	phosphoglycerate kinase	Oryza sativa Indica
SILVLGGAGGVGTQIIQLAK	2	9	gi 15220854	ref NP_173786.	quinone oxidoreductase-like protein	Arabidopsis thaliana
FEEKDGIDYAAVTVQLPGGER	4	8	gi 161728809	dbj BAF94218.	33kDa oxygen evolving protein of photosystem II	Chara braunii
DGIDYAAVTVQLPGGER	3	8	gi 161728809	dbj BAF94218.	33kDa oxygen evolving protein of photosystem II	Chara braunii
FILIGSGNPEEGELRPQLLDR	5	1	gi 163846014	ref YP_001634	magnesium chelatase ATPase subunit I	Chloroflexus
PSTLTALAEK	6	0	gi 164454783	dbj BAF96938.	flavanone 3-hydroxylase	Nicotiana tabacum
DYFGAHTYER	4	2	gi 168702634	ref ZP_027349	6-phosphogluconate dehydrogenase	Gemmata
SNIVGLPVSLLLK	1	5	gi 17017265	gb AAL33590.1	methylenetetrahydrofolate	Zea mays
ADAFEYADQVLEK	4	4	gi 18405061	ref NP_564667.	thylakoid lumen 18.3 kDa protein	Arabidopsis thaliana
AAEEVIFGEPEVTTGAAGDLQQITGLA	1	1	gi 187830110	ref NP_001120	filamentation temperature-sensitive H 2B	Zea mays
GLDVDALYISHIQVNQAQK	3	4	gi 192912960	gb ACF06588.1	ribosomal protein L17	Elaeis guineensis
ISALSEAGVIYVMTHDSIGLGEDGPTH	6	1	gi 194396261	gb ACF60500.1	plastid transketolase	Nicotiana tabacum
AFDEGPWPK	0	5	gi 195622136	gb ACG32898.	aldehyde dehydrogenase	Zea mays
AAVEEGIVPGGGVALLYASK	4	2	gi 22250	emb CAA78101	mitochondrial chaperonin-60	Zea mays
FGESVWFK	4	5	gi 224064850	ref XP_002301	light-harvesting complex II protein Lhcb2	Populus trichocarpa
STPFQPYSEVFGFLQR	3	2	gi 224090986	ref XP_002309	light-harvesting complex II protein Lhcb4	Populus trichocarpa
LTLEPASGGDQTTLEADVVLVSAGR	6	5	gi 224099079	ref XP_002311	dehydrogenase dihydrolipoamide dehydrogenase 1	Populus trichocarpa
EIVPGMIVTGMEVAEIDGAPR	7	4	gi 224831511	gb ACN66755.1	thiazole biosynthetic enzyme	Carica papaya
RLTYDEIQSK	5	0	gi 22530906	gb AAM96957.	33 kDa polypeptide of oxygen-evolving complex	Arabidopsis thaliana
LLVLLDDLCLK	3	2	gi 225428556	ref XP_002284	dolichyl-diphosphooligosaccharide--protein	Vitis vinifera
AGELSAEELDKLMTIVANPR	2	4	gi 225432662	ref XP_002282	40S ribosomal protein S18-like	Vitis vinifera
AMGIMNSFINDIFEKLAQEASR	1	0	gi 225435054	ref XP_002284	histone H2B-like	Vitis vinifera
KEWGVPADLFGTLPGR	7	5	gi 225444633	ref XP_002275	peroxiredoxin Q, chloroplastic	Vitis vinifera
AVALVLPTLK	4	5	gi 225451685	ref XP_002278	glyceraldehyde-3-phosphate dehydrogenase A,	Vitis vinifera

NDEELGKLLAGVTIAHGGVLPNINPVLL	1	0	gi 225455268	ref XP_002273	6-phosphogluconolactonase 4, chloroplastic	Vitis vinifera
IAIVEQSVSPGGGAWLGGQLFSAMVV	4	6	gi 225557533	gb EEH05819.1	thiazole biosynthetic enzyme	Ajellomyces
YIGSLVGDFHR	6	1	gi 226498474	ref NP_001150	fructose-1,6-bisphosphatase	Zea mays
ALDDTIITGVPTTIDYHK	5	2	gi 239759398	gb ACS14663.1	biotin carboxylase	Camellia oleifera
DSPLDVIAINDTGGVK	5	5	gi 251831338	gb ACT21568.1	glyceraldehyde-3-phosphate dehydrogenase	Bruguiera
LLALMGMPFR	2	3	gi 255541698	ref XP_002511	50S ribosomal protein L5, putative	Ricinus communis
AGRPIPIPNSGIQMTQLGHVK	4	4	gi 255542956	ref XP_002512	NAD dependent epimerase/dehydratase, putative	Ricinus communis
AFKDMEELEK	5	0	gi 255545916	ref XP_002514	glucose-6-phosphate isomerase, putative	Ricinus communis
KIELKPGFLMEHDKDVGAVVVGFD	2	4	gi 255547472	ref XP_002514	4-nitrophenylphosphatase, putative	Ricinus communis
MGGHFVQGHVDGTGVIVEK	1	4	gi 255553759	ref XP_002517	Riboflavin synthase alpha chain, putative	Ricinus communis
LLGDVTIANGGVMPNIHNLPLK	3	3	gi 255563635	ref XP_002522	histone h2a, putative	Ricinus communis
SDYKDAFVEK	4	4	gi 255578100	ref XP_002529	dihydrolipoamide succinyltransferase component of	Ricinus communis
AASMDEVLR	2	4	gi 255581842	ref XP_002531	hydroxypyruvate reductase, putative	Ricinus communis
SLPGITIDEEKIVSSTGALALK	7	0	gi 281206416	gb EFA80603.1	dihydrolipoamide:NAD oxidoreductase	Polysphondylium
MFGDFQKDTPEER	3	6	gi 28190676	gb AAO33154.1	transketolase	Oryza sativa
TPGHPENFETPGVEVTTGPLGQGAN	2	1	gi 297810173	ref XP_002872	chloroplast transketolase	Arabidopsis lyrata
GGSTGYDNAVALPAGGR	4	4	gi 30013657	gb AAP03871.1	oxygen evolving complex 33 kDa photosystem II	Nicotiana tabacum
AHGTAUGLPSEDDMGNSEVGHNALG	4	2	gi 302390383	ref YP_003826	phosphoglycerate mutase	Thermosediminibact
AFADIEIGIPFMETSAK	3	3	gi 303748	dbj BAA02115.	GTP-binding protein	Pisum sativum
LLEQDNPDLYDAK	5	5	gi 30685604	ref NP_850203.	chaperonin CPN60-like 1	Arabidopsis thaliana
NHYVEGGDAGNRENYINELIR	6	4	gi 307176212	gb EFN65855.1	60S ribosomal protein L7	Camponotus
VGLVAPIDVVPPGNTGLDPSQTSFFQ	9	2	gi 313586451	gb ADR71236.1	60S acidic ribosomal protein P0B	Hevea brasiliensis
AAPLIALANYIAYR	2	3	gi 336359762	gb AEI53622.1	chloroplast geranylgeranyl diphosphate synthase	Catharanthus
FKFPENGVELYAEK	4	4	gi 343887316	dbj BAK61862.	40S ribosomal protein S3	Citrus unshiu
AIGPQDVLATLLNNLK	5	3	gi 345489666	ref XP_001599	splicing factor 3B subunit 1-like	Nasonia vitripennis
AASAAGTIMTLSSWATSSVEEVASTG	1	1	gi 3493460	gb AAC33509.1	glycolate oxidase, partial	Nicotiana tabacum
RGAEVYQLGPLNEYER	3	3	gi 350536711	ref NP_001234	glyoxisomal malate dehydrogenase	Solanum
AHTDAGGIILLFQDDKVSGLQLLK	3	2	gi 350539815	ref NP_001234	1-aminocyclopropane-1-carboxylate oxidase 1	Solanum
MEITNVMEYEAIAK	2	2	gi 351726610	ref NP_001238	peroxisomal glycolate oxidase	Glycine max
AMPNILMLRPADGNETAGAYK	4	3	gi 356495111	ref XP_003516	transketolase, chloroplastic-like	Glycine max
AIKDGIABVEKPIYNHVTGLDPPPELIKPP	4	1	gi 356495988	ref XP_003516	phosphoribulokinase, chloroplastic-like	Glycine max

YAMLGAVGAIPEILGK	4	3	gi 356501755	ref XP_003519	chlorophyll a-b binding protein 3, chloroplastic-like	Glycine max
RFEGLGWHVIWVK	6	0	gi 356506190	ref XP_003521	transketolase, chloroplastic-like	Glycine max
IGIINAPQWYDAGK	1	6	gi 356516812	ref XP_003527	chlorophyll a-b binding protein P4, chloroplastic-like	Glycine max
GNLMSYFLDR	2	3	gi 356531959	ref XP_003534	50S ribosomal protein L5, chloroplastic-like	Glycine max
ILNMVEGVLLVDSVEGPMPQTR	9	9	gi 356534728	ref XP_003535	GTP-binding protein TypA/BipA homolog	Glycine max
VLPGLLGGSADLASSNMTLLK	1	3	gi 356536526	ref XP_003536	transketolase, chloroplastic-like	Glycine max
APLLIIAEDVTGEALATLVVNK	1	1	gi 356539334	ref XP_003538	ruBisCO large subunit-binding protein subunit alpha,	Glycine max
AAVLGFVGAPFTLASVVEGGSSK	1	1	gi 356540042	ref XP_003538	uroporphyrinogen decarboxylase, chloroplastic-like	Glycine max
DAPMFVVGVNEK	0	5	gi 356548623	ref XP_003542	aldehyde dehydrogenase family 2 member B4,	Glycine max
ALDAGIVAQAALDVFTVEPPPQDSK	5	0	gi 356550691	ref XP_003543	D-3-phosphoglycerate dehydrogenase, chloroplastic-	Glycine max
FDLMLVGMGPDGHVASLFGHPHLLK	5	8	gi 356551542	ref XP_003544	6-phosphogluconolactonase 4, chloroplastic-like	Glycine max
AEIPEEEKNLGPDR	4	1	gi 356554393	ref XP_003545	ubiquitin carboxyl-terminal hydrolase 12-like	Glycine max
SEGGILLPSSAQTKPQAGEVVAVGEG	5	4	gi 356556406	ref XP_003546	20 kDa chaperonin, chloroplastic-like	Glycine max
AIGFIDTMIESLANAPLQK	9	1	gi 356566218	ref XP_003551	enoyl-	...
YIDDLKDPGSPGKPYAR	1	6	gi 356568028	ref XP_003552	fructose-1,6-bisphosphatase, chloroplastic-like	Glycine max
ALPTYTPESPADATR	2	3	gi 356576867	ref XP_003556	transketolase, chloroplastic-like	Glycine max
AGTPSPTGFVMNLQPSEAEAGFDIR	3	2	gi 357438995	ref XP_003589	aminoacylase-1	Medicago truncatula
QNLGWPYEPFHPEDVK	1	1	gi 357445031	ref XP_003592	transketolase	Medicago truncatula
NLMEDEPEK	1	4	gi 357467543	ref XP_003604	60S ribosomal protein L5	Medicago truncatula
NLMEDEPEKYQSHFSEYIK	1	4	gi 357467543	ref XP_003604	60S ribosomal protein L5	Medicago truncatula
YAGTEVEFDGSSHLILK	0	5	gi 357479371	ref XP_003609	20 kDa chaperonin	Medicago truncatula
QVGNMYTASLYAAFASLLHNK	3	3	gi 357500885	ref XP_003620	hydroxymethylglutaryl-CoA synthase	Medicago truncatula
IELKPGFLMEHDKDVGAVVVGFR	4	4	gi 357507861	ref XP_003624	phosphoglycolate phosphatase	Medicago truncatula
FYWAPTRDDR	1	3	gi 358249078	ref NP_001240	ribulose bisphosphate carboxylase/oxygenase	Glycine max
YSGVGAAIEYAVLHLK	1	3	gi 359473182	ref XP_003631	carbonic anhydrase, chloroplastic-like isoform 1	Vitis vinifera
LKDSQNTFAMLTTFNEVDMTNLMK	4	2	gi 359483352	ref XP_002266	dihydrolipoyllysine-residue succinyltransferase	Vitis vinifera
VTTTIGYGSPNKANSYSVHGSALGAK	1	0	gi 359490179	ref XP_002266	transketolase, chloroplastic-like	Vitis vinifera
YLKPAVAGFLMQK	4	2	gi 3738259	dbj BAA33802.	cytosolic phosphoglycerate kinase 1	Populus nigra
NPLDLPWK	6	8	gi 379131258	emb CCF55355	glyceraldehyde-3-phosphate dehydrogenase A	Cucumis sativus
KDSPLDVIAINDTGGVK	5	3	gi 379131258	emb CCF55355	glyceraldehyde-3-phosphate dehydrogenase A	Cucumis sativus
ADELMEFYK	3	3	gi 380798339	gb AFE71045.1	adenylate cyclase type 1, partial	Macaca mulatta

DAVTHLTDAQEWAGGGSVMGAFVGS	1	2	gi 383100942	emb CCD7448	2-phosphoglycolate phosphatase 1	<i>Arabidopsis halleri</i>
ANTFVAEVLGLDPR	7	1	gi 383857729	ref XP_003704	probable malate dehydrogenase, mitochondrial-like	<i>Megachile rotundata</i>
ALGGSSLV DGQIELMLHR	1	5	gi 391346165	ref XP_003747	lysosomal alpha-mannosidase-like	<i>Metaseiulus</i>
AFGMELLR	5	2	gi 401662260	emb CCG0634	chloroplast lycopene epsilon cyclase precursor	<i>Mychonastes</i>
VFPNGEVQYLHPK	2	3	gi 413932723	gb AFW67274.	photosystem I reaction center subunit II	<i>Zea mays</i>
ELDYLVGAVSNPK	7	9	gi 413935730	gb AFW70281.	phosphoglycerate kinase	<i>Zea mays</i>
FLKPSVAGFLLQKELDYL VGAVSNPK	6	0	gi 413935730	gb AFW70281.	phosphoglycerate kinase	<i>Zea mays</i>
EAGSTVEV VAAQTK	4	1	gi 413954394	gb AFW87043.	putative serine/threonine protein phosphatase	<i>Zea mays</i>
MEQTFIMIKPDGVQR	4	3	gi 414886742	tpg DAA62756.	TPA: nucleotide diphosphate kinase1, partial	<i>Zea mays</i>
HTLEGPDDMPAHK	6	2	gi 424841562	ref ZP_182661	secondary thiamine-phosphate synthase enzyme	<i>Saprospira grandis</i>
LGGTTQYTVNNQMVNATLMNIADNPT	6	0	gi 428227229	ref YP_007111	ribulose 1,5-bisphosphate carboxylase small subunit	<i>Geitlerinema sp.</i>
AITVFSPDGHLFQVEYALEAVR	3	6	gi 440637274	gb ELR07193.1	20S proteasome subunit alpha 3	<i>Geomyces</i>
KVYSDLSVDYLK	2	3	gi 440740028	ref ZP_209195	phosphoenolpyruvate carboxykinase	<i>Pseudomonas</i>
AFIAGVADDNGYGWAIK	4	7	gi 443267280	ref YP_007374	NAD(P)H-quinone oxidoreductase subunit H	<i>Ophioglossum</i>
AGGYSSDGYQSVLDK	4	2	gi 448591017	ref ZP_216507	fructose-1,6-bisphosphatase	<i>Haloferax elongans</i>
HYAHVDCPGHADYVK	4	7	gi 449327205	gb AGE93761.1	translation elongation factor Tu (chloroplast)	<i>Pavlova lutheri</i>
LGGTTQYTVNNQMVNATLMNIADNPT	8	0	gi 449441384	ref XP_004138	ribulose bisphosphate carboxylase/oxygenase	<i>Cucumis sativus</i>
AKMELSSLTQTNISLPFITATADGPK	6	0	gi 449442347	ref XP_004138	stromal 70 kDa heat shock-related protein,	<i>Cucumis sativus</i>
VLQNDIDLLNPPELEK	4	3	gi 449445790	ref XP_004140	40S ribosomal protein S27-2-like	<i>Cucumis sativus</i>
GEYVDMVK	4	5	gi 449447855	ref XP_004141	chaperonin CPN60-like 2, mitochondrial-like	<i>Cucumis sativus</i>
AHVLAHESGYHQLISHWLR	1	3	gi 449448970	ref XP_004142	lipoxygenase 2, chloroplastic-like	<i>Cucumis sativus</i>
YGAGIGPGVYDIHSPR	5	5	gi 449452654	ref XP_004144	5-methyltetrahydropteroyltriglutamate-homocysteine	<i>Cucumis sativus</i>
ALASYMFNTEELVR	3	2	gi 449460533	ref XP_004148	chaperone protein ClpB3, chloroplastic-like	<i>Cucumis sativus</i>
GHPHAQAWDDALSK	1	4	gi 449465814	ref XP_004150	phosphomethylpyrimidine synthase, chloroplastic-like	<i>Cucumis sativus</i>
AEATDVANAVLDGSDAILGAETLR	5	0	gi 449482372	ref XP_004156	pyruvate kinase, cytosolic isozyme-like	<i>Cucumis sativus</i>
AAVLGFVGAPWTIATYIVEGGTTR	9	3	gi 449483285	ref XP_004156	uroporphyrinogen decarboxylase 1, chloroplastic-like	<i>Cucumis sativus</i>
NDFTPEEEEEVRR	4	7	gi 449495074	ref XP_004159	SKP1-like protein 1A-like	<i>Cucumis sativus</i>
EVDATRQNLGWPYEPFHVPEVK	6	0	gi 449498919	ref XP_004160	transketolase, chloroplastic-like	<i>Cucumis sativus</i>
GTMTTTHSYTGDQR	6	3	gi 449501309	ref XP_004161	glyceraldehyde-3-phosphate dehydrogenase A,	<i>Cucumis sativus</i>
GTVEITPVELIK	2	4	gi 449516071	ref XP_004165	60S acidic ribosomal protein P0-like	<i>Cucumis sativus</i>
AIFASGSPFDPVEYEGK	5	4	gi 46850202	gb AAT02534.1	NADP-dependent malic enzyme 2	<i>Hydrilla verticillata</i>

SGIYYTPEQEK	7	5	gi 4884033	emb CAB43186	peptide methionine sulfoxide reductase	Arabidopsis thaliana
LTYTLDEIEGPFVSPDGTIK	3	1	gi 5052366	gb AAD38521.1	33 kDa oxygen evolving protein of photosystem II	Brassica napus
LTYTLDEIEGPFVSPDGTIKFEEK	3	4	gi 5052366	gb AAD38521.1	33 kDa oxygen evolving protein of photosystem II	Brassica napus
FADENFIKK	6	3	gi 51235021	gb AAT98376.1	peptidyl-prolyl cis-trans isomerase	Populus trichocarpa
AEELGLLSLEK	6	5	gi 62319859	dbj BAD93901.	cytochrome P450 monooxygenase	Arabidopsis thaliana
EGPPTFEQPK	2	6	gi 68565781	sp Q7X999.1	ribulose biphosphate carboxylase/oxygenase	Larrea tridentata
AGVKDYKLTYYTPDYETK	7	0	gi 7240538	gb AAB63985.2	ribulose-1,5-bisphosphate carboxylase/oxygenase	Syzygium smithii
YAHEETVATASFAGK	4	1	gi 356545301	XP_003541082	4-hydroxy-3-methylbut-2-enyl diphosphate reductase-	Glycine max
FGFDEAFNYKEEPDLTAALKR	1	5	gi 414879280	tpg DAA56411.	TPA: putative alcohol dehydrogenase superfamily	Zea mays
MNDSFADGMLHLLYFK	5	0	gi 356518342	ref XP_003527	carotenoid 9,10(9',10)-cleavage dioxygenase-1 like	Glycine max
SEDLLAAFR	5	1	gi 495673466	WP_00839804	carboxypeptidase	Clostridium sp.
NMFVVHQYQSSTTPGAAAGSAESYSK	0	6	gi 386818449	ref ZP_101056	condensation domain protein	Thiothrix nivea DSM
WWHAKRTVLNLPGLGGGGSWLK	0	6	gi 414885330	tpg DAA61344.	TPA: putative trehalose phosphatase/synthase family	Zea mays
LWEPLRYWVTGVHRNGYSNLLLEWK	0	5	gi 440295971	gb ELP88818.1	coatomer subunit alpha-3, putative	Entamoeba
AVYYDPETGKEDPSK	5	0	gi 1706781	sp P53991.1	Ferredoxin-NADP reductase, chloroplastic;	Chlamydomonas
QELQLGSFPTILFFPK	5	4	gi 359483358	ref XP_003632	5'-adenylylsulfate reductase 3, chloroplastic-like	Vitis vinifera
LKYYGAMGGSLAFVGLKFWHYR	0	1	gi 502125880	ref XP_004499	nucleobase-ascorbate transporter 3-like	Cicer arietinum
TYRGNNELPTVLYVTER	0	7	gi 359481150	ref XP_003632	PHD finger protein ALFIN-LIKE 1 isoform 4	Vitis vinifera
SSKWYFWYMDTEDPPLSYEAFKEHR	0	1	gi 480330422	ref YP_007872	arabinose efflux permease family protein	Propionibacterium
YYNATSLKVQDLNLYNYRR	0	5	gi 417399509	gb JAA46757.1	Putative secreted protein precursor	Desmodus rotundus