Universidade Federal do Estado do Rio de Janeiro
Instituto de Biociências
Programa de Pós-Graduação em Ciências Biológicas
Mestrado em Biodiversidade Neotropical

Ellen Moura Lopes

Vanilla bahiana, fonte alternativa da Mata Atlântica para a produção de baunilha: uma abordagem proteômica através de *nanoLC-MS* de alta definição.

Rio de Janeiro

2018

ELLEN MOURA LOPES

Vanilla bahiana, fonte alternativa da Mata Atlântica para a produção de baunilha: uma abordagem proteômica através de *nanoLC-MS* de alta definição.

Dissertação apresentada ao Programa de Pós-Graduação, S*tricto sensu* em Ciências Biológicas do Centro de Ciências Biológicas e da Saúde da Universidade Federal do Estado do Rio de Janeiro, como requisito parcial para a obtenção do título de Mestre em Ciências Biológicas.

Orientadora: Profa. Dra. Andrea Furtado Macedo

Coorientadora: Profa. Dra. Maria Gabriela Bello Koblitz

Rio de Janeiro, RJ

2018

Catalogação informatizada pelo(a) autor(a)

L864 Lopes, Ellen Moura
Vanilla bahiana, fonte alternativa da Mata Atlântica para a produção de baunilha: uma abordagem proteômica através de nanoLC-MS de alta definição. / Ellen Moura Lopes. -- Rio de Janeiro, 2018.
85

Orientadora: Andrea Furtado Macedo.
Coorientadora: Maria Gabriela Bello Koblitz.
Dissertação (Mestrado) - Universidade Federal do Estado do Rio de Janeiro, Programa de Pós-Graduação em Ciências Biológicas, 2018.

1. Vanilla bahiana Hoehne. 2. compostos fenólicos. 3. Vanilina. 4. Dodecil sulfato de sódio. 5. ?-mercaptoetanol. I. Macedo, Andrea Furtado, orient. II. Koblitz, Maria Gabriela Bello, coorient. III. Título.

UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
INSTITUTO DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

Folha de aprovação

Ellen Moura Lopes

Vanilla bahiana, fonte alternativa da Mata Atlântica para a produção de baunilha: uma abordagem proteômica através de *nanoLC-MS* de alta definição

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Biológicas (Biodiversidade Neotropical) da Universidade Federal do Estado do Rio de Janeiro como requisito para obtenção do título de Mestre em Biodiversidade Neotropical.

Rio de Janeiro, 22 de fevereiro de 2018.

Banca Examinadora:

Prof.ª Dra. Andrea Furtado Macedo (Orientadora)
Universidade Federal do Estado do Rio de Janeiro

Prof.ª Dra. Maria Gabriela Bello Koblitz (Coorientadora)
Universidade Federal do Estado do Rio de Janeiro

Prof. Dr. Joel Campos de Paula
Universidade Federal do Estado do Rio de Janeiro

Prof.ª Dra. Eidy de Oliveira Santos
Fundação Centro Universitário Estadual da Zona Oeste

AGRADECIMENTOS

À Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), por financiar e tornar possível o desenvolvimento da pesquisa. Obrigada também à UNIRIO, PPGBIO e ao CNPQ, por me concederem uma bolsa de incentivo à ciência.

A minha orientadora, Professora Dr.ª Andrea Furtado Macedo, pela coordenação do projeto, pela oportunidade e apoio de anos. Agradeço a sua dedicação, paciência, competência, generosidade, revisões e sugestões, que foram fundamentais para a minha vida acadêmica e para a conclusão desta dissertação. Sou grata por todo seu carinho e todas as conversas.

A coorientação da Professora Dr.ª Maria Gabriela Bello Koblitz, pela dedicação, apoio, competência, generosidade e suas revisões e sugestões que também foram fundamentais para conclusão deste trabalho.

A todos os professores do mestrado PPGBIO que de alguma forma contribuíram positivamente para minha formação.

Aos meus colegas do mestrado pelo apoio ao longo desses anos. As minhas amigas Izabella Fontenelle e Aloma Nogueira pelas conversas e momentos de descontração.

A Unirio, seus funcionários e servidores do setor de transporte, biblioteca, segurança, limpeza e secretaria.

Aos meus amigos do Laboratório Integrado de Biologia Vegetal (LIBV) que me acompanham diariamente: Ana Carolina Pereira, Gustavo Bocayuva, Vinícius Portella e Fernanda D'Andrea, assim como outros amigos da Unirio. Em especial gostaria de agradecer as amigas Roberta Linhares e Joana Oliveira pelo apoio, conversas e risadas ao som Vanilla Ice. Não teria conseguido sem vocês ao meu lado, sempre me ajudando e compartilhando.

Aos coautores e colaboradores desse estudo.

Ao Instituto Chico Mendes de Conservação da Biodiversidade (ICMBIO) por conceder licenças para as coletas. A secretaria de meio ambiente pelas permissões de coleta e auxílio.

A minha família, meus pais Cristiane e Mauricio Lopes por toda a paciência, apoio de muitos anos, dedicação pela minha educação e crescimento. Ao meu irmão, Eric, por todas as nossas conversas e seu apoio. Ao Johnny por todo o companheirismo e carinho.

Gostaria de agradecer aos meus amigos, à Rayanne Luiz pelos anos de amizade, por me apoiar, ouvir e por todo o carinho. Gostaria de dedicar essa dissertação a memória de Cíntia Silva, uma amiga que estaria terminando o mestrado esse ano. "As vezes a vida é iluminada por pessoas tão especiais, que nos tornamos felizes só porque um dia fizemos parte de suas vidas".

RESUMO

O extrato natural de baunilha tem grande importância econômica, sendo a vanilina um de seus principais componentes. A produção desse extrato é cara, laboriosa e demorada para a demanda mundial. Atualmente, existe uma grande perda de variabilidade genética das espécies de *Vanilla*. Fatores como o desmatamento, mudanças climáticas, doenças e extrativismo predatório impactam na sobrevivência das espécies de *Vanilla*. Com isso, a caracterização bioquímica dessas espécies tem se mostrado uma alternativa para a conservação, produção e desenvolvimento de indivíduos mais resistentes. Sendo assim, o objetivo desse trabalho foi caracterizar a expressão proteica dos frutos maduros de *Vanilla bahiana*, avaliando a melhor metodologia de extração. Seis soluções foram selecionadas: Vb1 -Tris-HCl; Vb2 – solução Vb1 + 0,5% de β-mercaptoetanol (β-MT); Vb3 -Vb2 + 1% de dodecil sulfato de sódio (SDS); Vb4 – Vb2 + 0,1% de SDS; Vb5 - Vb1 + 1% de SDS e Vb6 -Vb1 + 0,1% de SDS. Após a extração, as proteínas digeridas foram analisadas por cromatografia liquida acoplada a espectrômetro de massas. O software *Progenesis QI* foi utilizado resultando na identificação de 2326 proteínas, sendo 135 relacionadas à floração e frutificação e 65 à biossíntese de metabólitos ligados ao aroma e sabor de baunilha. A maior diversidade de proteínas foi obtida nas extrações com 1% de SDS. As proteínas identificadas nos frutos de *V. bahiana* confirmam o potencial enzimático dessa espécie na produção de compostos, já descritos na literatura, como responsáveis pelo aroma e sabor do extrato natural de baunilha, potencial esse validado pela quantificação de vanilina nos frutos da espécie estudada.

Palavras-chave: Vanilla bahiana Hoehne; compostos fenólicos; Bottom-up; Vanilina; Dodecil sulfato de sódio; β-mercaptoetanol

ABSTRACT

The natural extract of vanilla has great economic importance, being vanillin one of its main components. The production of this extract is expensive, laborious and time consuming for world demand. Currently there is a countless loss of genetic variability of the producer species. Factors such as deforestation, climatic changes, diseases and predatory extractivism impact the survival of *Vanilla spp*. Thus, the biochemical characterization of these species has shown to be an alternative for the conservation, production and development of more resistant individuals. Therefore, the objective of this work is to characterize the protein expression of mature fruits of *Vanilla bahiana*, evaluating the best extraction methodology. Six solutions were selected: Vb1-Tris-HCl; Vb2 - Vb1 + 0.5% β-mercaptoethanol solution (β-MT); Vb3 -Vb2 + 1% sodium dodecyl sulfate (SDS); Vb4 - Vb2 + 0.1% SDS; Vb5-Vb1 + 1% SDS and Vb6 -Vb1 + 0.1% SDS. After extraction, the digested proteins were analyzed by liquid chromatography coupled to a mass spectrometer. The Progenesis QI software was used resulting in the identification of 2326 proteins, 135 related to flowering and fruiting, and to the biosynthesis of metabolites linked to the aroma and flavor of vanilla. The highest protein diversity was obtained in extractions with 1% SDS. The proteins identified in the fruits of *V. bahiana* point to the enzymatic potential of this species in the production of compounds, already described in the literature, as responsible for the aroma and flavor of the natural extract of vanilla, potential validated by the quantification of vanillin in fruits of the studied species.

Key-words: Vanilla bahiana Hoehne*; Bottom up; Vanillin; phenolic compounds; sodium dodecyl sulfate; β-mercaptoethanol*

LISTA DE FIGURAS

Introdução

Capítulo 1

Material suplementar

LISTA DE TABELAS

Capítulo 1

Material suplementar

LISTA DE SIGLAS E ABREVIATURAS

µL – Microlitro

nL - Nanolitro

mM – Millimolar

µm - Micrometro

fmol – Fentomol

spp. – Espécies

US$ - Dolar americado

UK – Reino Unido

USA – Estados Unidos da América

IMS- *Ion mobility separation* (separação de mobilidade iónica)

MS – Espectrometria de Massas

LC-MS – Cromatografia liquida acoplado a Espectrômetro de Massas

UPLC – Cromatografia líquida de Ultra Performace

HPLC – Cromatografia líquida de alta eficiência

TCA – Ácido tricloroacético

PCA – Análise de componentes principais

TIC - *Total ion account* (Contagem total de ions)

GFP – *[Glu1]-Fibrinopeptide B human (*Fibrinopeptídeo B humano)

T- wave – *Traveling-wave*TOF – *Time-of-flight* (Tempo de voo)

eV- eletron Volt

CID – *Collision-induced dissociation* (dissociação induzida por colisão)

KDa – Kilo Dalton

FDR – False Discovery Rate

CV- Coeficiente de variação

KEGG – Enciclopédia de Genes e Genoma de Kioto

SUMÁRIO

Introdução

Cultivo e produção de baunilha

A tribo Vanilleae pertence à família Orchidaceae e contém 10 gêneros, incluindo o gênero pantropical *Vanilla* Miller, popularmente conhecida como baunilha (Pansarin, Aguiar, and Ferreira 2012). As espécies de *Vanilla* são monofiléticas e representadas por aproximadamente 120 espécies, até então descritas (Ormerod and Cootes 2013). A maior diversidade dessas espécies está concentrada em regiões tropicais, principalmente nos biomas brasileiros (Chase et al. 2015; Pansarin, Aguiar, and Ferreira 2012). Muitas espécies deste gênero são consideradas raras ou ameaçadas devido ao: desmatamento de seu habitat original (tipicamente subcosmopolita), mudanças climáticas, exploração predatória e agentes patogênicos pandêmicos (Divakaran et al. 2015; Divakaran, Babu, and Peter 2006). No entanto, esse gênero é economicamente importante devido à presença de vanilina produzida nos frutos das *Vanilla spp*. (Rain and Group 2004).

A vanilina é o principal componente do sabor e aroma da baunilha, que por sua vez é um dos sabores naturais mais populares do mundo, devido à sua importância na indústria de alimentos, farmacêutica, perfumaria e cosméticos (Pansarin, Aguiar, and Ferreira 2012). Estudos vêm mostrando a eficiência da vanilina, e outros compostos derivados do extrato de baunilha, contra diversas doenças devido a suas características antioxidantes, anticancerígenas, antimutagênica, dentre outras (Anuradha, Shyamala, and Naidu 2013). A vanilina (4-hydroxy-3-methoxybenzal- dehyde) é um aldeído aromático, que pertencente ao grupo de compostos fenólicos simples (C6-C1) (Figura 1) e é encontrada em uma concentração mais elevada nos frutos maduros de *V. planifolia* (Palama 2014).

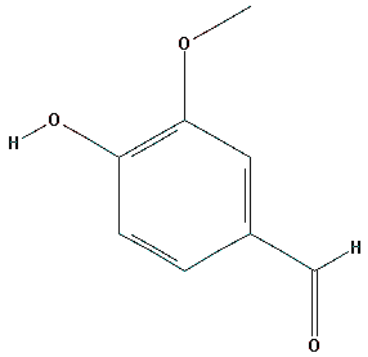


Figura 1. Estrutura química da Vanilina. Fonte: *https://pubchem.ncbi.nlm.nih.gov/compound/vanillin#section=2D-structure*

Baunilha é o condimento cultivado mais valioso e o terceiro mais caro do mundo após o açafrão e o cardamomo (Hrazdina 2006). Devido à sua alta demanda, a produção global chega a 5600 toneladas de frutos curados e o extrato natural de vanilina custa cerca de US$ 1.200 a US$ 4.000/kg (Rubert et al. 2016). A produção natural de baunilha é cara, laboriosa e demorada (Divakaran et al. 2015) (Figura 2). A polinização das flores de *Vanilla* é realizada à mão, onde para se obter 1 Kg de vanilina são necessários aproximadamente 500 kg do fruto, correspondente a 40.000 flores de *V. planifolia* polinizadas (Gallage and Møller 2018). O cultivo clonal é geralmente aplicado a duas espécies: *Vanilla planifolia* G. Jackson (syn. *Vanilla fragrans* Andrews) e *Vanilla tahitensis* Moore, com *V. planifolia* fornecendo 95% da produção mundial (Kahane et al. 2008). Apesar da grande importância, a produção clonal de *V. planifolia* provocou uma redução na variabilidade genética e deixou as espécies vulneráveis a doenças, que atualmente afetam negativamente a produção mundial de baunilha em 50-90% (Gallage and Møller 2018; Pinaria, Liew, and Burgess 2010). O preço do extrato natural vem sendo elevado devido à alta demanda e fontes naturais limitadas (Greule et al. 2015). Atualmente, o extrato natural de baunilha chega a custar em média US$600/Kg, preço que vem aumentando devido ao desmatamento e às catástrofes naturais, como ciclones, que destroem milhares de hectares de cultivos de *V. planifolia* no principal país produtor, Madagascar (Strong 2017).

Uma alternativa mais barata ao extrato natural é o uso da vanilina sintética, que usa o guaiacol e a lignina como compostos de partida para a produção de vanilina sintética (Gallage and Møller 2018). Sendo mais barata, essa é usada por cerca de 50% do mercado mundial para as mais diversas finalidades, como na indústria alimentícia (Walton, Mayer, and Narbad 2003). A síntese química da vanilina tem suas desvantagens, que atualmente não se encaixam na demanda consciente do uso de recursos e preservação ambiental. A síntese química da vanilina via lignina gera 160 Kg de resíduos por 1 Kg de vanilina obtida, consequentemente provocando um impacto ambiental negativo (Hocking 1997). O extrato natural de baunilha possui uma qualidade de sabor e aroma superior ao sintético, relacionado a uma mistura de diversos compostos. Até então, 200 compostos aromáticos foram identificados nos frutos curados de *Vanilla spp*. (Medina, Rodriguez Jiménes, and García 2009). Vinte e seis compostos fenólicos, com concentrações acima de 1 mg/Kg, foram identificados como responsáveis pelo aroma e sabor característico da baunilha, dentre os quais os mais frequentemente citados pela literatura são: vanilina (4-hidroxi-3-metoxibenzaldeído), álcool de vanilina, ácido vanílico, álcool 4-hidroxibenzilo, 4-

hidroxibenzaldeído, ácido 4-hidroxibenzóico, álcool anisílico, anisaldeído e ácido anísico (Sharma et al. 2007; Pérez-Silva et al. 2006).

Os frutos fermentados, ou curados, de *V. planifolia* contêm aproximadamente 2% de vanilina, dependendo do seu local de origem: México com cerca de 1,75%, Sri Lanka 1,5% e Indonésia 2,75% (Parthasarathy, Chempakam, and Zachariah 2008). Foi possível constatar que além da espécie, outros fatores podem interferir com a qualidade do extrato de baunilha natural como: o processo de cura, estágio de maturação do fruto, os nutrientes do solo, calor, incidência luminosa, regime de chuvas, microrganismos presentes no processo de cura dos frutos, dentre outros (Palama 2014; Gu et al. 2017; Baqueiro-Peña and Guerrero-Beltrán 2016). A maior concentração de vanilina e outros compostos fenólicos ligados ao sabor e aroma de baunilha podem ser usados como indicadores de sua qualidade para fins comerciais ou até mesmo marcadores de estágio de desenvolvimento (Sharma et al. 2007; Pérez-Silva et al. 2006; Greule et al. 2015).

Frutos maduros produzem uma concentração superior de compostos fenólicos (Medina, Rodriguez Jiménes, and García 2009). O processo de verificação da maturação dos frutos de *V. planifolia* deve ser feito de duas a três vezes por semana, uma vez que os frutos imaturos e maduros apresentam praticamente a mesma coloração e não apresentam nem tamanho, nem aroma distintos (Medina, Rodriguez Jiménes, and García 2009).

Com base nos argumentos acima relacionados à difícil produção, à susceptibilidade das espécies, à alta demanda e ao aumento dos preços, existe um esforço mundial para procurar novas espécies de *Vanilla*. Essas espécies poderiam melhorar a cultura, o sabor e o aroma de baunilha, visando aumentar a produção de ingredientes ativos e ampliar os recursos genéticos (Virol et al. 2016; Anuradha, Shyamala, and Naidu 2013). Para isso, é necessária uma caracterização química dos frutos das *Vanilla spp.*

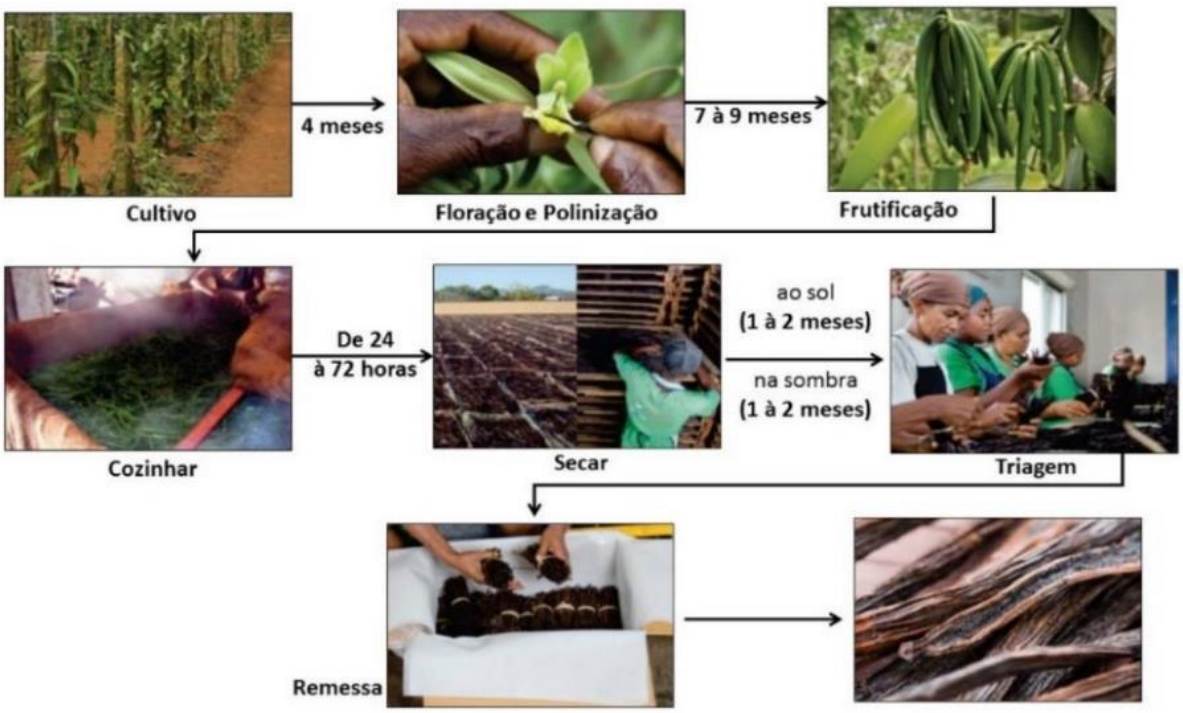


Figura 2. Processo de produção e cura dos frutos de baunilha natural. Fonte: http://www.provagourmet.us/the-vanilla-process

A espécie brasileira: Vanilla bahiana

Vanilla bahiana Hoehne é uma espécie endêmica da Floresta de Mata Atlântica brasileira, e ainda cientificamente e economicamente inexplorada (Figura 3). Existem poucos trabalhos publicado com *V. bahiana*, em geral estes estudos descrevem sua reprodução, filogenia e localização (de Fraga, Couto, and Pansarin 2017; Villanueva-Viramontes, Hernández-Apolinar, Fernández-Concha, et al. 2017; Sambin and Chiron 2015; Moreira, Barberena, and Lopes 2014; Odoux 2011). Esta espécie é filogeneticamente próxima a *V. planifolia* e ocorre nas regiões do Pará, Pernambuco, Bahia, Espírito Santo e Rio de Janeiro, especialmente em restingas, em áreas de caatinga, cerrado e na borda da Floresta Atlântica (Villanueva-Viramontes, Hernández-Apolinar, Carnevali Fernández-Concha, et al. 2017; Gigant et al. 2011; Bouetard et al. 2010).

A *V. bahiana* é autogâmica, como a maioria das *Vanilla spp*., mas depende de polinizadores para a reprodução. A *V. bahiana*, assim como grande parte das *Vanilla spp*., é hemiepífita (herbáceas). Essa apresenta flores com sépalas verdes e pétalas brancas levemente amareladas e oblanceoladas de ápice agudo (Figura 3). A floração dessa espécie

se estende por oito meses (de novembro a junho), com um pico em abril (Anjos, Barbarena, and Pigozzo 2016).

Como outras espécies do gênero, a *V. bahiana* enfrenta problemas quanto à redução da população, o que enfatiza a urgência de sua caracterização química, ainda não publicada (Moreira, Barberena, and Lopes 2014).

Figura 3. Flor e fruto de *Vanilla bahiana* no Monumento Natural do Pão de Açúcar e Urca, cidade do Rio de Janeiro (RJ, Brasil). Foto por: Ellen Lopes e Roberta Linhares.

Estudos proteômicos de Vanilla spp.

Recentes avanços em biotecnologia permitiram uma alternativa ao método de síntese química da vanilina, a bioengenharia de vanilina natural (Gallage and Møller 2015; Busconi et al. 2017; Gallage and Møller 2018; Chee et al. 2017). Embora, existam muitos esforços para estudar as *Vanilla spp*. e desenvolver a bioengenharia de vanilina, de acordo com Gallage e Møller (2018), ainda persiste a necessidade de se caracterizar e conhecer as vias de biossíntese, não apenas de vanilina, mas dos demais compostos característicos do aroma e sabor de baunilha (Gallage and Møller 2018). De acordo com o mesmo, as *Vanilla spp*. produzem em seus frutos vanilina em concentrações elevadas e até então, não existe outro organismo biológico conhecido na natureza, que consiga produzir vanilina nessa concentração (Gallage and Møller 2018, 2015).

A proteômica colabora no entendimento do funcionamento celular, permitindo a compreensão da função das proteínas nas células, fornecendo as informações das modificações pós-transducionais dos genes. O objetivo final da proteômica é identificar todas as proteínas em uma célula e determinar a função de cada uma, assim desvendando as

vias de biosíntese de compostos de interesse de um organismo em um determinado momento (Wilson and Walker 2010).

No presente trabalho, foi utilizado o método "*bottom-up*" que identifica proteínas digeridas por processo enzimático ou químico antes da análise por *LC-MS*. As proteínas foram digeridas diretamente em uma mistura complexa, e os peptídeos resultantes foram analisados (Martins Ferreira, Guest, and Martins-de-souza 2017; Bond et al. 2013).

Até então, os estudos proteômicos sobre o gênero *Vanilla* estão restritos ao desenvolvimento de calos através da cultura de tecidos, revelando o ineditismo e a importância do presente trabalho (Guerrero et al. 2011; Tan et al. 2013; Tan et al. 2014; Palama et al. 2010;). Especificamente, o estudo de Gallage et al. (2014) utilizou uma análise combinada de transcriptômica e proteômica em frutos de *V. planifolia*, com foco em algumas enzimas sugeridas pela literatura, com o intuito de propor uma via de biossíntese de vanilina mais completa (Gallage et al. 2014). De acordo com esse estudo, existe uma enzima chave, a vanilina sintase, responsável por catalisar a clivagem da dupla ligação de carbono do ácido ferúlico e de seu glicosídeo em vanilina e seu glicosídeo, respectivamente (Figura 4). A vanilina sintase, de acordo com o estudo, pertence à família de proteases de cisteínas, que são conhecidas por possuírem funções fisiológicas versáteis (Gallage et al. 2014). Apesar de trabalhos mais recentes terem sidos publicados, de acordo com Kundu (2017), a via biosintética proposta por Gallage et al (2014) continua sendo a mais aceita (Kundu 2017).

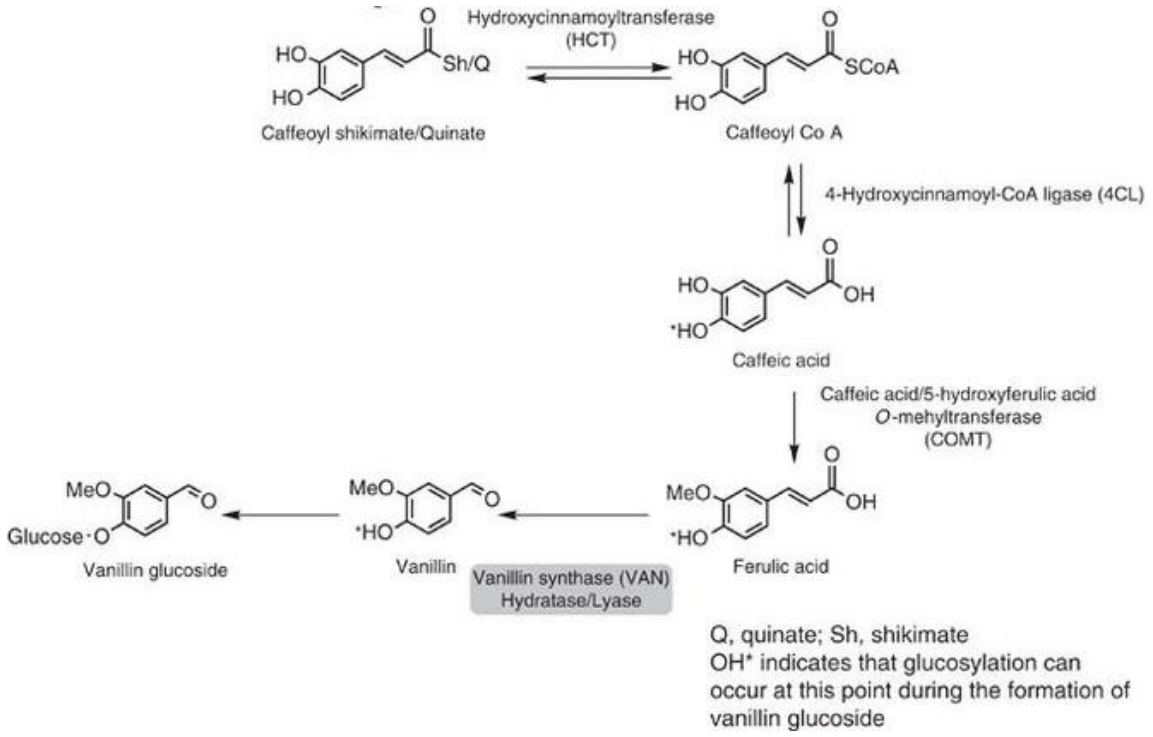


Figura 4. Parte final da via de biossíntese de vanilina nos frutos de *V. planifolia* proposta por Gallage et al. (Gallage et al. 2014).

Soluções de extração de proteínas

A necessidade de entender melhor a produção de composto do aroma e sabor dos frutos de *Vanilla spp.* é possível através do estudo das suas vias bissintéticas (Palama 2014). O perfil proteômico de uma espécie identifica os componentes reguladores que medeiam as diversas vias de biossíntese sendo assim, as proteínas podem servir como marcadores para melhorar a qualidade nutricional, sabor, resistência/tolerância à doença e a vida útil das moléculas de interesse (Kilambi et al. 2016).

Contudo, estudos do perfil proteômico tem um grande desafio, a variada amplitude dinâmica das proteínas que constituem o complexo proteico (Kilambi et al. 2016). Embora vários protocolos de extração de proteínas de tecidos de frutos estejam disponíveis como: tomate (Kilambi et al. 2016), pimenta (Choi and Hwang 2011), morango (Bianco et al. 2009), uva (Negri et al. 2015), banana (Toledo et al. 2012), maçã e pêra (Kiemer and Cesareni 2007), a maioria deles não está relacionado à extração tipo *shotgun*. Atualmente, esse método é o mais indicado para análise de perfis proteômicos apresentando grande amplitude dinâmica de proteínas (Garrido et al. 2016). A extração de proteínas tão diversas em concentrações também variadas, presentes nos frutos ou em outros tecidos vegetais,

apresenta substâncias interferentes como os pigmentos, carboidratos, polifenóis, polissacarídeos e amido, que podem levar a desnaturação, inativação de proteínas e atrapalhar a extração das mesmas (Song and Braun 2008).

Entre os tampões de extração de proteína mais comuns estão: os reguladores de pH (por exemplo, o Tris), agentes redutores (por exemplo, ditiotreitol - DDT, β-mercaptoetanol) e os desnaturantes (por exemplo, uréia, dodecil sulfato de sódio, CHAPS) (Song and Braun 2008). No presente estudo, foi sugerido o uso de um agente regulador de pH (Tris-HCl), um agente redutor (β-mercaptoetanol) e um detergente iônico (dodecil sulfato de sódio – SDS) com o intuito de aumentar a capacidade extratora das proteínas do fruto de *V. bahiana*. De acordo com Wilson & Walker (2010), o β-mercaptoetanol reduz as pontes de dissulfeto, que mantêm a estrutura terciária das proteínas, e o SDS se liga fortemente as proteínas auxiliando na solubilização das mesmas.

Protocolos de extração de proteínas que utilizam soluções de fenol também foram relatados como adequados para a extração de baixas concentrações de proteínas em frutas (Vincent, Wheatley, and Cramer 2006). Contudo, protocolos de extração de proteínas de membrana baseados no uso de fenol também precisam dos detergentes (como o SDS) para otimizar a extração de proteínas (Sun, Wang, and Li 2012; Lin et al. 2012; Hurkman and Tanaka 1986; Botelho et al. 2010; Wu and Wang 1984). O uso de SDS em análise de espectrometria de massas pode ser problemático, pois sua presença nas amostras pode ocasionar supressão de íons, assim reduzindo significativamente o número proteínas identificadas. Porém alguns estudos apontam, que após a retirada do SDS das amostras, esse desnaturante pode auxiliar e melhorar a extração de proteínas (Hurkman and Tanaka 1986; Sun, Wang, and Li 2012; Liu et al. 2012; Lin et al. 2012; Botelho et al. 2010; Song and Braun 2008).

Método HDMS^E

Para as análises proteômicas de diferentes condições de extração realizadas no presente estudo, e para alcançar a confiabilidade e reprodutibilidade dos resultados, utilizamos uma abordagem 2D nano-UPLC-HDMS^E (de alta definição), livre de marcadores com Aquisição de Dados Independentes (ADI). Os formatos multiplex de alta resolução MS^E e HDMS^E são métodos recomendados para proteômica de amostras complexas. No método HDMS^E os íons são separados com base na onda de voltagem da mobilidade iônica, e são submetidos à fragmentação, onde serão analisados os íons precursores e seus íons filhos,

garantindo maior confiabilidade nas identificações (Martins Ferreira, Guest, and Martins-de-souza 2017; Bond et al. 2013).

Objetivo Geral

Segundo a nossa hipótese de que *V. bahiana* pode apresentar potencial para produzir vanilina e outros fenóis relacionados ao aroma e sabor da baunilha, o objetivo desse trabalho é caracterizar a expressão proteica dos frutos maduros de *Vanilla bahiana* da Mata Atlântica do Rio de Janeiro.

Objetivos específicos

Avaliar qual solução é capaz de extrair: A) maior conteúdo de proteínas totais; B) maior diversidade de proteínas; C) maior número de proteínas ligadas a biossíntese de vanilina e outros fenóis encontrados no extrato natural de baunilha. Identificar, de acordo com a literatura, quais proteínas estão relacionadas à floração, amadurecimento do fruto, aroma e sabor de baunilha. E, finalmente, validar os dados proteômicos quantificando três compostos de interesse: vanilina, ácido *p*-coumárico e pirogalol.

Referências Bibliográficas

Anjos, A.M., F.F.V.A. Barbarena, and C.M. Pigozzo. 2016. "Biologia Reprodutiva de Vanilla Bahiana Hoehne (Orchidaceae)." *Orquidário* 30 (3–4):67–79.

Anuradha, Krushnamurthy, Bellur Nanjundaiah Shyamala, and Madeneni Madhava Naidu. 2013. "Vanilla- Its Science of Cultivation, Curing, Chemistry, and Nutraceutical Properties." *Critical Reviews in Food Science and Nutrition* 53 (12):1250–76. https://doi.org/10.1080/10408398.2011.563879.

Baqueiro-Peña, Itzamná, and José Ángel Guerrero-Beltrán. 2016. "Vanilla (Vanilla Planifolia Andr.), Its Residues and Other Industrial by-Products for Recovering High Value Flavor Molecules: A Review." *Journal of Applied Research on Medicinal and Aromatic Plants*. https://doi.org/10.1016/j.jarmap.2016.10.003.

Bianco, Linda, Loredana Lopez, Anna Grazia Scalone, Mariasole Di Carli, Angiola Desiderio, Eugenio Benvenuto, and Gaetano Perrotta. 2009. "Strawberry Proteome Characterization and Its Regulation during Fruit Ripening and in Different Genotypes." *Journal of Proteomics* 72 (4). Elsevier B.V.:586–607.

https://doi.org/10.1016/j.jprot.2008.11.019.

Botelho, Diane, Mark J. Wall, Douglas B. Vieira, Shayla Fitzsimmons, Fang Liu, and Alan Doucette. 2010. “Top-down and Bottom-up Proteomics of Sds-Containing Solutions Following Mass-Based Separation.” *Journal of Proteome Research* 9 (6):2863–70. https://doi.org/10.1021/pr900949p.

Bond, Nicholas J, Pavel V Shliaha, Kathryn S Lilley, and Laurent Gatto. 2013. “Improving Qualitative and Quantitative Performance for MS E -Based Label-Free Proteomics.” *Journal of Proteome Research* 12 (6):2340–53. https://doi.org/10.1021/pr300776t.

Bouetard, Anthony, Pierre Lefeuvre, Rodolphe Gigant, Séverine Bory, Marc Pignal, Pascale Besse, and Michel Grisoni. 2010. “Evidence of Transoceanic Dispersion of the Genus Vanilla Based on Plastid DNA Phylogenetic Analysis.” *Molecular Phylogenetics and Evolution* 55 (2). Elsevier Inc.:621–30. https://doi.org/10.1016/j.ympev.2010.01.021.

Busconi, Matteo, Luigi Lucini, Giovanna Soffritti, Jamila Bernardi, Letizia Bernardo, Christel Brunschwig, Sandra Lepers-Andrzejewski, Phila Raharivelomanana, and Jose A. Fernandez. 2017. “Phenolic Profiling for Traceability of Vanilla ×tahitensis.” *Frontiers in Plant Science* 8 (October):1–13. https://doi.org/10.3389/fpls.2017.01746.

Chase, Mark W., Kenneth M. Cameron, John V. Freudenstein, Alec M. Pridgeon, Gerado Salazar, Cássio van den Berg, and André Schuiteman. 2015. “An Updated Classification of Orchidaceae.” *Botanical Journal of the Linnean Society* 177 (2):151–74. https://doi.org/10.1111/boj.12234.

Chee, Marcus Jenn Yang, Grantley W. Lycett, Teng Jin Khoo, and Chiew Foan Chin. 2017. “Bioengineering of the Plant Culture of Capsicum Frutescens with Vanillin Synthase Gene for the Production of Vanillin.” *Molecular Biotechnology* 59 (1). Springer US:1–8. https://doi.org/10.1007/s12033-016-9986-2.

Choi, Du Seok, and Byung Kook Hwang. 2011. “Proteomics and Functional Analyses of Pepper Abscisic Acid-Responsive 1 (ABR1), Which Is Involved in Cell Death and Defense Signaling.” *The Plant Cell* 23 (2):823–42. https://doi.org/10.1105/tpc.110.082081.

Divakaran, Minoo, K. Nirmal Babu, and K. V. Peter. 2006. “Conservation of Vanilla Species, in Vitro.” *Scientia Horticulturae* 110 (2):175–80. https://doi.org/10.1016/j.scienta.2006.07.003.

Divakaran, Minoo, K Nirmal Babu, P N Ravindran, K V Peter, Asian J Plant, and Sci Res.

2015. “Biotechnology for Micropropagation and Enhancing Variations in Vanilla” 5 (2):52–62.

Fraga, Claudio Nicoletti de, Dayvid Rodrigues Couto, and Emerson Ricardo Pansarin. 2017. “Two New Species of Vanilla (Orchidaceae) in the Brazilian Atlantic Forest.” *Phytotaxa* 296 (1):063–072. https://doi.org/10.11646/phytotaxa.296.1.4.

Gallage, Nethaji J., and Birger Lindberg Møller. 2015. “Vanillin–Bioconversion and Bioengineering of the Most Popular Plant Flavor and Its De Novo Biosynthesis in the Vanilla Orchid.” *Molecular Plant* 8 (1). Elsevier Ltd:40–57. https://doi.org/10.1016/j.molp.2014.11.008.

Gallage, Nethaji J., and Birger Lindberg Møller. 2018. *Vanilla: The Most Popular Flavour*. https://doi.org/10.1007/978-3-319-67903-7_1.

Gallage, Nethaji J, Esben H Hansen, Rubini Kannangara, Carl Erik Olsen, Mohammed Saddik Motawia, Kirsten Jørgensen, Inger Holme, Kim Hebelstrup, Michel Grisoni, and Birger Lindberg Møller. 2014. “Vanillin Formation from Ferulic Acid in Vanilla Planifolia Is Catalysed by a Single Enzyme.” *Nature Communications* 5 (May):4037. https://doi.org/10.1038/ncomms5037.

Garrido, Bruno Carius, Gustavo H.M.F. Souza, Daniela C. Lourenço, and Maíra Fasciotti. 2016. “Proteomics in Quality Control: Whey Protein-Based Supplements.” *Journal of Proteomics* 147. Elsevier B.V.:48–55. https://doi.org/10.1016/j.jprot.2016.03.044.

Gigant, Rodolphe, Severine Bory, Michel Grisoni, and Pascale Besse. 2011. “Biodiversity and Evolution in the Vanilla Genus.” *The Dynamical Processes of Biodiversity - Case Studies of Evolution and Spatial Distribution*, no. figure 1:1–27. https://doi.org/10.5772/24567.

Greule, Markus, Armin Mosandl, John T. G. Hamilton, and Frank Keppler. 2015. “Comment on Authenticity and Traceability of *Vanilla* Flavors by Analysis of Stable Isotopes of Carbon and Hydrogen.” *Journal of Agricultural and Food Chemistry* 63 (21):5305–6. https://doi.org/10.1021/jf506172q.

Gu, Fenglin, Yonggan Chen, Yinghua Hong, Yiming Fang, and Lehe Tan. 2017. “Comparative Metabolomics in Vanilla Pod and Vanilla Bean Revealing the Biosynthesis of Vanillin during the Curing Process of Vanilla.” *AMB Express* 7 (1). Springer Berlin Heidelberg:116. https://doi.org/10.1186/s13568-017-0413-2.

Guerrero, A., S.E Valdés-Rodríguez, B Durán-Sánchez, M.T González-Arnao, B Jiménez-Francisco, and C.E Lázaro-Vallejo. 2011. “Cryopreservation and Proteomic Analysis

of Vanilla (V . Planifolia A .) Apices Treated with Osmoprotectants." *Acta Horticulturae*, no. 908:67–72. https://doi.org/10.17660/ActaHortic.2011.908.5.

Hocking, Martin B. 1997. "Vanillin: Synthetic Flavoring from Spent Sulfite Liquor." *Journal of Chemical Education* 74 (9):1055. https://doi.org/10.1021/ed074p1055.

Hrazdina, G. 2006. "Aroma Production by Tissue Cultures Aroma Production by Tissue Cultures." *Journal of Agricultural and Food Chemistry* 54 (February):1116–23. https://doi.org/10.1021/jf053146w.

Hurkman, William J, and Charlene K Tanaka. 1986. "Solubilization of Plant Membrane Proteins for Analysis by Two-Dimensional Gel Electrophoresis." *Plant Physiology* 81 (3):802–6. https://doi.org/10.1104/pp.81.3.802.

Kahane, R, P Besse, M Grisoni, F Le Bellec, and E Odoux. 2008. "Bourbon Vanilla: Natural Flavour with a Future." *Chronica Horticulturae* 48 (2):23–29.

Kiemer, Lars, and Gianni Cesareni. 2007. "Comparative Interactomics: Comparing Apples and Pears?" *Trends in Biotechnology* 25 (10):448–54. https://doi.org/10.1016/j.tibtech.2007.08.002.

Kilambi, Himabindu V., Kalyani Manda, Hemalatha Sanivarapu, Vineet K. Maurya, Rameshwar Sharma, and Yellamaraju Sreelakshmi. 2016. "Shotgun Proteomics of Tomato Fruits: Evaluation, Optimization and Validation of Sample Preparation Methods and Mass Spectrometric Parameters." *Frontiers in Plant Science* 7 (June):1–14. https://doi.org/10.3389/fpls.2016.00969.

Kundu, Anish. 2017. "Vanillin Biosynthetic Pathways in Plants." *Planta* 245 (6). Springer Berlin Heidelberg:1069–78. https://doi.org/10.1007/s00425-017-2684-x.

Lin, Yong, Huajun Jiang, Yujun Yan, Bin Peng, Jinhua Chen, Haiyan Lin, and Zhonghua Liu. 2012. "Shotgun Analysis of Membrane Proteomes by an Improved SDS-Assisted Sample Preparation Method Coupled with Liquid Chromatography-Tandem Mass Spectrometry." *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* 911. Elsevier B.V.:6–14. https://doi.org/10.1016/j.jchromb.2012.10.016.

Liu, Yi, Yong Lin, Yizhong Yan, Jianjun Li, Quanze He, Ping Chen, Xianchun Wang, and Songping Liang. 2012. "Electrophoretically Driven SDS Removal and Protein Fractionation in the Shotgun Analysis of Membrane Proteomes." *Electrophoresis* 33 (2):316–24. https://doi.org/10.1002/elps.201100364.

Medina, Javier De La Cruz, Guadalupe C. Rodriguez Jiménes, and Hugo S. García. 2009.

"VANILLA- Post-Harvest Operations." *Food and Agriculture Organization of the United Nations*.

Martins Ferreira, Henrique Gustavo, Paul C Guest, and Daniel Martins-de-souza. 2017. "Multiplex Biomarker Techniques." In *Methods in Molecular Biology*, 1546:57–73. https://doi.org/10.1007/978-1-4939-6730-8.

Moreira, Marina Muniz, FFVA Barberena, and RC Lopes. 2014. "Orchidaceae of the Grumari Restinga: Floristic and Similarity among Restingas in Rio de Janeiro State, Brazil." *Acta Botanica Brasilica* 28 (3):321–26. https://doi.org/10.1590/0102-33062014abb3173.

Negri, Alfredo S., Bhakti Prinsi, Osvaldo Failla, Attilio Scienza, and Luca Espen. 2015. "Proteomic and Metabolic Traits of Grape Exocarp to Explain Different Anthocyanin Concentrations of the Cultivars." *Frontiers in Plant Science* 6 (August). https://doi.org/10.3389/fpls.2015.00603.

Ormerod, Paul, and Jim Cootes. 2013. "Leafy Vanilla Species of the Philippines." *OrchideenJournal* 1:1–19.

Palama, Tony L. 2014. "Metabolomic Analysis of a Plant from the Indian Ocean : Vanilla Planifolia." In *Novel Plant Bioresources: Applications in Food, Medicine and Cosmetics*, 471–78.

Palama, Tony L, Patrice Menard, Isabelle Fock, Young H Choi, Emmanuel Bourdon, Joyce Govinden-Soulange, Muriel Bahut, Bertrand Payet, Robert Verpoorte, and Hippolyte Kodja. 2010. "Shoot Differentiation from Protocorm Callus Cultures of Vanilla Planifolia (Orchidaceae): Proteomic and Metabolic Responses at Early Stage." *BMC Plant Biology* 10:82. https://doi.org/10.1186/1471-2229-10-82.

Pansarin, Emerson, João Aguiar, and Alessandro Ferreira. 2012. "A New Species of Vanilla (Orchidaceae : Vanilloideae) from São." *The New York Botanical Garden Press* 64 (June):157–61.

Parthasarathy, V . A., B. Chempakam, and T. J. Zachariah, eds. 2008. *Chemistry of Spices*. IV. Massachusetts: CABI. https://doi.org/10.1079/9781845934057.0000.

Pérez-Silva, a., E. Odoux, P. Brat, F. Ribeyre, G. Rodriguez-Jimenes, V. Robles-Olvera, M. a. García-Alvarado, and Z. Günata. 2006. "GC-MS and GC-Olfactometry Analysis of Aroma Compounds in a Representative Organic Aroma Extract from Cured Vanilla (Vanilla Planifolia G. Jackson) Beans." *Food Chemistry* 99 (4):728–35. https://doi.org/10.1016/j.foodchem.2005.08.050.

Pinaria, A. G., E. C.Y. Liew, and L. W. Burgess. 2010. “Fusarium Species Associated with Vanilla Stem Rot in Indonesia.” *Australasian Plant Pathology* 39 (2):176–83. https://doi.org/10.1071/AP09079.

Rain, P, and Philip Lief Group. 2004. *Vanilla: The Cultural History of the World’s Most Popular Flavor and Fragrance*. Jeremy P. Tarcher/Penguin.

Rubert, Josep, Ondrej Lacina, Milena Zachariasova, and Jana Hajslova. 2016. “Saffron Authentication Based on Liquid Chromatography High Resolution Tandem Mass Spectrometry and Multivariate Data Analysis.” *Food Chemistry* 204 (2015). Elsevier Ltd:201–9. https://doi.org/10.1016/j.foodchem.2016.01.003.

Sambin, A, and Guy R Chiron. 2015. “Deux Nouvelles Espèces de Vanilla (Orchidaceae) de Guyane Française.” *Richardiana* 15:306–16.

Sharma, Upendra Kumar, Nandini Sharma, Ajai Prakash Gupta, Vinod Kumar, and Arun Kumar Sinha. 2007. “RP-HPTLC Densitometric Determination and Validation of Vanillin and Related Phenolic Compounds in Accelerated Solvent Extract of Vanilla Planifolia.” *Journal of Separation Science* 30 (18):3174–80. https://doi.org/10.1002/jssc.200700229.

Song, Jun, and Gordon Braun. 2008. “Application of Proteomic Techniques to Fruits and Vegetables.” *Current Proteomics* 5 (3):191–201. https://doi.org/10.2174/157016408785909659.

Strong, Andrea. 2017. “Pastry Chefs Forced to Get Creative as Vanilla Prices Soar.” Eater Business Reports. 2017.

Sun, Difei, Nan Wang, and Liang Li. 2012. “Integrated SDS Removal and Peptide Separation by Strong-Cation Exchange Liquid Chromatography for SDS-Assisted Shotgun Proteome Analysis.” *Journal of Proteome Research* 11 (2):818–28. https://doi.org/10.1021/pr200676v.

Tan, B. C., C. F. Chin, S. Liddell, and P. Alderson. 2014. “Protein Extraction for Callus and Node Cultures of Vanilla Planifolia Andrews.” *Minerva Biotecnologica* 26 (July):115–26.

Tan, Boon Chin, Chiew Foan Chin, Susan Liddell, and Peter Alderson. 2013. “Proteomic Analysis of Callus Development in Vanilla Planifolia Andrews.” *Plant Molecular Biology Reporter* 31 (6):1220–29. https://doi.org/10.1007/s11105-013-0590-3.

Toledo, Tatiana Torres, Silvia Beserra Nogueira, Beatriz Rosana Cordenunsi, Fábio César Gozzo, Eduardo Jorge Pilau, Franco Maria Lajolo, and João Roberto Oliveira Do

Nascimento. 2012. “Proteomic Analysis of Banana Fruit Reveals Proteins That Are Differentially Accumulated during Ripening.” *Postharvest Biology and Technology* 70. Elsevier B.V.:51–58. https://doi.org/10.1016/j.postharvbio.2012.04.005.

Villanueva-Viramontes, Sara, Mariana Hernández-Apolinar, Germán Carnevali Fernández-Concha, Alfredo Dorantes-Euán, Gabriel R. Dzib, and Jaime Martínez-Castillo. 2017. “Wild Vanilla Planifolia and Its Relatives in the Mexican Yucatan Peninsula: Systematic Analyses with ISSR and ITS.” *Botanical Sciences* 95 (2):169–87. https://doi.org/10.17129/botsci.668.

Villanueva-Viramontes, Sara, Mariana Hernández-Apolinar, Germán Carnevali Fernández-Concha, Alfredo Dorantes-Euán, Gabriel R Dzib, and Jaime Martínez Castillo. 2017. “Wild Vanilla Planifolia and Its Relatives in the Mexican Yucatan Peninsula: Systematic Analyses with ISSR and ITS.” *Botanical Sciences* 95 (2):169–87.

Vincent, Delphine, Matthew D. Wheatley, and Grant R. Cramer. 2006. “Optimization of Protein Extraction and Solubilization for Mature Grape Berry Clusters.” *Electrophoresis* 27 (9):1853–65. https://doi.org/10.1002/elps.200500698.

Virol, Arch, Christopher Puli, Subuhi Khan, and Wee-leong Chang Gardette. 2016. “First Complete Genome Sequence of Vanilla Mosaic Strain of Dasheen Mosaic Virus Isolated from the Cook Islands .” *Archives of Virology*. Springer Vienna. https://doi.org/10.1007/s00705-016-3133-z.

Walton, Nicholas J., Melinda J. Mayer, and Arjan Narbad. 2003. “Vanillin.” *Phytochemistry* 63 (5):505–15. https://doi.org/10.1016/S0031-9422(03)00149-3.

Wilson, Keith, and John Walker. 2010. *Principles and Techniques of Biochemistry and Molecular Biology*. https://doi.org/10.1093/nar/gkl830.

Wu, Fang-Sheng, and Menq-Yun Wang. 1984. “Extraction of Proteins for Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis from Protease-Rich Plant Tissues.” *Analytical Biochemistry* 139 (1):100–103. https://doi.org/10.1016/0003-2697(84)90394-4.

Capítulo 1

Manuscrito a ser submetido à revista Journal of Agricultural and Food Chemistry (ACS Publications)

Vanilla bahiana, an alternative source of the Atlantic Forest for the production of vanilla: a proteomic approach through high definition nano-LC-MS

[1]Ellen M. Lopes, [1]Roberta G. Linhares, [2]Lucas de Oliveira Pires, [2]Rosane Nora Castro, [3,4]Maria Gabriela B. Koblitz, [4]L. C. Cameron, [1,4] Andrea F. Macedo*.

[1]Integrated Laboratory of Plant Biology, Department of Botany, Institute of Biosciences, Federal University of Rio de Janeiro State, UNIRIO. Av. Pasteur, 458, Urca, 22290-240 Rio de Janeiro, Brazil.

[2]Universidade Federal Rural do Rio de Janeiro – ICE - Departamento de Química, BR 467 – Km 47, Seropédica, RJ, Brazil

[3]Food and Nutrition Graduate Program, Nutritional Biochemistry Center, Federal University of Rio de Janeiro State, UNIRIO. Av. Pasteur, 296, Urca, 22290-250 Rio de Janeiro, Brazil.

[4]Laboratory of Protein Biochemistry - Center of Innovation of Mass Spectrometry, Federal University of State of Rio de Janeiro, UNIRIO. Av. Pasteur, 296, Urca, 22290-250 Rio de Janeiro, Brazil.

*Corresponding author Tel. +55(21) 2530-2551; +55(21) 99198-5671; Email: andrea.macedo@unirio.br; andreafm22@yahoo.com.br

Abstract

Vanilla spp. presented few cultivated species applied to vanilla production, despite the great demand, predatory exploitation and poor genetic variability threatening the natural vanilla production. *Vanilla bahiana* pods from the Atlantic Forest may be an alternative source for natural vanilla. This study applied bottom-up and shotgun proteomics analysis to identify proteins related to flowering, fruitening and vanilla flavor production processes. Extraction solutions combining Tris-HCl buffer, β-mercaptoethanol and SDS were assayed. Progenesis QI software loaded with an *Orchidaceae* database was able to identify 2326 proteins in our samples. Amongst them, 75 were highlighted as relevant to the synthesis of compounds related to vanilla flavor compounds, such as vanillin synthase, which was successfully extracted with 1% of SDS, a condition that also improved the diversity of extracted protein. The proteins identified in *V. bahiana*'s pods point to the enzymatic potential of this species, validated by the quantification of vanillin in the samples.

Keywords: Bottom up proteomics; nano-LC-HDMS^E; vanillin; sodium dodecyl sulfate; β-mercaptoethanol*; vanillin synthase; Vanilla bahiana*

Introduction

Vanilla spp. (*Orchidaceae*) are monophyletic species with the highest concentrated diversity in tropical regions, mainly in the Brazilian biomes[1]. Many species of this genus are considered rare or endangered due to deforestation of their subcosmopolitan habitat, climate change, predatory exploitation and pandemic pathogenic agents[2]. This is an economically relevant genus due to the synthesis of vanillin, a phenolic compound present in the cured pods[3]. This compound is considered to be amongst the most valuable cultivated spices in the world[4], with further applications in health industry due to its antioxidant, anticarcinogenic, antimutagenic activities[4].

The global production of *Vanilla* spp. weights around 5600 tons and the natural vanillin extract costs around U$1,200 - U$4,000/kg. Today, natural vanilla production is expensive, time-consuming and depends on clonal cultivation of two species: *Vanilla planifoli*a G. Jackson (syn. *Vanilla fragrans* Andrews) and *Vanilla tahitensis* Moore, with the former providing 95% of world's production of natural extract[2,5]. Despite the great economic relevance, clonal production of *V. planifolia* has depleted the genetic variability and left the species vulnerable to diseases, which jeopardizes the world's supply[6]. The cheap synthetic vanillin compound has been widely commercialized since the 19th century, however, the demand for the natural extract has been increasing. This is due to the consumers demand for products free of artificial additives, in addition to the natural product providing superior flavor quality, as a result of a mixture of several different compounds[7].

The difficult production, the susceptibility of species, the high demand and increasing prices created a worldwide effort to search for new *Vanilla* species as alternative sources for the purpose of crop and flavor improvement, aiming to enhance the production of active ingredients and to amplify the genetic resources[8]. In the Brazilian Atlantic Forest, an endemic species is still scientifically and economically unexplored, the *Vanilla bahiana* Hoehne[9] and the hypothesis of this study was that this species may present potential to produce vanillin and other compounds related to vanilla flavor. To test this hypothesis, we applied a untargeted proteomics' strategy based on nanoUPLC and High Definition

Mass Spectrometry (HDMS^E) label-free quantitative approach was applied to characterize the protein expression of *Vanilla bahiana* from the Atlantic Forest of Rio de Janeiro, focusing on the proteins related to vanilla flavor production, flowering and fruit ripening.

Material and Methods

Plant material. Three mature *V. bahiana* pods, from three different individuals, were harvested from the Natural Monument of Pão de Açúcar and Urca, Rio de Janeiro city (RJ, Brazil). The plant material was identified by the Botanic Garden of Rio de Janeiro (catalog number: RB 646438, bar code number: 01111538, http://jabot.jbrj.gov.br/v2/consulta.php#).

Chemicals. All reagents were purchased from Sigma-Aldrich (St Louis, MO, USA) and were of LC-MS grade, unless otherwise stated. Ultrapure water (resistivity > 18.2 MΩ·cm) (Barnstead™ Smart2Pure™, Thermo Fisher Scientific, MA, USA) was also used.

Protein extraction and digestion. The pods were freeze-dried, powdered and pooled (**Figure S1**). Two hundred milligrams of this material were further pulverized in liquid nitrogen and extracted with 3 mL of different protein extraction solutions. Six different extraction solutions were assayed: 1 – TRIS-hydrochloride (HCl) buffer (pH 6.8; 125 nmol/L) (Vb1); 2 – solution 1 + 0.5% (v/v) β-mercaptoethanol (β-MT) (10 mmol/L) (Vb2); 3 – solution 2 + 1% of sodium dodecyl sulfate (SDS) (Vb3); 4 – solution 2 + 0.1% of SDS (Vb4); 5 – solution 1 + 1% of SDS (Vb5) and 6 – solution 1 + 0.1% of SDS (Vb6) (**Figure S1**). The samples were homogenized using an ultrasonic probe in 250W/10 min (Eco-sonic,Ultronique) and centrifuged (Thermo scientific, Sorvall Legend X1R) at 14,000x*g*, for 60 min, at 4 °C. The supernatants were transferred to an Amicon Ultra 3 KDa (Merck Millipore, Billerica, MA) in which the samples were washed and concentrated. Three washing steps were performed using 2 mL of ammonium bicarbonate buffer (pH 8.5, 50 mM). The protein concentration in each extracted sample was determined by the Bradford method[10] using a flat plate reader (FlexStation 3; Molecular Devices, CA). Each sample was diluted to a final concentration of 1 μg.μL^{-1} of protein. Protein digestion was performed according to Lobo et al.[11].

NanoUPLC and Label-Free Data-Independent Mass spectrometric analysis - HDMS^E analysis. The nanoUPLC-RP SYNAPT G2-S HDMS mass spectrometer instrument (Waters, Manchester, UK) was used with an orthogonal effective resolution from the ion mobility separation (nanoESI-Qq-*oa*TOF). The analytical reversed-phase column nanoACQUITY HSS T3 1.8-µm (100µm x 100 mm) (Waters, Manchester, UK) was used to separate the tryptic peptides. Prior to proteomic analysis, stoichiometric measurements based on scouting runs of the integrated total ion account (TIC), were conducted for each sample, to ensure standardized molar values across all conditions and normalize the injections on the column.

For the HDMS^E acquisitions, specific volumes were injected for each technical injection sample at a rate of 350 nL min^{-1}/ 90 min. All the parameters were set according to Victorio et al.[12] study. The separation gradient of phase A (ultra-pure water with 0.1 % formic acid) and B (acetonitrile with 0.1% formic acid) was performed according to following schedule: 0 min - 97 % A, 1.0 min - 86.9 % A, 5.0 min - 97 % A (flux 2000 µL min^{-1}).

The mass spectrometer was operated as described in Victorio et al.[12], in resolution mode (35,000 FWHM), using a nano-electrospray ionization in the positive ion mode and a NanoLockSpray (Waters, Manchester, UK) ionization source. Ions with mass between 50 and 2.000 Da were acquired within 0.5 second scanning time, with a transfer ramp and collision energy of 19 V to 55 V.

Proteomic data analysis. The Progenesis QI for Proteomics (v 2.1, Nonlinear Dynamics; Waters, Newcastle, UK) software was used to process and search proteins, using a label-free methodology, with an *Orchidaceae* database proteins, noted in UniProt online program (UniProt - http://www.uniprot.org) in FASTA form. The Uniprot *Orchidaceae* database released in 2017_03, with 26,899 reviewed entries was used. The parameters for database searching were: one missed cleavage, minimum fragment ion/ peptide equal to one, minimum fragment ion/ protein equal to three, minimum peptide/ protein equal to one, fixed modifications of carbamidomethyl C, variable modifications of oxidation M and phosphoryl STY, and a default false discovery rate (FDR) maximum

of 4%. The parameters set as default were peptide mass error of tolerance greater than 10 ppm, maximum protein mass of 600 KDa, fragment mass error tolerance of 20 ppm and score less than 4. Relative quantification was determined from the absolute intensities with the use of ion accounting Hi3 based quantification methods. The homologies were processed and homology filters quantification were performed by the software according to Li et al.[13].

Proteomics data processing and statistical analysis. The multiple sample test (ANOVA) with Tukey test ($p \leq 0.05$) was used in Bradford quantification data and protein relative quantitative analyses. Proteins that were present in, at least, two of the three technical replicates, and with a coefficient of variation (CV) ≤ 0.4 were considered as successfully identified. The cellular component and biological process annotation were performed using UniProtKB taxonomic.

Vanillin, pyrogallol and p-coumaric acid extraction and absolute quantification. One hundred milligrams of dry fruit were extracted with 15 mL of methanol (MeOH) in ultrasonic cleaner bath for 30 min, and evaporated. The samples were resuspended with 4 mL of MeOH. The experiments were performed in duplicate. The extracts were passed through a 25 mm PTFE 0.45 µL syringe filter prior to injection (20 µL) into the HPLC system, in technical triplicate.

The samples were analyzed on an HPLC-DAD Shimadzu Prominence system fitted with a LC-20AT pump module, a SPD- M20A diode array detector, a CBM-20A control center, a SIL-10A Autosampler injector, and a CTO-20A oven (Shimadzu, Japan). Data analysis were performed using LC Solution software (Shimadzu, Japan). The column used was a Betasil RP – 18 (250 x 4.6 mm, 5 µm, Thermo Fisher Scientific, Runcorn, UK), and the mobile phase was water–acetic acid (99: 1, solvent A) and B (acetonitrile). Elution was performed at a flow rate of 1 ml min^{-1} using a gradient starting with 35% B increasing to levels of 80% B at 7 min and 35% B at 11 min to re-equilibrate the column. The compounds were monitored at: 270 nm for vanillin and pyrogallol and 310nm for *p*-coumaric acid. The identification of the compounds was performed through the comparison of retention

time and ultraviolet spectrum between the samples and the controls (standard), and quantification was performed using external calibration data for the same compounds.

Results and Discussion

Selection of the extraction solution. Proteomics remains a challenge in recalcitrant plant tissues, like in fruits, and the stages of protein extraction and sample preparation are critical to eliminate interfering substances and to achieve high protein content. Protein protocols applying phenol solutions have been reported to be suitable for the extraction of low concentrations of protein in fruits[14–16]. Some of the limitations of phenol-based protocols are the lack of detergents in the extraction buffer, the need of many steps with too much sample manipulation and the high toxicicity[15]. On the other hand, the SDS containing procedures have been reported to be a viable alternative when a secondary clean-up step is applied[17]. In this study, six different non-phenol-based extraction conditions were tested (**Figure S1**).

According to our results, samples Vb5 (1% SDS) and Vb3 (β-MT + 1% SDS) contained significantly more total protein extracted and higher protein diversity than the other samples (**Figure 1A, B**), thus the addition of β-MT and 1% of SDS in the extraction solution was successful to obtain better protein solubilization when compared to other extraction solutions (**Figure 1A**). However, the presence of β-MT interfered with the SDS, resulting in the loss of proteins (**Figure 1A, B**). On the other hand, no SDS signal was found in the mass spectrum, meaning that the washing procedure with ammonium bicarbonate removed the detergent, as suggested by Garrido[18]. Samples Vb4 and Vb6 (0.1% of SDS) were able to extract only little concentrations of proteins (**Figure 1A**) and were not used in further evaluations of the extracted protein diversity (**Figure 1B**).

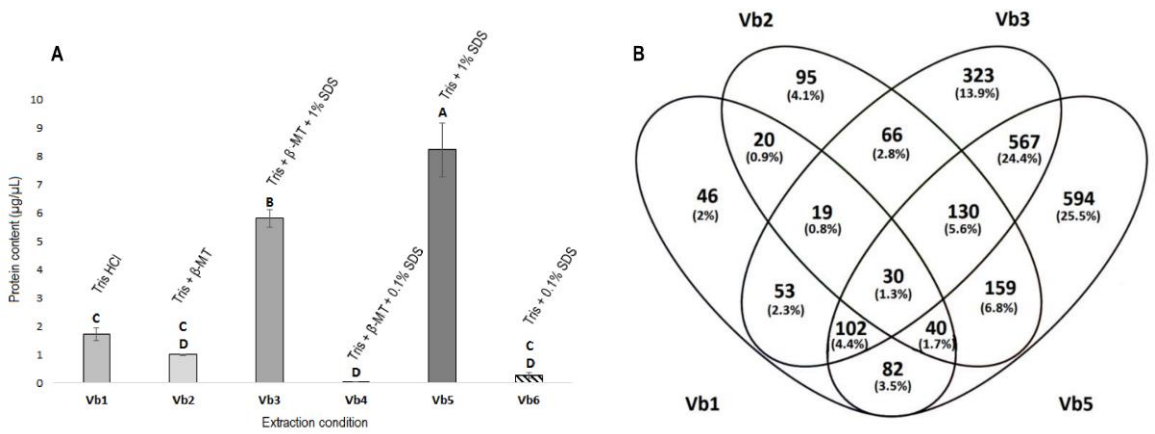


Figure 1. A – Total protein in µg µL^{-1} (mean ± SD). The data were analyzed using Statistics software v 7.0. The ANOVA with $p < 0.05$ was used, followed by the Tukey test. The same letters mean no statistical significance. B- Venn diagram compares the number of unique and common proteins identified in the dataset between extraction solutions. It shows the overlaps between proteins identified in each of the extraction condition.

The use of chemical reducers and detergents usually causes the opening of the protein's structure facilitating the protein elution and leading to greater amount and variety of proteins in solution[19,20]. However, according to Anand et al.[21], the presence of a reducing agent may interfere with the anionic surfactant. Although SDS is known to cause a better solubilization of proteins improving phase separations and protein recovery, it was also reported as a problem for ESI-based analytical methods, as its presence in the samples can lead to ion suppression and significantly reduce the number of peptides identifications[18,22]. Usually, methods applying SDS depend on different tactics to remove the detergent prior to ionization. The present strategy seemed to be less time consuming and less damaging to proteins than the most common methods (precipitation with organic solvents, column based approaches, dialysis)[23].

A total of 2326 proteins were collectively identified (**Table S1**) and the presence of SDS allowed the extraction of more diverse proteins, with 1704 identified in Vb5 and 1290 in Vb3 (**Figure 1B**). It was also possible to observe a similarity between Vb5 and Vb3 (with 829 proteins in common) both

conditions with 1% SDS (**Figure 1B**). Regarding biological process, most of the proteins were related to the biosynthetic process of organic substance. Vb5 was the extraction condition with highest biological process diversity. The Vb1 condition extracted proteins with less biological process diversity (**Figure 2**). Usually, SDS shows the ability to extract hydrophobic membrane proteins not readily soluble in water, resulting in a more complete protein sampling[24,25]. However, according to the cellular localization, a similar profile was observed among all conditions: most proteins extracted were located in the plastids, followed by proteins located in the intrinsic component of the membrane. In this case the use of SDS did not significantly improve the extraction of membrane proteins (**Figure S2)**. According to Gallage et al.[26] the vanillin synthase location, an important enzyme for vanillin biosynthesis, was restricted to plastids.

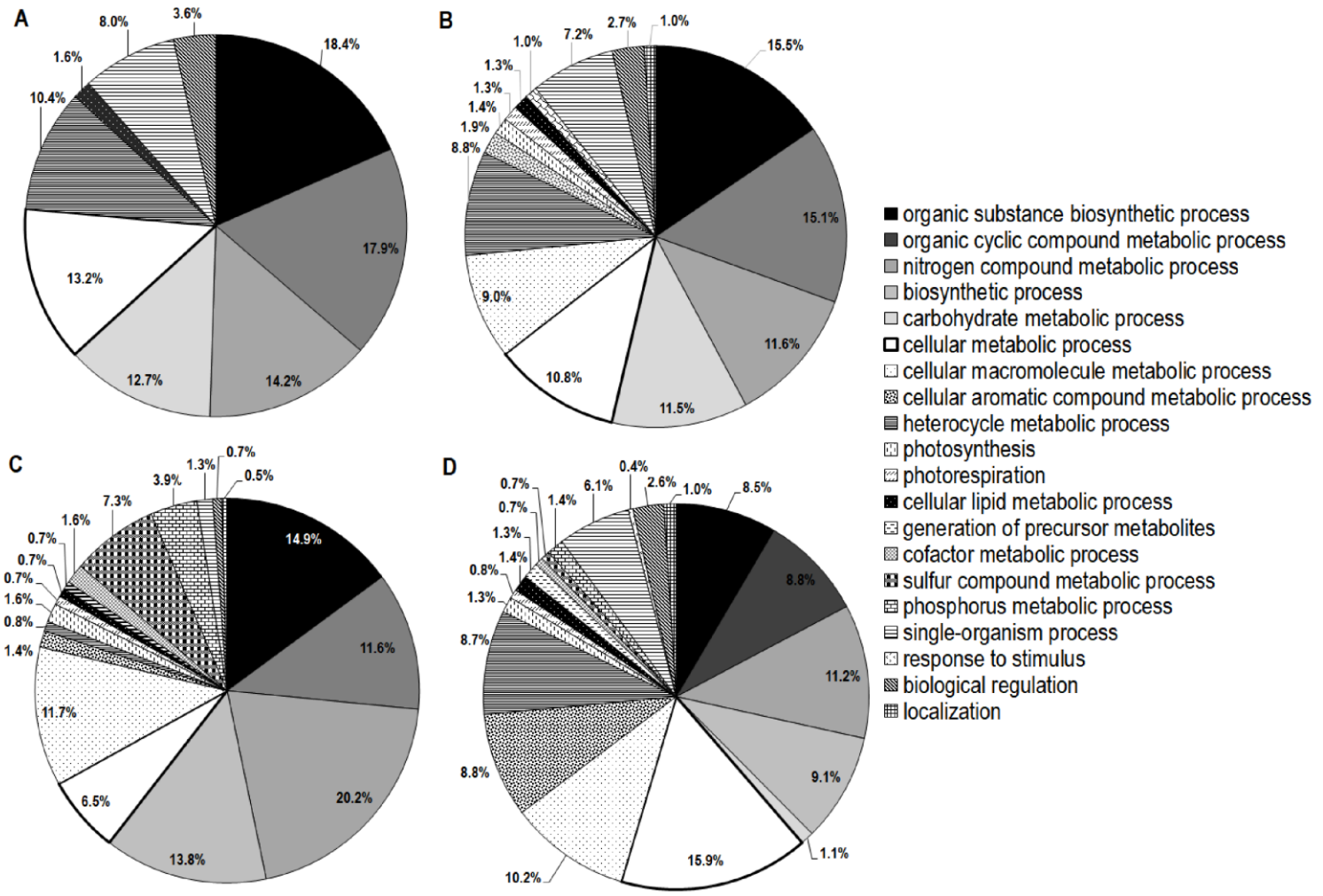


Figure 2. Biological process annotated from UniProtKB for the identified protein. (A) Vb1 extraction condition. (B) Vb2 extraction condition. (C) Vb3 extraction condition. (D) Vb5 extraction condition.

To date, proteomics data of the *Vanilla* genus are restricted to *in vitro* development of *V. planifolia* which focused on apical buds and calluses, identifying proteins related to embryogenesis[27–30]. The extraction protocols followed by these authors, without SDS or any other detergent, lead to the identification of roughly ten times less proteins than the present study. In Gallage et al.'s[3] study, only proteins related to five major enzymes families involved in the vanillin biosynthesis were identified. The larger the identified proteome is, the greater the chances of identifying proteins that can serve as markers for molecules of interest. This would enable the improvement of nutritional quality, taste and resistance or tolerance to diseases[14]. In the present work it was possible to identify a wide diversity of proteins related to vanilla flavor production, and to flowering and fruitening, therefore our extraction method can also be a good alternative to *Vanilla*'s protein extraction.

Identified proteins and classification. The taxonomic data obtained from UniprotKb showed that 88% of the identified proteins belong to the Vanilleae group and from this, 67% were related to *V. planifolia* (**Figure 3**).

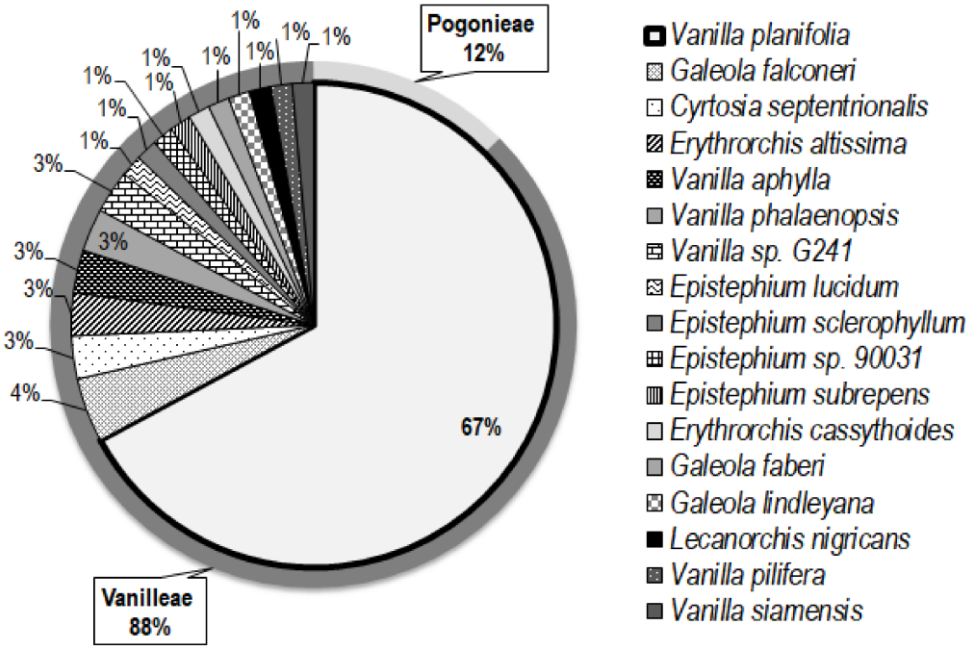


Figure 3. Taxonomy information from UniProt annotation results from identified proteins of *Vanilla spp*.

The group of identified proteins related to flowering and fruitening presented the highest protein diversity (**Figure 4**) in which 103 proteins were related to the development of flowers and 32 were related to fruit development (**Table S2** and **S3**). Most of these proteins were present in Vb3 and Vb5 rather than in Vb1 and Vb2 (**Figure 5**).

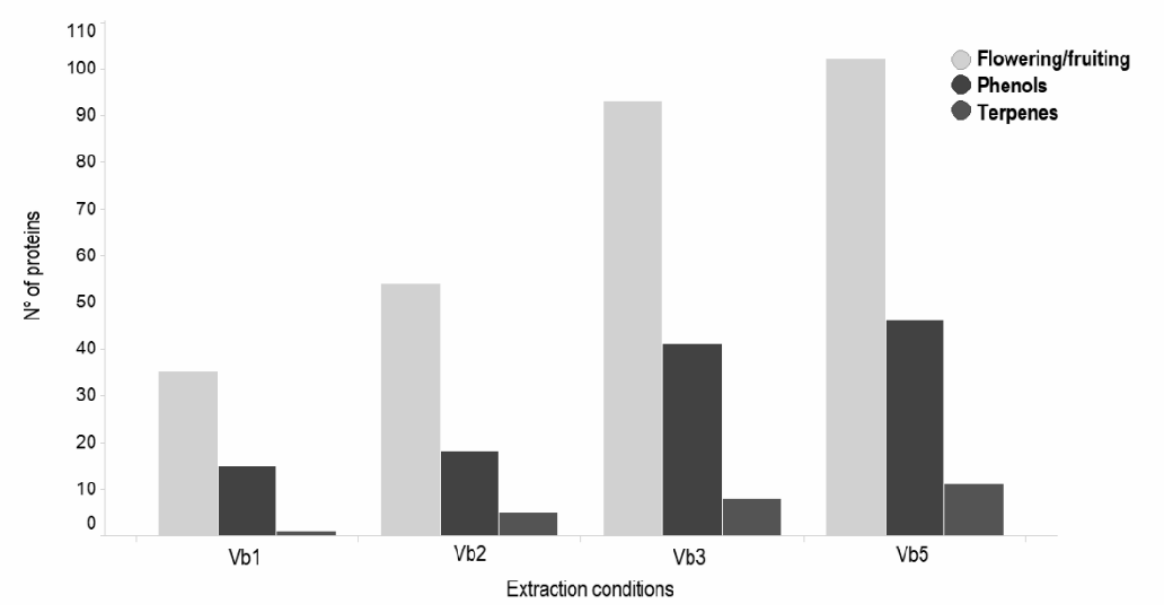


Figure 4. Identification results from nanoUPLC-MS/MS analysis showing the number of proteins identified in the dataset between extraction solutions that are involved in phenols, terpenes, flowering and fruiting.

Amongst the identified proteins, were found the following mads-box group proteins, previously reported in Vanilloideae[32] and peach[33]: 1 protein from agamous (AG), 13 deficiens (DEF), 1 pistillata (PI), 2 apetala (AP), 1 domads1 (DOM1), 1 sepallata (SEP) and 5 squamosa (SQUA). These proteins have diverse functions: encoded transcription factors that controls the development of the floral reproductive organ[34] and in the regulation of fruit development[35] (AG), mainly in *V. planifolia*[32] (DOM1); controls the growth and development of petals and stamens from A, B and C of floral ABCDE, and contribute to floral identity [36,37] (DEF, PI, AP and AG); determines the flower and floral meristem[38] (the SEP group proteins, like sepallata-like mads-box protein 4); regulates transition phase, flower and fruit

development, plant structure, gibberellin signaling and sporogenesis[39] (SQUA proteins and transcription factor TCP4) and regulates the onset of flowering through photoperiodism and the circadian cycle in response to long days[40] (7 CO-like proteins) (**Table S2**).

Specifically LFY is a master regulator, controlling the entire floral network, like mads-box group[41]. LFY proteins were not previously identified in the *Vanilla* genus, however the 12 LFY-like protein OrcLFY was identified in this experiment (**Table S2**).

Maturation and abscission processes in *Orchidaceae* are still unknown. Valadares et al.[42] showed an accumulation of ACC synthase and other ethylene biosynthesis related proteins in green protocorms of *Oncidium sphacelatum* in response to stress. The following ethylene pathway[43] related proteins were identified in the samples of *V. bahiana*: aminocyclopropane-1-carboxylic acid (ACC) synthase, ACC oxidase, ethylene insensitive 3 and ethylene response sensor 2. The putative ethylene response factor was observed in *Oncidium*[44], this protein presents an important role in fruit ripening, flower senescence and abscission[45] (**Tables S3**).

In this present study, 24 proteins intrinsically related to phenolic biosynthesis[46] were found:14 proteins of chalcone synthase, 2 chalcone-flavonone isomerase, 1 flavanone-3-hydroxylase (f3h), 1 flavonoid-3-hydroxylase, 4 flavonoid 3'5'hydroxylase, 1 flavonol synthase and 1 o-methyltransferase-2 (**Table S4**). These proteins had higher abundance in Vb5 or Vb3 with a wider dynamic range than in Vb1 e Vb2 (**Figura 5**, **Figure 6C**, **Table S4, Figure S3C, Figure S4C, Figure S5C)**. Most of the previously mentioned proteins are enzymes related to the biosynthesis of compounds that belong to the vanilla flavor, like: chalcone synthase involved on naringenin biosynthesis[46] and dihydroflavonol 4-reductases related to the biosynthesis of dihydroflanols[47] and of v-myb myeloblastosis (MYB), transcription factors key protein regulator of the synthesis of phenylpropanoid-derived compounds[32,33].

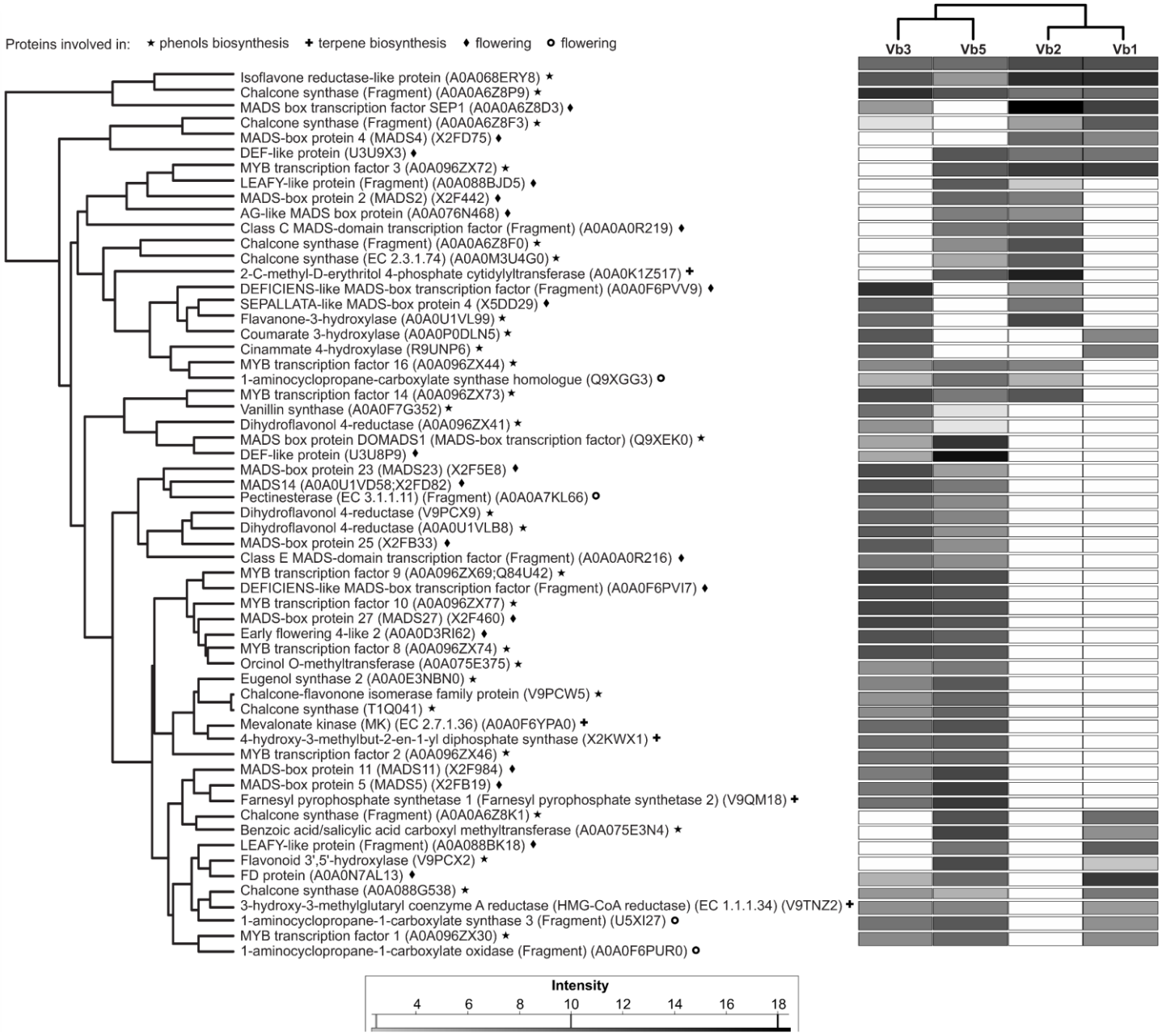
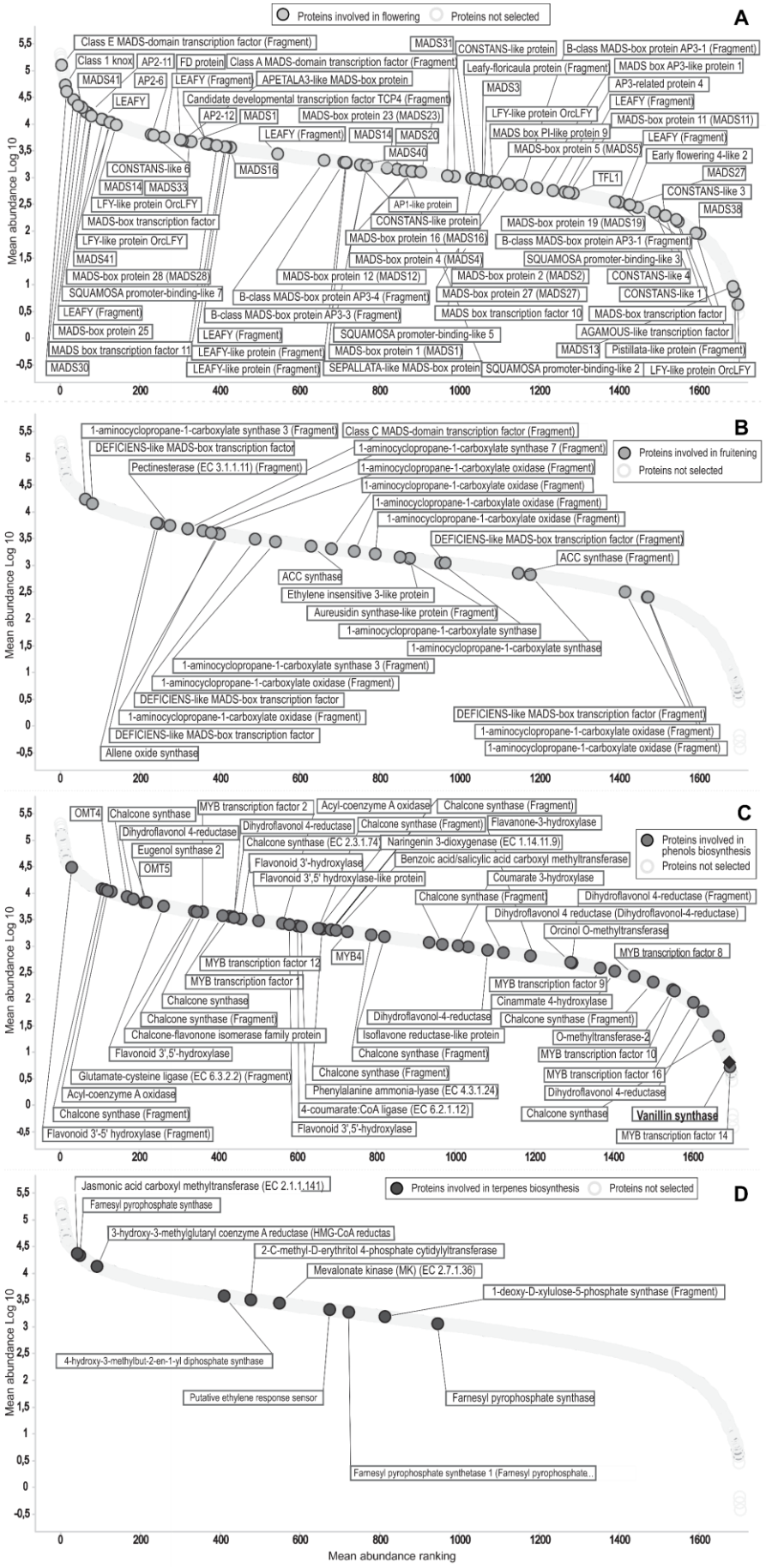


Figure 5. Heat map of the proteins identified and selected, with the ANOVA p-value<0.05. Symbols were used for classification. Star – Phenol pathway; circle – Flowering; diamond – Fruitening and cross – Terpene.

271 According to the literature, some enzymes families are involved specifically in vanillin biosynthesis[3].
272 In the preset study, most of these proteins were identified: 12 phenylalanine ammonia lyase (PAL) and
273 MYB, 7 related to cytochrome p450s family and 4 related to o-methyltransferases (OMTs) (**Figura 5**,
274 **Figure 6C**, **Table S4, Figure S3, Figure S4C, Figure S5C)**.
275 MYB transcription factor family regulates various stages of phenol production pathway and are
276 necessary for the activation of the gene that produces the chalcone synthase enzyme, as well as the
277 activation of the gene that produces PAL[48]. A recent study verified the direct action of MYB
278 transcription factor on the metabolism of flavonoids and phenylpropanoids[49].
279 The OMT and PAL enzymes were characterized in cured and uncured fruits of *V. planifolia* and
280 were responsible for transforming caffeic acid directly into ferulic acid[3]. The caffeic acid o-
281 methyltransferase, found in Vb1 and Vb3 extraction (**Figure 5**, **Table S4**, **Figure S3C**, **S5C**),
282 participates in the biosynthesis of caffeic acid[3,50,51].
283 Eugenol synthase was also identified (**Figure 5**, **Figure 6C**, **Table S4, Figure S3C**, **S4C**, **S5C**), a
284 protein responsible for the biosynthesis of eugenol described in vanilla flavor and in vanillin pathway.
285 This compound is bioconverted into ferulic acid by microorganisms in the curing process of vanilla
286 beans[52].
287 One of the most important findings of the current work was the identification of vanillin synthase,
288 significantly more abundant in Vb3 extraction than in Vb5 and better positioned in the dynamic range
289 (**Figure 5**, **Figure 6C**, **Table S4, Figure S5C**).This enzyme, described by Gallage et al.[3] as the most
290 relevant enzyme for vanillin biosynthesis, belong to the non-CoA-dependent non-β-oxidation
291 transformation of ferulic acid into vanillin. This pathway is still one of the more accepted via for vanillin
292 biosynthesis[51] with vanillin synthase being the only enzyme that converts ferulic acid and its glycoside
293 into vanillin and its glycoside[3,50].


A
Proteins involved in flowering
Proteins not selected
Class E MADS-domain transcription factor (Fragment)
Class 1 knox
AP2-11
FD protein
Class A MADS-domain transcription factor (Fragment)
MADS41
AP2-6
LEAFY (Fragment)
APETALA3-like MADS-box protein
LEAFY
Candidate developmental transcription factor TCP4 (Fragment)
AP2-12
MADS1
MADS-box protein 23 (MADS23)
LEAFY (Fragment)
MADS14
MADS20
MADS40
CONSTANS-like 6
MADS16
MADS14
MADS33
LFY-like protein OrcLFY
MADS-box transcription factor
LFY-like protein OrcLFY
MADS41
MADS-box protein 28 (MADS28)
SQUAMOSA promoter-binding-like 7
LEAFY (Fragment)
MADS-box protein 25
MADS box transcription factor 11
MADS30
LEAFY-like protein (Fragment)
LEAFY-like protein (Fragment)
MADS31
CONSTANS-like protein
B-class MADS-box protein AP3-1 (Fragment)
Leafy-floricaula protein (Fragment)
MADS box AP3-like protein 1
MADS3
AP3-related protein 4
LEAFY (Fragment)
LFY-like protein OrcLFY
MADS-box protein 11 (MADS11)
MADS box PI-like protein 9
LEAFY (Fragment)
MADS-box protein 5 (MADS5)
Early flowering 4-like 2
TFL1
MADS27
CONSTANS-like 3
MADS38
AP1-like protein
CONSTANS-like protein
MADS-box protein 19 (MADS19)
MADS-box protein 16 (MADS16)
B-class MADS-box protein AP3-1 (Fragment)
MADS-box protein 4 (MADS4)
SQUAMOSA promoter-binding-like 3
MADS-box protein 12 (MADS12)
MADS-box protein 2 (MADS2)
CONSTANS-like 4
B-class MADS-box protein AP3-4 (Fragment)
MADS-box protein 27 (MADS27)
CONSTANS-like 1
B-class MADS-box protein AP3-3 (Fragment)
MADS box transcription factor 10
MADS-box transcription factor
LEAFY (Fragment)
SQUAMOSA promoter-binding-like 5
AGAMOUS-like transcription factor
LEAFY-like protein (Fragment)
MADS-box protein 1 (MADS1)
MADS13
Pistillata-like protein (Fragment)
SEPALLATA-like MADS-box protein
SQUAMOSA promoter-binding-like 2
LFY-like protein OrcLFY
Mean abundance Log 10
B
Proteins involved in fruitening
Proteins not selected
1-aminocyclopropane-1-carboxylate synthase 3 (Fragment)
Class C MADS-domain transcription factor (Fragment)
DEFICIENS-like MADS-box transcription factor
1-aminocyclopropane-1-carboxylate synthase 7 (Fragment)
Pectinesterase (EC 3.1.1.11) (Fragment)
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
DEFICIENS-like MADS-box transcription factor (Fragment)
ACC synthase (Fragment)
ACC synthase
Ethylene insensitive 3-like protein
Aureusidin synthase-like protein (Fragment)
1-aminocyclopropane-1-carboxylate synthase
1-aminocyclopropane-1-carboxylate synthase
1-aminocyclopropane-1-carboxylate synthase 3 (Fragment)
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
DEFICIENS-like MADS-box transcription factor
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
DEFICIENS-like MADS-box transcription factor
Allene oxide synthase
DEFICIENS-like MADS-box transcription factor (Fragment)
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
Mean abundance Log 10
C
Proteins involved in phenols biosynthesis
Proteins not selected
OMT4
Chalcone synthase
MYB transcription factor 2
Acyl-coenzyme A oxidase
Chalcone synthase (Fragment)
Dihydroflavonol 4-reductase
Dihydroflavonol 4-reductase
Chalcone synthase (Fragment)
Flavanone-3-hydroxylase
Eugenol synthase 2
Chalcone synthase (EC 2.3.1.74)
Naringenin 3-dioxygenase (EC 1.14.11.9)
OMT5
Flavonoid 3'-hydroxylase
Benzoic acid/salicylic acid carboxyl methyltransferase
Flavonoid 3',5' hydroxylase-like protein
Coumarate 3-hydroxylase
Chalcone synthase (Fragment)
Dihydroflavonol 4-reductase (Fragment)
Dihydroflavonol 4 reductase (Dihydroflavonol-4-reductase)
Orcinol O-methyltransferase
MYB4
MYB transcription factor 8
MYB transcription factor 12
MYB transcription factor 1
MYB transcription factor 9
Cinammate 4-hydroxylase
Chalcone synthase
Chalcone synthase (Fragment)
Dihydroflavonol-4-reductase
Chalcone synthase (Fragment)
Chalcone-flavonone isomerase family protein
Isoflavone reductase-like protein
O-methyltransferase-2
Flavonoid 3',5'-hydroxylase
Chalcone synthase (Fragment)
MYB transcription factor 10
Glutamate-cysteine ligase (EC 6.3.2.2) (Fragment)
Chalcone synthase (Fragment)
MYB transcription factor 16
Acyl-coenzyme A oxidase
Phenylalanine ammonia-lyase (EC 4.3.1.24)
Dihydroflavonol 4-reductase
Chalcone synthase (Fragment)
4-coumarate:CoA ligase (EC 6.2.1.12)
Chalcone synthase
Vanillin synthase
Flavonoid 3'-5' hydroxylase (Fragment)
Flavonoid 3',5'-hydroxylase
MYB transcription factor 14
Mean abundance Log 10
D
Proteins involved in terpenes biosynthesis
Proteins not selected
Jasmonic acid carboxyl methyltransferase (EC 2.1.1.141)
Farnesyl pyrophosphate synthase
3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA reductas
2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase
Mevalonate kinase (MK) (EC 2.7.1.36)
1-deoxy-D-xylulose-5-phosphate synthase (Fragment)
4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase
Putative ethylene response sensor
Farnesyl pyrophosphate synthase
Farnesyl pyrophosphate synthetase 1 (Farnesyl pyrophosphate...
Mean abundance Log 10
Mean abundance ranking

295 **Figure 5.** Dynamic range of proteins identified in the dataset of the *V. bahiana*'s bean for Vb5
296 extraction condition. A- flowering; B- fruitening; C-phenolic and D- terpene pathway. The median
297 absolute expression value of each protein revealing the typical S-shaped distribution over the mean
298 abundance orders of dynamic range. The most abundant proteins (left) and the lowest abundance
299 (right).

300 Some terpenes have been described as part of the vanilla essential oil and flavor present in *Vanilla*
301 *spp*., such as: limonene, linalool, terpinen-4-ol, α-pinene, α-terpineol and β-pinene[7,53]. In our work, we
302 were able to find some proteins related to these compounds' biosynthesis[54], such as: terpene
303 synthase, 1-deoxy-D-xylulose-5-phosphate synthase (Fragment), allene oxide synthase, 3-hydroxy-3-
304 methylglutaryl coenzyme A reductase and 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase
305 (**Figure 5**, **Figure 6D**, **Figure S3D**, **S4D, S5D**). All of them were present and/or had higher abundance
306 in Vb5 (**Figure 5**, **Figure 6D**, **Table S5**).

307 The mevalonate kinase, statistically more abundant in Vb3 than Vb5 (**Figure 5**, **Figure 6D** , **Figure**
308 **S5D, Table S5**), is the first enzyme to act right after HMG-CoA in the mevalonate pathway [55] and it is
309 related to the same pathway of 4-hydroxy-3-methoxyphenyl-β-hydroxypropionyl compound, that can
310 be cleaved into vanillin and acetyl-CoA[56].

311 *Quantification of key vanilla flavor compounds.* To validate the presence of some phenolic
312 compounds in *V. bahiana* samples, vanillin, *p*-coumaric acid and pyrogallol were extracted and
313 quantified (**Table 1**). According to Perez-Silva et al.[57], acids and phenolic compounds were
314 characteristic predominates in *V. planifolia* cured fruit extracts, being vanillin the main phenolic
315 compound. Was quantified 1.55 mg/g of dried extract of vanillin (**Table 1**). The vanillin and related
316 phenolic compounds in cured *Vanilla* spp. pods can be an important indicators of fruit and extract
317 quality and origin[58]. According to Gallage and Møller[59] vanillin in the mature fruit has high
318 concentrations, 1kg of vanillin per 500kg of vanilla fruit, and cannot be compared to any other known
319 organism in nature. The comparison of the amount of vanillin per g of dried fruit obtained in our study

was compared to the other studies already published and the results are in **Table 2**. We observed that the differences between the vanillin content in the present work and those cited in the literature were at most approximately 0.02g/g of dry fruit. The *p*-coumaric acid is an important precursor of vanillin in fruit and cell cultures of *V. planifolia* and in *V. tahitensis* pod[29,47,60]. The following proteins, identified in the samples, are supposed to take part in its biosynthetic pathway[3]: PAL and cytochrome p450s family proteins like Flavonoid 3'-hydroxylase (Figure 5, 6D e material suplementar). Other phenolic compound quantified was pyrogallol, the most abundant, that was previously identified in the flavor of *V. planifolia*[53] and in *V. tahitensis* pod[47] (**Table 1**).

Table 1. Mean quantification of vanillin, p-coumaric acid and pyrogallol.

Compound	Mean concentration (mg/g)*
Vanillin	1.55 ± 0.00**
p-coumaric acid	0.18 ± 0.00
Pyrogallol	297.42 ± 7.72

**Mean ± Standard deviations
* Dry extract

Table 2. Comparation of vanillin content in fruit of *Vanilla spp*. according to literature.

Literature	Species	Vanillin Content (g/g of dry extract)	Vanillin content (g/g of dry fruit)
Present study	*V. bahiana*	0.0015	0.0155
Calva-Estrada et al, 2017	-*	0.0161	-
Wongsheree et al. 2013	*V. planifolia*	-	0.0368
Kumar et al. 2010	*V. planifolia*	-	0.0266
Gassenmeier et al, 2008	*V. planifolia***	-	0.0245
Gassenmeier et al, 2008	*V. planifolia****	-	0.0086
Sagrero-Nives & Schwartz, 1988	*V. planifolia*	0.0113	-

*natural Vanilla extract/ ***V. planifolia*** extract from Madagascar, Red non-Split color type/ *** *V. planifolia* extract from Madagascar, cuts shape type

The *V. bahiana* pod seems to be an alternative source of the natural vanilla extract since it proved to express some of the most important enzymes in the biosynthesis of the vanilla flavor compounds and produces vanillin. The extraction, digestion, acquisition and identification methods in MS used were adequate (**Figure S6 – S9**). The use of 1% of SDS was important, not only for the extraction of the most abundant proteins, in a wide dynamic range, but also for the extraction of a great protein diversity indicating that this could be used in protein studies of other *Vanilla spp*. pods.

Acknowledgements

The authors are grateful to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for master's scholarship; to Dr. Andrés Rodríguez Veja from PPGAN/UNIRIO, to Dr. André Ferreira from Oswaldo Cruz Foundation (FIOCRUZ) and to Gustavo Souza from Waters Corporation for logistical and technical support. This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), number 2010.3553.8, and Universidade Federal do Estado do Rio de Janeiro (Unirio).

References

(1) Pansarin, E.; Aguiar, J.; Ferreira, Al. A new species of Vanilla (Orchidaceae : Vanilloideae) from São. *New York Bot. Gard. Press* **2012**, *64* (June), 157–161.

(2) Divakaran, M.; Babu, K. N.; Peter, K. V. Conservation of Vanilla species, in vitro. *Sci. Hortic. (Amsterdam).* **2006**, *110* (2), 175–180.

(3) Gallage, N. J.; Hansen, E. H.; Kannangara, R.; Olsen, C. E.; Motawia, M. S.; Jørgensen, K.; Holme, I.; Hebelstrup, K.; Grisoni, M.; Møller, B. L. Vanillin formation from ferulic acid in Vanilla planifolia is catalysed by a single enzyme. *Nat. Commun.* **2014**, *5* (May), 4037.

(4) Hrazdina, G. Aroma Production by Tissue Cultures Aroma Production by Tissue Cultures. *J. Agric. Food Chem.* **2006**, *54* (February), 1116–1123.

(5) Greule, M.; Mosandl, A.; Hamilton, J. T. G.; Keppler, F. Comment on Authenticity and Traceability of *Vanilla* Flavors by Analysis of Stable Isotopes of Carbon and Hydrogen. *J. Agric. Food Chem.* **2015**, *63* (21), 5305–5306.

(6) Koyyappurath, S.; Atuahiva, T.; Le Guen, R.; Batina, H.; Le Squin, S.; Gautheron, N.; Edel Hermann, V.; Peribe, J.; Jahiel, M.; Steinberg, C.; et al. Fusarium oxysporum f. sp. radicis-vanillae is the causal agent of root and stem rot of vanilla. *Plant Pathol.* **2016**, *65* (4), 612–625.

(7) Anuradha, K.; Shyamala, B. N.; Naidu, M. M. Vanilla- Its science of cultivation, curing, chemistry, and nutraceutical properties. *Crit. Rev. Food Sci. Nutr.* **2013**, *53* (12), 1250–1276.

(8) Naidu, M. M.; Kumar, P. V. S.; Shyamala, B. N.; Sulochanamma, G.; Prakash, M.; Thakur, M. S. Enzyme-Assisted Process for Production of Superior Quality Vanilla Extracts from Green Vanilla Pods Using Tea Leaf Enzymes. *Food Bioprocess Technol.* **2012**, *5* (2), 527–532.

(9) Botânico, J. *Arquivos do Jardim botânico do Rio de Janeiro*; 1915; Vol. XVIII.

(10) Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72* (1–2), 248–254.

(11) Lobo, M. D. P.; Moreno, F. B. M. B.; Souza, G. H. M. F.; Verde, S. M. M. L.; Moreira, R. de A.; Monteiro-Moreira, A. C. de O. Label-Free Proteome Analysis of Plasma from Patients with Breast Cancer: Stage-Specific Protein Expression. *Front. Oncol.* **2017**, *7* (February), 1–12.

(12) Victorio, V. C. M.; Souza, G. H. M. F.; Santos, M. C. B.; Vega, A. R.; Cameron, L. C.; Ferreira, M. S. L. Differential expression of albumins and globulins of wheat flours of different technological qualities revealed by nanoUPLC-UDMSE. *Food Chem.* **2018**, *239*, 1027–1036.

(13) Li, G. Z.; Vissers, J. P. C.; Silva, J. C.; Golick, D.; Gorenstein, M. V.; Geromanos, S. J. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **2009**, *9* (6), 1696–1719.

(14) Song, J.; Braun, G. Application of Proteomic Techniques to Fruits and Vegetables. *Curr. Proteomics* **2008**, *5* (3), 191–201.

(15) Zhang, S.; Zhang, L.-L.; Zhou, K.-K.; Liu, Y.-J.; Zhao, Z. Evaluation of three types of protein extraction methods for tetraploid black locust (Robinia pseudoacacia L.) phloem tissue proteome analysis by two-dimensional electrophoresis. *Anal. Methods* **2015**, *7*, 1008–1017.

(16) Tamayo Tenorio, A.; Boom, R. M.; van der Goot, A. J. Understanding leaf membrane protein extraction to develop a food-grade process. *Food Chem.* **2017**, *217*, 234–243.

(17) Capriotti, A. L.; Cavaliere, C.; Piovesana, S.; Stampachiacchiere, S.; Ventura, S.; Zenezini Chiozzi, R.; Laganà, A. Characterization of quinoa seed proteome combining different protein precipitation techniques: Improvement of knowledge of nonmodel plant proteomics. *J. Sep. Sci.* **2015**, *38* (6), 1017–1025.

(18) Garrido, B. C.; Souza, G. H. M. F.; Lourenço, D. C.; Fasciotti, M. Proteomics in quality control: Whey protein-based supplements. *J. Proteomics* **2016**, *147*, 48–55.

(19) Wilson, K.; Walker, J. *Principles and techniques of biochemistry and molecular biology*; 2010.

(20) Chengang, R.; Lu, T.; Min, Z.; Shuntang, G. Structural characterization of heat-induced

protein particles in soy milk. *J. Agric. Food Chem.* **2009**, *57* (5), 1921–1926.

(21) Anand, U.; Ray, S.; Ghosh, S.; Banerjee, R.; Mukherjee, S. Structural Aspects of a Protein–Surfactant Assembly: Native and Reduced States of Human Serum Albumin. *Protein J.* **2015**, *34* (2), 147–157.

(22) Capriotti, A. L.; Caruso, G.; Cavaliere, C.; Samperi, R.; Stampachiacchiere, S.; Zenezini Chiozzi, R.; Laganà, A. Protein profile of mature soybean seeds and prepared soybean milk. *J. Agric. Food Chem.* **2014**, *62* (40), 9893–9899.

(23) Botelho, D.; Wall, M. J.; Vieira, D. B.; Fitzsimmons, S.; Liu, F.; Doucette, A. Top-down and bottom-up proteomics of sds-containing solutions following mass-based separation. *J. Proteome Res.* **2010**, *9* (6), 2863–2870.

(24) Liu, Y.; Lin, Y.; Yan, Y.; Li, J.; He, Q.; Chen, P.; Wang, X.; Liang, S. Electrophoretically driven SDS removal and protein fractionation in the shotgun analysis of membrane proteomes. *Electrophoresis* **2012**, *33* (2), 316–324.

(25) Lin, Y.; Jiang, H.; Yan, Y.; Peng, B.; Chen, J.; Lin, H.; Liu, Z. Shotgun analysis of membrane proteomes by an improved SDS-assisted sample preparation method coupled with liquid chromatography-tandem mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2012**, *911*, 6–14.

(26) Gallage, N. J.; Jørgensen, K.; Janfelt, C.; Nielsen, A. J. Z.; Naake, T.; Duński1, E.; Møller, B. L.; Grisoni, M.; Dalsten, L. The Intra-Cellular Localization of the Vanillin Biosynthetic Machinery in Pods of Vanilla planifolia Running head : Vanillin biosynthesis in chloroplasts The Intra-Cellular Localization of the Vanillin Biosynthetic Machinery in Pods of Vanilla planifolia. **2017**, No. November.

(27) Tan, B. C.; Chin, C. F.; Liddell, S.; Alderson, P. Protein extraction for callus and node cultures of Vanilla planifolia Andrews. *Minerva Biotecnol.* **2014**, *26* (July), 115–126.

(28) Guerrero, A.; Valdés-Rodríguez, S. .; Durán-Sánchez, B.; González-Arnao, M. .; Jiménez-

Francisco, B.; Lázaro-Vallejo, C. . Cryopreservation and Proteomic Analysis of Vanilla (V . planifolia A .) Apices Treated with Osmoprotectants. *Acta Hortic.* **2011**, No. 908, 67–72.

(29) Palama, T. L.; Menard, P.; Fock, I.; Choi, Y. H.; Bourdon, E.; Govinden-Soulange, J.; Bahut, M.; Payet, B.; Verpoorte, R.; Kodja, H. Shoot differentiation from protocorm callus cultures of Vanilla planifolia (Orchidaceae): proteomic and metabolic responses at early stage. *BMC Plant Biol.* **2010**, *10*, 82.

(30) Tan, B. C.; Chin, C. F.; Liddell, S.; Alderson, P. Proteomic Analysis of Callus Development in Vanilla planifolia Andrews. *Plant Mol. Biol. Report.* **2013**, *31* (6), 1220–1229.

(31) Villanueva-Viramontes, S.; Hernández-Apolinar, M.; Carnevali Fernández-Concha, G.; Dorantes-Euán, A.; Dzib, G. R.; Martínez-Castillo, J. Wild Vanilla planifolia and its relatives in the Mexican Yucatan Peninsula: Systematic analyses with ISSR and ITS. *Bot. Sci.* **2017**, *95* (2), 169–187.

(32) Acri-Nunes-Miranda, R.; Mondragón-Palomino, M. Expression of paralogous SEP-, FUL-, AG- and STK-like MADS-box genes in wild-type and peloric Phalaenopsis flowers. *Front. Plant Sci.* **2014**, *5* (March), 76.

(33) Tani, E.; Polidoros, A. N.; Flemetakis, E.; Stedel, C.; Kalloniati, C.; Demetriou, K.; Katinakis, P.; Tsaftaris, A. S. Characterization and expression analysis of AGAMOUS-like, SEEDSTICK-like, and SEPALLATA-like MADS-box genes in peach (Prunus persica) fruit. *Plant Physiol. Biochem.* **2009**, *47* (8), 690–700.

(34) Honma, T.; Goto, K. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* **2001**, *409* (6819), 525–529.

(35) Roy Choudhury, S.; Roy, S.; Nag, A.; Singh, S. K.; Sengupta, D. N. Characterization of an AGAMOUS-like MADS Box Protein, a Probable Constituent of Flowering and Fruit Ripening Regulatory System in Banana. *PLoS One* **2012**, *7* (9).

(36) Kramer, E. M.; Dorit, R. L.; Irish, V. F. Molecular evolution of genes controlling petal and

456 stamen development: Duplication and divergence within the APETALA3 and PISTILLATA
457 MADS-box gene lineages. *Genetics* **1998**, *149* (2), 765–783.
458 (37) Melzer, R.; Härter, A.; Rümpler, F.; Kim, S.; Soltis, P. S.; Soltis, D. E.; Theißen, G. DEF- and
459 GLO-like proteins may have lost most of their interaction partners during angiosperm
460 evolution. *Ann. Bot.* **2014**, *114* (7), 1431–1443.
461 (38) Wong, C. E.; Singh, M. B.; Bhalla, P. L. Novel members of the AGAMOUS LIKE 6 subfamily
462 of MIKCC-type MADS-box genes in soybean. *BMC Plant Biol.* **2013**, *13* (1), 105.
463 (39) Preston, J. C.; Hileman, L. C. Functional Evolution in the Plant SQUAMOSA-PROMOTER
464 BINDING PROTEIN-LIKE (SPL) Gene Family. *Front. Plant Sci.* **2013**, *4*.
465 (40) Valverde, F. Photoreceptor Regulation of CONSTANS Protein in Photoperiodic Flowering.
466 *Science (80-.).* **2004**, *303* (5660), 1003–1006.
467 (41) Lu, S.; Li, Z.; Zhang, J.; Yi, S.; Liu, L.; Bao, M.; Liu, G. Isolation and expression analysis of a
468 LEAFY/FLORICAULA homolog and its promoter from London plane (Platanus acerifolia
469 Willd.). *Plant Cell Rep.* **2012**, *31* (10), 1851–1865.
470 (42) Valadares, R. B. S.; Perotto, S.; Santos, E. C.; Lambais, M. R. Proteome changes in
471 Oncidium sphacelatum (Orchidaceae) at different trophic stages of symbiotic germination.
472 *Mycorrhiza* **2014**, *24* (5), 349–360.
473 (43) Pilati, S.; Perazzolli, M.; Malossini, A.; Cestaro, A.; Demattè, L.; Fontana, P.; Dal Ri, A.; Viola,
474 R.; Velasco, R.; Moser, C. Genome-wide transcriptional analysis of grapevine berry ripening
475 reveals a set of genes similarly modulated during three seasons and the occurrence of an
476 oxidative burst at vèraison. *BMC Genomics* **2007**, *8* (1), 428.
477 (44) Lashbrook, C. C.; Giovannoni, J. J.; Hall, B. D.; Fischer, R. L.; Bennett, A. B. Transgenic
478 analysis of tomato endo-β-1,4-glucanase gene function. Role of cel1 in floral abscission.
479 *Plant J.* **1998**, *13* (3), 303–310.
480 (45) Bäurle, I.; Dean, C. The Timing of Developmental Transitions in Plants. *Cell* **2006**, *125* (4),

655–664.

(46) Li, X.; Jackson, A.; Xie, M.; Wu, D.; Tsai, W. C.; Zhang, S. Proteomic insights into floral biology. *Biochim. Biophys. Acta - Proteins Proteomics* **2016**, *1864* (8), 1050–1060.

(47) Busconi, M.; Lucini, L.; Soffritti, G.; Bernardi, J.; Bernardo, L.; Brunschwig, C.; Lepers-Andrzejewski, S.; Raharivelomanana, P.; Fernandez, J. A. Phenolic Profiling for Traceability of Vanilla ×tahitensis. *Front. Plant Sci.* **2017**, *8* (October), 1–13.

(48) Liu, J.; Osbourn, A.; Ma, P. MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol. Plant* **2015**, *8* (5), 689–708.

(49) Medina-Puche, L.; Cumplido-Laso, G.; Amil-Ruiz, F.; Hoffmann, T.; Ring, L.; Rodríguez-Franco, A.; Caballero, J. L.; Schwab, W.; Muñoz-Blanco, J.; Blanco-Portales, R. MYB10 plays a major role in the regulation of flavonoid/phenylpropanoid metabolism during ripening of Fragaria × ananassa fruits. *J. Exp. Bot.* **2014**, *65* (2), 401–417.

(50) Negishi, O.; Negishi, Y. Phenylpropanoid 2,3-dioxygenase involved in the cleavage of the ferulic acid side chain to form vanillin and glyoxylic acid in Vanilla planifolia. *Biosci. Biotechnol. Biochem.* **2017**, *81* (9), 1732–1740.

(51) Kundu, A. Vanillin biosynthetic pathways in plants. *Planta* **2017**, *245* (6), 1069–1078.

(52) Havkin-frenkel, D.; Belanger, F. C. Biotechnological production of vanillin (Priefert 2001).pdf. **2016**, 1–121.

(53) Zhang, S.; Mueller, C. Comparative analysis of volatiles in traditionally cured Bourbon and Ugandan vanilla bean (Vanilla planifolia) extracts. *J. Agric. Food Chem.* **2012**, *60* (42), 10433–10444.

(54) Anuradha, K.; Shyamala, B. N.; Naidu, M. M. Vanilla--its science of cultivation, curing, chemistry, and nutraceutical properties. *Crit. Rev. Food Sci. Nutr.* **2013**, *53* (12), 1250–1276.

(55) Bouvier, F.; Rahier, A.; Camara, B. Biogenesis, molecular regulation and function of plant isoprenoids. *Prog. Lipid Res.* **2005**, *44* (6), 357–429.

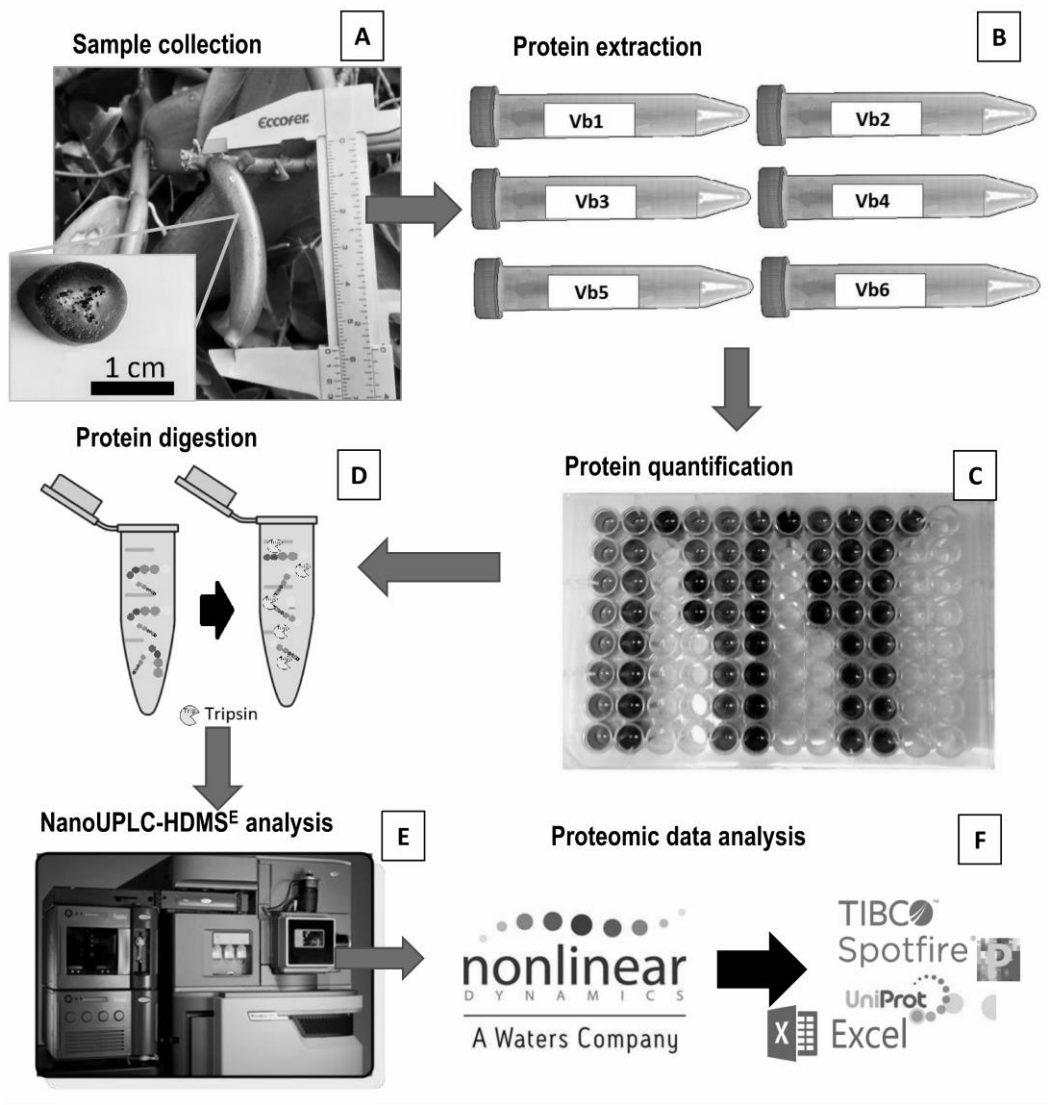
(56) Świzdor, A.; Panek, A.; Milecka-Tronina, N.; Kołek, T. Biotransformations utilizing β-oxidation cycle reactions in the synthesis of natural compounds and medicines. *Int. J. Mol. Sci.* **2012**, *13* (12), 16514–16543.

(57) Pérez-Silva, a.; Odoux, E.; Brat, P.; Ribeyre, F.; Rodriguez-Jimenes, G.; Robles-Olvera, V.; García-Alvarado, M. a.; Günata, Z. GC-MS and GC-olfactometry analysis of aroma compounds in a representative organic aroma extract from cured vanilla (Vanilla planifolia G. Jackson) beans. *Food Chem.* **2006**, *99* (4), 728–735.

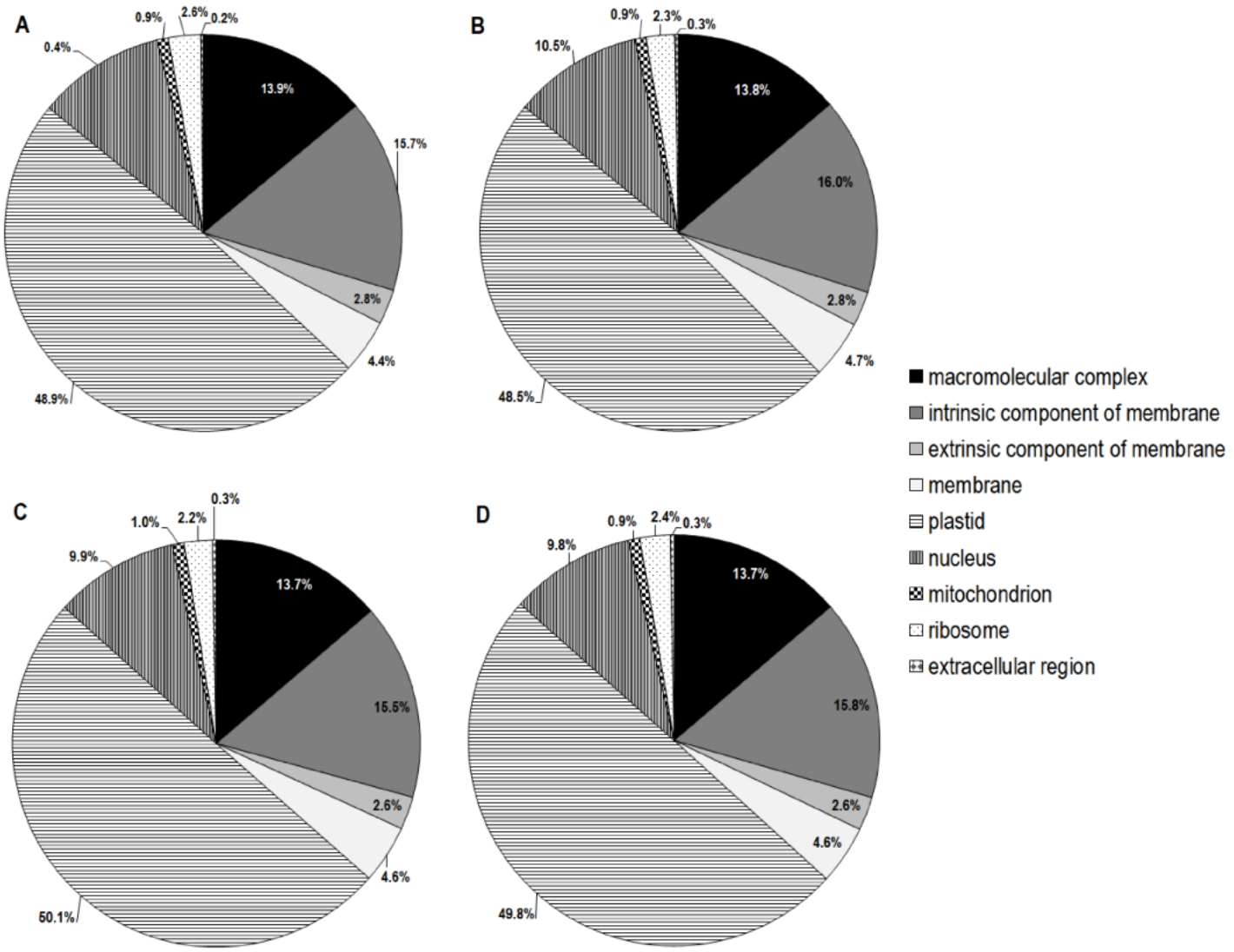
(58) Sharma, U. K.; Sharma, N.; Gupta, A. P.; Kumar, V.; Sinha, A. K. RP-HPTLC densitometric determination and validation of vanillin and related phenolic compounds in accelerated solvent extract of Vanilla planifolia. *J. Sep. Sci.* **2007**, *30* (18), 3174–3180.

(59) Gallage, N. J.; Møller, B. L. Vanillin–Bioconversion and Bioengineering of the Most Popular Plant Flavor and Its De Novo Biosynthesis in the Vanilla Orchid. *Mol. Plant* **2015**, *8* (1), 40–57.

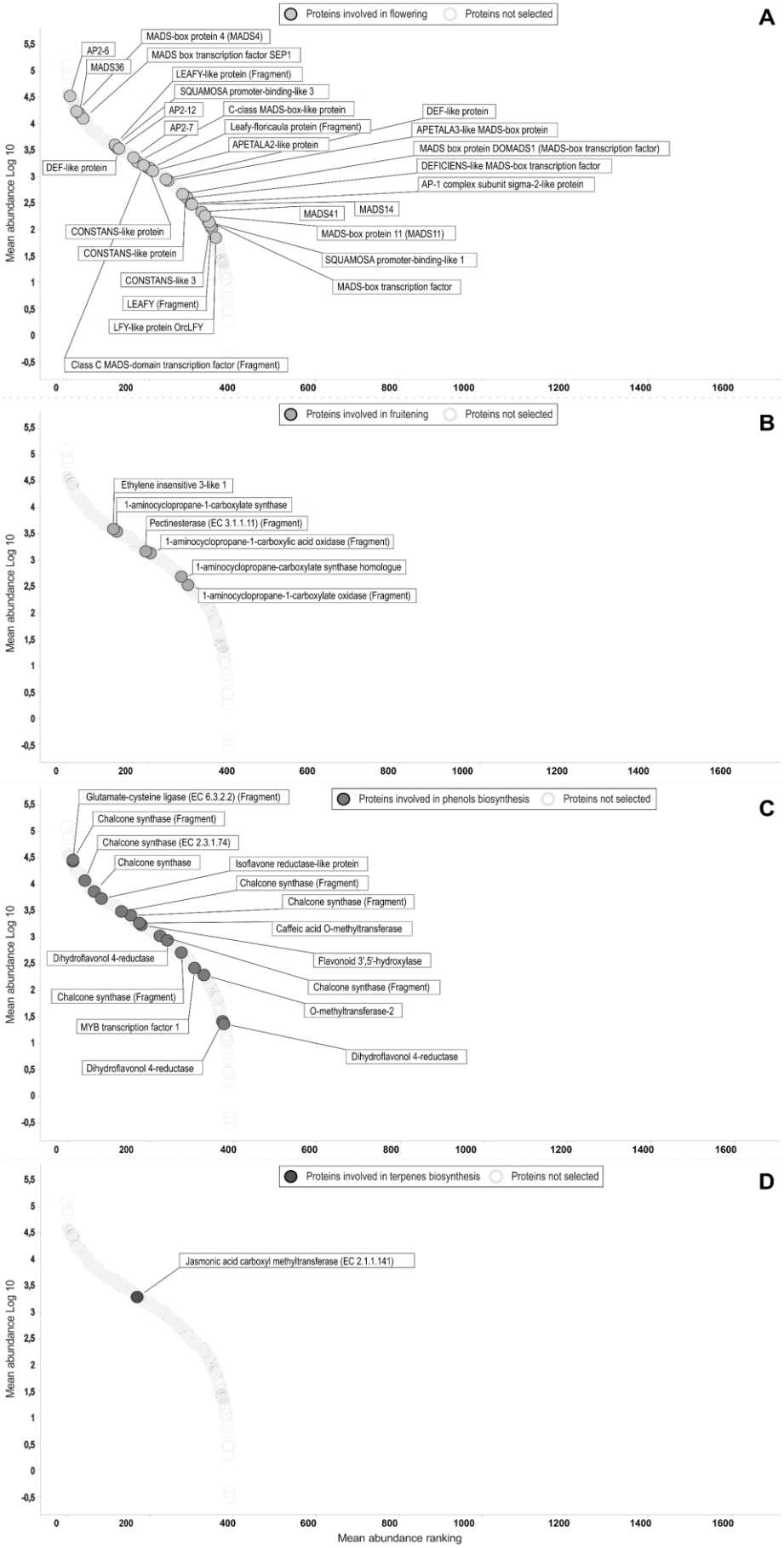
(60) Gu, F.; Chen, Y.; Hong, Y.; Fang, Y.; Tan, L. Comparative metabolomics in vanilla pod and vanilla bean revealing the biosynthesis of vanillin during the curing process of vanilla. *AMB Express* **2017**, *7* (1), 116.



Supplementary Figure 1. Workflow of the protein extraction and peptide and protein identification preparation of *V. bahiana*'s fruit using bottom-up and LC-MS/MS analysis. (A) Fruit collection in the Natural Monument of Pão de Açúcar and Urca, with a transversal cut of the fruit. (B) Six different extraction solutions. (C) The proteins samples were quantified. (D) The proteins were digested using trypsin. (E) The peptides were analyzed by a nanoUPLC-HDMSE system using a nanoUPLC-RP SYNAPT G2-S HDMS instrument (Waters, Manchester, UK). (F) The raw MS data were processed, statistical analysis and graphics with the results were created.

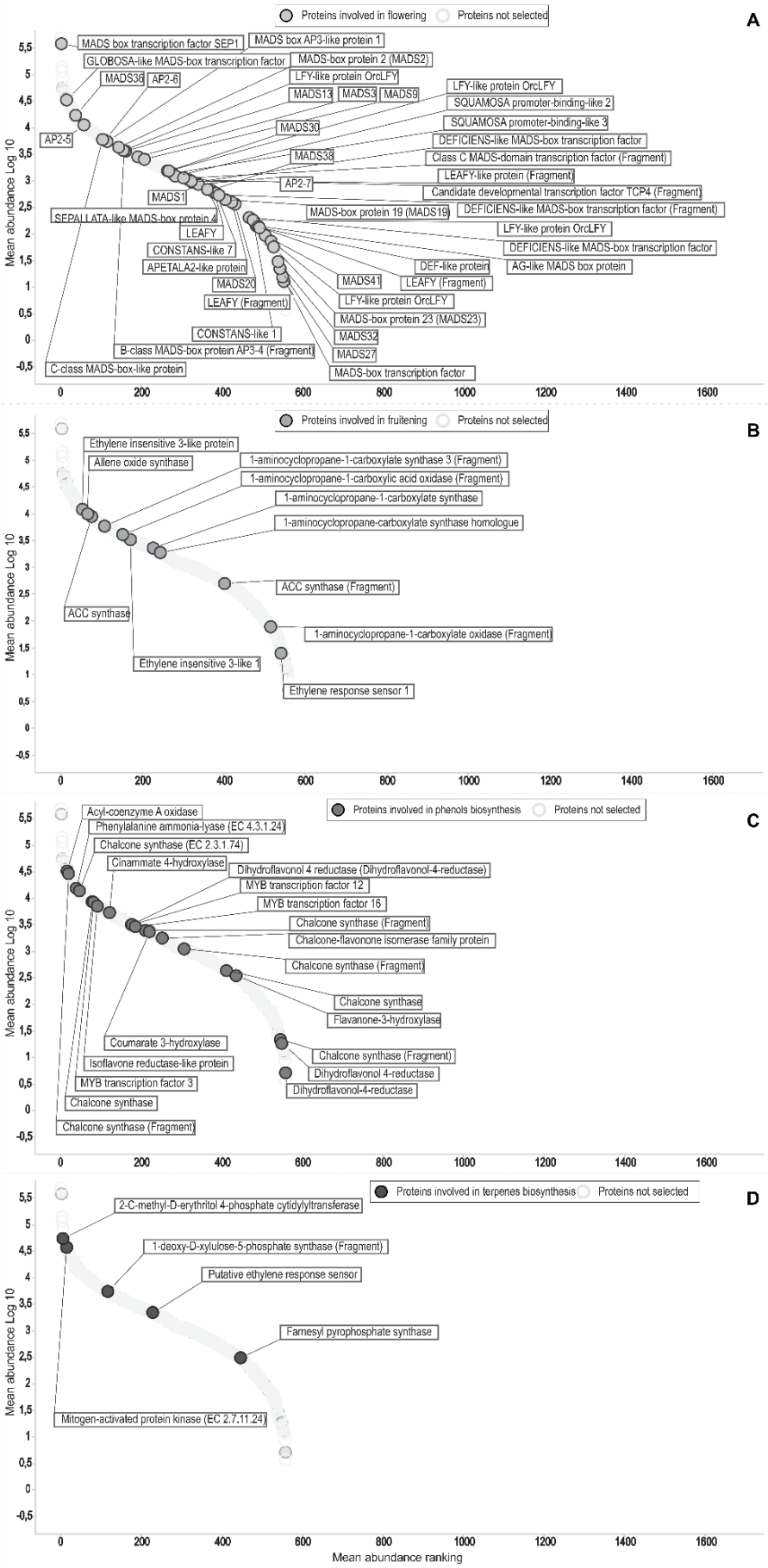


Supplementary Figure 2. Cellular component annotated from UniProtKB for the identified protein. (A) proteins from Vb1 extraction condition. (B) Vb2 extraction condition. (C) Vb3 extraction condition. (D) Vb5 extraction condition.


A
Proteins involved in flowering
Proteins not selected
MADS-box protein 4 (MADS4)
AP2-6
MADS box transcription factor SEP1
MADS36
LEAFY-like protein (Fragment)
SQUAMOSA promoter-binding-like 3
AP2-12
C-class MADS-box-like protein
DEF-like protein
AP2-7
Leafy-floricaula protein (Fragment)
APETALA3-like MADS-box protein
APETALA2-like protein
MADS box protein DOMADS1 (MADS-box transcription factor)
DEF-like protein
DEFICIENS-like MADS-box transcription factor
AP-1 complex subunit sigma-2-like protein
MADS41
MADS14
CONSTANS-like protein
MADS-box protein 11 (MADS11)
CONSTANS-like protein
SQUAMOSA promoter-binding-like 1
CONSTANS-like 3
MADS-box transcription factor
LEAFY (Fragment)
LFY-like protein OrcLFY
Class C MADS-domain transcription factor (Fragment)
Mean abundance Log 10
B
Proteins involved in fruitening
Proteins not selected
Ethylene insensitive 3-like 1
1-aminocyclopropane-1-carboxylate synthase
Pectinesterase (EC 3.1.1.11) (Fragment)
1-aminocyclopropane-1-carboxylic acid oxidase (Fragment)
1-aminocyclopropane-carboxylate synthase homologue
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
Mean abundance Log 10
C
Glutamate-cysteine ligase (EC 6.3.2.2) (Fragment)
Proteins involved in phenols biosynthesis
Proteins not selected
Chalcone synthase (Fragment)
Chalcone synthase (EC 2.3.1.74)
Chalcone synthase
Isoflavone reductase-like protein
Chalcone synthase (Fragment)
Chalcone synthase (Fragment)
Caffeic acid O-methyltransferase
Dihydroflavonol 4-reductase
Flavonoid 3',5'-hydroxylase
Chalcone synthase (Fragment)
Chalcone synthase (Fragment)
O-methyltransferase-2
MYB transcription factor 1
Dihydroflavonol 4-reductase
Dihydroflavonol 4-reductase
Mean abundance Log 10
D
Proteins involved in terpenes biosynthesis
Proteins not selected
Jasmonic acid carboxyl methyltransferase (EC 2.1.1.141)
Mean abundance Log 10
0 200 400 600 800 1000 1200 1400 1600
Mean abundance ranking

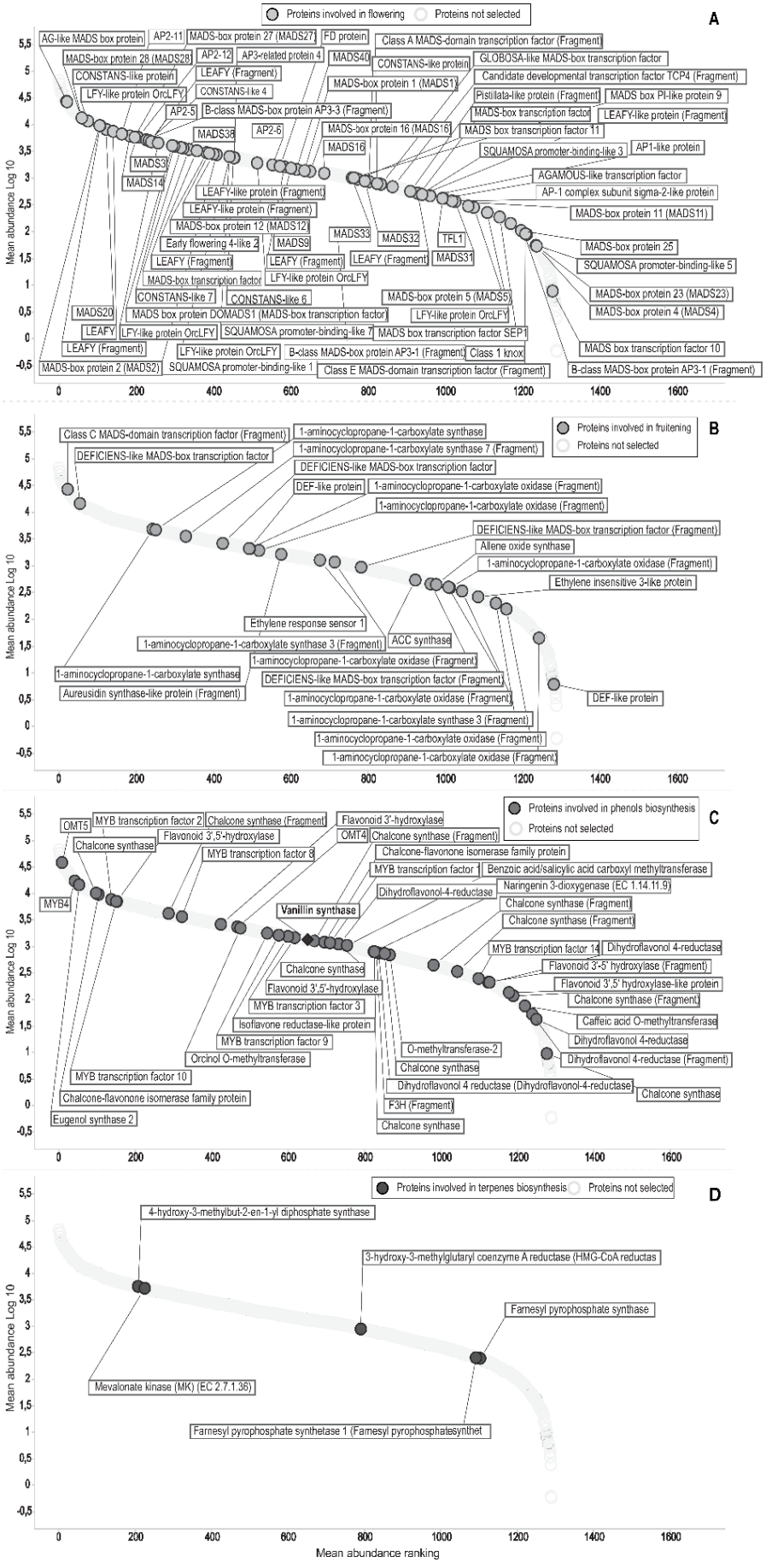
543

Supplementary Figure 3. Dynamic range of proteins identified in the dataset of the *V. bahiana*'s bean for Vb1 extraction condition. A- flowering; B- fruitening; C-phenolic and D- terpene pathway. The median absolute expression value of each protein revealing the typical S-shaped distribution over the mean abundance orders of dynamic range. Proteins with higher abundance are located in the left side of the graph and the ones with the lowest abundance are located in the right side.


A
Proteins involved in flowering
Proteins not selected
MADS box transcription factor SEP1
MADS box AP3-like protein 1
GLOBOSA-like MADS-box transcription factor
MADS-box protein 2 (MADS2)
MADS36
AP2-6
LFY-like protein OrcLFY
MADS13
MADS3
MADS9
LFY-like protein OrcLFY
SQUAMOSA promoter-binding-like 2
SQUAMOSA promoter-binding-like 3
MADS30
AP2-5
DEFICIENS-like MADS-box transcription factor
MADS38
Class C MADS-domain transcription factor (Fragment)
LEAFY-like protein (Fragment)
AP2-7
Candidate developmental transcription factor TCP4 (Fragment)
MADS1
MADS-box protein 19 (MADS19)
DEFICIENS-like MADS-box transcription factor (Fragment)
SEPALLATA-like MADS-box protein 4
LEAFY
LFY-like protein OrcLFY
CONSTANS-like 7
DEFICIENS-like MADS-box transcription factor
APETALA2-like protein
DEF-like protein
AG-like MADS box protein
MADS20
MADS41
LEAFY (Fragment)
LEAFY (Fragment)
LFY-like protein OrcLFY
MADS-box protein 23 (MADS23)
CONSTANS-like 1
MADS32
B-class MADS-box protein AP3-4 (Fragment)
MADS27
C-class MADS-box-like protein
MADS-box transcription factor
Mean abundance Log 10
B
Proteins involved in fruitening
Proteins not selected
Ethylene insensitive 3-like protein
Allene oxide synthase
1-aminocyclopropane-1-carboxylate synthase 3 (Fragment)
1-aminocyclopropane-1-carboxylic acid oxidase (Fragment)
1-aminocyclopropane-1-carboxylate synthase
1-aminocyclopropane-carboxylate synthase homologue
ACC synthase (Fragment)
ACC synthase
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
Ethylene insensitive 3-like 1
Ethylene response sensor 1
Mean abundance Log 10
C
Proteins involved in phenols biosynthesis
Proteins not selected
Acyl-coenzyme A oxidase
Phenylalanine ammonia-lyase (EC 4.3.1.24)
Chalcone synthase (EC 2.3.1.74)
Cinammate 4-hydroxylase
Dihydroflavonol 4 reductase (Dihydroflavonol-4-reductase)
MYB transcription factor 12
MYB transcription factor 16
Chalcone synthase (Fragment)
Chalcone-flavonone isomerase family protein
Chalcone synthase (Fragment)
Chalcone synthase
Flavanone-3-hydroxylase
Coumarate 3-hydroxylase
Chalcone synthase (Fragment)
Isoflavone reductase-like protein
Dihydroflavonol 4-reductase
MYB transcription factor 3
Dihydroflavonol-4-reductase
Chalcone synthase
Chalcone synthase (Fragment)
Mean abundance Log 10
D
Proteins involved in terpenes biosynthesis
Proteins not selected
2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase
1-deoxy-D-xylulose-5-phosphate synthase (Fragment)
Putative ethylene response sensor
Farnesyl pyrophosphate synthase
Mitogen-activated protein kinase (EC 2.7.11.24)
Mean abundance Log 10
0 200 400 600 800 1000 1200 1400 1600
Mean abundance ranking

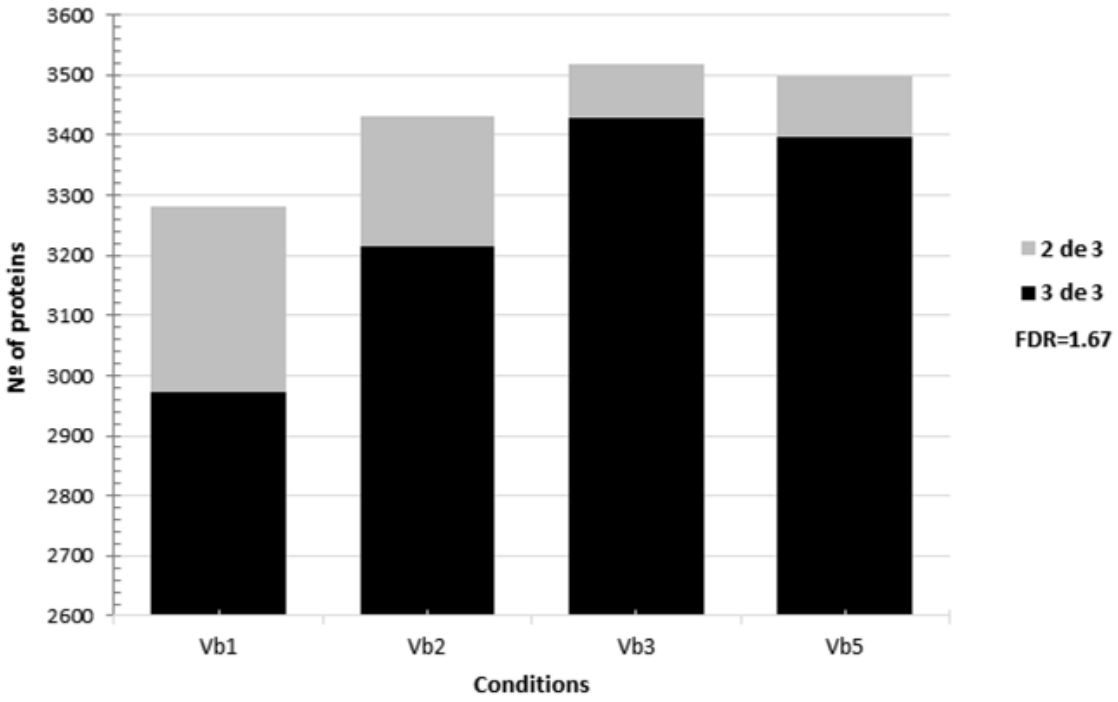
549

Supplementary Figure 4. Dynamic range of proteins identified in the dataset of the *V. bahiana*'s bean for Vb2 extraction conditionA- flowering; B- fruitening; C-phenolic and D- terpene pathway. The median absolute expression value of each protein revealing the typical S-shaped distribution over the mean abundance orders of dynamic range. Proteins with higher abundance are located in the left side of the graph and the ones with the lowest abundance are located in the right side.

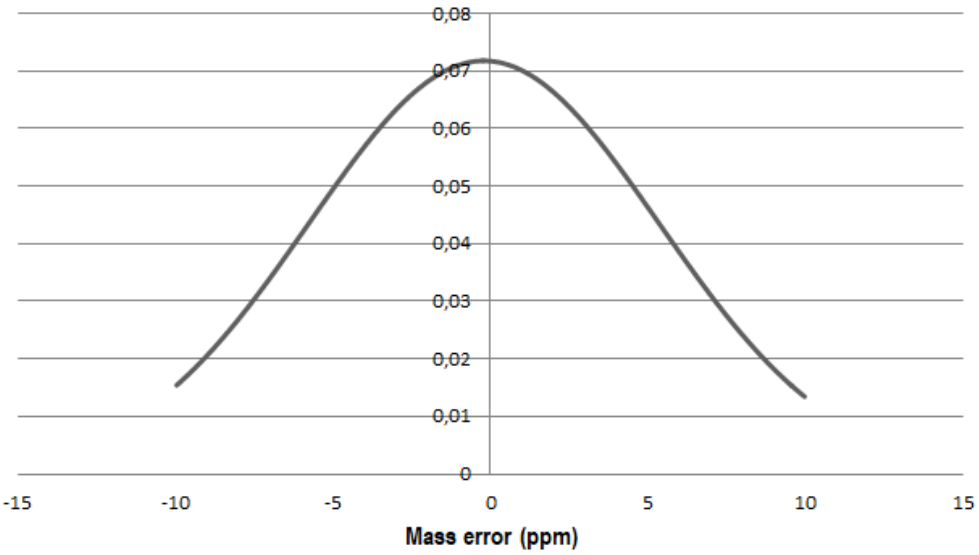

A
Proteins involved in flowering
Proteins not selected
AG-like MADS box protein
AP2-11
MADS-box protein 27 (MADS27)
FD protein
Class A MADS-domain transcription factor (Fragment)
MADS-box protein 28 (MADS28)
AP2-12
AP3-related protein 4
MADS40
CONSTANS-like protein
GLOBOSA-like MADS-box transcription factor
CONSTANS-like protein
LEAFY (Fragment)
MADS-box protein 1 (MADS1)
Candidate developmental transcription factor TCP4 (Fragment)
LFY-like protein OrcLFY
CONSTANS-like 4
Pistillata-like protein (Fragment)
MADS box PI-like protein 9
AP2-5
B-class MADS-box protein AP3-3 (Fragment)
MADS-box transcription factor
LEAFY-like protein (Fragment)
MADS38
AP2-6
MADS-box protein 16 (MADS16)
MADS box transcription factor 11
MADS16
SQUAMOSA promoter-binding-like 3
AP1-like protein
MADS3
MADS14
AGAMOUS-like transcription factor
LEAFY-like protein (Fragment)
AP-1 complex subunit sigma-2-like protein
LEAFY-like protein (Fragment)
MADS-box protein 11 (MADS11)
MADS-box protein 12 (MADS12)
MADS33
MADS32
TFL1
MADS-box protein 25
Early flowering 4-like 2
MADS9
SQUAMOSA promoter-binding-like 5
LEAFY (Fragment)
LEAFY (Fragment)
LEAFY (Fragment)
MADS31
MADS-box transcription factor
LFY-like protein OrcLFY
CONSTANS-like 7
CONSTANS-like 6
MADS-box protein 5 (MADS5)
MADS-box protein 23 (MADS23)
MADS20
MADS box protein DOMADS1 (MADS-box transcription factor)
LFY-like protein OrcLFY
MADS-box protein 4 (MADS4)
LEAFY
LFY-like protein OrcLFY
SQUAMOSA promoter-binding-like 7
MADS box transcription factor SEP1
LEAFY (Fragment)
LFY-like protein OrcLFY
B-class MADS-box protein AP3-1 (Fragment)
Class 1 knox
MADS box transcription factor 10
MADS-box protein 2 (MADS2)
SQUAMOSA promoter-binding-like 1
Class E MADS-domain transcription factor (Fragment)
B-class MADS-box protein AP3-1 (Fragment)
Mean abundance Log 10
B
Proteins involved in fruitening
Proteins not selected
Class C MADS-domain transcription factor (Fragment)
1-aminocyclopropane-1-carboxylate synthase
DEFICIENS-like MADS-box transcription factor
1-aminocyclopropane-1-carboxylate synthase 7 (Fragment)
DEFICIENS-like MADS-box transcription factor
DEF-like protein
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
DEFICIENS-like MADS-box transcription factor (Fragment)
Allene oxide synthase
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
Ethylene insensitive 3-like protein
Ethylene response sensor 1
1-aminocyclopropane-1-carboxylate synthase 3 (Fragment)
ACC synthase
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
1-aminocyclopropane-1-carboxylate synthase
DEFICIENS-like MADS-box transcription factor (Fragment)
Aureusidin synthase-like protein (Fragment)
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
DEF-like protein
1-aminocyclopropane-1-carboxylate synthase 3 (Fragment)
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
1-aminocyclopropane-1-carboxylate oxidase (Fragment)
Mean abundance Log 10
C
Proteins involved in phenols biosynthesis
Proteins not selected
OMT5
MYB transcription factor 2
Chalcone synthase (Fragment)
Flavonoid 3'-hydroxylase
Chalcone synthase
Flavonoid 3',5'-hydroxylase
OMT4
Chalcone synthase (Fragment)
MYB transcription factor 8
Chalcone-flavonone isomerase family protein
MYB transcription factor 1
Benzoic acid/salicylic acid carboxyl methyltransferase
MYB4
Vanillin synthase
Dihydroflavonol-4-reductase
Naringenin 3-dioxygenase (EC 1.14.11.9)
Chalcone synthase (Fragment)
Chalcone synthase (Fragment)
MYB transcription factor 14
Dihydroflavonol 4-reductase
Chalcone synthase
Flavonoid 3'-5' hydroxylase (Fragment)
Flavonoid 3',5'-hydroxylase
Flavonoid 3',5' hydroxylase-like protein
MYB transcription factor 3
Chalcone synthase (Fragment)
Isoflavone reductase-like protein
Caffeic acid O-methyltransferase
MYB transcription factor 9
Dihydroflavonol 4-reductase
O-methyltransferase-2
Orcinol O-methyltransferase
Dihydroflavonol 4-reductase (Fragment)
Chalcone synthase
MYB transcription factor 10
Dihydroflavonol 4 reductase (Dihydroflavonol-4-reductase)
Chalcone synthase
Chalcone-flavonone isomerase family protein
F3H (Fragment)
Eugenol synthase 2
Chalcone synthase
Mean abundance Log 10
D
Proteins involved in terpenes biosynthesis
Proteins not selected
4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase
3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA reductas
Farnesyl pyrophosphate synthase
Mevalonate kinase (MK) (EC 2.7.1.36)
Farnesyl pyrophosphate synthetase 1 (Farnesyl pyrophosphatesynthet
Mean abundance Log 10
Mean abundance ranking

555

Supplementary Figure 5. Dynamic range of proteins identified in the dataset of the *V. bahiana*'s bean for Vb3 extraction condition. A- flowering; B- fruitening; C-phenolic and D- terpene pathway. The median absolute expression value of each protein revealing the typical S-shaped distribution over the mean abundance orders of dynamic range. Proteins with higher abundance are located in the left side of the graph and the ones with the lowest abundance are located in the right side.



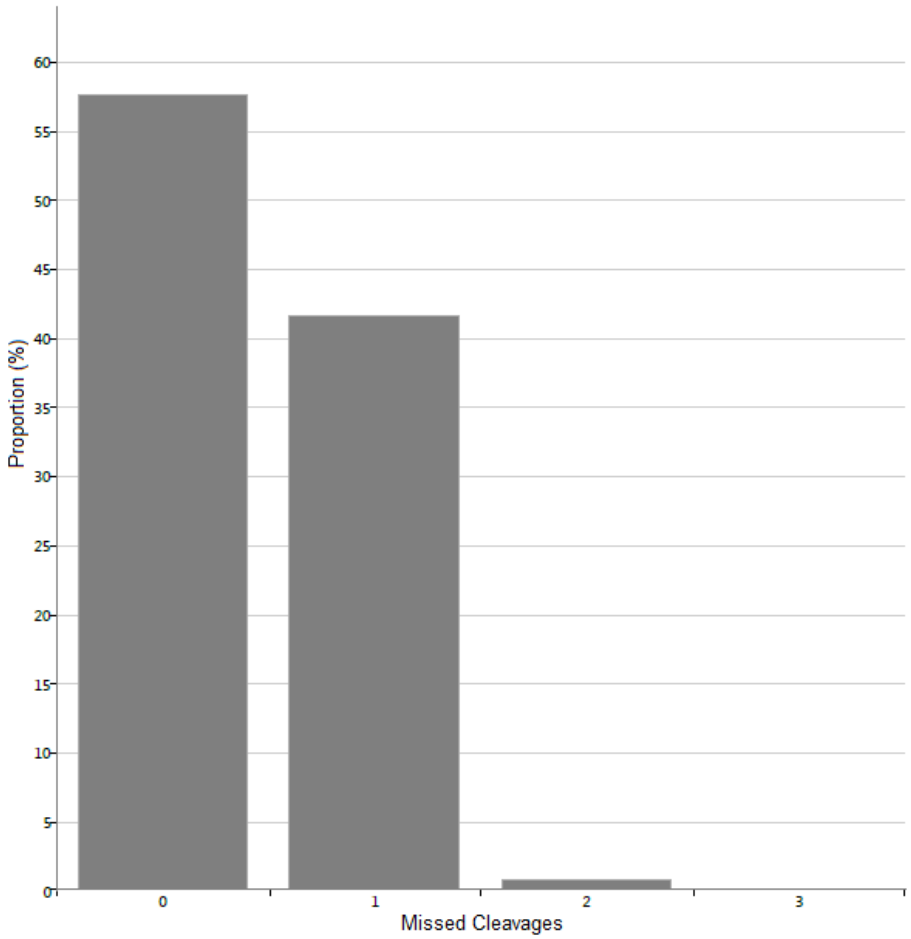
Supplementary Figure 6. Number of proteins identified present in two of the three and three of the three technical replicates, with FDR calculated. Most of the proteins were identified in the three technical replicates.



565

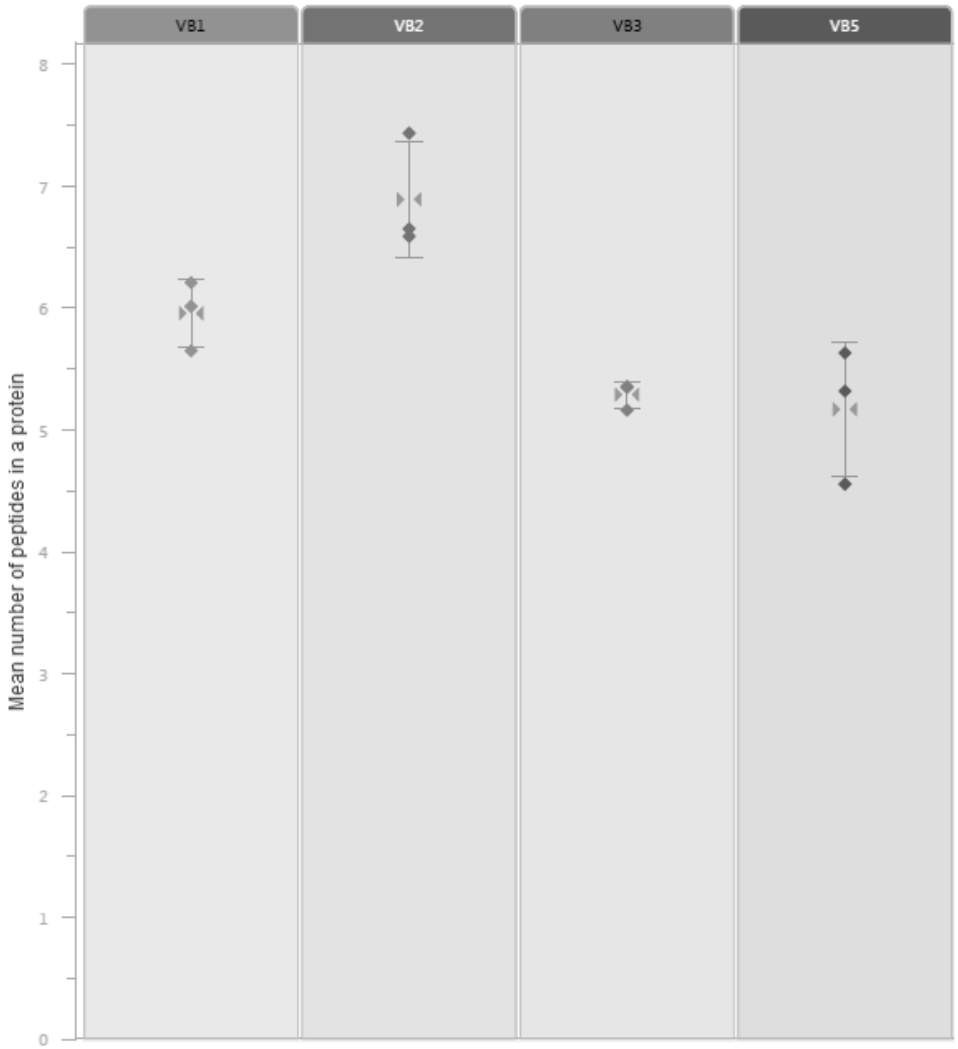
566 **Supplementary Figure 7.** Mass error distribution of *V. bahiana*. The normal distribution of mass

567 errors obeyed a normal curve.



568 **Supplementary Figure 8.** Missed cleavages of *V. bahiana* data. Approximately 60% of the missed

569 cleavages were around zero.



570

571 **Supplementary Figure 9.** Mean number of peptides/protein of *V. bahiana* data, with an average 6

572 peptides/ protein.

573 **Supplementary Table 1.** Proteins identified involved in Flowering.

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
A0A076N468	Vb2, Vb3	8	2	0.00*	Vb3	27257.2±1549.9	AG-like MADS box protein	*Oncidium hybrid cultivar*
A0A088BJD5	Vb2, Vb3	3	1	0.00*	Vb3	2958.2±251.5	LEAFY-like protein (Fragment)	*Dendrobium tosaense*
A0A088BJT1	Vb1, Vb3, Vb5	4	1	0.53	Vb1	3976.6±223.9	LEAFY-like protein (Fragment)	*Dendrobium lindleyi*
A0A088BK18	Vb3, Vb5	8	1	0.00*	Vb5	3987.7±1082.5	LEAFY-like protein (Fragment)	*Dendrobium hercoglossum*
A0A097PAK0	Vb3, Vb5	2	1	0.08	Vb5	506.8±25.1	TFL1	*Oncidium hybrid cultivar*
A0A0A0QXJ4	Vb3, Vb5	8	4	0.18	Vb5	1092.8±157.5	Class A MADS-domain transcription factor (Fragment)	*Gongora galeata*
A0A0A0R216	Vb3, Vb5	3	1	0.00*	Vb5	128033.5±10450.7	Class E MADS-domain transcription factor (Fragment)	*Vanilla planifolia* (Vanilla)
A0A0A0R219	ALL	12	2	0.00*	Vb3	27007.9±844	Class C MADS-domain transcription factor (Fragment)	*Gongora galeata*
A0A0A6Z8D3	Vb1, Vb2, Vb3	9	4	0.00*	Vb2	378765.1±69634.4	MADS box transcription factor SEP1	*Paphiopedilum concolor*
A0A0D3QNZ5	Vb3, Vb5	7	4	0.85	Vb5	14354.5±2841.6	MADS-box protein 28 (MADS28)	*Erycina pusilla*
A0A0D3RI62	Vb3, Vb5	7	3	0.00*	Vb3	3293.2±1129.2	Early flowering 4-like 2	*Doritaenopsis hybrid cultivar*
A0A0F6PVI7	Vb3, Vb5	4	1	0.01*	Vb3	931.3±152.2	DEFICIENS-like MADS-box transcription factor (Fragment)	*Paphiopedilum sp. 2 XQW-2015*

Supplementary Table 1. Proteins identified involved in Flowering (continue).

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
A0A0F6PVV9	Vb2, Vb3, Vb5	9	1	0.00*	Vb5	1099.9±183.3	DEFICIENS-like MADS-box transcription factor (Fragment)	*Paphiopedilum hangianum*
A0A0N7AL13	Vb3, Vb5	6	3	0.00*	Vb5	5076.7±659.6	FD protein	*Phalaenopsis aphrodite subsp. formosana*
A0A0U1VD58	Vb1, Vb5	5	1	0.00*	Vb5	9791.1±430.7	MADS14	*Erycina pusilla*
A0A142L043	Vb1, Vb2	15	9	1.00	Vb1	1288±425.1	APETALA2-like protein	*Cattleya trianae*
A0A1L1WKW6	Vb2, Vb3, Vb5	2	2	1.00	Vb3	3617.9±1311.9	MADS38	*Erycina pusilla*
A0A1L1WKX4	Vb1, Vb5	4	3	1.00	Vb5	20417.2±1862	MADS41	*Erycina pusilla*
A0A1L1WKY1	Vb3, Vb5	2	1	1.00	Vb5	1744.7±321.1	MADS40	*Erycina pusilla*
A0A1L1WKY3	Vb2, Vb3, Vb5	10	5	1.00	Vb3	5900.8±1729.9	MADS3	*Apostasia odorata*
A0A1L1WKY5	Vb2, Vb5	2	1	1.00	Vb5	4674.9±1642.2	MADS1	*Apostasia odorata*
A0A1L1WKZ4	Vb3, Vb5	7	4	1.00	Vb3	6243.8±1130.5	MADS14	*Apostasia odorata*
A0A1L1WKZ5	Vb3, Vb5	6	2	1.00	Vb5	3668.6±1305.2	MADS16	*Apostasia odorata*
A0A1L1WKZ7	Vb2, Vb5	15	5	1.00	Vb2	3666.6±1131.8	MADS13	*Apostasia odorata*

Supplementary Table 1. *Proteins identified involved in Flowering (continue).*

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
A0A1L1WL06	Vb2, Vb3	9	7	1.00	Vb2	2553.7±364.7	MADS9	*Apostasia odorata*
A0A1L1WL08	Vb2, Vb3, Vb5	5	2	1.00	Vb3	8221.4±2983.4	MADS20	*Apostasia odorata*
A0A1L1WL12	Vb3, Vb5	9	2	1.00	Vb5	1056.1±72.9	MADS31	*Apostasia odorata*
A0A1L1WL16	Vb3, Vb5	3	1	1.00	Vb5	4704.5±1289.3	MADS33	*Apostasia odorata*
A0A1L1WL19	Vb1, Vb2	5	2	1.00	Vb2	17151.1±6587.2	MADS36	*Apostasia odorata*
A0A1L1WL24	Vb2, Vb5	7	4	1.00	Vb5	40581.2±7565.8	MADS30	*Apostasia odorata*
A0A1L1WL26	Vb2, Vb5	2	2	1.00	Vb5	14416.7±1726.4	MADS41	*Apostasia odorata*
A0A1L1WP45	Vb2, Vb3	1	1	1.00	Vb3	882.6±194.4	MADS32	*Apostasia odorata*
A0A1L1WSB1	Vb2, Vb5	4	1	1.00	Vb5	310.3±22.9	MADS27	*Apostasia odorata*
A5I851	Vb1, Vb5	3	1	1.00	Vb1	1404.7±242	Leafy-floricaula protein (Fragment)	*Ophrys creticola*
B1WAN2	Vb2, Vb3, Vb5	12	1	1.00	Vb5	12528.9±1070.5	LEAFY	*Phalaenopsis hybrid cultivar*
B6ZDS5	Vb2, Vb3, Vb5	9	1	1.00	Vb5	10901.9±1999.8	MADS-box transcription factor	*Habenaria radiata*

Supplementary Table 1. *Proteins identified involved in Flowering (continue).*

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
B7T4F3	Vb2, Vb5	16	2	1.00	Vb2	5576.7±1498.2	MADS box AP3-like protein 1	*Dendrobium hybrid cultivar*
B9W016	Vb2, Vb5	5	1	1.00	Vb5	20750.2±5824.2	LEAFY (Fragment)	*Oncidium hybrid cultivar*
C5I9R2	Vb2, Vb3, Vb5	6	3	1.00	Vb5	5991.5±1035.8	DEFICIENS-like MADS-box transcription factor	*Gongora galeata*
C5I9S1	Vb1, Vb2, Vb5	8	1	1.00	Vb5	3932.8±164.3	DEFICIENS-like MADS-box transcription factor	*Phragmipedium longifolium*
C5I9S6	Vb3, Vb5	7	1	1.00	Vb3	14495.1±2651.6	DEFICIENS-like MADS-box transcription factor	*Spiranthes odorata*
C5I9S8	Vb2, Vb3	10	1	1.00	Vb2	33261.1±12474.6	GLOBOSA-like MADS-box transcription factor	*Spiranthes odorata*
D4N890	Vb3, Vb5	3	1	1.00	Vb3	2451.1±887.5	AP3-related protein 4	*Dendrobium moniliforme (Epidendrum moniliforme)*
D9IFM4	Vb3, Vb5	3	1	1.00	Vb5	963.3±156	MADS box transcription factor 10	*Oncidium hybrid cultivar*
D9IFM5	Vb3, Vb5	5	2	1.00	Vb5	28379.1±4195.5	MADS box transcription factor 11	*Oncidium hybrid cultivar*
E6Y7M5	Vb1, Vb3, Vb5	4	2	1.00	Vb3	13624.3±2111.2	CONSTANS-like protein	*Phalaenopsis hybrid cultivar*
E7BUG5	Vb1, Vb3,Vb5	2	2	1.00	Vb5	1305.7±410.7	CONSTANS-like protein	*Dendrobium loddigesii (Callista loddigesii)*

Supplementary Table 1. *Proteins identified involved in Flowering (continue).*

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
F5ANW4	Vb2, Vb3, Vb5	5	2	1.00	Vb5	4192.8±1482.2	Candidate developmental transcription factor TCP4 (Fragment)	*Phalaenopsis hybrid cultivar*
G2XK84	Vb3, Vb5	12	1	1.00	Vb3	3873.9±870.8	B-class MADS-box protein PI-2 (Fragment)	*Anoectochilus formosanus*
G2XK93	Vb3, Vb5	7	3	1.00	Vb3	27545.4±84.7	B-class MADS-box protein AP3-1 (Fragment)	*Galeola falconeri*
G2XKA8	Vb2, Vb5	8	4	1.00	Vb2	3790.1±1477.3	B-class MADS-box protein AP3-4 (Fragment)	*Oncidium hybrid cultivar*
G2XKB0	Vb3, Vb5	9	2	1.00	Vb5	282±62.1	B-class MADS-box protein AP3-1 (Fragment)	*Paphiopedilum hybrid cultivar*
G2XKB2	Vb3, Vb5	9	5	1.00	Vb3	5207.7±787.2	B-class MADS-box protein AP3-3 (Fragment)	*Paphiopedilum hybrid cultivar*
H2ERP3	Vb3, Vb5	4	2	1.00	Vb5	1746.3±239.7	AP1-like protein	*Cymbidium faberi*
H6U642	Vb1, Vb5	8	2	1.00	Vb5	4952.8±1323.9	APETALA3-like MADS-box protein	*Cymbidium ensifolium (Epidendrum ensifolium)*
I6Q0H7	ALL	6	1	1.00	Vb3	3498.3±543.3	LEAFY (Fragment)	*Apostasia sp. G244*
I6Q0K8	Vb3, Vb5	5	1	1.00	Vb5	540.1±159.7	LEAFY (Fragment)	*Cypripedium margaritaceum*
I6Q108	Vb3, Vb5	6	1	1.00	Vb3	4934.8±1617.6	LEAFY (Fragment)	*Paphiopedilum wardii*
I6Q436	Vb3, Vb5	4	1	1.00	Vb5	5715.1±408.6	LEAFY (Fragment)	*Neuwiedia zollingeri var. singapureana*
I6Q4J0	Vb3, Vb5	3	1	1.00	Vb3	9538.5±923.9	LEAFY (Fragment)	*Cypripedium margaritaceum*

Supplementary Table 1. *Proteins identified involved in Flowering (continue).*

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
J9Z4N7	Vb1, Vb3	2	2	1.00	Vb3	378.3±50.7	AP-1 complex subunit sigma-2-like protein	*Dendrobium catenatum*
K4JB92	Vb1, Vb2, Vb3	8	1	1.00	Vb2	6070±1393.7	C-class MADS-box-like protein	*Orchis italic*
M9QR13	Vb1, Vb5	4	2	1.00	Vb5	194±60.9	CONSTANS-like 3	*Erycina pusilla*
M9QR29	ALL	11	7	1.00	Vb1	34051.4±1674.5	AP2-6	*Erycina pusilla*
M9QR34	Vb3, Vb5	17	6	1.00	Vb5	15803.4±1174.5	AP2-11	*Erycina pusilla*
M9QS80	Vb2, Vb5	1	1	1.00	Vb2	186.6±20.3	CONSTANS-like 1	*Erycina pusilla*
M9QS85	Vb3, Vb5	1	1	1.00	Vb5	6283.9±349.6	CONSTANS-like 6	*Erycina pusilla*
M9QSA7	Vb2, Vb5	7	5	1.00	Vb5	1358±401.7	SQUAMOSA promoter-binding-like 2	*Erycina pusilla*
M9QSB4	Vb3, Vb5	13	10	1.00	Vb5	16606.8±1171.5	SQUAMOSA promoter-binding-like 7	*Erycina pusilla*
M9QTP3	Vb2, Vb3	2	1	1.00	Vb3	4014.8±413.1	CONSTANS-like 7	*Erycina pusilla*
M9QTQ4	Vb2, Vb3	3	2	1.00	Vb2	11502.9±4200.5	AP2-5	*Erycina pusilla*
M9QTR6	ALL	5	4	1.00	Vb1	3941.2±506.1	SQUAMOSA promoter-binding-like 3	*Erycina pusilla*
M9QXL3	Vb1, Vb3	1	1	1.00	Vb3	2767.3±962.6	SQUAMOSA promoter-binding-like 1	*Erycina pusilla*
M9QZ49	Vb3,Vb5	1	1	1.00	Vb3	5059±561	CONSTANS-like 4	*Erycina pusilla*

Supplementary Table 1. *Proteins identified involved in Flowering (continue).*

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
M9QZ62	Vb1, Vb2	7	3	1.00	Vb1	2305.4±608.4	AP2-7	*Erycina pusilla*
M9QZ67	Vb1, Vb3, Vb5	5	3	1.00	Vb3	5806.2±1736.5	AP2-12	*Erycina pusilla*
M9QZ71	Vb3, Vb5	3	2	1.00	Vb5	1702.4±172.8	SQUAMOSA promoter-binding-like 5	*Erycina pusilla*
Q0KKF5	Vb3, Vb5	4	1	1.00	Vb3	575.2±195.6	Pistillata-like protein (Fragment)	*Anacamptis morio*
Q2IA03	Vb3, Vb5	3	2	1.00	Vb3	420.9±107.7	AGAMOUS-like transcription factor	*Dendrobium crumenatum (Tropical pigeon orchid)*
Q539E6	Vb3, Vb5	5	1	1.00	Vb5	53113.6±18262.4	Class 1 knox	*Dendrobium nobile*
Q5XM80	Vb3, Vb5	8	1	1.00	Vb3	1013±75.7	MADS box PI-like protein 9	*Phalaenopsis hybrid cultivar*
Q84KZ6	Vb2, Vb3	12	1	1.00	Vb3	4629.4±1028.9	LFY-like protein OrcLFY	*Dactylorhiza romana* (Orchis romana)
Q84KZ9	Vb3, Vb5	13	2	1.00	Vb3	231.7±47.9	LFY-like protein OrcLFY	*Anacamptis laxiflora*
Q84L00	Vb2, Vb5	8	3	1.00	Vb5	10913.3±2153.4	LFY-like protein OrcLFY	*Anacamptis pyramidalis*
Q84L02	Vb3, Vb5	7	1	1.00	Vb3	11878.1±1830.8	LFY-like protein OrcLFY	*Neotinea maculate*
Q84L04	ALL	8	1	1.00	Vb5	12064.5±3322.7	LFY-like protein OrcLFY	*Ophrys tenthredinifera*
Q84L07	Vb2,Vb3	8	3	1.00	Vb3	1955.4±340.4	LFY-like protein OrcLFY	*Orchis anthropophora*

Supplementary Table 1. *Proteins identified involved in Flowering (continue).*

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
Q8RVW5	Vb1, Vb3, Vb5	11	4	1.00	Vb3	3733.8±534.5	MADS-box transcription factor	*Phalaenopsis equestris*
Q9XEK0	Vb1, Vb3	7	5	0.00*	Vb3	4074.7±1006.2	MADS box protein DOMADS1 (MADS-box transcription factor)	*Dendrobium grex Madame Thong-In*
U3U8P9	Vb1, Vb3	7	6	0.02*	Vb3	2134.5±807.1	DEF-like protein	*Orchis italic*
U3U9X3	Vb1, Vb2, Vb3	11	2	0.00*	Vb1	3413.3±916.1	DEF-like protein	*Orchis italic*
X2F442	Vb2, Vb3, Vb5	10	9	0.00*	Vb3	9552.7±2651.2	MADS-box protein 2 (MADS2)	*Erycina pusilla*
X2F450	Vb3, Vb5	6	1	0.06	Vb3	3000.6±618	MADS-box protein 12 (MADS12)	*Erycina pusilla*
X2F460	Vb3, Vb5	2	1	0.00*	Vb3	5941±1554.8	MADS-box protein 27 (MADS27)	*Erycina pusilla*
X2F5E8	Vb2, Vb3, Vb5	4	1	0.00*	Vb5	1266.3±260.2	MADS-box protein 23 (MADS23)	*Erycina pusilla*
X2F977	Vb3, Vb5	9	2	0.22	Vb5	1910.7±503.3	MADS-box protein 1 (MADS1)	*Erycina pusilla*
X2F984	Vb1, Vb3, Vb5	6	3	0.00*	Vb5	523.6±138.4	MADS-box protein 11 (MADS11)	*Erycina pusilla*
X2F988	Vb3, Vb5	9	1	0.97	Vb3	1343.5±471.7	MADS-box protein 16 (MADS16)	*Erycina pusilla*

Supplementary Table 1. *Proteins identified involved in Flowering (continue).*

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
X2FB19	Vb3, Vb5	7	2	0.00*	Vb5	847.3±192.4	MADS-box protein 5 (MADS5)	*Erycina pusilla*
X2FB33	Vb3, Vb5	5	3	0.00*	Vb5	22438.5±4680	MADS-box protein 25	*Erycina pusilla*
X2FD75	Vb1, Vb3, Vb5	7	6	0.00*	Vb1	17207±2413.2	MADS-box protein 4 (MADS4)	*Erycina pusilla*
X2FD86	Vb2, Vb5	5	1	0.32	Vb2	451.1±86.6	MADS-box protein 19 (MADS19)	*Erycina pusilla*
X5DD29	Vb2, Vb5	4	1	0.01*	Vb5	1913.6±118.5	SEPALLATA-like MADS-box protein 4	*Phalaenopsis equestris*

A extraction condition/**B** peptide count/**C** unique peptide/**D** data analyzes was performed using ANOVA ($p < 0.05$) followed by the Tukey significance test, were asterisk (*) denote statistical significance difference. /**E** mean value of the highest condition ± Standard deviation value.

Supplementary Table 2. *Proteins identified involved in Fruiting.*

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
A0A068FJK4	Vb3, Vb5	4	3	0.51	Vb5	4415.5±1526.5	1-aminocyclopropane-1-carboxylate synthase 7 (Fragment)	*Oncidium hybrid cultivar*
A0A0A0R219	ALL	12	2	0.00*	Vb3	27007.9±844	Class C MADS-domain transcription factor (Fragment)	*Gongora galeata*
A0A0A7KL66	Vb1, Vb5	2	1	0.00*	Vb5	5530.8±1025.4	Pectinesterase (Fragment)	*Dendrobium hybrid cultivar*
A0A0F6PUR0	Vb1, Vb3, Vb5	2	1	0.00*	Vb5	1841.9±611.2	1-aminocyclopropane-1-carboxylate oxidase (Fragment)	*Paphiopedilum helenae*

Supplementary Table 2. *Proteins identified involved in Fruiting (continue)*

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
A0A0F6PUT3	Vb3, Vb5	6	2	0.00*	Vb5	4138.8±442.4	1-aminocyclopropane-1-carboxylate oxidase (Fragment)	*Paphiopedilum sugiyamanum*
A0A0F6PVY2	Vb3, Vb5	6	1	0.15	Vb5	3903.6±1365	1-aminocyclopropane-1-carboxylate oxidase (Fragment)	*Paphiopedilum victoria-mariae*
A0A0F6PXG1	Vb3, Vb5	7	2	0.57	Vb5	2070.5±194.8	1-aminocyclopropane-1-carboxylate oxidase (Fragment)	*Paphiopedilum dayanum*
A0A0U1TTF5	Vb3, Vb5	9	2	0.26	Vb3	1651.2±323.8	Aureusidin synthase-like protein (Fragment)	*Oncidium hybrid cultivar*
A0A1D8DGS6	Vb1, Vb2	8	5	1.00	Vb1	3981.3±1367.1	Ethylene insensitive 3-like 1	*Phalaenopsis aphrodite subsp. formosana*
A0JBY6	Vb2, Vb5	7	3	1.00	Vb5	716.9±190.4	ACC synthase (Fragment)	*Cymbidium hybrid cultivar*
A4UTQ8	Vb2, Vb3, Vb5	9	1	1.00	Vb3	4833.5±482	1-aminocyclopropane-1-carboxylate synthase	*Dendrobium hybrid cultivar*
A8DC72	Vb1, Vb2	12	1	1.00	Vb2	4054.3±795.3	1-aminocyclopropane-1-carboxylic acid oxidase (Fragment)	*Dendrobium hybrid cultivar*
B0FYC7	Vb3, Vb5	14	1	1.00	Vb5	3113.3±309.6	1-aminocyclopropane-1-carboxylate oxidase (Fragment)	*x Brassolaeliocattleya 'Sung Ya Green'*
C0KXG8	Vb2, Vb3	11	1	1.00	Vb3	1271.8±483.9	Ethylene response sensor 1	*Dendrobium hybrid cultivar*
E2I9R2	Vb2, Vb3, Vb5	6	1	1.00	Vb2	5880.1±1578.9	1-aminocyclopropane-1-carboxylate synthase 3 (Fragment)	*Dendrobium hybrid cultivar*
G0Y289	Vb2, Vb3, Vb5	14	4	1.00	Vb2	12136.8±3913	Ethylene insensitive 3-like protein	*Oncidium hybrid cultivar*

Supplementary Table 2. *Proteins identified involved in Fruiting (continue)*

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
I0DHH0	Vb2, Vb3, Vb5	16	7	1.00	Vb2	9944.2±2500.3	Allene oxide synthase	*Cymbidium ensifolium (Epidendrum ensifolium)*
I6Q0N0	Vb2, Vb5	6	1	1.00	Vb5	253.4±94.3	1-aminocyclopropane-1-carboxylate oxidase (Fragment)	*Cypripedium acaule* (Pink lady's slipper orchid)
I6Q2U8	Vb3, Vb5	9	2	1.00	Vb5	1623.5±393.4	1-aminocyclopropane-1-carboxylate oxidase (Fragment)	*Cypripedium japonicum*
I6Q2V1	Vb3, Vb5	5	1	1.00	Vb5	253.9±86.9	1-aminocyclopropane-1-carboxylate oxidase (Fragment)	*Vanilla sp. G241*
Q6RCS3	Vb1, Vb3, Vb5	3	1	1.00	Vb3	4695.7±468.5	1-aminocyclopropane-1-carboxylate synthase	*Cattleya bicolor*
Q96413	Vb2, Vb3, Vb5	6	2	1.00	Vb2	8882.4±2734.8	ACC synthase	*Dendrobium crumenatum (*Tropical pigeon orchid)
Q9ZSD2	Vb2, Vb5	16	3	1.00	Vb2	2202.6±874	Putative ethylene response sensor	*Phalaenopsis hybrid cultivar*
Q9XGG3	Vb1, Vb2	6	5	0.00*	Vb2	1893.5±388.6	1-aminocyclopropane-carboxylate synthase homologue	*Phalaenopsis hybrid cultivar*
U5XI27	Vb3, Vb5	6	4	0.00*	Vb5	17341.8±5800.3	1-aminocyclopropane-1-carboxylate synthase 3 (Fragment)	*Oncidium hybrid cultivar*
A0A0F6PVI7	Vb3, Vb5	4	1	0.01*	Vb3	931.3±152.2	DEFICIENS-like MADS-box transcription factor (Fragment)	*Paphiopedilum sp. 2 XQW-2015*
A0A0F6PVV9	Vb2, Vb3, Vb5	9	1	0.00*	Vb5	1099.9±183.3	DEFICIENS-like MADS-box transcription factor (Fragment)	*Paphiopedilum hangianum*

Supplementary Table 2. *Proteins identified involved in Fruiting (continue)*

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
C5I9R2	Vb2, Vb3, Vb5	6	3	1.00	Vb5	5991.5±1035.8	DEFICIENS-like MADS-box transcription factor	*Gongora galeata*
C5I9S1	Vb1, Vb2, Vb5	8	1	1.00	Vb5	3932.8±164.3	DEFICIENS-like MADS-box transcription factor	*Phragmipedium longifolium*
C5I9S6	Vb3, Vb5	7	1	1.00	Vb3	14495.1±2651.6	DEFICIENS-like MADS-box transcription factor	*Spiranthes odorata*
U3U8P9	Vb1, Vb3	7	6	0.02*	Vb3	2134.5±807.1	DEF-like protein	*Orchis italica*
U3U9X3	Vb1, Vb2, Vb3	11	2	0.00*	Vb1	3413.3±916.1	DEF-like protein	*Orchis italica*

A extraction condition/**B** peptide count/**C** unique peptide/**D** data analyzes was performed using ANOVA ($p < 0.05$) followed by the Tukey significance test, were asterisk (*) denote statistical significance difference. /**E** mean value of the highest condition ± Standard deviation value.

Supplementary Table 3. *Proteins identified involved in phenols biosynthesis.*

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
A0A068ERY8	ALL	3	3	0.00*	Vb2	7072.7±2028.3	Isoflavone reductase-like protein	*Dendrobium catenatum*
A0A075E375	Vb3, Vb5	12	10	0.00*	Vb3	2234.8±62.8	Orcinol O-methyltransferase	*Vanda hybrid cultivar*
A0A075E3N4	Vb3, Vb5	9	5	0.05*	Vb5	1966.6±275.2	Benzoic acid/salicylic acid carboxyl methyltransferase	*Vanda hybrid cultivar*

Supplementary Table 34. Proteins identified involved in phenols biosynthesis (continue)

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
A0A088G538	Vb3, Vb5	7	1	0.00*	Vb5	8631.2±739.5	Chalcone synthase	*Cymbidium hybrid cultivar*
A0A088G4R1	Vb5	13	11	0.77	Vb5	1604.6±212.9	Flavonol synthase	*Cymbidium hybrid cultivar*
A0A096ZX30	Vb1, Vb3, Vb5	1	1	0.00*	Vb5	3775.8±458.5	MYB transcription factor 1	*Phalaenopsis equestris*
A0A096ZX41	Vb1, Vb3, Vb6	8	1	0.00*	Vb1	888±250.8	Dihydroflavonol 4-reductase	*Phalaenopsis equestris*
A0A096ZX44	Vb2, Vb5	5	2	0.00*	Vb2	2911.2±794.2	MYB transcription factor 16	*Phalaenopsis equestris*
A0A096ZX46	Vb3, Vb5	4	1	0.03*	Vb3	7696.8±916.1	MYB transcription factor 2	*Phalaenopsis equestris*
A0A096ZX55	Vb2, Vb5	11	7	0.48	Vb5	3635.1±595.6	MYB transcription factor 12	*Phalaenopsis equestris*
A0A096ZX66	Vb2, Vb3	7	1	0.09	Vb3	789.3±246.8	Chalcone synthase	*Phalaenopsis equestris*
A0A096ZX69	Vb3, Vb5	9	4	0.00*	Vb3	1809±287.4	MYB transcription factor 9	*Phalaenopsis equestris*
A0A096ZX71	Vb3, Vb5	16	13	0.38	Vb5	3295.2±757.8	Flavonoid 3'-hydroxylase	*Phalaenopsis equestris*

A0A096ZX72	Vb2, Vb3	2	2	0.00*	Vb2	8324.9±1929.2	MYB transcription factor 3	*Phalaenopsis equestris*

Supplementary Table 35. Proteins identified involved in phenols biosynthesis (continue)

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
A0A096ZX73	Vb3, Vb5	2	1	0.00*	Vb3	251.9±96.9	MYB transcription factor 14	*Phalaenopsis equestris*
A0A096ZX74	Vb3, Vb5	8	4	0.00*	Vb3	3618.3±123.2	MYB transcription factor 8	*Phalaenopsis equestris*
A0A096ZX77	Vb3, Vb5	10	3	0.00*	Vb3	7153±883.6	MYB transcription factor 10	*Phalaenopsis equestris*
A0A0A6Z8F0	Vb1, Vb2, Vb5	5	1	0.00*	Vb5	4485.6±315.9	Chalcone synthase (Fragment)	*Paphiopedilum armeniacum (*Golden slipper orchid)
A0A0A6Z8F3	Vb1, Vb5	6	2	0.02*	Vb1	3134.6±1108.6	Chalcone synthase (Fragment)	*Paphiopedilum purpuratum*
A0A0A6Z8K1	Vb3, Vb5	3	1	0.00*	Vb5	2158.1±813.3	Chalcone synthase (Fragment)	*Paphiopedilum micranthum*
A0A0A6Z8P9	ALL	10	5	0.00*	Vb1	29404.1±9770.7	Chalcone synthase (Fragment)	*Paphiopedilum x areeanum*
A0A0E3NBN0	Vb3, Vb5	8	1	0.04*	Vb3	14543.7±5572.9	Eugenol synthase 2	*Gymnadenia odoratissima* (Orchis odoratissima)
A0A0F7G352	Vb3, Vb5	11	10	0.00*	Vb3	1344.9±177.9	Vanillin synthase	*Vanilla planifolia* (Vanilla)

Supplementary Table 3. Proteins identified involved in phenols biosynthesis (continue)

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
A0A0M3U4G0	Vb1, Vb2, Vb5	12	2	0.00*	Vb2	13956.5±3504	Chalcone synthase	*Dendrobium catenatum*
A0A0P0DLN5	Vb2, Vb5	26	13	0.04*	Vb2	2345.8±785.5	Coumarate 3-hydroxylase	*Dendrobium catenatum*
A0A0U1TTF5	Vb3, Vb5	9	2	0.26	Vb3	1651.2±323.8	Aureusidin synthase-like protein (Fragment)	*Oncidium hybrid cultivar*
A0A0U1VL99	Vb2, Vb5	8	2	0.021*	Vb5	654.5±151.2	Flavanone-3-hydroxylase	*Phalaenopsis hybrid cultivar*
A0A0U1VLB8	Vb1, Vb5	14	1	0.00*	Vb5	7764±1098.1	Dihydroflavonol 4-reductase	*Phalaenopsis hybrid cultivar*
A0A126QG95	Vb3, Vb5	3	2	1.00	Vb5	491.9±87.9	Dihydroflavonol 4-reductase (Fragment)	*Dendrobium catenatum*
A0A142IGA4	Vb1, Vb5	10	8	1.00	Vb1	27916.9±6181.9	Glutamate-cysteine ligase (Fragment)	*Oncidium hybrid cultivar*
A0A191XZN5	Vb3, Vb5	6	5	1.00	Vb5	30827±3535	Flavonoid 3'-5' hydroxylase (Fragment)	*Phalaenopsis lueddemanniana*
A0A1I9K5X5	Vb5	5	2	1.00	Vb5	2386.7±447.9	4-coumarate.CoA ligase	*Dendrobium catenatum*
A7KTI4	Vb2, Vb3	5	3	1.00	Vb2	1778.5±421.7	Chalcone-flavonone isomerase family protein	*Oncidium hybrid cultivar*

Supplementary Table 36. Proteins identified involved in phenols biosynthesis (continue)

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
A7UCI5	Vb3, Vb5	3	1	1.00	Vb5	2027.5±234.9	Chalcone synthase (Fragment)	*Cymbidium floribundum*
A8QME9	ALL	5	1	1.00	Vb2	2502.3±556.8	Chalcone synthase (Fragment)	*Dactylorhiza viridis*
A8QMI6	Vb1, Vb3, Vb5	4	1	1.00	Vb1	2661.1±1009.7	Chalcone synthase (Fragment)	*Gymnadenia borealis*
A8QMJ9	Vb2, Vb5	4	1	1.00	Vb5	12221.1±306	Chalcone synthase (Fragment)	*Traunsteinera globosa*
B2LUN8	Vb3, Vb5	2	2	1.00	Vb5	3039.3±204.9	Flavonoid 3',5' hydroxylase-like protein	*Vanda coerulea*
B5MBV6	Vb2, Vb3, Vb5	9	1	1.00	Vb3	1109.3±80.5	Dihydroflavonol-4-reductase	*Dendrobium hybrid cultivar*
F2YP45	Vb3, Vb5	6	6	1.00	Vb5	10686.6±3023.3	OMT4	*Vanilla planifolia* (Vanilla)
F2YP46	Vb3, Vb5	4	2	1.00	Vb3	38082.6±3676.6	OMT5	*Vanilla planifolia* (Vanilla)
F5A635	Vb1, Vb2, Vb3	12	1	1.00	Vb2	8543.1±2144.9	Chalcone synthase	*Dendrobium moniliforme (Epidendrum moniliforme)*

Supplementary Table 37. Proteins identified involved in phenols biosynthesis (continue)

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
F5A637	Vb1, Vb3, Vb5	8	2	1.00	Vb3	7118.7±1387.1	Flavonoid 3',5'-hydroxylase	*Dendrobium moniliforme (Epidendrum moniliforme)*
H9TB32	Vb3, Vb5	1	1	1.00	Vb5	1161.9±133.2	F3H (Fragment)	*Oncidium hybrid cultivar*
L7SSS6	Vb5	21	12	1.00	Vb5	2354.1±370.6	Phenylalanine ammonia-lyase	*Dendrobium candidum*
O64902	Vb2, Vb3, Vb5	9	5	1.00	Vb2	3200±1047.3	Dihydroflavonol-4-reductase	*Cymbidium hybrid cultivar*
P93482	Vb5	25	2	1.00	Vb5	2671.6±569.1	Acyl-coenzyme A oxidase	*Phalaenopsis hybrid cultivar*
Q27I63	Vb1, Vb3, Vb5	5	1	1.00	Vb3	698±110.8	O-methyltransferase -2	*Vanilla planifolia (Vanilla)*
Q42609	Vb2, Vb5	20	8	1.00	Vb2	15520.4±5966.6	Phenylalanine ammonia-lyase	*Bromheadia finlaysoniana*
Q43741	Vb3, Vb5	10	5	1.00	Vb5	2148.7±609.4	Naringenin 3-dioxygenase	*Bromheadia finlaysoniana*
Q45RS8	Vb3, Vb5	5	2	1.00	Vb5	20.3±1.4	Chalcone synthase	*Oncidium hybrid cultivar*

Supplementary Table 3. Proteins identified involved in phenols biosynthesis (continue)

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
Q4VFZ0	Vb1, Vb3	10	1	1.00	Vb3	42.3±12.9	Dihydroflavonol 4-reductase	*Oncidium hybrid cultivar*
Q6LC98	Vb2, Vb5	29	4	1.00	Vb2	32996.8±8813.7	Acyl-coenzyme A oxidase	*Phalaenopsis hybrid cultivar*
Q6Q796	Vb1, Vb3	12	7	1.00	Vb1	1887.1±519.4	Caffeic acid O-methyltransferase	*Vanilla planifolia* (Vanilla)
Q84U50	Vb3, Vb5	3	1	1.00	Vb3	17224.2±6062.4	MYB4	*Dendrobium sp. XMW-2002-4*
R9UNP6	Vb2, Vb5	9	8	0.00*	Vb2	5430.6±1336.5	Cinammate 4-hydroxylase	*Dendrobium catenatum*
T1Q041	Vb3, Vb5	8	2	0.00*	Vb3	10219.5±1670.8	Chalcone synthase	*Paphiopedilum concolor*
V9PCW5	Vb3, Vb5	3	3	0.01*	Vb3	9643.4±1388	Chalcone-flavonone isomerase family protein	*Dendrobium hybrid cultivar*
V9PCX2	Vb3, Vb5	15	7	0.00*	Vb5	2547.6±329	Flavonoid 3',5'-hydroxylase	*Dendrobium hybrid cultivar*
V9PCX9	Vb2, Vb5	6	1	0.00*	Vb5	3481.6±950.8	Dihydroflavonol 4-reductase	*Dendrobium hybrid cultivar*

582 **A** extraction condition/**B** peptide count/**C** unique peptide/**D** data analyzes was performed using ANOVA ($p < 0.05$) followed by the Tukey significance test, were asterisk (*) denote
583 statistical significance difference. /**E** mean value of the highest condition ± Standard deviation value.

584

585 **Supplementary Table 4.** Proteins identified involved in terpene biosynthesis.

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
A0A0F6YPA0	Vb3, Vb5	2	1	0.00*	Vb3	5290.5±127.8	Mevalonate kinase (MK)	*Nervilia fordii*
A0A0K1Z517	Vb2, Vb5	7	5	0.00*	Vb2	55753.4±5228.4	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	*Dendrobium catenatum*
A0A159BKQ8	Vb2, Vb5	12	9	1.00	Vb2	5573±1563.7	1-deoxy-D-xylulose-5-phosphate synthase (Fragment)	*Nervilia fordii*
A9YCD2	Vb5	23	13	0.80	Vb5	7723.1±212.9	Flavonol synthase	*Cymbidium hybrid cultivar*
I0DHG9	Vb1, Vb5	7	5	1.00	Vb5	23722.2±3051.4	Jasmonic acid carboxyl methyltransferase	*Cymbidium ensifolium* (Epidendrum ensifolium)
I0DHH0	Vb2, Vb3, Vb5	16	7	1.00	Vb2	9944.2±2500.3	Allene oxide synthase	*Cymbidium ensifolium* (Epidendrum ensifolium)
J7EGE4	Vb5	19	10	0.76	Vb5	2084.8±398.5	Terpene synthase (Fragment)	Phalaenopsis bellina
J7EGE9	Vb5	8	3	1.00	Vb5	372±85.2	Terpene synthase	Phalaenopsis equestris (Moth orchid)
I7FHQ2	Vb2, Vb5	12	3	1.00	Vb5	1134.8±205.4	Farnesyl pyrophosphate synthase	Cymbidium goeringii
K7WTR3	Vb3, Vb5	11	1	1.00	Vb5	21842.8±2656.4	Farnesyl pyrophosphate synthase	*Dendrobium catenatum*
V9MW47	Vb2	11	9	1.00	Vb2	37505.7±3271.3	Mitogen-activated protein kinase	*Dendrobium catenatum*

Supplementary Table 4. Proteins identified involved in terpene biosynthesis (continue).

Accession	Extract cond.[A]	Pept[B]. count	Unique pept.[C]	ANOVA *p*-value[D]	Highest mean condition	Highest mean±SD[E]	Protein names	Organism
V9QM18	Vb3, Vb5	12	2	0.00*	Vb5	1893.8±349.8	Farnesyl pyrophosphate synthetase 1 (Farnesyl pyrophosphate synthetase 2)	*Dendrobium huoshanense*
V9TNZ2	Vb3, Vb5	8	3	0.00*	Vb5	13404.7±390.4	3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA reductase)	*Dendrobium huoshanense*
X2KWX1	Vb3, Vb5	27	16	0.02*	Vb3	5776±113.9	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	*Dendrobium catenatum*

A extraction condition/**B** peptide count/**C** unique peptide/**D** data analyzes was performed using ANOVA ($p < 0.05$) followed by the Tukey significance test, were asterisk (*) denote statistical significance difference. /**E** mean value of the highest condition ± Standard deviation value.

Supplementary Table 8. Identified proteins of *V. bahiana* fruit from Vb1, Vb2, Vb3 and Vb5 samples- **not show**

Conclusões Gerais

- As proteínas identificadas nos frutos de *V. bahiana* confirmam o potencial enzimático dessa espécie para a produção de compostos, já descritos na literatura, como responsáveis pelo aroma e sabor do extrato natural de baunilha.
- Nos frutos de *V. bahiana* foram identificadas proteínas ligadas à floração, frutificação e amadurecimento de frutos, proteínas essas que podem servir como indicadores de amadurecimento dos frutos de *Vanilla spp*.
- Foi possível identificar proteínas relacionadas à biossíntese de fenóis de importância econômica como: ácido cafeico, ácido coumárico, ácido ferúlico e vanilina.
- O uso de SDS na concentração de 1% foi essencial para extrair mais proteínas totais com grande amplitude dinâmica nos frutos maduros.
- 1% de SDS também se mostrou eficaz na extração de uma diversidade maior de proteínas, principalmente das proteínas de interesse.
- A metodologia desenvolvida neste trabalho se mostrou eficaz e fundamental para extrair e identificar uma grande diversidade proteínas, reduzindo a interferência de contaminantes como o SDS.
- As análises de quantificação absoluta comprovaram a presença da vanilina, pirogalol e ácido cumarico nos frutos maduros de *V. bahiana*, que já forma encontrados nos extratos de *V. planifolia* e *V. tahitian* ligados ao aroma e sabor de baunilha característico.
- Os resultados e a metodologia apresentados podem colaborar com a conservação das *Vanilla spp*. e na caracterização químicas de seus frutos, podendo o perfil proteico identificado servir como um marcador de origem do extrato assegurando a qualidade do produto ao consumidor.

Perspectivas Futuras

Uma vez que o extrato natural de baunilha possui uma qualidade de sabor e aroma superior ao sintético, e essa relação de qualidade está ligada a mistura complexa de metabólitos presentes no fruto, uma das perspectivas futuras desse trabalho é realizar estudo metabolômico dos frutos de *V. bahiana* coletados.

Outra perspectiva desse trabalho é continuar os estudos ômicos das espécies de *Vanilla* localizadas na cidade do Rio de Janeiro, principalmente de *V. chamissonis*, que já foi coletada. Futuras extrações de proteínas serão realizadas utilizando o SDS nas soluções, com o intuito de obter uma maior diversidade de proteínas.