UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE QUÍMICA

Programa de Pós-Graduação em Ciências Biológicas (Bioquímica)

EDUARDO SHIGUEO KITANO

**Estudo do Proteoma de Folhas de Cultivares e
Genótipos de Cana-de-Açúcar**

Versão original da Tese defendida

São Paulo

Data do Depósito na SPG:

22/02/2017

EDUARDO SHIGUEO KITANO

**Estudo do Proteoma de Folhas de Cultivares e
Genótipos de Cana-de-Açúcar**

Tese apresentada ao Instituto de Química da
Universidade de São Paulo para a obtenção do
Título de Doutor em Ciências (Bioquímica)

Orientadora: Dra. Solange Maria de Toledo Serrano

São Paulo

2017

Ficha Catalográfica

Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP.

K62e	<p>Kitano, Eduardo Shigueo</p> <p>Estudo do proteoma de folhas de cultivares e genótipos de cana-de-açúcar / Eduardo Shigueo Kitano. -- São Paulo, 2017.</p> <p>252p. +CD-Rom</p> <p>Tese (doutorado) - Instituto de Química da Universidade de São Paulo. Departamento de Bioquímica</p> <p>Orientador: Serrano, Solange Maria de Toledo</p> <p>1. Proteínas : Bioquímica 2. Biologia molecular I. T. II. Serrano, Solange Maria de Toledo, orientador.</p> <p>574.19245 CDD</p>
------	---

À minha amada esposa Carolina.

AGRADECIMENTOS

Nessas poucas linhas que tenho para os agradecimentos, espero poder contemplar todas as pessoas que fizeram parte de minha trajetória até aqui e que, direta ou indiretamente, contribuíram para o meu desenvolvimento pessoal e profissional.

Agradeço primeiramente à minha amada esposa, Carolina, que compartilhou comigo muitos momentos felizes ao longo desses quase 14 anos e também, algumas frustrações. A pessoa que sempre me direcionou para o caminho correto e que sempre admirei pela inteligência, pela doçura, sinceridade e pelo zelo incondicional à família. Espero fazer por você tudo o que você fez e faz por mim hoje. Você é meu exemplo de vida e a pessoa que mais admiro neste mundo. Obrigado por sempre acreditar em meu potencial, mas nunca se esqueça de seu próprio, que, como eu sempre disse (e você sabe disso), é maior do que o potencial de qualquer pessoa que conheço neste mundo. Te amo para sempre e obrigado por tudo!

Aos meus pais, Massanori Kitano e Kazumi Nakayama Kitano, não só pelo esforço em buscar oferecer sempre uma educação de qualidade para mim e para os meus irmãos, mas principalmente pelo esforço em estabelecer um ambiente familiar saudável e feliz. Muito do que sou hoje devo eternamente aos ensinamentos dos dois.

Aos meus irmãos, Massao e Sayuri, pelos incríveis momentos que vivenciamos durante a infância e também pelos bons momentos que compartilhamos hoje.

À minha sobrinha, Kaory, por ser sempre essa menina linda e maravilhosa.

Em memória aos meus avós, sempre presentes durante minha formação e também em memória ao meu cunhado Toti, que partiu cedo demais de nossas vidas.

Aos meus familiares, pelos bons momentos de convivência durante todos esses anos.

À minha sogra e sogro, Tereza e Toshikazu Sugimoto, pelos ensinamentos, por sempre me ajudar e pelos aconselhamentos construtivos para o meu desenvolvimento pessoal. Às minhas cunhadas, Natália e Sophia, por mais de uma década de convivência. Admiro sempre a força de vontade, esforço e inteligência de vocês.

Aos familiares de Londrina, pelos diversos momentos de alegria, pelos conselhos e que, receberam este novo agregado de braços abertos. Agradeço especialmente à batian, ditian e aos tios e tias.

À minha orientadora Dra. Solange Maria de Toledo Serrano, pelos ensinamentos e pela dedicação e amor ao que faz, pela conduta profissional sempre correta, pela confiança e amizade durante todos esses anos e por acreditar em meu potencial até mesmo quando eu mesmo duvidava.

À Dra. Glaucia Mendes Souza, pela oportunidade de desenvolver esse projeto e pelos ensinamentos.

Ao Dr. Sávio de Siqueira Ferreira, à Dra. Carolina Lembke, Érica Rodrigues Bandeira, Danielle Izilda Rodrigues e à Dra. Amanda de Souza pela valiosa ajuda nos estágios iniciais do projeto e pela parceria nos experimentos de campo e pela ajuda no desenvolvimento deste projeto.

Ao Dr. Milton Nishiyama-Júnior, não só pela ajuda profissional, mas também pela amizade e pelas conversas informais. Um dos poucos profissionais capazes de compreender e conciliar a biologia com a matemática.

Aos amigos André Zelanis e Alexandre Tashima, por todos os ensinamentos técnicos, pela parceria em projetos e experimentos e pelas boas conversas que tivemos durante esses anos. Sempre admirei a maneira como vocês entendem e dominam as técnicas de espectrometria de massas.

À Dra. Adriana F. Paes-Leme, pela amizade e por me orientar durante o período de Iniciação Científica, período em que aprendi a ter gosto pela ciência.

Às amigas Milene Menezes e Ana Karina, por todos esses anos de convivência, pelas boas risadas e por me acolherem tão bem no laboratório. Vocês são um exemplo de dedicação e companheirismo.

Às amigas Débora Andrade, Pollyana Fernandes, Maria Andrea Camarano e Carolina Brás, pelos momentos de descontração e colaborações em projetos. Espero ainda estar presente para sempre ajudar e ser ajudado por vocês.

Aos amigos que passaram pelo laboratório Edson Yamashiro, Luciana Nasciben, Ana Helena Pagotto, Gabriela Dias, Aline Lopes, Mara Hoshida e Amanda Asega, que estão sempre presentes.

Aos amigos Ismael Feitosa Lima, Ivan Novaski, Bruno Duzzi, Diego Dantas e Daniel Ferreira, pela amizade sincera e pelas muitas conversas descontraídas que tivemos durante esses anos de convivência. Muito obrigado pela força que vocês me deram nesse período.

Ao Dr. Leo Iwai, pela amizade durante esses poucos anos que nos conhecemos e por compartilhar comigo a incrível experiência de realizar um estágio no Japão.

Ao Dr. Yasushi Ishihama e seu grupo, do Departamento de Bioanálise Molecular e Celular da Universidade de Kyoto, pela indescritível experiência de realizar um estágio em seu laboratório e pela colaboração nas análises de fosfoproteômica.

Aos professores Dr. Sergio Verjovski-Almeida, Dra. Maria Julia Manso Alves e Dra. Suely Lopes Gomes, pela participação no meu exame de qualificação e pelas importantes discussões e questionamentos.

Ao meu assessor(a) da FAPESP, pelos pareceres sempre construtivos ao projeto.

Ao Dr. Hugo Armelin, coordenador do *Center of Toxins, Immune-response and Cell Signaling*-CEPID-FAPESP, pelo apoio a este projeto.

Aos pesquisadores, pós-graduandos, estagiários e funcionários do Laboratório Especial de Toxinologia Aplicada, Instituto Butantan, que participaram direta ou indiretamente neste projeto.

À Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP (Processo 2011/11308-0), pela concessão da bolsa de Doutorado direto e projeto regular de auxílio à pesquisa.

À Universidade de São Paulo, ao Instituto de Química e ao Instituto Butantan, por permitir o meu desenvolvimento acadêmico e profissional.

***“[...] let no feeling of
discouragement prey
upon you, and in the end
you are sure to succeed.”***

Abraham Lincoln

July 22, 1860.

Resumo

Kitano, ES. **Estudo do proteoma de folhas de cultivares e genótipos de cana-de-açúcar**. 2017. 252 p. Tese (Doutorado) - Programa e Pós-Graduação em Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo.

A cana-de-açúcar, uma das maiores culturas agrícolas no mundo, é uma antiga fonte de energia para os seres humanos e mais recentemente, tem sido utilizada como fonte para a produção de etanol. Em estudos moleculares sobre a cana-de-açúcar, sua complexidade genética e o fato de seu genoma não estar completamente sequenciado, dificultam a análise dos proteomas de seus diferentes tecidos. Recentemente, em um estudo de um grupo brasileiro sobre os transcriptomas de vários tecidos de cana-de-açúcar, 17.563 transcritos completos foram sequenciados, os quais correspondem a aproximadamente 53% do genoma predito desta planta. Entretanto, sabe-se que o conteúdo de mRNA nem sempre se correlaciona diretamente com os níveis de expressão de proteínas. Dessa forma, a caracterização do proteoma é importante para identificar as proteínas que são de fato expressas, em complementaridade aos estudos sobre o genoma, e fornece informação preditiva sobre a composição do proteoma de diferentes cultivares. Entretanto, estudos sobre folhas de cana-de-açúcar são raros. Neste estudo, com o objetivo de caracterizar o proteoma de folhas +1 de cana-de-açúcar, foram selecionadas duas espécies ancestrais (*Saccharum officinarum* [S. off] e *S. spontaneum* [S. sp]) e uma cultivar híbrida (SP80-3280 [SP]), as quais diferem-se na capacidade de produção de sacarose e na resistência a patógenos. A análise de parâmetros fisiológicos (taxa de fotossíntese, condutância estomática, taxa de transpiração e eficiência no uso de água), em folhas de 12 meses revelaram que em S. off e SP existe uma relação melhor entre o número de moléculas de água perdidas na transpiração e o número de moléculas de CO₂ assimiladas na fotossíntese do que em S. sp. Os mesmos parâmetros foram avaliados em folhas +1 de SP com 4, 8, 11 e 13 meses de crescimento e revelaram o papel dos estômatos na manutenção do potencial hídrico das folhas, atuando na modulação da taxa de respiração em resposta às mudanças ambientais, entretanto sem induzir alterações drásticas na taxa de fotossíntese, no período avaliado neste estudo. Com o objetivo de obter proteínas em condições ótimas para análise proteômica, foram testados diferentes métodos de extração de proteínas, que foram avaliados quanto ao rendimento e perfil eletroforético das proteínas extraídas, e selecionado um protocolo que utiliza uma combinação de fenol e dodecil sulfato de sódio para este estudo. Para caracterizar e comparar os proteomas de folhas +1 de S. off, S. sp e SP, foram utilizadas as abordagens proteômicas *Shotgun/bottom-up*, *GeLC-MS/MS* e *Multidimensional Protein Identification Technology* (MudPIT) (*off-line* e “*on-line*”), e as plataformas MaxQuant e Perseus, como ferramentas de bioinformática para análise dos dados. Os dados obtidos por meio destas abordagens foram combinados e resultaram na identificação de 5.825 grupos de proteínas em folhas +1 de S. off, S. sp e SP, e a análise de enriquecimento de termos pelo *Gene Ontology* destacou a natureza fotossintética desses proteomas, além da representatividade de organelas envolvidas

em processos de fotossíntese e metabolismo de carbono. A análise quantitativa comparativa dos 1.060 grupos de proteínas identificados por meio da abordagem *Shotgun/bottom-up* revelou que 120 proteínas encontravam-se diferencialmente expressas em folhas de *S. off*, *S. sp* e SP, e destas, aquelas que mostraram diferença em abundância de pelo menos quatro vezes, são relacionadas com processos biológicos envolvidos com os estresses biótico e abiótico (resposta a outros organismos, resposta ao íon cádmio, defesa contra fungos e defesa ao estresse oxidativo). Para a análise da variação ontogenética no proteoma de folhas +1 de SP (de 4, 8, 11 e 13 meses de crescimento) foi aplicada a abordagem *Shotgun/bottom-up*, que resultou na identificação de 5.780 peptídeos, correspondendo a 1.615 grupos de proteínas únicos. Sessenta e cinco proteínas foram identificadas com abundância diferencial entre as amostras, e a análise de agrupamento hierárquico revelou que as folhas de 13 meses mostraram um perfil mais distinto, enquanto que as outras formaram um grupo separado. A caracterização dos fosfoproteomas de folhas +1 de *S. off*, *S. sp* e SP utilizando a abordagem *Immobilized Metal Affinity Chromatography (IMAC)* como método de enriquecimento de fosfopeptídeos, e identificação por LC-MS/MS, resultou na identificação de 633 fosfopeptídeos únicos em 515 proteínas. A caracterização dos fosfoproteomas de folhas +1 de SP (de 4, 8, 11 e 13 meses de crescimento) utilizando as abordagens *IMAC* e *Hydroxy Acid-Modified Metal Oxide Chromatography (HAMMOG)* como métodos de enriquecimento de fosfopeptídeos, e identificação por LC-MS/MS, resultou na identificação de 558 fosfopeptídeos únicos em 413 proteínas. Em ambas análises, foram identificadas várias proteínas quinases fosforiladas, e de forma geral, a análise de enriquecimento de termos pelo *Gene Ontology* revelou o núcleo e o citoplasma, entre os componentes celulares, e a regulação da transcrição, a resposta ao estresse abiótico e a glicólise, entre os processos biológicos, como os termos mais enriquecidos. Este é o primeiro estudo objetivando a caracterização dos proteomas totais de folhas de cultivares e genótipos de cana-de-açúcar no Brasil, e seus resultados representam um significativo esforço para o uso de diferentes abordagens de proteômica para aprofundar o conhecimento sobre os complexos processos fisiológicos que modulam o crescimento e a performance em cana-de-açúcar.

Palavras-chave: folha de cana-de-açúcar, proteômica, fosfoproteômica, *Saccharum officinarum*, *Saccharum spontaneum*, SP80-3280.

Abstract

Kitano, ES. **Study on the leaf proteome of sugarcane cultivars and genotypes.** 2017. 252 p. PhD Thesis. Graduate Program in Biochemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

Sugarcane, one of the major crops worldwide, is an old energy source for humans and, more recently, has been used as a source for ethanol production. In sugarcane molecular studies, its complex genetic background and the lack of a complete genome sequencing hinder the analysis of proteomes of different tissues. Recently, in a Brazilian research study focused on the transcriptomes of different sugarcane tissues, 17,563 full-length transcripts were sequenced and covered approximately 53% of the predicted sugarcane genome. However, it is known that mRNA levels are not always directly correlated with protein expression levels. Therefore, the study of the sugarcane proteome is important to identify proteins that are actually expressed, complementing the genome studies, and provides predictive information on the proteome composition of different cultivars. However, studies on sugarcane leaves are rare. In this study, with the aim of characterizing the proteome of sugarcane +1 leaves, we selected two ancestral species (*Saccharum officinarum* [S. off] and *S. spontaneum* [S. sp]), and a hybrid cultivar (SP80-3280 [SP]), which differ in sucrose production and pathogen resistance. The analysis of physiological parameters (photosynthesis rate, stomatal conductance, transpiration and efficiency in water use) of 12 month-old +1 leaves revealed that in S. off and SP there is a better relationship between the number of H₂O molecules lost by transpiration and the number of CO₂ molecules assimilated upon photosynthesis than in S. sp. The same parameters were evaluated in +1 leaves of SP upon 4, 8, 11 and 13 months of growth and revealed the role of stomata in the maintenance of leaf water potential, acting in the modulation of the transpiration rate in response to environmental changes, however without inducing drastic changes in the photosynthesis rate, during the period evaluated in this study. In order to obtain proteins in optimal conditions for proteomic analysis, we carried out different protein extraction methods and evaluated the yield and electrophoretic profiles of extracted proteins, and selected a protocol using a combination of phenol and sodium dodecyl sulfate to carry out this study. To characterize and compare the +1 leaf proteomes of S. off, S. sp and SP we used Shotgun/bottom-up, GeLC-MS/MS and Multidimensional Protein Identification Technology (MudPIT) (off-line and “on-line”) as proteomic approaches, and the software MaxQuant and Perseus as bioinformatic tools for data analysis. The combined results of these approaches allowed for the identification of 5,825 protein groups in +1 leaves of S. off, S. sp and SP, and the Gene Ontology enrichment analysis highlighted the photosynthetic nature of these proteomes, along with high representativeness of organelles involved in the processes of photosynthesis and carbon metabolism. The comparative quantitative analysis of 1,060 protein groups

identified using the Shotgun/bottom-up approach revealed that 120 proteins were differentially expressed in +1 leaves of *S. off*, *S. sp* and *SP*, and of these, those that showed at least a 4-fold change are related to biological processes involved in both abiotic and biotic stress (response to other organism, response to cadmium ion, defense response to fungus, and defense to oxidative stress). For the analysis of the ontogenetic variation in the proteome of +1 leaves of *SP* (of 4, 8, 11 and 13 months old plants) the Shotgun/bottom-up proteomic approach was employed and resulted in the identification of a total of 5,780 peptides corresponding to 1,615 unique protein groups. Sixty five proteins were identified as differentially abundant between the samples and a hierarchical cluster analysis revealed that the 13 months-old leaves showed the most distinct profile, while the others clustered together. The characterization of the phosphoproteomes of +1 leaves of *S. off*, *S. sp* and *SP* using Immobilized Metal Affinity Chromatography (IMAC) as an enrichment method for phosphopeptides, and LC-MS/MS analysis, resulted in the identification of 633 unique phosphopeptides in 515 proteins. The characterization of the phosphoproteome of +1 leaves of *SP* (of 4, 8, 11 and 13 months old plants) using IMAC and Hydroxy Acid-Modified Metal Oxide Chromatography (HAMMOC) as enrichment methods for phosphopeptides, and LC-MS/MS analysis, resulted in the identification of 558 unique phosphopeptides in 413 proteins. In both analyses various phosphorylated kinases were identified, and the overall Gene Ontology enrichment analysis showed as most enriched terms nucleus and cytoplasm, as “cellular components”, and regulation of transcription, response to abiotic stress and glycolysis in “biological process”. This is the first study to characterize the whole proteome of leaves from sugarcane cultivars and genotypes in Brazil, and its results constitute a significant effort to use different proteomic approaches to advance the knowledge on the complex physiological processes modulating sugarcane growth and performance.

Keywords: sugarcane leaf, proteomics, phosphoproteomics, *Saccharum officinarum*, *Saccharum spontaneum*, SP80-3280.

LISTA DE FIGURAS

Figura 1. Mapa de distribuição do plantio de cana-de-açúcar em território brasileiro.

Figura 2. Representação esquemática da estrutura do genoma de cultivares modernas de cana-de-açúcar.

Figura 3. Morfologia da folha de cana-de-açúcar.

Figura 4. Representação anatômica de folhas de plantas C3 e C4.

Figura 5. Representação das vias de fotossíntese C4 e C3.

Figura 6. Representação simplificada da via de síntese de sacarose em plantas.

Figura 7. Esquema ilustrativo dos métodos de análise utilizados para a caracterização dos proteomas e fosfoproteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280.

Figura 8. Métodos de cromatografia líquida bidimensional utilizados no estudo dos proteomas de folhas +1 de cana-de-açúcar.

Figura 9. Passos de confecção de uma *StageTip*.

Figura 10. Representação esquemática da marcação isobárica dos fosfopeptídeos com o reagente TMT.

Figura 11. Parâmetros fisiológicos medidos em folhas + 1 de *S. officinarum*, *S. spontaneum* e SP80-3280.

Figura 12. Dados meteorológicos mensais referentes aos períodos de plantio e coletas de folhas +1 de SP80-3280 ao longo do desenvolvimento da planta.

Figura 13. Parâmetros fisiológicos medidos em folhas +1 de SP80-3280 coletadas em quatro estágios de desenvolvimento da planta.

Figura 14. Eletroforese em gel de SDS-poliacrilamida 12% das proteínas extraídas de folhas +1 de *S. officinarum* e *S. spontaneum* de acordo com os métodos de extração de proteínas TCA/acetona, Fenol e Fenol/SDS.

Figura 15. Número de bandas de proteínas observadas nos perfis eletroforéticos das amostras extraídas de folhas +1 de *S. officinarum* e *S. spontaneum* de acordo com os métodos de extração TCA/acetona, Fenol e Fenol/SDS.

Figura 16. Perfil eletroforético das proteínas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, extraídas de acordo com o método Fenol/SDS.

Figura 17. Espectros de fragmentação de peptídeos precursores ilustrando o critério adotado para a seleção de proteínas identificadas por somente um peptídeo.

Figura 18. Diagrama de Venn demonstrando a distribuição do número total de proteínas identificadas nas amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 a partir da abordagem *Shotgun/bottom-up proteomics*.

Figura 19. *Boxplots* mostrando as distribuições dos valores médios de intensidade e de contagem de espectros de MS/MS dos peptídeos associados às proteínas identificadas de forma exclusiva, compartilhadas entre duas ou entre as três amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280.

Figura 20. Perfil eletroforético das proteínas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, extraídas de acordo com o método Fenol/SDS.

Figura 21. Diagrama de Venn demonstrando a distribuição do número total de proteínas identificadas nas amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com a abordagem GeLC-MS/MS.

Figura 22. Distribuição do número de proteínas presentes em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 identificadas de acordo com as abordagens de análise *Shotgun/bottom-up* e GeLC-MS/MS, segundo o número de peptídeos.

Figura 23. Comparação entre as porcentagens de cobertura de sequência do conjunto de proteínas concomitantemente identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 a partir das abordagens de análise *Shotgun/bottom-up* e GeLC-MS/MS.

Figura 24. Diagrama de Venn demonstrando a distribuição do número total de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 a partir das análises de MudPIT *off-line* e “*on-line*”.

Figura 25. Comparação de número de peptídeos e % de cobertura de sequências obtidas com as abordagens MudPIT *off-line* e “*on-line*”.

Figura 26. *Boxplots* demonstrando as distribuições de GRAVY score dos peptídeos identificados em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 pelas metodologias de MudPIT *off-line* e “*on-line*”.

Figura 27. Números médios de proteínas e peptídeos identificados em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com as diferentes abordagens de análise utilizadas.

Figura 28. Comparação entre as coberturas de sequências dos conjuntos totais de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280

segundo as abordagens *Shotgun/bottom-up*, GeLC-MS/MS e análise MudPIT *off-line* e “*on-line*”.

Figura 29. Comparação entre as massas e GRAVY score dos peptídeos identificados em proteínas presentes em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 segundo as abordagens *Shotgun/bottom-up*, GeLC-MS/MS e análise MudPIT *off-line* e “*on-line*”.

Figura 30. Diagrama de Venn mostrando a comparação do número de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com as abordagens de análise proteômica *Shotgun/bottom-up*, GeLC-MS/MS e MudPIT *off-line* e “*on-line*”.

Figura 31. Distribuição do conjunto de proteínas identificadas em amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 considerando a combinação das identificações de proteínas de acordo com as abordagens de análise MudPIT *off-line* e “*on-line*”, *Shotgun/bottom-up* e GeLC-MS/MS.

Figura 32. Anotação funcional do conjunto de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 segundo as categorias Componente Celular, Processo Biológico e Função Molecular do *Gene Ontology*.

Figura 33. Termos enriquecidos no conjunto de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 segundo as categorias funcionais Componente Celular, Processo Biológico e Função Molecular do *Gene Ontology*.

Figura 34. Histogramas mostrando a distribuição dos dados de LFQ intensity resultantes da análise quantitativa *label-free* aplicada às amostras de proteínas extraídas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280.

Figura 35. Histogramas mostrando a distribuição dos dados de LFQ intensity após a utilização da ferramenta “*replace missing values from normal distribution*”.

Figura 36. *Heat Map* das proteínas diferencialmente expressas nos proteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280.

Figura 37. Perfil de expressão do conjunto de proteínas diferencialmente expressas nos proteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280.

Figura 38. Diagramas de Venn dos peptídeos fosforilados e fosfoproteínas totais identificados em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280.

Figura 39. Termos enriquecidos no conjunto de fosfoproteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 segundo as categorias funcionais Componente Celular, Processo Biológico e Função Molecular do *Gene Ontology*.

Figura 40. Perfil eletroforético das proteínas de folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta.

Figura 41. Diagrama de Venn mostrando a distribuição do número total de proteínas identificadas em amostras de folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta.

Figura 42. Termos enriquecidos no conjunto de proteínas identificadas em folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta segundo as categorias funcionais Componente Celular, Processo Biológico e Função Molecular do *Gene Ontology*.

Figura 43. *Heat Map* das proteínas diferencialmente expressas entre os proteomas de folhas +1 de SP80-3280, coletadas em quatro diferentes estágios de desenvolvimento da planta.

Figura 44. Perfil de expressão do conjunto de proteínas diferencialmente expressas entre os proteomas de folhas +1 de SP80-3280, coletadas em quatro diferentes estágios de desenvolvimento da planta.

Figura 45. Diagrama de Venn mostrando o número e a distribuição dos fosfopeptídeos identificados em amostras de folhas +1 de SP80-3280, coletadas em quatro diferentes estágios de desenvolvimento da planta, de acordo com os métodos de enriquecimento de fosfopeptídeos HAMMOC e IMAC.

Figura 46. Análise de enriquecimento de termos da população de fosfoproteínas identificadas em folhas +1 de SP80-3280, coletadas em quatro diferentes estágios de desenvolvimento da planta, segundo as categorias Componente Celular, Processo Biológico e Função Molecular do *Gene Ontology*.

LISTA DE TABELAS

Tabela 1. Massas dos reagentes de TMTsixplex utilizados para a derivatização dos fosfopeptídeos resultantes do processo de enriquecimento HAMMOc aplicado sobre a mistura de peptídeos tripsínicos presentes em proteínas de folhas +1 de SP80-3280, coletadas em quatro diferentes estágios de desenvolvimento da planta.

Tabela 2. Valores dos parâmetros fisiológicos medidos em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 sob condições de luz artificial.

Tabela 3. Valores dos parâmetros fisiológicos medidos em folhas +1 de SP80-3280 coletadas em quatro estágios de desenvolvimento da planta sob condições de luz artificial.

Tabela 4. Rendimento de extração de proteínas presentes nas folhas +1 de cana-de-açúcar de acordo com os diferentes métodos de extração utilizados.

Tabela 5. Rendimento de extração de proteínas presentes nas folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com o método de extração Fenol/SDS.

Tabela 6. Número de peptídeos e grupos únicos de proteínas identificados em amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP8032-80 a partir das análises de LC-MS/MS dos peptídeos gerados pela digestão tripsínica em solução.

Tabela 7. Número de peptídeos e grupos de proteínas identificados em amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP8032-80 a partir das análises por LC-MS/MS dos peptídeos gerados pela digestão tripsínica *in gel*.

Tabela 8. Número de peptídeos e grupos de proteínas identificados em amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP8032-80 de acordo com a metodologia MudPIT *off-line* e "*on-line*".

Tabela 9. Número de peptídeos e proteínas identificados em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com as abordagens *Shotgun/bottom-up*, GeLC-MS/MS e MudPIT *off-line* e "*on-line*".

Tabela 10. Categorias de maior ocorrência em proteínas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com as classificações presentes na biblioteca do banco de dados *Superfamily*.

Tabela 11. Termos enriquecidos no conjunto de proteínas identificadas concomitantemente em folhas +1 de *S. officinarum* e SP80-3280 e em folhas +1 de *S.*

spontaneum e SP80-3280 segundo as categorias Função Molecular e Processo Biológico do *Gene Ontology*.

Tabela 12. Termos enriquecidos no conjunto de proteínas identificadas de forma exclusiva em folhas +1 de *S. officinarum* *S. spontaneum* e SP80-3280 segundo as categorias Função Molecular e Processo Biológico do *Gene Ontology*.

Tabela 13. Proteínas diferencialmente expressas em folhas +1 de *S. officinarum* e *S. spontaneum* que apresentaram uma variação de abundância de pelo menos quatro vezes.

Tabela 14. Proteínas diferencialmente expressas em folhas +1 de *S. officinarum* e SP80-3280 que apresentaram uma variação de abundância de pelo menos quatro vezes.

Tabela 15. Proteínas diferencialmente expressas em folhas +1 de *S. spontaneum* e SP80-3280 que apresentaram uma variação de abundância de pelo menos quatro vezes.

Tabela 16. Sumário dos dados obtidos das análises dos fosfoproteomas dos conjuntos de proteínas extraídas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280.

Tabela 17. Número de peptídeos e grupos de proteínas identificados em amostras de folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta.

Tabela 18. Lista de proteínas identificadas de forma exclusiva em folhas +1 de SP80-3280 em diferentes estágios de desenvolvimento da planta.

Tabela 19. Proteínas diferencialmente expressas em folhas +1 de SP80-3280 em quatro diferentes estágios de desenvolvimento da planta que apresentaram uma variação de abundância de pelo menos quatro vezes.

Tabela 20. Sumário dos dados obtidos das análises dos fosfoproteomas dos conjuntos de proteínas extraídas de folhas +1 de SP80-3280 em quatro diferentes estágios de desenvolvimento da planta, de acordo com os métodos de enriquecimento de fosopeptídeos HAMMOC e IMAC.

LISTA DE ABREVIATURAS E SIGLAS

2-D PAGE: *Two-Dimensional Polyacrylamide Gel Electrophoresis*

2-D PAGE/MALDI-ToF/ToF-MS: *Two-Dimensional Polyacrylamide Gel Electrophoresis/ Matrix-assisted laser desorption/ionization-Time-of-Flight/Time-of-Flight-Mass Spectrometry*

A: Taxa fotossintética

ACN: Acetonitrila

ADPGlc-PPase: ADP-glicose pirofosforilase

AF: Ácido fórmico (do inglês *formic acid*)

ANOVA: Análise de variância (do inglês *Analysis of variance*)

ATP: *Adenosine triphosphate*

CA: Anidrase carbônica (do inglês *carbonic anhydrases*)

CAD1: *constitutively activated cell death 1*

cDNA: DNA complementar (do inglês *Complementary DNA*)

CHAPS: 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate

CID: Dissociação induzida por colisão (do inglês *Collision induced dissociation*)

cps: *Counts per scan*

DDA: Aquisição dependente de dados (do inglês *Data Dependent Acquisition*)

DNA: Ácido desoxirribonucleico (do inglês *Deoxyribonucleic acid*)

DTT: Ditioneitol

E: Taxa de transpiração

EDTA: Ácido etilenodiamino tetra-acético (do inglês *Ethylenediamine tetraacetic acid*)

ESI: *Electrospray ionization*

ESI-MS: *Electrospray ionization-Mass Spectrometry*

ESTs: *Expressed Sequence Tags*

FAD: *Flavin adenine dinucleotide*

FASP: *Filter-Aided Sample Preparation*

FDR: Taxa de falso positivo (do inglês *False discovery rate*)

FTMS: *Fourier transform mass spectrometry*

GO: *Gene Ontology*

Gs: Condutância estomática

GTP: *Guanosine-5'-triphosphate*

H⁺-PPase: H⁺ pirofosfatase

HAMMOC: *Hydroxy Acid-Modified Metal Oxide Chromatography*

HCD: *Higher-energy collision dissociation*

HFBA: Ácido heptafluorobutírico (do inglês *Heptafluorobutyric acid*)

HPLC: Cromatografia líquida de alta performance (do inglês *High performance liquid chromatography*)

HSPs: *Heat Shock Proteins*

IAA: Iodoacetamida

ID: Diâmetro interno (do inglês *Inner diameter*)

IgG: Imunoglobulina G

IMAC: *Immobilized Metal Affinity Chromatography*

ITMS: *Ion trap mass spectrometry*

iTRAQ: *isobaric Tag for Relative and Absolute Quantification*

LBT: *Lysis buffer thiourea*

LC-MS/MS: Cromatografia líquida acoplada à espectrometria de massas (do inglês *Liquid chromatography tandem-mass spectrometry*)

m/z: Relação massa/carga

MACPF: *Membrane attack complex component/perforin domain*

MAL: Malato

MALDI: *Matrix-assisted laser desorption/ionization*

MALDI-ToF-MS: *Matrix-assisted laser desorption/ionization-Time-of-Flight-Mass Spectrometry*

MDH: Malato desidrogenase (do inglês *Malate dehydrogenase*)

MetSO: Metionina sulfóxido

MOC: *Metal Oxide Chromatography*

MRM: *Multiple Reaction Monitoring*

mRNA: RNA mensageiro (do inglês *Messenger RNA*)

MS: *Mass Spectrometry*

MS/MS: *Tandem mass spectrometry*

MudPIT: *Multidimensional protein identification technology*

NAD: *Nicotinamide adenine dinucleotide*

NADP: *Nicotinamide adenine dinucleotide phosphate*

NADP-ME: Enzima málica

NTA: Ácido nitrilotriacético (do inglês *Nitrilotriacetic acid*)

NTP: *Nucleoside triphosphate*

OAA: Oxaloacetato

OD: Diâmetro externo (do inglês *Outside diameter*)

PEP: Fosfoenolpiruvato (do inglês *Phosphoenolpyruvate*)

PEP: Probabilidade de erro posterior (do inglês *Posterior error probability*)

PEPC: Fosfoenolpiruvato carboxilase (do inglês *Phosphoenolpyruvate carboxylase*)

PMSR: Peptídeo metionina sulfóxido redutase

PPDK: Piruvato fosfato diquinase (do inglês *Pyruvate, phosphate dikinase*)

PPi: Pirofosfato inorgânico

PRPs: Proteínas relacionadas à patogênese (do inglês *Pathogenesis-Related Proteins*)

PVA: Piruvato

RB: Réplica biológica

RNA: *Ribonucleic acid*

ROS: *Reactive Oxygen Species*

RT: Réplica técnica

RuBisCO: Ribulose-1,5-bisfosfato carboxilase/oxigenasse

RuBP: Ribulose-1,5-bisfosfato

S. off: *Saccharum officinarum*

S. sp: *Saccharum spontaneum*

SAR: *Systemic acquired resistance*

SAS: *Sugarcane Assembled Sequence*

SCX: *Strong cation exchange*

SDB-XC: *Styrene-divinylbenzene-XC Sorbent Particles*

SDS: Dodecil sulfato de sódio (do inglês *Sodium dodecyl sulfate*)

SP: SP80-3280

StageTip: Stop-and-go extraction tip

SUCAST: *Sugarcane Signal Transduction*

SUCEST: *Sugarcane EST Genome Project*

TCA: Ácido tricloroacético (do inglês *Trichloroacetic acid*)

TEMED: N,N,N',N'-Tetrametil etileno-diamino

TFA: Ácido trifluoroacético (do inglês *Trifluoroacetic acid*)

TMT: *Tandem Mass Tag*

TP: Trioses fosfato

TVD: *Top Visible Dewlap*

WUE: Eficiência no uso de água (do inglês *Water use efficiency*)

REPRESENTAÇÃO DOS AMINOÁCIDOS

Alanina	Ala	A
Cisteína	Cys	C
Ácido aspártico	Asp	D
Ácido glutâmico	Glu	E
Fenilalanina	Phe	F
Glicina	Gly	G
Histidina	His	H
Isoleucina	Ile	I
Lisina	Lys	K
Leucina	Leu	L
Metionina	Met	M
Asparagina	Asn	N
Prolina	Pro	P
Glutamina	Gln	Q
Arginina	Arg	R
Serina	Ser	S
Treonina	Thr	T
Valina	Val	V
Triptofano	Trp	W
Tirosina	Tyr	Y

SUMÁRIO

1. INTRODUÇÃO	26
1.1. A Cultura da Cana-de-Açúcar e sua Importância no Cenário Nacional Atual	26
1.2. Cana-de-Açúcar: Histórico do Cultivo e Melhoramento Genético	28
1.3. A Folha de Cana-de-Açúcar, Fotossíntese e Síntese de Sacarose	32
1.4. Metodologias “ômicas” Empregadas no Estudo da Cana-de-Açúcar	41
2. OBJETIVOS	54
3. MATERIAIS E MÉTODOS	55
3.1. Material de Estudo, Análise de Parâmetros Fotossintéticos e Tratamento Inicial das Amostras	55
3.2. Extração de Proteínas	56
3.2.1. Método de Extração TCA/Acetona	57
3.2.2. Método de Extração Fenol	57
3.2.3. Método de Extração Fenol/SDS	59
3.3. Solubilização das Amostras de Proteínas	60
3.4. Quantificação de Proteínas	61
3.5. Eletroforese em Gel de SDS-Poliacrilamida	61
3.6. Coloração para Proteínas Totais	62
3.6.1. Método de Coloração com Coomassie G-250	62
3.6.2. Método de Coloração com Nitrato de Prata	63
3.7. Digestão Tripsínica <i>in gel</i>	63
3.8. Digestão Tripsínica em Solução	65
3.9. Remoção de Sais Presentes nas Amostras de Peptídeos	66
3.10. Estratégias de Análise Proteômica de Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280	67
3.10.1. Análise dos Proteomas de Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280 por Digestão Tripsínica em Solução e Análise por LC-MS/MS (Estratégia <i>Shotgun/Bottom-Up Proteomics</i>)	70
3.10.2. Análise dos Proteomas de Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280 por Eletroforese em Gel de SDS-Poliacrilamida e Análise por LC-MS/MS (Estratégia <i>GeLC-MS/MS</i>)	71

3.10.3. Análise dos Proteomas de Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280 por Digestão Tripsínica em Solução, Cromatografia Ortogonal (MudPIT) e Análise por LC-MS/MS	73
3.10.3.1. MudPIT <i>off-line</i>	75
3.10.3.2. MudPIT “ <i>on-line</i> ”	77
3.10.4. Busca em Banco de Dados e Análise Bioinformática dos Dados Gerados pelas Análises dos Proteomas de folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280	78
3.10.4.1. Análise dos Dados Gerados pela Estratégia <i>Shotgun/Bottom-Up Proteomics</i>	79
3.10.4.2. Análise dos Dados Gerados pela Estratégia GeLC-MS/MS	82
3.10.4.3. Análise dos Dados Gerados pela Estratégia MudPIT <i>off-line</i> e “ <i>on-line</i> ”	84
3.10.4.4. Análise dos Dados Gerados pela Combinação das Estratégias <i>Shotgun/Bottom-Up Proteomics</i> , GeLC-MS/MS e MudPIT <i>off-line</i> e “ <i>on-line</i> ”	85
3.10.5. Análise dos Fosfoproteomas de Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280 pela Metodologia IMAC de Enriquecimento de Fosfopeptídeos e Análise por LC-MS/MS	86
3.10.5.1. Análise dos Dados Gerados pela Estratégia Baseada no Enriquecimento de Fosfopeptídeos pela metodologia IMAC em amostras de peptídeos provenientes de amostras de folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280	88
3.10.6. Análise do Proteoma de folhas +1 de SP80-3280 coletadas ao longo do desenvolvimento da planta pela Utilização da Estratégia <i>Shotgun/Bottom-Up Proteomics</i>	90
3.10.6.1. Análise dos Dados Gerados pela Estratégia <i>Shotgun/Bottom-Up Proteomics</i> no Estudo do Proteoma de folhas +1 de SP80-3280 coletadas ao longo do desenvolvimento da planta	91
3.10.7. Análise do Fosfoproteoma de folhas +1 de SP80-3280 coletadas ao longo do desenvolvimento da planta a partir das Metodologias IMAC e HAMMOC e Análise por LC-MS/MS	91
3.10.7.1. Método IMAC e Análise dos Fosfopeptídeos por LC-MS/MS	92
3.10.7.2. Método HAMMOC com Marcação Isobárica dos Peptídeos com TMT (<i>Tandem Mass Tag</i>) e Análise dos Fosfopeptídeos por LC-MS/MS	92
3.10.7.3. Análise dos Dados Gerados pelas Estratégias Baseadas no Enriquecimento de Fosfopeptídeos pelas metodologias IMAC e HAMMOC em amostras de folhas +1 de SP80-3280 Coletadas ao Longo do Desenvolvimento da Planta	97
4. RESULTADOS E DISCUSSÃO.....	99
4.1. Análise de Parâmetros Fisiológicos em Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280 Coletadas de Plantas de 12 Meses e em Folhas +1 de SP80-3280 Coletadas em Diferentes Estágios de Desenvolvimento da Planta.....	99
4.1.1. Análise de Parâmetros Fisiológicos em Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280 Coletadas de Plantas de 12 Meses	99
4.1.2. Análise de Parâmetros Fisiológicos em Folhas +1 de SP80-3280 Coletadas em Diferentes Estágios de Desenvolvimento da Planta	103

4.2. Extração de Proteínas das Amostras de Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280 segundo o Método Fenol/SDS	115
4.3. Análise Qualitativa dos Proteomas de Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280	117
4.3.1. Análise dos Proteomas de Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280 pela Abordagem <i>Shotgun/bottom-up proteomics</i>	118
4.3.2. Análise dos Proteomas de Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280 pela Abordagem GeLC-MS/MS.....	127
4.3.3. Análise dos Proteomas de Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280 pela Abordagem MudPIT <i>off-line</i> e “ <i>on-line</i> ”	136
4.3.4. Avaliação Global das Abordagens de Análise dos Proteomas de Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280.....	144
4.3.5. Agrupamento dos Dados de Identificação de Proteínas Provenientes das Abordagens de Análise <i>Shotgun/bottom-up</i> , GeLC-MS/MS e MudPIT <i>off-line</i> e “ <i>on-line</i> ” Aplicadas sobre o Estudo dos Proteomas de Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280.....	152
4.4. Análise Quantitativa dos Proteomas de Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280	173
4.5. Análise dos Fosfoproteomas de Folhas +1 de <i>S. officinarum</i> , <i>S. spontaneum</i> e SP80-3280	189
4.6. Análises Qualitativa e Quantitativa dos Proteomas de Folhas +1 de SP80-3280 durante 4 Estágios de Desenvolvimento da Planta.....	197
4.6.1. Análise Qualitativa	197
4.6.2. Análise Quantitativa <i>label-free</i>	206
4.7. Análise dos Fosfoproteomas de Folhas +1 de SP80-3280 durante 4 Estágios de Desenvolvimento da Planta	212
5. CONSIDERAÇÕES FINAIS	220
6. REFERÊNCIAS	222
7. LISTA DE ANEXOS (EM CD-ROM)	245
7.1. Figuras Suplementares.....	245
7.2. Tabelas Suplementares.....	245
8. ANEXOS.....	248
8.1. Súmula Curricular	248

1. INTRODUÇÃO

1.1. A CULTURA DA CANA-DE-AÇÚCAR E SUA IMPORTÂNCIA NO CENÁRIO NACIONAL ATUAL

A cana-de-açúcar é uma das principais culturas do setor agrícola brasileiro. O país se destaca como o maior produtor de cana do mundo, com a produção aproximada de 634 milhões de toneladas na safra 2014/15 em uma área total de plantio correspondendo a aproximadamente 11 milhões de hectares (Fig. 1). São Paulo é o estado que apresenta a maior produção nacional com 52% da área plantada, seguido por Goiás, Minas Gerais, Mato Grosso do Sul, Paraná, Alagoas e Pernambuco. Estes sete estados juntos são responsáveis por mais de 90% da produção nacional (CONAB, 2015; UNICA, 2016).

Figura 1. Mapa de distribuição do plantio de cana-de-açúcar em território brasileiro. Fonte: CONAB, 2015.

No mesmo período, o Brasil foi responsável pela produção de 20% de todo o açúcar produzido no mundo (FAO, 2015), com o total aproximado de 35,5 milhões de toneladas. Acompanhando a tendência de aumento no consumo de açúcar em nível mundial, estima-se que a produção deste derivado alcance um valor aproximado de 48 milhões de toneladas em 2024, representando um aumento de aproximadamente 35% em relação à produção observada em 2014/15. A mesma tendência de aumento de produção, porém de maneira mais acentuada, é projetada para o etanol obtido da cana-de-açúcar. Estima-se que a produção de etanol aumente aproximadamente 47% no mesmo período, atingindo cerca de 42 bilhões de litros em 2024 (OECD, 2015).

A indústria sucroalcooleira gera mais de um milhão de empregos no Brasil, considerando a mão-de-obra na lavoura, na indústria, na administração, na pesquisa, na comercialização e exportação dos produtos, destacando-se como um dos principais setores do agronegócio brasileiro. As exportações de açúcar e etanol foram responsáveis pela geração de US\$ 10 bilhões em divisas externas no ano de 2014. Além da importância econômica que representa, a cana-de-açúcar contribui significativamente na matriz energética do país, uma vez que seus derivados compreendem cerca de 16% da energia consumida no país, percentual este que representa um valor acima da média mundial (13,2%) no uso de energias renováveis (UNICA, 2016). Como resultado, o Brasil apresenta papel de destaque internacional em questões que concernem às estratégias energéticas e combate às mudanças climáticas. Ainda, o cultivo da cana-de-açúcar fomenta a geração de empregos, a interiorização do desenvolvimento e a inovação tecnológica (BNDES, 2008; Gordinho, 2010).

1.2. CANA-DE-AÇÚCAR: HISTÓRICO DO CULTIVO E MELHORAMENTO GENÉTICO

A cana-de-açúcar pertence ao gênero *Saccharum* da família das gramíneas (*Poaceae*), e se destaca pela sua alta capacidade de acumular sacarose nos colmos, podendo atingir concentrações de aproximadamente 0,7 M em entrenós maduros (Moore, 1995). O gênero foi inicialmente descrito por Linnaeus em 1753, na primeira edição do livro de sua autoria *Species Plantarum*, e tem origem associada às regiões do sul e sudeste asiático. Dentro deste gênero ocorrem principalmente seis espécies, que apresentam altos níveis de ploidia e diferentes números de cromossomos: *Saccharum officinarum* ($2n=80$), *S. robustum* ($2n=60-194$), *S. barberi* ($2n=81-124$), *S. sinense* ($2n=110-120$), *S. spontaneum* ($2n=40-128$) e *S. edule* ($2n=60-80$) (Daniels e Roach, 1987; D'Hont *et al.*, 1998; Paterson *et al.*, 2013), as quais podem ser subdivididas em duas categorias que compreendem às variedades tradicionais e às variedades selvagens (Grivet *et al.*, 2004). Os cruzamentos artificiais interespecíficos entre as variedades tradicionais, onde se encontram as espécies *S. officinarum*, *S. barberi*, *S. edule* e *S. sinense*, e as variedades selvagens (*S. spontaneum* e *S. robustum*) dão origem às variedades comerciais modernas, que combinam características fenotípicas de interesse para o cultivo e produção de sacarose pela planta. Atualmente, de forma geral, as cultivares modernas utilizadas na indústria sucroalcooleira são cultivadas como um híbrido, resultado do cruzamento entre *S. officinarum* e a variedade selvagem *S. spontaneum*, com menores contribuições de outras espécies do mesmo gênero (Daniels e Roach, 1987; Ming *et al.*, 2006).

A espécie *S. officinarum*, também conhecida como cana nobre devido a sua característica de acumular altas concentrações de sacarose em seus colmos, tem sua origem intimamente associada à domesticação do cultivo de cana-de-açúcar pelo homem, visto que ela não é encontrada na natureza (Sreenivasan *et al.*, 1987). Evidências apontam que a domesticação da cultura ocorreu em Papua Nova Guiné com a seleção de *S. officinarum* a partir da variedade selvagem *S. robustum* (Brandes, 1956; Daniels e Roach, 1987; Grivet *et al.*, 2004). Além da elevada capacidade de acúmulo de sacarose, *S. officinarum* se caracteriza pelo desenvolvimento de colmos grossos com baixo conteúdo de fibras, folhas longas e largas e raízes pouco profundas. Contudo, esta espécie apresenta alta suscetibilidade às doenças causadas principalmente por bactérias e vírus, e ao ataque de insetos. Acredita-se que a espécie selvagem *S. spontaneum* teve origem no sul da Ásia e é considerada uma espécie muito mais adaptada que a *S. officinarum*, apresentando a capacidade de crescer em diferentes habitats e altitudes, sendo encontrada em países como Japão, China, Índia, Paquistão, Egito, Sudão, entre outros (Tai e Miller, 2001; Daniels e Roach, 1987; Pursglove, 1972). É uma espécie altamente polimórfica e normalmente caracterizada pelo desenvolvimento de colmos finos com alto conteúdo de fibras e folhas relativamente mais finas. Ao contrário da *S. officinarum*, *S. spontaneum* apresenta baixa capacidade de acumular sacarose em seus colmos, porém é uma espécie que se caracteriza pela alta resistência às doenças que normalmente acometem *S. officinarum* e, além disso, apresenta maior tolerância quando submetida a algum tipo de estresse abiótico (Sreenivasan *et al.*, 1987; Paterson *et al.*, 2013).

Os primeiros programas de melhoramento genético da cana-de-açúcar foram registrados em meados do século XIX em Java e Barbados (Ming *et al.*, 2006). Inicialmente somente a espécie *S. officinarum* (cana nobre) foi utilizada, justamente pela alta capacidade de acúmulo de sacarose em seus colmos. A observação de que estas plantas eram altamente suscetíveis à diversas doenças e ataques de insetos, bem como a restrição de cultivo a determinados ambientes, levou à conclusão de que o arsenal genético destas plantas deveria ser expandido com o intuito de melhorar sua adaptabilidade e aumentar sua resistência às doenças e insetos. Diante dessas questões, iniciaram-se os cruzamentos entre *S. officinarum* e a variedade selvagem *S. spontaneum*, a partir de um processo denominado “nobilização”. Este processo é caracterizado pelo cruzamento inicial de variedades de *S. officinarum* com *S. spontaneum*, seguido de diversos retrocruzamentos das progênes resultantes com *S. officinarum*, permitindo a combinação entre os genomas e favorecimento das características fenotípicas da variedade tradicional, uma vez que este processo se caracteriza pela assimetria na transmissão cromossômica, evidenciada pela não redução do número de cromossomos do óvulo de *S. officinarum* ($2n$) adicionado ao número gamético do parental masculino (n) nas duas primeiras gerações (cruzamento dos parentais e primeiro retrocruzamento), gerando uma progênie $2n+n$. Assim, somente a partir do segundo retrocruzamento é que se observa a transmissão cromossômica simétrica ($n+n$). Dessa forma, este processo resulta em uma variedade “nobilizada” que apresenta número de cromossomos variável (de 100 a 130) e é constituída majoritariamente por cromossomos provenientes de *S. officinarum*. (Bremer, 1961; Ming *et al.*, 2006). De fato, este processo representou um grande avanço aos programas de

melhoramento genético, que passaram a desenvolver as variedades modernas a partir das plantas geradas pelo processo de nobilitação. Porém, do ponto de vista de estruturação do genoma, contribuiu para o aumento de sua complexidade devido ao aumento da poliploidia e aneuploidia nas cultivares modernas. A análise da estrutura do genoma de variedades modernas por técnicas citogenéticas demonstrou que 70-80% dos cromossomos são derivados de *S. officinarum*, 10-20% provenientes de *S. spontaneum* e 10% são produtos de recombinação entre as duas espécies (Fig. 2) (D'Hont *et al.* 1996; D'Hont, 2005).

Figura 2. Representação esquemática da estrutura do genoma de cultivares modernas de cana-de-açúcar. As barras pretas representam os cromossomos herdados de *S. officinarum*, enquanto que barras cinzas representam cromossomos herdados de *S. spontaneum* e as barras que apresentam a combinação de cores representam os cromossomos produtos de recombinação entre as duas espécies. Fonte: D'Hont *et al.*, 2005.

Cruzamentos entre variedades nobilizadas formaram a base para a produção de cultivares híbridas a partir da primeira metade do século XX, que por sua vez compõem o principal *pool* genético para a produção das cultivares comerciais, de forma que as variedades modernas apresentam uma base genética muito similar, provenientes de não mais do que 15-20 genótipos de cultivares nobilizadas (Roach, 1989). Diante disso, a partir de 1965, foi iniciado em Barbados um programa com o objetivo de aumentar a base genética da cana-de-açúcar utilizando clones diferentes daqueles utilizados inicialmente em Java. Diversos clones semicomerciais produzidos pelo programa têm sido introduzidos nos programas de melhoramento genético desde o final da década de 80, porém, a dificuldade em identificar a incorporação de um conjunto de genes, em uma população de um programa de melhoramento genético, apenas pela seleção baseada em manifestações fenotípicas (seleção visual) representou a maior limitação para o sucesso dos programas de ampliação da base genética. Dessa forma, os avanços recentes na área de biotecnologia podem auxiliar nos programas de melhoramento para produção de novas cultivares (Paterson *et al.*, 2013), de forma a acelerar o processo de geração de uma cultivar comercial, que atualmente, de acordo com o modelo clássico de melhoramento genético, pode levar de 12 a 15 anos desde o primeiro cruzamento até a sua comercialização (Roach, 1989).

1.3. A FOLHA DE CANA-DE-AÇÚCAR, FOTOSSÍNTESE E SÍNTESE DE SACAROSE

A cana-de-açúcar, assim como o milho (*Zea mays*) e o sorgo (*Sorghum bicolor*), é uma planta C4, que se caracteriza pelo alto rendimento fotossintético, convertendo até

6% da energia solar captada em biomassa, enquanto que em plantas C3, como a *Arabidopsis* (*Arabidopsis thaliana*) e o arroz (*Oryza sativa*), observa-se uma taxa de conversão na ordem de 4,6% (Zhu *et al.*, 2008). O tecido foliar é o tecido que apresenta maior atividade fotossintética em plantas superiores, pois contém elevado número de cloroplastos, onde a energia radiante absorvida pelos pigmentos fotossintéticos (clorofila e carotenoides) é utilizada para converter moléculas de dióxido de carbono (CO₂) e água em carboidratos e outras moléculas orgânicas. Em cana-de-açúcar, as folhas, normalmente longas, estreitas e lineares, distribuem-se de forma alternada ao longo do colmo e são sustentadas por uma espessa nervura central, de forma que essas características maximizam a captação da energia radiante utilizada para o processo de fotossíntese (Fig. 3A). Morfologicamente, as folhas de cana-de-açúcar podem ser divididas em dois segmentos: a bainha e a lâmina. A bainha corresponde à região que envolve o entrenó e é fisicamente ligada à lâmina por uma junção articulada denominada junção da lâmina (*blade joint*), que contém a lígula, a aurícula e o colarinho (*dewlap*) (Fig. 3B) (Moore, 1987). As folhas de cana-de-açúcar são numeradas de cima para baixo, de acordo com o sistema Kuijper (1915) de classificação, sendo que a primeira folha que apresenta o colarinho visível é denominada folha +1 ou folha TVD (*Top Visible Dewlap*) e representa neste estágio a primeira folha completamente expandida. As folhas relativamente mais novas, que ainda se encontram em estágio de expansão são numeradas 0, -1, -2, e as mais velhas, já completamente expandidas que podem ou não estar em estágio de senescência, +2, +3, +4, +5, +6 (Fig. 3C).

Figura 3. Morfologia da folha de cana-de-açúcar destacando as principais regiões (A e B) e sistema de classificação de folhas de cana-de-açúcar segundo Kuijper (C). Em (C) a seta indica o primeiro colarinho (*dewlap*) visível em relação ao topo, evidenciando a folha +1 ou folha TVD (*Top Visible Dewlap*). Fontes: A: Sávio de Siqueira Ferreira; B: Eduardo S. Kitano; C: Cheavegatti-Gianotto *et al.*, 2011.

A maior capacidade fotossintética das plantas C4 em comparação às plantas C3 decorre de diferenças encontradas entre suas anatomias foliares (Fig. 4), que resultam na maximização da atividade carboxilase da enzima ribulose-1,5-bisfosfato carboxilase/oxigenase (RuBisCO), em plantas que realizam a fotossíntese C4.

Figura 4. Representação anatômica de folhas de plantas C3 (A) e C4 (B). Fonte: Alberts *et al.*, 2002.

As diferenças anatômicas entre folhas C3 e C4 são observadas principalmente na organização das células da bainha do feixe vascular e do mesofilo em torno do feixe vascular. Em folhas C3 (Fig. 4A) as células do mesofilo apresentam uma distribuição dispersa em torno das células da bainha do feixe vascular e a presença de cloroplastos está restrita às células do mesofilo. Em folhas C4 (Fig. 4B) verifica-se a presença de cloroplastos tanto em células do mesofilo quanto em células da bainha do feixe vascular, que se distribuem circularmente em torno do feixe vascular e de forma compacta, caracterizando a anatomia Kranz (Sage *et al.*, 2012). Ainda, células da bainha do feixe vascular são encontradas com tamanhos relativamente maiores em folhas C4.

O processo de fotossíntese se inicia quando a luz solar incide sobre os pigmentos presentes nos complexos de captação de luz associados aos fotossistemas I e II na

membrana dos tilacóides, que operam de modo simultâneo e em série durante a fotossíntese, e são conectados pelo complexo citocromo b_6f e pelos carreadores plastoquinona e plastocianina. A energia dos fótons, transformada em fluxo de elétrons entre os fotossistemas resulta em um gradiente de prótons através da membrana dos tilacóides que leva à síntese de ATP e NADPH, com a participação da ATP sintase e ferredoxina-NADP redutase, respectivamente, na etapa fotoquímica da fotossíntese (Kerbaudy, 2004). Estes produtos, por sua vez, serão utilizados na etapa de fixação do CO_2 .

Na fotossíntese em plantas C4, a etapa de fixação de carbono é caracterizada pela integração das reações bioquímicas que ocorrem no citosol e nos cloroplastos de células do mesofilo e da bainha do feixe vascular (Fig. 5A). Dessa forma, o CO_2 proveniente da atmosfera difunde-se pelos estômatos, e no citosol das células do mesofilo é convertido em HCO_3^- (ânion bicarbonato) com a participação da anidrase carbônica, tornando-se substrato para a enzima fosfoenolpiruvato carboxilase (PEPC), resultando na formação do oxaloacetato (molécula constituída por quatro carbonos), motivo pelo qual plantas que produzem este intermediário durante a fixação de carbono são classificadas como plantas C4. Posteriormente, oxaloacetato é translocado para dentro do cloroplasto, onde é convertido a malato, que então se difunde para os cloroplastos das células da bainha do feixe vascular através de canais conhecidos como plasmodesmos. Nestas células, a descarboxilação do malato, catalisada pela enzima málica, leva à formação de piruvato e regeneração de CO_2 em concentrações aproximadamente 10 vezes maiores em relação à concentração atmosférica, favorecendo a atividade carboxilase da RuBisCO (de Siqueira Ferreira *et al.*, 2013). Esta

enzima apresenta a peculiaridade de ser bifuncional, promovendo tanto a carboxilação como a oxigenação de seu substrato, a ribulose-1,5-bisfosfato. De forma que em condições de alta concentração de CO_2 , situação encontrada no ambiente das células da bainha do feixe vascular em plantas C4, ela promove a síntese do intermediário 3-fosfoglicerato, que no ciclo de Calvin (ou ciclo C3), resulta na síntese de diversos carboidratos e na regeneração da ribulose-1,5-bisfosfato. Já em altas concentrações de O_2 , a RuBisCO favorece a oxigenação de seu substrato, resultando na síntese de 3-fosfoglicerato e também de 2-fosfoglicolato. Este último não é utilizado no ciclo de Calvin e é processado por uma via que promove a perda de carbono, em um processo conhecido como fotorrespiração. Dessa forma, a fotossíntese C4 é caracterizada pela supressão do evento de fotorrespiração, devido ao aumento da concentração de CO_2 nas células da bainha do feixe vascular, favorecendo a atividade carboxilase da RuBisCO (Maroco *et al.*, 1998; Sheen, 1999). Esse mecanismo encontrado em plantas C4, que maximiza a atividade carboxilase da RuBisCO pelo aumento da concentração de CO_2 , por sua vez, não está presente em plantas que seguem a fotossíntese C3 (Fig. 5B). Dessa forma, dependendo das concentrações relativas de CO_2 e de O_2 que difundem através das aberturas dos estômatos, a enzima RuBisCO, presente nas células do mesófilo em folhas de plantas C3, exerce atividade oxigenase pronunciada, que eventualmente resulta na perda de carbonos devido ao evento de fotorrespiração.

A

B

Figura 5. Representação das vias de fotossíntese C4 (A) e C3 (B). O processo de concentração de CO_2 na fotossíntese C4 (A) se inicia no citosol de células do mesófilo com a conversão de CO_2 para HCO_3^- e sucessiva reação com moléculas de fosfoenolpiruvato (PEP) para formação de um composto contendo 4 carbonos, o oxaloacetato (OAA), que por sua vez é convertido a malato (MAL) ainda na célula do mesófilo. Esse composto se difunde para as células da bainha do feixe vascular onde é descarboxilado, regenerando o CO_2 , NADPH e piruvato (PVA). O CO_2 , cuja concentração é aproximadamente 10 vezes maior no sítio de carboxilação da enzima RuBisCO em relação ao valor atmosférico, é então fixado pela ação desta enzima pertencente à via C3 nas células da bainha do feixe vascular, onde é utilizado para a síntese de trioses fosfato (TP) utilizadas para a síntese de sacarose. A via C3 (B), por sua vez, não apresenta o mecanismo responsável pelo aumento da concentração de CO_2 que leva à maximização da atividade carboxilase da RuBisCO. CA: Anidrase carbônica; PEPC: fosfoenolpiruvato carboxilase; MDH: malato desidrogenase; NADP-ME: enzima málica; RuBisCO: ribulose-1,5-bisfosfato carboxilase oxigenase; PPDK: piruvato fosfato diquinase; RuBP: ribulose-1,5-bisfosfato. Modificado de: Sage *et al.*, 2012.

Uma vez fixado pelo ciclo C3, que ocorre tanto em plantas que seguem a fotossíntese C3 quanto a C4, o carbono proveniente do CO_2 atmosférico é utilizado principalmente para a síntese de moléculas de sacarose e amido, que ocorrem no citosol e cloroplasto, respectivamente. A síntese de sacarose se inicia a partir da translocação das trioses fosfato (gliceraldeído-3-fosfato) produzidas nos cloroplastos pelo ciclo de

Calvin até o citosol das células do mesofilo (Fig. 6). Neste compartimento, as moléculas de trioses fosfato são transformadas em moléculas de frutose-1,6-bisfosfato, em uma reação catalisada pela enzima aldolase, e que são posteriormente convertidas em frutose-6-fosfato pela ação da frutose-1,6-bisfosfatase citosólica (Serrato *et al.* 2009). Frutose-6-fosfato pode ser convertida em glicose-6-fosfato pela enzima hexose fosfato isomerase, ou ser combinada com UDP-glicose para formar sacarose-6-fosfato pela ação da enzima sacarose fosfato sintase. A sacarose-6-fosfato é subsequentemente desfosforilada pela sacarose fosfato fosfatase, convertendo-a em sacarose, que será então direcionada ao floema e distribuída aos demais tecidos da planta.

Célula do mesofilo

Figura 6. Representação simplificada da via de síntese de sacarose em plantas. As trioses fosfato provenientes do ciclo de Calvin (ciclo C3) são convertidas, a partir de uma série de reações enzimáticas, em frutose-6-fosfato e UDP-glicose que reagem entre si em uma reação catalisada pela sacarose fosfato sintase para a formação de sacarose-6-fosfato que é então submetida à desfosforilação pela enzima sacarose fosfato fosfatase, gerando sacarose que será transferida para o floema e então aos tecidos. Modificado de: Taiz e Zeiger, 2002.

O acúmulo de sacarose nos tecidos-dreno é influenciado por processos relacionados ao metabolismo e ao transporte de açúcares e, apesar de ser alvo de diversos estudos de caracterização bioquímica e fisiológica, a relação entre o acúmulo de sacarose nos colmos (tecido-dreno) e a atividade fotossintética em folhas (tecido-

fonte) ainda não é bem compreendida (McCormick *et al.*, 2009). Irvine e colaboradores (1975) demonstraram que a taxa fotossintética não está diretamente correlacionada com a capacidade da planta em acumular sacarose ao constatar que folhas de *S. spontaneum*, variedade selvagem que apresenta baixa capacidade de acúmulo de sacarose, apresentavam uma taxa de fotossíntese aproximadamente 30% maior em relação às folhas de *S. officinarum*, que se caracteriza por acumular altas concentrações de sacarose nos colmos. Outros trabalhos demonstraram que o aumento da demanda de carboidratos pelos tecidos-dreno é acompanhado pelo aumento da taxa fotossintética nos tecidos-fonte, indicando que a taxa de fotossíntese nas folhas é limitada pelas necessidades energéticas do colmo (McCormick *et al.* 2006; McCormick *et al.* 2008). Ainda, o acúmulo de açúcares nas folhas causa a diminuição da taxa fotossintética (Hartt, 1963; Waldron *et al.*, 1967). Dessa forma, sugere-se a existência de um sistema de percepção de açúcares que envolve, além da sacarose, enzimas, como a hexoquinase (Dai *et al.* 1999), que apresentam papel importante na regulação da taxa de fotossíntese pela planta.

1.4. METODOLOGIAS “ÔMICAS” EMPREGADAS NO ESTUDO DA CANA-DE-AÇÚCAR

O advento e o constante desenvolvimento de metodologias *high throughput* aplicadas às análises de genomas, transcriptomas e proteomas, têm contribuído para o avanço do conhecimento sobre diversos organismos. Nas últimas décadas, os genomas de diferentes espécies foram sequenciados, incluindo desde os genomas de organismos mais simples como vírus e bactérias até os genomas de organismos mais complexos

como fungos, animais e plantas. Após a publicação do genoma de *Arabidopsis thaliana* (*The Arabidopsis Genome Initiative*, 2000), sucederam-se diversos outros genomas de plantas sequenciados, principalmente a partir da introdução dos sequenciadores de nova geração (*Next Generation Sequencing*), que permitiram a diminuição do tempo de análise e dos custos de sequenciamento (Margulies *et al.*, 2005). Atualmente, existem mais de 100 genomas sequenciados de plantas, sendo que a maioria corresponde às plantas que apresentam elevado interesse agrônomo, dentre as quais podemos citar o arroz, a soja, o milho e o sorgo (Michael e Jackson, 2013; Michael e VanBuren, 2015). Em conjunto, os dados gerados pelos sequenciamentos desses genomas têm contribuído para a identificação de genes responsáveis por características desejáveis, possibilitando, por exemplo, a manipulação dos genes de interesse a partir de técnicas de biologia molecular. Ainda, estes dados contribuem para o estudo funcional dos produtos gênicos, as proteínas, ao fornecerem informações preditivas de sequências de aminoácidos, que são o principal foco dos estudos proteômicos.

Devido à ocorrência de poliploidias, o estudo da cana-de-açúcar em nível genético e molecular se torna mais desafiador do que em genomas relativamente mais simples, como os organismos diploides (Henry, 2010), de forma que essa relativa complexidade representa o principal gargalo para o sequenciamento de seu genoma. Como consequência, observa-se a dificuldade no desenvolvimento biotecnológico e de ferramentas genéticas para essa cultura (Menossi *et al.*, 2008). Apesar de não dispormos do genoma completo da cana-de-açúcar, observou-se ao longo das duas últimas décadas um progresso significativo no estudo de seu genoma a partir da utilização de ferramentas de análise de ESTs (*Expressed Sequence Tags*). ESTs são sequências de

DNA que representam trechos de RNA mensageiros (mRNAs) que serão eventualmente traduzidos em proteínas, revelando quais são os genes que serão expressos em um tecido ou órgão em uma dada situação fisiológica ou patológica. A informação carregada pela coleção de ESTs se configura como um importante ponto de partida para determinação do genoma de um organismo (Menossi *et al.*, 2008; Souza e da Silva, 2002), além de auxiliar no estabelecimento de estudos funcionais.

O primeiro estudo sobre o transcriptoma de cana-de-açúcar foi conduzido na África do Sul por Carson e Botha (2000) e possibilitou a identificação de uma pequena coleção de ESTs provenientes de folhas da cultivar híbrida Nco376. No Brasil, o estudo do genoma da cana-de-açúcar iniciou-se com o projeto SUCEST (*Sugarcane EST Genome Project*) (Vettore *et al.*, 2001; Vettore *et al.*, 2003), que resultou no sequenciamento da maior coleção de ESTs disponível em cana-de-açúcar. Aproximadamente 238.000 ESTs foram geradas de 26 bibliotecas de cDNA de diversos tecidos da planta, e foram agrupadas em 43.141 transcritos putativos exclusivos da cana-de-açúcar, denominados SASs (*Sugarcane Assembled Sequences*), correspondentes a 33.620 genes. Aproximadamente 50% dos transcritos foram associadas com a dinâmica celular, resposta ao estresse, metabolismo de proteínas, bioenergética e transdução de sinal celular (Vettore *et al.*, 2003). Recentemente, Nishiyama-Jr *et al.* (2014), utilizando diferentes tecidos dos genótipos ancestrais *S. officinarum* e *S. spontaneum*, e da cultivar híbrida SP80-3280, construíram uma coleção de 17.563 transcritos *full-length*, correspondendo a aproximadamente 53% dos genes preditos em cana-de-açúcar.

A partir da biblioteca derivada do projeto SUCEST, catálogos de genes foram selecionados para a utilização em experimentos de microarranjos de cDNA com o intuito

de se avaliar o perfil de expressão gênica em diferentes tecidos de diversas variedades de cana-de-açúcar, bem como em populações submetidas a diferentes tratamentos. Dessa forma, no projeto SUCAST (*Sugarcane Signal Transduction*) foram realizadas a identificação e a anotação de 1.830 SASs envolvidas em vias de transdução de sinal. Esse catálogo foi utilizado em diversos experimentos, resultando na identificação de genes diferencialmente expressos em vias de transdução de sinal em diferentes tecidos da planta, relacionados principalmente às características de acúmulo de sacarose, resposta ao estresse biótico e abiótico, resposta a hormônios e níveis variados de CO₂ (Souza *et al.*, 2001; Papini-Terzi *et al.*, 2005; Rocha *et al.*, 2007; De Souza *et al.*, 2008; Papini-Terzi *et al.*, 2009).

A análise funcional do genoma de cana-de-açúcar baseada em experimentos de microarranjo de cDNA tem sido responsável pelo significativo avanço no conhecimento de vias que podem interferir com características agronômicas de interesse, como a capacidade de acúmulo de açúcar e resistência ao estresse biótico e abiótico, por exemplo. Muito embora o transcriptoma represente o conjunto de genes expressos por um tecido, célula ou sistema em determinada condição, os níveis de expressão proteica não podem ser correlacionados diretamente com os níveis de mRNA, de forma que o aumento dos níveis de mRNA nem sempre está associado com o aumento dos níveis de expressão de proteínas. Considerando isso, o estudo do proteoma, o complemento proteico do genoma, representa um importante indicador de expressão dos genes (Gygi *et al.*, 1999; Watt *et al.*, 2010). Está claro que o conhecimento apenas do transcriptoma e do genoma não será suficiente para entender fenômenos que ocorrem em nível celular, tecidual e sistêmico. Assim sendo, a proteômica emergiu como um conjunto de

tecnologias para a análise da estrutura, função e interação entre proteínas, que permite associar a informação dos genes com a informação proteica (Yarmush e Jayaraman, 2002). Os estudos sobre os genomas e a utilização dos dados gerados para a construção de banco de dados das proteínas preditas são essenciais para o desenvolvimento dos estudos voltados para o proteoma (Wienkoop *et al.*, 2010; Tyers e Mann, 2003; Zhang *et al.*, 2013).

Inicialmente, os estudos sobre os proteomas de plantas e de outros organismos baseavam-se na utilização da técnica de separação de proteínas por eletroforese bidimensional (*two-dimensional polyacrylamide gel electrophoresis*; *2-D PAGE*) (O'Farrell, 1975), combinada com a análise de sequenciamento de proteínas baseada na degradação de Edman. Essa abordagem foi utilizada por Komatsu *et al.* (1993) e Kamo *et al.* (1995), resultando na identificação de pequenos conjuntos de proteínas provenientes de diferentes tecidos de arroz e *Arabidopsis*, respectivamente. Nesse mesmo período, com a popularização da metodologia de separação de proteínas por eletroforese bidimensional, alguns trabalhos direcionados ao estudo comparativo do proteoma de cana-de-açúcar foram publicados utilizando amostras de tecidos meristemáticos (Ramagopal, 1990), de culturas celulares (Moisyadi e Harrington, 1989), e de folhas (Ramagopal, 1994; Sugiharto *et al.*, 2002), com o objetivo de se buscar marcadores moleculares, entender o papel fisiológico de proteínas induzidas pelo choque térmico e identificar proteínas envolvidas nos processos de morfogênese ou que participam da resposta ao estresse induzido pela seca.

Os trabalhos conduzidos por Tanaka *et al.* (1988) e Fenn *et al.* (1989), sobre os processos de ionização de biomoléculas por, respectivamente, MALDI (*Matrix-assisted*

laser desorption/ionization) e ESI (*electrospray ionization*), tornaram possível a análise de peptídeos e proteínas pela técnica de espectrometria de massas, de forma a permitir que as técnicas de sequenciamento de proteínas e peptídeos baseadas em dados adquiridos pelos espectrômetros de massas substituíssem completamente os métodos baseados na reação de degradação de Edman. Adicionalmente, a elucidação de diversos genomas de plantas, como por exemplo os genomas de *Arabidopsis thaliana* e *Oryza sativa*, propiciou a expansão dos estudos sobre os proteomas de tecidos vegetais no início do século XXI. Inúmeros trabalhos científicos foram publicados a partir desse novo horizonte de análise de proteomas, representado inicialmente pela combinação da técnica de eletroforese bidimensional com a análise dos peptídeos gerados pela digestão enzimática *in gel* (de proteínas presentes nos *spots*) por espectrometria de massas utilizando MALDI ou ESI, e que mais tarde foi sendo substituída pela abordagem *shotgun/bottom-up proteomics* (Yates, 1998; Zhang *et al.*, 2013) que supera muitas das limitações da metodologia baseada na utilização da eletroforese bidimensional, como por exemplo problemas referentes à baixa reprodutibilidade, sensibilidade e restrições técnicas impostas pelo alto *dynamic range* de concentração de proteínas em diferentes proteomas (López, 2007; Zhang *et al.*, 2013).

Embora a combinação da metodologia 2-D PAGE com MALDI-ToF-MS ou ESI-MS tenha se tornado cada vez mais comum a partir dos anos 2000, os primeiros artigos direcionados ao estudo do proteoma de cana-de-açúcar, utilizando essa plataforma de análise, foram publicados de forma tardia em relação às espécies vegetais cujos estudos genômicos já se encontravam em estágios mais avançados, baseando-se majoritariamente nas informações contidas nas coleções de ESTs derivadas dos estudos

transcriptômicos aplicados em diferentes tecidos de cana-de-açúcar (Carson e Botha, 2000; Vettore *et al.*, 2001; Vettore *et al.*, 2003; Nishiyama-Jr *et al.*, 2014). Dessa forma, diante da inexistência de um genoma sequenciado, a utilização dos bancos de dados transcriptômicos foi essencial para os estudos proteômicos em cana-de-açúcar.

A alteração de diversos processos celulares, ocasionados pelo déficit hídrico e pelo aumento da salinidade do solo, resulta na redução da produtividade da cultura de cana-de-açúcar. Diante disso, Jangpromma *et al.* (2010), Zhou *et al.* (2011), Ngamhui *et al.* (2012), Pacheco *et al.* (2013), Murad *et al.* (2014) e Rahman *et al.* (2015) avaliaram a influência do estresse abiótico causados pelo déficit hídrico e pela salinidade do solo sobre a expressão de proteínas em folhas e em raízes de cana-de-açúcar. Em conjunto, esses trabalhos demonstraram o papel de proteínas importantes, como as HSPs (*Heat shock proteins*), fator de alongação EF-Tu, ATP-sintase, enzimas envolvidas em vias antioxidantes e de fotossíntese, entre outras, na resposta induzida pelo estresse abiótico. Ainda, sugere-se que a tolerância ao estresse abiótico causado pelo aumento da salinidade do solo é resultado da rápida percepção e ativação de genes que conferem proteção enquanto que respostas tardias de expressão gênica caracterizam o fenótipo sensível. Utilizando técnicas de separação de proteínas por eletroforese bidimensional combinadas com a identificação de proteínas por espectrometria de massas e análise por *Western-blot*, um biomarcador para cultivares resistentes à seca, denominado p18, foi identificado sistematicamente com maior nível de expressão em cultivares de cana-de-açúcar resistentes à seca (Jangpromma *et al.*, 2010), demonstrando que ferramentas de análise proteômica podem contribuir na descoberta de proteínas-alvos que eventualmente podem ser utilizadas em processos de seleção de variedades resistentes

à seca nos programas de melhoramento genético de cana-de-açúcar. O estresse biótico também representa um importante fator que impacta negativamente na produtividade da cana-de-açúcar. Utilizando metodologias de análise proteômica baseadas na utilização da plataforma 2-D PAGE/MALDI-ToF/ToF-MS, as respostas induzida pela interação entre o fungo *Sporisorium scitamineum*, agente causador do carvão, e a cana-de-açúcar foram investigadas por Que *et al.* (2011), Song *et al.* (2013) e Barnabas *et al.* (2016), demonstrando que a infecção da planta pelo fungo resulta em alterações na expressão de proteínas envolvidas na resistência a doenças, transdução de sinal, fotossíntese, em processos envolvidos no estresse oxidativo, transporte, metabolismo energético, dentre outras. Su *et al.* (2016) avaliaram a população de proteínas diferencialmente expressas em tecidos meristemáticos de genótipos sensíveis e resistentes à infecção por *S. scitamineum* a partir das abordagens proteômicas quantitativas, utilizando iTRAQ (*isobaric tag for relative and absolute quantification*) e MRM (*multiple reaction monitoring*), demonstrando que a maioria das proteínas diferencialmente expressas corresponde às proteínas relacionadas à resposta ao estresse e de resistência a doenças, como por exemplo beta-1,3-glucanase, PRPs (*pathogenesis-related proteins*), peroxidases e HSPs. Diante do crescente interesse na obtenção de etanol celulósico a partir do processo de fermentação de açúcares presentes na parede celular, Calderan-Rodrigues *et al.* (2014) conduziram um estudo sobre as proteínas de parede celular provenientes de cultura celular da cultivar SP80-3280, identificando 69 proteínas pertencentes à parede celular, incluindo oxidoredutases, glicosil hidrolases, proteases e, proteínas relacionadas ao metabolismo de lipídeos e à transdução de sinal. Este estudo foi estendido às amostras de proteínas extraídas de forma destrutiva e não

destrutiva de colmos jovens da mesma cultivar, resultando na identificação de 84 proteínas pertencentes à parede celular, contribuindo tanto para o aumento da cobertura do proteoma da parede celular em cana-de-açúcar como na identificação de possíveis candidatos alvos de manipulação genética que eventualmente propicie a diminuição dos custos de produção do etanol celulósico (Calderan-Rodrigues *et al.*, 2016). Os autores ainda destacaram a importância dos métodos de extração de proteínas e seu impacto na identificação dos componentes proteicos. Amalraj *et al.* (2010) avaliaram diferentes metodologias de extração de proteínas do colmo de cana-de-açúcar (cv. CoC 671) a partir da comparação entre a quantidade de proteínas obtida ao final de cada processo de extração e entre os perfis eletroforéticos bidimensionais gerados de acordo com o protocolo de extração utilizado, estabelecendo um mapa bidimensional do proteoma de colmo de cana-de-açúcar. Quarenta e dois *spots* presentes nos géis foram selecionados e submetidos à digestão *in gel* seguida pela análise dos peptídeos por espectrometria de massas, identificando 36 proteínas não redundantes, sendo a maior parte destas composta por proteínas relacionadas aos mecanismos de defesa da planta. Um estudo do sub-proteoma nuclear de folhas +1 coletadas de plantas jovens de cana-de-açúcar (*Saccharum* spp) foi realizado por Silva (2012), utilizando metodologias de enriquecimento de proteínas nucleares combinadas com as técnicas de separação de proteínas por eletroforese em gel de SDS-poliacrilamida e identificação de peptídeos pela técnica de espectrometria de massas. Esse estudo destacou a alta representatividade de histonas no conjunto de proteínas enriquecidas no núcleo, além da identificação de proteínas que estão envolvidas no metabolismo de DNA e RNA, demonstrando a aplicabilidade da metodologia de enriquecimento utilizada. Outros

trabalhos voltados ao estudo do proteoma de cana-de-açúcar têm se focado na caracterização da enzima peroxidase durante o desenvolvimento do colmo e em cultura celular (Cesarino *et al.*, 2012; Cesarino *et al.*, 2013), em organelas (Fu *et al.*, 2011) e na interação simbiótica entre bactérias endofíticas e cana-de-açúcar, como por exemplo os estudos realizados sobre a interação entre *Gluconoacetobacter diazotrophicus* e cana-de-açúcar, que buscam entender os papéis fisiológicos que as proteínas bacterianas desempenham na interação bactéria-hospedeiro (Lery *et al.*, 2008; dos Santos *et al.*, 2009; Lery *et al.*, 2011).

Os eventos de fosforilação e desfosforilação de proteínas mediados pelas proteínas quinases e fosfatases, respectivamente, são de fundamental importância para a modulação das vias de sinalização celular (Hunter, 2000), interferindo na estrutura, atividade, estabilidade e localização de proteínas e na capacidade de interação proteína-proteína ou proteína-ligante (Nakagami *et al.*, 2010; Zhang *et al.*, 2013). A análise em larga escala dessa importante modificação pós-traducional, conhecida também como fosfoproteômica, foi beneficiada pelo desenvolvimento tecnológico dos espectrômetros de massas e pelo desenvolvimento de estratégias analíticas de enriquecimento de fosfopeptídeos, baseadas principalmente na afinidade dos grupos fosfatos por íons metálicos como Ti^{4+} e Fe^{3+} , conhecidas como MOC (*Metal Oxide Chromatography*) (Sano e Nakamura, 2004; Larsen *et al.*, 2005; Wolschin *et al.*, 2005; Sugiyama *et al.*, 2007; Mazanek *et al.*, 2007) e IMAC (*Immobilized Metal Affinity Chromatography*) (Stensballe *et al.*, 2001; Kokubu *et al.*, 2005), respectivamente. Essa etapa de seleção de fosfopeptídeos é essencial para o estudo da população de proteínas fosforiladas, uma vez que estas se encontram em concentrações relativamente baixas. Assim como já

observado em estudos proteômicos direcionados às espécies vegetais, os estudos fosfoproteômicos foram inicialmente executados sobre espécies cujas informações genômicas já se encontravam em estágios mais avançados. Se por um lado os estudos relativos ao proteoma de cana-de-açúcar estão em seus passos iniciais, os estudos sobre o fosfoproteoma estão apenas idealizados do ponto de vista metodológico, tendo como base os trabalhos publicados utilizando diferentes metodologias de análise fosfoproteômica em outras espécies vegetais, como *Arabidopsis*, arroz e milho, por exemplo. Em plantas, os estudos fosfoproteômicos foram iniciados sobre a planta modelo *Arabidopsis thaliana*, utilizando a abordagem baseada no enriquecimento de fosfopeptídeos por IMAC, identificando um número relativamente pequeno de fosfopeptídeos (Nuhse *et al.*, 2004). Com o desenvolvimento de novas metodologias de enriquecimento, bem como a otimização dos protocolos já existentes e melhorias técnicas dos espectrômetros de massas, o número de fosfopeptídeos identificados em tecidos de *Arabidopsis* tem aumentado progressivamente de algumas centenas para mais de 5.000 fosfopeptídeos (Benschop *et al.*, 2007; Sugiyama *et al.*, 2008; Reiland *et al.*, 2009; Kline *et al.*, 2010; Nakagami *et al.*, 2010), contribuindo para avanço no conhecimento sobre a influência dessa modificação pós-traducional em processos fisiológicos em plantas. Utilizando amostras de suspensões celulares derivadas de sementes de arroz (*Oryza sativa japonica*) e caules de *Arabidopsis thaliana* e uma combinação dos métodos IMAC e uma variação do método MOC, denominado HAMMOC (*hydroxy acid-modified metal oxide chromatography*) (Sugiyama *et al.*, 2007), Nakagami *et al.* (2010) identificaram aproximadamente 7.000 e 5.000 fosfopeptídeos em amostras de arroz e *Arabidopsis*, respectivamente, e, demonstraram a identificação de

sítios conservados de fosforilação entre proteínas ortólogas. Estudos voltados ao fosfoproteoma do arroz, investigando os mecanismos envolvidos no desenvolvimento da antera (Ye *et al.*, 2015) e no processo de germinação e desenvolvimento de sementes (Han *et al.*, 2014; Li *et al.*, 2015; Qiu *et al.*, 2015), revelaram o fundamental papel da fosforilação nesses processos. As respostas ao estresse abiótico, causado pela salinidade e déficit hídrico em folhas de milho (*Zea mays*), foram estudadas de acordo com a população de fosfoproteínas por Hu *et al.* (2013) e Bonhomme *et al.* (2012), respectivamente, enquanto Facette *et al.* (2013), utilizando abordagens de quantificação *label-free*, avaliaram paralelamente o proteoma e o fosfoproteoma em diferentes zonas de desenvolvimento de folhas de milho, identificando aproximadamente 12.000 proteínas e 14.000 sítios de fosforilação, demonstrando que o evento de fosforilação atua diferentemente em quatro diferentes estágios de desenvolvimento da folha. Dessa forma, diante dos recentes desenvolvimentos em metodologias de análise de fosfoproteínas aliados ao aumento do número de sequências de transcritos de diferentes tecidos de cana-de-açúcar nos bancos de dados transcriptômicos, a análise do fosfoproteoma direcionada aos tecidos de cana-de-açúcar seria agora exequível.

A Proteômica, como área da ciência, foi amplamente beneficiada pela introdução da espectrometria de massas na análise de biomoléculas. Hoje, o conjunto de ferramentas da proteômica representa o alicerce central das análises de diversos proteomas e, ao longo das últimas décadas, tem sido favorecido por um amplo desenvolvimento tecnológico que permitiu o aumento da resolução, acurácia de massa, sensibilidade e aumento da taxa de aquisição de espectros (Zhang *et al.*, 2013). Além disso, o desenvolvimento de métodos de preparação e separação de amostras de

proteínas/peptídeos, o acesso à diferentes bancos de dados de sequências derivadas de estudos transcriptômicos e genômicos e o desenvolvimento de ferramentas bioinformáticas para a análise de peptídeos e proteínas têm contribuído de forma substancial com o aumento do número de peptídeos e proteínas identificados e no aumento da precisão na quantificação de proteínas em amostras complexas (Agrawal *et al.*, 2009; Cox e Mann, 2008; Cox *et al.*, 2014). Dessa forma, a Proteômica se tornou uma área de fronteira altamente valiosa. A natureza dinâmica do proteoma e o papel das proteínas como componentes efetores dos processos celulares, participando por exemplo da catálise enzimática e dos processos de sinalização celular, posiciona a Proteômica em lugar de destaque no estudo em larga-escala das funções e da expressão de genes em diferentes sistemas submetidos à diferentes condições ou em diferentes estágios de desenvolvimento, com o objetivo de contribuir para o aumento do conhecimento sobre os processos celulares que governam ou influenciam eventualmente as características fenotípicas.

2. OBJETIVOS

O objetivo geral deste estudo foi analisar os proteomas de folhas +1 de duas espécies ancestrais de cana-de-açúcar, *S. officinarum* e *S. spontaneum*, e da cultivar híbrida SP80-3280, utilizando abordagens proteômicas.

Foram propostos os objetivos específicos:

- 1 – Avaliar a eficiência de diferentes métodos de extração de proteínas de folhas de cana-de-açúcar;
- 2 – Avaliar a complexidade do proteoma de folhas +1 de *S. officinarum*, *S. spontaneum* e da cultivar híbrida SP80-3280;
- 3 – Comparar os proteomas das espécies ancestrais (*S. officinarum* e *S. spontaneum*) e híbrida (SP80-3280);
- 4- Avaliar os proteomas de folhas +1 de SP80-3280 coletadas em diferentes estágios de desenvolvimento da planta.

3. MATERIAIS E MÉTODOS

3.1. MATERIAL DE ESTUDO, ANÁLISE DE PARÂMETROS FOTOSSINTÉTICOS E TRATAMENTO INICIAL DAS AMOSTRAS

As amostras de folhas +1, provenientes das espécies ancestrais *S. officinarum* e *S. spontaneum*, e da cultivar híbrida SP80-3280, cultivadas em campo, foram coletadas no Centro de Ciências Agrárias da Universidade Federal de São Carlos (CCA – UFSCar – Araras) no dia 26 de junho de 2012. As plantas apresentavam idade de 12 meses no momento da coleta. Para cada espécie e cultivar, as folhas +1 foram coletadas de quinze plantas independentes e agrupadas em três réplicas biológicas, cada qual composta por um *pool* de cinco folhas +1. As réplicas biológicas foram designadas da seguinte forma: Réplica Biológica 1 de folhas +1 de *S. officinarum*: S.off RB1; Réplica Biológica 2 de folhas +1 de *S. officinarum*: S.off RB2; Réplica Biológica 3 de folhas +1 de *S. officinarum*: S.off RB3. De forma similar, as réplicas biológicas de *S. spontaneum* e SP80-3280 foram nomeadas de S.sp RB1 a RB3 e SP RB1 a RB3, respectivamente.

Outro conjunto de amostras de folhas +1, constituído somente por folhas da cultivar híbrida SP80-3280 cultivada em campo, foi coletado ao longo do ciclo de maturação da planta por um período de 13 meses em quatro tempos distintos – 4, 8, 11 e 13 meses após a data de plantio (20/04/2012). Estas foram agrupadas da mesma forma descrita para as amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 coletadas de plantas de 12 meses, constituindo três réplicas biológicas por tempo de coleta, cada uma composta por um *pool* de duas folhas +1. Assim, as amostras de folhas +1 de SP80-3280 coletadas em quatro estágios de desenvolvimento da planta foram

nomeadas da seguinte forma: SP 4m – RB1, – RB2 e – RB3, correspondendo às réplicas biológicas 1, 2 e 3 da coleta realizada de plantas de 4 meses; SP 8m – RB1, – RB2 e – RB3, correspondendo às réplicas biológicas 1, 2 e 3 da coleta realizada de plantas de 8 meses; SP 11m – RB1, – RB2 e – RB3, correspondendo às réplicas biológicas 1, 2 e 3 da coleta realizada de plantas de 11 meses e; SP 13m – RB1, – RB2 e – RB3, correspondendo às réplicas biológicas 1, 2 e 3 da coleta realizada de plantas de 13 meses.

Antes da coleta, as folhas +1 foram submetidas à análise de eficiência de fotossíntese utilizando um sistema de medida de fotossíntese portátil LCI (ADC Bioscientific) entre 10 h e 12 h sob intensidade de luz artificial de $1.500 \mu\text{mol m}^{-2} \text{s}^{-1}$. Os parâmetros fisiológicos medidos pelo instrumento foram a taxa fotossintética (A), a condutância estomática (G_s), a taxa de transpiração (E) e a eficiência no uso de água (WUE). Após a coleta e limpeza das folhas, estas foram imediatamente congeladas em nitrogênio líquido e armazenadas a -80°C e, posteriormente maceradas com o auxílio de um moinho de bolas modelo TE-350 (TECNAL) e armazenadas a -80°C até o momento do uso. Os dados obtidos de A , G_s , E e WUE foram submetidos à análise de variância *one-way* ANOVA ($p\text{-value} < 0,05$) pelo software PASW Statistics v.17.0.2 e os valores médios foram comparados aplicando-se o teste de Tukey ($p\text{-value} < 0,05$).

3.2. EXTRAÇÃO DE PROTEÍNAS

A extração de proteínas presentes nas amostras de folhas +1 de cana-de-açúcar foi realizada a partir da utilização dos métodos descritos a seguir.

3.2.1. MÉTODO DE EXTRAÇÃO TCA/ACETONA

As folhas maceradas foram pesadas e submetidas ao processo de extração de proteínas, segundo método descrito por Wu e Wang (1984), com algumas modificações.

A amostra de folha +1 (200 mg) foi transferida para um tubo de centrifugação com capacidade volumétrica de 2 mL e a este foram adicionados 1,8 mL de solução de precipitação [ácido tricloroacético (TCA) 10% (m/v), 2-mercaptoetanol 0,07% (v/v) em acetona gelada]. Após agitação, a mistura foi mantida sob refrigeração a -20°C por 1 hora e em seguida centrifugada a 14.000 x g, por 20 minutos a 4°C. O sobrenadante foi removido e, ao material precipitado foram adicionados 1,8 mL de solução de lavagem [2-mercaptoetanol 0,07% (v/v) em acetona gelada]. A seguir, a amostra foi novamente submetida à refrigeração a -20°C por 1 hora, centrifugada a 14.000 x g por 20 minutos a 4°C e o sobrenadante foi descartado. O processo de lavagem foi repetido mais duas vezes e ao final do processo o material precipitado foi submetido à secagem em *speed-vac* para eliminação da acetona residual. A amostra foi armazenada a -80°C até o momento do uso.

3.2.2. MÉTODO DE EXTRAÇÃO FENOL

As folhas maceradas foram pesadas e submetidas ao processo de extração de proteínas, segundo método descrito por Hurkman e Tanaka (1986), com algumas modificações.

A amostra de folha +1 (200 mg) foi transferida para um tubo de centrifugação com capacidade volumétrica de 2 mL e em seguida foram adicionados 0,6 mL de solução de extração [Tris-HCl 500 mM, ácido etilenodiamino tetra-acético (EDTA) 50 mM, sacarose

700 mM, KCl 100 mM, 2-mercaptoetanol 2%, pH 8,0] acrescida de coquetel de inibidores de proteases e fosfatases (Halt™, Thermo Fisher Scientific), conforme instruções do fabricante. A amostra foi submetida à agitação em banho de gelo por 10 minutos. A seguir foram adicionados 0,6 mL de fenol equilibrado pH 7,5 – 7,8 (Invitrogen) e a amostra foi submetida à incubação por 10 minutos sob agitação à temperatura ambiente. Após essa etapa, a amostra foi centrifugada a 5.500 x g por 10 minutos a 4°C. A fase superior (fase fenol) foi coletada em um novo tubo e, sobre essa fase foram adicionados 0,6 mL de solução de extração. A mistura foi submetida à agitação e centrifugada novamente sob as mesmas condições descritas. A fase superior foi novamente coletada e transferida para um novo tubo e, sobre esta foram adicionados 4 volumes de solução de precipitação (acetato de amônio 0,1 M em metanol gelado). A seguir, a mistura foi submetida à incubação por um período de 18 horas a -20°C. No dia seguinte a amostra foi centrifugada a 5.500 x g por 10 minutos, a 4°C, e o sobrenadante foi removido. O precipitado obtido foi lavado três vezes com 1,5 mL de solução de precipitação. Ao final de cada etapa de lavagem, a amostra foi centrifugada a 5.500 x g por 5 minutos, a 4°C. Após a remoção do sobrenadante, o precipitado foi submetido a mais duas etapas de lavagem com acetona gelada e a amostra foi novamente submetida à centrifugação. O sobrenadante foi descartado e o precipitado foi submetido à secagem em *speed-vac* para eliminação da acetona residual. A amostra foi armazenada a -80°C até o momento do uso.

3.2.3. MÉTODO DE EXTRAÇÃO FENOL/SDS

As folhas maceradas foram pesadas e submetidas ao processo de extração de proteínas segundo método descrito por Wang *et al.* (2003), com algumas modificações.

A amostra de folha +1 (200 mg) foi transferida para um tubo de centrifugação com capacidade volumétrica de 2 mL e a este foram adicionados 1,5 mL de acetona gelada. A amostra foi agitada e mantida sob refrigeração por 5 minutos e a seguir, centrifugada a 10.000 x g por 5 minutos, a 4°C. O sobrenadante foi removido e, ao material precipitado, foram adicionados 1,5 mL de acetona gelada e a amostra submetida à agitação em banho de gelo por 5 minutos e em seguida centrifugada a 10.000 x g por 5 minutos, a 4°C. Esse procedimento foi repetido mais uma vez. O material precipitado resultante foi lavado mais sete vezes com 1,5 mL de uma solução de TCA 10% (m/v) em acetona gelada e por fim, duas vezes com acetona 80%. O material precipitado foi submetido à secagem em *speed-vac* e, posteriormente foram adicionados 0,8 mL de tampão SDS [sacarose 30% (m/v), dodecil sulfato de sódio (SDS) 2% (m/v), Tris-HCl 0,1 M, pH 8,0, 2-mercaptoetanol 5%] acrescido de coquetel de inibidores de proteases e fosfatases (Halt™, Thermo Fisher Scientific), conforme instruções do fabricante. A mistura foi submetida à incubação em banho de gelo sob agitação por 10 minutos. Após essa etapa, foram adicionados 0,8 mL de fenol equilibrado pH 7,5 – 7,8 (Invitrogen) e a amostra foi submetida à incubação por 10 minutos sob agitação, à temperatura ambiente. A amostra foi centrifugada por 5 minutos a 10.000 x g a 4°C e a fase fenol foi cuidadosamente recolhida em um novo tubo e, sobre esta, foram adicionados 5 volumes de uma solução de acetato de amônio 0,1 M em metanol gelado. As proteínas foram submetidas à precipitação por 18 horas a -20°C. Após essa etapa, a amostra foi centrifugada a 10.000

x g, a 4°C por 5 minutos e o sobrenadante foi descartado. O precipitado foi lavado duas vezes com a solução de acetato de amônio 0,1 M em metanol gelado e duas vezes com acetona 80% gelada. Em todos os passos de lavagem a amostra foi centrifugada por 5 minutos a 10.000 x g, 4°C. O precipitado foi submetido à secagem em *speed-vac* e guardado a -80°C até o momento do uso.

3.3. SOLUBILIZAÇÃO DAS AMOSTRAS DE PROTEÍNAS

As amostras extraídas de acordo com o método TCA/Acetona foram solubilizadas em uma solução contendo Tris-HCl 50 mM, pH 7,5, SDS 0,1% (m/v), EDTA 25 mM e coquetel de inibidores de proteases e fosfatases (Halt™, Thermo Fisher Scientific) de acordo com instruções do fabricante para permitir a solubilização das proteínas. As amostras proteicas foram submetidas à agitação e centrifugação a 14.000 x g por 20 minutos, a 4°C. O sobrenadante foi transferido para um novo tubo.

As amostras extraídas de acordo com os métodos Fenol e Fenol/SDS foram dissolvidas em tampão *Lysis Buffer Thiourea* (LBT), composto por uréia 7 M, tiouréia 2 M, 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS) 4% (m/v) e DTT 0,3% (m/v) acrescido de coquetel de inibidores de proteases e fosfatases (Halt™, Thermo Fisher Scientific) de acordo com instruções do fabricante.

3.4. QUANTIFICAÇÃO DE PROTEÍNAS

Após dissolução das amostras de proteínas extraídas pelos métodos de descritos acima, estas foram quantificadas de acordo com o método de Bradford (Bradford *et al.*, 1976). A curva padrão foi construída utilizando-se concentrações crescentes de soroalbumina bovina (BioRad). As leituras de absorbância foram realizadas em espectrofotômetro (Gene Quant – GE Healthcare) no comprimento de onda de 595 nm.

3.5. ELETROFORESE EM GEL DE SDS-POLIACRILAMIDA

A eletroforese em gel de SDS-poliacrilamida foi realizada de acordo com o método descrito por Laemmli (1970), utilizando mini-géis de 10 cm x 8 cm x 0,75 mm em cuba vertical (Electrophoresis System Hoefer™ miniVE). A eletroforese foi conduzida à temperatura ambiente, utilizando fonte elétrica *Electrophoresis Power Supply* EPS 601 (GE Healthcare). A voltagem e a corrente aplicada foram de 150 V e 40 mA/gel, respectivamente. O gel de SDS-poliacrilamida a 12% continha tampão Tris-HCl 0,38 M, pH 8,8, SDS 0,1% (m/v), acrilamida 12% (m/v), bisacrilamida 0,3% (m/v), N,N,N',N'-Tetrametil etileno-diamino (TEMED) 0,1% (v/v) e persulfato de amônio 0,06% (m/v). Após a polimerização do gel de separação, um gel superior de empacotamento foi preparado com acrilamida 4% (m/v) e bisacrilamida 0,1% (m/v), tampão Tris-HCl 0,1 M, pH 6,8, TEMED 0,1% (v/v) e persulfato de amônio 0,06% (m/v). As amostras foram submetidas à desnaturação e redução das pontes dissulfeto em tampão Tris-HCl 0,125 M, pH 6,8, contendo 2-mercaptoetanol 2% (v/v) e SDS 2% (m/v), aquecidas a 95°C durante 5 minutos e aplicadas sobre o gel de empacotamento. O tampão de corrida era composto

de Tris-HCl 0,025 M, glicina 0,18 M, pH 8,3, e SDS 1% (m/v) e, como padrão de massa molecular, foi utilizada uma mistura de proteínas, contendo albumina bovina (68 kDa), cadeias pesada e leve da IgG (50 e 24 kDa, respectivamente), ovoalbumina (43 kDa) e ribonuclease (14 kDa).

3.6. COLORAÇÃO PARA PROTEÍNAS TOTAIS

Após a eletroforese em gel de SDS-poliacrilamida, as bandas de proteínas foram reveladas de acordo com Neuhoﬀ *et al.* (1988) utilizando solução de azul de Coomassie G-250 ou a partir do método de coloração com nitrato de prata, de acordo com Mortz *et al.* (2001), como descrito abaixo.

3.6.1. MÉTODO DE COLORAÇÃO COM COOMASSIE G-250

Os géis de SDS-poliacrilamida foram mergulhados por um período de 10 horas em uma solução contendo Coomassie G-250 (*Pierce Coomassie Brilliant Blue G-250 Dye* - Thermo Fisher Scientific) 0,08% (m/v), metanol 25% (v/v), ácido fosfórico 1,6% (v/v) e sulfato de amônio 8% (m/v). Após essa etapa, os géis foram submetidos à descoloração a partir de sucessivas lavagens com água.

3.6.2. MÉTODO DE COLORAÇÃO COM NITRATO DE PRATA

Os géis foram colocados na solução de fixação (metanol 50%, ácido acético 12% e formaldeído 0,05%) por 18 horas. Em seguida os géis foram lavados três vezes com uma solução de etanol 35%, por 20 minutos, e incubados em solução de tiosulfato de sódio 0,02% (m/v), por 2 minutos. Após esse período os géis foram lavados duas vezes com água por 5 minutos e posteriormente incubados em solução de nitrato de prata 0,2% (m/v) contendo formaldeído 0,076%, por 20 minutos. Após duas lavagens com água por 1 minuto, os géis foram colocados em uma solução contendo carbonato de sódio 6% (m/v), formaldeído 0,05% e tiosulfato de sódio 0,0004% (m/v) até se atingir a coloração desejada das bandas de proteínas. Para interromper a reação de coloração utilizou-se uma solução de metanol 50% e ácido acético 12%.

3.7. DIGESTÃO TRIPSÍNICA *IN GEL*

Os géis de SDS-poliacrilamida foram recortados em regiões de interesse e os fragmentos foram submetidos ao processo de digestão tripsínica *in gel* segundo o protocolo descrito por Hanna *et al.* (2000).

Inicialmente, os fragmentos de gel recortados foram incubados em 500 µL de uma solução de metanol 50%/ácido acético 5% em água, por duas horas. Em seguida, essa solução foi removida por aspiração com pipeta e foram adicionados mais 500 µL da mesma solução aos fragmentos, que permaneceram por mais uma hora nesta solução. Em seguida, os fragmentos de gel foram desidratados pela incubação em 200 µL de acetonitrila (ACN) 100%, por 10 minutos (2 vezes de 5 minutos). A solução de ACN foi

aspirada com pipeta e o restante foi evaporado em *speed-vac*. Após esta etapa, os pedaços de gel foram reidratados por 30 minutos em 30 μ L da solução redutora [ditiotreitól (DTT) 10 mM em solução de bicarbonato de amônio 100 mM]. Decorrido o tempo de reidratação com DTT, a solução foi aspirada com pipeta e os fragmentos de gel foram incubados por 30 minutos em 30 μ L da solução alquilante [iodoacetamida (IAA) 50 mM em solução de bicarbonato de amônio 100 mM]. Em seguida, a solução alquilante foi removida por aspiração com pipeta e, os pedaços de gel foram submetidos à incubação em 100 μ L de uma solução de bicarbonato de amônio 100 mM por 10 minutos, à temperatura ambiente. Após esta etapa, a solução de bicarbonato de amônio foi retirada e 200 μ L de ACN 100% foram adicionados sobre os fragmentos de gel, que foram incubados por 5 minutos. A solução de ACN foi removida e os fragmentos foram submetidos à incubação por 10 minutos em 200 μ L da solução de bicarbonato de amônio 100 mM. Na última etapa de desidratação, a solução de bicarbonato de amônio foi retirada e os fragmentos foram incubados por 10 minutos (duas vezes de 5 minutos) em 200 μ L de ACN 100%. A solução de ACN foi retirada por aspiração com pipeta e o restante foi evaporado em *speed-vac*. Os fragmentos de gel foram então reidratados em 5 μ L de uma solução de tripsina (Sigma-Aldrich; 50 ng/ μ L em bicarbonato de amônio 50 mM), em banho de gelo, por 30 minutos. Em seguida foi adicionada uma solução de bicarbonato de amônio 50 mM em volume suficiente para cobrir os pedaços de gel, que foram incubados em banho termostatizado a 37°C por 18 horas.

A extração dos peptídeos do gel foi realizada com a adição de 30 μ L de uma solução de ácido fórmico (AF) 5% (solução de extração 1) e incubação por 10 minutos, à temperatura ambiente. Em seguida, a solução foi transferida para um novo tubo e,

sobre os fragmentos de gel foram adicionados 30 μ L de uma solução composta de AF 5% e ACN 50% (solução de extração 2). Após incubação por 20 minutos, a solução foi removida e transferida para o mesmo tubo contendo a solução de extração 1. A etapa de incubação dos fragmentos de gel com a solução de extração 2 foi repetida mais uma vez e as amostras de peptídeos extraídos foram submetidas à secagem em *speed-vac* e armazenadas a -80°C até o momento do uso.

3.8. DIGESTÃO TRIPSÍNICA EM SOLUÇÃO

As proteínas extraídas das amostras de folhas +1 de cana-de-açúcar foram submetidas à digestão tripsínica em solução conforme o protocolo adaptado de Wiśniewski *et al.* (2009), conhecido como *Filter-Aided Sample Preparation* (FASP).

As amostras, contendo 0,4 mg de proteínas, foram adicionadas em 200 μ L de uma solução de uréia 8 M em Tris-HCl 0,1 M pH 8,5 (solução UA) sobre aparatos de filtragem Microcon YM-10 (Millipore) previamente condicionados de acordo com as recomendações do fabricante. Após leve agitação, os aparatos foram centrifugados a 14.000 x g, por 15 minutos, seguidos de nova adição de 200 μ L de UA e centrifugação sob a mesma condição anterior. O volume coletado foi descartado do tubo coletor e, sobre o filtro, foram adicionados 100 μ L de solução de DTT 0,02 M (solubilizado em UA), e após leve agitação as amostras foram incubadas por 30 minutos à temperatura ambiente. Após a incubação com DTT, os aparatos foram centrifugados a 14.000 x g por 10 minutos e então foram adicionados 100 μ L de solução de IAA 0,05 M (solubilizada em UA). Após leve agitação, as amostras foram incubadas à temperatura ambiente por 30

minutos no escuro e então centrifugadas a 14.000 x g por 10 minutos. Os filtros foram então lavados três vezes com 100 µL de UA para remoção do excesso de DTT e IAA. Em cada etapa de lavagem os aparatos foram centrifugados a 14.000 x g por 15 minutos. Após o processo de lavagem, aos aparatos, foram adicionados 100 µL de solução de NH_4HCO_3 0,05 M em água (solução ABC) e então estes foram centrifugados a 14.000 x g por 10 minutos. Esse passo foi repetido mais duas vezes. Prosseguiu-se com a adição de 60 µL de solução ABC sobre os aparatos e a adição de tripsina (Sigma-Aldrich) na proporção de 1:50 (enzima:substrato; m/m). Os aparatos de filtragem foram submetidos à leve agitação e incubação a 37°C por 18 horas. Após o processo de digestão, os aparatos de filtragem foram transferidos para outro tubo coletor e submetidos à centrifugação a 14.000 x g por 10 minutos. Adicionaram-se mais 50 µL de solução ABC e a centrifugação a 14.000 x g foi repetida por 10 minutos. As amostras correspondentes à mistura de peptídeos tripsínicos foram acidificadas com ácido trifluoracético (TFA) até $\text{pH} \leq 3$ para posterior dessalinização.

3.9. REMOÇÃO DE SAIS PRESENTES NAS AMOSTRAS DE PEPTÍDEOS

As amostras de peptídeos resultantes do processo de digestão tripsínica *in gel* foram submetidas ao processo de dessalinização utilizando *Stop-and-go extraction tip* (StageTip) C-18, montados em ponteiros Tip p-200 (Axygen), de acordo com o método descrito por Rappsilber *et al.* (2007). As amostras dessalinizadas foram dissolvidas em

ácido fórmico 0,1% para as análises de cromatografia líquida acoplada à espectrometria de massas (LC-MS/MS).

Com relação às amostras de peptídeos provenientes do processo de digestão tripsínica em solução, estas, após acidificadas com TFA até $\text{pH} \leq 3$, foram submetidas ao processo de dessalinização utilizando colunas de extração em fase sólida Sep-Pak Light C-18 (Waters). Os peptídeos foram eluídos com 2 mL de uma solução contendo ACN 50% em TFA 0,1% e concentrados em *speed-vac*.

As amostras de peptídeos dessalinizados provenientes das digestões tripsínicas *in gel* e em solução foram posteriormente dissolvidas em ácido fórmico 0,1% para as análises de LC-MS/MS.

3.10. ESTRATÉGIAS DE ANÁLISE PROTEÔMICA DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280

Após a análise comparativa dos métodos de extração de proteínas aplicados às folhas +1 de cana-de-açúcar, os tecidos de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 coletadas de plantas de 12 meses, bem como de folhas +1 coletadas ao longo do desenvolvimento da planta SP80-3280, foram submetidos ao processo de extração de proteínas de acordo com o método Fenol/SDS. As proteínas extraídas foram posteriormente submetidas às diferentes estratégias de análises proteômicas esquematizadas na Figura 7.

Figura 7. Esquema ilustrativo dos métodos de análise utilizados para a caracterização dos proteomas e fosfoproteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280. As setas em cinza mostram as análises realizadas utilizando-se amostras de folhas +1 das espécies ancestrais e da cultivar híbrida coletadas de plantas de 12 meses; as setas na cor laranja representam as análises realizadas sobre as amostras de folhas +1 de SP80-3280 coletadas durante quatro estágios de desenvolvimento da planta. As análises de LC-MS/MS foram realizadas pelo sistema de nano-cromatografia líquida Easy-nLCII acoplado ao espectrômetro LTQ-Orbitrap Velos ou pelo sistema de nano-cromatografia Dionex Ultimate 3000 acoplado ao espectrômetro de massas Q-Exactive.

As amostras provenientes de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de plantas de 12 meses foram analisadas segundo três estratégias de análise proteômica: a primeira, baseada na metodologia *shotgun/bottom-up proteomics*; a

segunda, baseada na identificação de peptídeos derivados do processo de digestão tripsínica *in gel* de proteínas separadas em gel de SDS-poliacrilamida, abordagem conhecida como GeLC-MS/MS e; a terceira, baseada na metodologia *Multidimensional Protein Identification Technology* – MudPIT (Link *et al.*, 1999; Washburn *et al.*, 2001) *off-line* e “*on-line*”. Sobre o mesmo conjunto de amostras, foi realizado o estudo fosfoproteômico utilizando a estratégia metodológica de enriquecimento da população de fosfopeptídeos baseada na afinidade dos grupos fosfatos pelo cátion metálico Fe^{3+} , técnica conhecida como *Immobilized Metal Affinity Chromatography* – IMAC (Ficarro *et al.*, 2009) e posterior análise e identificação da população de fosfopeptídeos por LC-MS/MS. As diferentes estratégias utilizadas para o estudo dos proteomas e fosfoproteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de plantas de 12 meses podem ser visualizadas pelas linhas em cinza, no fluxo de trabalho simplificado, ilustrado na Figura 7.

Em relação aos proteomas de folhas +1 de SP80-3280 coletadas ao longo do desenvolvimento da planta, estes foram analisados segundo a estratégia *shotgun/bottom-up proteomics*. A população de fosfoproteínas, por sua vez, foi analisada a partir da utilização de dois métodos de enriquecimento de fosfopeptídeos. O primeiro foi o método IMAC e o segundo, conhecido como *Hydroxy Acid-Modified Metal Oxide Chromatography* (HAMMOC) (Nakagami, 2014), é baseado na afinidade dos peptídeos fosforilados por óxidos metálicos e posterior análise e identificação dos fosfopeptídeos por LC-MS/MS. As estratégias utilizadas para o estudo dos proteomas e fosfoproteomas de folhas +1 de SP80-3280 coletadas ao longo do desenvolvimento da planta podem ser visualizadas pelas linhas em laranja no fluxo de trabalho simplificado, ilustrado na Figura

7, e a descrição dessas estratégias experimentais utilizadas no estudo dos proteomas e fosfoproteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP803280 é descrita em detalhes abaixo.

3.10.1. ANÁLISE DOS PROTEOMAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280 POR DIGESTÃO TRIPSÍNICA EM SOLUÇÃO E ANÁLISE POR LC-MS/MS (ESTRATÉGIA *SHOTGUN/BOTTOM-UP PROTEOMICS*)

Os peptídeos provenientes do processo de digestão tripsínica em solução das proteínas extraídas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de plantas de 12 meses foram analisados no espectrômetro LTQ Orbitrap Velos (Thermo Scientific) a partir da abordagem de *shotgun/bottom-up proteomics*. Cada réplica biológica foi submetida a duas análises independentes de LC-MS/MS, resultando em um número total de 18 análises. Alíquotas equivalentes a 5 µg de peptídeos foram automaticamente injetadas em uma pré-coluna C-18 Jupiter (Phenomenex; diâmetro das esferas 10 µm; 100 µm ID x 50 mm) por um sistema de nano-cromatografia líquida Easy-nLCII (Thermo Scientific). Os peptídeos foram submetidos à cromatografia em fase reversa utilizando uma coluna analítica C-18 AQUA (Phenomenex; diâmetro das esferas 5 µm; 75 µm ID x 100 mm) confeccionada no próprio laboratório. Os peptídeos foram separados por um gradiente linear de 7 – 35% B (AF 0,1% em ACN) durante 95 minutos, sob fluxo de 200 nL/min. A fonte de ionização por *electrospray* foi ajustada para o modo de operação positivo, com a voltagem e temperatura ajustadas para 2,2 kV e 200°C, respectivamente. O intervalo de varredura de massas considerado para o *full scan* foi de 300 - 1650 m/z (resolução de 60.000 em 400 m/z), operando no modo de aquisição dependente de dados (DDA), onde os quinze íons mais intensos por *scan* foram

selecionados para o evento de fragmentação por dissociação induzida por colisão (CID) com energia de colisão normalizada de 35% e janela de isolamento de 2 m/z para o íon precursor. O sinal mínimo requerido para disparar eventos de fragmentação de determinado íon foi ajustado para 5.000 cps, e o tempo de exclusão dinâmica utilizado foi 90 segundos.

3.10.2. ANÁLISE DOS PROTEOMAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280 POR ELETROFORESE EM GEL DE SDS-POLIACRILAMIDA E ANÁLISE POR LC-MS/MS (ESTRATÉGIA GELC-MS/MS)

As proteínas (25 µg) extraídas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de plantas de 12 meses foram submetidas ao processo de desnaturação e redução das pontes dissulfeto e analisadas por eletroforese em gel de SDS-poliacrilamida 12%, de dimensões 10 cm x 8 cm x 0,75 mm sob voltagem de 150 V por aproximadamente 1,5 horas, segundo Laemmli (1970).

Após o término da eletroforese, as bandas de proteínas correspondentes às amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 foram reveladas de acordo com Neuhoff *et al.* (1988) e as faixas do gel correspondentes às amostras individuais foram divididas e seccionadas em 8 segmentos por amostra. Em seguida, as proteínas presentes nos segmentos de gel foram submetidas à digestão tripsínica *in gel* de acordo com o método descrito por Hanna *et al.* (2000) e as amostras de peptídeos resultantes do processo de digestão foram dessalinizadas utilizando *StageTip* C-18 e submetidas à secagem em *speed-vac*. Após a completa secagem das amostras de peptídeos, estes foram reconstituídos em 20 µL de ácido fórmico 0,1% para as análises por LC-MS/MS.

As amostras de peptídeos foram submetidas à análise por LC-MS/MS utilizando o espectrômetro de massas LTQ-Orbitrap Velos (Thermo Scientific) acoplado ao sistema de nano-cromatografia líquida Easy-nLCII (Thermo Scientific). Cinco microlitros de cada amostra foram injetados automaticamente em uma pré-coluna C-18 Jupiter (Phenomenex; diâmetro das esferas 10 μm ; 100 μm ID x 50 mm) acoplada a uma coluna analítica C-18 (Thermo Scientific; diâmetro das esferas 3 μm ; 75 μm ID x 100 mm). Os peptídeos foram submetidos à cromatografia em fase reversa utilizando um gradiente linear de 5-40% B (AF 0,1% em ACN) durante 30 minutos sob fluxo de 300 nL/min. A fonte de ionização por *electrospray* (ESI *source*) foi ajustada para o modo de operação positivo, com a voltagem e temperatura ajustadas para 2,2 kV e 200°C, respectivamente. O intervalo de varredura de massas considerado para o *full scan* foi de 200 - 2000 m/z (resolução de 30.000 em 400 m/z), operando no modo DDA, onde os dez íons mais intensos por *scan* foram selecionados para o evento de fragmentação por CID com energia de colisão normalizada de 35% e janela de isolamento de 2 m/z para o íon precursor. O sinal mínimo requerido para disparar eventos de fragmentação de determinado íon foi ajustado para 5.000 cps, e o tempo de exclusão dinâmica foi ajustado para 15 segundos.

3.10.3. ANÁLISE DOS PROTEOMAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280 POR DIGESTÃO TRIPSÍNICA EM SOLUÇÃO, CROMATOGRAFIA ORTOGONAL (MUDPIT) E ANÁLISE POR LC-MS/MS

As amostras de peptídeos tripsínicos provenientes das digestões em solução das amostras de proteínas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 foram submetidas ao processo de fracionamento de acordo com os métodos MudPIT *off-line* e “*on-line*” (Fig. 8). Os métodos de fracionamento de peptídeos por cromatografia líquida bidimensional descritos a seguir foram adaptados dos métodos originais descritos por Link *et al.* (1999), Washburn *et al.* (2001); Peng *et al.* (2003), Wolters *et al.* (2001) e Magdeldin *et al.* (2014).

Figura 8. Métodos de cromatografia líquida bidimensional utilizados no estudo dos proteomas de folhas +1 de cana-de-açúcar. As amostras de peptídeos foram submetidas ao fracionamento utilizando colunas trifásicas (C-18/SCX/C-18) ou bifásicas (C-18/SCX). As frações foram coletadas em tubos, injetadas automaticamente pelo *autosampler* do sistema de nano-HPLC e então analisadas por LC-MS/MS a partir da abordagem conhecida como MudPIT *off-line* (A). Em outra abordagem, nomeada MudPIT *on-line* (B), as amostras de peptídeos foram submetidas à separação ortogonal em coluna bifásica C-18/SCX. As frações foram coletadas diretamente em pré-colunas C-18 que foram subsequentemente acopladas ao sistema de nano-HPLC e então submetidas à análise de LC-MS/MS.

3.10.3.1. MUDPIT OFF-LINE

As amostras de peptídeos (100 µg) gerados pela digestão tripsínica em solução foram injetadas em colunas capilares trifásicas compostas por uma camada de resina de *strong cation-exchange* (SCX) entre duas camadas de resina de C-18 (C-18/SCX/C-18) (Fig. 8A) empacotadas e condicionadas no próprio laboratório da seguinte forma: Inicialmente foi preparado o *frit* de kasil em uma extremidade de um capilar de especificações 250 µm ID x 360 µm OD. Para o preparo do *frit*, o capilar foi brevemente mergulhado em uma solução composta por 88 µL de kasil 1624 (PQ Corporation) e 18 µL de formamida (Sigma) e então submetido a aquecimento de 80°C por 5 minutos. Após essa etapa, o capilar foi cortado de modo que o *frit* apresentasse uma espessura de 1,5 mm e então submetido aos passos de lavagens com ácido acético 0,1% e ACN 100%. A partir da extremidade que contém o *frit*, o capilar foi empacotado com 20 mm de resina C-18 AQUA (Phenomenex; diâmetro das esferas 5 µm), 20 mm de SCX Luna (Phenomenex; diâmetro das esferas 5 µm) e novamente com 30 mm de C-18 AQUA. A coluna confeccionada dessa forma foi submetida aos passos de lavagens com ácido acético 0,1% e ACN 100% e condicionadas com 10 pmol de angiotensina II (Sigma-Aldrich). Após o condicionamento, as amostras de peptídeos foram injetadas na coluna com o auxílio de uma bomba pressurizada de nitrogênio e então a coluna foi acoplada ao sistema Accella Pump 600 (Thermo Scientific) constituído por uma bomba quaternária. O fluxo de entrada na coluna cromatográfica foi ajustado para aproximadamente 2 µL/min utilizando um sistema de divisão de fluxo (*split-flow*). Os solventes utilizados para a cromatografia ortogonal foram: A – AF 0,1%; B – ACN em AF 0,1%; C – acetato de amônio 250 mM, ACN 5% e AF 0,1%; D – acetato de amônio 1000

mM, ACN 5% e AF 0,1%. Os peptídeos foram inicialmente eluídos do primeiro segmento de coluna de C-18 (localizada na extremidade contrária ao *frit*) a partir de um gradiente linear de 5-80% solvente B em 30 minutos e a fração zero (ou *flow-through*; fração de peptídeos que não é retido pela coluna de SCX) foi coletada em um tubo. A eluição subsequente dos peptídeos foi realizada utilizando-se 17 concentrações crescentes de acetato de amônio por 15 minutos, correspondendo a 10, 20, 30, 40, 50, 100, 150, 170, 180, 190, 200, 225, 250, 300, 500, 750 e 1000 mM. Cada etapa de eluição com acetato de amônio foi seguida por dessalinização com solvente A por 15 minutos e eluição dos peptídeos do segundo segmento de resina C-18 utilizando um gradiente linear de 5-50% B por 20 minutos. As frações correspondentes foram coletadas em tubos e submetidas à secagem em *speed-vac* e dissolução em 20 μ L de solvente A. O tempo total de cromatografia bidimensional por amostra foi de 17 horas. Para o experimento de LC-MS/MS, os peptídeos foram automaticamente injetados em uma pré-coluna C-18 Jupiter (Phenomenex; diâmetro das esferas 10 μ m; 100 μ m ID x 50 mm) pelo sistema de nano-HPLC EASY nLCII (Thermo Scientific) acoplado ao espectrômetro LTQ Orbitrap Velos (Thermo Scientific). A seguir, os peptídeos foram submetidos à cromatografia em fase reversa utilizando uma coluna analítica C-18 AQUA (Phenomenex; diâmetro das esferas 5 μ m; 75 μ m ID x 100 mm) confeccionada no próprio laboratório e separados por um gradiente linear de 5 – 35% B durante 60 minutos sob fluxo de 200 nL/min. A fonte de ionização por *electrospray* foi ajustada para o modo de operação positivo, com a voltagem e temperatura ajustadas para 2,2 kV e 200°C, respectivamente. O intervalo de varredura de massas considerado para o *full scan* foi de 300 - 1650 m/z (resolução de 60.000 em 400 m/z), operando no modo DDA, onde os quinze íons mais intensos por

scan foram selecionados para o evento de fragmentação por CID com energia de colisão normalizada de 35% e janela de isolamento de 2 m/z para o íon precursor. O sinal mínimo requerido para disparar eventos de fragmentação de determinado íon foi ajustado para 5.000 cps, e o tempo de exclusão dinâmica utilizado foi 90 segundos.

3.10.3.2. MUDPIT “ON-LINE”

O fracionamento dos peptídeos também foi realizado segundo a abordagem de MudPIT “*on-line*”. Nesta abordagem, os peptídeos (100 µg) foram carregados com o auxílio de uma bomba pressurizada de nitrogênio em colunas bifásicas de C-18/SCX (Fig. 8B) confeccionadas de maneira similar às colunas trifásicas descritas anteriormente para a abordagem de MudPIT *off-line*, porém empacotadas com 35 mm de cada resina. A partir de um protocolo modificado de cromatografia líquida e montagem de um sistema de acoplamento de colunas na saída da coluna bifásica, as frações foram coletadas diretamente em pré-colunas C-18 Jupiter (30 mm; Phenomenex; diâmetro dos *beads* 10 µm), confeccionadas em capilares de especificações 100 µm ID x 360 µm OD. Os peptídeos foram inicialmente eluídos da coluna de C-18 utilizando um gradiente linear de 5-80% B em 25 minutos e a fração zero (ou *flow-through*) foi coletada em um tubo. Esta fração foi submetida à secagem em *speed-vac* e os peptídeos, dissolvidos em AF 0,1%, foram injetados em uma pré-coluna com o auxílio de uma bomba pressurizada de nitrogênio (esta foi reservada para posterior conexão ao sistema de cromatografia acoplado ao espectrômetro de massas). Após essa etapa, a coluna foi equilibrada com o solvente A por 5 minutos. Ao final do equilíbrio da coluna, procedeu-se com o primeiro

step de injeção de acetato de amônio (10 mM) com imediato acoplamento de uma pré-coluna C-18 à saída da coluna bifásica (Fig. 8B) para reter os peptídeos da primeira fração. Essa etapa foi executada ao longo de 15 minutos, e foi seguida de uma etapa de dessalinização por 10 minutos. O acoplamento de pré-colunas se deu de maneira subsequente e imediatamente ao início das injeções de concentrações crescentes de acetato de amônio (17 concentrações diferentes entre 10 mM e 1000 mM), de acordo com a descrição do método de MudPIT *off-line*. O processo de cromatografia ortogonal nessa abordagem tem duração de aproximadamente 8 horas. As pré-colunas, contendo os peptídeos eluídos em diferentes concentrações de acetato de amônio, foram acopladas ao sistema de nano-HPLC EASY nLCII (Thermo Scientific) acoplado ao espectrômetro de massas LTQ Orbitrap Velos (Thermo Scientific) e os peptídeos foram submetidos aos mesmos parâmetros de análise de LC-MS/MS descritos na análise de MudPIT *off-line*.

3.10.4. BUSCA EM BANCO DE DADOS E ANÁLISE BIOINFORMÁTICA DOS DADOS GERADOS PELAS ANÁLISES DOS PROTEOMAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280

Os dados provenientes das análises dos proteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 coletadas de plantas de 12 meses, de acordo com as diferentes estratégias experimentais, foram submetidos à busca em banco de dados a partir do algoritmo Andromeda (Cox *et al.*, 2011), ferramenta de busca contida no software MaxQuant v.1.5.3.12 (Cox e Mann, 2008), utilizando o banco de dados transcriptômico traduzido de cana-de-açúcar contendo 40.494 sequências representando 17.563 transcritos completos provenientes de amostras de folhas e

entrenós de *S. officinarum*, *S. spontaneum* e SP80-3280 (Nishiyama-Jr *et al.*, 2014). Os resultados de identificação de peptídeos e proteínas gerados pelo software MaxQuant foram analisados pelas ferramentas bioinformáticas contidas no software Perseus v.1.5.0.15.

3.10.4.1. ANÁLISE DOS DADOS GERADOS PELA ESTRATÉGIA *SHOTGUN/BOTTOM-UP* *PROTEOMICS*

Os dados brutos gerados pelas 18 análises de LC-MS/MS dos peptídeos gerados pelo processo de digestão tripsínica em solução foram submetidos à busca em banco de dados utilizando o software MaxQuant. Cada arquivo bruto foi tratado de forma independente para a identificação de proteínas. Com o intuito de aprimorar a identificação de proteínas, foi utilizada a ferramenta “*match between runs*”, que se utiliza das medidas de alta acurácia de massa e de tempo de retenção para transferir a identificação de um peptídeo sequenciado em determinada análise de LC-MS/MS e que foi detectado, porém não identificado em uma outra análise correlata (ou por não ter sido selecionado para o evento de fragmentação ou por apresentar baixa qualidade de espectros de fragmentação, impossibilitando seu sequenciamento). Nesse sentido, esta ferramenta alinha os perfis cromatográficos entre diferentes análises de LC-MS/MS intra- e inter-experimentos e busca a ocorrência de precursores de mesmo valor de m/z dentro de um intervalo de tempo de retenção. Em caso de adequação aos critérios de tempo de retenção e massa de um determinado peptídeo precursor detectado em diferentes análises e que eventualmente foi identificado em somente uma delas, a identificação é

transferida às análises em que estes precursores não foram confirmados, a partir da interpretação dos espectros de fragmentação.

Os parâmetros utilizados para a busca em banco de dados foram:

- Enzima: tripsina;
- Tolerância de duas clivagens perdidas pela enzima;
- Tolerância de erro de massa para o peptídeo precursor de 4,5 ppm;
- Tolerância de erro de massa para os fragmentos de MS/MS de 0,5 Da;
- Modificação fixa: carbamidometilação de resíduos de cisteínas;
- Modificações variáveis: oxidação dos resíduos de metionina e a deamidação em resíduos de asparagina ou glutamina;
- Taxa de falso positivo em nível de peptídeos e proteínas: 1%.

A taxa de falso positivo (FDR) foi calculada a partir das estimativas realizadas a partir da construção de um banco de dados *decoy*, elaborado por meio da ordenação reversa das sequências traduzidas presentes no banco de dados original. Um banco de dados contendo 245 sequências traduzidas dos principais contaminantes encontrados nas análises por espectrometria de massas (como por exemplo, queratina, albumina e tripsina) foi utilizado adicionalmente aos outros bancos de dados para identificação de potenciais contaminantes nas amostras. Sobre os dados de LC-MS/MS, foi aplicada a ferramenta de quantificação *label-free* MaxLFQ (Cox *et al.*, 2014) para a comparação entre as abundâncias relativas de proteínas presentes nas amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280. Para a quantificação relativa, foi levada em consideração a intensidade de pelo menos dois peptídeos para os cálculos das intensidades normalizadas.

Os resultados de identificação de peptídeos e proteínas gerados pelo software MaxQuant foram analisados pelas ferramentas bioinformáticas contidas no software Perseus. As proteínas identificadas pelo banco de dados de contaminantes e banco de dados *decoy* foram removidas. Como critério de identificação de proteínas, foram considerados somente os peptídeos identificados com Probabilidade de Erro Posterior (PEP) $\leq 0,01$ em pelo menos uma réplica biológica, a identificação mínima de oito íons pertencentes às séries -b ou -y nos espectros de MS/MS, e a ocorrência de pelo menos um peptídeo único. No caso específico de proteínas identificadas por somente um peptídeo, além dos critérios supracitados, foram aceitas somente as proteínas cujo peptídeo identificado apresentou a ocorrência ininterrupta mínima de pelo menos três íons da série -b ou -y.

Os valores de intensidades normalizados de proteínas presentes nas amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, calculadas pela ferramenta MaxLFQ, foram agrupados segundo suas réplicas e logaritmizados na base 2. Os valores foram filtrados de acordo com a ocorrência mínima de pelo menos três valores válidos de intensidade em pelo menos um grupo, e os valores não válidos foram substituídos pela ferramenta “*replace missing values from normal distribution*” (Deeb *et al.*, 2012; Cox *et al.*, 2014) utilizando os parâmetros *default*. Os valores de intensidades normalizados foram comparados de acordo com o teste estatístico *one-way* ANOVA com *p-value* ajustado para 0,05, segundo Benjamini-Hochberg (Benjamini e Hochberg, 1995) para o controle de falsos positivos. Os valores quantitativos das proteínas diferencialmente expressas (*p-value* < 0,05) foram normalizados de acordo com a ferramenta Z-score e estes foram submetidos à clusterização hierárquica baseada nas distâncias Euclidianas

entre suas médias. O conjunto de proteínas diferencialmente expressas nas três amostras de folhas +1 foi analisado segundo as categorias Processo Biológico, Componente Celular e Função Molecular do *Gene Ontology* (Ashburner *et al.* 2000), de acordo com a análise de enriquecimento de termos que atribui pesos proporcionais às categorias, considerando o número de ocorrências de categorias dentro da população total e dentro da população de estudo, e posicionando as categorias de acordo com sua representatividade na população de estudo. A representatividade de cada categoria foi calculada de acordo com as ferramentas contidas no software GeneMerge v.1.4 (Castillo-Davis e Hartl, 2003), considerando a taxa de falso positivo de 0,05%.

3.10.4.2. ANÁLISE DOS DADOS GERADOS PELA ESTRATÉGIA GELC-MS/MS

Os dados brutos (.raw) gerados pelas análises de LC-MS/MS dos peptídeos tripsínicos extraídos dos 24 segmentos de gel foram submetidos à busca em banco de dados. Inicialmente, foi realizado o carregamento dos arquivos brutos no ambiente de análise do software MaxQuant. Os peptídeos identificados nos diferentes segmentos de gel foram combinados para gerar as listas de proteínas identificadas em *S. officinarum*, *S. spontaneum* e SP80-3280. A ferramenta “*match between runs*” foi utilizada para otimizar a identificação de peptídeos e proteínas nas amostras, alinhando-se os perfis cromatográficos entre os experimentos e avaliando a ocorrência de precursores de mesmo valor de m/z.

Os parâmetros utilizados para a busca em banco de dados foram:

- Enzima: tripsina;

- Tolerância de duas clivagens perdidas pela enzima;
- Tolerância de erro de massa para o peptídeo precursor de 4,5 ppm;
- Tolerância de erro de massa para os fragmentos de MS/MS de 0,5 Da;
- Modificação fixa: carbamidometilação de resíduos de cisteínas;
- Modificações variáveis: oxidação dos resíduos de metionina e a deamidação em resíduos de asparagina ou glutamina;
- Taxa de falso positivo em nível de peptídeos e proteínas: 1%.

A taxa de falso positivo (FDR) foi calculada a partir das estimativas realizadas a partir da construção de um banco de dados *decoy*, elaborado por meio da ordenação reversa das sequências traduzidas presentes no banco de dados original. Um banco de dados contendo 245 sequências traduzidas dos principais contaminantes encontrados nas análises por espectrometria de massas foi utilizado adicionalmente aos outros bancos de dados para identificação de potenciais contaminantes nas amostras.

Os resultados de identificação de peptídeos e proteínas foram analisados pelas ferramentas bioinformáticas contidas no software Perseus, onde as proteínas identificadas pelo banco de dados de contaminantes e banco de dados *decoy* foram removidas. Como critério de identificação de proteínas, foram considerados somente os peptídeos identificados com Probabilidade de Erro Posterior (PEP) $\leq 0,01$, a identificação mínima de oito íons pertencentes às séries -b ou -y nos espectros de MS/MS e a ocorrência de pelo menos um peptídeo único. No caso específico de proteínas identificadas por somente um peptídeo, além dos critérios supracitados, foram aceitas somente as proteínas cujo peptídeo identificado apresentou a ocorrência ininterrupta mínima de pelo menos três íons da série -b ou -y.

3.10.4.3. ANÁLISE DOS DADOS GERADOS PELA ESTRATÉGIA MUDPIT *OFF-LINE* E “*ON-LINE*”

Os dados brutos gerados pelas análises de LC-MS/MS dos peptídeos tripsínicos de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, resultantes das 54 frações obtidas pela utilização da metodologia de MudPIT *off-line* e das 54 frações obtidas pela metodologia MudPIT “*on-line*” foram submetidos à busca em banco de dados, iniciando pelo carregamento dos 108 arquivos brutos no ambiente de análise do software MaxQuant. Os peptídeos identificados nas diferentes frações cromatográficas foram combinados para gerar as listas de proteínas identificadas em *S. officinarum*, *S. spontaneum* e SP80-3280, de acordo o método utilizado (*off-line* ou “*on-line*”). A ferramenta “*match between runs*” foi aplicada sobre os dados de LC-MS/MS correspondentes às frações iguais ou adjacentes intra- e inter-experimentos para o aprimoramento da identificação de proteínas.

Os parâmetros utilizados para a busca em banco de dados foram:

- Enzima: tripsina;
- Tolerância de duas clivagens perdidas pela enzima;
- Tolerância de erro de massa para o peptídeo precursor de 4,5 ppm;
- Tolerância de erro de massa para os fragmentos de MS/MS de 0,5 Da;
- Modificação fixa: carbamidometilação de resíduos de cisteínas;
- Modificações variáveis: oxidação dos resíduos de metionina e a deamidação em resíduos de asparagina ou glutamina;
- Taxa de falso positivo em nível de peptídeos e proteínas: 1%.

A taxa de falso positivo (FDR) foi calculada a partir das estimativas realizadas a partir da construção de um banco de dados *decoy*, elaborado por meio da ordenação reversa das sequências traduzidas presentes no banco de dados original. Um banco de dados contendo 245 sequências traduzidas dos principais contaminantes encontrados nas análises por espectrometria de massas foi utilizado adicionalmente aos outros bancos de dados para identificação de potenciais contaminantes nas amostras.

Os resultados de identificação de peptídeos e proteínas foram analisados pelas ferramentas bioinformáticas contidas no software Perseus, onde as proteínas identificadas pela busca no banco de dados de contaminantes e banco de dados *decoy* foram removidas. Como critério de identificação de proteínas, foram considerados somente os peptídeos identificados com probabilidade de erro posterior (PEP) $\leq 0,01$, a identificação mínima de oito íons pertencentes às séries -b ou -y nos espectros de MS/MS e a ocorrência de pelo menos um peptídeo único. No caso específico de proteínas identificadas por somente um peptídeo, além dos critérios supracitados, foram aceitas somente as proteínas cujo peptídeo identificado apresentou a ocorrência ininterrupta mínima de pelo menos três íons da série -b ou -y.

3.10.4.4. ANÁLISE DOS DADOS GERADOS PELA COMBINAÇÃO DAS ESTRATÉGIAS SHOTGUN/BOTTOM-UP PROTEOMICS, GELC-MS/MS E MUDPIT OFF-LINE E "ON-LINE"

Os dados de identificação de proteínas, obtidos a partir das diferentes estratégias de análise dos proteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 foram combinados de forma a descrever o conjunto total de proteínas identificadas por amostra. Os conjuntos de proteínas identificadas, correspondendo aos proteomas de

folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 foram analisadas segundo as categorias Processo Biológico, Componente Celular e Função Molecular do *Gene Ontology* (Ashburner *et al.* 2000) e os resultados submetidos à análise de enriquecimento de termos e a representatividade de cada categoria foi calculada de acordo com as ferramentas contidas no software GeneMerge v.1.4 (Castillo-Davis e Hartl, 2003), considerando a taxa de falso positivo de 0,05%. Os conjuntos de proteínas identificadas foram classificados de acordo com as categorias presentes nos bancos de dados do *Superfamily* (Gough *et al.*, 2001).

3.10.5. ANÁLISE DOS FOSFOPROTEOMAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280 PELA METODOLOGIA IMAC DE ENRIQUECIMENTO DE FOSFOPEPTÍDEOS E ANÁLISE POR LC-MS/MS

Os peptídeos gerados pelo processo de digestão tripsínica em solução das amostras de proteínas extraídas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, coletadas de plantas de 12 meses, foram submetidas ao processo de enriquecimento de fosfopeptídeos segundo o método IMAC (Ficarro *et al.*, 2009). Para essa análise foram utilizadas duas réplicas biológicas de proteínas extraídas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280.

Após a digestão tripsínica de 200 µg de proteínas e posterior dessalinização das amostras utilizando cartuchos C-18 Sep-Pak light (Waters), estas foram submetidas ao processo de enriquecimento de fosfopeptídeos segundo o método conhecido como IMAC, baseado na afinidade dos peptídeos fosforilados pelo cátion metálico Fe³⁺ coordenados em resina de ácido nitrilotriacético (NTA), derivada de uma resina de Ni-

NTA superflow (QIAGEN), na qual o cátion metálico original (Ni^{2+}) foi substituído pelo Fe^{3+} a partir do protocolo modificado de Ficarro *et al.* (2009) descrito abaixo:

A. Preparo da resina de Fe-NTA: a resina de Ni-NTA (200 μL em tubo de 1,5 mL) foi lavada três vezes com 800 μL de água MilliQ. A cada passo de lavagem os tubos foram centrifugados a 6500 rpm por 3 minutos. Após essa etapa o níquel foi removido dos *beads* pela incubação com 100 μL de EDTA, pH 8,0, por 30 minutos, sob agitação e à temperatura ambiente. A resina foi lavada três vezes com 800 μL de água MilliQ seguida de incubação com solução de cloreto de ferro (III) 100 mM (Sigma-Aldrich) por 45 minutos à temperatura ambiente, sob agitação. Os *beads* de Fe-NTA foram então lavados mais três vezes com água MilliQ para remoção do excesso da solução de cloreto de ferro.

B. Enriquecimento de fosfopeptídeos: as amostras de peptídeos tripsínicos (volume de 25 a 50 μL) foram acidificadas com 25 μL de TFA 10% e adicionadas sobre a resina de Fe-NTA, à qual foram adicionados 200 μL de uma solução de ACN 80%/TFA 0,1%. As amostras foram submetidas à incubação à temperatura ambiente, por 1 hora sob agitação. Após essa etapa as amostras foram centrifugadas a 6500 rpm por 3 minutos e o sobrenadante foi removido. Os *beads* foram lavados duas vezes com 400 μL da solução de ACN 80%/TFA 0,1% para remoção dos peptídeos que não se ligaram à resina; duas vezes com 400 μL de uma solução de ACN 80%/ácido acético 0,1% para remoção do TFA e; mais duas vezes com 400 μL de solução de ácido acético 0,1% para remoção da acetonitrila.

C. Eluição dos fosfopeptídeos: Os fosfopeptídeos foram eluídos da resina com 80 μL de uma solução de fosfato de sódio 250 mM, pH 8,0, à temperatura ambiente, por 30

minutos sob agitação, e então carregados diretamente em uma pré-coluna confeccionada no próprio laboratório (C-18; Phenomenex; diâmetro dos *beads* 10 µm; 100 µm ID x 50 mm). Após o processo de carregamento, as pré-colunas foram lavadas com 50 µL de solução de ácido fórmico 0,1% e então acopladas ao sistema de nano HPLC Easy-nLCII (Thermo Scientific) para as análises de LC-MS/MS. Os peptídeos foram separados em um gradiente de 5 – 35% de B (AF 0,1% em ACN) durante 95 minutos sob fluxo de 200 nL/min. A fonte de ionização por *electrospray* foi ajustada para o modo de operação positivo, com a voltagem e temperatura ajustadas para 2,2 kV e 200°C, respectivamente. O intervalo de varredura de massas considerado para o *full scan* foi de 300 - 1650 m/z (resolução de 60.000 em 400 m/z), operando no modo DDA, onde os doze íons mais intensos por *scan* foram selecionados para o evento de fragmentação por CID com energia de colisão normalizada de 35% e janela de isolamento de 2 m/z para o íon precursor. O sinal mínimo requerido para disparar eventos de fragmentação de determinado íon foi ajustado para 5.000 cps, e o tempo de exclusão dinâmica utilizado foi 90 segundos.

3.10.5.1. ANÁLISE DOS DADOS GERADOS PELA ESTRATÉGIA BASEADA NO ENRIQUECIMENTO DE FOSFOPEPTÍDEOS PELA METODOLOGIA IMAC EM AMOSTRAS DE PEPTÍDEOS PROVENIENTES DE AMOSTRAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280

Os dados brutos gerados pelas análises de LC-MS/MS dos fosfopeptídeos foram submetidos à busca em banco de dados utilizando o software MaxQuant. Cada arquivo bruto foi tratado de forma independente para as buscas. A ferramenta “*match between runs*” foi utilizada para otimizar a identificação de fosfopeptídeos nas amostras.

Os parâmetros utilizados para a busca em banco de dados foram:

- Enzima: tripsina;
- Tolerância de duas clivagens perdidas pela enzima;
- Tolerância de erro de massa para o peptídeo precursor de 4,5 ppm;
- Tolerância de erro de massa para os fragmentos de MS/MS de 0,5 Da;
- Modificação fixa: carbamidometilação de resíduos de cisteínas;
- Modificações variáveis: fosforilação dos resíduos de serina, treonina ou tirosina, oxidação dos resíduos de metionina e a deamidação em resíduos de asparagina ou glutamina;
- Taxa de falso positivo em nível de peptídeos e proteínas: 1%.

A taxa de falso positivo (FDR) foi calculada a partir das estimativas realizadas a partir da construção de um banco de dados *decoy*, elaborado por meio da ordenação reversa das sequências traduzidas presentes no banco de dados original. Um banco de dados contendo 245 sequências traduzidas dos principais contaminantes encontrados nas análises por espectrometria de massas foi utilizado adicionalmente aos outros bancos de dados para identificação de potenciais contaminantes nas amostras.

Os resultados de identificação de fosfopeptídeos gerados pelo software MaxQuant foram analisados pelas ferramentas bioinformáticas contidas no software Perseus. As identificações obtidas a partir do banco de dados de contaminantes e do banco de dados *decoy* foram removidas. Como critério de identificação, foram considerados os fosfopeptídeos identificados em pelo menos uma réplica biológica com probabilidade mínima de localização de sítio de fosforilação igual a 75% (Olsen *et al.*, 2006).

3.10.6. ANÁLISE DO PROTEOMA DE FOLHAS +1 DE SP80-3280 COLETADAS AO LONGO DO DESENVOLVIMENTO DA PLANTA PELA UTILIZAÇÃO DA ESTRATÉGIA *SHOTGUN/BOTTOM-UP PROTEOMICS*

Os peptídeos provenientes do processo de digestão tripsínica em solução das proteínas extraídas de folhas +1 de SP80-3280 coletadas em diferentes estágios de desenvolvimento da planta foram analisados no espectrômetro LTQ Orbitrap Velos (Thermo Scientific) a partir da abordagem de *shotgun/bottom-up proteomics*. Cada réplica biológica foi submetida a duas análises independentes de LC-MS/MS, resultando em um número total de 18 análises. Alíquotas equivalentes a 5 µg de peptídeos foram automaticamente injetadas em uma pré-coluna C-18 Jupiter (Phenomenex; diâmetro dos *beads* 10 µm; 100 µm ID x 50 mm) por um sistema de nano-cromatografia líquida Easy-nLCII (Thermo Scientific). Os peptídeos foram submetidos à cromatografia em fase reversa utilizando uma coluna analítica C-18 (Dr. Maisch, Ammerbuch, Germany; diâmetro dos *beads* 3 µm; 100 µm x 150 mm) confeccionada no próprio laboratório. Os peptídeos foram separados por um gradiente linear de 7 – 35% de solvente B durante 95 minutos, sob fluxo de 200 nL/min. A fonte de ionização por *electrospray* foi ajustada para o modo de operação positivo, com a voltagem e temperatura ajustadas para 2,2 kV e 200°C, respectivamente. O intervalo de varredura de massas considerado para o *full scan* foi de 300 - 1650 m/z (resolução de 60.000 em 400 m/z), operando no modo DDA, onde os quinze íons mais intensos por *scan* foram selecionados para o evento de fragmentação por CID com energia de colisão normalizada de 35%, e janela de isolamento de 2 m/z para o íon precursor. O sinal mínimo requerido para disparar eventos de fragmentação de determinado íon foi ajustado para 5.000 cps, e o tempo de exclusão dinâmica utilizado foi de 90 segundos.

3.10.6.1. ANÁLISE DOS DADOS GERADOS PELA ESTRATÉGIA *SHOTGUN/BOTTOM-UP PROTEOMICS* NO ESTUDO DO PROTEOMA DE FOLHAS +1 DE SP80-3280 COLETADAS AO LONGO DO DESENVOLVIMENTO DA PLANTA

Os dados gerados pelas análises de LC-MS/MS dos peptídeos derivados das proteínas extraídas de folhas +1 de SP80-3280 em quatro estágios de desenvolvimento da planta (4, 8, 11 e 13 meses) foram submetidos aos mesmos parâmetros de busca e ferramentas bioinformáticas utilizadas para as análises dos dados referentes aos proteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de plantas de 12 meses. A análise quantitativa também foi realizada de acordo com os critérios descritos na seção 3.10.4.1.

3.10.7. ANÁLISE DO FOSFOPROTEOMA DE FOLHAS +1 DE SP80-3280 COLETADAS AO LONGO DO DESENVOLVIMENTO DA PLANTA A PARTIR DAS METODOLOGIAS IMAC E HAMMOC E ANÁLISE POR LC-MS/MS

Os peptídeos gerados pela digestão tripsínica em solução das amostras de proteínas extraídas de folhas +1 de SP80-3280 durante quatro estágios de desenvolvimento da planta (4, 8, 11 e 13 meses) foram submetidos à metodologia de enriquecimento de fosopeptídeos utilizando os métodos IMAC e HAMMOC (Nakagami, 2014). O método de enriquecimento HAMMOC foi executado em colaboração com o Departamento de Bioanálise Molecular e Celular, Escola Superior de Ciências Farmacêuticas, da Universidade de Kyoto, sob a supervisão do Prof. Dr. Yasushi Ishihama, no período compreendido entre 04/07/2016 e 15/07/2016.

3.10.7.1. MÉTODO IMAC E ANÁLISE DOS FOSFOPEPTÍDEOS POR LC-MS/MS

Após a digestão tripsínica de 200 µg de proteínas e posterior dessalinização das amostras utilizando cartuchos C-18 Sep-Pak light (Waters), estas foram submetidas ao processo de enriquecimento de fosfopeptídeos segundo o método IMAC, como descrito na seção 3.10.5. Para essa análise foram utilizadas três réplicas biológicas de cada amostra de proteínas extraídas de folhas +1 de SP80-3280 (4, 8, 11 e 13 meses).

3.10.7.2. MÉTODO HAMMOC COM MARCAÇÃO ISOBÁRICA DOS PEPTÍDEOS COM TMT (*TANDEM MASS TAG*) E ANÁLISE DOS FOSFOPEPTÍDEOS POR LC-MS/MS

As amostras de peptídeos (100 µg), provenientes do processo de digestão tripsínica em solução das amostras proteicas provenientes de folhas +1 de SP80-3280 coletadas durante quatro estágios de desenvolvimento da planta, foram submetidas ao processo de enriquecimento de fosfopeptídeos segundo o método HAMMOC seguido de marcação isobárica com o kit de TMT (Thermo Scientific). Para a confecção das colunas de dióxido de titânio (TiO₂) foram utilizadas membranas de C-8 (Empore, 3M) para montagem de *StageTip* em ponteiros Tip p-10, de forma análoga ao procedimento mostrado na Figura 9.

Figura 9. Passos demonstrando a confecção de uma *StageTip*. O produto final é mostrado em A, bem como uma ilustração da disposição dos *beads* na malha de Teflon. As *StageTips* podem ser confeccionadas em ponteiros de diferentes tamanhos (B), seguindo-se os passos descritos em C. Fonte: Rappsilber *et al.*, 2007.

As soluções A (ACN 5% em TFA 0,1%), B (ACN 80% em TFA 0,1%), C (ácido láctico 25%, dissolvido na solução B) e D (piperidina 0,5%) foram preparadas imediatamente antes do uso. Sobre a membrana de C-8 posicionada na extremidade do Tip foram adicionados 500 μg de micropartículas de TiO_2 (GL Sciences) suspensas em metanol e o aparato foi então submetido à centrifugação à temperatura ambiente por 5 minutos, a 500 x g. Após essa etapa foram adicionados 20 μL de solução B e o aparato foi centrifugado a 1500 x g por 5 minutos seguida de adição de 20 μL de solução C e centrifugação a 1500 x g por mais 5 minutos. Os peptídeos dessalinizados, previamente dissolvidos na solução C, foram adicionados sobre o aparato e este foi submetido à

centrifugação por 3 minutos a 1000 x g. As lavagens foram realizadas com a adição de 20 µL de solução C e 50 µL de solução B. Cada passo de lavagem foi seguido de centrifugação a 1500 x g por 5 minutos. Os fosfopeptídeos foram eluídos do aparato com 50 µL de solução D a 1000 x g por 8 minutos e coletados em um tubo contendo 10 µL de uma solução de ácido heptafluorobutírico (HFBA, Fluka) 10%.

Após o processo de enriquecimento da população de fosfopeptídeos, as amostras de peptídeos correspondentes às triplicatas biológicas das coletas 1 a 4 (4, 8, 11 e 13 meses) foram submetidas individualmente ao processo de derivatização utilizando o kit de marcação isobárica conhecido como *Tandem Mass Tags* (TMT; Thermo Scientific) para posterior combinação das amostras e análise por LC-MS/MS. As informações sobre os reagentes utilizados para a marcação dos peptídeos nas diferentes amostras são mostradas na Tabela 1.

Tabela 1. Massas dos reagentes de TMTsixplex utilizados para a derivatização dos fosfopeptídeos resultantes do processo de enriquecimento HAMMOC aplicado sobre a mistura de peptídeos tripsínicos presentes em proteínas de folhas +1 de SP80-3280, coletadas em quatro diferentes estágios de desenvolvimento da planta.

Amostra de peptídeo	Reagente	Massa da modificação (Da)	Massa monoisotópica do íon repórter (Da)
SP 4m RB1 SP 4m RB2 SP 4m RB3	TMT ⁶ -127 (127N)	229,162932	127,124761
SP 8m RB1 SP 8m RB2 SP 8m RB3	TMT ⁶ -128 (128C)	229,162932	128,134436
SP 11m RB1 SP 11m RB2 SP 11m RB3	TMT ⁶ -129 (129N)	229,162932	129,131471
SP 13m RB1 SP 13m RB2 SP 13m RB3	TMT ⁶ -130 (130C)	229,162932	130,141145

Para a derivatização dos peptídeos, inicialmente uma membrana de SDB-XC (*styrene-divinylbenzene*; Empore, 3M) foi posicionada em ponteira Tip p-10 de maneira análoga ao procedimento mostrado na Figura 9 e então 250 µg de esferas de InertSep RP C-18 (GL Sciences), suspensos em metanol, foram adicionados sobre membrana e o aparato foi submetido à centrifugação à temperatura ambiente por 4 minutos a 1500 x g. Após essa etapa foram adicionados 50 µL de uma solução de metanol 50% e o aparato foi centrifugado a 2500 x g por 3 minutos. A etapa de equilíbrio foi realizada com a adição de 20 µL de solução B e centrifugação a 1200 x g por 3 minutos, seguida de adição de 40 µL de solução A e centrifugação a 1200 x g por 3 minutos. Procedeu-se com o carregamento da amostra de peptídeos previamente eluídos com a solução D e o aparato foi centrifugado a 800 x g por 5 minutos, seguido de lavagens com 40 µL de solução A e 10 µL de solução E (ACN 2,5% em 50 mM de tampão fosfato KNa, pH 6,5). Em cada passo de lavagem o aparato foi submetido à centrifugação por 3 minutos a 1000 x g. A reação de marcação foi realizada adicionando-se 10 µL dos reagentes de TMT (dissolvidos em solução E na concentração de 1 µg/µL) sobre os aparatos, seguida de centrifugação a 500 x g por 3 minutos e incubação por 1 hora, à temperatura ambiente. Após o período de incubação, os aparatos foram lavados com 80 µL de solução A, centrifugados a 1000 x g por 3 minutos e os peptídeos foram eluídos com 50 µL de solução B, sob as mesmas condições de centrifugação. Os eluatos, correspondendo aos peptídeos das diferentes amostras marcados com os diferentes reagentes de TMT, foram combinados na proporção 1:1:1:1 da maneira mostrada na Figura 10.

Figura 10. Representação esquemática da marcação isobárica dos fosopeptídeos com o reagente TMT (Thermo Scientific). Após o processo de digestão tripsínica em solução, as amostras de peptídeos provenientes das proteínas extraídas de folhas +1 de SP80-3280 em diferentes tempos de desenvolvimento da planta [4 (col 1), 8 (col 2), 11 (col 3) e 13 (col 4) meses] foram submetidas ao processo de enriquecimento da população de fosopeptídeos pelo método HAMMOC. Os fosopeptídeos foram posteriormente submetidos à derivatização com diferentes *tags* do reagente TMT 6-plex (127N, 128C, 129N e 130C), combinados na proporção 1:1:1:1 e analisados por LC-MS/MS. Fonte: Adaptado de “User Guide: TMT Mass Tagging Kits and Reagents”, disponível no site <https://www.thermofisher.com/order/catalog/product/90061>.

Os peptídeos derivatizados e combinados de acordo com o esquema mostrado na Figura 10, após secagem em *speed-vac*, foram dissolvidos em 8 µL de ácido fórmico 0,1% e analisados por LC-MS/MS. As amostras foram injetadas automaticamente em uma coluna analítica C-18 (Dr. Maisch, Ammerbuch, Germany; diâmetro das esferas 3 µm; 100 µm x 150 mm) pelo sistema de nano-cromatografia Dionex Ultimate 3000 (Thermo Fisher Scientific) acoplado ao espectrômetro de massas Q-Exactive (Thermo Fisher Scientific). Os peptídeos foram fracionados utilizando-se um gradiente de 5 – 12,5% de B (ácido fórmico 0,1% em ACN) durante 2,5 minutos, 12,5 – 32,5% B em 60 minutos, 32,5 – 40% B em 2,5 minutos sob fluxo de 500 nL/min e os espectros de MS e

MS/MS foram adquiridos pelo analisador FTMS (*Fourier transform mass spectrometry*; *full scan range*: 300 – 1500 m/z), com resolução nominal de 70.000 e 35.000, respectivamente. A voltagem do *spray* foi ajustada para +2,4 kV e o instrumento foi operado no modo DDA em que os dez íons mais intensos por *scan* foram selecionados para o evento de fragmentação por HCD (*Higher energy Collision Dissociation*) utilizando energia de colisão normalizada de 32% e janela de isolamento do íon precursor de 1,4 m/z. O tempo de exclusão dinâmica foi ajustado para 24 segundos.

3.10.7.3. ANÁLISE DOS DADOS GERADOS PELAS ESTRATÉGIAS BASEADAS NO ENRIQUECIMENTO DE FOSFOPEPTÍDEOS PELAS METODOLOGIAS IMAC E HAMMOC EM AMOSTRAS DE FOLHAS +1 DE SP80-3280 COLETADAS AO LONGO DO DESENVOLVIMENTO DA PLANTA

Os dados gerados pelas análises de LC-MS/MS dos fosfopeptídeos provenientes do método de enriquecimento IMAC foram submetidos aos mesmos parâmetros de busca e análise bioinformática descritos na seção 3.10.5.1.

Os dados gerados pelas análises de LC-MS/MS dos fosfopeptídeos provenientes do método de enriquecimento HAMMOC foram submetidos à busca em banco de dados utilizando o software MaxQuant. A ferramenta “*match between runs*” foi utilizada para otimizar a identificação de fosfopeptídeos nas amostras. Os seguintes parâmetros de busca foram considerados:

- Enzima: tripsina;
- Tolerância de duas clivagens perdidas pela enzima;
- Tolerância de erro de massa para o peptídeo precursor de 4,5 ppm;

- Tolerância de erro de massa para os fragmentos de MS/MS de 0,02 Da;
- Modificação fixa: carbamidometilação de resíduos de cisteínas; TMT⁶-127 (127N), TMT⁶-128 (128C), TMT⁶-129 (129N) e TMT⁶-130 (130C);
- Modificações variáveis: fosforilação dos resíduos de serina, treonina ou tirosina, oxidação dos resíduos de metionina e a deamidação em resíduos de asparagina ou glutamina;
- Taxa de falso positivo em nível de peptídeos e proteínas: 1%.

Os resultados de identificação de fosfopeptídeos gerados pelo software MaxQuant foram analisados pelas ferramentas bioinformáticas contidas no software Perseus, onde as identificações obtidas a partir do banco de dados de contaminantes e do banco de dados *decoy* foram removidas. Como critério de identificação, foram considerados os fosfopeptídeos identificados em pelo menos uma réplica biológica, com probabilidade mínima de localização de sítio de fosforilação igual a 75% (Olsen *et al.*, 2006).

4. RESULTADOS E DISCUSSÃO

4.1. ANÁLISE DE PARÂMETROS FISIOLÓGICOS EM FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280 COLETADAS DE PLANTAS DE 12 MESES E EM FOLHAS +1 DE SP80-3280 COLETADAS EM DIFERENTES ESTÁGIOS DE DESENVOLVIMENTO DA PLANTA

As amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 coletadas de plantas de 12 meses, e de folhas +1 de SP80-3280 coletadas ao longo de diferentes estágios de desenvolvimento da planta foram submetidas às medidas de taxa fotossintética (*A*), de condutância estomática (*G_s*), da taxa de transpiração (*E*) e da eficiência no uso de água (*WUE*).

4.1.1. ANÁLISE DE PARÂMETROS FISIOLÓGICOS EM FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280 COLETADAS DE PLANTAS DE 12 MESES

As espécies ancestrais *S. officinarum* e *S. spontaneum* apresentam características fenotípicas muito distintas. *S. officinarum*, a espécie domesticada, apresenta maior capacidade de acumular sacarose em seus colmos e contém pouco teor de fibras, porém caracteriza-se pela maior suscetibilidade aos estresses bióticos e abióticos. *S. spontaneum*, a espécie selvagem, caracteriza-se por ser mais resistente, apresentar alto conteúdo de fibra e baixa capacidade de acumular sacarose (Sreenivasan *et al.*, 1987; Paterson *et al.*, 2013). Considerando essas diferenças fenotípicas entre *S. officinarum* e *S. spontaneum*, espécies utilizadas para a geração da

cultivar híbrida SP80-3280 a partir de cruzamentos inter-específicos, diferentes parâmetros fisiológicos foram medidos em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 coletadas de plantas de 12 meses, dentre eles, a taxa fotossintética ou assimilação de carbono (A), a condutância estomática (G_s), a taxa de transpiração (E) e eficiência no uso de água (WUE) (Tab. 2 e Fig. 11).

Tabela 2. Valores dos parâmetros fisiológicos medidos em folhas +1 (F +1) de *S. officinarum* (S. off), *S. spontaneum* (S. sp) e SP80-3280 (SP) sob condições de luz artificial.

Parâmetro	Amostra	Média	Erro padrão		Significância ANOVA	Unidade
Taxa Fotossintética (A)	F +1 S. off	17,37	0,70	a	9,49E-06	$\mu\text{mol CO}_2 \text{ m}^2 \text{ s}^{-1}$
	F +1 S. sp	11,67	1,04	b		
	F +1 SP	15,24	0,36	a		
Condutância Estomática (G_s)	F +1 S. off	0,42	0,04	a	0,295	$\text{mol CO}_2 \text{ m}^2 \text{ s}^{-1}$
	F +1 S. sp	0,48	0,06	a		
	F +1 SP	0,38	0,02	a		
Transpiração (E)	F +1 S. off	7,61	0,13	b	0,004	$\text{mol H}_2\text{O m}^2 \text{ s}^{-1}$
	F +1 S. sp	7,08	0,14	a		
	F +1 SP	6,92	0,17	a		
Eficiência no uso de H ₂ O (WUE)	F +1 S. off	2,28	0,07	a	3,30E-05	$\mu\text{mol CO}_2 / \text{mol H}_2\text{O}$
	F +1 S. sp	1,64	0,14	b		
	F +1 SP	2,22	0,07	a		

De acordo com o teste de Tukey, os valores de A , G_s , E e WUE que não apresentaram diferenças significativas entre si ($p\text{-value} > 0,05$) são representados por letras iguais, enquanto que valores que apresentaram diferenças significativas ($p\text{-value} < 0,05$) são representados por letras diferentes.

Figura 11. Parâmetros fisiológicos medidos em folhas + 1 (F +1) de *S. officinarum* (*S. off*), *S. spontaneum* (*S. sp*) e SP80-3280 (SP). As taxas de fotossíntese (A), de condutância estomática (B), de transpiração (C) e a eficiência de utilização da água (D) foram avaliadas estatisticamente de acordo com a análise de variância *one-way* ANOVA ($p\text{-value} < 0,05$) e posteriormente pelo teste Tukey ($p\text{-value} < 0,05$). Barras contendo letras iguais correspondem aos valores que não apresentaram diferenças significativas entre si ($p\text{-value} > 0,05$) e barras contendo letras diferentes correspondem aos valores que apresentaram diferenças significativas ($p\text{-value} < 0,05$).

A partir dos resultados da análise de variância *one-way* ANOVA mostrados na Tabela 2, verificou-se que existem diferenças significativas nas medidas da taxa de fotossíntese ($p = 9,49\text{E-}06$), taxa de transpiração ($p = 0,004$) e na eficiência no uso de água ($p = 3,30\text{E-}05$) em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280. Por outro lado, não foram observadas diferenças significativas na medida de condutância

estomática ($p = 0,295$). Constatou-se que folhas +1 de *S. officinarum* e SP80-3280 apresentam valores aproximadamente 49% e 30% maiores na taxa fotossintética quando comparados ao valor observado em folhas +1 de *S. spontaneum* ($11,67 \pm 1,04 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$), respectivamente (Tab. 2 e Fig. 11A). Estes resultados contrariam aqueles reportados por Irvine (1975), onde folhas de *S. spontaneum*, espécie que apresenta baixa capacidade de acúmulo de sacarose, apresentaram taxa fotossintética aproximadamente 30% maior em relação a espécies que apresentam alta capacidade de acúmulo de sacarose, demonstrando que a capacidade da planta em acumular sacarose não está necessariamente correlacionada com altas taxas fotossintéticas. Variações no potencial hídrico das folhas podem afetar a assimilação de carbono pela planta, de forma que a perda de água em níveis acima de sua capacidade de captação e transporte resulta no fechamento dos estômatos e diminuição da atividade fotossintética (Hsiao, 1973). Os valores da taxa de transpiração em folhas +1 de *S. spontaneum* ($7,08 \pm 0,14 \text{ mol H}_2\text{O m}^{-2} \text{ s}^{-1}$) e SP80-3280 ($6,92 \pm 0,17 \text{ mol H}_2\text{O m}^{-2} \text{ s}^{-1}$) foram 7% e 10% menores em relação ao valor observado em folhas +1 de *S. officinarum* ($7,61 \pm 0,13 \text{ mol H}_2\text{O m}^{-2} \text{ s}^{-1}$), enquanto que os valores de condutância estomática entre as espécies ancestrais e a cultivar híbrida não apresentaram diferenças significativas (Tab. 2 e Fig. 11B), indicando que plantas de *S. spontaneum* e SP80-3280 apresentam mecanismos de controle de perda de água por transpiração sem comprometimento das taxas de condutância estomática. Cowan e Troughton (1971) demonstraram que os mecanismos de regulação de abertura e fechamento dos estômatos apresentam pouca influência sobre o influxo de CO_2 e atuam principalmente no controle da perda de água e, conseqüentemente, na manutenção da eficiência no uso de água, dada pela relação

A / E. Folhas +1 de *S. spontaneum* e SP80-3280 perdem água por transpiração a uma taxa aproximadamente 9% menor em relação às folhas +1 de *S. officinarum* (Tab. 2 e Fig. 11C), o que não representou, no caso de *S. spontaneum*, uma melhor relação da eficiência no uso de água (Tab. 2 e Fig. 11D). *S. spontaneum* apresentou o menor valor de eficiência no uso de água ($1,64 \pm 0,14 \mu\text{mol CO}_2 / \text{mol H}_2\text{O}$), devido à baixa taxa de assimilação de CO_2 em relação aos valores observados em folhas +1 de *S. officinarum* ($2,28 \pm 0,07 \mu\text{mol CO}_2 / \text{mol H}_2\text{O}$) e SP80-3280 ($2,22 \pm 0,07 \mu\text{mol CO}_2 / \text{mol H}_2\text{O}$), que por sua vez não apresentaram diferenças significativas nos valores de *WUE*. Esses dados, em conjunto, revelam que *S. officinarum* e SP80-3280 apresentam uma melhor relação entre o número de moléculas de água perdidas por transpiração e o número de moléculas de CO_2 assimiladas pelo processo de fotossíntese, principalmente devido às maiores taxas de assimilação de carbono em folhas +1 de *S. officinarum* e SP80-3280 quando comparadas à taxa de assimilação em folhas +1 de *S. spontaneum*.

4.1.2. ANÁLISE DE PARÂMETROS FISIOLÓGICOS EM FOLHAS +1 DE SP80-3280 COLETADAS EM DIFERENTES ESTÁGIOS DE DESENVOLVIMENTO DA PLANTA

As folhas +1 de SP80-3280 coletadas ao longo do desenvolvimento da planta foram submetidas às mesmas medidas de parâmetros fisiológicos que as espécies ancestrais, antes da realização das coletas. Considerando o experimento de longo prazo, nas quais as condições meteorológicas podem variar bastante entre os diferentes tempos de coleta, optou-se pela utilização do instrumento de medida de fotossíntese portátil LCI com fonte de iluminação artificial. O plantio foi realizado em abril de 2012, em um período de pluviosidade média e temperatura amena. A primeira coleta foi realizada

quando as plantas apresentavam 4 meses, em condições de baixa temperatura e precipitação. Os meses subsequentes se caracterizaram pelo gradual aumento da temperatura e precipitação até o momento da segunda coleta (dezembro de 2012), quando as plantas apresentavam 8 meses de idade, seguido pela progressiva diminuição da temperatura e condições pluviométricas semelhantes no momento em que as plantas apresentavam 11 e 13 meses (Fig. 12).

Figura 12. Dados meteorológicos mensais referentes aos períodos de plantio e coletas (Col 1 a 4) de folhas +1 de SP80-3280 ao longo do desenvolvimento da planta. As medidas dos parâmetros fisiológicos foram realizadas em folhas +1 quando as plantas apresentavam 4, 8, 11 e 13 meses de idade. Fonte: <http://www.cca.ufscar.br/servicos/dados-climatologicos/>

Na ocasião da primeira coleta, o instrumento de medida de fotossíntese portátil LCI apresentou problemas na fonte de iluminação artificial e, portanto, as medidas foram realizadas sob regime de iluminação natural. Com relação aos outros tempos de coleta

considerados (8, 11 e 13 meses), as medidas foram executadas utilizando a fonte artificial de iluminação. Dessa forma, diante das variações introduzidas nas medidas realizadas sem a utilização da fonte de iluminação artificial, os valores de *A*, *G_s*, *E* e *WUE* referentes às medidas realizadas em folhas +1 de SP80-3280 de plantas de 4 meses não foram considerados no conjunto de dados mostrados na Tabela 3 e Figura 13.

Tabela 3. Valores dos parâmetros fisiológicos medidos em folhas +1 (F +1) de SP80-3280 (SP) coletadas em quatro estágios de desenvolvimento da planta (4, 8, 11 e 13 meses após o plantio) sob condições de luz artificial.

Parâmetro	Amostra	Média	Erro padrão		Significância ANOVA	Unidade
Taxa Fotossintética (<i>A</i>)	F +1 SP 4m	-	-	-	0,747	$\mu\text{mol CO}_2 \text{ m}^2 \text{ s}^{-1}$
	F +1 SP 8m	10,39	1,64	a		
	F +1 SP 11m	9,46	0,63	a		
	F +1 SP 13m	9,21	0,93	a		
Condutância Estomática (<i>G_s</i>)	F +1 SP 4m	-	-	-	0,031	$\text{mol CO}_2 \text{ m}^2 \text{ s}^{-1}$
	F +1 SP 8m	0,067	0,011	a		
	F +1 SP 11m	0,043	0,005	ab		
	F +1 SP 13m	0,040	0,006	b		
Transpiração (<i>E</i>)	F +1 SP 4m	-	-	-	2,26E-04	$\text{mol H}_2\text{O m}^2 \text{ s}^{-1}$
	F +1 SP 8m	1,80	0,25	a		
	F +1 SP 11m	0,87	0,11	b		
	F +1 SP 13m	0,86	0,10	b		
Eficiência no uso de H₂O (<i>WUE</i>)	F +1 SP 4m	-	-	-	0,01	$\mu\text{mol CO}_2 / \text{mol H}_2\text{O}$
	F +1 SP 8m	6,63	0,84	a		
	F +1 SP 11m	15,31	3,19	b		
	F +1 SP 13m	11,59	0,76	ab		

De acordo com o teste de Tukey, os valores de *A*, *G_s*, *E* e *WUE* que não apresentaram diferenças significativas entre si (*p-value* > 0,05) são representados por letras iguais, enquanto que valores que apresentaram diferenças significativas (*p-value* < 0,05) são representados por letras diferentes. Valores representados por "ab" correspondem aos valores intermediários e que não são estatisticamente diferentes em relação aos valores representados por "a" e "b"; os dados de *A*, *G_s*, *E* e *WUE* referentes às medidas realizadas em folhas de 4 meses foram omitidos devido à não utilização da fonte de iluminação artificial, excepcionalmente na ocasião da primeira coleta.

Figura 13. Parâmetros fisiológicos medidos em folhas + 1 (F +1) de SP80-3280 (SP) coletadas em quatro estágios de desenvolvimento da planta (4, 8, 11 e 13 meses após o plantio). As taxas de fotossíntese (A), de condutância estomática (B), de transpiração (C) e a eficiência de utilização da água (D) foram avaliadas estatisticamente de acordo com a análise de variância *one-way* ANOVA ($p\text{-value} < 0,05$) e posteriormente pelo teste Tukey ($p\text{-value} < 0,05$). Barras contendo letras iguais correspondem aos valores que não apresentaram diferenças significativas entre si ($p\text{-value} > 0,05$) e barras contendo letras diferentes correspondem aos valores que apresentaram diferenças significativas ($p\text{-value} < 0,05$). Barras representadas por “ab” correspondem aos valores intermediários e que não são estatisticamente diferentes em relação aos valores representados por “a” e “b”. As medidas realizadas em folhas de 4 meses não foram consideradas devido à não utilização da fonte de iluminação artificial, excepcionalmente nessa ocasião.

Os resultados da análise de variância *one-way* ANOVA mostraram que os valores de condutância estomática ($p = 0,031$) (Tab. 3 e Fig. 13B), taxa de transpiração ($p = 2,26\text{E-}04$) (Tab. 3 e Fig. 13C) e de eficiência no uso de água ($p = 0,01$) (Tab. 3 e Fig. 13D) apresentaram diferenças significativas em folhas +1 de SP80-3280 de plantas de

8, 11 e 13 meses. Por outro lado, os valores de taxa fotossintética não apresentaram diferenças significativas ($p = 0,747$) (Tab. 3 e Fig. 13A).

Em folhas +1 de plantas de 8 meses, o valor de condutância estomática ($0,067 \pm 0,011 \text{ mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$) foi aproximadamente 68% maior em relação ao valor observado em folhas +1 de plantas de 13 meses ($0,040 \pm 0,006 \text{ mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$). Apesar do teste estatístico não apontar diferenças entre os valores observados em folhas +1 de plantas de 8 e 11 meses no nível de significância determinado, a taxa de condutância estomática foi aproximadamente 58% maior em folhas de plantas de 8 meses. A maior taxa de influxo de CO_2 em folhas de plantas de 8 meses resultou em uma taxa de transpiração ($1,80 \pm 0,25 \text{ mol H}_2\text{O m}^{-2} \text{ s}^{-1}$) aproximadamente 2 vezes maior em relação ao valor médio observado em folhas de plantas de 11 e 13 meses, corroborando com os relatos de Cowan e Troughton (1971) sobre a maior influência do controle da abertura dos estômatos nos processos de perda de água por transpiração, e em menor grau nos processos de influxo de CO_2 . Considerando as condições de elevada precipitação e umidade em períodos próximos à coleta realizada no mês de dezembro de 2012, quando as plantas apresentavam 8 meses (Fig. 12), pode-se atribuir o alto valor relativo da taxa de transpiração e de condutância estomática à condição de alto potencial hídrico das folhas, de forma que a alta disponibilidade de água permitiu que sua captação pela planta fosse alta o suficiente para permitir uma maior abertura estomática, sem representar um prejuízo hídrico. Os baixos valores da taxa de transpiração observados em folhas +1 de plantas de 11 e 13 meses podem estar associados à diminuição da abertura dos estômatos em resposta à deficiência hídrica da planta, uma vez que nesses dois períodos considerados, a precipitação moderada (Fig. 12) poderia ativar mecanismos de

controle de abertura e fechamento dos estômatos para evitar a perda de água acima dos limites de captação pela planta. E, de maneira análoga aos efeitos resultantes da abertura dos estômatos nos valores das taxas de transpiração e de condutância estomática, o seu fechamento apresenta maior efeito sobre a inibição da transpiração do que sobre a diminuição do influxo de CO₂ (Taiz e Zeiger, 2002). Considerando que os valores de taxa fotossintética foram similares entre as folhas +1 coletadas de plantas de 8, 11 e 13 meses, os valores de eficiência no uso de água são influenciados principalmente pelas diferenças entre as taxas de transpiração observadas. Os baixos valores observados na taxa de transpiração em folhas +1 de plantas de 11 e 13 meses resultaram em valores maiores de eficiência no uso de água, enquanto que a alta taxa de transpiração em folhas +1 de plantas de 8 meses resultou em um menor valor de eficiência no uso de água ($6,63 \pm 0,84 \mu\text{mol CO}_2 / \text{mol H}_2\text{O}$).

Esses resultados evidenciam o importante papel dos estômatos na manutenção do potencial hídrico foliar, por meio de ajustes que levam ao controle das taxas de transpiração em resposta às variações do ambiente, sem contudo induzir alterações drásticas na taxa de fotossíntese, ao longo do período avaliado durante este estudo.

4.2. AVALIAÇÃO DOS MÉTODOS DE EXTRAÇÃO DE PROTEÍNAS

Considerando a fundamental importância que os processos de extração de proteínas e preparação de amostras representam para a boa condução das análises proteômicas, a primeira etapa do processo de caracterização dos proteomas de folhas

+1 de *S. officinarum*, *S. spontaneum* e SP80-3280 consistiu na avaliação comparativa de diferentes métodos de extração de proteínas aplicados às amostras. É conhecido que tecidos vegetais são particularmente desafiadores, devido ao relativamente baixo conteúdo proteico, à presença de proteases e à elevada quantidade de compostos interferentes presentes nas células vegetais, como pigmentos, ácidos nucleicos, polissacarídeos e lipídeos (Westermeyer e Naven, 2002; Amalraj *et al.*, 2010). Nessa primeira etapa foram realizados experimentos pilotos somente com algumas amostras de folhas +1 das espécies ancestrais *S. officinarum* e *S. spontaneum*.

As extrações de proteínas presentes em folhas +1 das espécies ancestrais *S. officinarum* e *S. spontaneum* foram então conduzidas segundo três métodos principais, que foram avaliados de acordo com os rendimentos de extração de proteínas e de acordo com os perfis eletroforéticos. O método TCA/acetona, descrito por Wu e Wang (1984), baseia-se nas propriedades de desnaturação de proteínas pelo agente 2-mercaptoetanol combinado com o processo de precipitação de proteínas com acetona gelada. O método Fenol, derivado do método descrito por Hurkman e Tanaka (1986), é baseado na captura das proteínas pelo fenol adicionado sobre a solução de extração, combinada com sua precipitação com acetato de amônio em metanol. E por fim, o método Fenol/SDS, descrito por Wang *et al.* (2003), integra os métodos de extração por TCA/acetona e Fenol e, ainda, se beneficia das propriedades do SDS, que contribuem para a solubilização das proteínas.

Nesta avaliação, folhas +1 foram submetidas aos diferentes processos de extração, iniciando pela pulverização e pesagem, e continuando com cada processo de extração de proteínas (TCA/acetona, Fenol e Fenol/SDS) em separado. As proteínas

extraídas de acordo com o método TCA/acetona foram solubilizadas em tampão Tris-HCl, enquanto que as proteínas derivadas dos processos de extração Fenol e Fenol/SDS foram dissolvidas em tampão LBT e as respectivas concentrações proteicas foram estimadas pelo método de Bradford. Para a determinação dos rendimentos de extração de proteínas foi considerada a massa total de proteínas extraídas em relação à massa total de folhas frescas submetidas aos processos de extração (Tab. 4).

Tabela 4. Rendimento de extração de proteínas presentes nas folhas +1 (F+1) de cana-de-açúcar de acordo com os diferentes métodos de extração utilizados.

Amostra	Rendimento de extração de proteínas (mg/g de folhas frescas)		
	TCA/ACETONA	FENOL	FENOL/SDS
F+1 <i>S. officinarum</i>	0,8	1,0	3,3
F+1 <i>S. spontaneum</i>	1,2	1,5	3,9
Média	1,0	1,2	3,6

Os valores de rendimento de extração de proteínas apresentados na Tabela 4 revelam que o método de extração Fenol/SDS foi aquele que apresentou o maior rendimento entre os métodos avaliados. Este se mostrou superior aos métodos TCA/acetona e Fenol tanto em folhas +1 de *S. officinarum* como em folhas +1 de *S. spontaneum*, apresentando um rendimento médio de extração aproximadamente 4 vezes maior em relação ao método TCA/acetona e 3 vezes maior em relação ao método Fenol. Os métodos Fenol e TCA/acetona, por outro lado, apresentaram rendimentos de extração semelhantes, com uma discreta superioridade observada no rendimento de proteínas extraídas pelo método Fenol em relação ao método TCA/acetona,

apresentando um rendimento médio de extração 20% maior em relação ao método TCA/acetona.

Além da avaliação dos rendimentos dos processos de extração de proteínas, foram também avaliados os perfis eletroforéticos das proteínas extraídas segundo os três métodos. A presença de interferentes em amostras proteicas extraídas de tecidos vegetais normalmente pode ser diagnosticada pela coloração difusa ao longo do gel de SDS-poliacrilamida e distorções das bandas de proteínas. Além disso, a remoção ineficiente desses componentes não proteicos pode ser evidenciada pela presença de bandas observadas na região superior dos géis que não exibem capacidade de migração (Saravanan e Rose, 2004). Dessa forma, após as extrações e quantificação das proteínas, as amostras proteicas foram submetidas à eletroforese em gel de SDS-poliacrilamida 12% e, após coloração dos géis, os perfis eletroforéticos das proteínas extraídas de folhas +1 de *S. officinarum* e *S. spontaneum* (Fig. 14) foram avaliados em relação à população de bandas e ao perfil de migração das proteínas.

Figura 14. Eletroforese em gel de SDS-poliacrilamida 12% das proteínas extraídas de folhas +1 de *S. officinarum* (**S. off**) e *S. spontaneum* (**S. sp**) de acordo com os métodos de extração de proteínas TCA/acetona (A), Fenol (B) e Fenol/SDS (C). As amostras de proteínas (5 µg) foram analisadas em duas réplicas técnicas (R1 e R2). Em verde destaca-se a região dos géis que contém componentes que não apresentam capacidade de migração e em vermelho regiões cuja complexidade da população de proteínas extraídas pelo método TCA/acetona é relativamente menor em relação aos métodos Fenol e Fenol/SDS. Os géis foram submetidos à coloração por nitrato de prata. Os valores à esquerda indicam a mobilidade de marcadores de massa molecular.

Ao analisar os perfis eletroforéticos apresentados na Figura 14 verificou-se que os três métodos foram capazes de extrair, em menor ou maior grau, proteínas presentes nas folhas de *S. officinarum* e *S. spontaneum*, sendo possível observar bandas de proteínas distribuídas em uma ampla faixa de massa molecular. Ainda, em todos os perfis observa-se a ocorrência de componentes que não apresentam capacidade de migração (regiões destacadas em verde). Novamente, o método de extração Fenol/SDS

(Fig. 14C) se sobressaiu em relação ao método Fenol (Fig. 14B) e TCA/acetona (Fig. 14A) no que diz respeito à complexidade da população de proteínas extraídas, sendo possível observar um número médio de 21 bandas de proteínas no perfil eletroforético das proteínas extraídas pelo método Fenol/SDS, enquanto que nos perfis das proteínas extraídas pelo método TCA/acetona e Fenol, o número de bandas observadas apresentou um valor médio de 13 e 14, respectivamente (Fig. 15), correspondendo a um aumento médio de aproximadamente 62% em relação ao método TCA/acetona e de aproximadamente 50% em relação ao método Fenol.

Figura 15. Número de bandas de proteínas observadas nos perfis eletroforéticos das amostras extraídas de folhas +1 de *S. officinarum* e *S. spontaneum* apresentados na Figura 14 de acordo com os métodos de extração TCA/acetona, Fenol e Fenol/SDS. As barras representam o desvio padrão de 4 contagens. O número de bandas foi determinado a partir da utilização do software ImageJ (Abramoff *et al.*, 2004).

É importante destacar que a elevada e difusa coloração de fundo, observada nos géis apresentados nas Figuras 14A e 14B, decorre da ineficiência dos métodos TCA/acetona e Fenol em remover os compostos interferentes, tornando particularmente difícil a visualização e distinção das bandas de proteínas.

No perfil eletroforético das proteínas extraídas pelo método TCA/acetona verificou-se a ocorrência de regiões de média e alta massa molecular (regiões destacadas em vermelho), cuja população de bandas de proteínas se apresentou relativamente menos complexa quando comparadas às regiões correspondentes nos perfis eletroforéticos das proteínas extraídas pelos métodos Fenol e Fenol/SDS. De maneira geral, os métodos de extração de proteínas com a utilização de fenol resultam em uma maior eliminação de compostos interferentes com consequente aumento na qualidade dos géis (Saravanan e Rose, 2004; Carpentier *et al.*, 2005). Nossos resultados mostraram que somente no caso do método Fenol/SDS isso se mostrou verdadeiro, uma vez que a coloração inespecífica de fundo, que foi observada de forma evidente nos perfis eletroforéticos das proteínas extraídas pelo método Fenol e TCA/acetona, contrasta com o perfil eletroforético das proteínas extraídas pelo método Fenol/SDS.

Em conjunto, os resultados dos rendimentos dos processos de extração e a avaliação dos perfis eletroforéticos das proteínas extraídas segundo os três métodos utilizados suportam o fato de que o método de extração Fenol/SDS é o mais adequado para ser aplicado às amostras de folhas +1 de cana-de-açúcar, uma vez que este foi capaz de conciliar alto rendimento de extração e elevado poder de remoção de compostos interferentes quando comparado aos métodos TCA/acetona e Fenol. Dessa forma, foi o método de extração de proteínas escolhido para este estudo.

4.2. EXTRAÇÃO DE PROTEÍNAS DAS AMOSTRAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280 SEGUNDO O MÉTODO FENOL/SDS

Uma vez constatadas as vantagens que o método Fenol/SDS apresenta na extração de proteínas presentes em folhas +1 de *S. officinarum* e *S. spontaneum*, este foi aplicado às amostras alvo deste estudo, que compreendiam três réplicas biológicas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280. Após a extração e solubilização das proteínas, as amostras foram submetidas à quantificação de proteínas utilizando o método de Bradford e os rendimentos do processo de extração são mostrados na Tabela 5.

Tabela 5. Rendimento de extração de proteínas presentes nas folhas +1 (F+1) de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com o método de extração Fenol/SDS.

Amostra	Rendimento de extração de proteínas (mg/g de folhas frescas)
F+1 <i>S. officinarum</i> RB1	2,8
F+1 <i>S. officinarum</i> RB2	2,4
F+1 <i>S. officinarum</i> RB3	2,5
F+1 <i>S. spontaneum</i> RB1	3,0
F+1 <i>S. spontaneum</i> RB2	3,0
F+1 <i>S. spontaneum</i> RB3	2,9
F+1 SP80-3280 RB1	3,1
F+1 SP80-3280 RB2	3,6
F+1 SP80-3280 RB3	3,4
Rendimento médio	3,0

RB: Réplica Biológica

Observou-se um rendimento médio de extração de 3 mg/g, que corresponde a um rendimento ligeiramente abaixo ao valor de rendimento médio de extração observado na avaliação do método Fenol/SDS em comparação com os métodos TCA/acetona e Fenol (Tab. 4). As proteínas extraídas de acordo com o método Fenol/SDS foram avaliadas por eletroforese em gel de SDS-poliacrilamida 12% (Fig. 16).

Figura 16. Perfil eletroforético das proteínas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, extraídas de acordo com o método Fenol/SDS. As amostras de proteínas (25 µg) foram analisadas por eletroforese em gel de SDS-poliacrilamida 12% segundo as réplicas biológicas (RB). O gel foi submetido à coloração por Coomassie G-250. Os valores à esquerda indicam a mobilidades dos marcadores de massa molecular.

A análise dos perfis eletroforéticos mostrou de forma sistemática que o processo de extração de proteínas Fenol/SDS foi eficiente na remoção de compostos interferentes e na capacidade de extração de proteínas, uma vez que é possível observar nitidamente diversas bandas de proteínas bem distintas em cada amostra, sem, contudo, observar a coloração difusa de fundo, características que denotam a co-extração de compostos interferentes. Observou-se que as proteínas extraídas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 estão distribuídas em uma ampla faixa de massa molecular e nesta análise não foram observadas alterações significativas entre os perfis eletroforéticos correspondentes às amostras de proteínas extraídas.

4.3. ANÁLISE QUALITATIVA DOS PROTEOMAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280

Os proteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 coletadas de plantas de 12 meses foram avaliados do ponto de vista qualitativo de acordo com as seguintes metodologias de análise proteômica:

- Digestão tripsínica em solução das proteínas extraídas de folhas +1 e análise direta dos peptídeos por LC-MS/MS, técnica conhecida como *Shotgun/bottom-up proteomics*;

- Eletroforese em gel de SDS-poliacrilamida das proteínas extraídas de folhas +1 seguida de digestão tripsínica *in gel* de segmentos do gel, e análise dos peptídeos por LC-MS/MS, técnica conhecida como GeLC-MS/MS;
- Digestão tripsínica em solução das proteínas extraídas de folhas +1 e análise dos peptídeos pelas técnicas de MudPIT *off-line* e “*on-line*”.

4.3.1. ANÁLISE DOS PROTEOMAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280 PELA ABORDAGEM *SHOTGUN/BOTTOM-UP PROTEOMICS*

Os proteomas de folhas +1 das espécies ancestrais *S. officinarum* e *S. spontaneum* e da cultivar híbrida SP80-3280 foram inicialmente avaliados de acordo com a técnica *Shotgun/bottom-up proteomics* (Yates, 1998; Zhang *et al.*, 2013), abordagem em que a população total de proteínas de determinado proteoma é submetida ao processo de proteólise em solução com enzimas de especificidade conhecida (normalmente tripsina) e os peptídeos resultantes são analisados por LC-MS/MS.

Nesta análise, as amostras correspondentes às três réplicas biológicas de folhas +1 foram submetidas à digestão tripsínica em solução e os peptídeos resultantes de cada amostra foram analisados em duplicata por LC-MS/MS em um sistema nano-HPLC EASY nLCII acoplado ao espectrômetro LTQ-Orbitrap Velos. O processamento dos arquivos de extensão .raw gerados pelas 18 análises de LC-MS/MS foi realizado pelo algoritmo de processamento e busca em banco de dados contido no software MaxQuant. Como critério de identificação de proteínas, foram consideradas a ocorrência de pelo menos um peptídeo único com $PEP \leq 0,01$ observado em pelo menos duas réplicas

biológicas e identificação mínima de oito íons pertencentes às séries -b ou -y nos espectros de MS/MS. No caso específico de proteínas identificadas somente por um peptídeo, sendo esse necessariamente único, além dos critérios de identificação já mencionados, foram aceitas somente as proteínas cujo peptídeo identificado apresentou a ocorrência ininterrupta mínima de pelo menos três íons da série -b ou -y. A Figura 17 ilustra as possíveis situações para a aceitação ou a rejeição de identificação de determinada proteína, no caso da detecção de somente um peptídeo.

A qualidade dos espectros de MS/MS é determinante para a correta associação dos picos de fragmentação com as massas dos resíduos de aminoácidos contidos em determinada sequência de proteína presente no banco de dados. Diante disso, fatores como abundância, estabilidade do peptídeo precursor ou dos fragmentos, ou mesmo os parâmetros de aquisição de dados podem influenciar a qualidade dos espectros de fragmentação e, conseqüentemente na confiabilidade de identificação de determinado peptídeo. Dessa forma, para minimizar a ocorrência de identificações espúrias daquelas proteínas dependentes de uma única evidência (condicionadas à identificação de somente um peptídeo) foi utilizado um critério adicional exemplificado na Figura 17. Nesses casos, os peptídeos identificados neste estudo, além de obedecer aos critérios já mencionados ($PEP \leq 0,01$ observado em pelo menos duas réplicas biológicas e identificação mínima de oito íons pertencentes às séries -b ou -y), deveriam obedecer aos critérios mínimos de três íons consecutivos pertencentes à série de fragmentação -b ou -y. Esse requisito é atendido idealmente no espectro de fragmentação apresentado na Figura 17A, onde é possível observar a anotação de 10 íons consecutivos na série -b e de 12 íons consecutivos na série -y, enquanto o espectro de fragmentação apresentado na Figura 17B atende minimamente este critério, com a anotação de três íons consecutivos (duas ocorrências) pertencentes à série -b. Por outro lado, na Figura 17C, apesar de o peptídeo identificado atender aos critérios de valor de PEP e de ocorrência mínima de oito íons pertencentes às séries -b ou -y, este não atende ao critério de ocorrência mínima de três íons consecutivos pertencentes às séries de fragmentação -b ou -y. Assim sendo, proteínas que seriam eventualmente identificadas por um peptídeo que não se ajusta ao critério de ocorrência mínima de três íons

consecutivos pertencentes à série de fragmentação -b ou -y foram removidas da lista final de proteínas identificadas.

A análise dos dados de aquisição pelo software MaxQuant resultou, de acordo com os critérios de identificação de peptídeos adotados neste estudo, na identificação de 4.922 peptídeos na amostra de *S. officinarum*, 4.876 em *S. spontaneum* e 4.901 peptídeos em SP80-3280, resultando na identificação de 1.030 grupos de proteínas em folhas +1 de *S. officinarum*, 1.010 em folhas +1 de *S. spontaneum* e 1.018 em folhas +1 de SP80-3280 (Tabela 6; Tabela Suplementar 1).

Tabela 6. Número de peptídeos e grupos únicos de proteínas identificados em amostras de folhas +1 (F+1) de *S. officinarum* (**S.off**), *S. spontaneum* (**S.sp**) e SP8032-80 (**SP**) a partir das análises de LC-MS/MS dos peptídeos gerados pela digestão tripsínica em solução.

Amostra	n. peptídeos	n. grupos únicos de proteínas*
F+1 S.off	4.922	1.030
F+1 S.sp	4.876	1.010
F+1 SP	4.901	1.018

* A presença de uma proteína em determinada amostra foi considerada de acordo com a identificação mínima de 1 peptídeo único com valor de PEP $\leq 0,01$ em pelo menos duas réplicas biológicas e ocorrência ininterrupta de pelo menos 3 íons das séries -y ou -b no caso da proteína ser identificada por somente 1 peptídeo.

A representação das identificações em grupos de proteínas tem por objetivo minimizar a redundância de identificação de proteínas. Este método de inferência, contido no algoritmo de busca do software MaxQuant, se baseia no agrupamento das proteínas cujas sequências de peptídeos identificados não permitem a distinção entre as proteínas contidas em um determinado grupo. Ou seja, na situação hipotética em que o conjunto de peptídeos identificados na proteína B está completamente contido no

conjunto de peptídeos identificados na proteína A, as proteínas A e B são agrupadas, constituindo assim um grupo de proteínas (Nesvizhskii *et al.*, 2003; Cox e Mann, 2008). A situação descrita é comumente observada na identificação de proteínas distintas que apresentam elevado grau de homologia de sequência, como por exemplo nos casos de famílias de proteínas e variantes de *splicing*, de forma que a distinção entre elas só é possível diante da identificação dos chamados peptídeos proteotípicos. Com o intuito de simplificar a discussão, os termos “grupos de proteínas” e “proteínas” serão considerados intercambiáveis no decorrer da descrição dos resultados deste estudo.

A Figura 18 mostra as distribuições das proteínas identificadas nas amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280. Verificou-se uma alta correlação entre as identificações de proteínas nas amostras de folhas +1 das espécies ancestrais e da cultivar híbrida. Das 1.060 proteínas únicas totais identificadas pela abordagem *Shotgun/bottom-up*, 972 (91,7%) foram identificadas concomitantemente nas três amostras, enquanto que 88 (8,3%) foram identificadas ou em duas das amostras, ou apenas em uma delas de forma exclusiva.

Figura 18. Diagrama de Venn demonstrando a distribuição do número total de proteínas identificadas (n= 1.060) nas amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 a partir das análises de LC-MS/MS dos peptídeos resultantes do processo de digestão tripsínica em solução, e busca em banco de dados. O diagrama de Venn foi construído utilizando o software eulerAPE (Micallef e Rodgers, 2014).

Diante da elevada sobreposição entre os proteomas, os conjuntos de proteínas identificadas de forma exclusiva ou compartilhada entre duas ou três amostras foram avaliadas segundo o número médio de peptídeos que compõe cada conjunto. A partir dessa análise, observou-se um valor médio de 5 peptídeos por proteína no conjunto de proteínas comuns a *S. officinarum*, *S. spontaneum* e SP80-3280 (972), 1,6 peptídeos por proteína no conjunto de proteínas identificadas concomitantemente em duas amostras (54) e 1,1 peptídeos por proteína no conjunto de proteínas identificadas de forma exclusiva (34). Essa tendência observada de diminuição do número médio de peptídeos

identificados à medida que as identificações de proteínas vão se tornando mais restritas evidencia que a ocorrência de proteínas exclusivas ou compartilhadas entre duas amostras pode ser resultado de flutuações intrínsecas na capacidade de detecção e identificação de peptídeos pouco abundantes nas amostras, de forma que estes podem ser eventualmente negligenciados pelas análises de LC-MS/MS. Para investigar essa suposta causa para as flutuações de identificação de peptídeos, foi realizada uma avaliação mais detalhada dos dados de identificação considerando as intensidades dos peptídeos precursores e a contagem de espectros de MS/MS. Dessa forma, foram calculados os valores médios de intensidade e de contagem de espectros referentes aos peptídeos associados às identificações exclusivas de proteínas (presentes somente em folhas +1 de *S. officinarum*, *S. spontaneum* ou SP80-3280), compartilhadas entre duas (*S. officinarum* e *S. spontaneum*; *S. officinarum* e SP80-3280; *S. spontaneum* e SP80-3280) ou três amostras (*S. officinarum*, *S. spontaneum* e SP80-3280). Os resultados dessa análise foram então avaliados segundo sua distribuição e dispersão (Fig. 19). Os valores referentes à contagem de espectros e de intensidades foram submetidos a uma etapa de normalização para permitir uma melhor visualização dos diagramas apresentados.

Figura 19. *Boxplots* mostrando as distribuições dos valores médios de intensidade (A) e de contagem de espectros de MS/MS (B) dos peptídeos associados às proteínas identificadas de forma exclusiva (n= 34), compartilhadas entre duas (n= 54) ou entre as três amostras (n=972) de folhas +1 de *S. officinarum* (**S.off**), *S. spontaneum* (**S.sp**) e SP80-3280 (**SP**). *x* corresponde aos valores individuais de intensidade em (A) ou de contagem de espectros de MS/MS em (B). Os *boxplots* foram gerados com as ferramentas contidas no sítio <http://boxplot.tyerslab.com/> (Spitzer *et al.*, 2014). No gráfico os números (n) se referem aos grupos de proteínas.

De acordo com a análise da Figura 19, as distribuições dos dados de intensidade (Fig. 19A) e de contagem de espectros de MS/MS (Fig. 19B) apresentaram níveis diferentes de acordo com o conjunto de proteínas considerado. O conjunto de proteínas identificadas concomitantemente entre todas as amostras (972 proteínas) se caracterizou pela ocorrência de valores médios de intensidade e de contagem de espectros de MS/MS relativamente maiores quando comparados aos valores de distribuição observados no caso de proteínas identificadas de forma compartilhada entre duas amostras (54 proteínas), que por sua vez apresentaram valores mais altos em relação às proteínas identificadas de forma exclusiva (34 proteínas) em folhas +1 de *S.*

officinarum, *S. spontaneum* ou SP80-3280. Dessa forma, de acordo com os perfis de distribuição dos dados de intensidade e de contagem de espectros, a eventual ocorrência de proteínas identificadas de forma exclusiva em folhas +1 de *S. officinarum*, *S. spontaneum* ou SP80-3280 decorre provavelmente da baixa abundância das mesmas nos proteomas estudados, uma vez que os peptídeos associados à identificação dessas proteínas apresentam valores de intensidade e de contagem de espectros relativamente baixos, sugerindo que estes peptídeos encontram-se próximos aos limites mínimos de detecção da técnica de análise por espectrometria de massas. Ainda, a identificação de proteínas exclusivas pode ser resultado de diferenças encontradas nos níveis de expressão de proteínas pouco abundantes entre os proteomas estudados, de forma que estas, quando relativamente mais expressas em uma determinada amostra, podem dar origem a peptídeos passíveis de detecção, ou seja, localizados acima dos limites mínimos de detecção do instrumento.

4.3.2. ANÁLISE DOS PROTEOMAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280 PELA ABORDAGEM GELC-MS/MS

Os proteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 foram também caracterizados pela análise por LC-MS/MS dos peptídeos tripsínicos resultantes da digestão *in gel* de fatias do gel de SDS-poliacrilamida. Após a coloração das bandas de proteínas, as áreas do gel correspondentes às amostras de proteínas de *S. officinarum*, *S. spontaneum* e SP80-3280 foram selecionadas e então separadas em 8

segmentos (Fig. 20, segmentos numerados de 1 a 8) de forma a cobrir todo o comprimento do gel.

Figura 20. Perfil eletroforético das proteínas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, extraídas de acordo com o método Fenol/SDS. As amostras de proteínas (25 µg) foram analisadas por eletroforese em gel de SDS-poliacrilamida 12% segundo as réplicas biológicas (RB). Trata-se de uma réplica da Figura 16, onde são destacadas as regiões delimitadas (1-8), que correspondem às fatias selecionadas para digestão tripsínica *in gel* e posterior análise dos peptídeos por LC-MS/MS. O gel foi submetido à coloração por Coomassie G-250. Valores à esquerda indicam a mobilidade dos marcadores de massa molecular.

Os segmentos selecionados foram seccionados e submetidos individualmente à digestão tripsínica *in gel*. A mistura de peptídeos tripsínicos, correspondente às proteínas presentes em cada segmento de gel seccionado, foi analisada por LC-MS/MS. Esse tipo

de abordagem, conhecida como GeLC-MS/MS (Dzieciatkowska *et al.*, 2014) é interessante para a análise de proteomas complexos, pois permite, ao selecionar regiões de interesse no gel, a diminuição da complexidade das amostras proteicas, minimizando os efeitos do *dynamic range* de concentração de proteínas. Exemplos clássicos deste fenômeno podem ser observados nos proteomas de células de *Saccharomyces cerevisiae* e do plasma humano, onde as diferenças de concentração entre as proteínas de maior e menor abundância se encontram na faixa de 6 a 10 ordens de magnitude, respectivamente (Anderson e Anderson, 2002; Ghaemmaghami *et al.*, 2003). Em espécies vegetais, a enzima RuBisCO, proteína mais abundante no planeta (Ellis, 1979) e responsável pelo processo de assimilação de CO₂ na fotossíntese, pode corresponder a mais da metade do total de proteínas presentes em folhas (Saravanan e Rose., 2004). Diante dessa questão, Mann e Mann (2011), em um trabalho direcionado ao estudo do proteoma de clara de ovo, cuja população de proteínas é representada majoritariamente por ovoalbumina, ovotransferrina e ovomucóide, demonstraram que a estratégia de descomplexação do proteoma alvo do estudo a partir da segmentação das proteínas presentes no gel de SDS-poliacrilamida, digestão *in gel* e análise dos peptídeos por LC-MS/MS foi capaz de identificar 158 proteínas, dentre as quais 79 foram identificadas pela primeira vez. Isso demonstra que abordagens de preparação de amostras que permitam a diminuição da complexidade do proteoma podem auxiliar na identificação de um maior número de proteínas, principalmente em proteomas dominados por uma ou mais proteínas majoritárias.

Dessa forma, a análise qualitativa dos proteomas de folhas +1 de *S. officinarum* e *S. spontaneum* e da cultivar híbrida SP80-3280 foi realizada segundo a abordagem

GeLC-MS/MS. O processamento dos 24 arquivos brutos gerados, derivados das análises de LC-MS/MS dos peptídeos tripsínicos provenientes das digestões *in gel* das proteínas presentes nos segmentos selecionados (Fig. 20) foi realizado utilizando-se o software MaxQuant, e resultou na identificação 4.185 peptídeos na amostra de *S. officinarum*, 4.163 peptídeos em *S. spontaneum* e 4.184 peptídeos em SP80-3280, resultando na identificação de 1.257 grupos de proteínas em folhas +1 de *S. officinarum* e em folhas +1 de *S. spontaneum* e, 1.245 grupos de proteínas em folhas +1 de SP80-3280 (Tab. 7; Tabela Suplementar 2).

Tabela 7. Número de peptídeos e grupos de proteínas identificados em amostras de folhas +1 (F+1) de *S. officinarum* (**S.off**), *S. spontaneum* (**S.sp**) e SP8032-80 (**SP**) a partir das análises por LC-MS/MS dos peptídeos gerados pela digestão tripsínica *in gel* das proteínas contidas nos segmentos indicados na Fig. 20.

Amostra	n. peptídeos	n. grupos únicos de proteínas*
F+1 S.off	4.185	1.257
F+1 S.sp	4.163	1.257
F+1 SP	4.184	1.245

* A presença de uma proteína em determinada amostra foi considerada de acordo com a identificação mínima de 1 peptídeo único com valor de PEP $\leq 0,01$ e ocorrência ininterrupta mínima de pelo menos 3 íons da série -y ou -b, no caso da proteína ser identificada por somente 1 peptídeo.

Quando comparada à abordagem *shotgun/bottom-up*, a abordagem GeLC-MS/MS resultou em um aumento médio de aproximadamente 23% (1.253/1.019) no número de proteínas identificadas, apesar da diminuição média de aproximadamente 15% (4.177/4.900) no número de peptídeos identificados. A Figura 21 mostra a distribuição das 1.360 proteínas identificadas nos conjuntos de segmentos de gel, correspondendo às amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280. Considerando o conjunto total de proteínas identificadas, 1.109 (81,5%) foram identificadas concomitantemente nas três amostras, enquanto que 251 (18,4%) foram identificadas ou em duas das amostras, ou apenas em uma delas de forma exclusiva.

importante nos resultados de identificação de peptídeos e proteínas provenientes da utilização das abordagens de análise *shotgun/bottom-up* e GeLC-MS/MS (Tabelas Suplementares 1 e 2). Na abordagem GeLC-MS/MS, foi constatado um aumento relativo sistemático no número de proteínas identificadas por 1, 2 ou mais peptídeos (Fig. 22) quando comparado ao número de proteínas identificadas a partir do método *Shotgun/bottom-up*.

Figura 22. Distribuição do número de proteínas presentes em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 identificadas de acordo com as abordagens de análise *shotgun/bottom-up* e GeLC-MS/MS, segundo o número de peptídeos.

Essa análise comparativa, entre as distribuições das proteínas identificadas em cada abordagem segundo o número de peptídeos, demonstrou que o uso da abordagem GeLC-MS/MS levou à identificação de um número expressivamente maior de proteínas identificadas por somente um peptídeo (542), quando comparada às identificações

derivadas da abordagem *shotgun/bottom-up* (323), correspondendo a um aumento de aproximadamente 68% no número de proteínas identificadas por um único peptídeo. Observou-se a continuidade da tendência de aumento para proteínas identificadas por dois ou mais peptídeos, que corresponderam, respectivamente, a aumentos na ordem de 36% (263/194) e 2% (555/543) no número de proteínas identificadas. Esses resultados indicam que a abordagem GeLC-MS/MS é adequada à identificação de proteínas que provavelmente se encontram em menor abundância nas amostras (normalmente identificadas por um conjunto pequeno de peptídeos), e demonstram que o processo de descomplexação do proteoma contribui de forma positiva para o aumento do número de proteínas identificadas, uma vez que o fracionamento realizado em nível de proteínas diminui os efeitos do amplo *dynamic range* de concentração de proteínas encontrado em proteomas complexos, de forma que a influência dominante dos peptídeos provenientes de proteínas mais abundantes ao longo das análises de LC-MS/MS é amenizada, favorecendo a identificação de peptídeos provenientes de proteínas presentes em menor concentração na amostra (Dzieciatkowska *et al.*, 2014).

Muito embora a abordagem GeLC-MS/MS propicie a identificação de um maior número de proteínas quando comparada com a abordagem *Shotgun/bottom-up*, uma das principais desvantagens da técnica está relacionada à baixa taxa de recuperação de peptídeos devido ao excesso de manipulação de amostra e da baixa eficiência de extração dos peptídeos da malha do gel de SDS-poliacrilamida. Estima-se que a perda de peptídeos durante todo o processo de digestão *in gel* seja algo em torno de 15-20% (Speicher *et al.*, 2000). O impacto ocasionado pela taxa mais baixa de recuperação de peptídeos observada na abordagem GeLC-MS/MS pode ser avaliado a partir da

comparação das identificações de proteínas concomitantemente identificadas pelas duas abordagens. Verificou-se a identificação concomitante de 871 proteínas pelas abordagens *Shotgun/bottom-up* e GeLC-MS/MS, identificadas por uma série de 4.735 peptídeos na abordagem *Shotgun/bottom-up* e 3.622 peptídeos na abordagem GeLC-MS/MS, indicando um decréscimo de 23,5% no número de peptídeos associados às identificações do mesmo conjunto de proteínas (dados não mostrados). Essa diferença demonstra uma tendência de diminuição média da cobertura de sequência nas identificações provenientes da abordagem GeLC-MS/MS, que pode ser observada em maiores detalhes quando as coberturas de sequência de cada proteína são comparadas entre as duas abordagens (Fig. 23).

Figura 23. Comparação entre as porcentagens de cobertura de sequência do conjunto de proteínas concomitantemente identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 ($n = 871$) a partir das abordagens de análise *shotgun/bottom-up* e GeLC-MS/MS. As proteínas cuja identificação se deu com o aumento da cobertura de sequência na abordagem *Shotgun/bottom-up* ($n = 512$) em relação à abordagem GeLC-MS/MS são representadas pelos pontos em azul ($y/x > 1$), enquanto que as proteínas cuja cobertura de sequência foi relativamente menor na abordagem *Shotgun/bottom-up* ($n = 280$) são representadas pelos pontos em vermelho ($y/x < 1$). Os pontos em verde mostram as proteínas ($n = 79$) que apresentaram o mesmo valor de cobertura de sequência de acordo com as duas abordagens. O gráfico foi gerado de acordo com a utilização das ferramentas contidas no software Perseus.

Como esperado, esta análise revelou que, para a maioria das proteínas (512) concomitantemente identificadas, houve um aumento na cobertura de sequência a partir da utilização da abordagem *Shotgun/bottom-up*, resultado do maior número de peptídeos identificados, enquanto que uma parcela significativamente menor (280) apresentou

valores de cobertura de sequência maiores quando se utilizou a abordagem GeLC-MS/MS. Somente 79 proteínas não apresentaram diferenças de cobertura de sequência de acordo com as duas abordagens utilizadas. Em conjunto, esses resultados reforçam mais uma vez que a abordagem GeLC-MS/MS, apesar de favorecer a identificação de proteínas presentes em menor abundância nas amostras, é responsável pela perda de uma parcela significativa de peptídeos, que influencia de forma negativa no grau de cobertura de sequência das proteínas.

4.3.3. ANÁLISE DOS PROTEOMAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280 PELA ABORDAGEM MUDPIT *OFF-LINE* E “*ON-LINE*”

Com o objetivo de aprimorar a caracterização dos proteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, foram realizadas análises segundo métodos alternativos de descomplexação de amostras, baseados na separação ortogonal de peptídeos combinada com a identificação por espectrometria de massas, utilizando a abordagem conhecida como MudPIT (do inglês, *Multi-Dimensional Protein Identification Technology*; Link *et al.*, 1999; Washburn *et al.*, 2001; Wolters *et al.*, 2001).

Inicialmente, os peptídeos tripsínicos derivados das digestões em solução das amostras de proteínas de folhas +1 das espécies ancestrais e da cultivar híbrida foram submetidos à metodologia MudPIT *off-line*, utilizando colunas capilares trifásicas (C-18/SCX/C-18), sem acoplamento direto ao espectrômetro de massas. Neste método os peptídeos foram separados de acordo com o aumento da força iônica do tampão de eluição e dessalinizados no segmento de C-18 localizado posteriormente ao segmento

de SCX. As 17 frações dessalinizadas, correspondentes aos conjuntos de peptídeos eluídos com concentrações diferentes de acetato de amônio (além da fração não retida), foram injetadas automaticamente pelo *autosampler* do sistema de nanocromatografia líquida acoplado ao espectrômetro de massas LTQ-Orbitrap Velos para a realização das análises por LC-MS/MS.

Adicionalmente à metodologia *off-line*, a metodologia MudPIT “*on-line*” foi realizada utilizando colunas capilares bifásicas (C-18/SCX) combinadas com o acoplamento sequencial de pré-colunas C-18 para coleta das 17 frações e, da mesma forma que na metodologia *off-line*, não houve o acoplamento direto do evento de separação ortogonal com o evento de aquisição de dados pelo espectrômetro de massas. Apesar da denominação “*on-line*” parecer fora de contexto, uma vez que não existe o acoplamento direto dos processos de separação ortogonal e aquisição dos dados, a metodologia usada neste estudo foi denominada dessa forma devido ao fato de ser conceitualmente similar à metodologia original MudPIT *on-line* (Link *et al.*, 1999; Washburn *et al.*, 2001), de forma que os peptídeos permaneceram durante todas as etapas, desde o início do fracionamento até o momento das análises por LC-MS/MS, inseridos no microambiente capilar e, portanto, não foram submetidos à exposição aos materiais que constituem as superfícies de tubos e placas utilizados na metodologia *off-line*, fato que potencialmente resultaria na retenção e perda de uma parcela significativa dos peptídeos, influenciando negativamente no número de peptídeos identificados (Magdeldin *et al.*, 2014).

Dessa forma, utilizando as abordagens MudPIT *off-line* e “*on-line*” foram identificados os peptídeos tripsínicos derivados das amostras de proteínas extraídas de

folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280. Em conjunto, 108 arquivos brutos foram gerados pelas metodologias de análise MudPIT *off-line* e “*on-line*”, sendo 54 provenientes de cada abordagem, e o processamento dos dados gerados pelas análises de LC-MS/MS foi realizado pelo software MaxQuant.

A abordagem *off-line* resultou na identificação de 17.998 peptídeos na amostra de folhas +1 de *S. officinarum*, 18.625 peptídeos em *S. spontaneum* e 18.803 peptídeos em SP80-3280, os quais permitiram a identificação de 3.328 grupos de proteínas em folhas +1 de *S. officinarum*, 3.662 em folhas +1 de *S. spontaneum* e, 3.771 em folhas +1 de SP80-3280 (Tab. 8; Tabela Suplementar 3).

Tabela 8. Número de peptídeos e grupos de proteínas identificados em amostras de folhas +1 (F+1) de *S. officinarum* (**S.off**), *S. spontaneum* (**S.sp**) e SP8032-80 (**SP**) de acordo com a metodologia MudPIT *off-line* e “*on-line*”.

Abordagem	Amostra	n. peptídeos	n. grupos únicos de proteínas*
MudPIT <i>off-line</i>	F+1 S.off	17.998	3.328
	F+1 S.sp	18.625	3.662
	F+1 SP	18.803	3.771
MudPIT “ <i>on-line</i> ”	F+1 S.off	27.133	4.351
	F+1 S.sp	28.708	5.113
	F+1 SP	28.207	4.860

* A presença de uma proteína foi considerada de acordo com a identificação mínima de 1 peptídeo único com PEP $\leq 0,01$. Em caso de identificação por apenas um peptídeo, adicionalmente ao critério de probabilidade de erro posterior, foi considerada a ocorrência ininterrupta mínima de pelo menos 3 íons da série -b ou -y nos espectros de MS/MS.

Com relação aos resultados obtidos pelo método “*on-line*”, foram identificados 27.133 peptídeos na amostra de *S. officinarum*, 28.708 peptídeos em *S. spontaneum* e 28.207 peptídeos em SP80-3280, que resultaram na identificação de 4.351 grupos de proteínas em folhas +1 de *S. officinarum*, 5.113 em folhas +1 de *S. spontaneum* e, 4.860

em folhas +1 de SP80-3280 (Tab. 8; Tabela Suplementar 4). As distribuições das proteínas identificadas a partir da utilização das metodologias MudPIT *off-line* e “*on-line*” são mostradas na Figura 24.

Figura 24. Diagrama de Venn demonstrando a distribuição do número total de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 a partir das análises de MudPIT *off-line* (A) (n= 4.211) e “*on-line*” (B) (n= 5.651). As identificações de peptídeos e proteínas foram realizadas pelo software MaxQuant a partir da utilização do banco de dados transcriptômico de cana-de-açúcar (Nishiyama-Jr *et al.*, 2014). O diagrama de Venn foi construído utilizando o software eulerAPE (Micallef e Rodgers, 2014).

A comparação entre os dados de identificação de peptídeos e proteínas de acordo com as duas abordagens de MudPIT utilizadas neste estudo revelou que a metodologia “*on-line*” foi capaz de identificar um número médio aproximadamente 52% maior de peptídeos (28.016/18.475) em relação à metodologia *off-line*, fato que resultou em um aumento de aproximadamente 33% (4.775/3.587) no número de proteínas identificadas.

Verificou-se que esse expressivo aumento no número de proteínas identificadas decorre da maior eficiência do método “*on-line*” na identificação de peptídeos em diferentes frações (Fig. 25A). Com exceção das frações 0 (não retida) e 1, a metodologia “*on-line*” apresentou valores médios de identificação de peptídeos por fração sempre maiores do que os valores médios de identificação observados de acordo com a metodologia *off-line*. Esse substancial aumento da capacidade de identificação de peptídeos em diferentes frações pela metodologia “*on-line*” contribuiu não somente para o aumento do número de proteínas identificadas, mas também para o aumento significativo nos valores de cobertura de sequência (Fig. 25B).

Figura 25. Comparação de número de peptídeos e % de cobertura de sequências obtidas com as abordagens MudPIT *off-line* e "on-line". (A) Número de peptídeos identificados nas frações eluídas com o aumento da concentração de acetato de amônio e (B) valores de cobertura de sequência dos grupos de proteínas identificados em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280. Os *boxplots* foram gerados segundo as ferramentas contidas no sítio <http://boxplot.tyerslab.com/> (Spitzer *et al.*, 2014).

Esses resultados de identificação de peptídeos e proteínas de acordo com as abordagens MudPIT *off-line* e “*on-line*”, que demonstram a maior eficiência relativa do método “*on-line*” na identificação de peptídeos e proteínas, indicam que os fatores associados à perda de peptídeos na abordagem *off-line* interferem negativamente na capacidade de identificação de proteínas. Observou-se que por meio desta abordagem foram identificadas 1.440 proteínas a menos, ou seja, um total 25,5% menor em relação à abordagem “*on-line*”. Perdas relativamente maiores, acima de 40%, já foram reportadas por Magdeldin *et al.* (2014) quando as abordagens *off-line* (com a injeção automática das amostras pelo *autosampler* do sistema de cromatografia) e *on-line* foram comparadas. Dessa forma, na metodologia MudPIT *off-line*, a excessiva manipulação das amostras e exposição dos peptídeos às superfícies constituídas de polipropileno resultam na perda de uma parcela significativa da população de peptídeos, constituída principalmente por peptídeos mais hidrofóbicos, devido à adsorção dos mesmos nas superfícies de tubos e frascos, além das perdas que podem ocorrer durante o processo de injeção das amostras de peptídeos em colunas de fase reversa (Stejskal *et al.*, 2013; van Midwoud *et al.*, 2007; Warwood *et al.*, 2013). Ao se comparar o conjunto de peptídeos identificados pelas abordagens *off-line* e “*on-line*” segundo o grau individual de hidrofobicidade (GRAVY ou Kyte–Doolittle score; Kyte e Doolittle, 1982), verificou-se que o conjunto de peptídeos identificados pelo método “*on-line*” apresentou valores levemente maiores de GRAVY score (Fig. 26), indicando uma significativa perda da população de peptídeos hidrofóbicos quando se utilizou a abordagem *off-line*.

Figura 26. Boxplots demonstrando as distribuições de GRAVY score dos peptídeos identificados em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 pelas metodologias de MudPIT *off-line* e “*on-line*”. Valores positivos de GRAVY score denotam maior hidrofobicidade enquanto que valores negativos denotam menor hidrofobicidade. Para os cálculos do GRAVY score, foi utilizada a ferramenta ProtParam presente nas ferramentas de análise proteômica do *Swiss Institute of Bioinformatics* no sítio <http://www.expasy.org/proteomics>. Os boxplots foram gerados segundo as ferramentas contidas no sítio <http://boxplot.tyerslab.com/> (Spitzer *et al.*, 2014)

Os resultados de identificação de peptídeos e de proteínas de acordo com a utilização das abordagens MudPIT *off-line* e “*on-line*” demonstraram que a metodologia “*on-line*” apresentou uma eficiência significativamente maior na identificação de peptídeos, influenciada principalmente pela minimização dos passos de manipulação das amostras e consequente diminuição da exposição dos peptídeos às superfícies e materiais que estão associados às perdas ocasionadas por adsorção.

4.3.4. AVALIAÇÃO GLOBAL DAS ABORDAGENS DE ANÁLISE DOS PROTEOMAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280

Uma vez que os dados derivados dos métodos *Shotgun/bottom-up*, GeLC-MS/MS e MudPIT (*off-line* e “*on-line*”), aplicados sobre os proteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, foram avaliados segundo critérios bem definidos de identificação de peptídeos e proteínas, estes foram submetidos a uma avaliação global de forma a obter um panorama dos resultados de identificação gerados. Os resultados sumarizados de identificação de peptídeos e proteínas referentes a cada abordagem de análise proteômica podem ser vistos na Tabela 9.

Tabela 9. Número de peptídeos e proteínas identificados em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com as abordagens *Shotgun/bottom-up*, GeLC-MS/MS e MudPIT *off-line* e “*on-line*”.

Amostra	Número de peptídeos/proteínas identificados de acordo com a abordagem de análise dos proteomas de folhas +1 de cana-de-açúcar			
	<i>Shotgun/bottom-up</i>	GeLC-MS/MS	MudPIT <i>off-line</i>	MudPIT “ <i>on-line</i> ”
F+1 <i>S. officinarum</i>	4.922/1.030	4.185/1.257	17.998/3.328	27.133/4.351
F+1 <i>S. spontaneum</i>	4.876/1.010	4.163/1.257	18.625/3.662	28.708/5.113
F+1 SP80-3280	4.901/1.018	4.184/1.245	18.803/3.771	28.207/4.860
n. total de proteínas*	1.060	1.360	4.211	5.651

* de acordo com o critério de identificação de pelo menos 1 peptídeo único com valor de PEP $\leq 0,01$ e pelo menos 8 íons anotados das séries -b ou -y. Adicionalmente, para proteínas identificadas por somente 1 peptídeo, foram consideradas as identificações com pelo menos 3 íons consecutivos das séries -b ou -y.

Comparando-se os resultados de identificação de peptídeos e proteínas apresentados na Tabela 9, fica clara a contribuição dos métodos de descomplexação das amostras de peptídeos ou proteínas para o aumento do número de proteínas identificadas. As abordagens que se utilizaram de alguma estratégia de descomplexação de amostras, seja em nível de proteínas ou peptídeos, levaram à identificação de uma população maior de proteínas, quando comparadas com a abordagem *Shotgun/bottom-up*. Em relação à esta última, observou-se um aumento discreto de aproximadamente 1,3 vezes no número total de proteínas identificadas quando se utilizou a abordagem GeLC-MS/MS, enquanto que aumentos expressivos, em torno de 4 a 5,3 vezes, foram observados quando se utilizaram as abordagens de análise MudPIT *off-line* e “*on-line*”, respectivamente. O gradual aumento observado deixa clara a necessidade da utilização de métodos de descomplexação das amostras de proteínas ou peptídeos para o aumento da capacidade e eficiência de identificação de proteínas em amostras complexas.

A abordagem *Shotgun/bottom-up* foi inicialmente escolhida para o estudo dos proteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 devido à relativa simplicidade metodológica. Nessa abordagem, os peptídeos derivados do processo de digestão tripsínica em solução das amostras de proteínas extraídas das folhas +1 de cana-de-açúcar foram analisados diretamente por LC-MS/MS, sem a utilização de passos adicionais de preparação de amostras e resultou na identificação de um conjunto de 1.060 proteínas presentes em folhas +1 das espécies ancestrais e da cultivar híbrida. Alternativamente, a abordagem GeLC-MS/MS foi realizada com o intuito de se avaliar a influência dos métodos de descomplexação de amostras de proteínas na eficiência e na

capacidade de identificação de peptídeos. Apesar da diminuição observada no número de peptídeos identificados quando comparada à abordagem *Shotgun/bottom-up*, por meio dessa metodologia foi possível identificar um número maior de proteínas (1.360), revelando que a diminuição da complexidade das amostras de proteínas e, conseqüentemente, das amostras de peptídeos submetidos às análises por LC-MS/MS contribuiu para a identificação de proteínas menos abundantes nos proteomas estudados. Diante dessas questões, as metodologias MudPIT *off-line* e “*on-line*” foram realizadas de forma a otimizar o processo de descomplexação de amostras de peptídeos, contornando de forma mais eficiente os problemas relacionadas aos efeitos do *dynamic range* de concentração de proteínas, e, ao mesmo tempo, de forma a diminuir a perda excessiva de peptídeos decorrentes do excesso de manipulação de amostras e da baixa eficiência de extração de peptídeos em abordagens baseadas na digestão *in gel* de proteínas.

Os resultados das identificações de peptídeos e proteínas nas amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 foram combinados segundo a abordagem de análise utilizada e os conjuntos de peptídeos e proteínas identificados em cada abordagem foram avaliados segundo sua distribuição (Fig. 27).

Figura 27. Números médios de proteínas (A) e peptídeos (B) identificados em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com as diferentes abordagens de análise utilizadas.

De acordo com os dados apresentados na Figura 27, em média, 1.019, 1.253, 3.587, e 4.775 proteínas (Fig. 27A) foram identificadas a partir dos conjuntos de 4.899, 4.177, 18.475, e 28.016 peptídeos identificados (Fig. 27B), de acordo com a utilização das abordagens *Shotgun/bottom-up*, GeLC-MS/MS, e MudPIT *off-line* e “*on-line*”, respectivamente. Os valores médios de identificação de proteínas e peptídeos observados nas abordagens MudPIT apresentaram diferenças significantes em relação aos outros métodos (*Shotgun/bottom-up* e GeLC-MS/MS). A abordagem MudPIT “*on-line*”, que propiciou a identificação de um número mais alto de peptídeos e proteínas

dentre todas as abordagens, apresentou diferenças significativas também em relação ao método correlato MudPIT *off-line*, com aumentos observados na ordem de 51% e 33% no número de peptídeos e proteínas identificados, respectivamente, corroborando mais uma vez com a superioridade observada dos métodos de descomplexação de amostras baseados na abordagem MudPIT acerca da capacidade de identificação de peptídeos e proteínas. Observa-se que a taxa de recuperação de peptídeos é significativamente maior na abordagem MudPIT “*on-line*” em comparação à abordagem *off-line*, demonstrando que a maior manipulação de amostra e maior exposição às superfícies plásticas de tubos e *vials* utilizados para as coletas de frações ou para a injeção de amostras configuram-se como reais fontes responsáveis pela perda significativa de peptídeos.

As abordagens MudPIT foram responsáveis pelo incremento expressivo no número de peptídeos identificados. Dessa forma, sobre os dados de identificação de peptídeos e proteínas de acordo com a abordagem utilizada, foi realizada uma análise comparativa para verificar qual seria a influência do maior número de peptídeos identificados sobre a cobertura de sequência de proteínas (Fig. 28).

Figura 28. Comparação entre as coberturas de sequências dos conjuntos totais de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 segundo as abordagens *Shotgun/bottom-up* (n= 1.060), *GeLC-MS/MS* (n= 1.360) e análise *MudPIT off-line* (n= 4.211) e "on-line" (n= 5.651). A distribuição total dos valores de cobertura de sequência de acordo com cada abordagem pode ser visualizada em (A) e, em (B), a distribuição relativa do número de proteínas identificadas por abordagem, de acordo com intervalos determinados de cobertura de sequência.

A análise sobre os valores de cobertura de sequência mostrou uma maior correlação entre as distribuições dos valores de cobertura observadas nas abordagens MudPIT *off-line* e *Shotgun/bottom-up*. O valor médio observado de cobertura de sequência nessas duas abordagens foi de aproximadamente 10,7%. A abordagem MudPIT *“on-line”*, por sua vez, apresentou valor médio de cobertura de sequência em torno de 12,9%, enquanto a abordagem GeLC-MS/MS, cujo perfil de distribuição de valores de cobertura diferiu significativamente das demais distribuições, resultou em uma cobertura de sequência média de 7,2%, que foi o menor valor dentre todos os valores médios observados (Fig. 28A). Dentre as proteínas identificadas por essa abordagem, observa-se a elevada contribuição de proteínas cuja cobertura apresentou valores até 10%. Aproximadamente 62% das proteínas identificadas pela abordagem GeLC-MS/MS apresentaram valores de cobertura de sequência menor ou igual a 10% enquanto que 38% apresentaram valores de cobertura distribuídos acima de 10%. Em situação oposta, na abordagem MudPIT *“on-line”*, somente 38% das proteínas apresentaram valores de cobertura abaixo ou igual a 10% enquanto que 62% apresentaram valores distribuídos acima de 10% (Fig. 28B). Perfis semelhantes foram observados para os conjuntos de proteínas identificadas pelas abordagens MudPIT *off-line* e *Shotgun/bottom-up*, indicando que as três abordagens resultam em uma tendência geral de aumento na cobertura de sequência quando comparadas à abordagem GeLC-MS/MS.

Ainda considerando os conjuntos de peptídeos identificados segundo as diferentes abordagens utilizadas neste estudo, foi realizada uma análise para avaliar a distribuição das massas e de GRAVY score dos peptídeos associados aos conjuntos de

proteínas identificadas em folhas +1 de cana-de-açúcar, a partir das abordagens *Shotgun/bottom-up*, GeLC-MS/MS e MudPIT *off-line* e “*on-line*” (Fig. 29).

Figura 29. Comparação entre as massas (A) e GRAVY score (B) dos peptídeos identificados em proteínas presentes em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 segundo as abordagens *Shotgun/bottom-up* (n= 4.954), GeLC-MS/MS (n= 4.326) e análise MudPIT *off-line* (n= 19.502) e “*on-line*” (n= 29.598). Valores positivos de GRAVY score denotam maior hidrofobicidade enquanto que valores negativos denotam menor hidrofobicidade das sequências de aminoácidos. Para os cálculos do GRAVY score, foi utilizada a ferramenta ProtParam presente nas ferramentas de análise proteômica do *Swiss Institute of Bioinformatics*, no sítio <http://www.expasy.org/proteomics>. Os *boxplots* foram gerados segundo as ferramentas contidas no sítio <http://boxplot.tyerslab.com/> (Spitzer *et al.*, 2014).

Verificou-se que as abordagens *Shotgun/bottom-up* e MudPIT *off-line* apresentaram perfis semelhantes de distribuição e dispersão dos valores de massas dos peptídeos identificados, apresentando médias de massa em torno de 1.600 Daltons. O valor médio de massa de peptídeos foi superior na abordagem MudPIT “*on-line*”

(aproximadamente 1.800 Da) e inferior na abordagem GeLC-MS/MS (aproximadamente 1.400 Da). Foi interessante notar que a distribuição e dispersão das massas dos peptídeos identificados a partir da abordagem GeLC-MS/MS estão preferencialmente localizadas em regiões correspondentes às massas relativamente mais baixas, quando comparadas às distribuições das massas de peptídeos identificados pelas abordagens *Shotgun/bottom-up* e MudPIT *off-line* e “*on-line*”. Uma vez que esta é a única abordagem que se baseia na digestão e extração de peptídeos *in gel*, uma possível explicação para este fato seria que a eficiência de extração de peptídeos se torna menor quanto maior a massa do peptídeo, uma vez que a capacidade de difusão dos mesmos se torna menor quando comparada a capacidade de difusão de peptídeos de menor massa (geralmente menores) (Fig. 29A). De uma maneira geral não foram observadas diferenças significantes no índice de hidrofobicidade (GRAVY score) dos peptídeos identificados de acordo com a utilização das diferentes abordagens (Fig. 29B). Um ligeiro deslocamento negativo da distribuição dos valores de GRAVY score, observado no conjunto de peptídeos identificados a partir da abordagem MudPIT *off-line*, pode sugerir a perda de peptídeos hidrofóbicos quando comparado aos valores de GRAVY score dos peptídeos identificados pela abordagem “*on-line*”.

4.3.5. AGRUPAMENTO DOS DADOS DE IDENTIFICAÇÃO DE PROTEÍNAS PROVENIENTES DAS ABORDAGENS DE ANÁLISE *SHOTGUN/BOTTOM-UP*, *GELC-MS/MS* E *MUDPIT OFF-LINE* E “*ON-LINE*” APLICADAS SOBRE O ESTUDO DOS PROTEOMAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280

As abordagens de análise dos proteomas de amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 foram avaliadas de forma criteriosa e,

comparadas de acordo com a eficiência e capacidade de identificação de peptídeos e proteínas, demonstrando que as abordagens baseadas na descomplexação de peptídeos (MudPIT *off-line* e “*on-line*”) são capazes de identificar um número significativamente maior de peptídeos e proteínas em relação às outras abordagens de análise proteômica utilizadas neste estudo (*Shotgun/bottom-up* e GeLC-MS/MS). Avançando um pouco mais nas análises que apontam a superioridade das abordagens MudPIT e, buscando permitir um eventual aumento do número de proteínas identificadas nos proteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, os conjuntos totais de proteínas identificadas pelas abordagens MudPIT *off-line* e “*on-line*”, GeLC-MS/MS e *Shotgun/bottom-up* (Tab. 9) foram comparados de forma a obter uma distribuição global das identificações de acordo com cada abordagem. Dessa forma, inicialmente, as identificações derivadas das abordagens MudPIT *off-line* e “*on-line*” foram combinadas, uma vez que estas foram responsáveis pelo maior número de proteínas identificadas entre todas as abordagens utilizadas, resultando em um número total de 5.785 proteínas identificadas. Essa lista de proteínas identificadas pela combinação das abordagens MudPIT foi utilizada como base para a comparação com as listas de identificações derivadas das abordagens *Shotgun/bottom-up* e GeLC-MS/MS uma vez que, como já discutido anteriormente, as abordagens MudPIT resultaram no aumento significativo do número de peptídeos e proteínas identificados, e que, foi acompanhado pela tendência de aumento das coberturas de sequências. Considerando isso, a lista resultante das identificações de proteínas das abordagens MudPIT *off-line* e “*on-line*” foi então submetida à comparação com as listas de identificações de proteínas derivadas das abordagens de análise *Shotgun/bottom-up* e GeLC-MS/MS. Esta análise

revelou que o conjunto de proteínas identificadas pelas abordagens MudPIT *off-line* e “*on-line*” foi suficientemente capaz de identificar quase que completamente os conjuntos de proteínas identificadas pelas outras duas abordagens (Fig. 30).

Figura 30. Diagrama de Venn mostrando a comparação do número de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com as diferentes abordagens de análise proteômica utilizadas neste estudo (*Shotgun/bottom-up*, GeLC-MS/MS e MudPIT *off-line* e “*on-line*”). Para esta análise foram utilizados os conjuntos de proteínas totais mostrados na Tabela 9 e Tabelas Suplementares 1, 2, 3 e 4. O diagrama de Venn foi construído utilizando o software eulerAPE (Micallef e Rodgers, 2014).

Sobre o conjunto de proteínas identificadas pelas abordagens MudPIT, foram adicionadas somente 40 proteínas, sendo grande parte proveniente das identificações

derivadas da utilização da abordagem GeLC-MS/MS (25) e o restante, derivado das identificações únicas, de acordo com a utilização da abordagem *Shotgun/bottom-up* (8) e, das identificações compartilhadas entre as duas abordagens (7). Dessa forma, este pequeno conjunto constituído por apenas 40 proteínas que foram somente identificadas pelas abordagens GeLC-MS/MS e *Shotgun/bottom-up*, foi adicionado ao conjunto total de 5.785 proteínas identificadas pela abordagem MudPIT, resultando em um número total de 5.825 proteínas identificadas (Tabela Suplementar 5), assim distribuídas: 4.784 proteínas identificadas em folhas +1 de *S. officinarum*, 5.400 em folhas +1 de *S. spontaneum* e 5.293 em folhas +1 de SP80-3280. Esses números representam o máximo de cobertura do proteoma de folhas +1 de cana-de-açúcar alcançado neste estudo, segundo os critérios adotados de identificação de peptídeos e proteínas. Considerando a estimativa de que o número de genes em cana-de-açúcar é de aproximadamente 33.000 (Vettore *et al.*, 2003) e, considerando a hipótese mínima, mas não real, de que um gene origina uma proteína, neste estudo foram identificados aproximadamente 18% (5.825/33.000) dos produtos gênicos esperados em cana-de-açúcar. Esse conjunto de proteínas, até o momento, corresponde ao maior conjunto de proteínas identificadas em cana-de-açúcar.

Dessa forma, o número total de proteínas identificadas em folhas +1 de cana-de-açúcar foi avaliada segundo a distribuição entre as amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 (Fig. 31).

Figura 31. Distribuição do conjunto de proteínas identificadas em amostras de folhas +1 de *S. officinarum* (n= 4.784), *S. spontaneum* (n= 5.400) e SP80-3280 (n= 5.293) considerando a combinação das identificações de proteínas de acordo com as abordagens de análise MudPIT *off-line* e “*on-line*”, Shotgun/*bottom-up* e GeLC-MS/MS.

Como já observado nas análises das proteínas identificadas pelas diferentes abordagens utilizadas neste estudo, verificou-se a alta concordância de identificação entre os proteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280. Das

5.825 proteínas identificadas, 4.303 (74%) correspondem às proteínas identificadas concomitantemente entre as três amostras. Uma parcela minoritária, correspondendo a aproximadamente 8% do número total de proteínas, foi identificada de forma exclusiva nas amostras de *S. officinarum* (130), *S. spontaneum* (190) e SP80-3280 (156). Foi interessante notar a alta proporção de proteínas compartilhadas entre a espécie selvagem *S. spontaneum* e a cultivar híbrida SP80-3280. Aproximadamente 12% (695) do número total de proteínas foram identificadas como pertencentes às duas amostras, enquanto apenas 2% (139) foram identificadas concomitantemente na espécie nobre (*S. officinarum*) e na cultivar híbrida. Essa expressiva diferença deve decorrer do menor número de proteínas identificadas em folhas +1 de *S. officinarum* em relação ao número de identificações derivado das amostras de *S. spontaneum* e SP80-3280. As espécies ancestrais compartilham entre si 212 proteínas, correspondendo a aproximadamente 4% do número total de proteínas identificadas.

Com o intuito de se verificar os possíveis papéis funcionais das proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, o conjunto total de proteínas foi submetido à classificação segundo as categorias Componente Celular, Processo Biológico e Função Molecular do *Gene Ontology* (GO) (Fig. 32).

Figura 32. Anotação funcional do conjunto de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 (n= 5.825) segundo as categorias Componente Celular, Processo Biológico e Função Molecular do *Gene Ontology*. Nesta análise foram mostrados somente os termos cuja representatividade foi maior ou igual a 1%.

De acordo com a análise da Figura 32, verifica-se que as proteínas identificadas em folhas +1 das espécies ancestrais e da cultivar híbrida estão distribuídas em diversos compartimentos celulares, destacando-se principalmente a localização citoplasmática, nuclear e cloroplástica. Uma parcela correspondente a 13% (denotada pelas classificações *no match* e *cellular component*) não apresentou anotação de localização conhecida ou determinada no período correspondente à realização das buscas. Os termos que apresentaram valores correspondentes a proporções menores que 1% foram agrupados e classificados como “*other*” para evitar a poluição visual nos gráficos. Apesar de ser possível observar as contribuições das proteínas identificadas em processos biológicos relacionados às vias das pentoses (*pentose-phosphate shunt*) e tradução de proteínas, por exemplo, e que atuam molecularmente na interação com proteínas e outros substratos e na catálise enzimática, as proporções relativamente altas de proteínas não anotadas ou não classificadas nas categorias Processo Biológico (29%) e Função Molecular (33%), interferem na interpretação funcional global, uma vez que não é possível determinar qual é a real contribuição dos termos dentro das categorias do *Gene Ontology* no conjunto de proteínas identificadas. Com o objetivo de contornar esse problema de inferência, o conjunto de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 foi submetido à análise de enriquecimento de termos segundo as diferentes categorias do *Gene Ontology*, utilizando as ferramentas do software GeneMerge v.1.4 (Castillo-Davis e Hartl, 2003). Este, basicamente realiza uma análise estatística sobre os genes ou proteínas que representam a população de estudo (no caso, as 5.825 proteínas), identificando categorias funcionais enriquecidas de acordo com a comparação entre as frequências dos termos encontrados na população

de estudo e na população total (idealmente representada pelo genoma). Nesse caso, a representatividade calculada em cada categoria é denotada por números que variam de zero a um (*p-value*), onde valores mais próximos de zero representam termos com maior enriquecimento na população de estudo.

Os resultados dessa análise, mostrando a lista completa dos termos enriquecidos no conjunto de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com as categorias do *Gene Ontology*, são apresentados na Tabela Suplementar 6 e, na Figura 33 estão destacados os 20 termos mais representativos de cada categoria do *Gene Ontology* no conjunto considerado de proteínas.

Figura 33. Termos enriquecidos no conjunto de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 segundo as categorias funcionais Componente Celular (A), Processo Biológico (B) e Função Molecular (C) do *Gene Ontology*. A análise de enriquecimento de termos, aplicada sobre o conjunto total de proteínas identificadas em folhas +1 ($n = 5.825$), foi realizada pelas ferramentas contidas no software GeneMerge v1.4 (Castillo-Davis e Hartl, 2003). São mostrados somente os 20 termos mais representativos em cada categoria.

De acordo com a análise dos resultados mostrados na Tabela Suplementar 6 e Figura 33, observa-se, em concordância com a natureza fotossintetizante dos tecidos dos quais as proteínas foram identificadas, a alta representatividade de organelas ou regiões relacionadas aos processos de fotossíntese e metabolismo de carbono, como por exemplo os cloroplastos, os plastídeos, estroma e vacúolo (Fig. 33A). Da mesma maneira, observa-se que os processos biológicos enriquecidos correspondem aos eventos associados ao metabolismo fotossintético, ao metabolismo de carboidratos, à tradução de proteínas e à resposta aos estresses biótico e abiótico (Fig. 33B) e, de forma mais específica, o conjunto de proteínas identificadas destaca-se pela atuação em interações não covalentes com outras proteínas ou moléculas (DNA ou RNA, por exemplo) e pela atividade enzimática (Fig. 33C), modulando diversos processos fisiológicos.

As proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 foram submetidas à anotação de acordo com as classificações contidas nas bibliotecas do banco de dados *Superfamily* (Gough *et al.*, 2001) para avaliar quais superfamílias são encontradas com maior frequência nos proteomas de folhas +1 das espécies ancestrais e da cultivar híbrida. De acordo com a análise dos resultados das buscas considerando a classificação do *Superfamily* (Tabela Suplementar 7), as proteínas identificadas apresentaram-se distribuídas em mais de 300 superfamílias, sendo que as cinco categorias de maior ocorrência nos três proteomas considerados são mostradas na Tabela 10.

Tabela 10. Categorias de maior ocorrência em proteínas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com as classificações presentes na biblioteca do banco de dados *Superfamily*.

Classificação <i>Superfamily</i>	<i>S. officinarum</i>	<i>S. spontaneum</i>	SP80-3280
<i>Unknown</i>	1581	1814	1782
<i>P-loop containing nucleoside triphosphate hydrolase superfamily</i>	293	321	316
<i>FAD/NAD(P)-binding Rossmann fold Superfamily</i>	278	306	299
<i>Tetratrico peptide repeat superfamily</i>	151	174	168
<i>Protein kinase superfamily</i>	126	170	155

Observou-se que, nos conjuntos de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, as principais categorias do *Superfamily* foram representadas por proteínas sem anotação conhecida (“*unknown*”) e por proteínas pertencentes às superfamílias “*P-loop containing nucleoside triphosphate hydrolase*”, “*FAD/NAD(P)-binding Rossmann fold*”, “*Tetratrico peptide repeat*” e “*Protein kinase*”, que juntas englobam famílias amplamente representadas por proteínas (incluindo enzimas) que exercem uma variedade de funções, que incluem desde processos relacionados ao enovelamento de proteínas e transporte, até processos envolvidos em vias de sinalização celular, a partir, principalmente, da atuação de proteínas quinases. A superfamília “*P-loop containing nucleoside triphosphate hydrolase*” é caracterizada estruturalmente por dois motivos, Walker A e Walker B (Walker *et al.*, 1982), que participam, respectivamente, na interação com moléculas de NTP (do inglês *nucleoside triphosphate*) e o cátion metálico magnésio. É representada por uma grande variedade de proteínas, incluindo quatro classes de enzimas, que apresentam capacidade de interagir com nucleotídeos (*nucleotide-binding proteins*) e exercer funções de regulação, de transporte, de transdução de sinal e funções catalíticas, entre outras (Kawamura *et al.*, 2003).

Outra superfamília representada no proteoma de folhas +1, denominada “*FAD/NAD(P)-binding Rossmann fold superfamily*”, é composta por uma gama de enzimas que não compartilham entre si homologia de sequência, mas apresentam motivos estruturais de ligação à co-enzimas, chamados de dobra de Rossmann (Rossmann *et al.*, 1975). Essas enzimas estão envolvidas na catálise de reações redox e não redox que utilizam as co-enzimas flavina adenina dinucleotídeo (FAD, do inglês *flavin adenine dinucleotide*), nicotinamida adenina dinucleotídeo (NAD, do inglês *nicotinamide adenine dinucleotide*) e a nicotinamida adenina dinucleotídeo fosfato (NADP, do inglês *nicotinamide adenine dinucleotide phosphate*) para exercer diversas atividades biológicas (Hua *et al.*, 2014). Dentro dessa superfamília encontram-se diferentes enzimas, incluindo a álcool desidrogenase, a gliceraldeído-3-fosfato desidrogenase, a lactato desidrogenase, a glicose-6-fosfato desidrogenase, diversas metiltransferases, oxidorreduções, dentre outras. A superfamília “*Tetratrico peptide repeat*” é representada pelo motivo estrutural de mesmo nome, constituído pela repetição (3 a 16 repetições) de uma sequência consenso de 34 resíduos de aminoácidos e que é caracterizado por ser um motivo de interação proteína-proteína, frequentemente encontrado em complexos multiproteicos (Hirano *et al.*, 1990; Sikorski *et al.*, 1990; Lamb *et al.*, 1995). Esta superfamília, da mesma forma que as outras citadas, inclui diversas proteínas com funções distintas envolvidas em processos de regulação do ciclo celular, controle transcricional, transporte e enovelamento de proteínas (Das *et al.*, 1998). As proteínas quinases compõem uma das superfamílias mais representadas em eucariotos. As enzimas que constituem essa superfamília utilizam ATP (ou GTP) em reações de transferência de grupos fosfato (fosforilação) em proteínas-alvos aos resíduos de serina,

treonina ou tirosina, modulando a atividade, localização e função dessas proteínas. A fosforilação se configura como um dos principais meios de regulação da sinalização celular em eucariotos (Hanks e Hunter, 1995; Champion *et al.*, 2004). A análise comparativa entre o genoma de *Arabidopsis* e o genoma humano revelou que a predição de proteínas quinases no genoma da espécie vegetal é cerca de duas vezes maior do que o número predito em humanos e estima-se que aproximadamente 30% das proteínas codificadas pelo genoma de *Arabidopsis* sejam alvos de fosforilação (Champion *et al.*, 2004). A alta representatividade dessa superfamília em folhas +1 de cana-de-açúcar, portanto, só reitera o papel fundamental da atividade de proteínas quinases em eventos de transdução de sinais em plantas.

Buscando avaliar a possível contribuição das espécies ancestrais no fenótipo apresentado pela cultivar híbrida, as proteínas identificadas concomitantemente entre *S. officinarum* e SP80-3280 (n= 139) e entre *S. spontaneum* e SP80-3280 (n= 695) (Fig. 31) foram submetidas à análise de enriquecimento de termos considerando as anotações presentes nas categorias Função Molecular e Processo Biológico do *Gene Ontology* (Tab. 11).

Tabela 11. Termos enriquecidos no conjunto de proteínas identificadas concomitantemente em folhas +1 de *S. officinarum* e SP80-3280 (n= 139) e em folhas +1 de *S. spontaneum* e SP80-3280 (n= 695) segundo as categorias Função Molecular e Processo Biológico do *Gene Ontology*.

<i>S. officinarum</i> e SP80-3280 (n= 139)			
Função Molecular	<i>p-value</i>	Processo Biológico	<i>p-value</i>
<i>molecular_function</i>	5.94E-03	<i>response to other organism</i>	6.54E-04
<i>S. spontaneum</i> e SP80-3280 (n= 695)			
Função Molecular	<i>p-value</i>	Processo Biológico	<i>p-value</i>
<i>molecular_function</i>	1.78E-12	<i>biological_process</i>	8.36E-12
<i>protein binding</i>	1.63E-11	<i>fatty acid beta-oxidation</i>	6.67E-09
<i>DNA binding</i>	4.47E-05	<i>protein import into peroxisome matrix</i>	2.07E-05
<i>chromatin binding</i>	4.35E-03	<i>N-terminal protein myristoylation</i>	1.88E-05

De acordo com essa análise, demonstrou-se que o único processo biológico que apresentou enriquecimento dentro do conjunto de proteínas identificadas concomitantemente em folhas +1 de *S. officinarum* e SP80-3280 (n= 139) foi o processo relacionado à “resposta a outros organismos”, principalmente representado pelas proteínas osmotina, lipoxigenase 1 e monooxigenase 1 dependente de flavina. Esse conjunto de proteínas está envolvido em processos de resposta de defesa, apresentando expressão induzida em resposta ao ataque de patógenos. A osmotina, proteína multifuncional isolada inicialmente em células de *Nicotiana tabacum* (Singh *et al.*, 1985) pertencente à família 5 das proteínas relacionadas à patogênese (PRPs) (van Loon, 1985; Sels *et al.*, 2008), apresenta atividade antifúngica sobre diversas espécies, atuando na regulação da resposta ao estresse a partir da modulação da atividade de fatores de transcrição ou na modulação da sinalização celular (Abdin *et al.*, 2011). As lipoxigenases são enzimas amplamente distribuídas em plantas e animais e catalisam a

reação de adição de oxigênio a moléculas de ácidos graxos polinsaturados. Em plantas, seus substratos são convertidos em uma variedade de moléculas bioativas envolvidas em diversos processos fisiológicos e de defesa (Brash, 1999; Porta e Rocha-Sosa, 2002). Em tabaco, foi observado um aumento relativamente antecipado nos níveis de expressão e de atividade dessa enzima durante a infecção por *Phytophthora parasitica* var *nicotianae* em plantas resistentes quando comparado com plantas suscetíveis, indicando a participação dessa enzima na defesa contra as infecções por fungo (Rancé *et al.*, 1998). A monooxigenase 1 dependente de flavina apresenta papel fundamental no evento de resistência sistêmica adquirida (SAR, do inglês *systemic acquired resistance*), de forma que plantas de *Arabidopsis knockout* para o gene que codifica essa enzima apresentam deficiência sistêmica na síntese de moléculas sinalizadoras e na expressão de genes de defesa (como proteínas pertencentes à família das PRPs) durante a infecção por cepas avirulentas e virulentas de bactérias *Pseudomonas syringae* pv *maculicola* (Mishina e Zeier, 2006). Diante disso, observa-se, no conjunto de proteínas identificadas compartilhadas entre folhas +1 de *S. officinarum* e SP80-3280, uma alta contribuição de proteínas relacionadas ao processo de resposta ao estresse biótico induzidos por fungos e bactérias.

Com relação às proteínas identificadas concomitantemente em *S. spontaneum* e SP80-3280 (n= 695) (Tab. 11), observou-se a alta representatividade dos processos biológicos “beta-oxidação de ácidos graxos” e “importação de proteínas para a matriz peroxissomal”, evidenciado a participação de proteínas envolvidas na atividade da organela peroxissomo nesse conjunto de proteínas comuns. Os peroxissomos são organelas que apresentam diferentes funções em células eucarióticas e contêm um

arsenal de enzimas e moléculas envolvidas na produção e metabolização das espécies reativas de oxigênio (ROS, do inglês *Reactive Oxygen Species*). A síntese de ROS nos peroxissomos, principalmente de H_2O_2 , se dá através da via respiratória ou pela metabolização dos ácidos graxos, a partir da reação de beta-oxidação, e do glicolato produzido pela fotorrespiração (Schlüter *et al.*, 2010). Sob a influência de diferentes agentes causadores de estresse, observa-se o aumento sistêmico na produção dessas espécies que, quando em excesso, causam danos oxidativos sobre diversos componentes celulares, incluindo os lipídeos, as proteínas e os ácidos nucleicos. Ao mesmo tempo, a síntese de ROS é acompanhada por mecanismos enzimáticos e não enzimáticos de controle dos níveis dessas espécies reativas, de forma a manter o equilíbrio de concentração intracelular e impedir seus efeitos oxidativos deletérios sobre a célula. Nesse contexto, a síntese e o controle dos níveis de H_2O_2 nos peroxissomos têm sido associados às vias de transdução de sinal em resposta ao estresse, interferindo na expressão de genes de defesa (Lopez-Huertas *et al.*, 2000). Portanto, o enriquecimento dos termos relacionados aos eventos de beta-oxidação e translocação de proteínas para dentro dos peroxissomos, evidencia a importância dos peroxissomos na síntese e metabolismo de H_2O_2 , sugerindo que a resposta ao estresse é um fator importante no conjunto de proteínas identificadas concomitantemente em *S. spontaneum* e SP80-3280.

Outro termo enriquecido na categoria Processos Biológicos, a “N-miristoilação de proteínas”, apresenta relação com os mecanismos de resposta ao estresse, principalmente aquele ocasionado por fatores abióticos. Essa modificação co-traducional é caracterizada pela ligação covalente do ácido mirístico (ácido tetradecanóico) ao

resíduo de glicina localizado na posição N-terminal do polipeptídeo nascente e pode ser encontrada em proteínas envolvidas na transdução de sinal, incluindo as proteínas serina/treonina quinases, tirosina quinases, fosfatases, proteínas G, proteínas ligadoras de cálcio e *ADP ribosylation factors* (Ishitani *et al.*, 2000). Em conjunto, os resultados da análise de enriquecimento de termos, realizada sobre o conjunto de proteínas identificadas concomitantemente em folhas +1 de *S. spontaneum* e SP80-3280, indicam que estas participam em processos primariamente envolvidos na resposta ao estresse.

A análise de proteínas identificadas de forma exclusiva pode colaborar no entendimento das características fenotípicas apresentadas pelas espécies ancestrais e pela variedade híbrida. Dessa forma, as proteínas identificadas exclusivamente em folhas +1 de *S. officinarum* (n= 130), de *S. spontaneum* (n= 190) e de SP80-3280 (n= 156) foram avaliadas de acordo com as categorias do *Gene Ontology* e submetidas à análise de enriquecimento de termos (Tab. 12).

Tabela 12. Termos enriquecidos no conjunto de proteínas identificadas de forma exclusiva em folhas +1 de *S. officinarum* (n= 130) *S. spontaneum* (n= 190) e SP80-3280 (n= 156) segundo as categorias Função Molecular e Processo Biológico do *Gene Ontology*.

<i>S. officinarum</i> (n= 130)			
Função Molecular	<i>p-value</i>	Processo Biológico	<i>p-value</i>
<i>linoleate 13S-lipoxygenase activity</i>	8.4E-06	<i>biological_process</i>	2.0E-05
-	-	<i>response to other organism</i>	4.9E-05
<i>S. spontaneum</i> (n= 190)			
Função Molecular	<i>p-value</i>	Processo Biológico	<i>p-value</i>
<i>molecular_function</i>	1.1E-07	<i>biological_process</i>	1.7E-07
<i>cysteine-type peptidase activity</i>	4.7E-07	<i>flavonoid biosynthetic process</i>	1.7E-06
<i>anthocyanidin reductase activity</i>	1.7E-06	-	-
SP80-3280 (n= 156)			
Função Molecular	<i>p-value</i>	Processo Biológico	<i>p-value</i>
<i>molecular_function</i>	1.0E-04	<i>biological_process</i>	2.1E-06
<i>hydrogen-translocating pyrophosphatase activity</i>	1.9E-04	<i>establishment or maintenance of transmembrane electrochemical gradient</i>	7.0E-05
-	-	-	-

Essa análise revelou que o conjunto de proteínas identificadas de forma exclusiva em folhas +1 de *S. officinarum* apresentou enriquecimento dos termos “*linoleate 13S-lipoxygenase activity*” e “resposta a outros organismos”, respectivamente nas categorias Função Molecular e Processo Biológico. Além da ocorrência de lipoxigenases exclusivas em folhas +1 nessa espécie ancestral, envolvidas em processos de resposta ao estresse biótico, foi identificada uma proteína contendo o domínio de complexo de ataque à membrana e perforina (MACPF, do inglês *Membrane attack complex component/perforin domain*) (MAC/perforin domain-containing protein; Tabela Suplementar 5). Em

Arabidopsis, o gene *CAD1* (*constitutively activated cell death 1*) codifica uma proteína contendo uma região de alta homologia com o domínio MACPF, encontrado em proteínas do sistema complemento e em perforina de mamíferos, que apresentam funções bem caracterizadas no processo de resposta imune inata (Esser, 1994; Trapani e Smyth, 2002). Dessa forma, Morita-Yamamuro *et al.* (2005) demonstraram o importante papel da proteína CAD1 em Arabidopsis, na regulação da resposta imune a partir de sua atuação como mediadora no reconhecimento de sinais relacionados à infecção por patógenos, e analogamente, em *S. officinarum*, a proteína *MAC/perforin domain-containing protein* poderia desempenhar papel semelhante.

Em relação às proteínas identificadas exclusivamente em *S. spontaneum*, as funções moleculares “*cysteine-type peptidase activity*” e “*anthocyanidin reductase activity*” foram identificadas como enriquecidas, enquanto que a “biossíntese de flavonóides” foi o termo enriquecido na categoria Processo Biológico. As cisteíno-peptidases (proteinases) participam em diversos processos fisiológicos essenciais para o crescimento e desenvolvimento da planta, mas também participam da resposta aos estresses biótico e abiótico. Em resposta ao déficit hídrico, sua atividade proteolítica aumenta significativamente em folhas de trigo (Zagdańska e Wiśniewski, 1996). Além desta resposta ao estresse abiótico, as cisteíno-peptidases atuam em mecanismos de defesa contra o ataque de insetos. Pechan *et al.* (2000) demonstraram o acúmulo dessa enzima em resposta ao ataque por larvas da ordem Lepidoptera em genótipos resistentes de milho.

Os flavonóides são metabólitos secundários que apresentam diferentes funções em plantas, incluindo a proteção contra a radiação ultravioleta, a coloração de flores como

atrativo visual para espécies polinizadoras, o transporte de hormônios e a resposta de defesa à patógenos (Dixon e Pasinetti, 2010). No que concerne às funções de defesa de patógenos, a representatividade do processo de “biossíntese de flavonóides”, bem como da atividade redutase sobre a antocianidina, cuja reação resulta na síntese do flavonóide proantocianidina (taninos condensados) (Falcone Ferreyra *et al.*, 2012), indicam que a via de síntese desse tipo de flavonóide é importante em folhas +1 de *S. spontaneum*. Sabe-se que os taninos condensados apresentam um proeminente efeito deletério sobre os insetos herbívoros, afetando seu crescimento e desenvolvimento. Devido às suas propriedades de interação com proteínas (incluindo as enzimas digestivas presentes nos herbívoros) e efeitos quelantes sobre metais, essas moléculas interferem na capacidade de digestibilidade e absorção de nutrientes pelos insetos (War *et al.*, 2012). Assim, essas proteínas identificadas de forma exclusiva, e que compõem o termo enriquecido “biossíntese de flavonóides”, podem ser específicas para a defesa contra insetos em folhas +1 de *S. spontaneum*, e estar relacionadas com o fenótipo de robustez desta planta.

No conjunto de proteínas identificadas de forma exclusiva em folhas +1 de SP80-3280, os termos “*hydrogen-translocating pyrophosphatase activity*” e “estabelecimento ou manutenção do gradiente eletroquímico transmembrana” foram identificados enriquecidos nas categorias Função Molecular e Processo Biológico do *Gene Ontology*, respectivamente. A proteína que representa esta categoria, a enzima H^+ pirofosfatase (H^+ -PPase), atua na manutenção do gradiente eletroquímico transmembrana e é responsável pela translocação de prótons aos vacúolos em células vegetais (Maeshima, 2000), apresentando papel importante no transporte de hormônios, no desenvolvimento

foliar e na resposta aos estresses hídrico e salino (Fukuda *et al.*, 2016; Gaxiola *et al.*, 2001).

Em conjunto, esses achados indicam que as proteínas identificadas como exclusivas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, de forma geral, estão envolvidas em processos de resposta ao estresse, sugerindo que, apesar da grande similaridade de seus proteomas (74% das proteínas foram identificadas nas três amostras), podem existir algumas peculiaridades na forma como essas diferentes plantas reagem às variações bióticas e abióticas.

4.4. ANÁLISE QUANTITATIVA DOS PROTEOMAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280

A análise quantitativa dos proteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 foi realizada com a ferramenta de quantificação *label-free*, MaxLFQ (Cox *et al.*, 2014), contida no pacote de ferramentas do software MaxQuant (Cox e Mann, 2008). Esta ferramenta de análise se baseia nas medidas de intensidade dos peptídeos, que, em conjunto, refletem as concentrações relativas das proteínas em diferentes amostras. Para essa análise foram utilizados os dados de LC-MS/MS gerados pela abordagem *shotgun/bottom-up proteomics* que resultou na identificação de 1.060 grupos de proteínas em *S. officinarum*, *S. spontaneum* e SP80-3280, uma vez que esta abordagem, quando comparada às demais abordagens utilizadas neste estudo, oferece maior robustez na quantificação de proteínas devido à menor manipulação das amostras.

Em análises quantitativas *label-free*, é de extrema importância a realização de réplicas experimentais, biológicas e técnicas, para minimizar os efeitos da manipulação individual de amostras e das flutuações inerentes às técnicas de cromatografia líquida e de aquisição de espectros de massas pelos espectrômetros (Bantscheff *et al.*, 2012). Dessa forma, as três réplicas biológicas das proteínas extraídas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 foram individualmente submetidas à digestão tripsínica em solução e os peptídeos resultantes de cada réplica biológica foram analisados em duplicata, totalizando 18 análises por LC-MS/MS (6 análises de cada amostra de folhas +1). Os dados quantitativos calculados pela ferramenta MaxLFQ foram categorizados em três grupos utilizando o software Perseus, sendo cada grupo constituído pelos valores de quantificação das 6 análises de LC-MS/MS referentes à *S. officinarum*, *S. spontaneum* e SP80-3280. Os valores de intensidade normalizados (LFQ *intensity*) foram transformados para a escala Log_2 e submetidos a uma etapa de filtragem, onde foram selecionadas para as análises subsequentes somente aquelas proteínas que apresentaram pelo menos 3 valores válidos (valores maiores que zero) no conjunto de 6 análises, em pelo menos um dos grupos. A aplicação deste critério resultou em um conjunto de 756 proteínas (Tabela Suplementar 8). Após essa etapa, os valores normalizados de intensidade foram avaliados segundo sua capacidade de se ajustar a uma distribuição normal (Fig. 34), condição necessária para as análises estatísticas subsequentes. A análise da Figura 34, indica que, sem exceção, os dados de intensidade normalizados obedeceram à distribuição normal, permitindo dessa forma o prosseguimento da análise dos dados quantitativos.

Figura 34. Histogramas mostrando a distribuição dos dados de LFQ intensity (Log_2) resultantes da análise quantitativa *label-free* aplicada sobre os dados gerados pelas análises de LC-MS/MS dos peptídeos tripsínicos provenientes do processo de digestão em solução das réplicas biológicas (RB 1 a 3) das amostras de proteínas extraídas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280. Para cada réplica biológica foram realizadas duas replicatas técnicas (RT 1 e 2) das análises de LC-MS/MS.

A seguir os dados de quantificação foram submetidos à ferramenta de imputação de valores, denominada “*replace missing values from normal distribution*”, que substitui os valores não válidos de intensidade. Esta ferramenta utiliza a distribuição normal com a média e o desvio padrão ajustados de modo a simular intensidades para proteínas de baixa abundância (Deeb *et al.*, 2012; Cox *et al.*, 2014). Essa estratégia, que confere maior robustez aos dados quantitativos que obedeceram à distribuição normal, é utilizada na plataforma MaxQuant, considerando que valores não válidos de intensidade provavelmente decorrem de proteínas pouco abundantes nas amostras, e não da eventual ausência das mesmas. Neste estudo, a ferramenta foi aplicada sobre os dados de quantificação contidos na matriz filtrada, composta por 756 proteínas. Após o uso dessa estratégia, as distribuições dos valores imputados demonstraram que estes foram ajustados em intervalos de valores de baixa intensidade (destacadas em amarelo na Fig. 35), simulando os valores de intensidade de proteínas pouco abundantes.

Figura 35. Histogramas mostrando a distribuição dos dados de LFQ *intensity* (Log_2) após a utilização da ferramenta “replace missing values from normal distribution”. As distribuições destacadas em amarelo representam os valores de intensidade substituídos pela ferramenta e em cinza as distribuições dos valores de intensidade medidos.

Dessa forma, a partir dos dados de quantificação das análises de LC-MS/MS das réplicas biológicas de *S. officinarum*, *S. spontaneum* e SP80-3280, foram calculados os valores médios de LFQ intensity das réplicas técnicas de cada grupo e esse conjunto de dados de quantificação das 756 proteínas foi submetido à análise de variância *one-way* ANOVA ($p\text{-value} < 0,05$, ajustado de acordo com o método de Benjamini-Hochberg), para determinar variações de expressão significativas entre os grupos (Tabela Suplementar 9). A partir dessa análise, observou-se que do total de proteínas submetidas à quantificação, 120 foram identificadas como diferencialmente expressas entre os conjuntos de proteínas extraídas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 (Tabela Suplementar 10). Os valores de quantificação das proteínas diferencialmente expressas foram submetidos a uma etapa de normalização dos dados pela ferramenta *Z-score* contida no software Perseus e então à análise de *cluster* hierárquico, mostrada na Figura 36.

Figura 36. Heat Map das proteínas diferencialmente expressas ($n=120$) nos proteomas de folhas +1 de *S. officinarum* (*S. off*), *S. spontaneum* (*S. sp*) e SP80-3280 (*SP*). Os valores de LFQ intensity das proteínas cujas abundâncias relativas variaram significativamente entre os grupos (one-way ANOVA, p -value ajustado $< 0,05$) foram normalizados (Z-score) e submetidos à análise de cluster hierárquico. Regiões do mapa em vermelho representam as proteínas com valores de expressão relativa maiores, enquanto regiões em azul, valores de expressão relativa menores. As réplicas biológicas de cada grupo são representadas no mapa por RB1 a RB3 e os grupos (6 clusters) formados de acordo com o perfil relativo de expressão são indicados no lado direito do mapa.

Esta análise mostrou concordância entre as intensidades normalizadas das réplicas biológicas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280. De forma geral, os seis clusters formados contém populações bem definidas de proteínas cujas expressões variaram significativamente. Ainda, observa-se que as espécies *S. officinarum* e SP80-3280 formaram um grupo separado de *S. spontaneum*, demonstrando uma maior correlação nos perfis de expressão das proteínas em suas folhas +1.

A Figura 37 mostra os seis *clusters* individuais formados e a distribuição dos perfis de expressão relativa das proteínas contidas em cada um deles. O *cluster* 1 (n= 10) é composto por proteínas cuja expressão é semelhante entre os proteomas de folhas +1 de *S. spontaneum* e SP80-3280 e significativamente mais alta quando comparada com os níveis de expressão em folhas +1 de *S. officinarum*. No *cluster* 2 (n= 29) encontram-se proteínas que se apresentam relativamente mais expressas somente em *S. spontaneum*, enquanto que nos *clusters* 3 (n= 22) e 4 (n= 11), observam-se conjuntos de proteínas de SP80-3280 que apresentam níveis de expressão relativa baixa e alta, respectivamente, quando comparados aos níveis de expressão nas espécies ancestrais. Níveis de expressão mais altos foram observados em *S. officinarum* e SP80-3280, em um conjunto de 24 proteínas que compõe o *cluster* 5 e, por fim, no *cluster* 6 (n= 24), observa-se a ocorrência de proteínas que apresentam níveis de expressão relativa mais alta em *S. officinarum*, quando comparados aos níveis de expressão observados em *S. spontaneum* e SP80-3280.

Figura 37. Perfil de expressão do conjunto de proteínas diferencialmente expressas nos proteomas de folhas +1 de *S. officinarum* (S. off), *S. spontaneum* (S. sp) e SP80-3280 (SP). Os *clusters* foram formados de acordo com os padrões de perfil de expressão relativa entre as proteínas. As distribuições dos valores de expressão relativa são mostradas ao lado direito dos mapas. Linhas representadas na coloração mais próxima do vermelho correspondem às distribuições mais próximas da distribuição central do *cluster*, enquanto que linhas verdes correspondem às distribuições mais distantes.

Considerando o conjunto de proteínas diferencialmente expressas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 (Fig. 36 e 37; Tabela Suplementar 10), foram selecionadas somente aquelas cujas abundâncias relativas variaram pelo menos quatro vezes na comparação entre as espécies ancestrais (Tab. 13) ou nas comparações entre as espécies ancestrais com a cultivar híbrida (Tab. 14 e 15).

Tabela 13. Proteínas diferencialmente expressas em folhas +1 de *S. officinarum* e *S. spontaneum*. Somente foram listadas proteínas cuja abundância relativa variou pelo menos quatro vezes.

Proteína	Cluster*	Fold-change**	Processo Biológico
Reguladas positivamente em <i>S. officinarum</i>			
<i>histone H1-3</i>	5	60.67	<i>nucleosome assembly</i>
<i>GDSL-like Lipase/Acylhydrolase superfamily protein</i>	6	55.94	<i>cell wall organization</i>
<i>polyamine oxidase 1</i>	6	37.04	<i>polyamine catabolic process</i>
<i>lipoxygenase 1</i>	6	19.25	<i>response to other organism</i>
<i>peptidomethionine sulfoxide reductase</i>	5	16.11	<i>response to oxidative stress</i>
<i>Ribonuclease T2 family protein</i>	6	15.75	-
<i>Protein phosphatase 2C family protein</i>	6	15.32	-
<i>HAD superfamily, subfamily IIIB acid phosphatase</i>	6	13.63	-
<i>ribosomal protein 5B</i>	5	10.62	<i>translation</i>
<i>Subtilase family protein</i>	5	8.32	-
<i>basic chitinase</i>	6	8.24	<i>defense response to fungus</i>
<i>osmotin 34</i>	5	6.45	<i>defense response to bacterium, incompatible interaction</i>
<i>Peroxidase superfamily protein</i>	5	5.67	-
<i>apyrase 1</i>	6	5.52	<i>D-xylose metabolic process</i>
<i>S-adenosylmethionine synthetase family protein</i>	5	4.76	<i>Golgi organization</i>
<i>basic chitinase</i>	5	4.27	<i>defense response to fungus</i>
Reguladas negativamente em <i>S. officinarum</i>			
<i>ADPGLC-PPase large subunit</i>	2	0.25	<i>starch biosynthetic process</i>
<i>NAD(P)-binding Rossmann-fold superfamily protein</i>	2	0.23	-
<i>Alanyl-tRNA synthetase</i>	2	0.22	<i>alanyl-tRNA aminoacylation</i>
<i>aspartate aminotransferase 3</i>	2	0.20	<i>nitrogen compound metabolic process</i>
<i>serine carboxypeptidase-like 27</i>	2	0.15	<i>proteolysis</i>
<i>beta-amylase 5</i>	2	0.12	<i>starch catabolic process</i>
<i>high mobility group B3</i>	2	0.08	<i>Golgi organization</i>
<i>lipoxygenase 2</i>	2	0.03	<i>response to other organism</i>

* Indicado nas Figuras 36 e 37.

** LFQ Intensity S.off / LFQ Intensity S.sp

No conjunto de proteínas que apresentaram variações de abundância relativa de pelo menos quatro vezes em folhas +1 de *S. officinarum* e *S. spontaneum* (Tab. 13), verificou-se, entre as proteínas que apresentaram maior expressão relativa em *S. officinarum*, a alta contribuição de proteínas relacionadas à resposta ao estresse biótico. A ocorrência de proteínas como a poliamina oxidase, a lipoxigenase, a peptídeo metionina sulfóxido redutase, a quitinase básica e a osmotina, que apresentam expressão relativa consideravelmente mais alta em *S. officinarum*, indica que o processo biológico de resposta a outros organismos é um componente importante para o fenótipo dessa espécie ancestral. Além do envolvimento das lipoxigenases em processos de defesa ao estresse (Brash, 1999; Porta e Rocha-Sosa, 2002), as poliamina oxidases têm papel importante na degradação enzimática de poliaminas, contribuindo diretamente na produção de ROS e no desencadeamento do processo de morte celular programada em plantas (Yoda *et al.*, 2006). Na ocorrência de altas taxas de produção de ROS, proteínas contendo resíduos de metionina são particularmente suscetíveis à reação de oxidação, resultando na geração de metionina sulfóxido (MetSO) que leva à alterações de estrutura e de atividade dessas proteínas (Gao *et al.*, 1998). Em situações de estresse oxidativo, enzimas como a peptídeo metionina sulfóxido redutase (PMSR) atuam na reação de redução de MetSO e regeneração da metionina, evidenciando a participação das PMSRs nos processos de controle de danos oxidativos (Vieira Dos Santos *et al.*, 2005). A osmotina, proteína multifuncional isolada inicialmente em células de *Nicotiana tabacum* (Singh *et al.*, 1985), pertencente à família 5 das proteínas relacionadas à patogênese (PRPs) (van Loon, 1985; Sels *et al.*, 2008), apresenta atividade antifúngica sobre diversas espécies, atuando na regulação da resposta ao estresse a partir da modulação

da atividade de fatores de transcrição ou na modulação da sinalização celular (Abdin *et al.*, 2011). A enzima quitinase, outra proteína pertencente às famílias das PRPs, atua na defesa contra invasões fúngicas, a partir de mecanismos de hidrólise de quitina, um polímero de (β -1,4) *N*-acetilglicosamina encontrado na parede celular de certas classes de fungos (Rakwal *et al.*, 2004). A co-expressão da quitinase e de outras proteínas envolvidas na defesa, como a osmotina, as peroxidases, as lipoxigenases e poliamina oxidases, pode configurar uma forma de potencialização da atividade antifúngica (Neuhaus, 1999). É interessante destacar que, apesar da espécie ancestral *S. officinarum* ser caracterizada pela suscetibilidade às doenças causadas principalmente por bactérias e vírus, e ao ataque de insetos, ela é resistente à infecção pelo fungo *Sporisorium scitamineum*, que é o agente causador da doença conhecida como carvão. *S. spontaneum*, por sua vez, é uma espécie suscetível a essa doença (Segalla, 1964; Cheavegatti-Gianotto *et al.*, 2011). Dessa forma, a ocorrência dessas proteínas de defesa, que atuam principalmente na resposta às infecções fúngicas, cujas expressões relativas foram mais altas em folhas +1 de *S. officinarum*, sugere que esta espécie ancestral pode estar mais bem adaptada para resistir à infecção por fungos.

Entre as proteínas que exibiram níveis de expressão relativa mais alta em folhas +1 de *S. spontaneum* (Tab. 13), encontram-se a subunidade maior da ADP-glicose pirofosforilase (*ADPGLC-PPase large subunit*) e a beta-amilase, que participam do metabolismo de amido. De um lado, a ADPGLC-PPase participa dos passos iniciais da síntese de amido e catalisa a reação de síntese de moléculas de ADP-glicose (ADPglic, substrato para a síntese de amido) e pirofosfato inorgânico (PPi) (Espada, 1962) e de outro, a beta-amilase participa do processo de degradação de amido em folhas (Smith

et al., 2005), apresentando nível de expressão mais alto em folhas +1 de *S. spontaneum*. Apesar de estes achados parecerem antagônicos, o fato de estas proteínas estarem aumentadas em *S. spontaneum*, sugere que o processo de geração de amido pode ser mais ativado nas folhas +1 desta planta, além de estar relacionado com sua taxa de fotossíntese, já que o aumento de açúcares nas folhas de cana-de-açúcar pode influenciar o processo da fotossíntese (McCormick *et al.*, 2008).

Tabela 14. Proteínas diferencialmente expressas em folhas +1 de *S. officinarum* e SP80-3280. Somente foram listadas proteínas cuja abundância relativa variou pelo menos quatro vezes.

Proteína	Cluster*	Fold-change**	Processo Biológico
Reguladas positivamente em <i>S. officinarum</i>			
<i>GDSL-like Lipase/Acylhydrolase superfamily protein</i>	6	56.21	<i>cell wall organization</i>
<i>polyamine oxidase 1</i>	6	38.03	<i>polyamine catabolic process</i>
<i>NAD(P)-linked oxidoreductase superfamily protein</i>	3	27.75	<i>response to cadmium ion</i>
<i>Ribonuclease T2 family protein</i>	6	17.58	-
<i>lipoxygenase 1</i>	6	13.55	<i>response to other organism</i>
<i>Protein phosphatase 2C family protein</i>	6	9.14	-
<i>basic chitinase</i>	6	7.63	<i>defense response to fungus</i>
<i>HAD superfamily, subfamily IIIB acid phosphatase</i>	6	5.19	-
<i>PLAT/LH2 domain-containing lipoxygenase family protein</i>	6	4.99	<i>response to other organism</i>

* Indicado nas Figuras 36 e 37.

** LFQ Intensity S.off / LFQ Intensity SP

Tabela 15. Proteínas diferencialmente expressas em folhas +1 de *S. spontaneum* e SP80-3280. Somente foram listadas proteínas cuja abundância relativa variou pelo menos quatro vezes.

Proteína	Cluster*	Fold-change**	Processo Biológico
Reguladas positivamente em <i>S. spontaneum</i>			
<i>lipoxygenase 2</i>	2	35.24	<i>response to other organism</i>
<i>NAD(P)-linked oxidoreductase superfamily protein</i>	3	16.09	<i>response to cadmium ion</i>
<i>high mobility group B3</i>	2	9.32	<i>Golgi organization</i>
<i>beta-amylase 5</i>	2	8.77	<i>starch catabolic process</i>
<i>serine carboxypeptidase-like 27</i>	2	8.54	<i>proteolysis</i>
<i>ADPGlc-PPase large subunit</i>	2	4.78	<i>starch biosynthetic process</i>
Reguladas negativamente em <i>S. spontaneum</i>			
<i>basic chitinase</i>	5	0.25	<i>defense response to fungus</i>
<i>Subtilase family protein</i>	5	0.24	-
<i>osmotin 34</i>	5	0.22	<i>defense response to bacterium, incompatible interaction</i>
<i>S-adenosylmethionine synthetase family protein</i>	5	0.18	<i>Golgi organization</i>
<i>glutathione S-transferase THETA 3</i>	4	0.17	<i>response to oxidative stress</i>
<i>ribosomal protein 5B</i>	5	0.13	<i>translation</i>
<i>peptidomethionine sulfoxide reductase 2</i>	5	0.11	<i>response to oxidative stress</i>
<i>histone H1-3</i>	5	0.03	<i>nucleosome assembly</i>

* Indicado nas Figuras 36 e 37.

** LFQ Intensity S.sp / LFQ Intensity SP

Proteínas como a poliamina oxidase, a quitinase e as lipoxigenases, que já foram observadas como diferencialmente expressas na comparação das espécies ancestrais (Tab. 13), estas apresentaram níveis de expressão relativa mais altos em *S. officinarum* em relação à SP80-3280, e como já discutido anteriormente, estão envolvidas em processos de resposta à infecção por organismos de origem fúngica. Por outro lado, na comparação entre os componentes de folhas +1 de *S. spontaneum* e SP80-3280, algumas proteínas relacionadas aos processos de defesa contra fungos e bactérias, e

resposta ao estresse oxidativo exibiram abundância mais baixa na espécie ancestral (Tab. 15).

Na comparação entre as proteínas diferencialmente abundantes nas três espécies de folhas, uma lipase (*GDSL-like Lipase/Acylhydrolase superfamily protein*), destacou-se por ter sido quantificada como cerca de 55 vezes mais abundante em *S. officinarum* (Tab. 13 e Tab. 14). Sobre esta proteína, foi reportado que ela desempenha um papel relevante na imunidade de plantas, e em *Arabidopsis*, regula a resistência sistêmica associada à sinalização por etileno (Kwon *et al.*, 2009). Sendo considerada como um componente do processo de organização da parede celular e apresentando atividade hidrolítica sobre lipídeos, esta proteína pode ter um espectro mais amplo de atividades e até desempenhar atividade proteolítica (Molgaard *et al.*, 2000). No entanto, neste estudo, o fato de ter sua expressão significativamente aumentada em folhas de *S. officinarum* não nos permite fazer nenhuma inferência sobre seu papel específico nas folhas +1 desta espécie ancestral. A histona H1.3, outra proteína que se destacou por apresentar níveis de expressão cerca de 61 vezes e 33 vezes maiores em *S. officinarum* e SP80-3280 em relação ao nível de expressão observado em *S. spontaneum*, respectivamente, participa no processo de compactação da cromatina em eucariotos. Em situações de estresse hídrico e em resposta ao ácido abscísico, essa variante de histona H1 apresenta expressão aumentada em plantas, sendo então associada aos eventos de resposta adaptativa ao estresse. Apesar de ainda não bem compreendidos, evidências apontam que os mecanismos pelos quais a histona H1.3 exerce seu papel na resposta adaptativa estão relacionados com sua atividade como fator facilitador da

acessibilidade de metil-transferases às regiões do DNA relacionadas ao estresse (Ascenzi e Gantt, 1997; Rutowicz *et al.*, 2015).

De forma geral, considerando que esta análise quantitativa comparativa entre as duas espécies ancestrais e a cultivar híbrida limita-se aos proteomas de suas folhas +1, em conjunto, estes achados permitem apenas algumas observações sobre como variam as abundâncias de algumas classes de proteínas e os processos biológicos aos quais estas estão relacionadas, não sendo possível correlacionar estes resultados com suas características de capacidade de produção de sacarose ou robustez no campo, ou mesmo com fenótipo das plantas como um todo.

4.5. ANÁLISE DOS FOSFOPROTEOMAS DE FOLHAS +1 DE *S. OFFICINARUM*, *S. SPONTANEUM* E SP80-3280

A fosforilação de proteínas exercida pela atividade de proteínas quinases é a modificação pós-traducional mais comum em eucariotos, e regula diversos processos fisiológicos, atuando fundamentalmente na modulação de vias de sinalização celular de forma a alterar a estrutura, a atividade, a estabilidade, a localização e a capacidade de interação de proteínas com outras moléculas (Hunter, 2000; Nakagami *et al.*, 2010; Zhang *et al.*, 2013). Os aminoácidos serina, treonina e tirosina são os alvos mais comuns de fosforilação. Embora seja uma modificação pós-traducional frequente, a abundância de proteínas fosforiladas é geralmente baixa nas células, fato que representa o principal gargalo para a análise voltada ao estudo de fosfoproteomas (Imamura *et al.*, 2012). Diante desse fato, diferentes metodologias de seleção de fosfopeptídeos têm sido

aplicadas para contornar os problemas de detecção de peptídeos fosforilados pela técnica de espectrometria de massas.

No estudo dos fosfoproteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, foi utilizado o método de enriquecimento de fosfopeptídeos IMAC (*Immobilized Metal Affinity Chromatography*) (Stensballe *et al.*, 2001; Kokubu *et al.*, 2005), que se baseia na afinidade do grupo fosfato ao íon metálico Fe^{3+} . A análise foi realizada em três réplicas biológicas das proteínas extraídas de folhas +1, que foram individualmente submetidas à digestão tripsínica em solução e os peptídeos resultantes de cada réplica biológica foram submetidos ao processo de enriquecimento de fosfopeptídeos IMAC. Para a identificação, a mistura de peptídeos resultante foi injetada diretamente em uma pré-coluna C-18, que foi acoplada ao sistema de cromatografia líquida conectado ao espectrômetro de massas, e analisada por LC-MS/MS. Os espectros de MS e MS/MS gerados resultaram em um total de 1.419 peptídeos identificados nas três amostras de folhas +1, dentre os quais 1.202 foram encontrados fosforilados (dados não mostrados), o que corresponde a aproximadamente 85% de eficiência no processo de enriquecimento. Com o objetivo de apresentar um panorama geral de identificação dos fosfopeptídeos nas amostras de folhas +1, as populações de fosfopeptídeos identificados a partir da utilização do método IMAC foram agrupadas e então filtradas de acordo com o sistema de classificação desenvolvido por Olsen *et al.* (2006). De acordo com este sistema, considera-se como identificados aqueles peptídeos cuja probabilidade mínima de localização dos sítios de fosforilação é maior ou igual a 75% (sítios de fosforilação classe I). A Tabela 16 mostra um resumo dos dados obtidos no estudo do fosfoproteoma de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280,

e as listas de identificação de fosfopeptídeos e fosfoproteínas podem ser consultadas nas Tabelas Suplementares 11 e 12.

Tabela 16. Sumário dos dados obtidos das análises dos fosfoproteomas dos conjuntos de proteínas extraídas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280.

Descrição	Total*	<i>S. officinarum</i>	<i>S. spontaneum</i>	SP80-3280
n. de fosfopeptídeos identificados**	751	695	665	546
n. de fosfoproteínas	515	491	416	470
n. de peptídeos fosforilados em serina	624	582	552	453
n. de peptídeos fosforilados em treonina	120	107	106	88
n. de peptídeos fosforilados em tirosina	7	6	7	5
n. de fosfopeptídeos únicos	633	596	570	478
Proporção Ser:Thr:Tyr (%)	624:120:7 (83,1%:16%:0,9%)	582:107:6 (83,7%:15,4%:0,9%)	552:106:7 (83%:16%:1%)	453:88:5 (83%:16,1%:0,9%)

* As quantidades totais não correspondem à somatória dos números individuais de cada grupo devido às eventuais ocorrências repetidas em diferentes grupos;

** O critério de identificação dos sítios de fosforilação baseou-se na probabilidade mínima de localização de 75% (Olsen *et al.*, 2006) e delta score > 6.

Esta análise resultou na identificação de 751 sítios de fosforilação, distribuídos em 515 fosfoproteínas. Considerando o número total de fosfopeptídeos identificados, 624 (83,1%) foram encontrados fosforilados em resíduos de serina, 120 (16%) em resíduos de treonina e 7 (0,9%) em resíduos de tirosina. Estes achados estão de acordo com as proporções encontradas em uma ampla análise fosfoproteômica de cultura de células de arroz e Arabidopsis, conduzida por Nakagami *et al.* (2010), onde 84,8% e 82,7% dos peptídeos foram identificados fosforilados em resíduos de serina, respectivamente em arroz e Arabidopsis, e 12,3% e 13,1% em resíduos de treonina, respectivamente em arroz e Arabidopsis. Por outro lado, neste estudo, observou-se que a quantidade de peptídeos fosforilados em tirosina (7, correspondendo a 0,9%) foi significativamente menor em relação ao que foi reportado sobre células de arroz e Arabidopsis (2,9% e

4,2%, respectivamente). Diferenças nas proporções de peptídeos fosforilados em tirosina foram associadas aos diferentes tipos de amostras em que as análises fosfoproteômicas foram realizadas. Reiland *et al.* (2009), utilizando tecidos meristemáticos de *Arabidopsis*, verificaram que apenas 0,3% dos fosfopeptídeos identificados eram fosforilados em tirosina e relacionou a baixa proporção encontrada ao fato de que a grau de fosforilação em tirosina em culturas celulares não reflete a situação encontrada em células diferenciadas. De fato, neste estudo, onde se utilizou o tecido foliar de duas espécies ancestrais e uma cultivar híbrida de cana-de-açúcar, verificou-se uma fração significativamente menor de peptídeos fosforilados em tirosina (0,9%), e este resultado é aqui interpretado com cautela, uma vez que o número total de fosfopeptídeos identificados foi significativamente menor em comparação aos estudos realizados por Sugiyama *et al.* (2008) e Nakagami *et al.* (2010), em que foram identificados, respectivamente, 2.597 e 5.143 fosfopeptídeos em *Arabidopsis*, ainda que estas análises tenham sido realizadas com material extraído de suspensões de células. Além disso, deve-se levar em consideração que esses trabalhos se referem às análises de fosfoproteomas de plantas C3, que apresentam diferenças fisiológicas em relação às plantas C4, fato que poderia refletir em padrões de fosforilação diferentes. Dessa forma, a proporção entre peptídeos fosforilados em resíduos de serina, treonina e tirosina encontrada neste estudo foi comparada com a proporção observada em um estudo direcionado ao proteoma e ao fosfoproteoma de folhas de milho (Facette *et al.*, 2013), uma planta C4, que identificou 85,7%, 13,5% e 0,8% de peptídeos fosforilados em serina, treonina e tirosina, respectivamente, sugerindo que, entre plantas C4, a proporção entre peptídeos fosforilados em serina, treonina e tirosina é semelhante.

Uma razão para a relativamente baixa quantidade de fosfopeptídeos identificados no presente estudo, pode ter sido o método de extração de proteínas aplicado. Apesar do uso de inibidores de enzimas proteolíticas e de fosfatases logo após a extração e solubilização das proteínas obtidas pelo método Fenol/SDS, a comparação entre o número de fosfopeptídeos identificados em folhas +1 de cana-de-açúcar e em folhas de outras plantas, indica que houve baixa eficiência na extração desta classe de peptídeos. Em um estudo recente sobre métodos de extração de proteínas para análise proteômica de folhas, Luís *et al.* (2016) testaram cinco abordagens, e demonstraram que um método baseado em um tampão de lise contendo ureia e tioureia, além do detergente CHAPS, se mostrou mais adequado para a análise proteômica e fosfoproteômica de folhas de milho, uma planta C4, já que, segundo os autores, beneficiou a extração de isoformas ácidas de proteínas.

As distribuições dos fosfopeptídeos e fosfoproteínas identificados entre as amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 são mostradas na Figura 38.

Figura 38. Diagramas de Venn dos peptídeos fosforilados e fosfoproteínas totais identificados em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280. (A) Número total de sítios fosforilados, exclusivos e compartilhados. (B) Proteínas identificadas como fosforiladas, exclusivas e compartilhadas. Os diagramas de Venn foram construídos utilizando o software *eulerAPE* (Micallef e Rodgers, 2014), disponível no sítio <http://www.eulerdiagrams.org/eulerAPE/>.

Verificou-se que no conjunto de fosfopeptídeos identificados (n= 751), 473 foram identificados em todas as amostras (Fig. 38A), correspondendo a 381 proteínas fosforiladas (Fig. 38B). Uma parcela minoritária de fosfopeptídeos (9,2%) foi identificada de forma exclusiva entre as amostras, e estes foram responsáveis pela identificação de 17, 11 e 6 fosfoproteínas exclusivas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, respectivamente.

O conjunto total de fosfoproteínas identificadas (n= 515) foi submetido à busca nas categorias Componente Celular, Processo Biológico e Função Molecular do *Gene Ontology* e submetido à análise de enriquecimento de termos (Fig. 39).

Figura 39. Termos enriquecidos no conjunto de fosfoproteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 segundo as categorias funcionais Componente Celular (A), Processo Biológico (B) e Função Molecular (C) do *Gene Ontology*. A análise de enriquecimento de termos, aplicada sobre o conjunto total de fosfoproteínas identificadas ($n = 515$), foi realizada pelas ferramentas contidas no software GeneMerge v1.4 (Castillo-Davis e Hartl, 2003).

De acordo com a análise mostrada na Figura 39, verifica-se que o conjunto de fosfoproteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 apresenta enriquecimento principalmente no núcleo, citosol, e citoplasma, (Fig. 39A) e destaca-se pela participação em processos de regulação da transcrição, de resposta ao estresse, e no metabolismo energético (Fig. 39B). A observação de que as categorias “*regulation of transcription, DNA-dependent*”, “*DNA binding*”, “*nucleic acid binding*” e “*RNA binding*” estão enriquecidas (Fig. 39, B e C) indica a participação importante de fosfoproteínas nos eventos transcripcionais e traducionais. A categoria “*protein binding*” também se destaca pelo alto enriquecimento relativo (Fig. 39 C) evidenciando a

importância dos eventos de fosforilação de proteínas na modulação das interações proteína-proteína, como já documentado por Nishi *et al.* (2011), que avaliaram o efeito do evento de fosforilação de proteínas sobre a energia de ligação de complexos proteicos, e demonstraram uma evidente correlação entre a fosforilação em resíduos de serina, treonina e tirosina presentes no sítio de ligação e a alteração da energia de ligação em complexos proteicos. Além do efeito energético decorrente da fosforilação de resíduos localizados em regiões do sítio de ligação, a fosforilação em regiões não adjacentes ao sítio de ligação pode levar a alterações conformacionais na estrutura da proteína expondo seu sítio de ligação e, conseqüentemente, modificando sua interação com outras proteínas. Essa alteração na interação proteína-proteína, decorrente dos efeitos alostéricos induzidos pela fosforilação, já foi observada na ativação da enzima glicogênio fosforilase, que participa do metabolismo de carboidratos e é responsável pela degradação do glicogênio (Lin *et al.*, 1997; Johnson, 2009).

Entre as proteínas identificadas como fosforiladas em folhas + 1 de *S. officinarum*, *S. spontaneum* e SP80-3280 encontram-se 39 proteínas quinases (Tabela Suplementar 12), sendo que apenas 4 delas foram identificadas como exclusivas em uma das plantas. Esse achado sugere um certo grau de conservação do conjunto de quinases e que se encontravam fosforiladas no momento da extração de proteínas das amostras para o estudo. Ainda, o conjunto de proteínas quinases fosforiladas identificadas (n= 40) corresponde a cerca de 30% do total de proteínas quinases (Tabela 10; *Protein kinase superfamily*) identificadas pela análise do proteoma total das folhas +1.

4.6. ANÁLISES QUALITATIVA E QUANTITATIVA DOS PROTEOMAS DE FOLHAS +1 DE SP80-3280 DURANTE QUATRO ESTÁGIOS DE DESENVOLVIMENTO DA PLANTA

4.6.1. ANÁLISE QUALITATIVA

Outro enfoque deste estudo sobre o proteoma de folhas +1 de cana-de-açúcar, foi a análise da cultivar híbrida SP80-3280 em diferentes estágios de desenvolvimento. Dessa forma, a análise qualitativa dos proteomas de folhas +1 coletadas de plantas que apresentavam 4, 8, 11 e 13 meses de idade foi conduzida utilizando a abordagem proteômica *Shotgun/bottom-up*. Inicialmente, após a realização do processo de extração de proteínas das réplicas biológicas de folhas +1 de SP80-3280 de acordo com o método Fenol/SDS, as proteínas extraídas foram submetidas à separação em gel de SDS-poliacrilamida para verificar os perfis de distribuição das proteínas em diferentes estágios de desenvolvimento da planta (Fig. 40).

Figura 40. Perfil eletroforético das proteínas de folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta (4, 8, 11 e 13 meses). As proteínas foram extraídas de acordo com o método Fenol/SDS. As amostras de proteínas (25 µg) foram analisadas por eletroforese em gel de SDS-poliacrilamida 12% segundo as réplicas biológicas (RB). O gel foi submetido à coloração por Coomassie G-250. Os valores à esquerda indicam a mobilidade de marcadores de massa molecular.

A análise visual dos perfis eletroforéticos apresentados na Figura 40 permite concluir que não existem diferenças significativas na quantidade, na intensidade e na distribuição das bandas de proteínas. Dessa forma, optou-se pela utilização da abordagem de análise proteômica *Shotgun/bottom-up* para a investigação dos proteomas de folhas +1 de SP80-3280 em diferentes estágios de desenvolvimento da planta.

De acordo com essa análise, utilizando os mesmos critérios de seleção de peptídeos adotados para as análises dos proteomas de folhas +1 de *S. officinarum*, *S.*

spontaneum e SP80-3280, foram identificados 5.780 peptídeos que resultaram na identificação de 1.615 grupos de proteínas (Tab. 17; Tabela Suplementar 13).

Tabela 17. Número de peptídeos e grupos de proteínas identificados em amostras de folhas +1 (F+1) de SP80-3280 (SP) coletadas em quatro diferentes estágios de desenvolvimento da planta (4, 8, 11 e 13 meses).

Amostra	n. peptídeos	n. grupos únicos de proteínas*
F+1 SP 4 m	5.693	1.550
F+1 SP 8 m	5.700	1.558
F+1 SP 11 m	5.698	1.558
F+1 SP 13 m	5.664	1.531
Total	5.780	1.615

* A presença de uma proteína em determinada amostra foi considerada de acordo com a identificação mínima de 1 peptídeo único com valor de PEP $\leq 0,01$ em pelo menos duas réplicas biológicas e ocorrência ininterrupta mínima de pelo menos 3 íons da série -y ou -b, no caso da proteína ser identificada por somente 1 peptídeo.

A Figura 41 mostra as distribuições das proteínas identificadas nas amostras de folhas +1 de SP80-3280 segundo os diferentes estágios de desenvolvimento da planta. Verificou-se aproximadamente 90% (1.451/1.615) das proteínas foram identificadas concomitantemente nas quatro amostras (4, 8, 11 e 13 meses). O restante (164 proteínas) corresponde às proteínas que foram identificadas de forma compartilhada entre duas ou três amostras ou de forma exclusiva entre os diferentes estágios de desenvolvimento da planta. Destas, 35 proteínas foram identificadas de forma compartilhada entre as amostras de folha coletada de plantas de 4, 8 e 11 meses (coletas 1, 2 e 3), ao passo que 32 foram identificadas nas amostras de plantas de 8, 11 e 13

meses (coletas 2, 3 e 4). Poucas proteínas foram identificadas de forma exclusiva e, juntas, estas correspondem a apenas 1,7% do número total de proteínas identificadas.

Figura 41. Diagrama de Venn mostrando a distribuição do número total de proteínas identificadas em amostras de folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta (4, 8, 11 e 13 meses). As identificações de peptídeos e proteínas foram realizadas pelo software MaxQuant a partir da utilização do banco de dados transcriptômico de cana-de-açúcar (Nishiyama-Jr *et al.*, 2014). O diagrama de Venn foi construído de acordo com as ferramentas contidas no sítio <http://bioinfogp.cnb.csic.es/tools/venny/index.html> (Oliveros, 2007-2015).

O conjunto de proteínas identificadas em folhas +1 de SP80-3280 coletadas nos diferentes estágios de desenvolvimento da planta foi submetido à análise de enriquecimento de termos segundo as diferentes categorias do *Gene Ontology*, utilizando as ferramentas do software GeneMerge v.1.4 (Castillo-Davis e Hartl, 2003). Os resultados dessa análise, mostrando a lista completa dos termos enriquecidos nesse conjunto de proteínas identificadas são apresentados na Tabela Suplementar 14 e, na Figura 42 estão destacados os 20 termos mais representativos de cada categoria do *Gene Ontology* no conjunto considerado de proteínas.

Figura 42. Termos enriquecidos no conjunto de proteínas identificadas em folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta segundo as categorias funcionais Componente Celular (A), Processo Biológico (B) e Função Molecular (C) do *Gene Ontology*. A análise de enriquecimento de termos, aplicada sobre o conjunto total de proteínas identificadas em folhas +1 ($n = 1.615$), foi realizada pelas ferramentas contidas no software GeneMerge v1.4 (Castillo-Davis e Hartl, 2003). São mostrados somente os 20 termos mais representativos em cada categoria.

De acordo com esta análise, observa-se que os termos enriquecidos na análise do conjunto de proteínas identificadas em folhas +1 de SP80-3280 são constituídos basicamente pelos mesmos termos enriquecidos no conjunto de proteínas identificadas nas amostras de folhas +1 das espécies ancestrais e na cultivar híbrida (Tabela Suplementar 6 e Figura 33). Mais uma vez observou-se a alta representatividade de organelas ou regiões relacionadas aos processos de fotossíntese e ao metabolismo de carbono (Fig. 42A). Além dos termos enriquecidos associados à fotossíntese, destacam-se os termos relacionados à resposta ao estresse, ao metabolismo de carboidratos e à tradução de proteínas (Fig. 42B). As funções moleculares enriquecidas reiteram a alta representatividade de proteínas envolvidas no processo de tradução, além de mostrar o enriquecimento de termos relacionados às interações não covalentes com outras proteínas ou moléculas (DNA ou RNA, por exemplo) e relacionados à atividade enzimática (Fig. 42C).

Com o objetivo de verificar possíveis diferenças nos proteomas de folhas +1 de SP80-3280 nos diferentes estágios de desenvolvimento da planta, os conjuntos de proteínas identificadas de forma exclusiva nas amostras coletadas de plantas de 4, 8, 11 e 13 meses (Tab. 18) foram avaliadas segundo os processos biológicos de que participam.

Tabela 18. Lista de proteínas identificadas de forma exclusiva em folhas +1 de SP80-3280 em diferentes estágios de desenvolvimento da planta (4, 8, 11 e 13 meses).

Exclusivas 4 meses (n= 12)	
Proteína	Processo Biológico
<i>Aminoacyl-tRNA synthetase</i>	<i>tRNA aminoacylation for protein translation</i>
<i>unknown</i>	-
<i>Translation initiation factor IF2</i>	<i>amino acid transport</i>
<i>Ribosomal protein S24e family protein</i>	<i>biological_process</i>
<i>NB-ARC domain-containing disease resistance protein</i>	-
<i>phosphoenolpyruvate carboxylase 3</i>	<i>carbon fixation</i>
<i>Transducin/WD40 repeat-like superfamily protein</i>	<i>biological_process</i>
<i>transporter associated with antigen processing protein 1</i>	<i>iron ion homeostasis</i>
<i>polyribonucleotide nucleotidyltransferase, putative</i>	<i>carotene biosynthetic process</i>
<i>Prolyl oligopeptidase family protein</i>	<i>proteolysis</i>
<i>shaggy-like protein kinase 32</i>	<i>response to stress</i>
<i>auxin transport protein (BIG)</i>	<i>unidimensional cell growth</i>
Exclusivas 8 meses (n= 5)	
Proteína	Processo Biológico
<i>DCD (Development and Cell Death) domain protein</i>	-
<i>Oxidoreductase family protein</i>	<i>metabolic process</i>
<i>Leucine-rich receptor-like protein kinase family protein</i>	<i>protein phosphorylation</i>
<i>nuclear shuttle interacting</i>	<i>pentose-phosphate shunt</i>
<i>Galactose oxidase/kelch repeat superfamily protein</i>	<i>glucosinolate catabolic process</i>
Exclusivas 11 meses (n= 5)	
Proteína	Processo Biológico
<i>RNA-binding (RRM/RBD/RNP motifs) family protein</i>	<i>biological_process</i>
<i>Class II aaRS and biotin synthetases superfamily protein</i>	<i>tRNA aminoacylation for protein translation</i>
<i>NUP50 (Nucleoporin 50 kDa) protein</i>	-
<i>Haloacid dehalogenase-like hydrolase (HAD) superfamily protein</i>	-
<i>unknown</i>	-
Exclusivas 13 meses (n= 7)	
Proteína	Processo Biológico
<i>Photosystem II reaction center PsbP family protein</i>	<i>positive regulation of transcription, DNA-dependent</i>
<i>Uncharacterised protein family (UPF0114)</i>	-
<i>stromal cell-derived factor 2-like protein precursor</i>	-
<i>Transducin/WD40 repeat-like superfamily protein</i>	<i>biological_process</i>
<i>Pyruvate kinase family protein</i>	<i>response to cadmium ion</i>
<i>Protein of unknown function (DUF1350)</i>	<i>biological_process</i>
<i>Plant protein of unknown function (DUF827)</i>	<i>response to blue light</i>

No conjunto de proteínas identificadas exclusivamente em folhas +1 de plantas de 4 meses (n= 12), destacam-se proteínas que estão envolvidas no processo de tradução de proteínas, como a aminoacil-tRNA sintetase, o fator de iniciação de tradução IF2 e a proteína ribossomal S24e. As aminoacil-tRNA sintetases são componentes essenciais para o processo de síntese de proteínas ao catalisar a reação de acoplamento dos aminoácidos ao RNA transportador correspondente (O'Donoghue e Luthey-Schulten, 2003). O fator de iniciação de tradução IF2, em combinação com outros fatores e proteínas que fazem parte do complexo ribossomal, como a proteína ribossomal S24e, apresenta papel fundamental no processo de iniciação da tradução. Considerando que a planta se encontra em estágios iniciais de desenvolvimento, sugere-se que, ao identificar tais proteínas de forma exclusiva em plantas de 4 meses, o processo de síntese de proteínas nessa fase é importante para suportar o crescimento e o desenvolvimento da planta. Ainda nesse contexto, foi identificada uma proteína transportadora de auxina, que está relacionada ao crescimento celular. Além dessas proteínas, destaca-se a fosfoenolpiruvato carboxilase 3, uma enzima chave envolvida no processo de fotossíntese em plantas C4. A classe de proteínas relacionadas ao estresse também foi observada nesse conjunto de proteínas exclusivas em folhas +1 de plantas de 4 meses. Foram identificadas, por exemplo, a proteína “*NB-ARC domain-containing disease resistance protein*”, envolvida no reconhecimento de patógenos e ativação de respostas imunes em plantas (van Ooijen *et al.*, 2008) e a “*Prolyl oligopeptidase family protein*”, que participa de processos de defesa ao estresse abiótico (Tan *et al.*, 2013).

Com relação às proteínas identificadas como exclusivas nas outras fases de desenvolvimento de folhas +1 (8, 11 e 13 meses) de SP80-3280, considerando que estas

desempenham funções diversas e participam de processos biológicos variados, não foi possível correlacionar o fato de estas terem sido identificadas como exclusivas e o estágio de desenvolvimento da planta. No entanto, essas observações devem ser consideradas com cuidado, já que a identificação de proteínas como exclusivas pode estar relacionada à abundância individual destas nas diferentes amostras. Assim, cabe aqui a seguinte ressalva: o fato de que uma proteína constitutiva tenha sido identificada exclusivamente em uma amostra, não indica que esta esteja ausente em outra, mas sugere que esta pode estar presente em menor abundância, e não foi detectada nas condições experimentais utilizadas.

4.6.2. ANÁLISE QUANTITATIVA *LABEL-FREE*

Ao conjunto de dados, composto por 24 análises de LC-MS/MS dos peptídeos (duas réplicas técnicas) derivados das três réplicas biológicas de proteínas extraídas de plantas de SP80-3280 com 4, 8, 11 e 13 meses, foi aplicada a ferramenta de análise quantitativa *label-free* MaxLFQ, utilizando a plataforma de análise de dados MaxQuant (Cox and Mann, 2008; Cox *et al.*, 2014). Os valores de intensidade normalizados (LFQ *intensity*) foram transformados para a escala Log_2 e agrupados de acordo com os respectivos períodos de coleta. Subsequentemente, os valores transformados foram submetidos a uma etapa de filtragem, como descrito no item 4.4. Os dados de distribuição dos valores de $\text{Log}_2(\text{LFQ } intensity)$ mostraram perfis de distribuição normal (Figura Suplementar 1). No total, 478 proteínas quantificadas entre os diferentes tempos de coleta foram submetidas à análise de variância *one-way* ANOVA. Considerando que

as réplicas biológicas nesse experimento eram compostas pelas proteínas extraídas de duas folhas +1 de SP80-3280 por tempo de coleta, enquanto que nos experimentos com as folhas +1 das espécies ancestrais e da cultivar híbrida cada réplica biológica era composta pelo material oriundo de cinco folhas +1, optou-se por utilizar, nesse caso, um parâmetro mais estrigente para a análise quantitativa. Dessa forma foi considerado o *p-value* ajustado (Benjamini-Hochberg) $< 0,01$ ao invés de 0,05, utilizado nas análises de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280.

A partir dessa análise quantitativa, 65 proteínas apresentaram diferenças significativas nos níveis de expressão considerando os diferentes estágios de desenvolvimento da planta (Tabela Suplementar 15). Os valores de LFQ intensity foram submetidos a uma etapa de normalização (*Z-score*) e então submetidos à análise de *clustering* hierárquico resultando no perfil de expressão de proteínas presentes em folhas +1 de SP80-3280 em diferentes estágios de desenvolvimento da planta (Figura 43).

Figura 43. *Heat Map* das proteínas diferencialmente expressas ($n= 65$) entre os proteomas de folhas +1 de SP80-3280, coletadas em quatro diferentes estágios de desenvolvimento da planta (4, 8, 11 e 13 meses). Os valores de LFQ *intensity* das proteínas cujas abundâncias relativas variaram significativamente entre os grupos (one-way ANOVA, p -value ajustado $< 0,01$) foram normalizados (Z-score) e submetidos à análise de *cluster* hierárquico. Regiões do mapa em vermelho representam as proteínas com valores de expressão relativa maiores, enquanto regiões em azul, valores de expressão relativa menores. As réplicas biológicas de cada grupo são representadas no mapa por RB1 a RB3 e os grupos (5 *clusters*) formados de acordo com o perfil relativo de expressão são mostrados ao lado direito do mapa.

De acordo com essa análise, verificou-se a formação de cinco *clusters* considerando os valores de expressão relativa entre as proteínas nos diferentes tempos de coleta. Ainda, no que concerne ao perfil de expressão das proteínas com diferentes abundâncias, a análise do *heat map* mostrado na Figura 43 demonstrou a maior correlação entre os perfis de expressão das proteínas presentes em folhas +1 de plantas

de 4, 8 e 11 meses, que formaram um grupo distinto daquelas presentes em folhas +1 de plantas de 13 meses. A análise individual dos cinco *clusters* hierárquicos, cada qual apresentando padrões de perfil médio de expressão distintos, está mostrada na Fig. 44.

Figura 44. Perfil de expressão do conjunto de proteínas diferencialmente expressas entre os proteomas de folhas +1 de SP80-3280, coletadas em quatro diferentes estágios de desenvolvimento da planta. Os *clusters* foram formados de acordo com os padrões de perfil de expressão relativa entre as proteínas nos quatro grupos considerados. As distribuições dos valores de expressão relativa são mostradas ao lado direito dos mapas, onde as linhas representadas pela coloração mais avermelhada correspondem às distribuições mais próximas da distribuição central de cada grupo, enquanto que linhas verdes correspondem às distribuições menos próximas à distribuição central.

O *cluster* 1 (n= 23) é composto por proteínas cuja expressão é semelhante entre os proteomas de folhas +1 coletadas em 8, 11 e 13 meses, e significativamente mais alta quando comparada com os níveis de expressão em folhas coletadas de plantas de 4 meses (Tabela Suplementar 15). No *cluster* 2 (n= 24) encontram-se proteínas com níveis mais altos de expressão em folhas coletadas no 13º mês, enquanto que no *cluster* 3 (n= 7), as proteínas que apresentaram níveis de expressão mais altos em folhas coletadas no 4º mês. Com relação ao *cluster* 4 (n= 3), este é composto por proteínas que apresentaram níveis de expressão mais altos em folhas coletadas quando as plantas apresentavam 4 e 8 meses de idade. Por fim, no *cluster* 5 (n= 8), observa-se a ocorrência de proteínas que apresentam níveis de expressão relativa mais altos em folhas coletadas no 4º, 8º e 11º mês.

Desse conjunto de proteínas diferencialmente expressas (Fig. 43 e Fig. 44; Tabela Suplementar 15), foram selecionadas as proteínas cujas abundâncias relativas variaram pelo menos quatro vezes na comparação entre os diferentes tempos em que as folhas foram coletadas (Tab. 19).

Tabela 19. Proteínas diferencialmente expressas em folhas +1 de SP80-3280 (SP) em quatro diferentes estágios de desenvolvimento da planta (4, 8, 11 e 13 meses). Somente foram listadas proteínas cuja abundância relativa variou pelo menos quatro vezes.

Proteína	Cluster	Processo Biológico	Valores médios de Log ₂ (LFQ Intensity)			
			SP 4m	SP 8m	SP 11m	SP 13m
<i>Cleavage/polyadenylation specificity factor, 25kDa subunit</i>	1	<i>biological_process</i>	15.99	17.48	17.23	18.26
<i>alanine:glyoxylate aminotransferase</i>	1	<i>photorespiration</i>	21.41	22.48	22.92	23.41
<i>alpha-L-arabinofuranosidase 1</i>	1	<i>xylan biosynthetic process</i>	19.73	21.83	21.47	21.36
<i>copper/zinc superoxide dismutase 1</i>	1	<i>removal of superoxide radicals</i>	22.25	24.16	24.31	24.50
<i>solaneyl diphosphate synthase 2</i>	2	<i>ubiquinone biosynthetic process</i>	16.07	16.87	17.98	19.02
<i>thiazole biosynthetic enzyme, chloroplast (ARA6) (THI1) (THI4)</i>	3	<i>cellular response to DNA damage stimulus</i>	24.09	22.92	21.91	22.85
<i>31-kDa RNA binding protein</i>	3	<i>base conversion or substitution editing</i>	24.53	21.53	22.25	21.83
<i>Protein kinase superfamily protein</i>	4	<i>nuclear-transcribed mRNA catabolic process</i>	21.60	22.01	19.47	19.08

A maior parte dessas proteínas participa de processos biológicos básicos, como a fotorrespiração, biossíntese de carboidratos, respiração aeróbica, e edição de RNA, com exceção de duas (*copper/zinc superoxide dismutase* e *thiazole biosynthetic enzyme, chloroplast (ARA6) (THI1) (THI4)*), que estão relacionadas com resposta ao estresse.

Foi interessante notar que apenas oito proteínas apresentaram variação de abundância relativa de quatro vezes entre as amostras de folhas de SP80-3280 dentro do período de 13 meses. Por outro lado, esse achado está em concordância com o resultado de identificação de proteínas exclusivas nas diferentes fases de desenvolvimento dessas folhas, que mostrou apenas 29 proteínas exclusivas. Em conjunto, esses resultados indicam que, de forma geral, os proteomas de folhas +1 não

apresentaram variações significativas, pelo menos com relação às proteínas identificadas neste estudo, e que possam ser correlacionadas com aspectos fenotípicos do tecido foliar.

4.7. ANÁLISE DOS FOSFOPROTEOMAS DE FOLHAS +1 DE SP80-3280 DURANTE QUATRO ESTÁGIOS DE DESENVOLVIMENTO DA PLANTA

Para a análise de seus fosfoproteomas, as amostras de proteínas de folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta (4, 8, 11 e 13 meses) foram submetidas à digestão tripsínica em solução e os peptídeos resultantes foram submetidos ao processo de enriquecimento de fosfopeptídeos, segundo os métodos HAMMOC e IMAC.

Com o objetivo de apresentar um panorama geral de identificação dos fosfopeptídeos nas amostras de folhas +1 de SP80-3280, as populações de fosfopeptídeos identificados pelo uso dos métodos HAMMOC e IMAC foram agrupadas e então filtradas de acordo com o sistema de classificação desenvolvido por Olsen *et al.* (2006), de forma a considerar a identificação somente daqueles peptídeos cuja probabilidade mínima de localização dos sítios de fosforilação foi maior ou igual a 75% (sítios de fosforilação classe I). A partir desta análise, foram identificados 580 fosfopeptídeos, sendo 558 únicos e pertencentes a 411 grupos de proteínas (Tabela Suplementar 16). Dentre os 580 fosfopeptídeos únicos identificados, 378 resultaram das análises de LC-MS/MS dos peptídeos enriquecidos pelo método HAMMOC, enquanto

que 285 foram identificados a partir do enriquecimento de fosfopeptídeos pelo método IMAC. No entanto, somente 83 fosfopeptídeos foram identificados concomitantemente pelos dois métodos (Figura 45; Tabela Suplementar 17), demonstrando que esses métodos de enriquecimento geraram populações distintas de fosfopeptídeos.

Figura 45. Diagrama de Venn mostrando o número e a distribuição dos fosfopeptídeos identificados em amostras de folhas +1 de SP80-3280, coletadas em quatro diferentes estágios de desenvolvimento da planta, de acordo com os métodos de enriquecimento de fosfopeptídeos HAMMOC e IMAC. Os peptídeos foram identificados segundo o critério de probabilidade mínima de 75% de localização do sítio de fosforilação e andromeda *delta score* maior ou igual a 6.

Considerando o número total de fosfopeptídeos identificados pelos dois métodos, 478 (82,4%) foram encontrados fosforilados em resíduos de serina, 89 (15,3%) em resíduos de treonina e 13 (2,2%) em resíduos de tirosina (Tab. 20). Como mencionado anteriormente (item 4.5) estes resultados concordam com as proporções encontradas

em uma ampla análise fosfoproteômica em células de arroz e *Arabidopsis*, conduzida por Nakagami *et al.* (2010).

Tabela 20. Sumário dos dados obtidos das análises dos fosfoproteomas dos conjuntos de proteínas extraídas de folhas +1 de SP80-3280 em quatro diferentes estágios de desenvolvimento da planta, de acordo com os métodos de enriquecimento de fosfopeptídeos HAMMOC e IMAC.

Categoria	HAMMOC + IMAC	HAMMOC	IMAC
n. de fosfopeptídeos*	580	378	285
n. de fosfoproteínas	411	300	216
n. de fosfopeptídeos únicos	558	371	265
Resíduos fosforilados (Ser:Thr:Tyr)	478:89:13 (82,4%, 15,3%, 2,2%)	338:36:4 (89,42%, 9,52%, 1,06%)	215:61:9 (75,44%, 21,4%, 3,16%)

* O critério de identificação dos sítios de fosforilação baseou-se na probabilidade mínima de localização de 75% (Olsen *et al.*, 2006) e delta score > 6.

Considerando de forma individual os resultados obtidos a partir da utilização do método HAMMOC, que foi conduzido no Departamento de Bioanálise Molecular e Celular, Escola Superior de Ciências Farmacêuticas, da Universidade de Kyoto, sob a supervisão do Prof. Yasushi Ishihama, observou-se a identificação total de 1.081 peptídeos, dos quais 649 foram identificados fosforilados (dados não mostrados), mostrando uma eficiência de enriquecimento de aproximadamente 60%. Uma possível causa para a baixa eficiência de enriquecimento observada pode estar relacionada à perda dos grupos fosfato ligados à cadeia peptídica, uma vez que estas ligações são relativamente lábeis. Como estratégia para tentar minimizar a perda dos grupos fosfato, as amostras de proteínas de folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta foram previamente submetidas à digestão tripsínica em solução em nosso laboratório e as amostras de peptídeos foram submetidas à completa secagem e, dessa forma transportadas até o laboratório do Prof.

Ishihama. Essa estratégia pode ser útil para se abolir a atividade de fosfatases, porém, não impede a perda espontânea dos grupos fosfato presentes nos fosfopeptídeos. Além disso, como discutido no item 4.5, o método de obtenção de proteínas de folhas +1 utilizado pode não ter privilegiado a extração das proteínas fosforiladas.

Dos 649 fosfopeptídeos, 378 foram identificados pertencentes à classe I de Olsen (Olsen *et al.*, 2006), sendo 371 únicos e pertencentes a 300 grupos de proteínas (Tab. 20; Tabela Suplementar 17; Fig. 45). Desse total, 338 (89,42%) foram encontrados fosforilados em resíduos de serina, 36 (9,52%) em resíduos de treonina e 4 (1,06%) em resíduos de tirosina. Ainda, 6 peptídeos foram identificados contendo dois sítios de fosforilação, enquanto que os 372 restantes foram identificados contendo somente um sítio de fosforilação.

Em relação às identificações geradas a partir da utilização do método IMAC, um total de 1463 peptídeos foram identificados, sendo que 829 correspondiam a fosfopeptídeos (dados não mostrados). Dessa forma, a eficiência de enriquecimento foi de aproximadamente 57%, que é menor do que aquela observada quando se utilizou o processo de enriquecimento HAMMOC sobre as amostras de peptídeos provenientes das digestões em solução das proteínas de folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta. De acordo com o critério de seleção de fosfopeptídeos classe I (Olsen *et al.*, 2006), foram identificados 285 fosfopeptídeos (265 únicos), pertencentes a 216 grupos de proteínas (Tab. 20; Tabela Suplementar 16; Fig. 45). Desse total, 215 (75,44%) foram encontrados fosforilados em resíduos de serina, 61 (21,4%) em resíduos de treonina e 9 (3,16%) em resíduos de tirosina. Observou-se um aumento nas proporções relativas de peptídeos identificados

fosforilados em treonina e tirosina quando comparado com as proporções observadas para os fosfopeptídeos identificados a partir do método HAMMOC. Entre os 285 fosfopeptídeos identificados a partir da utilização do método IMAC, 252 continham um sítio, 30 continham dois e 3 continham três sítios de fosforilação. Observou-se que a quantidade de peptídeos multi-fosforilados (duplamente ou triplamente fosforilados) foi relativamente maior quando se utilizou o método IMAC. No método HAMMOC, dos 378 fosfopeptídeos identificados, somente 6 continham dois sítios de fosforilação e nenhum foi identificado contendo três sítios (Tabela Suplementar 17). Pode-se atribuir esse resultado às diferenças de especificidade dos métodos baseados na interação dos fosfopeptídeos com os cátions metálicos Ti^{4+} e Fe^{3+} . Pelo fato da interação dos fosfopeptídeos pelo cátion metálico ser maior em métodos que utilizam Ti^{4+} quando comparados com os métodos baseados na utilização de Fe^{3+} , a eluição dos peptídeos multi-fosforilados, no primeiro caso, se torna um processo menos favorecido devido à alta afinidade desses peptídeos pelo Ti^{4+} , de forma que, quanto maior o número de grupos fosfatos por molécula, maior sua afinidade pelo cátion metálico e, conseqüentemente, maior sua estabilidade, tornando mais difícil seu desacoplamento. Essa diferença de especificidade inerente aos métodos pode explicar o aumento de 5 vezes no número de peptídeos duplamente fosforilados, além da observação de identificações exclusivas de peptídeos triplamente fosforilados quando se utilizou o método IMAC (Bodenmiller *et al.*, 2007; Yue *et al.*, 2015).

Nesse ponto caberia uma análise comparativa das proteínas fosforiladas identificadas em cada estágio de desenvolvimento das folhas +1 de SP80-3280 (4, 8, 11 e 13 meses) segundo os dois métodos de enriquecimento utilizados. No entanto, o fato

de que em uma das etapas de processamento dos peptídeos enriquecidos com TiO_2 e submetidos à marcação com os *tags* isobáricos TMT no método HAMMOC, as quatro amostras foram combinadas previamente à análise por LC-MS/MS, o que inviabilizou a análise da variação do fosfoproteoma ao longo dos 13 meses de desenvolvimento.

Dessa forma, as proteínas fosforiladas, identificadas a partir das abordagens de enriquecimento HAMMOC e IMAC em conjunto, foram submetidas à análise de busca nas categorias do *Gene Ontology*, combinada à análise de enriquecimento de termos, com o objetivo de verificar a representatividade relativa das fosfoproteínas identificadas dentro das categorias Componente Celular, Processo Biológico e Função Molecular (Fig. 46; Tabela suplementar 18).

Figura 46. Análise de enriquecimento de termos da população de fosfoproteínas identificadas em folhas +1 de SP80-3280, coletadas em quatro diferentes estágios de desenvolvimento da planta, segundo as categorias Componente Celular (A), Processo Biológico (B) e Função Molecular (C) do *Gene Ontology*.

De acordo com esta análise, verifica-se que os componentes celulares enriquecidos em fosfoproteínas identificadas neste estudo são principalmente núcleo, citosol, membranas e cloroplasto (Fig. 46A), enquanto que as fosfoproteínas identificadas participam de processos importantes no metabolismo energético celular, metabolismo de carboidratos, fotossíntese, controle dos processos de transcrição e tradução, e resposta à estresse, entre outros (Fig. 46, B e C). A observação de que as categorias “*regulation of transcription*” e “*transcription factor activity*” estão enriquecidas (Fig. 46, B e C) indica a participação importante de fosfoproteínas nos eventos de transcrição de DNA. A categoria “*protein binding*” também se destaca pelo alto enriquecimento relativo (Fig. 46C), evidenciando a importância dos eventos de fosforilação de proteínas na modulação das interações proteína-proteína como discutido no item 4.5.

Em conjunto, os resultados obtidos neste estudo, sobre os fosfoproteomas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280, caracteriza-se como uma fonte inicial de informação sobre o estado e os sítios de fosforilação das proteínas, a qual poderá ser ampliada com a utilização de um método de extração de proteínas otimizado para este tipo de tecido foliar, que privilegie a extração de proteínas fosforiladas e consequentemente resulte na identificação de um número mais alto de fosfopeptídeos.

5. CONSIDERAÇÕES FINAIS

O presente estudo atingiu sua meta principal, que era a aplicação de abordagens avançadas de proteômica para a análise do proteoma total e do fosfoproteoma de folhas de cana-de-açúcar.

Este estudo, de caráter pioneiro no Brasil, possibilitou a experimentação de diversas abordagens proteômicas, incluindo algumas que não estão mencionadas nesta tese, para encontrar as melhores condições de preparação de amostra, descomplexação de misturas de peptídeos tripsínicos, enriquecimento de fosopeptídeos, e diversas ferramentas de bioinformática, para explorar profundamente os proteomas de duas espécies ancestrais e uma cultivar híbrida de cana-de-açúcar.

Os resultados obtidos sobre a identificação do proteomas e fosfoproteomas de folhas +1 de *S. off*, *S. sp* e *SP* de 12 meses de idade, e quatro amostras de folhas +1 de *SP* coletadas ao longo de 13 meses de crescimento, são discutidos exaustivamente quanto à eficiência e aplicabilidade das abordagens de proteômica utilizadas para a análise de folhas, servindo portanto como fonte de informação sólida sobre estudos futuros nesta mesma área.

As análises proteômicas comparativas, qualitativa e quantitativa, permitiram a identificação de proteínas compartilhadas pelas diferentes folhas +1 e também proteínas exclusivas, além daquelas que se mostraram em diferentes abundâncias nos proteomas, de forma que estes achados oferecem pistas sobre a fisiologia de folhas de diferentes espécies e cultivares, e em diferentes estágios de desenvolvimento da planta.

Assim, investigações futuras sobre a composição proteica de folhas +1 de cana-de-açúcar poderão se beneficiar dos resultados mostrados aqui, que constituem o mais abrangente repertório de proteínas deste tecido foliar identificado até o momento, permitindo sua ampliação no sentido de avançar o conhecimento sobre os complexos processos fisiológicos que modulam o crescimento e a performance em cana-de-açúcar.

6. REFERÊNCIAS

Abdin MZ, Kiran U, Alam A. Analysis of osmotin, a PR protein as metabolic modulator in plants. *Bioinformation*. 2011;5(8):336-40.

Abramoff MD, Magalhaes PJ, Ram SJ. "Image Processing with ImageJ". *Biophotonics International*. 2004;11(7),36-42.

Agrawal GK, Jwa NS, Rakwal R. Rice proteomics: ending phase I and the beginning of phase II. *Proteomics*. 2009;9(4):935-63.

Alberts B, Johnson A, Lewis J, *et al*. *Molecular Biology of the Cell*. 4th edition. New York: Garland Science. 2002.

Amalraj RS, Selvaraj N, Veluswamy GK, Ramanujan RP, Muthurajan R, Palaniyandi M, Agrawal GK, Rakwal R, Viswanathan R. Sugarcane proteomics: establishment of a protein extraction method for 2-DE in stalk tissues and initiation of sugarcane proteome reference map. *Electrophoresis*. 2010;31(12):1959-74.

Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics*. 2002;1(11),845-867.

Ascenzi R, Gantt JS. A drought-stress-inducible histone gene in *Arabidopsis thaliana* is a member of a distinct class of plant linker histone variants. *Plant Mol Biol*. 1997;34(4):629-41.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25(1):25-9.

Bantscheff M, Lemeer S, Savitski MM, Kuster B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem*. 2012;404(4):939-65.

Barnabas L, Ashwin NM, Kaverinathan K, Trentin AR, Pivato M, Sundar AR, Malathi P, Viswanathan R, Rosana OB, Neethukrishna K, Carletti P, Arrigoni G, Masi A, Agrawal GK, Rakwal R. Proteomic analysis of a compatible interaction between sugarcane and *Sporisorium scitamineum*. *Proteomics*. 2016;16(7):1111-22.

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Stat Soc (Methodological)*. 1995;57(1):289-300.

Benschop JJ, Mohammed S, O'Flaherty M, Heck AJ, Slijper M, Menke FL. Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. *Mol Cell Proteomics*. 2007;6(7):1198-214.

BNDES. Banco Nacional de Desenvolvimento Econômico e Social. Bioetanol de Cana-de-açúcar. Energia para o Desenvolvimento Sustentável. BNDES, Rio de Janeiro; 2008.

Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat Methods* 2007;4:231-7.

Bonhomme L, Valot B, Tardieu F, Zivy M. Phosphoproteome dynamics upon changes in plant water status reveal early events associated with rapid growth adjustment in maize leaves. *Mol Cell Proteomics*. 2012;11(10):957-72.

Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*. 1976;7(72):248-54.

Brandes EW. Origin, dispersal and use in breeding of the Melanesian garden sugarcane and their derivatives, *Saccharum officinarum* L. *Proc Int Soc Sugar Cane Technol*. 1956 9:709–50.

Brash AR. Lipoxygenases: occurrence, functions, catalysis, and acquisition of substrate. *J Biol Chem*. 1999;274(34):23679-82.

Bremer G. Problems in breeding and cytology of sugar cane. 4. Origin of increase of chromosome number in species hybrids of *Saccharum*. *Euphytica*. 1961;10(59–78):325–42.

Calderan-Rodrigues MJ, Jamet E, Bonassi MB, Guidetti-Gonzalez S, Begossi AC, Setem LV, Franceschini LM, Fonseca JG, Labate CA. Cell wall proteomics of sugarcane cell suspension cultures. *Proteomics*. 2014;14(6):738-49.

Calderan-Rodrigues MJ, Jamet E, Douché T, Bonassi MB, Cataldi TR, Fonseca JG, San Clemente H, Pont-Lezica R, Labate CA. Cell wall proteome of sugarcane stems: comparison of a destructive and a non-destructive extraction method showed differences in glycoside hydrolases and peroxidases. *BMC Plant Biol*. 2016;16:14.

Carpentier SC, Witters E, Laukens K, Deckers P, Swennen R, Panis B. Preparation of protein extracts from recalcitrant plant tissues: an evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics*. 2005;5(10):2497-507.

Carson DL, Botha FC. Preliminary analysis of expressed sequence tags for sugarcane. *Crop Sci*. 2000;40:1769–79.

Castillo-Davis CI, Hartl DL. GeneMerge--post-genomic analysis, data mining, and hypothesis testing. *Bioinformatics*. 2003;19(7):891-2.

Cesarino I, Araújo P, Sampaio Mayer JL, Paes Leme AF, Mazzafera P. Enzymatic activity and proteomic profile of class III peroxidases during sugarcane stem development. *Plant Physiol Biochem*. 2012;55:66-76.

Cesarino I, Araújo P, Paes Leme AF, Creste S, Mazzafera P. Suspension cell culture as a tool for the characterization of class III peroxidases in sugarcane. *Plant Physiol Biochem*. 2013;62:1-10.

Champion A, Kreis M, Mockaitis K, Picaud A, Henry Y. Arabidopsis kinome: after the casting. *Funct Integr Genomics*. 2004;4(3):163-87.

Cheavegatti-Gianotto A, de Abreu HM, Arruda P, Bessalho Filho JC, Burnquist WL, Creste S, di Ciero L, Ferro JA, de Oliveira Figueira AV, de Sousa Filgueiras T, Grossi-de-Sá MD, Guzzo EC, Hoffmann HP, de Andrade Landell MG, Macedo N, Matsuoka S, de Castro Reinach F, Romano E, da Silva WJ, de Castro Silva Filho M, César Ulian E. Sugarcane (*Saccharum X officinarum*): A Reference Study for the Regulation of Genetically Modified Cultivars in Brazil. *Trop Plant Biol*. 2011;4(1):62-89.

CONAB. Companhia Nacional de Abastecimento. Acompanhamento da Safra Brasileira: Cana-de-Açúcar, terceiro levantamento. 2015.

Cowan IR and Troughton JH. The relative role of stomata in transpiration and assimilation. *Planta*. 1971;97(4):325-36.

Cox J and Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*. 2008;26:1367-72.

Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res*. 2011;10(4):1794-805.

Cox J, Hein MY, Lubner CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics*. 2014;13(9):2513-26.

D'Hont A, Grivet L, Feldmann P, Rao P, Berding N, Glaszmann JC: Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. *Mol Gen Genet*. 1996;250:405–13.

D'Hont A. Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenet. Genome Res*. 2005;109:27–33.

Dai N, Schaffer A, Petreikov M, Shahak Y, Giller Y, Ratner K, Levine A, Granot D. Overexpression of Arabidopsis hexokinase in tomato plants inhibits growth, reduces photosynthesis, and induces rapid senescence. *Plant Cell*. 1999;11(7):1253-66.

D'Andrea LD, Regan L. TPR proteins: the versatile helix. *Trends Biochem Sci*. 2003;28(12):655-62.

Das AK, Cohen PW, Barford D. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J*. 1998;17(5):1192-9.

de Siqueira Ferreira S, Nishiyama M, Paterson A, Souza G. Biofuel and energy crops: high-yield Saccharinae take center stage in the post-genomics era. *Genome Biol*. 2013;14(6):210.

De Souza AP, Gaspar M, Da Silva EA, Ulian EC, Wacławovsky AJ, Nishiyama MY Jr, Dos Santos RV, Teixeira MM, Souza GM, Buckeridge MS. Elevated CO₂ increases photosynthesis, biomass and productivity, and modifies gene expression in sugarcane. *Plant Cell Environ*. 2008;31(8):1116-27.

Deeb SJ, D'Souza RC, Cox J, Schmidt-Supprian M, Mann M. Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles. *Mol Cell Proteomics*. 2012 May;11(5):77-89.

Dixon RA, Pasinetti GM. Flavonoids and isoflavonoids: from plant biology to agriculture and neuroscience. *Plant Physiol*. 2010;154(2):453-7.

dos Santos MF, Muniz de Pádua VL, de Matos Nogueira E, Hemerly AS, Domont GB. Proteome of *Gluconacetobacter diazotrophicus* co-cultivated with sugarcane plantlets. *J Proteomics*. 2010;73(5):917-31.

Dzieciatkowska M, Hill R, Hansen KC. GeLC-MS/MS analysis of complex protein mixtures. *Methods Mol Biol*. 2014;1156:53-66.

Ellis RJ. Most abundant protein in the world. *Trends in Biochem Sci*. 1979;4(11):241-4.

Espada J. Enzymic synthesis of adenosine diphosphate Glc from Glc 1-phosphate and adenosine triphosphate. *J Biol Chem*. 1962;237:3577–81.

Esser AF. The membrane attack complex of complement. Assembly, structure and cytotoxic activity. *Toxicology*. 1994;87(1-3):229-47.

Facette MR, Shen Z, Björnsdóttir FR, Briggs SP, Smith LG. Parallel proteomic and phosphoproteomic analyses of successive stages of maize leaf development. *Plant Cell*. 2013;25(8):2798-812.

Falcone Ferreyra ML, Rius SP, Casati P. Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front Plant Sci*. 2012;3:222.

Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science*. 1989;246(4926):64-71.

Ficarro SB, Adelmant G, Tomar MN, Zhang Y, Cheng VJ, Marto JA. Magnetic bead processor for rapid evaluation and optimization of parameters for phosphopeptide enrichment. *Anal Chem*. 2009;81(11):4566-75.

Fu S, Qiao-ling H, Xing H, Bao-qing Z, MingHui C, LiTao Y, YangRui L. Isolation of chloroplasts from leaf tissue of sugarcane and extraction of their protein. *Journal of Southern Agriculture*. 2011;42:463-7.

Fukuda M, Segami S, Tomoyama T, Asaoka M, Nakanishi Y, Gunji S, Ferjani A, Maeshima M. Lack of H(+)-pyrophosphatase Prompts Developmental Damage in Arabidopsis Leaves on Ammonia-Free Culture Medium. *Front Plant Sci*. 2016;7:819.

Gao J, Yin DH, Yao Y, Sun H, Qin Z, Schöneich C, Williams TD, Squier TC. Loss of conformational stability in calmodulin upon methionine oxidation. *Biophys J*. 1998;74(3):1115-34.

Gaxiola RA, Li J, Undurraga S, Dang LM, Allen GJ, Alper SL, Fink GR. Drought- and salt-tolerant plants result from overexpression of the AVP1 H⁺-pump. *Proc Natl Acad Sci U S A*. 2001;98(20):11444-9.

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. Global analysis of protein expression in yeast. *Nature*. 2003 Oct 16;425(6959):737-41.

Gordinho MC. Do álcool ao etanol: trajetória única. São Paulo. Ed. Terceiro Nome, 2010.

Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol*. 2001;313(4):903-19.

Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol*. 1999;1720–30.

Han C, Yang P, Sakata K, Komatsu S. Quantitative proteomics reveals the role of protein phosphorylation in rice embryos during early stages of germination. *J Proteome Res*. 2014;13(3):1766-82.

Hanks SK, Hunter T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J*. 1995;9(8):576-96.

Hanna SL, Sherman NE, Kinter MT, Goldberg JB. Comparison of proteins expressed by *Pseudomonas aeruginosa* strains representing initial and chronic isolates from a cystic fibrosis patient: an analysis by 2-D gel electrophoresis and capillary column liquid chromatography-tandem mass spectrometry. *Microbiology*. 2000;146(Pt 10):2495-508.

Hartt CE. 1963. Translocation as a factor in photosynthesis. *Naturwissenschaften* 50, 666–67.

Hatch MD, Agostino A, Burnell JN. Photosynthesis in phosphoenolpyruvate carboxykinase-type C4 plants: activity and role of mitochondria in bundle sheath cells. *Arch Biochem Biophys*. 1988;261(2):357-67.

Henry RJ, Kole C. Basic information on the sugarcane plant. In: Henry RJ and Kole C (eds) *Genetics, Genomics and Breeding of Sugarcane*, Science Publishers, Enfield, NH. 2000;1-8.

Hirano T, Kinoshita N, Morikawa K, Yanagida M. Snap helix with knob and hole: essential repeats in *S. pombe* nuclear protein nuc2+. *Cell*. 1990;60(2):319-28.

Hsiao TC. Plant responses to water stress. *Annu Rev Plant Physiol*. 1973;24(1):519-70.

Hu Y, Guo S, Li X, Ren X. Comparative analysis of salt-responsive phosphoproteins in maize leaves using Ti(4+)-IMAC enrichment and ESI-Q-TOF MS. *Electrophoresis*. 2013;34(4):485-92.

Hua YH, Wu CY, Sargsyan K, Lim C. Sequence-motif detection of NAD(P)-binding proteins: discovery of a unique antibacterial drug target. *Sci Rep*. 2014;4:6471.

Hunter T. Signaling—2000 and beyond. *Cell*. 2000;100, 113–27.

Hurkman WJ, Tanaka CK. Solubilization of plant membrane proteins for analysis by two-dimensional gel electrophoresis. *Plant Physiol*. 1986;81(3):802-6.

Imamura H, Wakabayashi M, Ishihama Y. Analytical strategies for shotgun phosphoproteomics: status and prospects. *Semin Cell Dev Biol*. 2012;23(8):836-42.

Initiative AG: Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 2000;408:796.

Irvine JE. Relations of Photosynthetic Rates and Leaf and Canopy Characters to Sugarcane Yield. *Crop Sci*. 1975;15(5):671-76.

Ishitani M, Liu J, Halfter U, Kim CS, Shi W, Zhu JK. SOS3 function in plant salt tolerance requires N-myristoylation and calcium binding. *Plant Cell*. 2000;12(9):1667-78.

Jangpromma N, Kitthaisong S, Lomthaisong K, Daduang S, Jaisil P, Thammasirirak S. A proteomics analysis of drought stress-responsive proteins as biomarker for drought-tolerant sugarcane cultivars. *Am J Biochem Biotechnol*. 2010;6: 89–102.

Johnson LN. The regulation of protein phosphorylation. *Biochem Soc Trans*. 2009;37:627–641.

Kamo M, Kawakami T, Miyatake N, Tsugita A. Separation and characterization of *Arabidopsis thaliana* proteins by two-dimensional gel electrophoresis. *Electrophoresis*. 1995;16(3):423-30.

Kawamura Y, Asai K, Ishii S, Nagano N. Systematic analyses of P-loop containing nucleotide triphosphate hydrolase superfamily based on sequence, structure and function. *Genome Informatics*. 2003;14: 581–2.

Kerbaux GB. *Fisiologia Vegetal*. 1a. ed. Rio de Janeiro: Guanabara Koogan. 2004. v1.

Kline KG, Barrett-Wilt GA, Sussman MR. In planta changes in protein phosphorylation induced by the plant hormone abscisic acid. *Proc Natl Acad Sci U S A*. 2010;107(36):15986-91.

Kokubu M, Ishihama Y, Sato T, Nagasu T, Oda Y. Specificity of immobilized metal affinity-based IMAC/C18 tip enrichment of phosphopeptides for protein phosphorylation analysis. *Anal Chem*. 2005;77(16):5144-54.

Komatsu S, Kajiwarra H, Hirano H. A rice protein library: a data-file of rice proteins separated by two-dimensional electrophoresis. *Theor Appl Genet*. 1993;86(8):935-42.

Kuijper J. DeGroeij van Bladschijf, Bladscheede em Stengel van het suikerriet. *Arch Suikerind Ned Indië*. 1915;23:528–56.

Kwon SJ, Jin HC, Lee S, Nam MH, Chung JH, Kwon SI, Ryu CM, Park OK. GDSL lipase-like 1 regulates systemic resistance associated with ethylene signaling in *Arabidopsis*. *Plant J*. 2009;58(2):235-45.

Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 1982;157(1):105-32.

Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. 1970;227(5259):680-5.

Lamb JR, Tugendreich S, Hieter P. Tetratrico peptide repeat interactions: to TPR or not to TPR? Trends Biochem Sci. 1995;20(7):257-9.

Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, Jørgensen TJ. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Mol Cell Proteomics. 2005;4(7):873-86.

Lery LM, Coelho A, von Kruger WM, Gonçalves MS, Santos MF, Valente RH, Santos EO, Rocha SL, Perales J, Domont GB, Teixeira KR, Bertalan M, Ferreira PC, Bisch PM. Protein expression profile of *Gluconacetobacter diazotrophicus* PAL5, a sugarcane endophytic plant growth-promoting bacterium. Proteomics. 2008;8(8):1631-44.

Lery LM, Hemerly AS, Nogueira EM, von Krüger WM, Bisch PM. Quantitative proteomic analysis of the interaction between the endophytic plant-growth-promoting bacterium *Gluconacetobacter diazotrophicus* and sugarcane. Mol Plant Microbe Interact. 2011;24(5):562-76.

Li M, Yin X, Sakata K, Yang P, Komatsu S. Proteomic Analysis of Phosphoproteins in the Rice Nucleus During the Early Stage of Seed Germination. J Proteome Res. 2015;14(7):2884-96.

Lin K, Hwang PK, Fletterick RJ. Distinct phosphorylation signals converge at the catalytic center in glycogen phosphorylases. Structure. 1997;5:1511-23.

Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR 3rd. Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol. 1999;17(7):676-82.

López JL. Two-dimensional electrophoresis in proteome expression analysis. J Chromatogr B Analyt Technol Biomed Life Sci. 2007;849(1-2):190-202.

Lopez-Huertas E, Charlton WL, Johnson B, Graham IA, Baker A. Stress induces peroxisome biogenesis genes. EMBO J. 2000;19(24):6770-7.

Luís IM, Alexandre BM, Oliveira MM, Abreu IA. Selection of an Appropriate Protein Extraction Method to Study the Phosphoproteome of Maize Photosynthetic Tissue. *PLoS One*. 2016;11(10):e0164387.

Maeshima M. Vacuolar H(+)-pyrophosphatase. *Biochim Biophys Acta*. 2000;1465(1-2):37-51.

Magdeldin S, Moresco JJ, Yamamoto T, Yates JR 3rd. Off-Line Multidimensional Liquid Chromatography and Auto Sampling Result in Sample Loss in LC/LC-MS/MS. *J Proteome Res*. 2014;13(8):3826-36.

Mann K, Mann M. In-depth analysis of the chicken egg white proteome using an LTQ Orbitrap Velos. *Proteome Sci*. 2011;9(1):7.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005;437(7057):376-80.

Maroco JP, Ku MSB, Lea PJ, Dever LV, Leegood RC, Furbank RT, Edwards GE. Oxygen requirement and inhibition of C4 photosynthesis: An analysis of C4 plants deficient in the C3 and C4 cycles. *Plant Physiol*. 1998;116:823–32.

Mazanek M, Mituloviae G, Herzog F, Stingl C, Hutchins JR, Peters JM, Mechtler K. Titanium dioxide as a chemo affinity solid phase in offline phosphopeptide chromatography prior to HPLC-MS/MS analysis. *Nat Protoc*. 2007;2(5):1059-69.

McCormick AJ, Cramer MD, Watt DA. Sink strength regulates photosynthesis in sugarcane. *New Phytol*. 2006;171(4):759-70.

McCormick AJ, Cramer MD, Watt DA. Changes in photosynthetic rates and gene expression of leaves during a source-sink perturbation in sugarcane. *Ann Bot.* 2008;101(1):89-102.

McCormick AJ, Watt DA, Cramer MD. Supply and demand: sink regulation of sugar accumulation in sugarcane. *J Exp Bot.* 2009;60(2):357-64.

Menossi M, Silva-Filho MC, Vincentz M, Van-Sluys MA, Souza GM. Sugarcane functional genomics: gene discovery for agronomic trait development. *Int J Plant Genomics.* 2008;2008:458732.

Micallef L, Rodgers P. eulerAPE: drawing area-proportional 3-Venn diagrams using ellipses. *PLoS One.* 2014;9(7):e101717.

Michael TP, Jackson S: The first 50 plant genomes. *Plant Genome* 2013;6.

Michael TP, VanBuren R. Progress, challenges and the future of crop genomes. *Curr Opin Plant Biol.* 2015;24:71-81.

Mishina TE, Zeier J. The Arabidopsis flavin-dependent monooxygenase FMO1 is an essential component of biologically induced systemic acquired resistance. *Plant Physiol.* 2006;141(4):1666-75.

Moisyadi S, Harrington HM. Characterization of the heat shock response in cultured sugarcane cells : I. Physiology of the heat shock response and heat shock protein synthesis. *Plant Physiol.* 1989;90(3):1156-62.

Mølgaard A, Kauppinen S, Larsen S. Rhamnogalacturonan acetyltransferase elucidates the structure and function of a new family of hydrolases. *Structure.* 2000;8(4):373-83.

Moore PH. Anatomy and morphology. In: Heinz DJ (ed) *Sugarcane improvement through breeding.* Elsevier, Amsterdam. 1987;85–142.

Moore PH. Temporal and spatial regulation of sucrose accumulation in the sugarcane stem. *Austr J Plant Physiol.* 1995;22:661–79.

Morita-Yamamuro C, Tsutsui T, Sato M, Yoshioka H, Tamaoki M, Ogawa D, Matsuura H, Yoshihara T, Ikeda A, Uyeda I, Yamaguchi J. The Arabidopsis gene *CAD1* controls programmed cell death in the plant immune system and encodes a protein containing a MACPF domain. *Plant Cell Physiol.* 2005;46(6):902-12.

Mortz E, Krogh TN, Vorum H, Görg A. Improved silver staining protocols for high sensitivity protein identification using matrix-assisted laser desorption/ionization-time of flight analysis. *Proteomics.* 2001;1(11):1359-63.

Murad AM, Molinari HB, Magalhães BS, Franco AC, Takahashi FS, de Oliveira- NG, Franco OL, Quirino BF. Physiological and proteomic analyses of *Saccharum* spp. grown under salt stress. *PLoS One.* 2014;9(6):e98463.

Nakagami H, Sugiyama N, Mochida K, Daudi A, Yoshida Y, Toyoda T, Tomita M, Ishihama Y, Shirasu K. Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. *Plant Physiol.* 2010;153(3):1161-74.

Nakagami H. *StageTip*-based HAMMOC, an efficient and inexpensive phosphopeptide enrichment method for plant shotgun phosphoproteomics. *Methods Mol Biol.* 2014;1072:595-607.

Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem.* 2003;75(17):4646-58.

Neuhaus JM. Plant Chitinases (PR-3, PR-4, PR-8, PR-11). In: Datta SK and Muthukrishnan S. *Pathogenesis-Related Proteins in Plants*. CRC Press. 1999; 77-105.

Neuhoff V, Arold N, Taube D, Ehrhardt W. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis.* 1988;9(6):255-62.

Ngamhui NO, Akkasaeng C, Zhu YJ, Tantisuwichwong N, Roytraku S, Sansayawichai T. Differentially expressed proteins in sugarcane leaves in response to water deficit stress. *Plant Omics* 2012;5:365–71.

Nishi H, Hashimoto K, Panchenko AR. Phosphorylation in protein-protein binding: effect on stability and function. *Structure*. 2011;19(12):1807-15.

Nishiyama-Jr MY, de Siqueira Ferreira S, Tang P-Z, Becker S, Pörtner-Taliana A, Souza GM. Full-length enriched cDNA libraries and ORFeome analysis of sugarcane hybrid and ancestor genotypes. *Plos One*. 2014;9(9):e107351.

Nühse TS, Stensballe A, Jensen ON, Peck SC. Phosphoproteomics of the Arabidopsis plasma membrane and a new phosphorylation site database. *Plant Cell*. 2004;16(9):2394-405.

O'Donoghue P, Luthey-Schulten Z. On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol Mol Biol Rev*. 2003;67(4):550-73.

OECD/Food and Agriculture Organization of the United Nations. OECD-FAO Agricultural Outlook 2015, OECD Publishing; 2015.

O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J Biol Chem*. 1975;250(10):4007-21.

Oliveros JC (2007-2015) Venny. An interactive tool for comparing lists with Venn's diagrams. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.

Olsen JV, Blagoev B, Gnäd F, Macek B, Kumar C, Mortensen P, Mann M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*. 2006;127: 635–48.

Pacheco CM, Pestana-Calsa MC, Gozzo FC, Mansur Custodio Nogueira RJ, Menossi M, Calsa T Jr. Differentially delayed root proteome responses to salt stress in sugar cane varieties. *J Proteome Res*. 2013;12(12):5681-95.

Papini-Terzi FS, Rocha FR, Vêncio RZ, Oliveira KC, Felix Jde M, Vicentini R, Rocha Cde S, Simões AC, Ulian EC, di Mauro SM, da Silva AM, Pereira CA, Menossi M, Souza GM. Transcription profiling of signal transduction-related genes in sugarcane tissues. *DNA Res*. 2005;12(1):27-38.

Papini-Terzi FS, Rocha FR, Vêncio RZ, Felix JM, Branco DS, Wacławovsky AJ, Del Bem LE, Lembke CG, Costa MD, Nishiyama MY Jr, Vicentini R, Vincentz MG, Ulian EC, Menossi M, Souza GM. Sugarcane genes associated with sucrose content. *BMC Genomics*. 2009;10:120.

Paterson AH, Moore PH, Tew TL. The gene pool of *Saccharum* species and their improvement. *Genomics of the Saccharinae*. Edited by: Paterson AH. Springer. 2013;11:43-72.

Pechan T, Ye L, Chang Y, Mitra A, Lin L, Davis FM, Williams WP, Luthe DS. A unique 33-kD cysteine proteinase accumulates in response to larval feeding in maize genotypes resistant to fall armyworm and other Lepidoptera. *Plant Cell*. 2000;12(7):1031-40.

Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res*. 2003;2(1):43-50.

Porta H, Rocha-Sosa M. Plant lipoxygenases. Physiological and molecular features. *Plant Physiol*. 2002;130(1):15-21.

Pursglove JW. Tropical crops: Monocotyledons. Longman Scientific and Technical. New York. 1972.

Qiu J, Hou Y, Tong X, Wang Y, Lin H, Liu Q, Zhang W, Li Z, Nallamilli BR, Zhang J. Quantitative phosphoproteomic analysis of early seed development in rice (*Oryza sativa* L.). *Plant Mol Biol*. 2016;90(3):249-65.

Que Y, Xu L, Lin J, Ruan M, Zhang M, Chen R. Differential protein expression in sugarcane during sugarcane-*Sporisorium scitamineum* interaction revealed by 2-DE and MALDI-TOF-TOF/MS. *Comp and Func Genomics*. 2011;Article ID 989016.

Rahman MA, Ren L, Wu W, Yan Y. Proteomic analysis of PEG-induced drought stress responsive protein in TERF1 overexpressed sugarcane (*Saccharum officinarum*) leaves. *Plant Mol Biol Rep*. 2015;33(3):716-30.

Rakwal R, Yang G, Komatsu S. Chitinase induced by jasmonic acid, methyl jasmonate, ethylene and protein phosphatase inhibitors in rice. *Mol Biol Rep.* 2004;31(2):113-9.

Ramagopal S. Protein polymorphism in sugarcane revealed by two-dimensional gel analysis. *Theor Appl Genet.* 1990;79(3):297-304.

Ramagopal S. Protein variation accompanies leaf dedifferentiation in sugarcane (*Saccharum officinarum*) and is influenced by genotype. *Plant Cell Rep.* 1994;13(12):692-6.

Rancé I, Fournier J, Esquerré-Tugayé MT. The incompatible interaction between *Phytophthora parasitica* var. *nicotianae* race 0 and tobacco is suppressed in transgenic plants expressing antisense lipoxygenase sequences. *Proc Natl Acad Sci U S A.* 1998;95(11):6554-9.

Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using *StageTips*. *Nat Protoc.* 2007;2(8):1896-906.

Reiland S, Messerli G, Baerenfaller K, Gerrits B, Endler A, Grossmann J, Gruissem W, Baginsky S. Large-scale Arabidopsis phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. *Plant Physiol.* 2009;150(2):889-903.

Roach BT. Origin and improvement of the genetic base of sugarcane. *Proc Aust Soc Sugar Cane Technol.* 1989;11:34–47.

Rocha FR, Papini-Terzi FS, Nishiyama MY Jr, Vêncio RZ, Vicentini R, Duarte RD, de Rosa VE Jr, Vinagre F, Barsalobres C, Medeiros AH, Rodrigues FA, Ulian EC, Zingaretti SM, Galbiatti JA, Almeida RS, Figueira AV, Hemerly AS, Silva-Filho MC, Menossi M, Souza GM. Signal transduction-related responses to phytohormones and environmental challenges in sugarcane. *BMC Genomics.* 2007;8:71.

Rossmann MG, Liljas A, Branden CI, Banaszak LT. Evolutionary and structural relationships among dehydrogenases. *The Enzymes.* 1975;11:61–102.

Rutowicz K, Puzio M, Halibart-Puzio J, Lirski M, Kotliński M, Kroteń MA, Knizewski L, Lange B, Muszewska A, Śniegowska-Świerk K, Kościelniak J, Iwanicka-Nowicka R, Buza K, Janowiak F, Żmuda K, Jõesaar I, Laskowska-Kaszub K, Fogtman A, Kollist H, Zielenkiewicz P, Tiuryn J, Siedlecki P, Swiezewski S, Ginalski K, Kobłowska M, Archacki R, Wilczynski B, Rapacz M, Jerzmanowski A. A Specialized Histone H1 Variant Is Required for Adaptive Responses to Complex Abiotic Stress and Related DNA Methylation in Arabidopsis. *Plant Physiol.* 2015;169(3):2080-101.

Sage RF, Sage TL, Kocacinar F. Photorespiration and the evolution of C4 photosynthesis. *Annu Rev Plant Biol.* 2012;63:19-47.

Sano A, Nakamura H. Chemo-affinity of titania for the column-switching HPLC analysis of phosphopeptides. *Anal Sci.* 2004;20(3):565-6.

Saravanan RS, Rose JK. A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. *Proteomics.* 2004;4(9):2522-32.

Schlüter A, Real-Chicharro A, Gabaldón T, Sánchez-Jiménez F, Pujol A. PeroxisomeDB 2.0: an integrative view of the global peroxisomal metabolome. *Nucleic Acids Res.* 2010;38(Database issue):D800-5.

Segalla AL. Botânica, melhoramento e variedades. In: Malavolta *et al* (eds) *Cultura e adubação da cana-de-açúcar*. Instituto Brasileiro de Potassa. 1964;61–98.

Sels J, Mathys J, De Coninck BM, Cammue BP, De Bolle MF. Plant pathogenesis-related (PR) proteins: a focus on PR peptides. *Plant Physiol Biochem.* 2008;46(11):941-50.

Serrato AJ, de Dios Barajas-López J, Chueca A, Sahrawy M. Changing sugar partitioning in FBPase-manipulated plants. *J Exp Bot.* 2009;60(10):2923-31.

Sheen J. C4 Gene Expression. *Annu Rev Plant Physiol Plant Mol Biol.* 1999;50:187-217.

Sikorski RS, Boguski MS, Goebel M, Hieter P. A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. *Cell*. 1990;60(2):307-17.

Silva DIR. Caracterização do proteoma nuclear de folhas de cana-de-açúcar (*Saccharum* spp) de 1 e 4 meses de idade. Dissertação. Piracicaba, Escola Superior de Agricultura Luiz de Queiroz. 2012; 170p.

Singh NK, Handa AK, Hasegawa PM, Bressan RA. Proteins Associated with Adaptation of Cultured Tobacco Cells to NaCl. *Plant Physiol*. 1985;79(1):126-37.

Smith AM, Zeeman SC, Smith SM. Starch degradation. *Annu Rev Plant Biol*. 2005;56:73-98.

Song X, Huang X, Tian D, Yang L, Li Y. Proteomic analysis of sugarcane seedling in response to *Ustilago scitaminea* infection. *Life Sci Journal* 2013;10:3026-35.

Souza GM, Simoes ACQ, Oliveira KC, Garay HM, Fiorini LC, Gomes FdS, Nishiyama-Junior MY, Silva AMd. The sugarcane signal transduction (SUCAST) catalogue: prospecting signal transduction in sugarcane. *Genet Mol Biol*. 2001;24(1-4):25-34.

Souza GM, da Silva AM. Desvendando as vias de transdução de sinal da cana-de-açúcar. *Revista Biotecnologia*. 2002;25:58-63.

Speicher KD, Kolbas O, Harper S, Speicher DW. Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. *J Biomol Tech*. 2000;11(2):74-86.

Spitzer M, Wildenhain J, Rappsilber J, Tyers M. BoxPlotR: a web tool for generation of box plots. *Nat Methods*. 2014;11(2):121-2.

Sreenivasan TV, Ahloowalia BS, Heinz DJ. Cytogenetics. In Sugarcane improvement through Breeding. DJ Heinz, ed. Elsevier. 1987;211–53.

Stejskal K, Potěšil D, Zdráhal Z. Suppression of peptide sample losses in autosampler vials. *J Proteome Res.* 2013;12(6):3057-62.

Stensballe A, Andersen S, Jensen ON. Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. *Proteomics.* 2001;1(2):207-22.

Su Y, Xu L, Wang Z, Peng Q, Yang Y, Chen Y, Que Y. Comparative proteomics reveals that central metabolism changes are associated with resistance against *Sporisorium scitamineum* in sugarcane. *BMC Genomics.* 2016;17(1):800.

Sugiharto B, Ermawati N, Mori H, Aoki K, Yonekura-Sakakibara K, Yamaya T, Sugiyama T, Sakakibara H. Identification and characterization of a gene encoding drought-inducible protein localizing in the bundle sheath cell of sugarcane. *Plant Cell Physiol.* 2002;43(3):350-4.

Sugiyama N, Masuda T, Shinoda K, Nakamura A, Tomita M, Ishihama Y. Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications. *Mol Cell Proteomics.* 2007;6(6):1103-9.

Sugiyama N, Nakagami H, Mochida K, Daudi A, Tomita M, Shirasu K, Ishihama Y. Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in *Arabidopsis*. *Mol Syst Biol.* 2008;4:193.

Tai PYP, Miller JD. A core collection for *Saccharum spontaneum* L. from the world collection of sugarcane. *Crop Science Society of America.* 2001;41:879-85.

Taiz L, Zeiger E. *Plant Physiology*. Edition ed.: Sinauer Associates. 2002.

Tan CM, Chen RJ, Zhang JH, Gao XL, Li LH, Wang PR, Deng XJ, Xu ZJ. OsPOP5, a prolyl oligopeptidase family gene from rice confers abiotic stress tolerance in *Escherichia coli*. *Int J Mol Sci.* 2013;14(10):20204-19.

Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T, Matsuo T. Protein and polymer analyses up to m/z 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom*. 1988;2:151–53.

Trapani JA, Smyth MJ. Functional significance of the perforin/granzyme cell death pathway. *Nat Rev Immunol*. 2002;2(10):735-47.

Tyers M, Mann M. From genomics to proteomics. *Nature*. 2003;422(6928):193-7.

UNICA. União da Indústria de Cana-de-Açúcar. Disponível em <<http://www.unicadata.com.br/>>. Acesso em: 26 de out. 2016.

van Loon LC. Pathogenesis-related proteins. *Plant Mol Biol*. 1985;4(2-3):111-6.

van Midwoud PM, Rieux L, Bischoff R, Verpoorte E, Niederländer HA. Improvement of recovery and repeatability in liquid chromatography-mass spectrometry analysis of peptides. *J Proteome Res*. 2007;6(2):781-91.

van Ooijen G, Mayr G, Kasiem MM, Albrecht M, Cornelissen BJ, Takken FL. Structure-function analysis of the NB-ARC domain of plant disease resistance proteins. *J Exp Bot*. 2008;59(6):1383-97.

Vettore AL, da Silva FR, Kemper EL, Arruda P. The libraries that made SUCEST. *Genet Mol Biol*. 2001;24:1–7.

Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, Ferro MI, Henrique-Silva F, Giglioti EA, Lemos MV, Coutinho LL, Nobrega MP, Carrer H, França SC, Bacci Júnior M, Goldman MH, Gomes SL, Nunes LR, Camargo LE, Siqueira WJ, Van Sluys MA, Thiemann OH, Kuramae EE, Santelli RV, Marino CL, Targon ML, Ferro JA, Silveira HC, Marini DC, Lemos EG, Monteiro-Vitorello CB, Tambor JH, Carraro DM, Roberto PG, Martins VG, Goldman GH, de Oliveira RC, Truffi D, Colombo CA, Rossi M, de Araujo PG, Sculaccio SA, Angella A, Lima MM, de Rosa Júnior VE, Siviero F, Coscrato VE, Machado MA, Grivet L, Di Mauro SM, Nobrega FG, Menck CF, Braga MD, Telles GP, Cara FA, Pedrosa G, Meidanis J, Arruda P. Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. *Genome Res*. 2003;13(12):2725-35.

Vieira Dos Santos C, Cuiné S, Rouhier N, Rey P. The Arabidopsis plastidic methionine sulfoxide reductase B proteins. Sequence and activity characteristics, comparison of the expression with plastidic methionine sulfoxide reductase A, and induction by photooxidative stress. *Plant Physiol.* 2005;138(2):909-22.

Waldron JC, Glasziou KT, Bull TA. The physiology of sugarcane.IX. Factors affecting photosynthesis and sugar storage. *Australian Journal of Biological Sciences.* 1967;20:1043–52.

Walker JE, Saraste M, Runswick MJ, Gay NJ. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* 1982;1(8):945-51.

Wang W, Scali M, Vignani R, Spadafora A, Sensi E, Mazzuca S, Cresti M. Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. *Electrophoresis.* 2003;24(14):2369-75.

War AR, Paulraj MG, Ahmad T, Buhroo AA, Hussain B, Ignacimuthu S, Sharma HC. Mechanisms of plant defense against insect herbivores. *Plant Signal Behav.* 2012;7(10):1306-20.

Warwood S, Byron A, Humphries MJ, Knight D. The effect of peptide adsorption on signal linearity and a simple approach to improve reliability of quantification. *J Proteomics.* 2013;85:160-4.

Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol.* 2001;19(3):242-7.

Watt D, Butterfield M, Hockett B. Proteomics and Metabolomics. In: Henry RJ and Kole C (eds) *Genetics, genomics and breeding of sugarcane.* CRC, USA. 2010;193-227.

Westermeier R and Naven T. Expression Proteomics. In *Proteomics in Practice: A Laboratory Manual of Proteome Analysis*, Wiley-VCH Verlag GmbH, Weinheim, FRG. 2002.

Wienkoop S, Baginsky S, Weckwerth W. *Arabidopsis thaliana* as a model organism for plant proteome research. J Proteomics. 2010;73:2239–48.

Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods. 2009;6(5):359-62.

Wolschin F, Wienkoop S, Weckwerth W. Enrichment of phosphorylated proteins and peptides from complex mixtures using metal oxide/hydroxide affinity chromatography (MOAC). Proteomics. 2005;5(17):4389-97.

Wolters DA, Washburn MP, Yates JR 3rd. An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem. 2001;73(23):5683-90.

Wu FS, Wang MY. Extraction of proteins for sodium dodecyl sulfate-polyacrylamide gel electrophoresis from protease-rich plant tissues. Anal Biochem. 1984;139(1):100-3.

Yarmush ML and Jayaraman A. Advances in Proteomic Technologies. Annu Rev Biomed Eng. 2002;4:349–73.

Yates JR 3rd. Mass spectrometry and the age of the proteome. J Mass Spectrom. 1998;33(1):1-19.

Ye J, Zhang Z, Long H, Zhang Z, Hong Y, Zhang X, You C, Liang W, Ma H, Lu P. Proteomic and phosphoproteomic analyses reveal extensive phosphorylation of regulatory proteins in developing rice anthers. Plant J. 2015;84(3):527-44.

Yoda H, Hiroi Y, Sano H. Polyamine oxidase is one of the key elements for oxidative burst to induce programmed cell death in tobacco cultured cells. Plant Physiol. 2006;142(1):193-206.

Yue X, Schunter A, Hummon AB. Comparing Multi-Step IMAC and Multi-Step TiO₂ Methods for Phosphopeptide Enrichment. Anal Chem. 2015;87(17):8837-44.

Zagdańska B, Wiśniewski K. Endoproteinase activities in wheat leaves upon water deficit. Acta Biochim Pol. 1996;43(3):515-9.

Zhang Y, Fonslow BR, Shan B, Baek MC, Yates JR 3rd. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev.* 2013;113(4):2343-94.

Zhou G, Yang LT, Li YR, Zou CL. *et al.* Proteomic analysis of osmotic stress responsive proteins in sugarcane leaves. *Plant Mol Biol Rep.* 2012;30:349–59.

Zhu XG, Long SP, Ort DR. What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Curr Opin Biotechnol.* 2008;19(2):153-59.

7. LISTA DE ANEXOS (EM CD-ROM)

7.1. FIGURAS SUPLEMENTARES

Figura Suplementar 1. Histogramas mostrando as distribuições dos dados de Lfq *intensity* (escala Log2) oriundos das análises de LC-MS/MS dos peptídeos provenientes do processo de digestão tripsínica em solução das amostras de proteínas extraídas de folhas +1 de SP80-3280 coletadas em diferentes estágios de desenvolvimento da planta.

7.2. TABELAS SUPLEMENTARES

Tabela Suplementar 1. Lista de proteínas identificadas em amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 a partir das análises de LC-MS/MS dos peptídeos gerados pela digestão tripsínica em solução e busca em banco de dados.

Tabela Suplementar 2. Lista de proteínas identificadas em amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 a partir das análises de LC-MS/MS dos peptídeos gerados pela digestão tripsínica *in gel* e busca em banco de dados.

Tabela Suplementar 3. Lista de proteínas identificadas em amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 a partir da metodologia MudPIT *off-line* aplicada aos peptídeos gerados pela digestão tripsínica em solução.

Tabela Suplementar 4. Lista de proteínas identificadas em amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 a partir da metodologia MudPIT "*on-line*" aplicada aos peptídeos gerados pela digestão tripsínica em solução.

Tabela Suplementar 5. Lista de proteínas identificadas em amostras de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com a combinação dos resultados de identificação de proteínas pelas abordagens *Shotgun/bottom-up*, GeLC-MS/MS e MudPIT *off-line* e "*on-line*".

Tabela Suplementar 6. Termos enriquecidos no conjunto de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 segundo as categorias Função Molecular, Processo Biológico e Componente Celular do *Gene Ontology*.

Tabela Suplementar 7. Lista das superfamílias identificadas nos conjuntos de proteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com as classificações presentes na biblioteca do banco de dados *Superfamily*.

Tabela Suplementar 8. Valores de intensidade normalizados [$\text{Log}_2(\text{LFQ intensity})$] das proteínas presentes em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280.

Tabela Suplementar 9. Valores de intensidade normalizados [$\text{Log}_2(\text{LFQ intensity})$] das proteínas presentes em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 após a substituição dos valores não válidos (NaN) pela ferramenta "*replace missing values from normal distribution*".

Tabela Suplementar 10. Lista de proteínas diferencialmente expressas em amostras de proteínas extraídas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com a análise quantitativa *label-free* MaxLFQ.

Tabela Suplementar 11. Lista de fosfopeptídeos identificados nas amostras de proteínas extraídas de folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 de acordo com a utilização do método de enriquecimento de fosfopeptídeos IMAC e análise por LC-MS/MS.

Tabela Suplementar 12. Lista de fosfoproteínas identificadas em folhas +1 de *S. officinarum*, *S. spontaneum* e SP80-3280 a partir da análise de LC-MS/MS dos fosfopeptídeos derivados do processo de enriquecimento IMAC.

Tabela Suplementar 13. Lista de proteínas identificadas em amostras de folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta de acordo com a abordagem de análise *Shotgun/bottom-up*.

Tabela Suplementar 14. Lista de termos enriquecidos nas categorias Componente Celular, Processo Biológico e Função Molecular do *Gene Ontology* de acordo com a

análise de enriquecimento de termos das proteínas identificadas em folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta.

Tabela Suplementar 15. Lista de proteínas diferencialmente expressas em amostras de proteínas extraídas de folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta de acordo com a análise quantitativa *label-free* MaxLFQ.

Tabela Suplementar 16. Lista de fosfoproteínas identificadas em folhas +1 de SP80-3280 de acordo com a utilização dos métodos de enriquecimento HAMMOC e IMAC e análise dos peptídeos por LC-MS/MS.

Tabela Suplementar 17. Lista de fosfopeptídeos identificados por LC-MS/MS a partir da utilização dos métodos de enriquecimento HAMMOC e IMAC sobre peptídeos provenientes da digestão tripsínica em solução de amostras de proteínas extraídas de folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta.

Tabela Suplementar 18. Lista de categorias enriquecidas do *Gene Ontology* (Componente Celular, Processo Biológico e Função Molecular) de acordo com a análise de enriquecimento de termos das proteínas identificadas a partir da análise do fosfoproteoma de folhas +1 de SP80-3280 coletadas em quatro diferentes estágios de desenvolvimento da planta.

8. ANEXOS

8.1. SÚMULA CURRICULAR

Eduardo Shigueo Kitano

Nascimento 26/03/1983 – São Caetano do Sul/SP – Brasil

Endereço eletrônico eduardosh.kitano@gmail.com

Formação:

2011 Doutorado em Ciências Biológicas (Bioquímica) – Instituto de Química, Universidade de São Paulo, USP, Brasil.

2004 – 2010 Bacharel em Farmácia e Bioquímica. Faculdade de Ciências Farmacêuticas da Universidade de São Paulo, FCF-USP, Brasil.

Formação Complementar:

2015 2nd International Mass Spectrometry School, Natal, Brasil.

2015 VI Proteomics Workshop. Centro Nacional de Pesquisa em Energia e Materiais (CNPEM) em Campinas – SP. Evento realizado nos dias 4 e 5 de novembro de 2015.

2015 Proteomics Informatics / TPP Course. Centro Nacional de Pesquisa em Energia e Materiais (CNPEM) em Campinas – SP. Curso organizado pelo ISB (Institute for Systems Biology).

2014 1^a Escola Brasileira de Espectrometria de Massas. (Carga horária: 40h). Sociedade Brasileira de Espectrometria de Massas.

2012 MaxQuant Summer School 2012. (Carga horária: 48h). Max-Planck-Institut für Biochemie.

2012 Curso de Interpretação de Espectros de Massas. (Carga horária: 24h). Fundação Oswaldo Cruz, RJ, Brasil.

2011 Abordagens proteômicas para a Biologia Sistêmica. (Carga horária: 12h). Instituto Butantan, IBU, Brasil.

2010 Proteomics Methods and Approaches for Protein Identification. (Carga horária: 12h). Universidade de São Paulo, USP, Brasil.

2009 Proteômica Quantitativa e Interpretação de Espectros de Massas (Carga horária: 8h). Sociedade Brasileira de Espectrometria de Massas.

Atuação Profissional:

1. Instituto Butantan, IBU, Brasil

2011 – Atual Vínculo: Aluno de Doutorado; Enquadramento Funcional: Aluno de Doutorado; Orientadora: Dra. Solange M. T. Serrano

2009 – 2010 Vínculo: estagiário; Enquadramento Funcional: Aluno Iniciação Científica; Orientadora: Dra. Solange M. T. Serrano

2006 – 2007 Vínculo: estágio; Enquadramento Funcional: Aluno Iniciação Científica; Orientadora: Dra. Solange M. T. Serrano

2. Universidade de São Paulo, USP, Brasil

2008 – 2009 Vínculo: estagiário; Enquadramento Funcional: Aluno Iniciação Científica; Orientador: Dr Flavio Finardi Filho

3. Universidade de Kyoto, Escola Superior de Ciências Farmacêuticas

2016 – 2016 Vínculo: Aluno de Doutorado; Enquadramento Funcional: *Visiting Graduate Student*; Carga horária: 80 horas; Supervisor: Dr. Yasushi Ishihama

Artigos Completos Publicados e Periódicos:

1. Folle AM, **Kitano ES**, Lima A, Gil M, Cucher M, Mourglia-Ettlin G, Iwai LK, Rosenzvit M, Batthyány C, Ferreira AM. Characterisation of Antigen B Protein Species Present in the Hydatid Cyst Fluid of *Echinococcus canadensis* G7 Genotype. *PLoS Negl Trop Dis*. 2017;11(1):e0005250.
2. Augusto-de-Oliveira C, Stuginski DR, **Kitano ES**, Andrade-Silva D, Liberato T, Fukushima I, Serrano SM, Zelanis A. Dynamic Rearrangement in Snake Venom Gland Proteome: Insights into *Bothrops jararaca* Intraspecific Venom Variation. *J Proteome Res*. 2016;15(10):3752-3762.
3. Abreu TF, Sumitomo BN, Nishiyama MY Jr, Oliveira UC, Souza GH, **Kitano ES**, Zelanis A, Serrano SM, Junqueira-de-Azevedo I, Silva PI Jr, Tashima AK. Peptidomics of *Acanthoscurria gomesiana* spider venom reveals new toxins with potential antimicrobial activity. *J Proteomics*. 2017;151:232-242.
4. Andrade-Silva D, Zelanis A, **Kitano ES**, Junqueira-de-Azevedo IL, Reis MS, Lopes AS, Serrano SM. Proteomic and Glycoproteomic Profilings Reveal That Post-translational Modifications of Toxins Contribute to Venom Phenotype in Snakes. *J Proteome Res*. 2016;15(8):2658-75.
5. Menezes MC, Imbert L, **Kitano ES**, Vernet T, Serrano SM. Recombinant expression of the precursor of the hemorrhagic metalloproteinase HF3 and its non-catalytic domains using a cell-free synthesis system. *Amino Acids*. 2016;48(9):2205-14.
6. Zelanis A, Menezes MC, **Kitano ES**, Liberato T, Tashima AK, Pinto AF, Sherman NE, Ho PL, Fox JW, Serrano SM. Proteomic identification of gender molecular markers in *Bothrops jararaca* venom. *J Proteomics*. 2016;139:26-37.

7. Silva-Álvarez V, Folle AM, Ramos AL, **Kitano ES**, Iwai LK, Corraliza I, Córscico B, Ferreira AM. Echinococcus granulosus Antigen B binds to monocytes and macrophages modulating cell response to inflammation. *Parasit Vectors*. 2016;9:69.
8. Dias MH, **Kitano ES**, Zelanis A, Iwai LK. Proteomics and drug discovery in cancer. *Drug Discov Today*. 2016;21(2):264-77.
9. Kodama RT, Cajado-Carvalho D, Kuniyoshi AK, **Kitano ES**, Tashima AK, Barna BF, Takakura AC, Serrano SM, Dias-Da-Silva W, Tambourgi DV, Portaro FV. New proline-rich oligopeptides from the venom of African adders: Insights into the hypotensive effect of the venoms. *Biochim Biophys Acta*. 2015;1850(6):1180-7.
10. Yamashiro ET, Oliveira AK, **Kitano ES**, Menezes MC, Junqueira-de-Azevedo IL, Paes Leme AF, Serrano SM. Proteoforms of the platelet-aggregating enzyme PA-BJ, a serine proteinase from *Bothrops jararaca* venom. *Biochim Biophys Acta*. 2014;1844(12):2068-76.
11. Castilho DG, Chaves AF, Xander P, Zelanis A, **Kitano ES**, Serrano SM, Tashima AK, Batista WL. Exploring potential virulence regulators in *Paracoccidioides brasiliensis* isolates of varying virulence through quantitative proteomics. *J Proteome Res*. 2014;13(10):4259-71.
12. Dias GS, **Kitano ES**, Pagotto AH, Sant'anna SS, Rocha MM, Zelanis A, Serrano SM. Individual variability in the venom proteome of juvenile *Bothrops jararaca* specimens. *J Proteome Res*. 2013;12(10):4585-98.
13. **Kitano ES**, Garcia TC, Menezes MC, Tashima AK, Zelanis A, Serrano SM. Cotiarinase is a novel prothrombin activator from the venom of *Bothrops cotiara*. *Biochimie*. 2013;95(8):1655-9.
14. Tashima AK, Zelanis A, **Kitano ES**, Ianzer D, Melo RL, Rioli V, Sant'anna SS, Schenberg AC, Camargo AC, Serrano SM. Peptidomics of three *Bothrops* snake venoms: insights into the molecular diversification of proteomes and peptidomes. *Mol Cell Proteomics*. 2012;11(11):1245-62.

15. Paes Leme AF, **Kitano ES**, Furtado MF, Valente RH, Camargo AC, Ho PL, Fox JW, Serrano SM. Analysis of the subproteomes of proteinases and heparin-binding toxins of eight *Bothrops* venoms. *Proteomics*. 2009;9(3):733-45.