**PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA
UNIVERSIDADE DE SÃO PAULO**

**Bioinformática aplicada à biologia sistêmica para a identificação dos fatores
regulatórios do acúmulo de sacarose no colmo da cana-de-açúcar**

Fabício Edgar de Moraes

Tese de doutorado apresentada ao Programa
Interunidades de Pós-Graduação em
Bioinformática para a obtenção do título de Doutor
em Ciências.

**São Paulo
2016**

PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA
UNIVERSIDADE DE SÃO PAULO

Bioinformática aplicada à biologia sistêmica para a identificação dos fatores regulatórios do acúmulo de sacarose no colmo da cana-de-açúcar

Fabício Edgar de Moraes

Tese de doutorado apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática para a obtenção do título de Doutor em Ciências.

Área de concentração: Bioinformática

Orientador: Carlos Alberto Labate

São Paulo
2016

Aos meus pais João Edgar de Moraes e Teresinha Rossi de Moraes, por todo carinho e compreensão, mesmo sem saber direito o que eu faço até hoje.

Aos meus irmãos Guilherme Edmar de Moraes e Priscila Augusta de Moraes, pelo carinho, amizade, apoio, críticas e ensinamentos.

Aos recém-chegados na família, Edlena, Conrado, Théo e Beatriz, por aceitarem fazer parte desta família.

Ao bebê Willy, por toda sua alegria e carinho.

À minha grande namorada Ana Paula Cavalcante Quaresma, por todo carinho, amor, dedicação e por conseguir compreender as necessidades de meu trabalho e ter ficado ao meu lado todo esse tempo. Te AMO Paulinha.

DEDICO

AGRADECIMENTOS

Ao Prof. Dr. Carlos Alberto Labate pela orientação, confiança e oportunidade em participar de seu grupo de pesquisa.

À Dra. Mônica Tereza Veneziano Labate pelos ensinamentos principalmente na área de proteômica e por cuidar tão bem de nosso laboratório.

À coordenação do programa de Bioinformática interunidades, pelos esforços contínuos para a melhoria do programa.

Ao DVATCOM pela moradia durante meus dois primeiros anos de doutorado.

Às secretárias do programa de Bioinformática (Patrícia e Cristiane), pela atenção, disponibilidade e pelas várias respostas por email.

À Thaís pela grandíssima ajuda em áreas que eu tinha pouco ou nenhum conhecimento, por nossas discussões científicas, por todo seu carinho com minhas amostras e pela total disposição no auxílio para a condução deste projeto.

À Ilara, por todas discussões científicas, ensinamentos, ajuda na coleta das amostras, por compartilhar todas suas experiências em metabolômica, proteômica, viagens e diversas outras áreas que eu tive dúvidas.

Ao Fernando, por compartilharmos as mesmas amostras e as mesmas dúvidas nos experimentos, pelas discussões científicas, pela ajuda na coleta das amostras e convívio.

Ao Thiago Falda, Felipão, Juliana Calderan, Leonardo, Ana Paula Silveira pelas conversas apesar do breve convívio.

À Lívia por todo seu suporte em informática.

Ao Felipinho, por nossas discussões científicas, aulas de inglês, por organizar o “bonde do rucas”, pela grande ajuda em todas coletas e por todas as risadas durante esses anos.

Ao Ivan pelas ajudas nas coletas, convívio, aulas de inglês, “Os mais legais”, cervejas e boas risadas.

À Andressa, Marisângela, Simone, Janaína, Mariana, Danielle, Ana Paula Chiaverini, Ana Cláudia, Jaqueline, Nathalia, Bruna, Alessandra, Vanessa e Aline pela convivência.

À Hana e Sarina pelo convívio e grande ajuda nas coletas.

À Flavia e Maria Letícia (que acabou de chegar no lab.) pelo convívio, cervejas e “post-it”.

Ao David e Carol pelas discussões científicas e conversas aleatórias.

À Juliana Guimarães Fonseca, por todas as perguntas que nunca me deixavam parar de pensar, pelo convívio em todos esses anos, por sua amizade e carinho.

À FAPESP pelo auxílio financeiro (Processo 2012/12521-2), que foi essencial para meu desenvolvimento como pesquisador e para uma ótima condução deste trabalho.

SUMÁRIO

SUMÁRIO.....	7
RESUMO.....	9
ABSTRACT.....	11
1. INTRODUÇÃO	13
2. REVISÃO BIBLIOGRÁFICA	16
2.1. Cana-de-açúcar	16
2.2. Acúmulo de sacarose na cana-de-açúcar	17
2.3. Metabolômica	19
2.4. Proteômica.....	21
2.5. Biologia Sistêmica.....	24
2.6. Integração dos dados pela Análise de Correlação Canonica Regularizada.....	26
3. MATERIAL E MÉTODOS	28
3.1. Material vegetal e condições de cultivo.....	28
3.2. Análise do Metabolôma.....	29
3.2.1. Teste de solvente de extração	29
3.2.2. Teste de padrões internos	29
3.2.3. Extração de metabólitos e análise por LC-MS	30
3.2.4. Análise do perfil metabólico e estatística dos dados de LC-MS	30
3.2.5. Identificação dos metabólitos por MS/MS	31
3.3. Análise do Proteôma.....	32
3.3.1. Extração das proteínas	32
3.3.2. Dessalinização das amostras.....	33
3.3.3. Quantificação das proteínas.....	33
3.3.4. Verificação da qualidade da extração de proteínas	33
3.3.5. Digestão das amostras de proteínas totais	34
3.3.6. Purificação das amostras de proteínas	34
3.3.7. LC-MSE	34
3.3.8. Análise dos dados.....	35
3.4. Integração dos dados de metabolômica e proteômica	36
3.4.1. Normalização dos dados.....	36
3.4.2. Integração por rCCA	37
4. RESULTADOS E DISCUSSÃO	39
4.1. Coleta das amostras	39
4.2. Análise do perfil metabólico	40
4.2.1. Teste de solventes de extração	40

4.2.2. Teste de padrões internos.....	42
4.2.3. Perfil metabólico da folha +1	46
4.2.4. Perfil metabólico do entrenó 1.....	48
4.2.5. Perfil metabólico do entrenó 5.....	50
4.2.6. Perfil metabólico do entrenó 9.....	52
4.2.7. Identificação dos metabólitos por MS/MS	54
4.2.7.1. VIPs anotados da análise da folha +1 em todos os tempos de coleta	54
4.2.7.2. VIPs anotados da análise do entrenó 1 com 8, 11 e 13 meses após o plantio	59
4.2.7.3. VIPs anotados da análise do entrenó 5 com 11 e 13 meses após o plantio	62
4.2.7.4. VIPs anotados da análise do entrenó 9 com 11 e 13 meses após o plantio	65
4.3. Proteômica.....	67
4.3.1. Extração das Proteínas	67
4.3.2. Perfil proteico da folha +1 com 11 e 13 meses após o plantio.....	68
4.3.3. Perfil proteico do entrenó 5 com 11 e 13 meses após o plantio.....	71
4.3.4. Perfil proteico do entrenó 9 com 11 e 13 meses após o plantio.....	74
4.4. Integração dos dados de metabolômica e proteômica	77
4.4.1. Rede biológica da folha +1 com 11 e 13 meses após o plantio	78
4.4.2. Rede biológica do entrenó 5 com 11 e 13 meses após o plantio.....	89
4.4.3. Rede biológica do entrenó 9 com 11 e 13 meses após o plantio.....	108
5. CONSIDERAÇÕES FINAIS.....	117
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	120
ANEXO.....	132

RESUMO

A cana-de-açúcar (*Saccharum* spp.) é uma das principais gramíneas cultivadas do mundo e o Brasil é seu maior produtor, ela se tornou uma importante cultura devido às altas taxas de assimilação de carbono permitindo a síntese e acumulação de grandes quantidades de sacarose em seus entrenós. Com isso, faz-se necessário uma melhor compreensão dos mecanismos moleculares que regulam o acúmulo de sacarose nesta planta. Tais mecanismos têm sido estudados em vários níveis, tais como, identificação e localização de genes, identificação de loci controladores de características quantitativas, transcriptoma, proteoma, caracterização e identificação de metabólitos. Com todos esses estudos é evidente a necessidade de uma abordagem holística para o entendimento global da planta durante o acúmulo de sacarose. Assim este trabalho teve por objetivo integrar dados de metabolômica e proteômica de tecidos da cana-de-açúcar da variedade SP80-3280 durante o desenvolvimento e o acúmulo de sacarose, utilizando a bioinformática para unir esses resultados por meio da análise de correlação canônica regularizada em uma abordagem de biologia sistêmica. Os resultados obtidos indicam diferenças no perfil metabólico e proteico da cana-açúcar durante o desenvolvimento e acúmulo de sacarose. Foram propostas classes de metabólitos que podem estar relacionados com o acúmulo de sacarose na cana-de-açúcar tais como glicerolipídeos, glicerofosfolipídeos, cumarinas e derivados, esteroides e derivados de esteroides e acil graxos. Também foram propostas proteínas que podem estar relacionadas com o acúmulo de sacarose, onde as histonas foram as que mais se destacaram. Nas redes biológicas de correlações também foram observadas correlações entre possíveis metabólitos e proteínas que podem estar correlacionadas com o acúmulo de sacarose na cana-de-açúcar.

ABSTRACT

Sugarcane (*Saccharum* spp.) is one of the most important cultivated grasses of the world and Brazil is the largest producer, it has become an important crop due to high carbon assimilation rates allowing the synthesis and accumulation of large amounts of sucrose in their internodes. Thus, it is necessary a high understanding of the molecular mechanisms involved in the regulation of sucrose accumulation in this plant. These mechanisms have been studied at various levels, such as gene identification and localization, identification of quantitative trait locus controlling, transcriptome, proteome, characterization and metabolites identification. With all these studies is evident the necessity for a holistic approach to global understanding of the plant during the sucrose accumulation. Thus, this work aims to integrate metabolomics and proteomics data from tissues of sugarcane variety SP80-3280, during plant development and sucrose accumulation, using bioinformatics to link these results by regularized canonical correlation analysis in a systems biology approach. The results indicate differences in the metabolic and protein profile of sugarcane during development and sucrose accumulation. Metabolites classes have been proposed that may be related to sugarcane sucrose accumulation as glycerolipids, glycerophospholipids, coumarins and derivatives, steroids and steroid derivatives and fatty acyl. In addition, some proteins have been proposed that may be related to sucrose accumulation, where the most highlighted were the histones. In the biological correlations networks, have been also observed correlations between possible metabolites and proteins that can be correlated with the accumulation of sucrose in sugarcane.

1. INTRODUÇÃO

O Brasil é hoje o maior produtor mundial de cana-de-açúcar e atualmente a área destinada à atividade sucroalcooleira brasileira está estimada em 8,6 milhões de hectares distribuídas em todos os estados produtores, sendo que a região sudeste concentra cerca de 63% da produção. O estado de São Paulo é o maior produtor de cana-de-açúcar do Brasil com uma área de 4,5 milhões de hectares disponíveis para colheita, representando quase 52% de toda a área utilizada na produção de cana-de-açúcar (CONAB, 2016).

A cana-de-açúcar se tornou uma importante cultura devido às altas taxas de assimilação de carbono permitindo a síntese e acumulação de grandes quantidades de sacarose em seus entrenós (MOORE, 1995).

As principais variedades modernas de cana-de-açúcar são híbridas provenientes de cruzamentos entre *Saccharum officinarum* e *Saccharum spontaneum*. Deste modo as variedades híbridas mantêm as características favoráveis de seus ancestrais possuindo um genoma poliploide com alta frequência de aneuploidia (NISHIYAMA et al., 2014). Devido à complexidade do seu genoma, o sucesso atual do cultivo desta cultura, deve-se à obtenção de variedades adaptadas aos diferentes ambientes edafoclimáticos por meio do melhoramento genético clássico, principalmente pelos centros de pesquisa e estações experimentais. No entanto, o lançamento de uma variedade para uma cultura pelo método de melhoramento clássico ocorre lentamente, demorando entre 14 a 15 anos para serem lançadas comercialmente (BUTTERFIELD; THOMAS, 1996). Assim faz-se necessário o melhor entendimento dos processos moleculares associados ao acúmulo de sacarose na cana-de-açúcar, para acelerar o processo de melhoramento dessa planta.

Dentre as técnicas moleculares consideradas *high-throughput*, que receberam esta denominação pelo grande volume de dados gerados, a metabolômica é um campo emergente. O metaboloma (conjunto completo de pequenas moléculas em um organismo) representa o fenótipo final de uma célula ou tecido em resposta a alterações genéticas ou ambientais. Além disso, o metaboloma também pode influenciar a expressão gênica e a função das proteínas (SAITO; MATSUDA, 2010).

A proteômica também tem se tornado uma poderosa ferramenta nos programas de melhoramento genético, pelo fato de fornecer informações ao nível molecular da variabilidade genética que é realmente expressa (PENNINGTON; DUNN, 2001), pois permite análises qualitativas e quantitativas de um grande número de proteínas que influenciam diretamente no sistema celular, proporcionando uma análise precisa das alterações na bioquímica celular, com isso é possível analisar um sistema ou organismo durante seu crescimento, desenvolvimento ou durante respostas a fatores ambientais (CHEN; HARMON, 2006).

Os mecanismos genéticos que controlam o acúmulo de sacarose na cana-de-açúcar têm sido estudados em vários níveis, tais como, identificação de genes da via de sacarose e de transporte (GROF; CAMPBELL, 2001; MOORE, 1995, 2005), proteoma (AMALRAJ et al., 2010), mapeamento gênico e identificação de locos controladores de características quantitativas (MING et al., 2002), transcriptoma (PAPINI-TERZI et al., 2009), caracterização e identificação de metabólitos (GLASSOP et al., 2007; UYS et al., 2007). Com isso faz-se necessário uma abordagem holística para maior compreensão do acúmulo de sacarose na cana-de-açúcar.

A biologia sistêmica pode ser definida como o estudo de interações entre componentes biológicos utilizando modelos e/ou redes para integrar genes, metabólitos, proteínas, elementos reguladores e outros componentes biológicos (YUAN et al., 2008). De acordo com Saito e Matsuda (2010), a biologia sistêmica baseada em dados de metabolômica, genômica e outras tecnologias moleculares *high-throughput*, tende a desempenhar um papel fundamental na compreensão dos sistemas de plantas e desenvolvimento de aplicações biotecnológicas. Com isso, existe a necessidade imediata da integração dos sistemas biológicos, considerando os grandes conjuntos de dados gerados a partir das diferentes tecnologias *ômicas* (YUAN et al., 2008).

A utilização do proteoma e metaboloma no estudo de algum sistema biológico permite a visualização de modificações ocorridas nas células ou tecidos de uma forma holística, elevando nossa compreensão sobre as redes metabólicas e seus componentes ao nível de biologia sistêmica (GLINSKI e WECKWERTH, 2006). Uma abordagem que vem crescendo para a integração de dois conjuntos de dados é a Análise de Correlação Canônica Regularizada (rCCA), que destaca a correlação entre dois conjuntos de dados adquiridos nas mesmas unidades experimentais (GONZALEZ et al., 2008). A rCCA é uma modificação da Análise de Correlação Canônica (CCA) clássica, que permite a análise de conjuntos de dados com um pequeno número de observações e um grande número de variáveis em ambos conjuntos de dados (WALLER et al., 2015).

Com isso, este trabalho teve por objetivo a integração de dados de metabolômica e proteômica durante o desenvolvimento e acúmulo de sacarose da cana-de-açúcar, utilizando uma abordagem no campo da biologia sistêmica. Assim, foi utilizado a variedade SP80-3280 da cana-de-açúcar plantadas em campo, as coletas foram realizadas com 4, 8, 11 e 13 meses após o plantio. As análises de metabolômica foram realizadas em folhas +1 (com 4, 8, 11 e 13 meses), entrenó 1 (com 8, 11 e 13 meses), entrenó 5 (com 11 e 13 meses) e entrenó 9 (com 11 e 13 meses) da cana-de-açúcar. Também foi realizado a proteômica da folha +1 (com 11 e 13 meses), do entrenó 5 (com 11 e 13 meses) e do entrenó 9 (com 11 e 13 meses) durante o desenvolvimento da planta. Para integrar estes dados, foi utilizado a rCCA e a

construção de redes biológicas de correlações da folha +1, entrenó 5 e entrenó 9, todos com 11 e 13 meses após o plantio.

Com os resultados obtidos, foi possível a identificação de diversos metabólitos e proteínas relacionados com o acúmulo de sacarose em folhas e entrenós de cana-de-açúcar. Também foi possível a integração dos dados obtidos, gerando uma visão holística a respeito do acúmulo de sacarose na cana ao longo do tempo. Ademais, os resultados apresentados servirão de base para estudos mais detalhados no âmbito de um melhor entendimento do acúmulo de sacarose em cana-de-açúcar, possibilitando o desenvolvimento de variedades com um maior rendimento.

2. REVISÃO BIBLIOGRÁFICA

2.1. Cana-de-açúcar

A cana-de-açúcar é uma monocotiledônea alógama e semi-perene, pertencente à família Poaceae e ao gênero *Saccharum*, e que provavelmente é originária do Sudeste Asiático, da região central da Nova Guiné e Indonésia (DANIELS; ROACH, 1987).

A cana-de-açúcar se tornou uma importante cultura devido às altas taxas de assimilação de carbono que permitem a síntese de grandes quantidades de sacarose nas folhas, que posteriormente são transportadas e armazenadas nos colmos (MOORE, 1995). Os entrenós imaturos da parte superior da planta têm baixa concentração de sacarose, porém com o desenvolvimento da planta, os entrenós inferiores podem acumular até 50% do peso seco total de sacarose (BOTHÁ; BLACK, 2000).

Atualmente o Brasil é o maior produtor mundial de cana-de-açúcar, também é o maior produtor de açúcar e é o segundo maior produtor mundial de etanol, ficando atrás somente dos Estados Unidos (UNICA, 2016). A produção da cana-de-açúcar na safra de 2015/2016 foi de 665,6 milhões de toneladas, com um crescimento de 4,9% em relação à safra anterior. Sozinho o estado de São Paulo produziu um total de 367,6 milhões de toneladas, gerando 55% da produção de cana-de-açúcar do Brasil. A produção de açúcar no Brasil, nesta mesma safra, foi de 33,5 milhões de toneladas e a produção de etanol total foi de 30,5 bilhões de litros (CONAB, 2016).

Nos dias atuais a área destinada à atividade sucroalcooleira brasileira está estimada em 8,6 milhões de hectares, distribuídas em todos estados produtores, onde o Sudeste concentra 63% do total dessa área. O estado de São Paulo possui a maior área destinada ao cultivo da cana com 4,5 milhões de hectares, representando aproximadamente 52% de toda a área utilizada na produção de cana-de-açúcar (CONAB, 2016).

O Brasil é um país modelo no desenvolvimento e uso comercial da cana-de-açúcar tanto no seu uso como biocombustível quanto na geração de energia elétrica a partir de sua biomassa (LAM et al., 2009). Também é possível destacar outros produtos de extrema importância que utilizam a cana-de-açúcar como matéria-prima, como a alimentação animal, cachaça, xarope de cana-de-açúcar, entre outros (CHEAVEGATTI-GIANOTTO et al., 2011).

A cana-de-açúcar é uma planta do tipo C4, portanto possui uma alta eficiência fotossintética e ponto de compensação luminoso elevado (MCCORMICK; CRAMER; WATT, 2008). Durante a fotossíntese ocorre a produção de carboidratos que serão usados primeiramente no desenvolvimento de folhas e raízes, depois os carboidratos são utilizados para a formação estrutural da planta e para acúmulo no colmo na forma de açúcares. Assim, quanto maior a saturação luminosa, maior a taxa fotossintética, aumentando

consequentemente a produção e acúmulo de sacarose durante a maturação dos entrenós, gerando um maior rendimento econômico (MCCORMICK; CRAMER; WATT, 2008; MOORE, 2005)

Uma das dificuldades no melhoramento genético da cana-de-açúcar é o fato de possuir um genoma altamente complexo, apresentando poliploidia, aneuploidia e um número variável de cromossomos. Também possui um genoma grande quando comparado com outras culturas, dificultando sua compreensão. O tamanho do DNA de uma única célula somática de *S. officinarum* é de aproximadamente 7440 Mpb. Como esta espécie é octaplóide, o tamanho do complemento haplóide é 8 vezes menor, medindo aproximadamente 930 Mpb, valor comparável ao genoma inteiro de uma espécie diplóide, como por exemplo o sorgo, que apresenta cerca de 760 Mpb (GRIVET; ARRUDA, 2002).

Nos dias atuais o sucesso do cultivo da cana-de-açúcar deve-se à obtenção de variedades adaptadas aos diferentes ambientes edafoclimáticos por meio de melhoramento genético clássico, principalmente pelos centros de pesquisa e estações experimentais. Este tipo de melhoramento é um processo relativamente demorado, visto que, dentre as espécies cultivadas, a cana-de-açúcar possui um dos genomas mais complexos, o que dificulta a aplicação de técnicas convencionais de melhoramento (VETTORE et al., 2001). Assim faz-se necessário um melhor entendimento dos processos moleculares na cana-de-açúcar, para acelerar o processo de melhoramento dessa planta.

2.2. Acúmulo de sacarose na cana-de-açúcar

A cana-de-açúcar se tornou importante pelo seus colmos que acumulam sacarose (AMALRAJ et al., 2010). Esse acúmulo se deve às altas taxas de assimilação de carbono que permitem a síntese de grandes quantidades de sacarose nas folhas e o armazenamento desse açúcar não redutor nos entrenós (MOORE, 1995). A sacarose por si só, pode ser uma importante reserva de energia em folhas, caules, tubérculos e frutas, podendo também ajudar a proteger as plantas de estresses ambientais, tais como frio e seca (LUNN; FURBANK, 1999). Também deve ser observado que a sacarose é o principal produto da fotossíntese em plantas e a principal forma em que os carboidratos são translocados das folhas para o resto da planta, fornecendo carbono e energia para o crescimento e também para o acúmulo de reservas energéticas (LUNN; FURBANK, 1999). Na cana-de-açúcar a sacarose é produzida nas folhas, depois translocada pelo floema e armazenada nos entrenós (GROF; CAMPBELL, 2001; MOORE, 2005).

O acúmulo de sacarose na cana-de-açúcar vem sendo estudado desde os anos 60, inicialmente foram realizados estudos em plantas inteiras e colmos. No entanto, a complexidade a este nível organizacional era grande demais para revelar detalhes sobre

como o acúmulo de sacarose era regulado a nível enzimático, celular ou molecular (MOORE, 1995).

Os tecidos que acumulam sacarose na cana-de-açúcar fazem isso contra um gradiente de concentração, usando energia fornecida pela respiração (BIELESKI, 1960), isto acontece da base da cana-de-açúcar para a parte mais apical do colmo. No início do desenvolvimento da planta os colmos estão crescendo rapidamente e a quantidade de açúcar é baixa em todos os entrenós. Já na fase madura os entrenós mais antigos atingem o nível máximo de acúmulo de açúcares, e os entrenós mais jovens vão gradativamente aumentando a quantidade de açúcar, assim, a concentração de açúcares é regulada no aspecto espacial e temporal (MOORE, 1995). Os entrenós imaturos da parte superior da cana têm baixa concentração de sacarose, porém com o desenvolvimento da planta, os entrenós nas partes inferiores podem acumular até 50% do peso seco total de sacarose (BOTHÁ; BLACK, 2000).

Sabe-se que a cana-de-açúcar acumula sacarose em seus entrenós, mas existe pouco conhecimento sobre a regulação deste processo. A sacarose pode ser armazenada como sacarose ou compartilhada entre a respiração, incluindo ambos os processos catabólico (glicólise) e anabólicos (gliconeogênese) e como um componente da parede celular insolúvel que consiste de celulose e lignina (WACLAWOVSKY et al., 2010).

Os mecanismos genéticos que controlam o acúmulo de sacarose na cana-de-açúcar têm sido estudados em vários níveis, tais como, identificação de genes na via de sacarose e de transporte (GROF; CAMPBELL, 2001; MOORE, 1995, 2005), proteôma (AMALRAJ et al., 2010), mapeamento gênico e identificação de locos controladores de características quantitativas (MING et al., 2002), transcriptoma (PAPINI-TERZI et al., 2009), caracterização e identificação de metabólitos (GLASSOP et al., 2007; UYS et al., 2007).

Moore, (2005) verificou que o crescimento e a maturação de colmos de cana-de-açúcar são associados com a expressão de genes que não possuem relação direta com o metabolismo de sacarose. Os fatores que regulam a quantidade de sacarose estocada na cana-de-açúcar ainda precisam ser melhor estudados (WACLAWOVSKY et al., 2010) mas estão incluídos relações entre água e célula e propriedades da molécula de sacarose, tais como a sua solubilidade e o transporte por membranas (MOORE; COSGROVE, 1991).

Em um estudo comparando a maturação dos colmos de cana-de-açúcar em 30 genótipos diferente, foram identificados genes da sinalização hormonal, resposta ao estresse, transporte de açúcar, biossíntese de lignina, teor de fibra, proteínas kinases, proteínas fosfatases e fatores de transcrição que estariam relacionados ao acúmulo de sacarose (PAPINI-TERZI et al., 2009). A **figura 1** apresenta um modelo proposto para as reações presentes no acúmulo de sacarose.

Figura 1: Esquema de reações do modelo de acumulo de sacarose na cana-de-açúcar. Os metabólitos são representados por retângulos cinza (metabólitos variáveis) ou círculos cinza (metabólitos fixos externos). As enzimas são representadas por retângulos brancos. O número 2 ao lado de uma linha indica o coeficiente estequiométrico. HK = hexoquinase; FRK = frutoquinase; UDPGDH = UDP-glicose desidrogenase; PFK = fosfofrutoquinase; PFP = pirofosfato-dependente; NI = invertase neutra; SuSy = sacarose sintase; ALD = aldolase; SPS = sacarose fosfato sintetase; SPase = sacarose fosfatase; VAC = importador de sacarose vacuolar; Glc = glicose; Fru = Frutose; Suc = sacarose; S6P = sacarose-6-fosfato; HexP = hexose fosfato; FbP = frutose-1,6-bifosfato; UDPGA, ácido UDP-glucurônico; Trp = triose fosfato. (Adaptado de UYS et al. (2007)).

As enzimas citosólicas frutose 1,6-bisfosfatase e a sacarose fosfato sintetase são as que mais influenciam na biossíntese da sacarose. A frutose 1,6-bisfosfatase catalisa a produção de frutose-6-fosfato e em conjunto com outras enzimas, a 6-fosfofrutoquinase e a pirofosfato: frutose-6-fosfato 1-fosfotransferase regulam a formação de um dos monossacarídeos necessários na síntese de sacarose. A sacarose fosfato sintetase catalisa o penúltimo passo na biossíntese de sacarose, com isso tem-se atribuído um papel chave na produção de sacarose (HUBER; HUBER, 1992).

2.3. Metabolômica

Dentre as técnicas moleculares consideradas *high-throughput*, que receberam esta denominação pelo grande volume de dados gerados, a metabolômica é um campo

emergente. Um metaboloma (conjunto completo de moléculas pequenas num organismo) representa o fenótipo final de células deduzidas pela perturbação da expressão do gene e da modulação das funções das proteínas, que são causadas pelo meio ambiente ou mutações. Além disso, o metaboloma também pode influenciar direta e indiretamente a expressão gênica e a função das proteínas (SAITO; MATSUDA, 2010).

Os metabólitos são os produtos finais derivados das diferenças na sequência de DNA, controle da expressão, tradução de proteínas e atividade enzimática, portanto podem ser utilizados como marcadores fenotípicos auxiliando no entendimento da base de características agronomicamente importantes (GLASSOP et al., 2007).

O número de metabólitos que podem ser detectados na metabolômica é dependente da amostra, do processo de extração e o método de análise, por esse fato, não existe uma única tecnologia (processo de extração, separação, detecção e análise) que possa identificar e quantificar todos os metabólitos presentes em uma amostra (FIEHN; WECKWERTH, 2003).

Com o avanço nos estudos de metabolômica, vários métodos de tecnologias *high-throughput* de análise e comparação de metabólitos em plantas foram estabelecidos, incluindo a cromatografia gasosa acoplada a espectrometria de massas (GC-MS), espectrometria de massa por injeção em fluxo (DFI-MS), cromatografia líquida acoplada a espectrometria de massas (LC-MS), eletroforese capilar acoplada a espectrometria de massas (CE-MS) e ressonância magnética nuclear (NMR), tendo um grande destaque para LC-MS, devido à bioquímica altamente rica das plantas, que abrange muitos compostos semi-polares, incluindo os principais grupos de metabólitos secundários, que são separados e detectados mais eficientemente por esta técnica (DE VOS et al., 2007).

Nos últimos anos a análise de metabólitos por espectrometria de massas (MS) tornou-se mais avançada (DETTMER; ARONOV; HAMMOCK, 2007; SAITO; MATSUDA, 2010), devido ao contínuo avanço dos espectrômetros de massas, técnicas de separação cromatográfica, desenvolvimento de softwares livres e comerciais, permitindo a extração de dados e processamento inicial (BENTON; WANT; EBBELS, 2010; KATAJAMAA; ORESIC, 2007; PLUSKAL et al., 2010; SMITH et al., 2006). Assim estes avanços serviram de base para a extração eficiente e confiável de informações dos dados brutos de MS (GIAVALISCO et al., 2011). Com isso as tecnologias existentes foram aperfeiçoadas, e dentre elas, a técnica por GC-MS, que identifica preferencialmente metabólitos primários (KUSANO et al., 2007) e a técnica por LC-MS, que identifica preferencialmente metabólitos secundários (MATSUDA et al., 2009).

A metabolômica foi denominada como um novo campo das *ômicas*, juntando-se a genômica, proteômica e transcriptômica, e como uma ciência utilizada para a compreensão da biologia sistêmica (ROCHFORD, 2005). Em geral, a espectrometria de massas faz

contribuições significativas para o desenvolvimento da biologia sistêmica, onde os processos biológicos complexos são tratados em termos de redes moleculares ao invés da análise molécula por molécula (SABIDÓ; SELEVSEK; AEBERSOLD, 2012). A ideia central por trás das abordagens *ômicas* é que um sistema biológico só pode ser compreendido quando observado de uma forma holística. Portanto, a ambição da genômica, transcriptômica e metabolômica é verificar o perfil de cada gene, transcrito e metabólito disponível, a fim de um entendimento completo do sistema estudado (REDESTIG et al., 2011).

Atualmente, a metabolômica vem sendo utilizada com sucesso em diversas áreas. No campo do melhoramento genético de plantas vários estudos estão obtendo sucesso com a utilização dessa tecnologia. Como no trabalho realizado por Hu et al. (2014) que estudaram a variação metabólica entre duas variedades de arroz. Neste trabalho os autores conseguiram identificar o perfil de 121 metabólitos nos grãos das duas variedades de arroz, que revelaram correlações entre o fenótipo metabólico e origem geográfica dos grãos.

Em outro estudo Baniasadi et al. (2014) verificando os efeitos do ambiente e genótipo em híbridos comerciais de milho, por meio da metabolômica, encontraram um total de 286 e 857 metabólitos em amostras de grão e forragem, respectivamente. Utilizando estatística multivariada, eles conseguiram identificar que o ambiente tem um efeito maior que o genótipo sobre o metaboloma.

Chebrolu et al. (2016) estudando o impacto do estresse térmico no metaboloma de sementes de soja, utilizando um genótipo tolerante e outro sensível ao calor, conseguiram identificar diferenças na abundância de metabólitos relacionados ao genótipo e ao estresse térmico. Foi identificado um aumento na abundância de um conjunto diversificado de metabólitos antioxidantes, incluindo tocoferóis, flavonoides, fenilpropanoides e precursores de ascorbato nas sementes do genótipo tolerante. Essa maior abundância indica que esses compostos são prováveis responsáveis, pelo menos em parte, pela tolerância a altas temperaturas durante o desenvolvimento das sementes de soja.

Vários outros trabalhos de metabolômica também obtiveram resultados satisfatórios na análise do perfil metabólico de plantas, como em rosas (SAINT-LARY et al., 2016), *Arabidopsis* (WIDEMANN et al., 2015), uva (CUADROS-INOSTROZA et al., 2016), *citrus* (WANG et al., 2016) e tomate (ROLDAN et al., 2014).

2.4. Proteômica

O termo proteômica se originou do termo proteoma que foi proposto primeiramente em 1996 por Wilkins et al. (1996), que seria a junção das palavras proteína e genoma. O proteoma é definido por todas as proteínas expressadas por um genoma, tecido ou célula, este conceito tem algumas diferenças quanto ao genoma, pois um genoma é específico e definitivo para um

do organismo, já o proteoma pode se alterar sobre diferentes condições, podendo também ser diferente nos diversos tecidos do mesmo organismo (WILKINS et al., 1996). Outra singularidade é que o proteoma nem sempre é um produto direto do genoma, podendo ter um número bem maior de proteínas do que o número de genes de um organismo, devido as proteínas serem expressas por *splicing* alternativos ou diferentes modificações pós-tradução (WILKINS et al., 1996). A proteômica também pode ser definida como o uso de medidas quantitativas do nível de proteínas oriundas da expressão gênica, para caracterizar processos biológico e os mecanismos de controle da expressão gênica (ANDERSON; ANDERSON, 1998).

A proteômica tem se tornado uma poderosa ferramenta nos programas de melhoramento genético, pelo fato de fornecer informações ao nível molecular da variabilidade genética que é realmente expressa (PENNINGTON; DUNN, 2001), pois permite medições qualitativas e quantitativas de um grande número de proteínas que influenciam diretamente no sistema celular, proporcionando uma análise precisa das alterações na bioquímica celular, com isso é possível a análise de um sistema ou organismo durante seu crescimento, desenvolvimento ou durante respostas a fatores ambientais (CHEN; HARMON, 2006).

Uma das técnicas mais utilizadas na proteômica foi a eletroforese em gel bidimensional (2DE), e os pesquisadores da área reconhecem a importância da técnica de 2DE na história da proteômica, mas com a atual evolução da proteômica baseadas em espectrometria de massa, muitos pesquisadores acabaram por aposentar a 2DE em seus laboratórios (OLIVEIRA; COORSEN; MARTINS-DE-SOUZA, 2014).

A espectrometria de massa (MS) tem ganhado popularidade pela sua habilidade de trabalhar com as complexidades associadas ao proteoma. Outras tecnologias como 2DE e microarranjos de proteínas falham em conseguir a profundidade de informação conseguido pela análise do proteoma por MS (HAN; ASLANIAN; YATES, 2008). Na espectrometria de massa as moléculas são analisadas em função da razão massa/carga (m/z). O processo consiste na ionização de moléculas para que elas recebam ou deem prótons, ficando assim carregadas positivamente ou negativamente, após a ionização as moléculas são expostas a um campo elétrico que conduz os íons ao interior do analisador de acordo com as características químicas dessas moléculas (BUDZINSKI et al., 2014).

Dois técnicas de ionização revolucionaram a proteômica por espectrometria de massas, essas técnicas foram o desenvolvimento da ionização por *electrospray* (ESI) (FENN et al., 1981) e a ionização por dessorção a laser assistida por matriz (MALDI) (KARAS; HILLENKAMP, 1988).

Já dentre os analisadores, quatro se destacaram no estudo da proteômica, o quadrupolo (Q), o analisador de massa *ion trap* (quadrupolo *ion trap*, QIT; *ion trap* linear, LIT

ou LTQ), o analisador por tempo-de-voo (TOF) e o analisador de massas de ressonância ciclôtrica de íons com transformada de Fourier (FTICR), esses analisadores variam em seus princípios físicos e desempenho analítico. Com isso, a partir desses quatro analisadores, surgiram os analisadores híbridos para combinar as diferentes capacidades destes analisadores, que incluem o Q-q-Q, Q-q-LIT, Q-TOF, TOF-TOF e o LTQ-FTICR (HAN; ASLANIAN; YATES, 2008).

A proteômica quantitativa pode ser separada em duas abordagens principais, que seriam proteômica com a utilização de isótopos estáveis para a marcação (*stable isotope labelling*) e a técnica sem marcação conhecida como proteômica *label-free* (NEILSON et al., 2011)

Como alternativa a abordagem por 2DE, a cromatografia líquida multidimensional começou a ser utilizada para a proteômica (WOLTERS; WASHBURN; YATES, 2001), esta inovação juntamente com o desenvolvimento da tecnologia de *shotgun* para proteômica *label-free* favoreceu a análise pela tecnologia de identificação multidimensional de proteínas (MudPIT) conseguir uma rápida expansão em sua utilização nas análises quantitativas de proteínas utilizando MS. Assim este avanço tem sido estimulado pelas mudanças instrumentais em MS e a necessidade de quantificar um número cada vez maior de proteínas por análise (NEILSON et al., 2011).

A proteômica por *shotgun* é uma metodologia onde não é utilizado o gel 2D para a separação das proteínas ou peptídeos, essa metodologia é baseada na separação de uma mistura complexa de peptídeos por cromatografia multidimensional acoplada a um espectrômetro de massa (SWANSON; WASHBURN, 2005). O gel 2D pode ser substituído pela tecnologia de MudPIT que é um método automatizado, que combina cromatografia líquida multidimensional com ionização por *electrospray* e espectrometria de massas em *tandem*. O MudPIT utiliza uma coluna de troca catiônica (SCX), seguida de coluna de fase reversa (RP). As amostras proteicas digeridas por tripsina são carregadas para a coluna SCX e eluídas por um gradiente de aumento de concentração de sal, após cada fração de amostra é carregada para uma coluna de RP e eluída por um gradiente de aumento da hidrofobicidade diretamente na fonte ESI para ionização, seguindo para o analisador (WOLTERS; WASHBURN; YATES, 2001).

A proteômica por *shotgun* é análoga ao sequenciamento de DNA genômico por *shotgun*, onde o DNA é fragmentado, sequenciado e remontado computacionalmente afim de determinar o genoma do organismo estudado. Assim, na proteômica por *shotgun*, uma mistura complexa de proteínas é digerida em peptídeos por ação enzimática (tripsina), após é feita a separação por cromatografia multidimensional e eluição dos peptídeos diretamente em um espectrômetro de massas, depois os algoritmos de computador mapeiam os peptídeos em

proteínas para determinar o conteúdo original da amostra complexa de proteínas (YATES, 1998).

A proteômica por *shotgun* tem sido utilizada vastamente para um melhor entendimento de fatores que possam influenciar determinadas funções ou características das plantas. Como no trabalho realizado por Heringer et al. (2015), que estudaram o proteoma, por *shotgun*, de calos de cana-de-açúcar durante a embriogênese somática, neste caso, a proteômica foi utilizada para investigar diferenças na aquisição de competências e desenvolvimento de embriões somáticos entre calos embriogênicos e não embriogênicos de cana-de-açúcar. Os autores conseguiram identificar 65 proteínas únicas nos calos embriogênicos no final do tratamento de maturação.

Kim et al. (2015) utilizando a proteômica quantitativa por *shotgun* para analisar proteínas da antera de arroz, após a exposição a altas temperaturas, identificaram 1944 proteínas em sua análise comparativa. Destas, 139 foram diferencialmente expressas, sendo que 95 aumentaram sua expressão e 44 diminuíram em resposta a condição de alta temperatura. Proteínas como a *Heat shock*, família da *DnaK* e chaperonas mostraram um grande aumento da expressão sugerindo que a tolerância a altas temperaturas é conseguida por uma estabilização das proteínas nas células de pólen, já o aumento da expressão de trealose sugere um papel na prevenção da desnaturação de proteínas através de dessecação.

Em outro trabalho Calderan-Rodrigues et al. (2016) estudando o proteoma da parede celular de colmos de cana-de-açúcar, conseguiram encontrar dez glicosídeo hidrolases indicadas como possíveis alvos para obtenção de paredes celulares menos recalcitrantes a desconstrução, gerando assim proteínas alvo que possam ser utilizadas em pesquisas futuras para facilitar a produção de etanol de segunda geração.

Além destes, existem atualmente vários trabalhos que utilizaram a proteômica *shotgun* em plantas de forma eficiente, como em soja (FERCHA et al., 2016), sorgo (HAN et al., 2016), amendoim (KOTTAPALLI et al., 2013), tomate (SHAH et al., 2012), *Arabidopsis thaliana* (MANN et al., 2013), cevada (KASPAR-SCHOENEFELD et al., 2016), soja (FERCHA et al., 2016), entre outras.

2.5. Biologia Sistêmica

Biologia sistêmica pode ser definida como o estudo de interações entre componentes biológicos utilizando modelos e/ou redes para integrar genes, metabólitos, proteínas, elementos reguladores e outros componentes biológicos (YUAN et al., 2008). Conforme Saito e Matsuda, (2010), a biologia sistêmica, baseado em dados de metabolômica, genômica e outras tecnologias moleculares *high-throughput*, tendem a desempenhar um papel fundamental na compreensão dos sistemas de plantas e desenvolvimento de aplicações biotecnológicas.

Com isso existe a necessidade imediata para a integração dos sistemas biológicos em plantas, considerando os grandes conjuntos de dados gerados a partir das diferentes tecnologias *ômicas* (YUAN et al., 2008).

Além disso, em termos de biologia sistêmica, os dados obtidos por metodologias *ômicas* sobre os padrões de acumulação de transcritos, proteínas e metabólitos, nos permite uma visão panorâmica sobre os fenômenos fisiológicos das plantas. Assim, a integração dessas técnicas abre o caminho para a compreensão da planta como um sistema complexo (REDESTIG et al., 2011). Para que isso aconteça, a bioinformática direcionada para a biologia sistêmica é um dos segmentos que precisa de mais atenção, pois é essencial que os dados biológicos sejam normalizados, padronizados e visualizados para construir modelos integrados (DI VENTURA et al., 2006).

Para a integração dos dados “ômicos”, existem dois tipos de abordagens principais, a integração intra-metabolômica, que tem por objetivo a integração de dados de metaboloma de diferentes plataformas (LC-MS, GC-MS, CE-MS, etc.) e integração “intra-ômicas” que visa a integração de dados de transcriptoma, proteoma, metaboloma, etc (REDESTIG et al., 2011).

A utilização do proteoma e metaboloma no estudo de algum sistema biológico, permite a visualização de modificações ocorridas nas células ou tecidos de uma forma holística, elevando nossa compreensão sobre as redes metabólicas e seus componentes ao nível de biologia sistêmica (GLINSKI; WECKWERTH, 2006). Em um de seus estudos, Weckwerth, (2008) utilizou com sucesso a estatística multivariada para integrar dados de proteômica e metabolômica em fisiologia molecular de plantas. Em um outro estudo conduzido por Yizhak et al. (2010) foram utilizados dados quantitativos de proteômica e metabolômica em uma abordagem integrativa, com isso os autores conseguiram uma predição correta no fluxo metabólico de *Escherichia coli* sob diferentes genes nocauteados. Já Delmotte et al. (2010) utilizaram a proteômica e transcriptômica na busca de novos conhecimentos sobre o metabolismo do *Bradyrhizobium japonicum* em nódulos de raiz de soja.

Mesmo com a necessidade de integração dos dados entre plataformas *ômicas*, nos últimos anos muitas ferramentas foram desenvolvidas para a análise da expressão gênica e dados de proteômica, mas até agora são poucas ferramentas que permitem ao usuário analisar os dados de metabolômica e integrá-los com dados de outras tecnologias *ômicas* (KARNOVSKY et al., 2011). Uma abordagem recente para a integração de dois conjuntos de dados é a análise de correlação canônica regularizada (rCCA), que destaca a correlação entre dois conjuntos de dados adquiridos nas mesmas unidades experimentais (GONZALEZ et al., 2008).

2.6. Integração dos dados pela Análise de Correlação Canônica Regularizada

A rCCA é uma modificação da clássica análise de correlação canônica (CCA) que permite a análise de conjuntos de dados com um pequeno número de observações e um grande número de variáveis em ambos conjuntos de dados (WALLER et al., 2015).

A CCA foi introduzida primeiramente por Hotelling (1936), ela é uma técnica da estatística multivariada que busca a maximização da correlação entre combinações lineares de variáveis (variáveis canônicas) nos conjuntos de dados (HONG et al., 2013). Os conjuntos de dados na CCA são representados como duas matrizes $X(n \times p)$ e $Y(n \times q)$, onde n é o número de observações e p e q são os números de variáveis em cada conjunto, em vez de analisar as semelhanças de pares de variáveis individuais, como na correlação de Pearson, essa análise busca por combinações lineares de todas as variáveis de X com uma correlação máxima nas combinações lineares com todas as variáveis em Y , onde estas combinações lineares são chamadas de variáveis canônicas (GUO et al., 2014; WAAIJENBORG; ZWINDERMAN, 2009).

O primeiro estágio na CCA consiste em encontrar dois vetores $a^1 = (a_1^1, \dots, a_p^1)^T$ e $b^1 = (b_1^1, \dots, b_q^1)^T$ que maximizam a correlação entre as combinações lineares $U^1 = Xa^1 = a_1^1X^1 + a_2^1X^2 + \dots + a_p^1X^p$ e $V^1 = Yb^1 = b_1^1Y^1 + b_2^1Y^2 + \dots + b_q^1Y^q$ em outras palavras o problema consiste em resolver $\rho_1 = \text{cor}(U^1, V^1) = \max \text{cor}_{a,b}(Xa, Yb)$, onde os resultados são as primeiras variáveis canônicas (U^1 e V^1) e a primeira correlação canônica (ρ_1) (GONZALEZ et al., 2008). Com isso, a matemática da CCA pode ser definida como:

$$P_X = X(X^T X)^{-1} X^T = \frac{1}{n} X S_{XX}^{-1} X^T \text{ e } P_Y = Y(Y^T Y)^{-1} Y^T = \frac{1}{n} Y S_{YY}^{-1} Y^T$$

Onde P são os projetores ortogonais sobre as extensões lineares das matrizes X e Y , e S_{xx} e S_{yy} as matrizes de covariância das amostra para os conjuntos de variáveis X e Y , respectivamente (GONZALEZ et al., 2008).

Para ser resolvido, a CCA necessita calcular a matriz de covariância inversa de X e Y , onde $p > n$ (LE CAO et al., 2009). Assim ao realizar uma CCA, os argumentos matemáticos obrigam os conjuntos de dados terem mais unidades do que variáveis em cada conjunto ($p > n$ e $q > n$). Hoje em dia isso nem sempre é possível, principalmente devido as técnicas biológicas *high throughput*, onde normalmente tem-se poucas unidades (amostras) e um grande número de variáveis (genes, proteínas, metabólitos, etc.) (GONZALEZ et al., 2008). Devido a esse fato, quando se deseja realçar correlações em dois conjuntos de dados nos quais foram utilizadas metodologias *ômicas*, a clássica CCA não pode ser utilizada. Uma solução para essa questão é incluir uma regularização no processamento dos dados (BICKEL; LI, 2006), podendo assim ser feito uma rCCA (GONZALEZ et al., 2008). Com isso Vinod (1976) e Gonzalez et al. (2008) introduziram uma penalidade às matrizes de covariância, de

modo a torná-las invertidas em um rCCA (LÊ CAO; GONZÁLEZ; DÉJEAN, 2009). Assim a regularização consiste na substituição das matrizes de covariância S_{xx} e S_{yy} por $S_{xx} + \lambda_1 I_p$ e $S_{yy} + \lambda_2 I_q$, respectivamente (GONZALEZ et al., 2008).

Nos últimos anos a rCCA tem sido utilizada com sucesso para verificar a correlação entre dois conjuntos de dados através de redes de correlações. Wallace et al. (2014) estudando a relação entre o lipidoma, marcadores inflamatórios e resistência à insulina em humanos, utilizaram redes de relevância, a partir da matriz de similaridade derivada da rCCA, para observar de forma holística o lipidoma correlacionado com parâmetros bioquímicos e o lipidoma correlacionado com lipoproteínas.

Em outro estudo Morine et al. (2010), estudando vias sensíveis a dieta e biomarcadores da síndrome metabólica, que aumenta o risco de diabetes e doenças cardiovasculares, usaram a análise de correlação canônica para dados de expressão hepática, para avaliar padrões de correlação entre a expressão do gene e os marcadores plasmáticos da síndrome metabólica, onde observaram vários genes correlacionados com vários marcadores fenotípicos de saúde metabólica.

Patel et al. (2010) estudando a análise de proteínas de membrana em metagenômica, usaram a rCCA entre as características ambientais e famílias de proteínas de membrana. Com isso eles puderam observar concordância das características ambientais e famílias de proteínas de membrana, onde esta análise é justificada e pode ser capaz de revelar as características correlacionadas mais específicas.

3. MATERIAL E MÉTODOS

3.1. Material vegetal e condições de cultivo

Para a condução do experimento foi utilizada a variedade de cana-de-açúcar SP80-3280, pelo fato de seu genoma estar sendo sequenciado e ser uma das variedades utilizada no projeto SUCEST (VETTORE et al., 2003). As plantas foram cultivadas no campo, na Universidade Federal de São Carlos, Campus de Araras, sob a supervisão da Profa. Dra. Monalisa Sampaio Carneiro, o desenho experimental pode ser observado na **figura 2**. O solo da área experimental era um latossolo vermelho eutroférico, que foi adubado com 400 kg/ha de NPK (5-25-25) e irrigado para formação do *stand*.

Figura 2: Desenho experimental de plantio da variedade SP80-3280 de cana-de-açúcar. C = número da coleta; R = número da repetição da parcela; Linha contínua = fileiras de plantio.

Cada parcela do experimento possuía 80 touceiras, distribuídas em quatro fileiras simples de plantio, com 10 m de comprimento e com espaçamento de 1,35 m entre os sulcos, os quais produziram cerca de 400 colmos por parcela. O conjunto de parcelas foi cercado por duas fileiras simples de cana de bordadura, para que o experimento refletisse o que acontece em um campo de plantio. Os toletes de cana foram plantados no dia 20 de abril de 2012 e as coletas das amostras da folha +1, entrenó 1, 5 e 9 foram feitos conforme a **tabela 1**. O entrenó 1 foi coletado junto com o entrenó 2, pois seu tamanho reduzido impedia a separação dos mesmos (**figura 3**). Após a coleta, as amostras foram imediatamente congeladas com nitrogênio líquido e mantidas a -80 °C até sua utilização.

Tabela 1: Tempo de plantio e número de amostras coletadas de cada tecido.

Tempo de plantio	Folha + 1	Entrenó 1 e 2	Entrenó 5	Entrenó 9
4 meses	12	-	-	-
8 meses	12	12	-	-
11 meses	12	12	12	12
13 meses	12	12	12	12

Figura 3: entrenós 1 e 2.

3.2. Análise do Metabolôma

3.2.1. Teste de solvente de extração

Para um teste preliminar do melhor método para extração dos metabólitos, foram extraídos metabólitos da folha +1 de seis repetições biológicas da primeira coleta, utilizando dois diferentes solventes de extração, e para cada repetição biológica foram feitas três repetições técnicas. As extrações foram realizadas a partir de 100 mg de tecido pulverizados e congelados. As extrações de metabólitos das amostras foram feitas com 300 µl do solvente de extração FF (99,875% (v/v) de metanol acidificada com 0,125% (v/v) de ácido fórmico) seguindo o protocolo utilizado por De Vos et al. (2007) e também com o solvente de extração FC (60% (v/v) de metanol, 20% (v/v) de clorofórmio e 20% (v/v) de água) seguindo o protocolo utilizado por Gullberg et al. (2004). Após a extração, as amostras foram filtradas com filtros de 0,22 µm.

3.2.2. Teste de padrões internos

Para a verificação da qualidade de extração e normalização da intensidade dos prováveis metabólitos da metabolômica, foram selecionados compostos que possivelmente possuíssem uma boa ionização e se mostrassem estáveis durante as corridas cromatográficas. Também foram selecionados alguns compostos com base na literatura

recente. Os compostos utilizados foram a alanina (massa molecular = 89,08 g.mol⁻¹), ácido 4-aminobenzoico (massa molecular = 137,14 g.mol⁻¹), ácido cafeico (massa molecular = 180,16 g.mol⁻¹), cisteína (massa molecular = 121,16 g.mol⁻¹), leucina (massa molecular = 131,17 g.mol⁻¹), ácido oleico (massa molecular = 282,46 g.mol⁻¹) e a quercetina (massa molecular = 302,23 g.mol⁻¹). Todos esses compostos foram testados em diferentes concentrações afim de encontrar a melhor ionização.

3.2.3. Extração de metabólitos e análise por LC-MS

Todos os tecidos das amostras de cana-de-açúcar foram pulverizados com auxílio de um almofariz e pistilo utilizando nitrogênio líquido para manter as amostras congeladas. Logo após a pulverização as amostras foram mantidas a -80 °C até a etapa de extração de metabólitos.

Após a seleção do melhor solvente de extração, foram utilizadas um total de seis repetições biológicas para cada tecido (folha +1, entrenós 1, 5 e 9) e tempo de coleta (4, 8, 11 e 13 meses após o plantio). As extrações foram realizadas a partir de 100 mg de tecido pulverizado e congelado. A extração de metabólitos das amostras foi feita com 300 µl do solvente de extração FF (99,875% (v/v) de metanol acidificada com 0,125% (v/v) de ácido fórmico) com a adição de 2,5 µmol.L⁻¹ de quercetina, seguindo o protocolo utilizado por De Vos et al. (2007). Após a extração, as amostras foram filtradas com filtros de 0,22 µm.

A separação cromatográfica por UPLC Acquity (Waters) foi realizada como descrito por De Vos et al. (2007). Injetou-se 5 µL de amostra em coluna de fase reversa (100 x 2.1 mm Acquity 1.7 µm C18). Dois eluentes foram utilizados como fase móvel: A (água ultra pura contendo 0,1% de ácido fórmico) e B (acetonitrila ultra pura contendo 0,1% ácido fórmico). O gradiente da fase móvel utilizado foi: 3 minutos 95% A + 5% B, 3 minutos 75% A + 25% B, 3 minutos 5% A + 95% B e 3 minutos 95% A + 5% B.

Para a identificação das massas dos metabólitos, foi utilizado um espectrômetro de massas QTOF Ultima API (Waters Technologies, UK), a aquisição da massa/carga (*m/z*) dos compostos foi feita em modo positivo com tensão do capilar de 3.000 V, tensão do cone de 35 V, temperatura da fonte de 150 °C, temperatura de dessolvatação de 450 °C, fluxo de nitrogênio no cone de 50 L/h e fluxo de nitrogênio na dessolvatação de 550 L/h.

3.2.4. Análise do perfil metabólico e estatística dos dados de LC-MS

A aquisição dos dados foi realizada pelo software MassLynx 4.1 (Waters Technologies, UK) e o processamento dos dados foi realizado por meio do software XCMS (SMITH et al., 2006) para a linguagem e ambiente de programação estatística R (<http://www.R-project.org>). Os parâmetros utilizados para a identificação dos marcadores foram o método centWave,

ppm=10, peakwidth de 5-20, snthresh igual a 5, prefilter de 3-50, método obiwarp para correção do tempo de retenção (rt) e a análise dos grupos de marcadores com bw igual a 5, minfrac igual a 0.6 e mzwid igual a 0.1.

Foram utilizadas duas normalizações para os marcadores obtidos, uma correção para a massa obtida durante a pesagem das amostras e a normalização da intensidade de todos os prováveis metabólitos relativos ao padrão interno. A análise do perfil metabólico e as normalizações dos dados foram feitas a partir de scripts para a linguagem R (<http://www.R-project.org>) (**ANEXO**).

A análise estatística dos dados foi realizada utilizando-se a análise de componentes principais (PCA, Principal Component Analysis) e a análise por PLS-DA (Partial Least Squares Discriminant Analysis) para poder mostrar as variações entre os diferentes tecidos juntamente com os tempos de coletas utilizados. Para isso, foi utilizado o *web server* MetaboAnalyst (XIA et al., 2009), na PCA foi utilizada a transformação logarítmica das variáveis e o escalonamento por pareto, que utiliza a raiz quadrada do desvio padrão para fazer o escalonamento de cada variável encontrada na PCA (VAN DEN BERG et al., 2006). Já na PLS-DA foram observados os valores de R^2 e Q^2 , onde o R^2 indica a variação descrita pelo modelo e Q^2 a acurácia, juntos esses parâmetros mostram a habilidade preditiva do modelo (SALEK et al., 2008), quanto mais próximo do número um, melhores os valores para R^2 e Q^2 . Além destes parâmetros, também foram observadas as variáveis mais importantes na projeção dos componentes da PLS-DA (VIPs).

A PCA é uma análise estatística multivariada não supervisionada, onde nenhuma classe é atribuída às amostras a priori. Ela permite uma visualização de qualquer padrão presente nos dados e também possibilita a identificação de *outliers* (SENGUPTA et al., 2011). Já a PLS-DA é uma análise estatística multivariada supervisionada que usa a regressão linear múltipla para encontrar a direção da máxima covariância entre um conjunto de dados e os membros das classes (XIA et al., 2009). Ambas metodologias, PCA e PLS-DA, tentam simplificar visualmente os resultados, diminuindo o número de variáveis usando suas médias ponderadas (XIA et al., 2009)

3.2.5. Identificação dos metabólitos por MS/MS

Os principais VIPs de cada análise realizada (folha +1, entrenó 1, entrenó 5 e entrenó 9) foram selecionados para a fragmentação e identificação por MS/MS, dando prioridade para os cinco primeiros prováveis metabólitos. A fragmentação por MS/MS auxilia na elucidação estrutural dos possíveis metabólitos, pois uma anotação confiável dos picos cromatográficos e uma identificação exata dos metabólitos ainda é o maior desafio na metabolômica *untargeted* (DUNN et al., 2013).

Foram utilizadas as mesmas configurações da corrida cromatográfica e ionização como já descrito, seguido de fragmentação com energias de colisão de 20 eV, 30 eV e 50 eV. Quando os resultados da fragmentação não foram satisfatórios, foram utilizadas energias de colisão intermediárias a esses valores, como 28 eV, 35 eV e 40 eV.

A busca dos possíveis metabólitos foi feita utilizando-se o m/z do precursor no banco de dados do HMDB 3.6 (WISHART et al., 2013), utilizando os possíveis adutos do precursor com H^+ , Na^+ ou K^+ e um erro de massa de até 100 ppm. Em seguida foram importadas as estruturas das moléculas encontradas no banco de dados do HMDB (WISHART et al., 2013) para a comparação do espectro de fragmentação dos VIPs selecionados com os metabólitos candidatos fragmentados *in silico* utilizando-se a ferramenta MetFrag 2.2 CL (WOLF et al., 2010). A anotação dos possíveis metabólitos foi aceita ou refutada de acordo com suas similaridades, levando em conta o *score* da anotação do MetFrag (WOLF et al., 2010). Como a fragmentação *in silico* também é passível de erros, optou-se pela discussão das classes dos metabólitos anotados. Assim esta anotação é considerada de nível 3, onde são identificadas as classes por semelhança espectral com compostos conhecidos de uma classe química (SUMNER et al., 2007). Os gráficos de alinhamento dos espectros foram construídos por meio de *scripts* para a linguagem de programação estatística R (<http://www.R-project.org>), com base nos resultados fornecidos pelo MetFrag 2.2 CL (WOLF et al., 2010) (**ANEXO**).

3.3. Análise do Proteoma

3.3.1. Extração das proteínas

Para a análise do proteoma foram utilizadas amostras de cana-de-açúcar referentes à folha +1, entrenó 5 e entrenó 9, nos tempos de coleta de 11 e 13 meses após o plantio. Não foi possível realizar a proteômica para todos os tempos de coleta, devido ao alto custo e tempo necessário para a análise. No total foram utilizadas seis repetições biológicas de cada tecido e período de coleta. Foi feito um *bulk* de duas amostras em um total de 200 mg, gerando três repetições para cada tecido e período de coleta.

Todos os tecidos das amostras foram pulverizados com auxílio de um almofariz e pistilo utilizando nitrogênio líquido para manter as amostras congeladas. Logo após a pulverização as amostras foram mantidas a -80 °C até a etapa de extração das proteínas.

Para a extração das proteínas foi adicionado 0,8 mL de tampão de extração de proteínas (0,5 M Tris-HCl pH 7,5; 0,7 M sacarose; 0,1 M cloreto de potássio; 50 mM EDTA; 1 mM PMSF; 2% (v/v) β -mercaptoetanol e 1% (m/v) PVPP), seguido de agitação em gelo, por 30 minutos. Após, foi adicionado 0,8 mL de solução fenólica saturada em Tris-HCl pH 7,5 sob agitação em gelo por mais 30 minutos, seguido de centrifugação por 30 minutos a 10.000 g e 4 °C. Recuperou-se o sobrenadante, adicionando-se 0,8 mL de tampão de extração de

proteínas, seguido de agitação em gelo por 30 minutos e centrifugação a 10.000 g e 4 °C, esta etapa foi repetida mais duas vezes.

Após a última centrifugação, o sobrenadante foi recuperado e adicionado 1,5 mL de acetato de amônio 0,1 M em metanol 100%, e mantidos a -20 °C *overnight*, para precipitação das proteínas. Em seguida as amostras foram submetidas a centrifugação por 20 minutos a 16.000 g e 4 °C. Terminada esta etapa, foi lavado o *pellet*, duas vezes com acetona 100% a -20 °C e uma com metanol 100% a -20 °C. Em todas as etapas de lavagem as amostras eram mantidas por 2 horas a 4 °C seguidos de centrifugação por 30 minutos a 16.000 g e 4 °C e descarte do sobrenadante. Após, os *pellets* foram secos em dissecador, a 4 °C, por 24 horas, e as proteínas totais foram ressuspensas em tampão de solubilização (7 M ureia, 2 M tiureia, 10 mM DTT e 0,4% v/v Triton X-100).

3.3.2. Dessalinização das amostras

Para dessalinizar as amostras foram utilizados filtros Amicom 3 KDa (Millipore), substituindo-se o tampão de solubilização por tampão de bicarbonato de amônio 50 mM com pH 8,5 (AmBic), seguindo o protocolo do fabricante. Primeiramente a coluna foi hidratada com água deionizada e centrifugada por 10 minutos a 10.000 g e 15 °C. Após isso, as amostras contendo as proteínas totais foram adicionadas à coluna e adicionado 300 µL de tampão AmBic, seguido de centrifugação a 14000 g e 15 °C até atingir o volume de 100 µL. Esta etapa foi repetida 3 vezes. Após, as amostras foram recolhidas e armazenadas a -80 °C.

3.3.3. Quantificação das proteínas

A quantificação das proteínas totais foi realizada pelo método de Bradford (BRADFORD, 1976), utilizando-se um kit comercial de quantificação de proteínas (Bio-Rad, Califórnia, USA). Após, foi verificada a absorbância das amostras no comprimento de onda de 595 nm em espectrofotômetro U-3300 (Hitachi).

3.3.4. Verificação da qualidade da extração de proteínas

Para verificar a qualidade da extração de proteínas foi feita uma eletroforese das amostras em gel de poliacrilamida 12% (LAEMMLI, 1970). Para isso foram utilizados 5 µg de proteína total juntamente com a quantidade suficiente de tampão de desnaturação (0,5 M Tris-HCl pH 6,8, 10% de glicerol, 10% de SDS, 5% de β-mercaptoetanol e 1% azul de bromofenol) para completar 15 µL. As amostras foram mantidas a 100 °C por 60 segundos e aplicadas nos géis com tensão de 100 volts. Foi utilizado 5 µg de BSA (*Bovine Serum Albumin*) como padrão para checar a qualidade da quantificação.

Após a eletroforese, os géis foram banhados em uma solução fixadora (40% (v/v) etanol e 10% (v/v) ácido acético) por uma hora, para imobilização das proteínas nos géis. Após a fixação, os géis foram banhados em solução corante de *Comassie Blue G250*, por 18 horas, seguindo o protocolo de Candiano et al. (2004), e posteriormente lavados em água deionizada.

3.3.5. Digestão das amostras de proteínas totais

Foram utilizados 50 µg de proteína total de cada amostra, primeiramente as amostras foram desnaturadas com 0,2% de RapiGest SF em água (RapiGest SF, Waters, Illinois, EUA) e mantidas por 15 minutos a 80 °C. Em seguida, as amostras foram centrifugadas (pulso), e 2,5 µL de ditioneitol (DTT) 100 mM foram adicionados. As amostras foram aquecidas a 60 °C por 30 minutos, resfriadas à temperatura ambiente e novamente centrifugadas (pulso). Em seguida, adicionou-se 2,5 µL de iodoacetamida 300 mM. Após, as amostras foram mantidas por 30 minutos, à temperatura ambiente, protegidas da luz. A digestão foi realizada por meio da adição de 10 µL de uma solução de tripsina em 50 mM de bicarbonato de amônio, a uma razão de 1:100 enzima/proteína e a amostra foi mantida a 37 °C por 16 horas. Após a digestão, realizou-se a hidrólise do RapiGest SF, por meio da adição de 10 µL de ácido trifluoroacético 5%. As amostras foram mantidas a 37 °C por 90 minutos e na sequência, foram centrifugadas a 14.000 x g, por 30 min a 6 °C. O sobrenadante foi transferido para frascos de vidro e liofilizado.

3.3.6. Purificação das amostras de proteínas

Para a purificação das amostras foi adicionada 50 µL de solução 0,1% TFA, em seguida cada amostra foi fracionada em 5 tubos com 10 µL cada. Após as amostras foram purificadas utilizando micro colunas de fase reversa (*Reverse phase Zip-Tip C18, Millipore*). Para isso, as colunas foram equilibradas em solução A (100% acetonitrila (ACN), 0,1% (v/v) TFA), solução B (50% (v/v) ACN:H₂O, 0,1% (v/v) TFA) e solução C (100% H₂O 0,1% (v/v) TFA), respectivamente. Após o equilíbrio, as amostras de proteínas digeridas foram aplicadas à coluna, seguindo-se lavagem com solução D (95% H₂O, 5% (v/v) metanol, 0,1% (v/v) TFA). Para a eluição dos peptídeos foi utilizado 10 µL de solução B juntando-se novamente as alíquotas fracionadas inicialmente, seguidas de concentração a vácuo.

3.3.7. LC-MSE

As amostras foram ressuspensas em 32 µL de Formiato de amônio 20 mM pH10 e 8 µL de *Phosphorylase B* (SwissProt P00489, padrão interno - 100 fmo.µL⁻¹). Os peptídeos foram sequenciados no espectrômetro de massas Synapt G2 HDMS (Waters, Manchester,

UK), acoplado ao sistema Acquity UPLC MClass, com tecnologia 2D (Waters, Manchester, UK). Os peptídeos foram capturados por uma coluna Trap 2D Symmetry C18 (5 μm , 180 μm x 20 mm) (Waters, Manchester, UK). A separação foi realizada com uma coluna de primeira dimensão Acquity UPLC M-Class peptide BEH C18 (5 μm , 300 μm x 50 mm) (Waters, Manchester, UK) e uma coluna analítica Acquity UPLC M-Class peptide CSH C18 (1.7 μm , 75 μm x 150 mm) (Waters, Manchester, UK). As fases móveis da primeira dimensão são: solução de formiato de amônio 20 mM pH 10 e acetonitrila. Já na segunda dimensão os dois eluentes foram: A (100% água contendo 0,1% de ácido fórmico) e B (100% acetonitrila contendo 0,1% ácido fórmico). Essa técnica combina a separação das proteínas usando-se um pH de 10, na separação da primeira dimensão e um pH de 2 na segunda dimensão, com auxílio de colunas de fase reversa (GILAR et al., 2005). A eluição foi realizada usando-se 5 gradientes binários diferentes de 20 mM de formiato de amônio pH 10 e acetonitrila, a um fluxo de 2 $\mu\text{L} \cdot \text{min}^{-1}$. A proporção de acetonitrila da primeira fração até a quinta variou de 11,4 a 50%. Os peptídeos eluídos da coluna da primeira dimensão foram diluídos em linha com 0,1% de ácido fórmico em acetonitrila a um fluxo de 20 $\mu\text{L} \cdot \text{min}^{-1}$ e capturados por uma coluna Trap 2D Symmetry. A separação, na segunda dimensão, foi realizada com um gradiente binário de 7 a 85% de acetonitrila, em 0,1% (v:v) de ácido fórmico durante 75 minutos, a um fluxo de 500 nL $\cdot \text{min}^{-1}$.

A aquisição dos dados foi realizada em um espectrômetro de massas do tipo quadrupolo-tempo de voo (Q-TOF) Synapt G2 MS, equipado com uma fonte *nanolockspray* operando no modo positivo (Waters, Manchester, UK), com *ion mobility*. Para todas as análises, o espectrômetro de massas operou no modo “V”, com a cela de *ion mobility* ativada. O espectrômetro de massas foi calibrado com (Glu1) fibrinopeptídeo B (GFP) humana 1 pmol $\cdot \mu\text{L}^{-1}$ e a mesma solução foi empregada para o *lock mass* utilizando-se o íon de dupla carga com uma amostragem a cada 30 segundos. Os experimentos foram realizados no modo HDMS^E (análise independente de dados) que consiste na aquisição alternada, entre espectros obtidos à baixa (3 eV) e alta (15-50 eV) energia de colisão, aplicadas ao módulo *trap* do *T-wave CID*, em presença de gás argônio, que produz íons precursores e produtos em sequência. O tempo de varredura foi de 0,8 segundos em cada modo, no intervalo de *m/z* entre 50 e 2000.

3.3.8. Análise dos dados

Os espectros de massas foram processados utilizando o programa *ProteinLynx GlobalServer* (PLGS) versão 3.1 (Waters, Manchester, UK), com o banco de dados de ESTs de cana-de-açúcar do SUCEST (VETTORE et al., 2001), contando com 237.954 ESTs que originaram 43.141 *clusters*. Os parâmetros de processamento incluíram: tolerância

automática para precursores e íons-produto; mínimo de três íons-fragmento correspondentes por peptídeo; mínimo de 7 íons-fragmento correspondentes por proteína; mínimo de 2 peptídeos correspondentes por proteína; um possível erro de clivagem pela tripsina; carbamidometilação de cisteína como modificação fixa e oxidação de metionina como modificação variável; taxa máxima de descoberta de falso-positivos (FDR) a 4%.

Para identificar e quantificar as proteínas, as intensidades do espectro foram calculadas pelo método estequiométrico, de acordo com o padrão interno, por análise de MS^E e normalizado pela função de “auto-normalização” do próprio software. A quantidade e sequência dos peptídeos correspondentes e femtomols (fmols) foram obtidos baseados na média dos três peptídeos mais abundantes, determinado em cada experimento individual (SILVA et al., 2006), considerando os dados de 3 repetições biológicas por amostra. A taxa de descoberta de falso-positivos para identificação de peptídeos e proteínas foi determinada baseada na busca em um banco de dados reverso, gerado automaticamente pelo programa PLGS 3.1 (Waters, Manchester, UK). Todas as proteínas encontradas foram identificadas com uma confiança acima de 95%. Somente as proteínas identificadas em pelo menos duas das três repetições biológicas foram consideradas para as análises posteriores.

As proteínas que foram apenas identificadas, mas não quantificadas em todas as amostras, foram excluídas para simplificar o processo de análise. Posteriormente, as proteínas foram anotadas utilizando-se o software Blast2GO 3.2.7 (CONESA et al., 2005) com a atualização do banco de dados de fevereiro de 2016. Para o *blastx* (ALTSCHUL et al., 1990) foi utilizando o banco *nr*, *E-Value* de 1.0E-5, número de *hits* igual a 20 e *word size* igual a 6. Na anotação foi utilizado o *E-Value-Hit-Filter* igual a 1.0E-6, *cutOff* de 55 e *GO weight* de 5. Todos os demais parâmetros foram utilizados os *defaults* do software.

Para a análise estatística dos dados foi utilizada a PCA e a PLS-DA para poder mostrar as variações entre os diferentes tecidos juntamente com os tempos de coletas utilizados. Também foram observadas as variáveis mais importantes na projeção dos componentes da PLS-DA (VIPs). Toda a análise estatística seguiu a metodologia descrita para a análise dos perfis metabólicos. Todos os gráficos, tabelas e arquivos necessários para a análise estatística foram construídas por meio de scripts para a linguagem de programação estatística R (<http://www.R-project.org>) (**ANEXO**).

3.4. Integração dos dados de metabolômica e proteômica

3.4.1. Normalização dos dados

Como a integração dos dados foi realizada utilizando-se dois conjuntos de dados distintos, com amplitudes diferentes, realizou-se a normalização desses conjuntos para se fazer a integração. A normalização adotada ($x_{normalizado}$) foi obtida dividindo-se a intensidade do

metabólito ou proteína ($x_{\text{metabólito ou proteína}}$) pela intensidade do padrão interno ($x_{\text{padrão interno}}$) naquela amostra (REDESTIG et al., 2011).

$$x_{\text{normalizado}} = \frac{x_{\text{metabólito ou proteína}}}{x_{\text{padrão interno}}}$$

3.4.2. Integração por rCCA

Para verificar a correlação entre os conjuntos de dados de metabólitos (X) e proteínas (Y), é necessário o mesmo número de amostras. Como cada repetição biológica da proteômica foi composta por um *bulk* de duas plantas e para a metabolômica cada repetição biológica foi composta por apenas uma planta, foi feito a média das duas repetições biológicas da metabolômica correspondentes as mesmas amostras utilizadas para a proteômica. Portanto cada conjunto de dados contém as mesmas três repetições biológicas, onde cada repetição é composta por um *bulk* de duas plantas.

A integração dos dados foi feita utilizando-se a análise de rCCA (GONZALEZ et al., 2008), por meio do pacote mixOmics 5.2 (LÊ CAO; GONZÁLEZ; DÉJEAN, 2009), desenvolvido para a linguagem e ambiente de programação estatística R (<http://www.R-project.org>).

A metodologia de rCCA requer a escolha ou ajuste dos parâmetros λ_1 e λ_2 , que são os parâmetro de regularização da rCCA, que substituirão as matrizes de covariância S_{xx} e S_{yy} da CCA por $S_{xx} + \lambda_1 I_p$ e $S_{yy} + \lambda_2 I_q$, respectivamente (GONZALEZ et al., 2008). Para ambos os parâmetros foi escolhido um conjunto de valores positivos para avaliar o *score* da validação cruzada (*cv-score*) em cada ponto da rede, conseguindo-se assim um valor ótimo para λ_1 e λ_2 que forneceu o maior *cv-score*. Essa etapa foi feita com a função *tune.rcc()* do pacote mixOmics (LÊ CAO; GONZÁLEZ; DÉJEAN, 2009). Após escolhido os melhores valores dos parâmetros para o rCCA, foi utilizada a função *rcc()*, que faz a rCCA entre os dois conjuntos de dados (metabolômica e proteômica).

Para a visualização das correlações canônicas obtidas, foi utilizada a função *network()* com um *threshold* de 0,9. Esse *threshold*, está acima dos geralmente utilizados (GONZÁLEZ et al., 2012; PIWOWAR; JURKOWSKI, 2015; WALLACE et al., 2014), assim as redes biológicas conterão somente informações altamente correlacionadas (negativamente e positivamente). O *threshold* utilizado foi escolhido por proporcionar redes biológicas não muito densas e por conter somente altas correlações. De acordo com Wallace et al. (2014) o valor de *threshold* para rCCA é arbitrariamente escolhido na forma de obter redes biológicas interpretáveis, que não sejam nem muito densa, nem muito esparsa.

Para a melhor visualização das redes de regulação, elas foram exportadas para o

4. RESULTADOS E DISCUSSÃO

4.1. Coleta das amostras

O aspecto das canas durante os períodos de coleta pode ser observado na **figura 5**. As medições fisiológicas das plantas foram feitas pelo grupo de pesquisa da profa. Glaucia Mendes Souza. **A figura 6** mostra a média de temperatura na cidade de Araras/SP, onde foi realizado o experimento. De acordo com Lembke et al. (2015) esse experimento mostrou um aumento significativo do grau brix da cana-de-açúcar durante os períodos de coleta de 8, 11 e 13 meses após o plantio. Aos 4 meses após o plantio, as plantas ainda só possuíam folhas, não permitindo a medição do grau brix.

Figura 5: aspecto visual das canas nos períodos de coleta. A = 4 meses de plantio; B = 8 meses de plantio; C = 11 meses de plantio; D = 13 meses de plantio.

Figura 6: média da temperatura na cidade de Araras/SP durante o plantio e desenvolvimento da cana-de-açúcar. Setas azuis indicam as coletas realizadas.

4.2. Análise do perfil metabólico

4.2.1. Teste de solventes de extração

Visando identificar o melhor solvente de extração de metabólitos, foi feita uma PCA com as 36 corridas do LC-MS (seis repetições biológicas, três repetições técnicas e dois diferentes solventes de extração) provenientes da folha +1 da 1ª coleta (cana com 4 meses de desenvolvimento). Esta primeira análise conseguiu identificar 114 prováveis metabólitos e a PCA resultante (**figura 7**) mostra uma tendência de separação entre as amostras de acordo com o solvente de extração utilizado (solvente FF e FC). A soma dos dois componentes principais possui 63% da variância das amostras, deixando clara a diferença no perfil metabólico, de acordo com o solvente utilizado na extração.

Figura 7: PCA das amostras da folha +1 da variedade SP80-3280 de cana-de-açúcar com 4 meses após o plantio, mostrando uma tendência de separação de acordo com o solvente de extração de metabólitos utilizado (FF e FC).

Como existe uma diferença no perfil metabólico decorrente do solvente utilizado, foi feita uma análise individual para cada solvente de extração, para visualizarmos o comportamento das repetições biológicas e técnicas dentro de cada extração realizada. Na análise contendo somente as amostras do solvente FC foram identificados 63 prováveis metabólitos e a PCA resultante dessa análise (**figura 8**) mostra uma grande dispersão das

amostras. A soma dos dois componentes principais possui 78% da variância entre as amostras.

Figura 8: PCA das amostras da folha +1 da variedade SP80-3280 de cana-de-açúcar com 4 meses após o plantio, utilizando o solvente de extração de metabólitos FC.

Já na análise das amostras utilizando o solvente FF foram identificados 104 prováveis metabólitos e a soma dos dois componentes principais de sua PCA possui 67% da variância entre as amostras (**figura 9**). Apesar da dispersão observada na PCA, as amostras se mostraram mais agrupadas do que aquelas extraídas com o solvente FC e o número de prováveis metabólitos identificados foi 65% maior, utilizando-se os mesmos padrões de análise. Assim o solvente de extração FF mostrou-se mais adequado para o nosso tipo de amostra.

Figura 9: PCA das amostras da folha +1 da variedade SP80-3280 de cana-de-açúcar com 4 meses após o plantio, utilizando o solvente de extração de metabólitos FF.

4.2.2. Teste de padrões internos

Após a análise das amostras da folha +1, foram avaliados alguns compostos que poderiam ser utilizados como padrões internos para melhorar a qualidade da análise dos tecidos da cana-de-açúcar. Os compostos utilizados foram a alanina (massa molecular = 89,08 g.mol⁻¹), ácido 4-aminobenzoico (massa molecular = 137,14 g.mol⁻¹), ácido cafeico (massa molecular = 180,16 g.mol⁻¹), cisteína (massa molecular = 121,16 g.mol⁻¹), leucina (massa molecular = 131,17 g.mol⁻¹), ácido oleico (massa molecular = 282,46 g.mol⁻¹) e a quercetina (massa molecular = 302,23 g.mol⁻¹). A **figura 10** mostra os cromatogramas desses compostos nas concentrações de 1,12E-05 mol.ml⁻¹, 6,29E-06 mol.ml⁻¹, 5,55E-06 mol.ml⁻¹, 8,26E-06 mol.ml⁻¹, 7,62E-06 mol.ml⁻¹, 3,54E-06 mol.ml⁻¹ e 3,31E-06 mol.ml⁻¹ respectivamente, também é apresentado o cromatograma de uma amostra de água.

Figura 10: Cromatograma dos compostos: A = água, B = alanina, C = ácido 4-aminobenzoico, D = ácido cafeico, E = cisteína, F = leucina, G = ácido oleico e a H = quercetina.

Após essa primeira análise, comparando os possíveis padrões, foi feita uma diluição em série de todos os compostos, pois com a primeira concentração utilizada, todos os padrões tiveram seus sinais saturados, impedindo uma identificação precisa no espectrômetro de massas. Pela qualidade da identificação do composto e pela precisão da m/z observada, o composto que mais se destacou foi a quercetina na concentração de $33,2 \text{ pmol.ml}^{-1}$, podendo ser utilizado como um padrão interno para as nossas amostras. **A figura 11** mostra a quercetina na concentração de $33,2 \text{ pmol.ml}^{-1}$ juntamente com seu espectro com uma contagem de aproximadamente 1050 íons.

4.2.3. Perfil metabólico da folha +1

No perfil metabólico da folha +1 foram identificados 662 diferentes prováveis metabólitos. Na **figura 13** é apresentado a PCA do perfil metabólico das amostras da folha +1 com 4, 8, 11 e 13 meses após o plantio. A soma dos dois primeiros componentes principais foi de 53,4%. É possível observar uma maior diferenciação das amostras referente à coleta de 4 meses em relação às demais coletas. Já na PLS-DA na **figura 14** é possível observar que além das amostras referentes à coleta de 4 meses, as amostras com 13 meses após o plantio também se destacaram das demais coletas, isso mostra uma diferença no perfil metabólico relacionado entre os tempos de coleta. Os valores apresentados na PLS-DA são de $R^2 = 0,95$ e $Q^2 = 0,91$ nos três primeiros componentes, garantindo assim uma variância e acurácia do modelo confiáveis.

Devido ao fato da sacarose ser produzida nas folhas, depois transportada pelo floema e armazenada nos entrenós (MOORE, 1995), e também ao fato da planta acumular altas concentrações de sacarose em entrenós maduros e uma menor concentração em entrenós jovens (MCCORMICK; CRAMER; WATT, 2011), essa maior diferença no perfil metabólico entre as idades avaliadas pode estar relacionada ao fato da planta ser muito jovem aos 4 meses. Como indicado anteriormente, nessa idade a planta ainda não tinha entrenós formados, com isso também não pode ser mensurado o grau brix para este período de coleta.

Figura 13: PCA das amostras da folha +1. C1F1 = 4 meses (vermelho), C2F1 = 8 meses (verde), C3F1 = 11 meses (azul escuro) e C4F1 = 13 meses após o plantio (azul claro).

Figura 14: PLS-DA 2D (**A**) e 3D (**B**) das amostras da folha +1. C1F1 = 4 meses (vermelho), C2F1 = 8 meses (verde), C3F1 = 11 meses (azul escuro) e C4F1 = 13 meses após o plantio (azul claro).

É possível observar na **figura 15** os 15 principais prováveis metabólitos que influenciaram na projeção do primeiro componente da PLS-DA (VIPs) e seus níveis de intensidade em cada período de coleta. Também são apresentadas as classes dos compostos anotados por MS/MS, que foram: glicerolipídeos, glicerofosfolipídeos, prenol lipídeos, carboidratos e carboidratos conjugados, esteroides e derivados de esteroides.

Figura 15: VIPs da PLS-DA da folha +1 durante os tempos de coleta. C1F1 = 4 meses, C2F1 = 8 meses, C3F1 = 11 meses e C4F1 = 13 meses após o plantio.

4.2.4. Perfil metabólico do entrenó 1

No perfil metabólico do entrenó +1 de cana-de-açúcar foram identificados 471 diferentes prováveis metabólitos. Na **figura 16** é apresentado a PCA do perfil metabólico do entrenó 1, que foi coletado juntamente com o entrenó 2 devido ao seu tamanho reduzido, com 8, 11 e 13 meses após o plantio. A soma dos dois primeiros componentes foi de 78.8% da variância.

Foi possível observar uma maior diferença no perfil metabólico entre as amostras de 13 meses e os demais períodos de coleta. O entrenó 1 é considerado imaturo por ter um baixo acúmulo de sacarose (PAPINI-TERZI et al., 2009). Deste modo, é possível que esta diferença metabólica esteja relacionada com o desenvolvimento da planta ao longo do tempo e não com o acúmulo de sacarose, pois entrenós imaturos possuem baixos níveis de açúcar (PAPINI-TERZI et al., 2009). Essa diferença no perfil metabólico também pode ser observada na PLS-DA da **figura 17** que apresentou um $R^2 = 0,95$ e um $Q^2 = 0,92$ nos três primeiros componentes.

Figura 16: PCA das amostras do entrenó 1. C2E1 = 8 meses (vermelho), C3E1 = 11 meses (verde) e C4E1 = 13 meses após o plantio (azul).

Figura 17: PLS-DA 2D (A) e 3D (B) das amostras do entrenó 1. C2E1 = 8 meses (vermelho), C3E1 = 11 meses (verde) e C4E1 = 13 meses após o plantio (azul).

É possível observar na **figura 18** os 15 principais prováveis metabólitos que influenciaram na projeção do primeiro componente da PLS-DA (VIPs) do entrenó 1, juntamente com o nível de intensidade em cada tempo de coleta. Também são apresentadas as classes dos compostos anotados por MS/MS que foram: cumarinas e derivados, ácidos carboxílicos e derivados, glicerolípídeos, carboidratos e carboidratos conjugados.

Figura 18: VIPs da PLS-DA do entrenó 1 durante os tempos de coleta. C2E1 = 8 meses, C3E1 = 11 meses e C4E1 = 13 meses após o plantio.

4.2.5. Perfil metabólico do entrenó 5

Na análise do perfil metabólico do entrenó 5 da cana-de-açúcar foram identificados 172 prováveis metabólitos. Na **figura 19** é apresentado a PCA do perfil metabólico do entrenó 5 com 11 e 13 meses. A soma dos dois primeiros componentes é de 89,4%. É possível observar na figura uma separação entre estes dois tempos de coleta. Na PLS-DA (**figura 20**) também é observada esta separação entre estes dois tempos de coleta ($R^2 = 0,98$ e $Q^2 = 0,93$ nos três primeiros componentes)

O entrenó 5 é considerado um entrenó intermediário em relação ao acúmulo de sacarose (PAPINI-TERZI et al., 2009). Como nestes períodos de coleta (11 e 13 meses) a cana já está acumulando açúcar, essa diferença no perfil metabólico pode estar relacionada a um maior acúmulo de sacarose no entrenó 5.

Figura 19: PCA das amostras do entrenó 5. C3E5 = 11 meses (vermelho) e C4E5 = 13 meses após o plantio (verde).

Figura 20: PLS-DA 2D (A) e 3D (B) das amostras do entrenó 5. C3E5 = 11 meses (vermelho) e C4E5 = 13 meses após o plantio (verde).

É possível observar na **figura 21** os 15 principais prováveis metabólitos que influenciaram na projeção do primeiro componente da PLS-DA (VIPs) do entrenó 5 juntamente com o nível de intensidade em cada tempo de coleta. Também são apresentadas as classes dos compostos anotados por MS/MS, que foram: esteroides e derivados de esteroides,

benzenos e derivados substitutos, harmanos alcaloides, ácidos carboxílicos e derivados, glicerofosfolídeos.

Figura 21: VIPs da PLS-DA do entrenó 5 durante os tempos de coleta. C3E5 = 11 meses e C4E5 = 13 meses após o plantio.

4.2.6. Perfil metabólico do entrenó 9

Na análise do perfil metabólico do entrenó 9 da cana-de-açúcar foram encontrados 51 prováveis metabólitos. Esse número é inferior aos encontrados na análise dos outros tecidos (folha +1, entrenós 1 e 5), isso deve-se ao fato do entrenó 9 ser um entrenó maduro que contém altas concentrações de sacarose, (PAPINI-TERZI et al., 2009), podendo acumular até 50% de seu peso seco total de sacarose (BOTHÁ; BLACK, 2000). Também deve-se ao fato da técnica de LC-MS identificar preferencialmente metabólitos secundários (MATSUDA et al., 2009), onde açúcares como a sacarose, são considerados metabólitos primários (URBANCZYK-WOCHNIAK et al., 2003). Na **figura 22** é apresentado a PCA do perfil metabólico do entrenó 9, com 11 e 13 meses após o plantio. A soma dos dois primeiros componentes foi de 56,8% da variância. Nessa figura não existe uma grande diferenciação no perfil metabólico entre os dois períodos de coleta estudados. Já na PLS-DA (**figura 23**) pode-se observar uma nítida separação, com altos valores de $R^2 = 0,95$ e $Q^2 = 0,92$ nos três primeiros componentes.

Figura 22: PCA das amostras do entrenó 9. C3E9 = 11 meses (vermelho) e C4E9 = 13 meses após o plantio (verde).

Figura 23: PLS-DA 2D (A) e 3D (B) das amostras do entrenó 9. C3E9 = 11 meses (vermelho) e C4E9 = 13 meses após o plantio (verde).

O entrenó 9 é considerado um entrenó maduro e um dos que possuem a maior concentração de sacarose (BOTHÁ; BLACK, 2000). Na **figura 24** são apresentados os 15 principais prováveis metabólitos que influenciaram na projeção do primeiro componente da PLS-DA (VIPs) do entrenó 9 e seus respectivos níveis de intensidade em cada período de

coleta. Também são apresentadas as classes dos compostos anotados por MS/MS, que foram: acils graxos, esfingolipídios, ácidos carboxílicos e derivados.

Figura 24: VIPs da PLS-DA do entrenó 9 durante os tempos de coleta. C3E9 = 11 meses e C4E9: 13 meses após o plantio.

4.2.7. Identificação dos metabólitos por MS/MS

4.2.7.1. VIPs anotados da análise da folha +1 em todos os tempos de coleta

A anotação dos prováveis metabólitos que mais se destacaram na PLS-DA da folha +1 de cana-de-açúcar, em todos os tempos de coleta (4, 8, 11 e 13 meses após o plantio) estão apresentados na **tabela 2**. Na **figura 25** é apresentado o melhor resultado da fragmentação *in silico* para os prováveis metabólitos selecionados e fragmentados por MS/MS.

Tabela 2: Anotação dos principais VIPs da PLS-DA da folha +1 em todos os tempos de coleta.

Posição do VIP	<i>m/z</i> precursor	Tempo de retenção (s)	Composto identificado	<i>m/z</i> composto	Aduto	Classe
2	871.7293	689.8	TG(18:3(9Z,12Z,15Z)/14:1(9Z)/2:6(4Z,7Z,10Z,13Z,16Z,19Z))	870.6737	H+	Glicerolipídeos
3	871.7248	675.04	TG(18:3(9Z,12Z,15Z)/14:1(9Z)/2:6(4Z,7Z,10Z,13Z,16Z,19Z))	870.6737	H+	Glicerolipídeos
5	872.731	691.66	PC(18:1(11Z)/24:0)	871.7030	H+	Glicerofosfolipídeos
6	872.7349	675.87	PC(18:1(11Z)/24:0)	871.7030	H+	Glicerofosfolipídeos
7	493.2285	320.56	<i>Methyl lucidenate F</i>	470.2668	Na+	Prenol lipídeos
8	637.3287	211.65	<i>Neomycin</i>	614.3122	Na+	Carboidratos e carboidratos conjugados
9	605.2965	319	<i>25-Cinnamoyl-vulgaroside</i>	566.3243	K+	Prenol lipídeos
14	659.3228	366.89	CE(17:1)	636.5845	Na+	Esteroides e derivados de esteroides

Figura 25: Comparação dos espectros dos melhores resultados da fragmentação *in silico* dos compostos selecionados com os prováveis metabólitos correspondentes fragmentados por MS/MS da folha +1. **A:** m/z = 871.7293, rt = 689.8; **B:** m/z 871.7248, rt = 675.04; **C:** m/z = 872.731, rt = 691.66; **D:** m/z = 872.7349, rt = 675.87; **E:** m/z = 493.2285, rt = 320.56; **F:** m/z = 637.3287, rt = 211.65; **G:** m/z = 605.2965, rt = 319; **H:** m/z = 659.3228, rt = 366.89.

Dentre os 15 primeiros VIPs da PLS-DA da folha +1, oito foram fragmentados e anotados por MS/MS. Os VIPs de número 1, 4 e 10, com m/z de 885,7311 e rt de 689,1s, m/z de 621,3305 com rt de 174,9s, m/z de 886,4116 e rt de 522,2s respectivamente, foram fragmentados mas não obtiveram uma anotação confiável utilizando-se o banco de dados HMDB 3.6 (WISHART et al., 2013) e a fragmentação *in silico* com o MetFrag 2.2 CL (WOLF et al., 2010). Os metabólitos anotados foram:

- TG(18:3(9Z,12Z,15Z)/14:1(9Z)/22:6(4Z,7Z,10Z,13Z,16Z,19Z)) da classe metabólica dos glicerolípídeos para os VIPs de número 2 e 3 com m/z de 871.7293 e rt de 689.8s e m/z de 871.7248 e rt de 675s respectivamente. Este metabólito foi identificado em dois distintos tempos de retenção podendo ser algum tipo de isômeros de massa, onde não é possível a visualização e diferenciação pelo espectro de fragmentação.

- PC(18:1(11Z)/24:0) da classe metabólica dos glicerofosfolípídeos para os VIPs de número 5 e 6 com m/z de 872,731 e rt 691,66s e m/z de 872,7349 e rt de 675,87s. Este metabólito também foi identificado em dois distintos tempos de retenção podendo ser algum tipo de isômero de massa, onde não é possível a visualização e diferenciação pelo espectro de fragmentação.

- *Methyl lucidenate F* da classe metabólica dos prenol lipídeos para o VIP de número 7 com m/z de 493,2285 e rt de 320,56s.

- *Neomycin* da classe metabólica dos carboidratos e carboidratos conjugados para o VIP de número 8 com m/z de 637,3287 e rt de 211,65s.

- *25-Cinnamoyl-vulgaroside* da classe metabólica dos prenol lipídeos para o VIP de número 9 com m/z de 605,2965 e rt de 319s.

- CE(17:1) da classe metabólica dos Esteroides e derivados de esteroides para o VIP de número 14 com m/z de 659,3228 e rt de 366,89s.

De acordo com a **figura 13** a classe dos glicerolípídeos e glicerofosfolípídeos apresentaram uma maior intensidade aos 11 meses e a menor intensidade aos 13 meses, a classe dos prenol lipídeos apresentou uma maior abundância aos 11 (*methyl lucidenate F*) e também aos 13 meses (*25-Cinnamoyl-vulgaroside*), a classe dos carboidratos e conjugados foi mais intensa aos 11 meses e a classe dos esteroides e derivados teve um aumento da intensidade durante todos os períodos de coleta.

Os metabólitos anotados que mais se destacaram entre os principais VIPs foram o TG(18:3(9Z,12Z,15Z)/14:1(9Z)/22:6(4Z,7Z,10Z,13Z,16Z,19Z)) e o PC(18:1(11Z)/24:0) que pertencem às classes dos glicerolípídeos e dos glicerofosfolípídeos.

Os glicerolipídeos são uma classe de compostos contendo um esqueleto 3-glicerol carbono com uma ou duas cadeias acil esterificadas nas posições C1 e C2, e uma cabeça polar a posição C3 (BOUDIÈRE et al., 2014), eles são constituintes das folhas das plantas e também de outros tecidos (ROUGHAN; BATT, 1969) e sua composição nas membranas fotossintéticas foram muito bem conservadas durante a evolução, desde as cianobactérias aos cloroplastos das plantas superiores atuais (BLOCK et al., 2007; JOUHET; MARÉCHAL; BLOCK, 2007). Boudière et al. (2014) estudando os glicerolipídeos na fotossíntese evidenciaram a importância de vários compostos desta classe na estabilização de componentes do fotossistema II.

Apesar dos glicerolipídeos estarem relacionados com a estabilização dos componentes do fotossistema II (BOUDIÈRE et al., 2014), eles tiveram uma menor intensidade no tempo de coleta de 13 meses, período que possui o maior grau brix, comparando-o com o tempo de coleta de 11 meses, mostrando assim que o fotossistema II pode não estar tão eficaz no tecido mais maduro e com uma maior quantidade de açúcares.

Já os glicerofosfolipídeos são compostos por uma molécula de glicerol contendo um radical polar de fósforo em uma variedade de classes (HARRABI et al., 2009), eles são os compostos mais abundantes nas membranas de plantas, animais e fungos (KERWIN; TUININGA; ERICSSONT, 1994) e possuem um papel importante na bioquímica de todas as células vivas sendo o principal constituinte da membrana celular, protegendo as células de poluentes tóxicos e radicais livres (TRABELSI et al., 2014) e também funcionando como sinalizadores em plantas (COWAN, 2006).

Os glicerofosfolipídeos seguem o mesmo perfil de intensidade dos glicerolipídeos, sugerindo uma atuação conjunta. Apesar desta análise ser da folha +1, esta diminuição na abundância desses metabólitos aos 13 meses pode estar relacionada com o acúmulo de sacarose durante o desenvolvimento da cana-de-açúcar, pois as atividades enzimáticas e expressão de transcritos relacionados com a fotossíntese são modificados pela demanda dos tecidos dreno (MCCORMICK; CRAMER; WATT, 2011; PAUL; FOYER, 2001; ROLLAND; BAENA-GONZALEZ; SHEEN, 2006; SHEEN, 1990). Assim esses resultados sugerem que a fotossíntese pode estar sendo limitada pela diminuição da necessidade de sacarose pelo colmo, devido a um maior grau brix em 13 meses, pois com o amadurecimento do entrenó, o descarregamento de carboidratos por meio do floema deixa de ocorrer gradualmente (UYS et al., 2007).

4.2.7.2. VIPs anotados da análise do entrenó 1 com 8, 11 e 13 meses após o plantio

A anotação dos prováveis metabólitos que mais se destacaram na PLS-DA do entrenó 1 de cana-de-açúcar nos tempos de coleta de 8, 11 e 13 meses após o plantio, estão apresentados na **tabela 3**. Na **figura 26** é apresentado o melhor resultado da fragmentação *in silico* para os prováveis metabólitos selecionados e fragmentados por MS/MS.

Tabela 3: Anotação dos principais VIPs da PLS-DA do entrenó 1 com 8, 11 e 13 meses após o plantio.

Posição do VIP	<i>m/z</i> precursor	Tempo de retenção (s)	Composto identificado	<i>m/z</i> composto	Aduto	Classe
1	443.1382	119.67	<i>Pectachol</i>	442.2355	H+	Cumarinas e derivados
2	574.242	498.29	<i>Aliskiren</i>	551.3934	Na+	Ácidos carboxílicos e derivados
3	873.1993	413.51	TG(14:0/18:3(6Z,9Z,12Z)/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	872.6893	H+	Glicerolípídeos
4	496.1277	90.2	<i>Dhurrin 6'-glucoside</i>	473.1533	Na+	Carboidratos e carboidratos conjugados
6	554.1502	275.26	<i>Vignatic acid A</i>	553.2788	H+	Ácidos carboxílicos e derivados

Figura 26: Comparação dos espectros dos melhores resultados da fragmentação *in silico* dos compostos selecionados com os prováveis metabólitos correspondentes fragmentados por MS/MS do entrenó 1. **A:** $m/z = 443.1382$, $rt = 119.67$; **B:** $m/z = 574.242$, $rt = 498.29$; **C:** $m/z = 873.1993$, $rt = 413.51$; **D:** $m/z = 496.1277$, $rt = 90.2$; **E:** $m/z = 554.1502$, $rt = 275.26$.

Dentre os 15 primeiros VIPs da PLS-DA do entrenó 1 com 8, 11 e 13 meses após o plantio, cinco foram fragmentados e anotados por MS/MS. O VIP de número 5 com m/z de 874,2027 e rt de 413,5s, foi fragmentado mas sem uma anotação confiável utilizando-se o banco de dados HMDB 3.6 (WISHART et al., 2013) e a fragmentação *in silico* com o MetFrag 2.2 CL (WOLF et al., 2010). Os metabólitos anotados foram:

- *Pectachol* da classe metabólica das cumarinas e derivados para o VIP de número 1 com m/z de 443,1382 e rt de 119,67s.
- *Aliskiren* da classe metabólica dos ácidos carboxílicos e derivados para o VIP de número 2 com m/z de 574,242 e rt de 498,29s.
- TG(14:0/18:3(6Z,9Z,12Z)/22:6(4Z,7Z,10Z,13Z,16Z,19Z)) da classe metabólica dos glicerolípídeos para o VIP de número 3 com m/z de 873,1993 e rt de 413,51s.
- *Dhurrin 6'-glucoside* da classe metabólica dos carboidratos e carboidratos conjugados para o VIP de número 4 com m/z de 496,1277 e rt de 90,2s.
- *Vignatic acid A* da classe metabólica dos ácidos carboxílicos e derivados para o VIP de número 6 com m/z de 554,1502 e rt de 275,26s.

Na **figura 16** é possível observar que dentre os VIPs anotados a classe das cumarinas e derivados, ácidos carboxílicos e derivados, carboidratos aumentam sua intensidade de acordo com o tempo, apresentando menor intensidade aos 8 meses e maior aos 13 meses após o plantio, já a classe dos glicerolípídeos apresenta um perfil contrário, tendo maior intensidade com 8 meses e menor com 13 meses após o plantio.

O metabólito mais importante para a projeção da PLS-DA é o *Pectachol* da classe metabólica das cumarinas e derivados. As cumarinas são compostos aromáticos policíclicos contendo um radical 1-benzopirano com um grupo cetona no átomo de carbono C2. Elas são encontradas em plantas superiores originadas da via dos fenilpropanoides (HARBORNE, 1999) e possuem diversas bioatividades, podendo atuar desde o crescimento normal de plantas a resposta a diferentes tipos de estresses (HASHEMITABAR; KOLAH; TABANDEH, 2014), como na defesa contra microrganismos (DIXON; PAIVA, 1995; RADULOVIC et al., 2013; RIBERA; ZUÑIGA, 2012).

O metabólito *Pectachol* possui um aumento de intensidade gradativo acompanhando o envelhecimento do tecido e consequentemente o aumento do acúmulo de sacarose na cana-de-açúcar. Pelo fato da classe deste metabólito poder estar atuando em diferentes respostas ao estresse (DIXON; PAIVA, 1995; HASHEMITABAR; KOLAH; TABANDEH, 2014; RADULOVIC et al., 2013; RIBERA; ZUÑIGA, 2012), assim como já foi observado o aumento de expressão de proteínas relacionadas ao estresse durante o acúmulo de sacarose (CASU

et al., 2004; GLASSOP et al., 2007; PAPINI-TERZI et al., 2009) é bem provável que este metabólito esteja relacionado ao desenvolvimento e acúmulo de sacarose na cana-de-açúcar.

4.2.7.3. VIPs anotados da análise do entrenó 5 com 11 e 13 meses após o plantio

A anotação dos prováveis metabólitos que mais se destacaram na PLS-DA do entrenó 5 de cana-de-açúcar nos tempos de coleta de 11 e 13 meses após o plantio, estão apresentados na **tabela 4**. Na **figura 27** é apresentado o melhor resultado da fragmentação *in silico* para os prováveis metabólitos selecionados e fragmentados por MS/MS.

Tabela 4: Anotação dos principais VIPs da PLS-DA do entrenó 5 com 11 e 13 meses após o plantio.

Posição do VIP	m/z precursor	Tempo de retenção (s)	Composto identificado	m/z composto	Aduto	Classe
1	905.1752	337.51	25-Acetyl-6,7-didehydrofevicordin F 3-[glucosyl-(1->6)-glucoside]	866.3936	K+	Esteroides e derivados de esteroides
2	630.0248	208.06	Zafirlukast metabolite M5	591.2039	K+	Benzenos e derivados substitutos
3	273.0658	130.28	(2S,4R)-4-(9H-Pyrido[3,4-b]indol-1-yl)-1,2,4-butanetriol	272.1160	H+	Harmanos alcalóides
5	271.0528	131.98	Histidiny-Aspartate	270.0964	H+	Ácidos carboxílicos e derivados
11	978.1199	230.74	CDP-DG(16:0/18:2(9Z,12Z))	977.5142	H+	Glicerofosfolípidos

Figura 27: Comparação dos espectros dos melhores resultados da fragmentação *in silico* dos compostos selecionados com os prováveis metabólitos correspondentes fragmentados por MS/MS do entrenó 5. **A:** m/z = 905.1752, rt = 337.51; **B:** m/z = 630.0248, rt = 208.06; **C:** m/z = 273.0658, rt = 130.28; **D:** m/z = 271.0528, rt = 131.98; **E:** m/z = 978.1199, rt = 230.74.

Dentre os 15 VIPs da PLS-DA do entrenó 5 com 11 e 13 meses após o plantio, cinco foram fragmentados e anotados por MS/MS. O VIP de número 4 com m/z de 950,088 e rt de 218,77s, foi fragmentado, mas sem uma anotação confiável utilizando-se o banco de dados HMDB 3.6 (WISHART et al., 2013) e a fragmentação *in silico* com o MetFrag 2.2 CL (WOLF et al., 2010). Os metabólitos anotados foram:

- *25-Acetyl-6,7-didehydrofevicordin F 3-[glucosyl-(1->6)-glucoside]* da classe metabólica dos esteroides e derivados de esteroides para o VIP de número 1 com m/z de 905,1752 e rt de 337,51s.

- *Zafirlukast metabolite M5* da classe metabólica dos benzenos e derivados substitutos para o VIP de número 2 com m/z de 630,0248 e rt de 208,06s.

- *(2S,4R)-4-(9H-Pyrido[3,4-b]indol-1-yl)-1,2,4-butanetriol* da classe metabólica dos Harmanos alcaloides para o VIP de número 3 com m/z de 273,0658 e rt de 130,28s.

- *Histidinyl-Aspartate* da classe metabólica dos ácidos carboxílicos e derivados para o VIP de número 5 com m/z de 271,0528 e rt de 131,98s.

- *CDP-DG(16:0/18:2(9Z,12Z))* da classe metabólica dos glicerofosfolídeos para o VIP de número 11 com m/z de 978,1199 e rt de 230,74s

Na **figura 21** pode ser observado que todos os 15 primeiros VIPs tiveram uma maior intensidade aos 13 meses após o plantio, portanto todas as classes identificadas (esteroides e derivados, benzenos e derivados, harmanos alcaloides, ácidos carboxílicos e derivados e glicerofosfolídeos) foram menos abundantes aos 11 meses do que aos 13 meses.

O metabólito que mais influenciou na projeção da PLS-DA foi o *25-Acetyl-6,7-didehydrofevicordin F 3-[glucosyl-(1->6)-glucoside]* da classe metabólica dos esteroides e derivados de esteroides. Os esteroides e derivados são compostos por um grande grupo de substâncias que possuem em comum um anel baseado em *1,2-cyclopentanoperhydrophenanthrene*. Este grupo inclui vários compostos, como os ácidos biliares, corticosteroides, hormônios sexuais, esteróis e vários esteroides de planta (SMITH, 1997)

Os esteroides são comumente classificados como metabólitos secundários em plantas, mas são compostos de vital importância para os constituintes celulares de plantas, atuando sobre a germinação, crescimento, floração, expressão sexual de plantas (HEFTMANN, 1975). Muitas das funções dos esteroides são conservadas entre plantas e animais, como o auxílio ao desenvolvimento, aceleração do envelhecimento e a diferenciação sexual, mas também existe esteroides com funções específicas em cada reino, como hormônios esteroides que promovem a diferenciação vascular e estomatal, alongamento

celular e imunidade em plantas (HARTWIG; WANG, 2015; RIGAL; MA; ROBERT, 2014). Estes compostos já foram encontrados em cana-de-açúcar (DESHMANE; DEV, 1971).

Como o entrenó 5 é considerado um entrenó intermediário em relação ao acúmulo de sacarose (PAPINI-TERZI et al., 2009) e considerando as funções dos esteroides em plantas, este metabólito pode estar relacionado com o desenvolvimento do entrenó e indiretamente preparando este tecido para um maior acúmulo de sacarose, pois a concentração de sacarose é menor em entrenós imaturos, mas tende a aumentar durante o desenvolvimento, chegando a 50% do total de seu peso seco (BOTHÁ; BLACK, 2000).

4.2.7.4. VIPs anotados da análise do entrenó 9 com 11 e 13 meses após o plantio

A anotação dos prováveis metabólitos que mais se destacaram na PLS-DA do entrenó 9 de cana-de-açúcar nos tempos de coleta de 11 e 13 meses após o plantio, estão apresentados na **tabela 5**. Na **figura 28** é apresentado o melhor resultado da fragmentação *in silico* para os prováveis metabólitos selecionados e fragmentados por MS/MS.

Tabela 5: Anotação dos principais VIPs da PLS-DA do entrenó 9 com 11 e 13 meses após o plantio.

Posição do VIP	<i>m/z</i> precursor	Tempo de retenção (s)	Composto identificado	<i>m/z</i> composto	Aduto	Classe
1	575.489	695.73	<i>Montecristin</i>	574.4961	H+	Acils graxos
2	520.3345	679.98	<i>Ceramide (d18:1/12:0)</i>	481.4494	K+	Esfoliolipídios
3	163.038	95.28	<i>3-Mercapto-3-methyl-1-butyl acetate</i>	162.0714	H+	Ácidos carboxílicos e derivados
4	381.1095	142.67	<i>Deacetylisovaltrate</i>	380.1835	H+	Acils graxos

Figura 28: Comparação dos espectros dos melhores resultados da fragmentação *in silico* dos compostos selecionados com os prováveis metabólitos correspondentes fragmentados por MS/MS do entrenó 9. **A:** $m/z = 575.489$, $rt = 695.73$; **B:** $m/z = 520.3345$, $rt = 679.98$; **C:** $m/z = 163.038$, $rt = 95.28$; **D:** $m/z = 381.1095$, $rt = 142.67$.

Dentre os 15 primeiros VIPs da PLS-DA do entrenó 9 com 11 e 13 meses após o plantio, quatro foram fragmentados e anotados por MS/MS. O VIP de número 5 com m/z de 411,1151 e rt de 157,9s, foi fragmentado, mas sem uma anotação confiável utilizando-se o banco de dados HMDB 3.6 (WISHART et al., 2013) e a fragmentação *in silico* com o MetFrag 2.2 CL (WOLF et al., 2010). Os metabólitos anotados foram:

- *Montecristin* da classe metabólica dos acils graxos para o VIP de número 1 com m/z de 575,489 e rt de 695,73s.

- *Ceramide (d18:1/12:0)* da classe metabólica dos esfingolipídios para o VIP de número 2 com m/z de 520,3345 e rt de 679,98s.

- *3-Mercapto-3-methyl-1-butyl acetate* da classe metabólica dos ácidos carboxílicos e derivados para o VIP de número 3 com m/z de 163,038 e rt de 95,28s.

- *Deacetylisovaltrate* da classe metabólica dos Acils graxos para o VIP de número 4 com m/z de 381,1095 e rt de 142,67s.

Na **figura 24** também é possível observar que todos os 15 primeiros VIPs tiveram uma maior intensidade com 13 meses após o plantio, assim todas as classes (acils graxos, esfingolipídios e ácidos carboxílicos e derivados) tiveram uma menor intensidade aos 11 meses e maior aos 13 meses após o plantio.

O VIP mais importante para a diferenciação do entrenó 9 relacionado com o tempo de coleta, foi a *Montecristin*. Este metabólito possivelmente está envolvido na biossíntese de acetogeninas (GLEYE et al., 1997) e pertence à classe dos acil graxos, que são um grupo diversificado de moléculas sintetizadas pelo alongamento da cadeia de um iniciador de acetil-CoA com grupos de malonil-CoA (ou metilmalonil-CoA) que podem conter uma funcionalidade cíclica podendo ser substituídas com heteroátomos (FAHY et al., 2005).

Os acil graxos fazem parte da super classe dos lipídeos (WISHART et al., 2013) que é a classe com os principais constituintes de membranas de organismos procariotos e eucariotos. Eles também são componentes estruturais de membranas intracelulares e possuem diversas funções biológicas, como o armazenamento de energia e de carbono, transdução de sinal, respostas ao estresse e sinalização (SIEBERS et al., 2016; WELTI et al., 2007).

A classe dos acil graxos não possui uma função definida e a super classe dos lipídeos possuem diversas funções. A função biológica da *montecritin*, que é a participação na biossíntese de acetogeninas não está relacionada diretamente com o acúmulo de açúcares em plantas, mas consequentemente, a identificação de metabólitos que aumentam sua abundância seguindo o aumento da sacarose na planta, pode levar a um melhor entendimento do acúmulo de sacarose em cana-de-açúcar (GLASSOP et al., 2007). Assim este metabólito pode estar correlacionado direta ou indiretamente com o acúmulo de sacarose na cana-de-açúcar.

4.3. Proteômica

4.3.1. Extração das Proteínas

A quantidade de proteínas totais obtidas pelas extrações dos tecidos de cana-de-açúcar (folha +1, entrenó 5 e entrenó 9) foram entre 60 μ g a 300 μ g dependendo da amostra, as amostras da folha +1 apresentaram uma maior quantidade de proteínas extraídas, já os

entrenós 9 foram as amostras que apresentaram o menor rendimento da extração. A qualidade da extração das proteínas pode ser observada na **Figura 29**. É possível observar a homogeneidade entre os *bulks* de mesmo tecido e época de coleta, mostrando uma boa qualidade da extração de proteínas e quantificação.

Figura 29: Géis de poliacrilamida das amostras de cana-de-açúcar da folha +1, entrenó 5 e entrenó 9 com 11 e 13 meses após o plantio. BSA (*Bovine Serum Albumin*) utilizado como padrão; A1, A2 e A3 = folha +1 com 11 meses após o plantio; B1, B2 e B3 folha +1 com 13 meses após o plantio; C1, C2 e C3 = entrenó 5 com 11 meses após o plantio; D1, D2 e D3 entrenó 5 com 13 meses após o plantio; E1, E2 e E3 = entrenó 9 com 11 meses após o plantio; F1, F2 e F3 = entrenó 9 com 13 meses após o plantio.

4.3.2. Perfil proteico da folha +1 com 11 e 13 meses após o plantio

Na análise do proteoma da folha +1 de cana-de-açúcar com 11 e 13 meses após o plantio, foram encontradas e quantificadas 1631 proteínas. Na **figura 30** é apresentado a PCA do perfil proteico da folha +1 de cana-de-açúcar com 11 e 13 meses após o plantio. A soma dos dois primeiros componentes principais foi de 73,6%. É nítida a diferença das amostras de acordo com o tempo de coleta. A **figura 31** mostra a PLS-DA da folha +1, onde também é possível observar a diferença entre as amostras de acordo com o tempo de coleta. Os valores apresentados na PLS-DA são de $R^2 = 0,99$ e $Q^2 = 0,90$ nos dois primeiros componentes.

Figura 30: PCA do perfil proteico das amostras da folha +1 com 11 e 13 meses após o plantio. C3F1 = 11 meses (vermelho), C4F1 = 13 meses após o plantio (verde).

Figura 31: PLS-DA do perfil proteico das amostras da folha +1 com 11 e 13 meses após o plantio. C3F1 = 11 meses (vermelho), C4F1 = 13 meses após o plantio (verde).

Na **figura 32** é possível observar as 15 principais proteínas que influenciaram na projeção (VIPs) do primeiro componente da PLS-DA do perfil proteico da folha +1, juntamente com o nível de intensidade em cada tempo de coleta. Também é apresentada a anotação dessas proteínas, que foi feita utilizando-se o software Blast2GO (CONESA et al., 2005).

Figura 32: VIPs da PLS-DA do perfil proteico das amostras da folha +1 com 11 e 13 meses após o plantio com a intensidade da proteína em cada período de coleta. C3F1 = 11 meses, C4F1 = 13 meses após o plantio.

Os resultados dos VIPs da PLS-DA, evidenciaram que a maior influência para a variância nessa análise foram duas histonas, a *histone h4-like* (SCAGLR2026F05), e a *histone* (SCVPLR1049C05). Os processos biológicos nos quais essas histonas participam de acordo com a anotação são os de metilação do DNA em citosina e o de montagem de nucleossomo, respectivamente.

Histonas são um grupo de proteínas altamente conservadas durante a evolução e possuem como função o empacotamento do DNA na cromatina. As principais histonas podem ser classificadas dentre cinco tipos de histonas, a H1, H2A, H2B, H3 e H4 (SMITH, 1997). Elas são altamente conservadas entre animais e plantas, especialmente as H3 e H4, também foi demonstrado que além de sua importância estrutural, as histonas estariam envolvidas na condensação e descondensação da cromatina durante o ciclo celular e regulação, reparo, e recombinação da expressão gênica, (BELYAEV et al., 1997; MORAES et al., 2015; TURNER, 2002).

As histonas H2A, H2B, H3 e H4 (duas moléculas de cada) se unem formando um octâmero chamado de nucleossomo, que é a estrutura básica da cromatina de eucariotos. A histona H1 não participa da formação do nucleossomo, mas está associada à ele (FELSENFELD, 1978).

CASU et al. (2004) observaram em seu trabalho, estudando transcritos no amadurecimento de colmos de cana-de-açúcar, uma diminuição de transcritos que codificam histonas em colmos maduros comparando-os com imaturos, indicando assim que o tecido pode tornar-se menos ativo em relação à transcrição e a tradução, devido a maturação. Em outro trabalho CASU et al. (2007) estudando transcritos associados com o metabolismo da parede celular e o desenvolvimento do colmo de cana-de-açúcar, identificaram que alguns tipos de histonas (H2A, H3 e H4) possuem uma menor expressão com o aumento da maturidade do colmo, reforçando a ideia de uma menor atividade nos colmos maduros.

Neste trabalho, ambas histonas possuem uma maior concentração com 11 meses de coleta, período no qual a planta possui um grau brix menor que o da coleta com 13 meses após o plantio. Apesar dessas histonas localizarem-se em tecido foliar, é possível que elas indiquem um maior amadurecimento da cana-de-açúcar aos 13 meses após o plantio, significando uma maior quantidade de sacarose na planta, assim as histonas teriam uma relação com o acúmulo de sacarose na cana-de-açúcar.

4.3.3. Perfil proteico do entrenó 5 com 11 e 13 meses após o plantio

Na análise do proteoma do entrenó 5 de cana-de-açúcar com 11 e 13 meses após o plantio, foram encontradas e quantificadas 1635 proteínas. Na **figura 33** é apresentado a PCA do perfil proteico do entrenó 5 de cana-de-açúcar com 11 e 13 meses após o plantio. A soma dos dois primeiros componentes principais foi de 73,4% e pode-se observar as diferenças entre as amostras de acordo com o tempo de coleta. A **figura 34** mostra a PLS-DA do entrenó 5, onde também é possível observar a diferença entre as amostras de acordo com o tempo de coleta. Os valores apresentados na PLS-DA são de $R^2 = 0,97$ e $Q^2 = 0,89$ nos dois primeiros componentes.

Figura 33: PCA do perfil proteico das amostras do entrenó 5 com 11 e 13 meses após o plantio. C3E5 = 11 meses (vermelho), C4E5 = 13 meses após o plantio (verde).

Figura 34: PLS-DA do perfil proteico das amostras do entrenó 5 com 11 e 13 meses após o plantio. C3E5 = 11 meses (vermelho), C4E5 = 13 meses após o plantio (verde).

Na **figura 35** é possível observar as 15 principais proteínas que influenciaram na projeção (VIPs) do primeiro componente da PLS-DA do perfil proteico do entrenó 5, juntamente com o nível de intensidade em cada tempo de coleta. Também é apresentada a anotação dessas proteínas, que foi feita utilizando-se o software Blast2GO (CONESA et al., 2005).

Figura 35: VIPs da PLS-DA do perfil proteico das amostras do entrenó 5 com 11 e 13 meses após o plantio com a intensidade da proteína em cada período de coleta. C3E5 = 11 meses, C4E5 = 13 meses após o plantio.

Os resultados dos VIPs da PLS-DA mostram que a proteína que mais influenciou a variância entre os tempos de coleta foi a *cold acclimation protein* (SCCCLR1001F04) e o processo biológico de que ela participa é o de resposta a água, que está dentro da categoria de processos de resposta ao estresse.

CASU et al. (2004) em seu trabalho sobre transcritos expressos durante a maturação de colmos de cana-de-açúcar, evidenciaram uma expressão maior de proteínas relacionadas ao estresse durante a maturação da cana-de-açúcar. Outros trabalhos também citam a correlação de proteínas relacionadas ao estresse com o acúmulo de sacarose em cana-de-açúcar (GLASSOP et al., 2007; PAPINI-TERZI et al., 2009).

A proteína *cold acclimation protein* é uma proteína induzida durante a aclimação ao frio (THOMASHOW, 1999), portanto ela é uma proteína relacionada ao estresse. Essa proteína teve uma maior intensidade nas amostras com 13 meses após o plantio, este aumento de intensidade pode estar relacionado com o acúmulo de sacarose, principalmente pelo fato dela ser a proteína que mais contribuiu para a variância do perfil proteico do entrenó 5 durante seu desenvolvimento, acompanhando assim o aumento da quantidade de açúcar (grau brix) na cana. Também devido ao fato da cana não estar sobre influência de frio nesse período de coleta, conforme apresentado na **figura 6**.

Também é possível observar a presença de uma histona (*histone h4-like* - SCAGLR2026F05) entre os cinco primeiros VIPs, apresentando uma intensidade maior no entrenó mais novo, reforçando assim a relação destas proteínas com o acúmulo de sacarose já discutido nos resultados do perfil proteico da folha +1 com 11 e 13 meses após o plantio.

4.3.4. Perfil proteico do entrenó 9 com 11 e 13 meses após o plantio

Na análise do proteoma do entrenó 9 de cana-de-açúcar, foram encontradas e quantificadas 1020 proteínas. Na **figura 36** é apresentado a PCA do perfil proteico do entrenó 9 de cana-de-açúcar com 11 e 13 meses após o plantio. A soma dos dois primeiros componentes principais foi de 71,5% e é possível observar a diferença entre as amostras de acordo com o tempo de coleta, apesar da dispersão das repetições biológicas dentro do grupo. A **figura 37** mostra a PLS-DA do entrenó 9, onde também é possível observar a diferença entre as amostras de acordo com o tempo de coleta. Os valores apresentados na PLS-DA são de $R^2 = 0,99$ e $Q^2 = 0,63$ nos dois primeiros componentes.

Figura 36: PCA do perfil proteico das amostras do entrenó 9 com 11 e 13 meses após o plantio. C3E9 = 11 meses (vermelho), C4E9 = 13 meses após o plantio (verde).

Figura 37: PLS-DA do perfil proteico das amostras do entrenó 9 com 11 e 13 meses após o plantio. C3E9 = 11 meses (vermelho), C4E9 = 13 meses após o plantio (verde).

Na **figura 38** é possível observar as 15 principais proteínas que influenciaram na projeção (VIPs) do primeiro componente da PLS-DA do perfil proteico do entrenó 9, juntamente com o nível de intensidade em cada tempo de coleta. Também é apresentada a anotação dessas proteínas, que foi feita utilizando-se o software Blast2GO (CONESA et al., 2005).

Figura 38: VIPs da PLS-DA do perfil proteico das amostras do entrenó 9 com 11 e 13 meses após o plantio com a intensidade da proteína em cada período de coleta. C3E9 = 11 meses , C4E9 = 13 meses após o plantio.

Os resultados dos VIPs da PLS-DA mostram que a proteína que mais influenciou para a variância entre os tempos de coleta foi a *plasma membrane-associated cation-binding protein 1 (PCaP1)* (SCCCRZ2004B02) e o processo biológico de que ela participa é o de resposta celular ao estímulo.

A proteína PCaP1 é uma proteína cátion-ligante encontrada primeiramente em *Arabidopsis thaliana*, ela pode se ligar a cátions de Ca^{2+} e Cu^{2+} , (IDE et al., 2007; NAGASAKI; TOMIOKA; MAESHIMA, 2008) e também a fosfatidilinositol (KATO et al., 2010). É uma proteína que se liga firmemente a membrana plasmática por meio de N-miristoilação (IDE et al., 2007), que é a introdução de um ou mais grupos de miristoilo em um componente orgânico, como as proteínas, por meio de uma ligação amida ao resíduo N-terminal de glicina (SMITH, 1997). Apesar de diversos trabalhos sobre esta proteína, sua função fisiológica ainda é desconhecida (NAGATA et al., 2016).

Vijayapalani et al. (2012) estudando a interação de *Potyvirus* com a proteína PCaP1, verificou que esta proteína é essencial para uma infecção eficiente, mostrando uma relação da PCaP1 com a movimentação intercelular do *Potyvirus*. Em outro trabalho MOHAMMADI; MOIENI; KOMATSU (2012) comparando o proteoma de raízes de colza durante a seca,

verificaram um aumento da proteína PCaP1 em linhagens tolerantes a seca, portanto estando relacionada a tolerância a seca na colza.

Outros estudos sugerem que a proteína PCaP1 possa estar relacionada com algumas funções, como o fechamento estomático (NAGATA et al., 2016), resposta a patógenos (IDE et al., 2007), sinalização intracelular (NAGASAKI; TOMIOKA; MAESHIMA, 2008) e estabilidade da membrana (IDE et al., 2007).

Apesar de não existir trabalhos relacionando a proteína PCaP1 com o acúmulo de sacarose, é possível que ela esteja relacionada ao acúmulo, pois foi a principal proteína (VIP) que influenciou na projeção da PLS-DA do entrenó 9 da cana-de-açúcar, com uma maior intensidade no entrenó mais jovem.

Também foi possível observar a presença de uma histona (*histone h4-like* - SCAGLR2026F05) entre os cinco primeiros VIPs, apresentando uma intensidade maior no entrenó mais novo. Essa proteína foi encontrada entre os cinco primeiros VIPs em todos os tecidos de cana-de-açúcar analisados (folha +1, entrenó 5 e entrenó 9), sempre apresentando uma redução em seu nível de intensidade no tecido mais velho, reforçando assim a relação e a importância da histona h4 com o acúmulo de sacarose já discutido nas análises da folha +1 e do entrenó 5, onde também foi presenciado por outros trabalhos (CASU et al., 2004, 2007).

4.4. Integração dos dados de metabolômica e proteômica

Para uma análise mais profunda dos dados de metabolômica e proteômica visando uma abordagem na área da biologia sistêmica, foi realizado uma integração desses dois conjuntos de dados. Assim foi possível uma melhor visualização das correlações entre metabólitos e proteínas com o acúmulo de sacarose na cana-de-açúcar durante seu desenvolvimento.

A integração dos dados foi feita utilizando-se a análise de correlação canônica regularizada (rCCA) que consegue lidar com a integração de dois conjuntos de dados com um número maior de variáveis que o número de unidades ($p > n$ e $q > n$) (GONZALEZ et al., 2008).

As redes biológicas foram exportadas para o software de visualização Cytoscape (SHANNON et al., 2003b), que possibilitou uma exploração das redes na busca de *hubs*, que são nós altamente conectados, onde normalmente exercem funções importantes em redes biológicas (DONG; HORVATH, 2007; GUO et al., 2014; HONG et al., 2013; MONTASTIER et al., 2015). Também foi feita a busca por correlações entre as enzimas participantes da via do metabolismo de amido e sacarose (map00500), que foram identificadas utilizando-se o software Blast2GO (CONESA et al., 2005) com o banco de dados de vias metabólicas do KEGG (KANEHISA et al., 2016). Esta via metabólica é importante por conter enzimas ligadas ao metabolismo de carboidratos (CASU et al., 2007) e enzimas importantes na produção de

sacarose, como a sacarose fosfato sintase (EC:2.4.1.14), sacarose sintase (EC: 2.4.1.13) e invertase (EC:3.2.1.26) (BOTHÁ; BLACK, 2000; SUI et al., 2015).

4.4.1. Rede biológica da folha +1 com 11 e 13 meses após o plantio

Para a criação da rede biológica de correlações da folha +1, foi utilizado como parâmetros para a rCCA um λ_1 de 0,1029388 e um λ_2 de 0,3475918 resultando em um *cv-score* de 0,97713. A rede F1A obtida a partir desses parâmetros contém 591 nós e 2025 arestas (**figura 39**). Ainda na **figura 39** é possível observar dois clusters, um com três grandes *hubs* (**figura 39-A**) e outro com dois grandes *hubs* (**figura 39-B**)

Figura 39: Rede biológica de integração F1A dos metabólitos e proteínas da folha +1, onde os nós de cor azul são os metabólitos e os de cor cinza são as proteínas, também se

encontram destacados os principais *hubs*. As cores das arestas indicam se a correlação é negativa (verde) ou positiva (vermelha). **A:** *cluster* com três grandes *hubs*. **B:** *cluster* com dois grandes *hubs*.

Na **figura 39-A** os três principais *hubs* (hub_F1A1, hub_F1A2 e hub_F1A3) são os possíveis metabólitos com *mz/rt* de 806,4367/358,65, 936,3652/216,89 e 337,1734/622,57, respectivamente, e juntos eles estão correlacionados diretamente com 233 proteínas. O hub_F1A1 está correlacionado positivamente com 120 proteínas e negativamente com 65 proteínas e os processos biológicos das proteínas de acordo com a anotação são apresentados na **figura 40** (correlações positivas) e **figura 41** (correlações negativas). O hub_F1A2 está correlacionado positivamente com 155 proteínas e negativamente com 19 proteínas e os processos biológicos das proteínas de acordo com a anotação são apresentados na **figura 42** (correlações positivas) e **figura 43** (correlações negativas). O hub_F1A3 está correlacionado positivamente com 96 proteínas e negativamente com 64 proteínas e os processos biológicos das proteínas de acordo com a anotação são apresentados na **figura 44** (correlações positivas) e **figura 45** (correlações negativas).

Figura 40: Processos biológicos das proteínas correlacionadas positivamente com o *hub* hub_F1A1 da rede de correlações F1A de metabólitos e proteínas da folha +1 com 11 e 13 meses.

Figura 41: Processos biológicos das proteínas correlacionadas negativamente com o *hub* hub_F1A1 da rede de correlações F1A de metabólitos e proteínas da folha +1 com 11 e 13 meses.

Figura 42: Processos biológicos das proteínas correlacionadas positivamente com o *hub_F1A2* da rede de correlações F1A de metabólitos e proteínas da folha +1 com 11 e 13 meses.

Figura 43: Processos biológicos das proteínas correlacionadas negativamente com o *hub* hub_F1A2 da rede de correlações F1A de metabólitos e proteínas da folha +1 com 11 e 13 meses.

Figura 44: Processos biológicos das proteínas correlacionadas positivamente com o *hub_F1A3* da rede de correlações F1A de metabólitos e proteínas da folha +1 com 11 e 13 meses.

Figura 45: Processos biológicos das proteínas correlacionadas negativamente com o *hub* hub_F1A3 da rede de correlações F1A de metabólitos e proteínas da folha +1 com 11 e 13 meses.

Os processos biológicos que mais se destacaram nas correlações positivas dos *hubs* hub_F1A1, hub_F1A2 e hub_F1A3 foram a fotossíntese (*photosynthesis*) e a tradução (*translation*). Nas correlações negativas os principais processos foram a tradução (*translation*), regulação do transporte de próton (*regulation of proton transport*), processo de oxidação-redução (*oxidation-reduction process*) e processo glicolítico (*glycolytic process*).

Na **figura 39-B** os dois principais *hubs* (hub_F1A4 e hub_F1A5) são os possíveis metabólitos com *mz/rt* de 607,3091/220.81 e 783,4114/353.47, respectivamente, e juntos eles estão correlacionados diretamente com 124 proteínas. O hub_F1A4 está correlacionado positivamente com 102 proteínas e negativamente somente com uma proteína, os processos biológicos das proteínas com correlação positiva de acordo com a anotação são apresentados na **figura 46** (correlações positivas), a única proteína correlacionada negativamente com este

hub faz parte do processo biológico chamado processo metabólico (*metabolic process*). O hub_F1A5 está correlacionado positivamente com 114 proteínas e não possui nenhuma correlação negativa e os processos biológicos das proteínas de acordo com a anotação são apresentados na **figura 47** (correlações positivas).

Figura 46: Processos biológicos das proteínas correlacionadas positivamente com o hub hub_F1A4 da rede de correlações F1A de metabólitos e proteínas da folha +1 com 11 e 13 meses.

Figura 47: Processos biológicos das proteínas correlacionadas positivamente com o *hub* hub_F1A5 da rede de correlações F1A de metabólitos e proteínas da folha +1 com 11 e 13 meses.

Os processos biológicos que mais se destacaram nas correlações positivas dos *hubs* hub_F1A4 e hub_F1A5 foram os processos de oxidação-redução (*oxidation-reduction process*) e o de tradução (*translation*), já nas correlações negativas somente o hub_F1A4 está correlacionado com uma proteína.

O processo de fotossíntese existe em organismos tais como, plantas superiores, algas verdes e cianobactérias, que consiste na síntese de compostos químicos orgânicos, especialmente carboidratos, a partir de CO₂ utilizando energia obtida a partir da luz, em vez da oxidação de compostos químicos (SMITH, 1997).

A tradução é o processo metabólico pelo qual uma sequência específica de bases do RNA mensageiro (mRNA) determina uma sequência de aminoácidos durante a biossíntese de proteínas e de peptídeos (SMITH, 1997).

A regulação do transporte de próton é qualquer processo que modula a frequência, a velocidade e a extensão de transporte de prótons para fora ou dentro de uma célula, ou entre células utilizando-se algum agente, como um transportador ou poro (BLAKE et al., 2015)

O processo metabólico de oxidação-redução resulta na remoção ou adição de um ou mais elétrons para uma ou de uma substância, podendo ser realizado com ou sem a remoção ou adição de próton ou os prótons (BLAKE et al., 2015)

O processo glicolítico é o conjunto de reações químicas e as vias que resultam na quebra de um carboidrato em piruvato, produzindo também ATP e a redução de NAD⁺ a NADH (NELSON; COX, 2012).

Com esses resultados é possível observar que o principal *hub* da folha +1, que pode estar relacionado com o acúmulo de sacarose na cana-de-açúcar é o *hub_F1A1*, pois possui uma forte correlação positiva com várias proteínas que fazem parte do processo biológico da fotossíntese. A sacarose é o principal produto da fotossíntese em plantas e a principal forma em que os carboidratos são translocados das folhas para o resto da planta, fornecendo carbono e energia para o crescimento e também para o acúmulo de reservas energéticas (GROF; CAMPBELL, 2001; LUNN; FURBANK, 1999; MOORE, 2005).. Assim o *hub hub_F1A1* pode estar influenciando positivamente a produção de sacarose e consequentemente o seu acúmulo nos entrenós.

Para observar melhor as correlações entre as proteínas e metabólitos relacionados com o acúmulo de sacarose, foi feita a busca das enzimas participantes do metabolismo de amido e sacarose (map00500 – KEGG) (KANEHISA et al., 2016) na rede metabólica da folha +1. As enzimas encontradas nesta rede de correlações foram a *sucrose synthase* (EC:2.4.1.13 - SCCCLR1001A05), *UDP glucose pyrophosphorylase* (EC:2.7.7.9 - SCQGLR1062D04), *ADP glucose pyrophosphorylase* (EC:2.7.7.27 - SCRFAM1028H01), *fructokinase* (EC:2.7.1.4 - SCRFLR1012E06), *sucrose-phosphate phosphatase* (EC:3.1.3.24 - SCSBST3096E12), *beta-amylase* (EC:3.2.1.2 - SCUTAM2089E05) e a *alpha-amylase* (EC:3.2.1.1 - SCUTAM2089E05).

Com isso pode ser observado os prováveis metabólitos com correlações diretas a estas enzimas, que foram os *mz/rt* de 337,1734/622,57, 599,5197/691,71, 607,3091/220,81, 639,304/320,59 e 783,4114/353,47, que estão correlacionados positivamente com estas enzimas, enquanto o 1348,6552/681,45 está correlacionado negativamente (**figura 48**).

Figura 48: Correlações diretas das enzimas participantes do metabolismo do amido e sacarose na rede de integração da folha +1 de cana-de-açúcar. SCCCLR1001A05 = *sucrose synthase* (EC:2.4.1.13); SCQGLR1062D04 = *UDP glucose pyrophosphorylase* (EC:2.7.7.9); SCRFAM1028H01 = *ADP glucose pyrophosphorylase* (EC:2.7.7.27); SCRFLR1012E06 = *fructokinase* (EC:2.7.1.4); SCSBST3096E12 = *sucrose-phosphate phosphatase* (EC:3.1.3.24); SCUTAM2089E05 = *beta-amylase; alpha-amylase* (EC:3.2.1.2; EC:3.2.1.1).

Dentre estes prováveis metabólitos que estão altamente correlacionados com o metabolismo de amido e sacarose, três são *hubs* de grande importância para a rede: o hub_F1A3 (337,1734/622,57), hub_F1A4 (607,3091/220,81) e hub_F1A5 (783,4114/353,47), todos estão correlacionados positivamente com esta via metabólica.

4.4.2. Rede biológica do entrenó 5 com 11 e 13 meses após o plantio

Para a criação da rede biológica de correlações do entrenó 5, foi utilizado como parâmetros para a rCCA um λ_1 de 0,9796122 e um λ_2 de 0,2048776 resultando em um *cv-score* de 0,1245354, com isso foram obtidas três sub-redes: a sub-rede E5A (**figura 49**), que contém a maior parte de metabólitos e proteínas utilizados na construção da rede, com 498 nós, 1636 arestas e seis grandes *hubs*; a sub-rede E5B (**figura 50**) com 12 nós e 11 arestas e um *hub*; a sub-rede E5C (**figura 51**) com 10 nós e 13 arestas.

Figura 49: Sub-rede biológica de integração E5A dos metabólitos e proteínas do entrenó 5, onde os nós de cor azul são os metabólitos e os de cor cinza são as proteínas, também se encontram destacados os principais *hubs*. As cores das arestas indicam se a correlação é negativa (verde) ou positiva (vermelha).

Figura 50: Sub-rede biológica de integração E5B dos metabólitos e proteínas do entrenó 5, onde os nós de cor azul são os metabólitos e os de cor cinza são as proteínas, também se encontra destacado o *hub* presente na sub-rede. As cores das arestas indicam se a correlação é negativa (verde) ou positiva (vermelha).

Figura 51: Sub-rede biológica de integração E5C dos metabólitos e proteínas do entrenó 5, onde os nós de cor azul são os metabólitos e os de cor cinza são as proteínas. As cores das arestas indicam se a correlação é negativa (verde) ou positiva (vermelha).

Na **figura 49** é possível observar seis grandes *hubs* na sub-rede E5A que possuem a maior parte das correlações, esses seis principais *hubs* (hub_E5A1, hub_E5A2, hub_E5A3,

hub_E5A4, hub_E5A5, hub_E5A6) são os possíveis metabólitos com mz/rt de 158,1513/518,57, 439,1424/152,77, 561,1746/184,93, 163,0345/142,97, 377,052/133,93 e 289,0607/208,7, juntos eles estão correlacionados diretamente com 444 proteínas de um total de 498 nós na sub-rede E5A. O hub_E5A1 está correlacionado positivamente com 22 proteínas e negativamente com 200 proteínas e os processos biológicos das proteínas de acordo com a anotação são apresentados na **figura 52** (correlações positivas) e **figura 53** (correlações negativas). O hub_E5A2 está correlacionado positivamente com 58 proteínas e negativamente com 120 proteínas e os processos biológicos das proteínas de acordo com a anotação são apresentados na **figura 54** (correlações positivas) e **figura 55** (correlações negativas). O hub_E5A3 está correlacionado positivamente com 101 proteínas e negativamente com 107 proteínas e os processos biológicos das proteínas de acordo com a anotação são apresentados na **figura 56** (correlações positivas) e **figura 57** (correlações negativas). O hub_E5A4 está correlacionado positivamente com 88 proteínas e negativamente com 92 proteínas e os processos biológicos das proteínas de acordo com a anotação são apresentados na **figura 58** (correlações positivas) e **figura 59** (correlações negativas). O hub_E5A5 está correlacionado positivamente com 75 proteínas e negativamente com 79 proteínas e os processos biológicos das proteínas de acordo com a anotação são apresentados na **figura 60** (correlações positivas) e **figura 61** (correlações negativas). O hub_E5A6 está correlacionado positivamente com 76 proteínas e negativamente com 58 proteínas e os processos biológicos das proteínas de acordo com a anotação são apresentados na **figura 62** (correlações positivas) e **figura 63** (correlações negativas).

Figura 52: Processos biológicos das proteínas correlacionadas positivamente com o *hub* hub_E5A1 da sub-rede de correlações E5A de metabólitos e proteínas do entrenó 5 com 11 e 13 meses.

hub_E5A1 (correlação negativa)

Figura 53: Processos biológicos das proteínas correlacionadas negativamente com o *hub* hub_E5A1 da sub-rede de correlações E5A de metabólitos e proteínas do entrenó 5 com 11 e 13 meses.

Figura 54: Processos biológicos das proteínas correlacionadas positivamente com o *hub* hub_E5A2 da sub-rede de correlações E5A de metabólitos e proteínas do entrenó 5 com 11 e 13 meses.

Figura 55: Processos biológicos das proteínas correlacionadas negativamente com o *hub* hub_E5A2 da sub-rede de correlações E5A de metabólitos e proteínas do entrenó 5 com 11 e 13 meses.

Figura 56: Processos biológicos das proteínas correlacionadas positivamente com o *hub* hub_E5A3 da sub-rede de correlações E5A de metabólitos e proteínas do entrenó 5 com 11 e 13 meses.

Figura 57: Processos biológicos das proteínas correlacionadas negativamente com o *hub* hub_E5A3 da sub-rede de correlações E5A de metabólitos e proteínas do entrenó 5 com 11 e 13 meses.

Figura 58: Processos biológicos das proteínas correlacionadas positivamente com o *hub* hub_E5A4 da sub-rede de correlações E5A de metabólitos e proteínas do entrenó 5 com 11 e 13 meses.

Figura 59: Processos biológicos das proteínas correlacionadas negativamente com o *hub* hub_E5A4 da sub-rede de correlações E5A de metabólitos e proteínas do entrenó 5 com 11 e 13 meses.

Figura 60: Processos biológicos das proteínas correlacionadas positivamente com o *hub* hub_E5A5 da sub-rede de correlações E5A de metabólitos e proteínas do entrenó 5 com 11 e 13 meses.

Figura 61: Processos biológicos das proteínas correlacionadas negativamente com o *hub* hub_E5A5 da sub-rede de correlações E5A de metabólitos e proteínas do entrenó 5 com 11 e 13 meses.

Figura 62: Processos biológicos das proteínas correlacionadas positivamente com o *hub* hub_E5A6 da sub-rede de correlações E5A de metabólitos e proteínas do entrenó 5 com 11 e 13 meses.

Figura 63: Processos biológicos das proteínas correlacionadas negativamente com o *hub_E5A6* da sub-rede de correlações E5A de metabólitos e proteínas do entrenó 5 com 11 e 13 meses.

Os processos biológicos que mais se destacaram nas correlações positivas dos *hubs* *hub_E5A1*, *hub_E5A2*, *hub_E5A3*, *hub_E5A4*, *hub_E5A5*, *hub_E5A6* foram o processo de biossíntese de pigmentos (*pigment biosynthetic process*), proteólise (*proteolysis*), splicing do mRNA via spliceossomo (*mRNA splicing via spliceosome*), processo metabólico (*metabolic process*), resposta a proteína dobrada incorretamente (*response to misfolded protein*), processo de oxidação-redução (*oxydation-reduction process*), enovelamento de proteínas (*protein folding*). Já nas correlações negativas, em todos os *hubs* o processo que mais se

destacou foi o processo de oxidação-redução (*oxidation-reduction process*), que se sobressaiu amplamente sobre os demais processos, diferente da ampla variedade de processos correlacionados positivamente.

Na **figura 50** é possível observar uma pequena sub-rede (E5B) com um único *hub*, esse *hub* (hub_E5B1) é o possível metabólito com *mz/rt* de 762,0808/133,61. O hub_E5B1 está correlacionado positivamente com uma proteína e negativamente com 10 proteínas. O processo biológico da qual a proteína correlacionada positivamente participa é a tradução (*translation*), e os processos biológicos das correlacionadas negativamente são apresentados na **figura 64**.

Figura 64: Processos biológicos das proteínas correlacionadas negativamente com o *hub* hub_E5B1 da sub-rede de correlações E5B de metabólitos e proteínas do entrenó 5 com 11 e 13 meses.

O processo biológico que mais se destacou nas correlações negativas do *hub* hub_E5B1 foi o processo de oxidação-redução (*oxidation-reduction process*).

Na **figura 51** é possível observar uma pequena sub-rede (E5C) que não possui nenhum hub. Ela é composta por 4 possíveis metabólitos (146,053/107,15, 188,0534/107,1, 189,0694/107,06, 783,5372/697,42) e 6 proteínas. As proteínas que fazem parte dessa rede são as *thioredoxin reductase 2* (EC:1.8.1.9 - SCBGLR1120D04), *adenosine kinase 2* (EC:2.7.1.20 - SCCCLB1004H08), *sucrose synthase* (EC:2.4.1.13 - SCCCLR1001A05), *plasma membrane-associated cation-binding protein 1-like* (SCCCLR1079C11), *fructose-bisphosphate aldolase* (EC:4.1.2.13 - SCCCRZ2001F03), *t-complex protein 1 subunit delta* (SCSFLR2009H09), e o processo biológico de que participam de acordo com os resultados do Blast2GO (CONESA et al., 2005), são o processo de oxidação-redução, fosforilação, processo metabólico de sacarose e resposta celular ao estímulo.

Dentre as proteínas participantes da sub-rede E5C, somente uma está relacionada ao metabolismo de amido e sacarose, a *sucrose synthase* (EC:2.4.1.13 - SCCCLR1001A05), que está correlacionada negativamente com três possíveis metabólitos (146,053/107,15, 188,0534/107,1, 189,0694/107,06).

As enzimas participantes da via do metabolismo do amido e sacarose encontradas nesta rede de correlações foram a *pectinesterase* (EC:3.1.1.11 - SCCCFL6001E11), *beta-glucosidase* (EC:3.2.1.21 - SCCCLB1004E10), *sucrose synthase* (EC:2.4.1.13 - SCCCLR1001A05), *glucose phosphomutase* (EC:5.4.2.2 - SCCCRZ1001D04), *sucrose synthase* (EC:2.4.1.13 - SCCCRZ1002G07), *UDP-glucuronate decarboxylase* (EC:4.1.1.35 - SCCCRZ2001D09), *sucrose alpha-glucosidase* (EC:3.2.1.48 - SCCCST3006A04), *beta-fructofuranosidase* (EC:3.2.1.26 - SCCCST3006A04), *glucan 1,3-beta-glucosidase* (EC:3.2.1.58 - SCEQLR1093F09), *beta-glucosidase* (EC:3.2.1.21 - SCEZLB1007A09), *glucokinase* (EC:2.7.1.2 - SCEZLB1008B10), *fructokinase* (EC:2.7.1.4 - SCEZLB1008B10), *beta-xylosidase* (EC:3.2.1.37 - SCEZRZ3015A12), *UDP-glucose 6-dehydrogenase* (EC:1.1.1.22 - SCQGLR1019G02).

Os prováveis metabólitos com mz/rt de 146,053/107,15, 158,1513/518,57, 163,0345/142,97, 177,0441/148,43, 188,0534/107,1, 189,0694/107,06, 241,2001/430,89, 289,0607/208,7, 375,0523/91,05, 377,052/133,93, 439,1424/152,77, 485,077/231,03, 561,1746/184,93, 568,2669/498,33, 569,2835/498,33, 782,5292/695,53, estão correlacionados com as enzimas identificadas (**figura 65**).

Figura 65: Correlações diretas das enzimas participantes da via do metabolismo do amido e sacarose na rede de integração do entrenó 5 de cana-de-açúcar. SCCCF6001E11 = *pectinesterase* (EC:3.1.1.11); SCCCLB1004E10 = *beta-glucosidase* (EC:3.2.1.21); SCCCLR1001A05 = *sucrose synthase* (EC:2.4.1.13); SCCCRZ1001D04 = *glucose phosphomutase* (EC:5.4.2.2); SCCCRZ1002G07 = *sucrose synthase* (EC:2.4.1.13); SCCCRZ2001D09 = *UDP-glucuronate decarboxylase* (EC:4.1.1.35); SCCST3006A04 = *sucrose alpha-glucosidase*; *beta-fructofuranosidase* (EC:3.2.1.48; EC:3.2.1.26); SCEQLR1093F09 = *glucan 1,3-beta-glucosidase* (EC:3.2.1.58); SCEZLB1007A09 = *beta-glucosidase* (EC:3.2.1.21); SCEZLB1008B10 = *glucokinase*; *fructokinase* (EC:2.7.1.2; EC:2.7.1.4); SCEZRZ3015A12 = *beta-xylosidase* (EC:3.2.1.37); SCQGLR1019G02 = *UDP-glucose 6-dehydrogenase* (EC:1.1.1.22).

O processo de biossíntese de pigmentos são as reações químicas e vias metabólicas que resultam na produção de um pigmento (SMITH, 1997).

O processo de proteólise refere-se à degradação de uma proteína, normalmente por meio de hidrólise em uma ou mais ligações peptídicas (SMITH, 1997)

O processo de splicing do mRNA via spliceossomo é basicamente a remoção dos íntrons e junção dos exons a partir de um ou mais transcritos primários de mRNA mensageiro,

através de um mecanismo spliceossomo, produzindo assim um mRNA somente com exons (SMITH, 1997).

O processo biológico chamado processo metabólico abrange todas reações e as vias metabólicas que incluem o anabolismo e catabolismo. Os processos metabólicos incluem a transformação de pequenas moléculas, reparo e replicação de DNA e a síntese e degradação de proteínas (BLAKE et al., 2015)

O processo de resposta a proteína dobrada incorretamente é qualquer efeito ou reações em uma célula ou organismo vivo que tem como origem um estímulo por uma proteína com enovelamento incorreto (BLAKE et al., 2015).

O processo metabólico de oxidação-redução resulta na remoção ou adição de um ou mais elétrons para uma ou de uma substância, podendo ser realizado com ou sem a remoção ou adição de próton ou os prótons (BLAKE et al., 2015)

O processo de enovelamento de proteínas é a montagem covalente e não covalente de polipeptídeos de cadeia simples ou complexa em uma estrutura terciária correta (BLAKE et al., 2015)

A tradução é o processo metabólico pelo qual uma sequência específica de bases do RNA mensageiro (mRNA) determina uma sequência de aminoácidos durante a biossíntese de proteínas e de peptídeos (SMITH, 1997).

De acordo com os resultados obtidos dos processos biológicos da anotação do Blast2GO (CONESA et al., 2005), nenhum processo que se destacou nas correlações positivas e negativas dos *hubs* de todas as sub-redes possuem uma relação direta com o acúmulo de sacarose no entrenó. Já na análise do metabolismo de amido e sacarose nesta rede biológica, é possível observar vários possíveis metabólitos com fortes correlações (positivas e negativas) com estas enzimas (**figura 65**). Dentre todas estas correlações que podem estar influenciando o acúmulo de sacarose no entrenó 5, pode-se observar a importância da correlação negativa de três possíveis metabólitos (146,053/107,15, 188,0534/107,1, 189,0694/107,06) com a *sucrose synthase* (EC:2.4.1.13 - SCCCLR1001A05), que é uma enzima de grande importância na produção de sacarose (BOTHÁ; BLACK, 2000; SUI et al., 2015). As correlações com a *sucrose synthase* foram encontradas na sub-rede E5C.

4.4.3. Rede biológica do entrenó 9 com 11 e 13 meses após o plantio

Para a criação da rede biológica de correlações do entrenó 9, foi utilizado como parâmetros para a rCCA um λ_1 de 0,004098 e um λ_2 de 0,052074 resultando em um valor de validação cruzada (*cv-score*) de 0.9932612, com isso foram obtidas duas sub-redes: a sub-rede E9A (**figura 66**) que destaca a presença de dois *clusters* contendo a maioria dos

metabólitos e proteínas utilizados para a construção das redes, com 87 nós, 372 arestas; a sub-rede E9B (**figura 67**) com 24 nós e 23 arestas e um *hub*.

Figura 66: Sub-rede biológica de integração E9A dos metabólitos e proteínas do entrenó 9, onde os nós de cor azul são os metabólitos e os de cor cinza são as proteínas, também se encontram destacados os principais *hubs*. As cores das arestas indicam se a correlação é negativa (verde) ou positiva (vermelha).

Figura 67: Sub-rede biológica de integração E9B dos metabólitos e proteínas do entrenó 9, onde os nós de cor azul são os metabólitos e os de cor cinza são as proteínas, também se encontra destacado o único *hub*. As cores das arestas indicam se a correlação é negativa (verde) ou positiva (vermelha).

Na **figura 66** é possível observar que a sub-rede E9A possui dois *cluster*. O *cluster* da **figura 66-A** possui o maior *hub* dessa sub-rede. Este *hub* (hub_E9A1) é o possível metabólito com *mz/rt* de 158,1549/518,34. O hub_E9A1 está correlacionado positivamente com 23 proteínas e negativamente com cinco proteínas e os processos biológicos das proteínas de acordo com a anotação são apresentados na **figura 68** (correlações positivas) e **figura 69** (correlações negativas). O *cluster* da **figura 66-B** possui um único *hub* (hub_E9A2), onde este *hub* é o possível metabólito com *mz/rt* de 365,1357/688,74. O hub_E9A2 está correlacionado positivamente com somente uma proteína e negativamente com 15 proteínas. O processo biológico da qual a proteína correlacionada positivamente participa é o de resposta a proteína dobrada incorretamente (*response to misfolded protein*), e os processos biológicos das correlacionadas negativamente são apresentados na **figura 70**.

Figura 68: Processos biológicos das proteínas correlacionadas positivamente com o *hub* hub_E9A1 da sub-rede de correlações E9A de metabólitos e proteínas do entrenó 9 com 11 e 13 meses.

Figura 69: Processos biológicos das proteínas correlacionadas negativamente com o *hub* hub_E9A1 da sub-rede de correlações E9A de metabólitos e proteínas do entrenó 9 com 11 e 13 meses.

Figura 70: Processos biológicos das proteínas correlacionadas negativamente com o *hub* hub_E9A2.

O processo biológico que mais se destacou nas correlações positivas do *hub* hub_E9A1 foi o processo de oxidação-redução (*oxidation-reduction process*). Nas correlações negativas, nenhum processo biológico se destacou dos demais. Para o *hub* Hub_E9A2 o processo biológico que mais se destacou nas correlações negativas foi o de tradução (*translation*).

Na **figura 67** é possível observar uma pequena sub-rede (E9B) com um único *hub* (hub_E9B1), o metabólito *montecristin*, que pertence à classe dos acil graxos com *mz/rt* de 575,489/695,73, este metabólito possivelmente está envolvido na biossíntese de acetogeninas (GLEYE et al., 1997).

Este metabólito *hub* foi o VIP de número um na análise do perfil metabólico por PLS-DA do entrenó 9 com 11 e 13 meses após o plantio, ou seja, foi o metabólito que mais influenciou na variância destes dois tempos de coleta.

O hub_E9B1 está correlacionado positivamente com 22 proteínas e negativamente com apenas uma proteína. O processo biológico da qual a proteína correlacionada

negativamente participa é o de respiração aeróbica (*aerobic respiration*), e os processos biológicos das correlacionadas positivamente são apresentados na **figura 71**.

Figura 71: Processos biológicos das proteínas correlacionadas positivamente com o *hub* hub_E9B1 da sub-rede de correlações E9B de metabólitos e proteínas do entrenó 9 com 11 e 13 meses.

Os processos biológicos que mais se destacaram nas correlações positivas do *hub* hub_E9B1 foram o processo catabólico ubiquitina-dependente (*ubiquitin-dependent catabolic process*) e o de enovelamento de proteína (*protein folding*).

Dentre as enzimas participantes da via do metabolismo de amido e sacarose, somente uma foi encontrada na rede de integração do entrenó 9 de cana-de-açúcar, que é a *glucan 1,3-beta-glucosidase* (EC:3.2.1.58 - SCEQLB1066E08), e os possíveis metabólitos com *mz/rt* de 119,0829/610,62, 149,0585/622,54, 158,1549/518,34, 209,1488/607,5, 210,1623/607,62, 241,2066/430,34, 325,2228/454,46, 331,0793/464,73, 415,1923/609,36, 416,207/608,74,

437,1739/608,74, 480,2161/251,3, 480,2162/264,65, 557,1596/522,44, estão correlacionados negativamente com esta enzima (**figura 72**).

Figura 72: Correlações diretas das enzimas participantes da via do metabolismo do amido e sacarose na rede de integração do entrenó 9 de cana-de-açúcar com 11 e 13 meses após o plantio.

O processo de enovelamento de proteínas é a montagem covalente e não covalente de polipeptídeos de cadeia simples ou complexa em uma estrutura terciária correta (BLAKE et al., 2015)

O processo metabólico de oxidação-redução resulta na remoção ou adição de um ou mais elétrons para uma ou de uma substância, podendo ser realizado com ou sem a remoção ou adição de próton ou os prótons (BLAKE et al., 2015)

A tradução é o processo metabólico pelo qual uma sequência específica de bases do RNA mensageiro (mRNA) determina uma sequência de aminoácidos durante a biossíntese de proteínas e de peptídeos (SMITH, 1997).

O processo catabólico ubiquitina-dependente consiste de reações bioquímicas que resultam na degradação de uma proteína ou péptido (hidrólise), sendo iniciado pela ligação covalente de um grupo de ubiquitina ou vários grupos de ubiquitina (BLAKE et al., 2015).

Com base nos processos biológicos que mais se sobressaíram nos *hubs* analisados em todos as sub-redes, não foi possível identificar uma relação direta com o acúmulo de

sacarose. Já na análise das enzimas que participam da via metabólica do amido e sacarose, foi possível identificar uma enzima, a *glucan 1,3-beta-glucosidase* (EC:3.2.1.58 - SCEQLB1066E08) que possui correlações negativa diretas com 14 possíveis metabólitos que podem estar associados ao acúmulo de sacarose no entrenó 9 de cana-de-açúcar.

5. CONSIDERAÇÕES FINAIS

A quercetina mostrou ser um ótimo padrão interno para a metabolômica utilizando-se LC-MS, pois demonstrou estabilidade tanto em seu valor de m/z como em seu r_t durante as corridas cromatográficas. Levando em conta que para sua funcionalidade a amostra complexa não pode conter naturalmente este composto. Assim ela pode ser usada para a normalização da intensidade dos prováveis metabólitos e verificação da qualidade de extração dos metabólitos.

A cana-de-açúcar SP80-3280 apresenta perfis metabólicos distintos durante seu desenvolvimento e acúmulo de sacarose, essa diferença pode ser observada em todos tecidos estudados, folha +1, entrenó 1, entrenó 5 e entrenó 9.

Na folha +1 os principais metabólitos responsáveis pela diferença no perfil metabólico pertencem a classe dos glicerolipídeos e glicerofosfolipídeos, que apresentaram uma maior intensidade aos 11 meses e a menor intensidade com 13 meses, apesar dessas classes participarem da estabilização dos componentes do fotossistema II e como constituinte da membrana celular, respectivamente, eles podem estar relacionados com o acúmulo de sacarose durante o desenvolvimento da cana-de-açúcar, pelo fato da cana ter um nível maior de açúcar aos 13 meses de idade.

No entrenó 1 o metabólito que mais influenciou na diferença do perfil metabólico, durante o desenvolvimento da planta e acúmulo de sacarose, pertence à classe metabólica das cumarinas e derivados, que possuem diversas bioatividades, podendo atuar desde o crescimento normal de plantas a resposta a diferentes tipos de estresses. Como já observado por diversos autores existe um aumento da expressão de proteínas relacionadas ao estresse durante o acúmulo de sacarose, com isso, apesar do tecido do entrenó 1 ser jovem é bem provável que este metabólito esteja relacionado ao desenvolvimento e acúmulo de sacarose na cana-de-açúcar.

No entrenó 5 o metabólito que mais influenciou na diferença do perfil metabólico durante o desenvolvimento da planta e acúmulo de sacarose pertence à classe metabólica dos esteroides e derivados de esteroides. Esta classe de compostos está relacionada com várias funções conservadas em plantas, como o auxílio ao desenvolvimento e aceleração do envelhecimento. Considerando as funções dos esteroides em plantas e a sua maior abundância no tecido mais velho (13 meses após o plantio), este metabólito pode estar relacionado ao desenvolvimento desse entrenó, indiretamente preparando este tecido para um maior acúmulo de sacarose, pois a concentração de sacarose é menor em entrenós imaturos, mas tende a aumentar durante o desenvolvimento.

Já no entrenó 9 o metabólito que mais influenciou na diferença do perfil metabólico durante o desenvolvimento da planta e acúmulo de sacarose foi a *Montecristin*, que pertence à classe metabólica dos acil graxos. Este metabólito não está relacionado diretamente com o acúmulo de açúcares em planta, mas apresentou maior intensidade com 13 meses após o plantio, consequentemente a identificação de metabólitos que aumentam sua abundância conjuntamente com o aumento da sacarose na planta, pode levar a um melhor entendimento do acúmulo de sacarose em cana-de-açúcar.

A cana-de-açúcar SP80-3280 também apresenta perfis proteicos distintos durante seu desenvolvimento e acúmulo de sacarose, essa diferença pode ser observada em todos tecidos estudados, folha +1, entrenó 5 e entrenó 9.

Na folha +1 da cana-de-açúcar destacaram-se duas histonas entre as proteínas que mais influenciaram nas diferenças entre os tempos de coletas de 11 e 13 meses após o plantio. As histonas identificadas tiveram diminuição em sua abundância aos 13 meses, fato que já foi relatado na literatura como relacionado à maturação da cana-de-açúcar, portanto as histonas são prováveis proteínas relacionadas ao acúmulo de sacarose.

No entrenó 5 foi observado a presença da proteína *cold acclimation protein* entre as que mais influenciaram na diferença do perfil proteico nesse tecido. Esta proteína está relacionada com estresse ao frio e teve uma maior abundância no tecido mais velho. Como diversos trabalhos citam proteínas relacionada a estresse durante o acúmulo de sacarose, provavelmente esta proteína está relacionada também com o acúmulo de sacarose.

Já no entrenó 9, a proteína que mais influenciou nas diferenças do perfil proteico entre os tempos de coleta de 11 e 13 meses após o plantio foi a *membrane-associated cation-binding protein 1 (PCaP1)*, apesar desta proteína não ter uma função biológica definida, é possível que ela esteja relacionada ao acúmulo de sacarose, pois foi a principal proteína (VIP) que influenciou na projeção da PLS-DA do entrenó 9 da cana-de-açúcar, com uma maior intensidade no entrenó mais jovem.

Na análise do proteoma da cana-de-açúcar, foi observado a presença de histonas em todos os tecidos analisados (folha +1, entrenó 5 e entrenó 9) e entre os cinco principais VIPs que influenciaram nas diferenças entre os tempos de coleta de 11 e 13 meses. Em todas as análises as histonas mantiveram seu padrão de abundância, diminuindo-a no tecido mais velho. Como esse comportamento das histonas já foi citado na literatura, então provavelmente elas estão relacionadas com o acúmulo de sacarose na cana-de-açúcar.

Também é possível observar a presença de uma histona (*histone h4-like* - SCAGLR2026F05) entre os cinco primeiros VIPs, apresentando uma intensidade maior no entrenó mais novo. Essa proteína foi encontrada entre os cinco primeiros VIPs em todos os tecidos de cana-de-açúcar analisados (folha +1, entrenó 5 e entrenó 9), sempre apresentando

uma redução em seu nível de intensidade no tecido mais velho, reforçando assim a relação e a importância da histona h4 com o acúmulo de sacarose já discutido nas análises da folha +1 e do entrenó 5, onde também foi presenciado por outros trabalhos (CASU et al., 2004, 2007).

Na rede de correlações entre os metabólitos e proteínas da folha +1, foi identificado o *hub* o hub_F1A1, que é o possível metabólito com valor de m/z e r_t de 806,4367/358,65, que pode estar influenciando positivamente a produção de sacarose e consequentemente o seu acúmulo nos entrenós, pois ele possui uma forte correlação positiva com várias proteínas que fazem parte do processo biológico da fotossíntese. Também foi observado possíveis metabólitos que possuem correlação direta com enzimas participantes da via do metabolismo de amido e sacarose. Esses metabólitos possuem valor de m/z e r_t de 337,1734/622,57, 599,5197/691,71, 607,3091/220,81, 639,304/320,59 e 783,4114/353,47, que estão correlacionados positivamente e o 1348,6552/681,45, que está correlacionado negativamente. Portanto são possíveis metabólitos que provavelmente influenciam no acúmulo de sacarose da cana-de-açúcar.

Na rede de correlações entre os metabólitos e proteínas do entrenó 5, foi identificado a correlação negativa direta de três possíveis metabólitos (146,053/107,15, 188,0534/107,1, 189,0694/107,06) com a *sucrose synthase* (EC:2.4.1.13 - SCCCLR1001A05), que é uma enzima de grande importância na produção de sacarose, portanto esses possíveis metabólitos podem estar associados com o acúmulo de sacarose da cana-de-açúcar.

Na rede de correlações entre os metabólitos e proteínas do entrenó 9, foi possível identificar 14 possíveis metabólitos correlacionados negativamente com a enzima *glucan 1,3-beta-glucosidase* (EC:3.2.1.58 - SCEQLB1066E08), esta enzima faz parte do metabolismo de amido e sacarose. Os valores de m/z e r_t dos possíveis metabólitos que estão correlacionados com esta enzima são 119,0829/610,62, 149,0585/622,54, 158,1549/518,34, 209,1488/607,5, 210,1623/607,62, 241,2066/430,34, 325,2228/454,46, 331,0793/464,73, 415,1923/609,36, 416,207/608,74, 437,1739/608,74, 480,2161/251,3, 480,2162/264,65, 557,1596/522,44. Como são correlações altas e de forma direta, estes possíveis metabólitos podem estar associados ao acúmulo de sacarose no entrenó da cana-de-açúcar.

O método de integração de dados por rCCA se mostrou bastante eficiente para a integração do metaboloma e proteoma da cana-de-açúcar, possibilitando uma visão holística entre os dois conjuntos de dados, favorecendo a identificação de correlações entre metabólitos e proteínas durante o desenvolvimento e acúmulo de sacarose nesta planta.

Os resultados apresentados neste trabalho servirão de base para estudos mais detalhados e direcionados a respeito das vias de regulação do acúmulo temporal e espacial de sacarose em cana-de-açúcar.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, out. 1990.

AMALRAJ, R. S. et al. Sugarcane proteomics: Establishment of a protein extraction method for 2-DE in stalk tissues and initiation of sugarcane proteome reference map. **Electrophoresis**, v. 31, n. 12, p. 1959–1974, 18 jun. 2010.

ANDERSON, N. L.; ANDERSON, N. G. Proteome and proteomics: new technologies, new concepts, and new words. **Electrophoresis**, v. 19, n. 11, p. 1853–1861, 1998.

BANIASADI, H. et al. Effect of environment and genotype on commercial maize hybrids using LC/MS-based metabolomics. **Journal of Agricultural and Food Chemistry**, v. 62, n. 6, p. 1412–1422, 2014.

BELYAEV, N. D. et al. Histone H4 acetylation in plant heterochromatin is altered during the cell cycle. **Chromosoma**, v. 106, n. 3, p. 193–197, 1997.

BENTON, H. P.; WANT, E. J.; EBBELS, T. M. D. Correction of mass calibration gaps in liquid chromatography-mass spectrometry metabolomics data. **Bioinformatics (Oxford, England)**, v. 26, n. 19, p. 2488–9, 1 out. 2010.

BICKEL, P. J.; LI, B. Regularization in statistics. **Test**, v. 15, n. 2, p. 271–303, 2006.

BIELESKI, R. L. The Physiology of sugar-cane III. Characteristics of Sugar Uptake in Slices of Mature and Immature Storage Tissue. **Australian Journal of Biological Sciences**, v. 13, n. 3, p. 203 – 220, 1960.

BLAKE, J. A. et al. Gene ontology consortium: Going forward. **Nucleic Acids Research**, v. 43, n. D1, p. D1049–D1056, 2015.

BLOCK, M. A. et al. Chloroplast envelope membranes: A dynamic interface between plastids and the cytosol. **Photosynthesis Research**, v. 92, n. 2, p. 225–244, 2007.

BOTHA, F. C.; BLACK, K. G. Sucrose phosphate synthase and sucrose synthase activity during maturation of internodal tissue in sugarcane. **Australian Journal of Plant Physiology**, v. 27, n. 1, p. 81–85, 2000.

BOUDIÈRE, L. et al. Glycerolipids in photosynthesis: Composition, synthesis and trafficking. **Biochimica et Biophysica Acta - Bioenergetics**, v. 1837, n. 4, p. 470–480, 2014.

BRADFORD, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. **Analytical Biochemistry**, v. 72, n. 1-2, p. 248–254, 1976.

- BUDZINSKI, I. G. F. et al. Proteomics. In: FRITSCHÉ-NETO, R.; BORÉM, A. (Eds.). . **Omics in Plant Breeding**. 1ed. ed. Hoboken: Wiley-Blackwell, 2014. p. 59–80.
- BUTTERFIELD, M. K.; THOMAS, D. W. **Sucrose , yield and disease resistance characteristics of sugarcane varieties under test in the sasex selection programme**. Proceedings of African Sugarcane Technology Association. **Anais...**1996
- CALDERAN-RODRIGUES, M. J. et al. 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 Biology**, v. 16, n. 1, p. 14, 2016.
- CANDIANO, G. et al. Blue silver: A very sensitive colloidal Coomassie G-250 staining for proteome analysis. **Electrophoresis**, v. 25, n. 9, p. 1327–1333, 2004.
- CASU, R. E. et al. Identification of differentially expressed transcripts from maturing stem of sugarcane by in silico analysis of stem expressed sequence tags and gene expression profiling. **Plant Molecular Biology**, v. 54, n. 4, p. 503–517, 2004.
- CASU, R. E. et al. Identification of transcripts associated with cell wall metabolism and development in the stem of sugarcane by Affymetrix GeneChip Sugarcane Genome Array expression profiling. **Functional & integrative genomics**, v. 7, n. 2, p. 153–67, abr. 2007.
- CHEAVEGATTI-GIANOTTO, A. et al. Sugarcane (*Saccharum X officinarum*): A Reference Study for the Regulation of Genetically Modified Cultivars in Brazil. **Tropical Plant Biology**, v. 4, n. 1, p. 62–89, 2011.
- CHEBROLU, K. K. et al. Impact of heat stress during seed development on soybean seed metabolome. **Metabolomics**, v. 12, n. 2, p. 1–14, 2016.
- CHEN, S.; HARMON, A. C. Advances in plant proteomics. **Proteomics**, v. 6, n. 20, p. 5504–5516, 2006.
- CONAB - Companhia Nacional de Abastecimento. **Acompanhamento da safra brasileira de cana-de-açúcar - quarto levantamento safra 2015/16**. Disponível em: <<http://www.conab.gov.br>>. Acesso em: 15 abr. 2016.
- CONESA, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, v. 21, n. 18, p. 3674–3676, 15 set. 2005.
- COWAN, A. K. Phospholipids as plant growth regulators. **Plant Growth Regulation**, v. 48, n. 2, p. 97–109, 2006.
- CUADROS-INOSTROZA, A. et al. GC–MS metabolic profiling of Cabernet Sauvignon and Merlot cultivars during grapevine berry development and network analysis reveals a stage- and cultivar-dependent connectivity of primary metabolites. **Metabolomics**, v. 12, n. 2, p. 39,

2016.

DANIELS, J.; ROACH, B. T. Taxonomy and Evolution. In: **Sugarcane Improvement through Breeding**. [s.l.] Elsevier B.V., 1987. v. 11p. 7–84.

DE VOS, R. C. H. et al. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. **Nature protocols**, v. 2, n. 4, p. 778–91, jan. 2007.

DELMOTTE, N. et al. An integrated proteomics and transcriptomics reference data set provides new insights into the *Bradyrhizobium japonicum* bacteroid metabolism in soybean root nodules. **Proteomics**, v. 10, n. 7, p. 1391–400, abr. 2010.

DESHMANE, S. S.; DEV, S. Higher isoprenoids—II. **Tetrahedron**, v. 27, n. 6, p. 1109–1118, jan. 1971.

DETTMER, K.; ARONOV, P. A.; HAMMOCK, B. D. Mass spectrometry-based metabolomics. **Mass Spectrometry Reviews**, v. 26, n. 1, p. 51–78, 2007.

DI VENTURA, B. et al. From in vivo to in silico biology and back. **Nature**, v. 443, n. 7111, p. 527–33, 5 out. 2006.

DIXON, R.; PAIVA, N. Stress-Induced Phenylpropanoid Metabolism. **The Plant cell**, v. 7, n. 7, p. 1085–1097, 1995.

DONG, J.; HORVATH, S. Understanding network concepts in modules. **BMC systems biology**, v. 1, n. 1, p. 24, 2007.

DUNN, W. B. et al. Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. **Metabolomics**, v. 9, n. S1, p. 44–66, 26 mar. 2013.

FAHY, E. et al. A comprehensive classification system for lipids. **Journal of lipid research**, v. 46, p. 839–861, 2005.

FELSENFELD, G. Chromatin. **Nature**, v. 271, n. 5641, p. 115–22, 1978.

FENN, J. B. et al. Tandem mass spectrometry. **Science**, v. 214, n. 4518, p. 280–287, 1981.

FERCHA, A. et al. Shotgun proteomic analysis of soybean embryonic axes during germination under salt stress. **Proteomics**, p. 1–10, 2016.

FIEHN, O.; WECKWERTH, W. Deciphering metabolic networks. **European Journal of Biochemistry**, v. 270, n. 4, p. 579–588, 2003.

GIAVALISCO, P. et al. Elemental formula annotation of polar and lipophilic metabolites using (13) C, (15) N and (34) S isotope labelling, in combination with high-resolution mass spectrometry. **The Plant journal : for cell and molecular biology**, v. 68, n. 2, p. 364–76, out. 2011.

2011.

GILAR, M. et al. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. **Journal of Separation Science**, v. 28, n. 14, p. 1694–1703, 2005.

GLASSOP, D. et al. Changes in the sugarcane metabolome with stem development. Are they related to sucrose accumulation? **Plant & cell physiology**, v. 48, n. 4, p. 573–84, abr. 2007.

GLEYE, C. et al. Isolation of montecristin, a key metabolite in biogenesis of acetogenins from *Annona muricata* and its structure elucidation by using tandem mass spectrometry. **Journal of Organic Chemistry**, v. 62, n. 3, p. 510–513, 1997.

GLINSKI, M.; WECKWERTH, W. The role of mass spectrometry in plant systems biology. **Mass Spectrometry Reviews**, v. 25, n. 2, p. 173–214, 2006.

GONZALEZ, I. et al. CCA: An R Package to Extend Canonical Correlation Analysis. **Journal Of Statistical Software**, v. 23, n. 12, p. 1–14, 2008.

GONZÁLEZ, I. et al. Visualising associations between paired “omics” data sets. **BioData mining**, v. 5, n. 1, p. 19, 2012.

GRIVET, L.; ARRUDA, P. Sugarcane genomics: depicting the complex genome of an important tropical crop. **Current opinion in plant biology**, v. 5, n. 2, p. 122–7, abr. 2002.

GROF, C. P. L.; CAMPBELL, J. A. Sugarcane sucrose metabolism: scope for molecular manipulation. **Australian Journal of Plant Physiology**, v. 28, n. 1, p. 1–12, 2001.

GULLBERG, J. et al. Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of *Arabidopsis thaliana* samples in metabolomic studies with gas chromatography/mass spectrometry. **Analytical biochemistry**, v. 331, n. 2, p. 283–295, 15 ago. 2004.

GUO, K. et al. An integrated genomic and metabolomic framework for cell wall biology in rice. **BMC genomics**, v. 15, n. 1, p. 596, 2014.

HAN, P. et al. Proteomic analysis of heterosis in the leaves of sorghum–sudangrass hybrids. **Acta Biochimica et Biophysica Sinica**, v. 48, n. 2, p. 161–173, 2016.

HAN, X.; ASLANIAN, A.; YATES, J. R. Mass spectrometry for proteomics. **Current Opinion in Chemical Biology**, v. 12, n. 5, p. 483–490, 2008.

HARBORNE, J. B. Classes and functions of secondary products from plants. In: **Chemicals from plants**. [s.l.] Imperial College Press London, 1999. p. 1–25.

HARRABI, S. et al. Liquid chromatographic-mass spectrometric analysis of

- glycerophospholipids in corn oil. **Food Chemistry**, v. 114, n. 2, p. 712–716, 2009.
- HARTWIG, T.; WANG, Z.-Y. The molecular circuit of steroid signalling in plants. **Essays In Biochemistry**, v. 58, n. 0, p. 71–82, 2015.
- HASHEMITABAR, M.; KOLAHI, M.; TABANDEH, M. R. cDNA cloning , Phylogenic Analysis and Gene Expression Pattern of Phenylalanine ammonia-lyase in Sugarcane (*Saccharum officinarum* L .). v. 57, n. August, p. 456–465, 2014.
- HEFTMANN, E. Functions of steroids in plants. **Phytochemistry**, v. 14, n. 4, p. 891–901, 1975.
- HERINGER, A. S. et al. Label-Free Quantitative Proteomics of Embryogenic and Non-Embryogenic Callus during Sugarcane Somatic Embryogenesis. **PLoS ONE**, v. 10, n. 6, p. e0127803, 2015.
- HONG, S. et al. Canonical correlation analysis for RNA-seq co-expression networks. **Nucleic Acids Research**, v. 41, n. 8, p. 1–15, 2013.
- HOTELLING, H. Relations between two sets of variates. **Biometrika**, v. 28, n. 3-4, p. 321–377, dez. 1936.
- HU, C. et al. Metabolic variation between *japonica* and *indica* rice cultivars as revealed by non-targeted metabolomics. **Scientific reports**, v. 4, p. 5067, 2014.
- HUBER, S. C.; HUBER, J. L. Role of sucrose-phosphate synthase in sucrose metabolism in leaves. **Plant physiology**, v. 99, n. 4, p. 1275–1278, 1992.
- IDE, Y. et al. Molecular properties of a novel, hydrophilic cation-binding protein associated with the plasma membrane. **Journal of Experimental Botany**, v. 58, n. 5, p. 1173–1183, 2007.
- JOUHET, J.; MARÉCHAL, E.; BLOCK, M. A. Glycerolipid transfer for the building of membranes in plant cells. **Progress in Lipid Research**, v. 46, n. 1, p. 37–55, 2007.
- KANEHISA, M. et al. KEGG as a reference resource for gene and protein annotation. **Nucleic Acids Research**, v. 44, n. D1, p. D457–D462, 2016.
- KARAS, M.; HILLENKAMP, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. **Analytical chemistry**, v. 60, n. 20, p. 2299–2301, 1988.
- KARNOVSKY, A. et al. Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. **Bioinformatics**, v. 28, n. 3, p. 373–380, 30 nov. 2011.
- KASPAR-SCHOENEFELD, S. et al. Label-free proteome profiling reveals developmental-dependent patterns in young barley grains. **Journal of Proteomics**, 2016.
- KATAJAMAA, M.; ORESIC, M. Data processing for mass spectrometry-based metabolomics.

Journal of chromatography. A, v. 1158, n. 1-2, p. 318–28, 27 jul. 2007.

KATO, M. et al. PCaPs, possible regulators of PtdInsP signals on plasma membrane. **Plant signaling & behavior**, v. 5, n. 7, p. 848–850, 2010.

KERWIN, J. L.; TUININGA, A. R.; ERICSSONT, L. H. Identification of molecular species of glycerophospholipids and sphingomyelin using electrospray mass spectrometry. **Journal of lipid research**, v. 35, p. 1102–1114, 1994.

KIM, M. et al. Quantitative Shotgun Proteomics Analysis of Rice Anther Proteins after Exposure to High Temperature. **International Journal of Genomics**, v. 2015, p. 1–9, 2015.

KOTTAPALLI, K. R. et al. Shotgun Label-Free Quantitative Proteomics of Water-Deficit-Stressed Midmature Peanut (*Arachis hypogaea* L.) Seed. **Journal of Proteome Research**, v. 12, n. 11, p. 5048–5057, nov. 2013.

KUSANO, M. et al. Unbiased characterization of genotype-dependent metabolic regulations by metabolomic approach in *Arabidopsis thaliana*. **BMC systems biology**, v. 1, n. 53, p. 1–17, jan. 2007.

LAEMMLI, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. **Nature**, v. 227, n. 5259, p. 680–685, 15 ago. 1970.

LAM, E. et al. Improving sugarcane for biofuel: engineering for an even better feedstock. **Global Change Biology Bioenergy**, v. 1, n. 3, p. 251–255, jun. 2009.

LE CAO, K. A. et al. Sparse canonical methods for biological data integration: application to a cross-platform study. **BMC Bioinformatics**, v. 10, p. 34, 2009.

LÊ CAO, K.-A. A.; GONZÁLEZ, I.; DÉJEAN, S. integrOmics: an R package to unravel relationships between two omics datasets. **Bioinformatics**, v. 25, n. 21, p. 2855–2856, 1 nov. 2009.

LEMBKE, C. G. et al. **Growth and maturation of sugarcane SP80-3280 in two different seasons**. XXIII International Plant & Animal Genome Conference. **Anais...**San Diego: 2015

LUNN, J. E.; FURBANK, R. T. Sucrose biosynthesis in C4 plants. **New Phytol**, v. 143, n. 105, p. 221–237, 1999.

MANN, G. W. et al. Proteome coverage of the model plant *Arabidopsis thaliana*: Implications for shotgun proteomic studies. **Journal of Proteomics**, v. 79, p. 195–199, 2013.

MATSUDA, F. et al. MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. **The Plant journal : for cell and molecular biology**, v. 57, n. 3, p. 555–77, fev. 2009.

MCCORMICK, A. J.; CRAMER, M. D.; WATT, D. A. Culm sucrose accumulation promotes physiological decline of mature leaves in ripening sugarcane. **Field Crops Research**, v. 108, n. 3, p. 250–258, 2008.

MCCORMICK, A. J.; CRAMER, M. D.; WATT, D. A. Strength Regulates in Sugarcane. **New Phytologist**, v. 171, n. 4, p. 759–770, 2011.

MING, R. et al. Construction of a Consensus Genetic Map from Two Interspecific Crosses. **Crop science**, v. 42, p. 570–583, 2002.

MOHAMMADI, P. P.; MOIENI, A.; KOMATSU, S. Comparative proteome analysis of drought-sensitive and drought-tolerant rapeseed roots and their hybrid F1 line under drought stress. **Amino Acids**, v. 43, n. 5, p. 2137–2152, 2012.

MONTASTIER, E. et al. System Model Network for Adipose Tissue Signatures Related to Weight Changes in Response to Calorie Restriction and Subsequent Weight Maintenance. **PLOS Computational Biology**, v. 11, n. 1, p. e1004047, 2015.

MOORE, P. H. Temporal and Spatial Regulation of Sucrose Accumulation in the Sugarcane Stem. **Australian Journal of Plant Physiology**, v. 22, n. 4, p. 661, 1995.

MOORE, P. H. Integration of sucrose accumulation processes across hierarchical scales: towards developing an understanding of the gene-to-crop continuum. **Field Crops Research**, v. 92, n. 2-3, p. 119–135, jun. 2005.

MOORE, P. H.; COSGROVE, D. J. Developmental-Changes in Cell and Tissue Water Relations Parameters in Storage Parenchyma of Sugarcane. **Plant Physiology**, v. 96, n. 3, p. 794–801, 1991.

MORAES, I. et al. Analysis of histones H3 and H4 reveals novel and conserved post-translational modifications in sugarcane. **PLoS ONE**, v. 10, n. 7, p. 1–25, 2015.

MORINE, M. J. et al. Bi-directional gene set enrichment and canonical correlation analysis identify key diet-sensitive pathways and biomarkers of metabolic syndrome. **BMC bioinformatics**, v. 11, n. 1, p. 499, 2010.

NAGASAKI, N.; TOMIOKA, R.; MAESHIMA, M. A hydrophilic cation-binding protein of *Arabidopsis thaliana*, AtPCaP1, is localized to plasma membrane via N-myristoylation and interacts with calmodulin and the phosphatidylinositol phosphates PtdIns(3,4,5)P3 and PtdIns(3,5)P2. **FEBS Journal**, v. 275, n. 9, p. 2267–2282, 2008.

NAGATA, C. et al. A novel-type phosphatidylinositol phosphate-interactive, Ca-binding protein PCaP1 in *Arabidopsis thaliana*: stable association with plasma membrane and partial involvement in stomata closure. **Journal of Plant Research**, p. 1–12, 15 mar. 2016.

- NEILSON, K. A. et al. Less label, more free: Approaches in label-free quantitative mass spectrometry. **Proteomics**, v. 11, n. 4, p. 535–553, 2011.
- NELSON, D. L.; COX, M. M. **Lehninger Principles of Biochemistry**. 4 th ed. [s.l.] W. H. Freeman, 2012.
- NISHIYAMA, M. Y. et al. Full-Length Enriched cDNA Libraries and ORFeome Analysis of Sugarcane Hybrid and Ancestor Genotypes. **PLoS ONE**, v. 9, n. 9, p. e107351, 2014.
- OGATA, H. et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic acids research**, v. 27, n. 1, p. 29–34, 1 jan. 1999.
- OLIVEIRA, B. M.; COORSEN, J. R.; MARTINS-DE-SOUZA, D. 2DE: The Phoenix of Proteomics. **Journal of Proteomics**, v. 104, p. 140–150, 2014.
- PAPINI-TERZI, F. S. et al. Sugarcane genes associated with sucrose content. **BMC genomics**, v. 10, p. 120, jan. 2009.
- PATEL, P. V et al. Analysis of membrane proteins in metagenomics: Networks of correlated environmental features and protein families. **Genome Research**, v. 20, n. 7, p. 960–971, 1 jul. 2010.
- PAUL, M. J.; FOYER, C. H. Sink regulation of photosynthesis. **Journal of experimental botany**, v. 52, n. 360, p. 1383–1400, 2001.
- PENNINGTON, S. R.; DUNN, M. J. Proteomics: from protein sequence to function. In: TWYMAN, R. M. (Ed.). . **Principles of proteomics**. New York: Springer-Verlag: BIOS Scientific, 2001. p. 1–21.
- PIWOWAR, M.; JURKOWSKI, W. ONION: Functional Approach for Integration of Lipidomics and Transcriptomics Data. **Plos One**, v. 10, n. 6, p. e0128854, 2015.
- PLUSKAL, T. et al. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. **BMC bioinformatics**, v. 11, p. 395, jan. 2010.
- RADULOVIĆ, N. S. et al. Antimicrobial plant metabolites: structural diversity and mechanism of action. **Current medicinal chemistry**, v. 20, n. 7, p. 932–52, 2013.
- REDESTIG, H. et al. Data integration, metabolic networks and systems biology. In: HALL, R. D. (Ed.). . **Annual Plant Reviews Volume 43: Biology of Plant Metabolomics**. Oxford, UK: Wiley-Blackwell, 2011. v. 43p. 261–316.
- RIBERA, A. E.; ZUÑIGA, G. Induced plant secondary metabolites for phytopatogenic fungi control : a review. **J. Soil Sci. Plant Nutr. [online]**, v. 12, n. 4, p. 893–911, 2012.

RIGAL, A.; MA, Q.; ROBERT, S. Unraveling plant hormone signaling through the use of small molecules. **Frontiers in Plant Science**, v. 5, n. July, p. 1–20, 2014.

ROCHFORT, S. Metabolomics Reviewed: A New “Omics” Platform Technology for Systems Biology and Implications for Natural Products Research. **Journal of Natural Products**, v. 68, n. 12, p. 1813–1820, 2005.

ROLDAN, M. V. G. et al. Metabolomics reveals organ-specific metabolic rearrangements during early tomato seedling development. **Metabolomics**, p. 1–17, 2014.

ROLLAND, F.; BAENA-GONZALEZ, E.; SHEEN, J. Sugar sensing and signaling in plants: Conserved and Novel Mechanisms. **Annual Review of Plant Biology**, v. 57, n. 1, p. 675–709, 2006.

ROUGHAN, P. G.; BATT, R. D. The glycerolipid composition of leaves. **Phytochemistry**, v. 8, p. 363–369, 1969.

SABIDÓ, E.; SELEVSEK, N.; AEBERSOLD, R. Mass spectrometry-based proteomics for systems biology. **Current opinion in biotechnology**, v. 23, n. 4, p. 591–7, 12 ago. 2012.

SAINT-LARY, L. et al. Metabolomics as a tool for the authentication of rose extracts used in flavour and fragrance area. **Metabolomics**, v. 12, n. 3, p. 49, 2016.

SAITO, K.; MATSUDA, F. Metabolomics for functional genomics, systems biology, and biotechnology. **Annual review of plant biology**, v. 61, n. January, p. 463–89, jan. 2010.

SALEK, R. M. et al. A metabolomic study of brain tissues from aged mice with low expression of the vesicular monoamine transporter 2 (VMAT2) gene. **Neurochemical Research**, v. 33, n. 2, p. 292–300, mar. 2008.

SENGUPTA, A. et al. Global host metabolic response to *Plasmodium vivax* infection: a ¹H NMR based urinary metabonomic study. **Malaria journal**, v. 10, p. 384, 2011.

SHAH, P. et al. Proteomic analysis of ripening tomato fruit infected by *botrytis cinerea*. **Journal of Proteome Research**, v. 11, n. 4, p. 2178–2192, 2012.

SHANNON, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. **Genome research**, v. 13, n. 11, p. 2498–504, nov. 2003a.

SHANNON, P. et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. **Genome research**, v. 13, n. 11, p. 2498–2504, nov. 2003b.

SHEEN, J. Metabolic repression of transcription in higher plants. **The Plant cell**, v. 2, n. 10, p. 1027–1038, 1990.

SIEBERS, M. et al. Lipids in plant–microbe interactions. **Biochimica et Biophysica Acta**

(BBA) - Molecular and Cell Biology of Lipids, v. 12, n. 1, p. 2494, fev. 2016.

SILVA, J. C. et al. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. **Molecular & cellular proteomics : MCP**, v. 5, n. 1, p. 144–156, 2006.

SMITH, A. D. **Oxford Dictionary of Biochemistry and Molecular Biology**. [s.l.] Oxford University Press, 1997.

SMITH, C. A et al. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. **Analytical chemistry**, v. 78, n. 3, p. 779–87, 1 fev. 2006.

SUI, N. et al. Identification and transcriptomic profiling of genes involved in increasing sugar content during salt stress in sweet sorghum leaves. **BMC Genomics**, v. 16, n. 1, p. 534, 2015.

SUMNER, L. W. et al. Proposed minimum reporting standards for chemical analysis. **Metabolomics**, v. 3, n. 3, p. 211–221, 19 set. 2007.

SWANSON, S. K.; WASHBURN, M. P. The continuing evolution of shotgun proteomics. **Drug Discovery Today**, v. 10, n. 10, p. 719–725, 2005.

THOMASHOW, M. F. Plant cold accliation: Freezing tolerance genes and regulatory mechanisms. **Annual review of plant physiology and plant molecular biology**, v. 50, p. 571–599, 1999.

TRABELSI, H. et al. Triacylglycerol and glycerophospholipid identification and accumulation during ripening of *Pistacia lentiscus* L. (Lentisc) fruit. **JAOCs, Journal of the American Oil Chemists' Society**, v. 91, n. 7, p. 1189–1196, 2014.

TURNER, B. M. Cellular memory and the histone code. **Cell**, v. 111, n. 3, p. 285–291, 2002.

UNICA. **União da indústria de cana-de-açúcar**. Disponível em: <<http://www.unica.com.br/>>. Acesso em: 16 abr. 2016.

URBANCZYK-WOCHNIAK, E. et al. Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. **EMBO reports**, v. 4, n. 10, p. 989–93, out. 2003.

UYS, L. et al. Kinetic model of sucrose accumulation in maturing sugarcane culm tissue. **Phytochemistry**, v. 68, n. 16-18, p. 2375–92, 2007.

VAN DEN BERG, R. A et al. Centering, scaling, and transformations: improving the biological information content of metabolomics data. **BMC genomics**, v. 7, p. 142, jan. 2006.

VETTORE, A. L. et al. The libraries that made SUCEST. **Genetics and Molecular Biology**, v. 24, n. 1-4, p. 1–7, dez. 2001.

VIJAYAPALANI, P. et al. Interaction of the trans-frame potyvirus protein P3N-PIPO with host

protein PCaP1 facilitates potyvirus movement. **PLoS Pathogens**, v. 8, n. 4, 2012.

VINOD, H. D. Canonical ridge and econometrics of joint production. **Journal of Econometrics**, v. 4, n. 2, p. 147–166, 1976.

WAAIJENBORG, S.; ZWINDERMAN, A. H. Sparse canonical correlation analysis for identifying, connecting and completing gene-expression networks. **BMC bioinformatics**, v. 10, p. 315, 2009.

WACLAWOVSKY, A. J. et al. Sugarcane for bioenergy production: an assessment of yield and regulation of sucrose content. **Plant biotechnology journal**, v. 8, n. 3, p. 263–76, abr. 2010.

WALLACE, M. et al. Relationship between the lipidome, inflammatory markers and insulin resistance. **Molecular bioSystems**, v. 10, n. 6, p. 1586–95, 2014.

WALLER, T. et al. DNA microarray integromics analysis platform. **BioData Mining**, v. 8, n. 1, p. 18, 2015.

WANG, S. et al. Spatio-temporal distribution and natural variation of metabolites in citrus fruits. **Food Chemistry**, v. 199, p. 8–17, maio 2016.

WECKWERTH, W. Integration of metabolomics and proteomics in molecular plant physiology - coping with the complexity by data-dimensionality reduction. **Physiologia plantarum**, v. 132, n. 2, p. 176–89, fev. 2008.

WELTI, R. et al. Plant lipidomics: Discerning biological function by profiling plant complex lipids using mass spectrometry. **Frontiers in Bioscience**, v. 12, n. 1, p. 2494–2506, 2007.

WIDEMANN, E. et al. Identification of the 12-oxojasmonoyl-isoleucine, a new intermediate of jasmonate metabolism in Arabidopsis, by combining chemical derivatization and LC–MS/MS analysis. **Metabolomics**, v. 11, n. 4, p. 991–997, 2015.

WILKINS, M. R. et al. From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Amino Acid Analysis. **Biotechnology**, v. 14, n. 1, p. 61–65, jan. 1996.

WISHART, D. S. et al. HMDB 3.0-The Human Metabolome Database in 2013. **Nucleic Acids Research**, v. 41, n. D1, p. 801–807, 2013.

WOLF, S. et al. In silico fragmentation for computer assisted identification of metabolite mass spectra. **BMC Bioinformatics**, v. 11, p. 148, 2010.

WOLTERS, D. A.; WASHBURN, M. P.; YATES, J. R. An automated multidimensional protein identification technology for shotgun proteomics. **Analytical Chemistry**, v. 73, n. 23, p. 5683–5690, 2001.

XIA, J. et al. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. **Nucleic acids research**, v. 37, n. Web Server issue, p. W652–W660, jul. 2009.

YATES, J. R. 3RD. Mass Spectrometry and the Age of the Proteome. **Journal of Mass Spectrometry**, v. 33, p. 1–19, 1998.

YIZHAK, K. et al. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. **Bioinformatics (Oxford, England)**, v. 26, n. 12, p. i255–60, 15 jun. 2010.

YUAN, J. S. et al. Plant systems biology comes of age. **Trends in plant science**, v. 13, n. 4, p. 165–71, abr. 2008.

ANEXO

Anexo – Scripts utilizados neste trabalho

Script para processamento dos dados brutos (metabolômica)

```
extract_peaks <- function() {  
  myDir <-  
  "/media/Arquivos/Max_Feffer/Fabricio/Dados_netCDF/cana_netCDF/analise/todos"  
  analise <- "Todos_03062014" # nome da sua analise  
  timerange <- c(185, 2145) # ajuste de acordo com o tempo de corrida que queira  
  analisar (em scans)  
  setwd(myDir)  
  library(xcms)  
  
  xset <- xcmsSet(method="centWave", ppm=10, peakwidth=c(5, 20), snthresh=5,  
  prefilter=c(3, 50), scanrange=timerange, mzCenterFun="wMean", integrate=1, mzdiff=-  
  0.001, fitgauss=FALSE, nSlaves=2)  
  xset  
  xset <- group(xset)  
  xset2 <- retcor(xset, method="obiwarp", plotttype="deviation")  
  xset2 <- group(xset2, bw=5, minfrac=0.6, mzwid=0.1, max=50, minsamp=1)  
  xset3 <- fillPeaks(xset2)  
  xset3  
}
```

Script para a normalização pela massa obtida na pesagem

```
norm_fw <- function(xset, xset2, xset3) {  
  fw <- c(51.7, 51.7, 51.7, 48.6, 48.6, 48.6, 57, 57, 57, 47.4, 47.4, 47.4, 47.8,  
  47.8, 47.8, 50.7, 50.7, 50.7, 54.1, 54.1, 54.1, 62.2, 62.2, 62.2, 46.9, 46.9, 46.9,  
  43.2, 43.2, 43.2, 62.5, 62.5, 62.5, 42.1, 42.1, 42.1, 49, 49, 49, 58.5, 58.5, 58.5,  
  47, 47, 47, 55.1, 55.1, 55.1, 61.7, 61.7, 61.7, 47.7, 47.7, 47.7, 65.5, 65.5, 65.5,  
  44.2, 44.2, 44.2, 53.9, 53.9, 53.9, 60.5, 60.5, 60.5, 41.1, 41.1, 41.1, 58.9, 58.9,  
  58.9, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50,  
  50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50, 50,  
  50, 50, 50, 50) # massa fresca das amostras  
  massa <- 50 # quantidade que deveria ser pesada  
  factors1 <- rep(massa/fw, times=table(peaks(xset2)[,"sample"]))  
  nrorigpeaks<-nrow(xset2@peaks)+1  
  nrtotalpeaks<-nrow(xset3@peaks)  
  factors2<-rep(massa/fw,  
  times=table(peaks(xset3)[nrorigpeaks:nrtotalpeaks,"sample"]))  
  factors<-c(factors1, factors2)  
  intensitycolumns <- c("into", "maxo") # verificar se existe o intb no  
  xset3@peaks, para poder usa-lo tambem  
  xset3@peaks[,intensitycolumns] <- xset3@peaks[,intensitycolumns] * factors  
  reporttab <- diffreport(xset3, filebase = analise, eicmax = 20, metlin = 0.15,  
  value = "into", sortpval=FALSE)  
  reporttab2 <- reporttab  
  reporttab2 <- data.frame(reporttab[, -  
  c(which(colnames(reporttab)=="fold"):which(colnames(reporttab2)=="metlin"))],  
  stringsAsFactors=FALSE)  
  reporttab2$name <- paste(round(reporttab$mzmed, 4), round(reporttab$rtmed, 2),  
  sep="/")  
}
```

Script para a normalização da intensidade relativa ao padrão interno

```
Norm_is <- function(reporttab, reporttab2, reporttab3) {  
  lineintstd <- 64 # marcador quercetina  
  startcol = which(colnames(reporttab)=="metlin")+1  
  reporttab3 <- matrix(data=NA, ncol=dim(reporttab2)[2], nrow=dim(reporttab2)[1])  
  colnames(reporttab3) <- colnames(reporttab2)
```

```

reporttab3[, 1] <- reporttab2[, 1]
for(i in 2:dim(reporttab3)[2]) { # numero de colunas
  intstd <- as.numeric(as.character(reporttab[lineintstd, (startcol-2+i)])) #
localizacao do padrao interno em reporttab2
  for(k in 1:dim(reporttab3)[1]) { # numero de linhas
    reporttab3[k,i] <- as.numeric(as.character(reporttab[k, (startcol-2+i)])) /
intstd # normalizaÃo pelo padrÃo interno
  }
}
reporttab3 = reporttab3[-c(64, 80),] #entreno
write.table(reporttab2, file = paste(analise, "-orig_pca.csv", sep = ""),
row.names = F, quote = F, sep = ",")
write.table(reporttab3, file = paste(analise, "-norm_pca.csv", sep = ""),
row.names = F, quote = F, sep = ",")
}

```

Script para construo das tabelas de protemica para a anlise estatstica dos dados

COMO USAR:

```

#### inicie o R e digite:
#### source("caminho e nome deste arquivo entre aspas")
#### ex: source("/media/Arquivos/Max_Feffer/Scripts/script_pca-MassPivot-v2.R")
####
#### tab_pca("Diretorio somente com as tabelas, nome da tabela q sera gerada")
#### ex: tab_pca("/media/Arquivos/Max_Feffer/Tabelas/", "Tabela_PCA_7MI9.txt")
####

##### funcao para fazer uma lista de todas CLUSTER em todas tabelas
l_prot = function(files, tab_orig) {
  lista = NULL
  for(i in 1:length(files)) {
    tab = tab_orig[[i]]
    lista = c(lista, tab$protein.Accession)
  }
  lista = sort(lista) # lista com todas as proteinas de todas as tabelas
  lista = unique(lista) # lista das proteinas sem repeticao
  return(lista)
}

##### funcao para buscar uma proteina em uma tabela
s_prot = function(tab, prot) {
  fmol = 0
  c = 0
  for(i in 1:dim(tab)[1]) { # todas as linhas da tabela
    if(is.na(tab$protein.Accession[i])) { # verifica se a proteina da tabela e NA
    } else {
      if(prot == tab$protein.Accession[i]) { # verifica se a prot = prot[i] da tabela
        if(tab$protein.score[i] > 70) { # verifica se o score da prot[i] > 70
          if(tab$protein.AutoCurate[i] == "Green") { # verifica se a cor da prot[i]
e Green
            if(is.na(tab$protein.fmolOnColumn[i])) { # verifica se o fmol da prot[i]
e NA
              } else {
                if(tab$protein.fmolOnColumn[i] > 0 ) { # verifica se o fmol da prot[i] >
0
                  fmol = fmol + tab$protein.fmolOnColumn[i] # soma os fmol[i] com os
anteriores
                  c = c + 1
                }
              }
            }
          }
        }
      }
    }
  }
}

```



```

    }
  }
  if(c == 0) {
    return(0) # se nao encontrar a proteina na tabela, retorna 0
  } else {
    return(fmol/c) # retorna a media de fmol da prot de entrada
  }
}

##### funcao para ler as tabelas de um diretorio e montar uma tabela
tab_pca = function(pasta, saida) { # pasta = diretorio das tabelas, saida = tabela
output
  setwd(pasta)
  files = dir()
  tab_orig = NULL
  tab_pca = NULL
  for(i in 1:length(files)) {
    tab_orig[[i]] = read.csv(files[i], sep=",", stringsAsFactors=F) # cada
tab_orig[[i]] contem uma planilha do diretorio
    x = tab_orig[[i]]
    tab_orig[[i]] = x[order(x$protein.Accession, x$protein.score,
x$protein.fmolOnColumn),] # organizando as colunas por proteina seguida do fmol
  }

  proteins = l_prot(files, tab_orig) # gerando a lista de todas as proteinas da tabela

  tab_final = matrix(ncol = length(tab_orig) + 1, nrow = length(proteins)) # criando
a tabela final

  for(j in 1:length(proteins)) {
    tab_final[j, 1] = proteins[j]
    colnames(tab_final) = c("accession", files) # criando o nome das colunas na tabela
final
    for(k in 1:length(tab_orig)) {
      prot = s_prot(tab = tab_orig[[k]], prot = proteins[j]) # buscar a proteina j na
tabela k
      tab_final[j, (k + 1)] = prot
    }
  }
  del_prot = NULL
  for(l in 1:dim(tab_final)[1]) { # retirar as proteinas com 0 de fmol para todas
amostras da tabela final
    if(regexpr("REVERSE+", tab_final[l, 1]) == 1) { # verificar se a proteina e uma
"Reversed Sequence+"
      del_prot = c(del_prot, l) # vetor com proteinas a serem deletadas com sinal "-"
    } else if(sum(as.numeric(tab_final[l, 2:dim(tab_final)[2]])) == 0) { # verificar
proteinas com 0 de fmol para todas amostras da tabela final
      del_prot = c(del_prot, l) # vetor com proteinas a serem deletadas com sinal "-"
    }
  }
  tab_final_limpa = tab_final[-del_prot,] # exclui as linhas contidas no vetor
del_prot
  write.table(tab_final_limpa, saida, row.names = F, sep = ",")
}

```

Script para construção dos gráficos de barra dos GO Terms

```

grafico_golist <- function(graph_name, lista) {
  setwd(dirname(lista))
  mylist <- scan(lista, what="character", sep="\n")
  mytable <- NULL
  for(i in 1:length(mylist)) { #percorre linhas do vetor
    lmy <- NULL
    mmy <- NULL
    lmy <- strsplit(mylist[i], split=";")
  }
}

```

```

lmy[[1]] <- gsub("^[:space:]", "", lmy[[1]])
mmy <- grep("P:.", lmy[[1]]) #busca todos os processos biológicos P:
if(length(mmy)>0) {
  mytable <- c(mytable, lmy[[1]][max(mmy)]) #trocar min() e max() para go basal
ou específico
}
}
x <- table(mytable)
x <- sort(x)
jpeg(paste(graph_name, ".jpg", sep=""), width=800, height=640, quality=75)
par(mai=c(1,8,1,0.5))
barplot(x, main=graph_name, horiz=TRUE, cex.names=0.8, las=1, xlab="Nº Proteínas")
dev.off()
}

```

Script para construção dos gráficos de espectros com os resultados da fragmentação *in silico*.

```

grafico_metfrag <- function() {
  winDialog(type="ok", message= "Escolha o arquivo texto dos fragmentos .txt")
  myfragfile <- file.choose()
  myfrag <- read.table(myfragfile, header=FALSE, sep="\t")
  winDialog(type="ok", message= "Escolha o resultado do Metfrag . csv")
  myFile <- file.choose()
  myData <- read.csv(myFile, sep = "|", header=TRUE)
  setwd(dirname(myFile))
  if(myData$NoExplPeaks > 0) {
    explpeaks <- strsplit(as.character(myData$ExplPeaks), ";")
    peaks_intens <- strsplit(explpeaks[[1]], "_")
    peaks <- NULL
    intens <- NULL
    for(i in 1:length(peaks_intens)) {
      peaks <- c(peaks, as.numeric(peaks_intens[[i]][1]))
      intens <- c(intens, as.numeric(peaks_intens[[i]][2]))
    }
  }
  ### escalonar intensidades originais
  max_ifrag <- max(myfrag$V2)
  int_scal <- round(myfrag$V2*999/max_ifrag, 1)
  max_num <- max(myfrag$V1)
  file_name <- strsplit(basename(myFile), split=".csv")
  jpeg(paste(file_name, ".jpg", sep=""))
  plot(x=myfrag$V1, y=int_scal, type="h", xlab="m/z", ylab="Intensidade",
main=myData$HMDB_ID, col="red", lwd=2)
  if(myData$NoExplPeaks > 0) {
    points(x=peaks, y=intens, col="blue", type="h", lwd=2)
  }
  pi <- myData$EXACT_MASS # massa exata do match do HMDB
  i_pi <- which.min(abs(myfrag$V1-pi)) # procura o index do pico correspondente ao
match do HMDB
  points(x=myfrag$V1[i_pi], y=int_scal[i_pi], col="deepskyblue", type="h", lwd=2)
  legend("topright", legend=c("não identificados", "identificados", "precursor"),
col=c("red", "blue", "deepskyblue"), lty=c(1,1), lwd=2)
  dev.off()
  winDialog(type="ok", message= "Pronto!!!")
}

grafico_metfrag()

```

Script para localização de sequências fasta de proteínas em arquivos multi-fasta

COMO USAR:

a lista e o FASTA tem que ser txt

```

####
#### inicie o R e digite:
#### source("caminho e nome deste arquivo entre aspas")
#### ex: > source("/home/fabricio/Área de Trabalho/merge_probmetab-xcms/merge_probm-
xcms.R")
####
#### digite o comando no R
#### select_fasta("LISTA", "FASTA", "SAIDA")
#### ex: > select_fasta("D:/Max_Feffer/Juliana_Fonseca/contaminantes/lista.txt",
"D:/Max_Feffer/Juliana_Fonseca/contaminantes/fasta.txt",
"D:/Max_Feffer/Juliana_Fonseca/contaminantes/tabela_saida.txt")

select_fasta <- function(lista, fasta, saida) {
  lista_orig <- read.table(lista, header = FALSE, stringsAsFactors = F)
  fasta_orig <- read.table(fasta, header = FALSE, sep = "\n", stringsAsFactors = F)
  tabela = NULL
  print("Running...")
  for(i in 1:dim(lista_orig)[1]) {
    sel_lista <- lista_orig[i, 1]
    for(j in 1:dim(fasta_orig)[1]) {
      sel_fasta <- fasta_orig[j, 1]
      sel_fasta_name <- regexec(">([[:alnum:]]+\\.[b,g]*)", sel_fasta)
      sel_fasta_name2 <- regmatches(sel_fasta, sel_fasta_name)
      if(is.na(sel_fasta_name2[[1]][2])) {
      } else if(sel_lista == sel_fasta_name2[[1]][2]) {
        tabela <- c(tabela, sel_fasta_name2[[1]][1])
        tabela <- c(tabela, fasta_orig[(j+1), 1])
      }
    }
  }
  write.table(tabela, file=saida, row.names=F ,col.names=F, quote=F)
  print("Done!")
}

```