O que foi feito:

Foi feita uma anotação funcional de transcritos relacionados à parede celular de cana-de-açúcar visando categorizá-los em domínios que indicam sua ação na parede (essa ação se refere à síntese ou degradação de compostos da parede). Os transcritos foram obtidos a partir dos trabalhos feitos pela Amanda e pelo David que triaram os transcritos que foram classificados como sendo relacionados à parede celular. Os domínios utilizados foram: H-hemicellulose, C-cellulose, P-pectin, L-lignin, E-expansin, H/P-hemicellulose and pectin, H/C-hemicellulose and celulose, NCW-non cell wall e M-metabolism (NCW e M são sinônimos, mas ambas nomenclaturas acabaram sendo usadas). NCW e M foram os domínios utilizados para transcritos que não agiam sobre a parede celular, mas que também foram analisados (em muitos caso essas enzimas foram classificadas como relacionadas à parede por ficarem nela ancoradas, entretanto elas não agem sobre os polissacarídeos de parede ou a lignina). Para poder fazer essa categorização foram utilizadas informações do Blast2go, SUCEST e adicionais (coluna J do arquivo “Validation tables final” do Excel) referentes a cada transcrito que também foram adquiridas a partir do trabalho do David e Amanda. Cruzando as informações dessas diferentes fontes buscou-se um consenso quanto à classificação daquele transcrito. Se todas as fontes de informação (Blast2go, SUCEST e adicionais) classificassem o transcrito da mesma forma era feita então uma pesquisa na literatura para identificar qual a função dessa enzima. Descobrindo-se a função da enzima era possível dizer a qual domínio ela pertencia e então era feita sua classificação dentro dos domínios. Entretanto, em muitos casos as fontes de informação apresentavam classificações conflitantes ou insuficientes e seria impossível dizer qual era o transcrito em questão somente com aqueles dados. Para tentar descobrir qual a classificação daquele transcrito foram feitos dois tipos de buscas. A primeira delas constava em usar a identificação do transcrito (SCACAD1039B10.g, por exemplo) em uma busca no arquivo de Excel “blast com EC das proteínas e identidade lado a lado” (também obtido dos trabalhos da Amanda e do David). Neste arquivo se encontram os nomes de todos os transcritos e, quando encontrados, os EC numbers (Enzyme Comission numbers) a eles associados (para ver a quais enzimas os EC numbers correspondem olhar o ExplorEnz - <http://www.enzyme-database.org/index.php> para números completos ou o Enzyme Nomenclature - <http://www.chem.qmul.ac.uk/iubmb/enzyme/>, para números parciais). Os EC numbers foram usados como uma outra fonte de identificação para as enzimas e caso fossem de encontro com as classificações do Blast2go do SUCEST ou das informações adicionais era possível determinar que essa era a classificação correta para a enzima em questão. O segundo tipo de busca foi feito utilizando as sequências de aminoácidos correspondentes a cada transcrito (ver arquivo “SUCEST”), em bancos de dados referentes a proteínas (UniProt - <http://www.uniprot.org/blast/>) e a domínios de proteínas (NCBI Conserved Domains - <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi> e InterPro - <https://www.ebi.ac.uk/interpro/>). Ao olhar os resultados das buscas no UniProt foram consideradas os *matches* com maior identidade e que apareciam ao menos mais de 5 vezes. Atendendo a esses requisitos essa classificação era considerada como a apontada por essa busca. Já o NCBI CD e o InterPro são bastante categóricos, dando o resultado de forma direta. Tal qual para as buscas com EC numbers, as classificações obtidas pelos sites eram confrontadas com as classificações do Blast2go do SUCEST ou das informações adicionais para ver se havia confirmação de alguma destas classificações. Entretanto, em muitos casos, mesmo após essas duas buscas, não era possível dizer de forma precisa qual era a enzima em questão. Para mostrar se foi possível ou não classificar os transcritos e quais foram as análises feitas as análises dos transcritos foram categorizadas desta forma: T-true, para os transcritos classificados diretamente a partir das informações do Blast2go, do SUCEST e das informações adicionais ou com auxílio do EC number; TW-true after web, para os transcritos classificados após as buscas nos sites supracitados (UniProt, NCBI CD e InterPro) e DW-doubt after web, para os casos em que mesmo após a busca nos sites não foi possível identificar o transcrito. Feitas todas essas etapas, foi obtida a lista com a classificação dos transcritos em domínios.

OBS: A forma como eu fiz a análise dos *matches* do UniProt é uma análise incompleta. Além de olhar a identidade das proteínas é necessário também observar a cobertura das mesmas, o que não foi feito por mim.

Dúvidas e direcionamentos futuros:

Ao passar por essas etapas muitos DWs podem ter suas dúvidas sanadas e assim passar a ser classificados dentro dos domínios de parede ou como NCW.

Dúvidas:

-- Dirigente protein: é só lignano mesmo?

-- glucuronosyltransferase: tem como definir o que ela faz? Ou é um termo muito genérico? Acredito que seja abrangente demais, não sendo possível classificar.

-- Questão das extensinas como novo domínio

-- Questão do M e do NCW serem ou não diferentes

-- Beta-galactosidase: H, P ou M?

-- Quando há ação de degradação sobre oligos eu sei que dá pra considerar como degradação do polissacarídeo, determinando os domínios normalmente, entretanto, eu não sei se o mesmo se aplica para os monos. Nesse caso não dá para classificar, uma vez que os monos podem vir da parede mas também de açúcares de reserva, transporte, dentre outros OU classifica como sendo uma enzima que age sobre diversos/todos os domínios? –creio que a primeira opção seja a mais coerente, mas é passível de discussão-

-- Preciso achar um ponto comum para minhas classificações de enzimas basais, pelo que eu vi tem horas que eu considero que se é basal não está relacionado à síntese do componente por não estar ligado diretamente com sua síntese e nem com sua degradação. Mas tem momentos em que por ser basal eu entendo que é fundamental para a formação da substância, então coloco como relacionada à sua síntese. Preciso tomar cuidado com isso para não haver incongruência!!!! As enzimas que entram nessa dúvida são:

-4-coumarate—CoA ligase (OBS: classifiquei o 1119 e o 1120 como L para manter o padrão de classificação da 4-coumarate—CoA ligase e não ficar incongruente com os outros que eu já havia classificado);

- cinnamyl alcohol dehydrogenase;

- cinnamoyl-coA reductase que pelo visto eu coloquei como L pelo fato de o artigo destacar sua importância para formação da lignina, mas ela está uns dois passos atrás na via, sendo um ótimo exemplo para trazer outra questão à tona: o quão longe o gene deve estar para ser considerado como basal e não relacionado e a partir de que ponto ele já deve ser considerado como diretamente relacionado à síntese do composto (polissacarídeos ou lignina) em questão? Ou ainda, faz diferença se ele está relacionado a uma reação mais basal da via ou uma reação mais próxima do composto?.

- trans-cinnamate 4-monooxygenase/cinnamate 4-hydroxylase/C4H;

- fenilalanina ammonia liase/phenylalanine ammonia-lyase (PAL);

- GDP-mannose 3,5-epimerase, relacionada à produção de GDP-galactose (ainda há dúvida sobre sua função, mas se enquadra na questão de ser basal);

- UDP-glucose 6-dehydrogenase;

- rhamnose biosynthetic enzyme/RHM;

- UDP-Sugar Pyrophosphorylase;

- UDP-D-apiose/UDP-D-xylose synthase;

- UDP-arabinopyranose mutase;

- UDP-Sugar Pyrophosphorylase/ UTP-monosaccharide-1-phosphate uridylyltransferase;

- UTP-glucose-1-phosphate uridylyltransferase/ UDP glucose pyrophosphorylase;

- GDP-mannose 4,6-dehydratase/ GDP-D-mannose-4,6-dehydratase;

- GDP-L-fucose synthase/GDP-4-keto-6-deoxy-D-mannose 3,5-epimerase-4-reductase (GER1).

-- kobito: coloca como C ou deixa sem classificar uma vez que não se sabe sobre sua atividade?

-- xylose isomerase: é NCW?

-- beta-glicosidase: é H/C ou não?

-- As proteínas notum, por apresentarem sequência similar às pectinacetylesterase podem ser usadas como evidência de sua presença? Eu acho que não seria muito correto, pois ainda que elas tenham similaridade na sequência é uma evidência muito indireta.

-- Questão de usar um domínio para auxiliar na classificação de uma enzima o que pode levar a erro (artigo que apontou os falsos positivos, vide arquivo “Informações sobre as enzimas e referências bibliográficas”, é uma das referências dentro de “artigos mais gerais”, no final do arquivo).

-- glucan endo-1,3-β-D-glucosidase/ beta-1,3-glucanase (EC 3.2.1.39) e também 1,3-β-glucan synthase e/ou callose synthase agem sobre a calose. Deve haver um domínio só para a calose? Ou ela entra em H? No momento estão todas como DW.

-- Enzimas que não achei em plantas e por isso não soube como classificar:

- mannosyl-oligosaccharide 1,2-α-mannosidase;

- galactosylxylosylprotein 3-β-galactosyltransferase/ galactosyltransferase II/ b3GALT6/ β3GalT6;

- UDP-glucose:glycoprotein glucosyltransferase.

-- annexin: apesar dos indícios, não tenho certeza de sua ação sobre a callose, como classificar?

-- GDP-mannose 3,5-epimerase: tem ou não influência sobre H e P? A dúvida é quanto a haver ou não relevância para a produção de precursores destes polissacarídeos, que é afirmada por um artigo e negada por outro.

-- Laccase: relacionada à lignina ou não?

Direcionamentos:

Aqui estão apontadas pesquisas que ficaram pendentes ou que poderiam ser aprofundadas

!!!Antes de fazer as pesquisas a seguir checar se elas vão interferir em alguma classificação no Excel, caso não houver nenhuma relação não há necessidade de pesquisá-las!!!

Ver se há ou não relação entre xyloglucan glycosyltransferase e celulose synthase like, porque se houver posso trocar alguns DW por H TW.

Pesquisar beta glucanase EC 3.2.1.4 e classificar o 415, 438, 457, 508, 367, outros.

Pesquisar mannan synthase – 224,

Pesquisar hydroquinone glucosyltransferase – 283,

Pesquisar alpha-1,3/1,6-mannosyltransferase ALG2 e ver por que há essa confusão entre os dois, qual a diferença entre eles e por que o gene para os dois é um só (ALG2). – 277,

Pesquisar peptidase S8/S53 domain/superfamily Depois rever 455.

# Averiguar o fato da kobito ser sinônimo da elongation defective 1 protein

# <https://www.ncbi.nlm.nih.gov/protein/NP_187467.1>

Pesquisar **Armadillo/beta-catenin-like repeats**. An approximately 40 amino acid long tandemly repeated sequence motif first identified in the Drosophila segment polarity gene armadillo; these repeats were also found in the mammalian armadillo homolog beta-catenin, the junctional plaque **protein plakoglobin**, the adenomatous polyposis coli (APC) tumor suppressor protein, and a number of other proteins. ARM has been implicated in mediating protein-protein interactions, but no common features among the target proteins recognized by the ARM repeats have been identified; related to the HEAT domain; three consecutive copies of the repeat are represented by this alignment model. VER 242.

Kobito – 722; 988; e todas as outras que tiver, fazer busca -- pesquisar glycosyltransferase-like KOBITO

Ver relação entre 4-coumarate-CoA ligase e AMP dependent CoA ligase.

Pesquisar laccase mais a fundo para ver se já descobriram outras relações que não com a lignificação. Tentar ter acesso a esse artigo <http://www.sciencedirect.com/science/article/pii/S0168945202000353>

Checar se isoflavone reductase like é o mesmo que isoflavone reductase

Pesquisar endo-1,4-β-xylanase (EC 3.2.1.8) pq aparentemente eu classifiquei como H só pelo que diz o ExplorEnz.

Pesquisar UDP-glucose 4-epimerase e HECT (Homologous to the E6-AP Carboxyl Terminus) domain and and RCC1-like domain (RLD) containing E3 ubiquitin protein ligase family member 1

<http://www.enzyme-database.org/query.php?ec=5.1.3.2>

Após a pesquisa classificar o 1122 e o 882.

Pesquisar galactinol synthase/inositol 3-α-galactosyltransferase

<http://www.enzyme-database.org/query.php?ec=2.4.1.123>

Após a pesquisa classificar o 1112.

Pesquisar mannan synthase (no Cazy está escrito β-1,4-mannan synthase). Buscar mannan synthase no Excel e ver casos em que eu assumi que por ser da família GT 2 era um glucomannan 4-beta-mannosyltransferase. Em alguns casos, pode até ser ele, quando ele é mencionado e talz, mas acho que em alguns casos ele sequer aparece e eu assumi que era ele. Depois da pesquisa: classificar o 1100 que deixei em aberto e os que classifiquei errado.

Pesquisar xyloglucan galactosyltransferase para ver se ele é de fato a KATAMARI ou se ele é algo à parte.

Pesquisar xyloglucan endotransglycosylase/hydrolase (XTH); laminin G domain e concanavalin A-like lectin/glucanase domain. É para o 366.

Pesquisar β-mannosidase e ver quais são as diferenças frente à mannan endo-1,4-β-mannosidase

<http://www.enzyme-database.org/query.php?ec=3.2.1.25>

<http://www.enzyme-database.org/query.php?ec=3.2.1.78>

Pesquisar/reler xylanase inhibitor pra ver certinho qual sua ação. Pode dar só uma olhada no chitinase inhibitor, mas essa o nome já deixa claro que é NCW.

Pesquisar polygalacturonase inhibitor

Calose sintase <http://cdn.intechopen.com/pdfs-wm/44293.pdf>

Finalizar pesquisa: alpha-galactosidase

Pesquisar: (N-acetylglucosaminyl-proteoglycan 4-β-glucuronosyltransferase e glucuronosyl-N-acetylglucosaminyl-proteoglycan 4-α-N-acetylglucosaminyltransferase). Pelas buscas existem muitas possibilidades, mesmo se eu pesquisar não sei se vai dar pra classificar categoricamente.

Pulei:

Pesquisa inacabada da beta-glucosidase

<https://scholar.google.com.br/scholar?q=%CE%B2-glucosidase+plant&btnG=&hl=pt-BR&as_sdt=0%2C5>

<http://www.jbc.org/content/279/30/31796.full>

<http://www.plantphysiol.org/content/94/2/401.full.pdf+html>

<http://www.jbc.org/content/270/26/15789.full.pdf>

<http://science.sciencemag.org/content/sci/262/5136/1051.full.pdf>

Tentar acessar, reler e buscar novos artigos para elucidar ação da Hydroxycinnamoyl-CoA transferase (HCT)

Tentar elucidar dúvidas sobre a mannosyltransferase ALG2

Tentar achar mais informações sobre a xylose isomerase