



Anotação funcional

Dr^a Desirrê Petters-Vandresen

Por que realizar uma anotação funcional?

- Atribuir categorias e funções aos genes anotados e detectados
- Facilidade de estudo dentro das categorias
- Melhor entendimento do modo de vida do organismo analisado
- Possibilidade de comparações funcionais entre organismos de modo de vida similar ou diferente
- Facilidade de seleção de genes candidatos para estudos funcionais com base em categorias de interesse

Abordagens gerais e aplicações

- Abordagens:
 - *Ab initio*
 - Baseada em homologia
- Aplicações:
 - Classificação e identificação (nomenclatura)
 - Atribuição de função
 - Localização celular
 - Processos biológicos relacionados

CAZymes

- Baseada em homologia:
 - Comparação com CAZymes já caracterizadas em bases de dados de CAZymes (ex: CAZy)
 - Comparação com domínios conservados das classes e famílias de CAZymes:
 - GTs: glycosyltransferases
 - GHs: glycoside hydrolases
 - PL: polysaccharide lyases
 - CE: carbohydrate esterases
 - CBM: carbohydrate-binding modules
 - AA: enzimas auxiliares (transformação de lignina)

Arquivo de saída (lista)

1	Gene ID	HMMER	Hotpep	DIAMOND	Signalp	#ofTools
2	scaffold_1.1012	GT4(512-682)	N	GT4	N	2
3	scaffold_1.1057	GT90(587-853)	N	N	N	1
4	scaffold_1.1058	GH47(45-507)	GH47	GH47	N	3
5	scaffold_1.1091	AA7(105-310)	N	N	Y(1-23)	1
6	scaffold_1.1104	GH3(90-311)	GH3+CBM1	GH3	Y(1-30)	3
7	scaffold_1.1174	GH5_49(118-421)	N	GH5_49	N	2
8	scaffold_1.1226	GT8(4-224)	GT8	GT8	N	3
9	scaffold_1.1230	GT90(708-991)	GT90	GT90	N	3
10	scaffold_1.1258	AA7(59-483)	N	N	N	1
11	scaffold_1.1296	GH16_18(73-229)	N	N	Y(1-24)	1

- Lista dos genes avaliados, classificação em classe e família de CAZymes e resultados para diferentes ferramentas de identificação utilizadas

Efetores

- *Ab initio*:
 - Proteínas pequenas (< 300 aa)
 - Ricas em cisteína e pobres em serinas
 - Secretadas (presença de peptídeo sinal, ausência de domínios transmembrana)
 - Menor proporção de aminoácidos alifáticos (apolares e hidrofóbicos)
 - Maior proporção de aminoácidos básicos (polares, positivamente carregados e hidrofílicos)
- Baseada em homologia: bases de dados de efetores ou sequências descritas em estudos prévios
- Alta variabilidade inter e intra-específica: dificuldade de detectar homologia em muitos casos

Arquivo de saída (lista)

- Em geral, colunas de identificação do gene avaliado, e colunas com a classificação e valor de confiabilidade ou probabilidade

1	# Identifier	Prediction	Probability
2	scaffold_10.6	Non-effector	0.872
3	scaffold_10.13	Non-effector	0.991
4	scaffold_10.23	Non-effector	0.981
5	scaffold_10.40	Non-effector	0.989
6	scaffold_10.41	Effector	0.924
7	scaffold_10.60	Non-effector	0.97
8	scaffold_10.61	Non-effector	0.973
9	scaffold_10.100	Non-effector	0.977
10	scaffold_10.101	Non-effector	0.99
11	scaffold_10.117	Non-effector	0.991
12	scaffold_10.124	Non-effector	0.991
13	scaffold_10.144	Non-effector	0.99
14	scaffold_10.152	Non-effector	0.555
15	scaffold_10.172	Non-effector	0.736
16	scaffold_10.199	Non-effector	0.97
17	scaffold_10.219	Non-effector	0.864
18	scaffold_10.230	Non-effector	0.931
19	scaffold_10.249	Effector	0.873

Função molecular, localização celular, processo biológico

- Gene Ontology

- **Função molecular (*molecular function*)**: atividades realizadas pelas proteínas de forma mais abrangente, sem especificar a localização celular ou etapa do desenvolvimento em que a atividade ocorre (ex: transporte, atividade catalítica)
- **Localização celular (*cellular component*)**: localização ou compartimento celular em que as proteínas desempenham uma atividade ou atuam de forma estrutura (ex: mitocôndria, ribossomo, citoplasma)
- **Processo biológico (*biological process*)**: processos biológicos abrangentes que são realizados através de atividades específicas (ex: reparo de DNA, transdução de sinal, processos biossintéticos específicos)

Função molecular, localização celular, processo biológico

- Entender o organismo e seus genes em termos mais abrangentes
- Compreender a interação entre diferentes proteínas ou produtos gênicos em uma mesma função molecular ou processo biológico
- Relacionar funções e processos ao modo de vida do organismo e utilizar esta informação em análises comparativas

Função molecular, localização celular, processo biológico

- Comparação com domínios conservados de genes associados à cada categoria
- Comparação com ortólogos caracterizados em cada categoria, e evolutivamente relacionados ao grupo analisado
 - Por exemplo: Ao analisar uma linhagem de uma espécie da classe Dothideomycetes, utilizar os ortólogos desta classe e não ortólogos de Sordariomycetes

Anotação
gênica

Anotação de
elementos
transponíveis

Anotação funcional:
- Efetores
- CAZymes
- Clusters de
metabólitos
secundários

Anotação funcional:
- Função molecular
- Localização celular
- Processo biológico

Análises
comparativas

- Responder à pergunta inicial
 - Testar as hipóteses
- Sugerir novas perspectivas
 - Fornecer bases para estudos futuros