

Acessando o potencial bioativo de *Streptomyces* sp. MAD39, isolada de sedimentos do rio Madeira baseado em análises *in* *silico* do genoma

Pereira, Sílvia Vitória Cruz Gonçalves¹; Yamagishi, Michel Eduardo Beleza²;
Silva, Gilvan Ferreira³; Caniato, Fernanda Fatima⁴

¹Universidade Federal do Amazonas, ²Embrapa Informática Agropecuária,³Embrapa Amazônia Ocidental, ⁴Universidade Federal do Amazonas.

Email: silvagf@gmail.com

Resumo

A microbiota amazônica tem sido foco de pesquisas usando diferentes abordagens, e os resultados preliminares têm revelado grande diversidade e possibilitado a identificação de inúmeras novas espécies. A diversa microbiota amazônica tem sido também investigada quanto ao seu potencial bioativo seja para a produção de enzimas de interesse biotecnológico quanto para à atividade antimicrobiana contra patógenos que comprometem a produção de culturas de importância para o estado do Amazonas como o guaraná e açaí. Neste estudo, conduzimos o sequenciamento do genoma do *Streptomyces* sp. MAD39 isolado de sedimentos do Rio Madeira, visando identificar fatores associados com potencial bioativo em seu genoma, com ênfase em CAZymes. O tamanho do genoma *Streptomyces* sp. MAD39 foi de 15,44 Mb e conteúdo de G+C de 72,11%. Foram preditas 455 Cazymes: 44,01% glicosil hidrolases (GH = 224), 10,02% carboidratos esterases (CE = 51), 15,72% módulo de ligação a carboidratos (CBM = 80), 6,29% enzimas com atividades auxiliares (AA = 32), 2,16% polissacarídeos liases (PL = 11) e 21,81% glicosil transferase (GT = 111). *Streptomyces* sp. MAD39 parece capaz de degradar vários substratos, entre eles celulose, hemicelulose, quitinas, quitosanas, amido, etc. *Streptomyces* sp. MAD39 apresentou um número notável de 37 cópias de enzimas relacionadas a hidrólise de amido das famílias GH13 e GH15, enzimas estas com aplicação conhecida na indústria de alimentos. Os resultados obtidos precisam ainda ser relacionados com dados funcionais conduzidos em diferentes substratos.

Palavras-Chave: Microbiota amazônica; Cazymes; Recursos naturais

Introdução

O interesse humano na exploração sustentável dos recursos naturais disponíveis visando atender às necessidades da agricultura, meio ambiente e indústrias de fármacos/cosméticos tem crescido nos últimos anos (Gupta et al. 2019; Ma 2019; Pham et al. 2019; Avila et al. 2021). Neste contexto, o potencial

de uso de microrganismos para suprir essas demandas tem sido intensamente investigado e seus empregos já são realidades nestes setores. A agricultura já conta com produtos comerciais que promovem a fixação biológica de nitrogênio como os Rizoliq LLI e Rizoliq TOP HC (<https://rizobacter.com.br>) à base de *Bradyrhizobium japonicum* e Masterfix (<https://www.stoller.com.br>) à base de *Azospirillum brasilense*. Outro exemplo bem sucedido da utilização de microrganismos no manejo agrícola tem sido o emprego de *Trichoderma* spp. como agente de biocontrole contra vários patógenos, como *Rhizoctonia solani*, *Fusarium moniliforme*, *Alternaria triticina*, *Ustilago segetum*, *Fusarium udum*, entre outros (Puyam 2016).

Em consonância com a tendência de exploração sustentável dos recursos naturais, esforços têm sido direcionados para a identificação e caracterização de centenas de microrganismos isolados da microbiota amazônica, incluindo microrganismos de sedimentos coletados em 7000 km de diferentes rios amazônicos. A coleção de microrganismos resultante dos esforços de isolamento, identificação e caracterização encontra-se hospedada na Embrapa Amazônia Ocidental, Manaus, Amazonas, e conta com representantes entre outros do versátil gênero *Streptomyces*.

Streptomyces são bactérias gram-positivas, filamentosas e produtoras de esporos, reconhecidas pelas inúmeras habilidades as quais destacam a produção de antibióticos, promoção de crescimento de plantas e biocontrole de patógenos de interesse agrícola (Sousa e Olivares 2016). Pontuando alguns exemplos, temos o *Streptomyces roseocinereus* MS1B15 que possui inúmeras habilidades, as quais destacam a de solubilizar fósforo, produzir sideróforos, sintetizar ácido indol acético, atividade antimicrobiana contra patógenos *Fusarium*, *Botrytis* e *Phytophthora* (Chouyia et al. 2020). *Streptomyces* sp. CB-75 isolado de solo de rizosfera de bananeira exibiu atividade antifúngica contra 11 diferentes patógenos e promoveu o crescimento de plantas de banana (Chen et al. 2018). Em sua revisão, Quinn et al. (2020) apresentam uma lista de moléculas importantes do ponto de vista clínico e econômico produzida por membros do gênero *Streptomyces*, com atividades anticâncer, antibiótica, antifúngica, herbicida, etc. *Streptomyces* sp. MAD39 isolado de sedimentos do Rio Madeira, integrante da coleção de isolados da Embrapa Amazônia Ocidental, tem chamado atenção diante dos resultados preliminares que indicam

potencial atividade antimicrobiana, com inibição *in vitro* acima de 73% contra os patógenos *Colletotrichum siamense*, *Fusarium decemcellulare* e *Moniliophthora perniciosa* bem como a identificação de 124 clusters biossintéticos no seu genoma. Diante do potencial bioativo do *Streptomyces* sp. MAD39 conduzimos a anotação do seu genoma para enzimas ativas de carboidratos (CAZymes) visando identificação de fatores com potencial bioativo entre os membros deste grupo de enzimas.

O objetivo desse trabalho foi realizar análises *in silico* no genoma do isolado de *Streptomyces* sp. MAD 39 visando a identificação de fatores associados com potencial bioativo em seu genoma, com ênfase em CAZymes.

Material e Métodos

O *Streptomyces* sp. MAD39, isolado de sedimentos do Rio Madeira (SISGEN Nº AB6B14F), teve o seu genoma sequenciado por *shotgun* Illumina com *reads paired-end* de 150 bp e cobertura de 100X. As inúmeras *reads* obtidas do sequenciamento foram pré-processadas para remoção de adaptadores utilizando o software Ktrim (Sun 2020). As *paired-reads* foram unidas usando o FLASH (Magoc e Salzberg 2011) e então montadas em 4740 *contigs* usando o software SPAdes (Prijbelski et al. 2020). Foi feito o *upload* do arquivo de montagem na workspace da KBase (US Department of Energy Knowledge Database (Arkin et al. 2018) para a anotação do genoma do *Streptomyces* sp. MAD39, que foi conduzida em duas diferentes plataformas: Prokka v1.14.5 (Seemann 2014) e RASTtk v1.073 (Brettin et al. 2015).

As CDSs, "*CoDing Sequence genoma*" do isolado de *Streptomyces* sp. MAD39 foram submetidas ao servidor da web dbCAN2 (<http://bcbl.unl.edu/dbCAN2/>) utilizando a ferramenta HMMER (e-value < 1e⁻¹⁵, coverage > 0,35) para a anotação de enzimas relacionadas ao metabolismo de carbono de acordo com a classificação da CAZy database (<http://www.cazy.org>): glicosil transferase (GT), glicosídeo hidrolases (GH), polissacarídeos liases (PL), carboidratos esterases (CE), enzimas com atividades auxiliares (AA) e módulo de ligação a carboidratos (CBM). A apresentação gráfica dos dados obtidos foi feita usando o pacote do R "*ggplot2*" v.3.4.0 (<https://cran.r-project.org/web/packages/ggplot2/index.html>).

Resultados e Discussão

A anotação automática do genoma do *Streptomyces* sp. MAD39 pelo Prokka e RASTtk foram similares: tamanho do genoma de 15,44 Mb, conteúdo de G+C de 72,11%. O genoma do *Streptomyces* sp. MAD39 consistiu de 14371 e 17359 CDSs anotadas respectivamente pelo Prokka e RASTtk. Notavelmente, *Streptomyces* sp. MAD39 apresentou tamanho de genoma superior ao dos demais genomas de *Streptomyces* depositado no NCBI, sendo o maior deles do *Streptomyces rapamycinicus* NRRL 5491 com 12,70 Mb (Baranasic et al. 2013). Um total de 455 CDSs codificando CAZymes foram identificadas no genoma do *Streptomyces* sp. MAD39, assim distribuídas: 44,01% glicosil hidrolases (GH = 224), 10,02% carboidratos esterases (CE = 51), 15,72% módulo de ligação a carboidratos (CBM = 80), 6,29% enzimas com atividades auxiliares (AA = 32), 2,16% polissacarídeos liases (PL = 11), e 21,81% glicosil transferase (GT = 111) (Figura 1). A predição de CAZymes no genoma de duas linhagens de *Streptomyces* sp. denominadas F1 e F7, previamente selecionadas quanto ao potencial celulolítico e hemicelulolítico, resultou na predição de 175 e 193 CAZymes, respectivamente (Melo et al. 2018). Em termos absolutos, o número de CAZymes detectadas no genoma do MAD39 foi bastante superior aos do F1 e F7 (MAD39 = 455 vs. F1 = 175 e F7 = 193). Relativo ao número de genes codificantes, embora a diferença tenha sido nivelada, MAD39 ainda apresentou maior porcentagem de Cazymes que F1 e F7 (MAD39 = 3,17% vs. F1 = 2,36% e F7 = 2,92%).

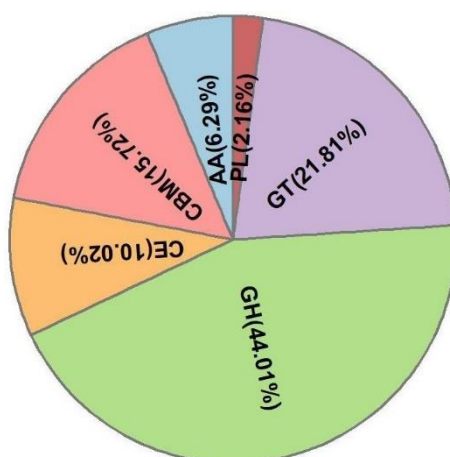


Figura 1. Distribuição das classes de CAZymes no genoma do *Streptomyces* sp. MAD39. Abreviações: GH, glicosil hidrolases. CE, carboidratos esterases; CBM, módulo de ligação a carboidratos; AA, enzimas com atividades auxiliares; PL, polissacarídeos liases e GT, glicosil transferase.

A classe das GHs é amplamente conhecida por conter atividades catalíticas envolvidas na degradação da celulose, hemicelulose, lignina, etc. No genoma do *Streptomyces* sp. MAD39 a classe das GHs constituiu a maior fração das CAZymes anotadas. As 224 GHs identificadas foram distribuídas em 52 famílias. Encontramos 80 CBMs distribuídas em 15 famílias, 111 GTs distribuídas em 14 famílias, 51 CEs distribuídas em 9 famílias, 32 AAs e 11 PLs cada uma distribuídas em 7 famílias (Figura 2).

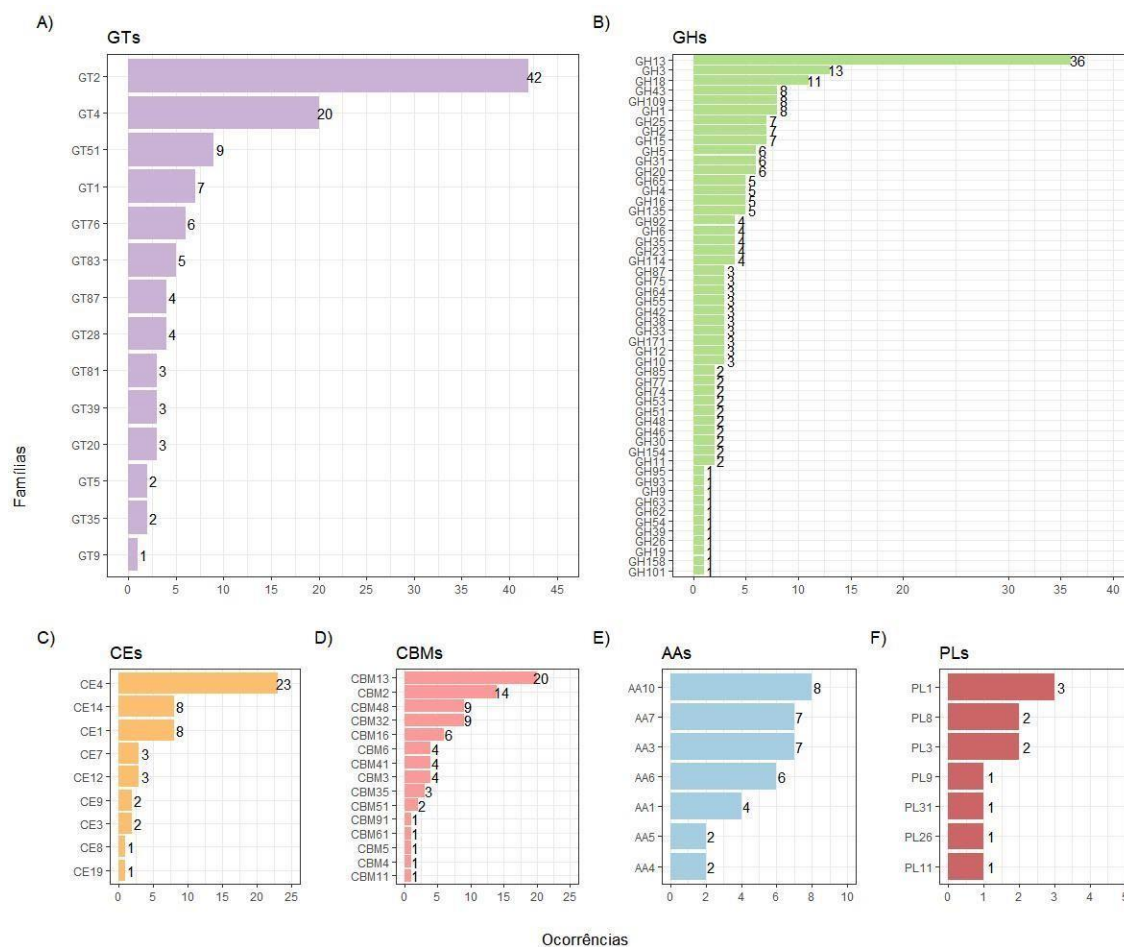


Figura 2. Famílias de CAZymes no genoma do *Streptomyces* sp. MAD39: A) GT, glicosil transferase; B) GH, glicosil hidrolases; C) CE, carboidratos esterases; D) CBM, módulo de ligação a carboidratos; E) AA, enzimas com atividades auxiliares e F) PL, polissacarídeos liases.

Streptomyces sp. MAD39 parece capaz de degradar vários substratos (Figura 3A-L). Entre as CDS do genoma de *Streptomyces* sp. MAD39, nenhuma foi categorizada como celulosoma por não apresentarem *hits* com domínios coesinas, doquerinas e SLH (*surface layer homology*). Entretanto, encontramos cópias de GHs que compõem o complexo multi-enzimático de celulasas formado pelas celobiohidrolases (CBH, EC 3.2.1.91), endoglucanase (EGL, EC 3.2.1.4) e β - glucosidase (BGL, EC 3.2.1.21) (Figura 3A). Onze das 27 enzimas com

envolvimento putativo na degradação de celulose estavam associadas com CBMs (CBM2: 5, CBM3: 4, CBM4: 1 e CBM11: 1). Em *Streptomyces* consideradas celulolíticas foram encontrados os módulos CBM2, CBM3, CBM4, e CBM46 associados com domínio catalítico GH, sendo CBM2 o mais comum (Sidar et al. 2020).

Encontramos no genoma do *Streptomyces* sp. MAD39 hemicelulases putativa (Figura 3B, 3C e 3E). Considerando a hidrólise enzimática de xilanas encontramos sete cópias de GH43 (Figura 3B), para xiloglucano encontramos seis cópias da GH1, duas cópias de cada umas das GH31, GH35 e GH74 (Figura 3C) e para manana quatro cópias das GH92 e três cópias de cada GH5 e GH38 (Figura 3E).

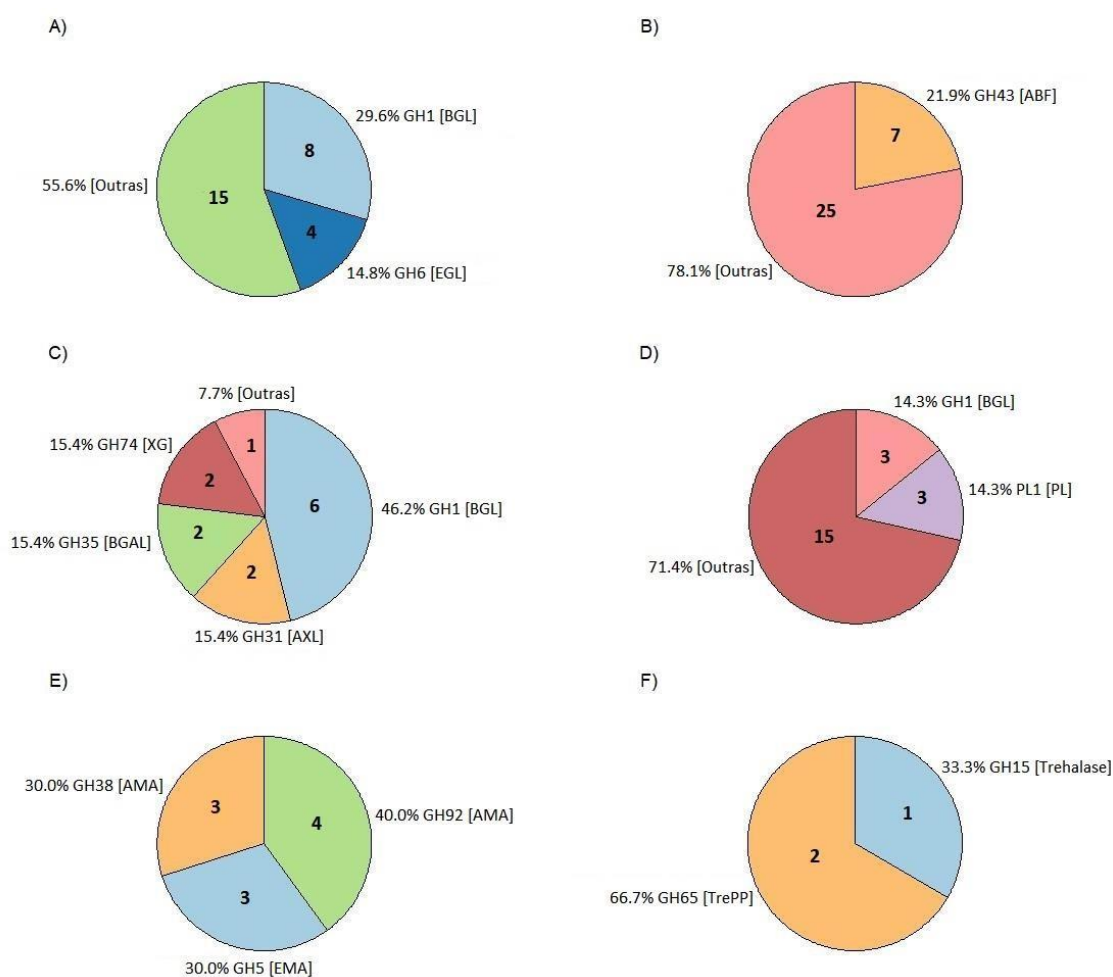


Figura 3 (parte 1). Famílias de CAZymes no genoma do *Streptomyces* sp. MAD39 de acordo com substratos: A) Celulose, B) Xilana, C) Xiloglucano, D) Pectina, E) Manana e F) Trehalose. Abreviações: BGL, β -glucosidase; EGL, endo- β -1,4-glucanase; ABF, α -arabinofuranosidase; XG, xyloglucanase; BGAL, β -arabinofuranosidase; AXL, α -xylosidase; PL, pectate lyase; AMA, α -mannosidase; EMA, endo-1,4- β -mannanase; TrePP, trehalose-6-phosphate phosphorylase; Trehalase, α,α -trehalase.

A hidrólise enzimática de quitina pode ser mediada pela ação das monooxigenases líticas de polissacarídeos (LPMOs), quitinases e β -N-acetylhexosaminidase resultando em quito-oligossacarídeos. No genoma do *Streptomyces* sp. MAD39, encontramos sete LPMO da família AA10, quatro delas apresentando domínio único e outras três associadas com CBM2 ou CBM5. As quitinases do *Streptomyces* sp. MAD39 pertencem à família GH18 e GH19. Das dez GH18 encontradas, seis delas estavam associadas com CBM2 ou CBM16. A única quitinase GH19 do *Streptomyces* sp. MAD39 apresentou um único domínio. As β -N-acetylhexosaminidase do *Streptomyces* sp. MAD39 são da família GH3 e foram encontradas contendo um único domínio (Figura 3G). A hidrólise enzimática da quitosana é medida pelas quitosanases, quatro quitosanases foram encontradas no genoma do *Streptomyces* sp. MAD39, duas cópias de cada uma das famílias GH46 e GH75 (Figura 3H). Entre as potenciais aplicações de quitinases e quitosanases destaca-se o tratamento de águas residuais (Nechita 2017).

Foi notável o número de enzimas relacionadas à hidrólise de amido no genoma *Streptomyces* sp. MAD39, principalmente representadas pelas Cazymes das famílias GH13 (n=32) e GH15 (n=5), com atividades preditas de isoamilase (EC: 3.2.1.68), α -glucosidase (3.2.1.20), glucoamilase (EC: 3.2.1.3) e outras (Figura 3K). As isoamilases do MAD39 foram das famílias GH13_11 e GH13_13 presentes com um único domínio ou associadas com CBM41 ou CBM48. As isoamilases têm aplicação no processamento industrial de amido para a produção de glicose ou maltose (Harada 1984). Outra enzima relacionada à hidrólise de amido encontrada no genoma do MAD39 foi a α -glucosidase, todas da família da GH13_30 apresentando um único domínio. A α -glucosidase de *Geobacillus stearothermophilus* U2 apresentou potencial na biotransformação de amido em isomalto oligossacarídeos (IMO), com conteúdo resultante dos oligossacarídeos de 41,74% e de glicose foi de 9,81% (Chen et al. 2018). As glucoamilases encontradas no genoma do MAD39 foram todas da família GH15. Esta enzima tem sido utilizada na produção de dextrose pela indústria de panificação, na fabricação de cerveja de baixa caloria e hidrólise de grãos integrais pela indústria do álcool (James e Lee 1997). Encontramos também enzimas de ramificação da família da GH13_9 que possuem envolvimento na

síntese de glicogênio, sendo capazes de usar seus próprios produtos alongados como substratos (Gaenssle et al. 2021).

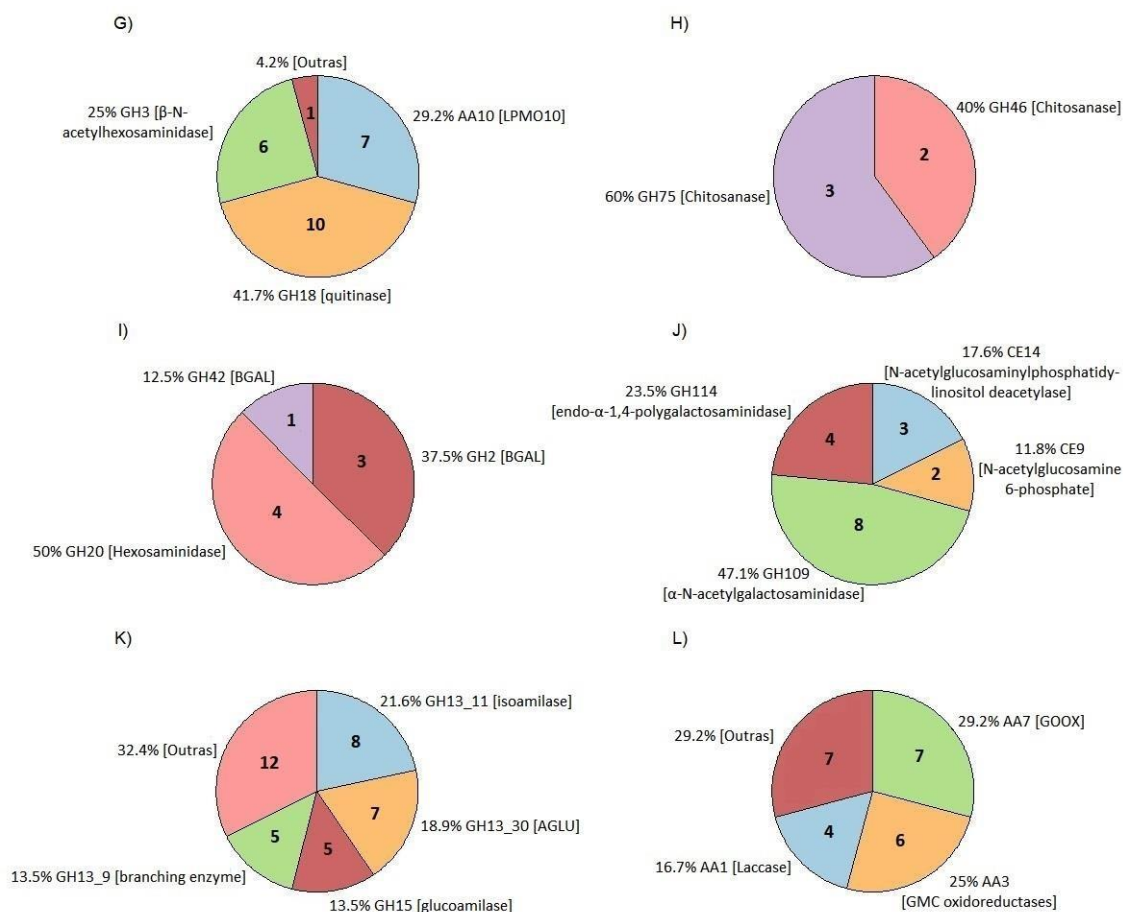


Figura 3 (parte 2). Famílias de CAZymes no genoma do *Streptomyces* sp. MAD39 de acordo com substratos: G) Quitina, H) Quitosana, I) Oligossacarídeos, J) Polissacarídeos, K) Amido e L) Lignina. Abreviações: LPMO10, Lytic chitin monooxygenase; BGAL, β-galactosidase; Hexosaminidase, β-hexosaminidase; AGLU, α-glucosidase; GOOX, glucooligosaccharide oxidase; GMC oxidoreductases, glucose methanol choline oxidoreductases.

Conclusões

A predição de CAZymes no genoma *Streptomyces* sp. MAD39 resultou em 455 enzimas, sendo as GHs a classe dominante com 224 representantes.

Streptomyces sp. MAD39 parece capaz de degradar vários substratos, entre eles a celulose, devido à presença de cópias de membros do complexo multi-enzimático de celulasas: celobiohidrolases, endoglucanase e β-glucosidase. Foram também encontradas cópias de enzimas relacionadas à hidrólise de quitinas e quitosanas, entre outras.

Streptomyces sp. MAD39 apresentou um número notável de 37 cópias de enzimas relacionadas à hidrólise de amido das famílias GH13 e GH15.

Os resultados obtidos precisam ainda ser relacionados com dados funcionais conduzidos em diferentes substratos.

Agradecimentos

Os autores agradecem ao Conselho Nacional de Aperfeiçoamento de Pessoal de Nível Superior CAPES-Procad AmazonMicro e à Universidade Federal do Amazonas pela concessão de bolsa de iniciação científica para a primeira autora.

Referências

Arkin, A.P.; Cottingham, R.W.; Henry, C.S.; Harris, N.L.; Stevens, R.L.; Maslov, S.; et al. 2018. The United States department of energy systems biology knowledgebase. *Nature Biotechnology* 36(7): 566-569.

Avila, G.M.A.; Gabardo, G.; Clock, D.C.; Junior, O.S.L. 2021. Use of efficient microorganisms in agriculture. *Research, Society and Development* 10(8).

Baranasic, D.; Gacesa, R.; Starcevic, A.; Zucko, J.; Blazic, M.; Horvat, M.; et al. 2013. Draft genome sequence of *Streptomyces rapamycinicus* strain NRRL 5491, the producer of the immunosuppressant rapamycin. *Genome Announcements* 1(4): e00581-13.

Brettin, T.; Davis, J.J.; Disz, T.; Edwards, R.A.; Gerdes, S.; Olsen, G.J.; et al. 2015. A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports* 5(10).

Chen, P.; Xu, R.; Wang, J.; Wu, Z.; Yan, L.; Zhao, W.; et al. 2018. Starch biotransformation into isomaltooligosaccharides using thermostable alpha-glucosidase from *Geobacillus stearothermophilus*. *PeerJ* 6: e5086.

Chen, Y.; Zhou, D.; Qi, D.; Gao, Z.; Xie, J.; Luo, Y. 2018. Growth promotion and disease suppression ability of a *Streptomyces* sp. CB-75 from banana rhizosphere soil. *Frontiers in Microbiology* 8: 2704.

Chouyia, F.E.; Romano, I.; Fechtali, T.; Fagnano, M.; Fiorentino, N.; Visconti, D.; et al. 2020. P-solubilizing *Streptomyces roseocinereus* ms1b15 with multiple plant growth-promoting traits enhance barley development and regulate rhizosphere microbial population. *Frontiers in Plant Science* 11: 1137.

Gaenssle, A.L.O.; Bax, H.H.M.; Van der Maarel, E.C.; Jurak, E. 2021. GH13 Glycogen branching enzymes can adapt the substrate chain length towards their preferences via α -1, 4-transglycosylation. *Enzyme and Microbial Technology* 150(109882).

- Grupta, P.L.; Rajput, M.; Oza, T.; Trivedi, U.; Sanghvi, G. 2019. Eminence of microbial products in cosmetic industry. *Natural Products and Bioprospecting Springer* 9: 267- 278.
- Harada, T. 1984. Isoamylase and its industrial significance in the production of sugars from starch. *Biotechnology and Genetic Engineering Reviews* 1(1): 39-64.
- Huang. L.; Zhang, H.; Wu, P.; Entwistle, S.; Li, X.; Yohe, T.; et al. 2018. dbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Research* 46: 516-521.
- James, J.A.; Lee, B.H. 1997. Glucoamylases: Microbial sources, industrial applications and molecular biology - A review. *Journal of Food Biochemistry* 21(6): 1-52.
- Ma, Y.; Rensing, C. 2019. Editorial: Biotechnological potential of plant-microbe interactions in environmental decontamination. *Frontiers in Plant Science* 10: 1519.
- Magoc, T.; Salzberg, S.L. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27(21): 2957-2963.
- Melo, R.R.; Tomazetto, G.; Persinoti, G.F.; Sato, H.H.; Ruller, R.; Squina, F.M. 2018. Unraveling the cellulolytic and hemicellulolytic potential of two novel *Streptomyces* strains. *Annals of Microbiology* 68: 677-688.
- Nechita, P. 2017. Applications of chitosan in wastewater treatment. In: Shalaby, E. (ed). *Biological Activities and Application of Marine Polysaccharides*. IntechOpen, Cairo.
- Pham, J.V.; Yilma, M.A.; Feliz, A.; Majid, M.T.; Maffetone, N.; Walker, J.R.; et al. 2019. A review of the microbial production of bioactive natural products and biologics. *Frontiers in Microbiology* 10: 1404.
- Prjibelski, A.; Antipov, D.; Meleshko, D.; Lapidus, A.; Korobeynikov, A. 2020. Using SPAdes de novo assembler. *Current protocols in bioinformatics* 70(1): e102.
- Puyam, A. 2016. Advent of *Trichoderma* as a bio-control agent-A review. *Journal of Applied and Natural Science* 8(2): 1100 -1109.
- Quinn, G.A.; Banat, A.M.; Abdelhameed, A.M.; Banat, I.M. 2020. *Streptomyces* from traditional medicine: sources of new innovations in antibiotic discovery. *Journal of Medical Microbiology* 69: 1040-1048.
- Seemann, T. 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30(14): 2068-2069.
- Sidar, A.; Albuquerque, E.D.; Voshol, G.P.; Ram, A.F.J.; Vijgenboom, E.; Punt, P.J. 2020. Carbohydrate binding modules: diversity of domain architecture in amylases and cellulases from filamentous microorganisms. *Frontiers in Bioengineering and Biotechnology* 8(871): 1-15.

Sousa, J.A.J.; Olivares, F.L. 2016. Plant growth promotion by Streptomyces: Ecophysiology, mechanisms and applications. *Chemical and Biological Technologies in Agriculture* 3: 24.

Sun, K. 2020. Ktrim: an extra-fast and accurate adapter- and quality-trimmer for sequencing data. *Bioinformatics* 36(11): 3561-3562.

Yin, Y.; Mao, X.; Yang, J.; Chen, X.; Mao, F.; Xu, Y. 2012. dbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 40 (W1): W445-W451.