

Projeto de Iniciação Científica - PIBIC

Recuperação de genomas a partir de metagenomas visando explorar novas estratégias de desconstrução de biomassa lignocelulósica

Orientadora: Dra. Gabriela Felix Persinoti

Laboratório Nacional de Ciência e Tecnologia do Bioetanol – CTBE

Centro Nacional de Pesquisa em Energia e Materiais – CNPEM

Campinas, Abril de 2019

Introdução e estado da arte

A crescente demanda global por energia, em conjunto com políticas públicas mundiais visando a sustentabilidade, refletiram em um aumento do número de pesquisas focadas no desenvolvimento de rotas alternativas de produção de energia. A biotecnologia tem um papel fundamental no desenvolvimento de novas tecnologias e pode ser aplicada às mais diversas áreas, como design de drogas, medicina, melhoramento genético de plantas e animais, produção de biocombustíveis e bioplásticos entre outros bioprodutos.

A biomassa lignocelulósica é considerada a principal fonte de carboidratos renováveis do mundo e tem potencial para se tornar uma plataforma sustentável para a produção de bioprodutos como biocombustíveis, blocos químicos de base biológica e biopolímeros [1], com potencial de substituir produtos de origem não renovável amplamente utilizados atualmente. Nesse contexto, biocombustíveis de segunda geração como o etanol de segunda geração (2G) surge como uma alternativa viável de produção de energia, uma vez que o etanol 2G pode ser produzido a partir dos polissacarídeos que compõem a parede celular vegetal, uma matriz renovável e com potencial de ser utilizada em larga escala [2].

A lignocelulose é principalmente composta por celulose, hemicelulose e lignina, juntamente com pequenas quantidades de pectina, grupos acetil, minerais e substituintes fenólicos [3]. Um dos principais desafios da bioconversão de biomassa lignocelulósica em bioprodutos como por exemplo biocombustíveis e/ou biopolímeros, é superar sua recalcitrância, ou seja, a transformação dos polissacarídeos presentes na biomassa em açúcares fermentescíveis. Para superar a recalcitrância intrínseca da biomassa, é proposto a realização de uma etapa de pré-tratamento seguida de uma etapa de hidrólise enzimática. A etapa de pré-tratamento pode ser realizada química ou fisicamente, com o intuito de quebrar a estrutura do material lignocelulósico, a fim de aumentar a acessibilidade das enzimas celulolíticas aos polímeros de

celulose e hemicelulose. O material resultante é então tratado com hidrolases responsáveis pela hidrólise das ligações glicosídicas presentes na celulose e hemicelulose para gerar monossacarídeos.

Vários microrganismos tais como bactérias e fungos filamentosos são capazes de degradar a biomassa naturalmente a fim de obter nutrientes para sobreviverem nos microambientes em que se encontram. Além disso, comunidades microbianas presente em ambientes naturais como solos recoberto por biomassa e intestino de animais, como cupins e ruminantes, possuem naturalmente grande capacidade de obter energia a partir de biomassa lignocelulósica. Essa capacidade encontra-se codificada no potencial genético dos microrganismos presentes nestes ambientes, porém, muitas vezes esse potencial é pouco ou apenas superficialmente explorado. Neste sentido, o recente desenvolvimento das tecnologias de sequenciamento de ácidos nucleicos em larga escala, impulsionaram o uso de abordagens independentes de cultivo, como a metagenômica, a qual se tornou uma poderosa ferramenta que permite acessar o potencial genético de comunidade microbianas por meio do sequenciamento direto do DNA ambiental. Dessa forma, a metagenômica é uma técnica promissora para prospecção de enzimas degradadoras de biomassa vegetal, pois além de permitir estudos de diversidade microbiana, permite acessar o genoma desses organismos incultiváveis possibilitando a descoberta de genes envolvidos em variadas atividades enzimáticas de sistemas biológicos até então desconhecidos, que podem ser de interesse biotecnológico.

Objetivos

O presente projeto tem como objetivo recuperar genomas a partir de metagenomas de comunidades microbianas naturalmente presentes em amostras ambientais, afim de para explorar o potencial genético de forma a aprofundar o conhecimento acerca da diversidade de micro-organismos e enzimas presentes em ambientes naturais, visando compreender as estratégias naturalmente empregadas pelas comunidades microbianas para desconstrução de materiais lignocelulósicos.

Dentre os objetivos específicos destacam-se:

- Recuperar Genomas a partir de Metagenomas (MAGs)
- Realizar análises filogenéticas e taxonômica dos genomas recuperados
- Realizar a predição gênica e anotação funcional dos genomas recuperados
- Análise das vias metabólicas e enzimas para identificar clusters de genes com potencial interesse biotecnológico

Materiais e Métodos

Amostras

Serão utilizadas amostras de comunidades microbianas naturalmente presentes em solo recoberto com bagaço de cana-de-açúcar. Essas amostras foram submetidas à extração de DNA e bibliotecas de amplicons, focadas na região hipervariável V4 do gene 16S rRNA, foram

construídas e sequenciadas na instalação aberta de sequenciamento de alto desempenho NGS do CTBE, utilizando o equipamento Illumina Miseq, gerando reads paired-end de 2x250 pb. As mesmas amostras foram submetidas ao sequenciamento metagenômico utilizando o equipamento HiSeq 2500 também disponível na instalação aberta de sequenciamento de alto desempenho NGS do CTBE, que gerou reads paired-end de 2x250pb.

Sequenciamento de reads longos

A fim de complementar as análises do perfil funcional das comunidades microbianas, de forma a facilitar a recuperação de genomas completos a partir de metagenomas (MAGs) e identificação de lócus gênicos de interesse, como os PULs (Polysaccharide Utilization Loci), o sequenciamento de reads longos é uma abordagem interessante. Os reads longos podem ser utilizados como moldes para a montagem de reads curtos (obtidos pelo sequenciamento Illumina), o que irá impactar significativamente a qualidade da montagem de novo dos metagenomas, pois permite a realização de montagens mais completas, com menor fragmentação, melhorando a qualidade da predição gênica das comunidades de interesse. O sequenciamento de reads longos será realizado por meio da plataforma minION (Oxford Nanopore Technologies).

Recuperação de genomas a partir de Metagenomas

Os dados provenientes do sequenciamento de 16S serão utilizados para a avaliação da diversidade e identificação de quais microrganismos estão presentes nas amostras de comunidades microbianas naturalmente presentes em solo recoberto com bagaço de cana-de-açúcar. Estes dados serão posteriormente comparados com os dados obtidos a partir dos genomas recuperados a partir dos metagenomas.

Os dados de metagenômica foram submetidos a montagem *de novo*, usando os algoritmos, IDBA_UD [4] e MEGAHIT [5]. Os reads longos que serão obtidos por meio do sequenciamento Nanopore serão utilizados para fazer o scaffolding e melhorar as montagens obtidas. As montagens resultantes serão avaliadas quanto a diversas métricas com relação à completeza e fragmentação usando os softwares checkM [6] e Quast [7].

As montagens *de novo* dos metagenomas serão submetidas ao processo de binning, usando diversos algoritmos como CONCOCT [8] e MaxBin 2.0 [9], seguidos de processo de dereplicação para remover redundâncias, com o intuito de identificar Metagenomes Assembled Genomes (MAGs), os quais serão avaliados quanto à qualidade e completeza. MAGs com contaminação < 5% e completeza > 90% serão considerados de boa qualidade.

MAGs de alta qualidade serão submetidos ao processo de genome polishing usando o pipeline Pilon [10] empregando tanto short e long reads nesse processo, a fim de aumentar a qualidade da montagem final dos mesmos. Estes genomas serão avaliados quanto a taxonomia usando o banco de dados Gtdbtk [11] e filogeneticamente usando o pipeline UBCG [12] visando a identificação de possíveis novos genomas.

Os genomas recuperados à partir dos metagenomas serão submetidos ao processo de predição e anotação gênica, utilizando o pipeline Prokka [13], a fim de compreender as estratégias de degradação de biomassa lignocelulósica e identificação de enzimas e vias metabólicas com potenciais aplicações biotecnológicas.

Referências

- [1] F.H. Isikgor, C.R. Becer, Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers, *Polym. Chem.* 6 (2015) 4497–4559. doi:10.1039/C5PY00263J.
- [2] Utilization of bioresources for sustainable biofuels: A Review, *Renew. Sustain. Energy Rev.* 73 (2017) 205–214. doi:10.1016/J.RSER.2017.01.070.
- [3] S.P.S. Chundawat, G.T. Beckham, M.E. Himmel, B.E. Dale, Deconstruction of Lignocellulosic Biomass to Fuels and Chemicals, *Annu. Rev. Chem. Biomol. Eng.* 2 (2011) 121–145. doi:10.1146/annurev-chembioeng-061010-114205.
- [4] Y. Peng, H.C.M. Leung, S.M. Yiu, F.Y.L. Chin, IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth., *Bioinformatics.* 28 (2012) 1420–8. doi:10.1093/bioinformatics/bts174.
- [5] D. Li, C.-M. Liu, R. Luo, K. Sadakane, T.-W. Lam, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph, *Bioinformatics.* 31 (2015) 1674–1676. doi:10.1093/bioinformatics/btv033.
- [6] D.H. Parks, M. Imelfort, C.T. Skennerton, P. Hugenholtz, G.W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes., *Genome Res.* 25 (2015) 1043–1055. doi:10.1101/gr.186072.114.
- [7] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: Quality assessment tool for genome assemblies, *Bioinformatics.* 29 (2013). doi:10.1093/bioinformatics/btt086.
- [8] J. Alneberg, B.S. Bjarnason, I. de Bruijn, M. Schirmer, J. Quick, U.Z. Ijaz, L. Lahti, N.J. Loman, A.F. Andersson, C. Quince, Binning metagenomic contigs by coverage and composition, *Nat. Methods.* 11 (2014) 1144–1146. doi:10.1038/nmeth.3103.
- [9] Y.-W. Wu, B.A. Simmons, S.W. Singer, MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets, *Bioinformatics.* 32 (2016) 605–607. doi:10.1093/bioinformatics/btv638.
- [10] B.J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C.A. Cuomo, Q. Zeng, J. Wortman, S.K. Young, A.M. Earl, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement., *PLoS One.* 9 (2014) e112963. doi:10.1371/journal.pone.0112963.
- [11] D.H. Parks, M. Chuvochina, D.W. Waite, C. Rinke, A. Skarshewski, P.-A. Chaumeil, P. Hugenholtz, A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life, *Nat. Biotechnol.* 36 (2018) 996. doi:10.1038/nbt.4229.
- [12] S.-I. Na, Y.O. Kim, S.-H. Yoon, S. Ha, I. Baek, J. Chun, UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction, *J. Microbiol.* 56 (2018) 280–285. doi:10.1007/s12275-018-8014-6.
- [13] T. Seemann, Prokka: rapid prokaryotic genome annotation, *Bioinformatics.* 30 (2014) 2068–2069. doi:10.1093/bioinformatics/btu153.