Universidade de São Paulo

Programa Interunidades de Pós-Graduação em Bioinformática

**Análise do transcrito de *Haemophilus influenzae* tipo b durante o processo
fermentativo em biorreator**

Dissertação de Mestrado

Carlos Eduardo Madureira Trufen

Orientador: Prof. Dr. Anatoli Iambartsev

Co-orientador: PqC Dr Enéas de Carvalho

São Paulo

2017

Universidade de São Paulo

Programa Interunidades de Pós-Graduação em Bioinformática

Análise do transcrito de *Haemophilus influenzae* tipo b durante o processo fermentativo em biorreator

Dissertação apresentada ao Programa de Pós-Graduação em Bioinformática Interunidades da Universidade de São Paulo para a obtenção do Título de Mestre em Bioinformática.

Área de concentração: Bioinformática

Orientador: Prof Dr. Anatoli Iambartsev
Co-Orientador: Dr. Enéas de Carvalho

São Paulo

2017

Agradeço ao meu orientador Prof. Dr. Anatoli Iambartsev por todo o apoio. Ao meu co-orientador Dr. Enéas de Carvalho pela oportunidade, atenção, paciência e amizade. À Dr^a Mickie Takagi pela oportunidade. A toda a equipe do Laboratório de Desenvolvimento de Processos, do Laboratório Especial de Piloto de Processos, do Laboratório de Bacteriologia II e do Laboratório de Genômica e Transcritômica do Laboratório Especial de Toxinologia Aplicada, todos do Instituto Butantan. À minha mãe, Maria Zelia Madureira e à minha avó, Elza Carrião Madureira, sem as quais não haveria alcançado meus objetivos. À minha esposa Camila Antonia Rosa por todo o apoio, paciência e compreensão. A todos que direta ou indiretamente fizeram parte da minha formação, muito obrigado.

“Logic is the beginning of wisdom...not the end”

Mr. Spock

“Once you have eliminated the impossible, whatever remains,
however improbable, must be the truth”

Sherlock Holmes (Sir Arthur Conan Doyle)

“Science in its purest form is the reduction of the
behavior of a system to a set of mathematical rules that define it”

Siliang Zhang *et. al*, 2006.

Resumo

Haemophilus influenzae (Hi) é uma bactéria Gram-negativa comensal da nasofaringe e um patógeno oportunista cujo único hospedeiro natural conhecido é o ser humano. As cepas de Hi que possuem cápsula de polissacarídeo estão associadas a doenças invasivas mais graves, sendo as de sorotipo b (Hib) as principais causadoras da meningite bacteriana em populações não vacinadas. Para produzir a vacina contra Hib, o polissacarídeo purificado desta bactéria é conjugado quimicamente ao toxóide tetânico. Industrialmente, a produção do polissacarídeo é realizada cultivando esse micro-organismo em biorreatores, entretanto o rendimento em polissacarídeo é baixo, mesmo com fornecimento de nutrientes, controle de pH e outros ajustes das condições no decorrer do cultivo. O estudo dos diferentes perfis fisiológicos da população bacteriana de Hib no decorrer do cultivo através da transcritômica traz a possibilidade de aprofundar o conhecimento sobre o metabolismo desse micro-organismo. As taxas de transcrição dos genes expressos em diferentes momentos considerados como pontos metabolicamente significativos do cultivo de Hib linhagem GB 3291 em batelada alimentada conduzido em Biorreator de 10 L, com aeração submersa e controles de pH (7,0) e temperatura (30° C) foram obtidas através de sequenciamento de RNA paralelo massivo (RNA-seq). A análise de co-expressão dos genes foi realizada com WGCNA, em que oito módulos de genes co-expressos foram identificados, quatro dos quais apresentaram correlação alta com dados fenotípicos dos cultivos, inclusive produtividade de acetato e de polissacarídeo. Análise de enriquecimento funcional identificou vias metabólicas associadas a ribossomo, síntese de parede celular, transportadores e consumo de carbono. A análise de expressão diferencial permitiu observar o comportamento desta bactéria durante o cultivo. Através da análise das taxas de transcrição dos genes foi possível identificar as principais vias de síntese de acetato e de polissacarídeo capsular, sendo esta última feita principalmente através da via de pentose fosfato, em detrimento da via de interconversão pentose-glucuronato. Nossos dados mostram que as diferentes etapas do cultivo de Hib leva à ação conjunta de vários grupos de genes, com destaque àqueles ligados às funções celulares básicas, como a síntese de proteínas e de parede celular, o transporte e a síntese de aminoácidos. Esses resultados contribuem para o entendimento dos processos bioquímicos e celulares que ocorrem durante o processo de cultivo de Hib, possibilitando que sejam feitas sugestões de modificações genéticas em Hib e alteração no processo de cultivo com propósito de diminuir produção de acetato e aumentar produção do polissacarídeo.

Abstract

Haemophilus influenzae (Hi) is a nasopharynx commensal Gram-negative bacterium and an opportunistic pathogen whose only known natural host is human being. Hi strains with polysaccharide capsule are related to more severe invasive diseases, wherein type b capsule (Hib) strains are the main cause of bacterial meningitis in unvaccinated population. To produce Hib vaccine, purified polysaccharide of this bacterium is chemically conjugated to tetanus toxoid protein. Industrially, polysaccharide production is performed by cultivating this micro-organism in bioreactors; however, the yield of polysaccharide is low, even with supply of nutrients, pH control and further adjustments of the fermentation conditions during cultivation. The study of different physiological profiles of Hib bacterial population during cultivation by transcriptomics brings the possibility to deepen the knowledge about the metabolism of this micro-organism. Transcription of genes expressed at different times considered metabolically significant points of Hib strain GB 3291 grown in fed-batch conducted in a 10 L bioreactor with submerged aeration and pH (7.0) and temperature (30 °C) control rates were obtained through massive parallel RNA sequencing (RNA-seq). Gene co-expression analysis was performed with WGCNA, in which eight modules of co-expressed genes were identified, four of which showed high correlation with cultivation data traits, including acetate and polysaccharide productivity. Enrichment analysis identified pathways related to ribosome, cell wall synthesis, transports and carbon consumption. Differential expression analysis allowed to observe this bacteria behaviour during cultivation. Through transcription rate analysis, it was possible to identify the main pathways for acetate and polysaccharide synthesis, which is through pentose phosphate pathway instead of glucuronate-pentose pathway. Our data show that different stages in Hib cultivation leads to joint action of several gene groups, highlighting genes related to basic cellular roles, like protein and cell wall synthesis, transport and aminoacid synthesis. These results contribute to the understanding of biochemical and cellular processes that occur during Hib cultivation process, allowing suggestions to be made to modify Hib gene circuitry and to change cultivation process in order to decrease acetate production and to decrease acetate production and increase polysaccharide production.

Lista de Figuras

Figura 1 – Organização genética da cópia truncada do locus <i>cap b</i> de <i>Haemophilus influenzae</i> tipo b.....	13
Figura 2 - Perfil cinético de biomassa, glicose consumida, formação de acetato, síntese de PRP e produção de CO ₂	38
Figura 3 – Perfil cinético de produtividade de PRP e produtividade de acetato.....	39
Figura 4 – Resultados do teste de integridade.....	40
Figura 5 – Curvas da saturação da detecção de genes em função do número de fragmentos, para cada amostra	42
Figura 6 – <i>Boxplots</i> da razão logarítmica da contagem de <i>reads</i> de cada gene de cada amostra em relação à mediana entre as amostras	43
Figura 7 – Análise de topologia de rede para potências de <i>soft-thresholding</i> entre 2 e 30.....	46
Figura 8 – Ajuste à topologia sem escala, com parâmetro $\beta = 7$	47
Figura 9 – Árvore de clusterização hierárquica.....	48
Figura 10 – Visualização da rede gênica com <i>heatmap</i>	49
Figura 11 – <i>Barplot</i> do autogene de cada módulo.....	50
Figura 12 – Associações módulo-dados associados aos cultivos.....	51
Figura 13 – Comparação de enriquecimento funcional nos módulos de genes.....	52
Figura 14 – Importância do gene para produtividade de PRP (eixo y) vs. conectividade intra-modular (eixo x) plotado separadamente para cada módulo encontrado.....	53
Figura 15 – Importância do gene para produtividade de acetato (eixo y) vs. conectividade intra-modular (eixo x) plotado separadamente para cada módulo encontrado.....	54
Figura 16 - Comparação de enriquecimento de genes diferencialmente expressos.....	68
Figura 17 – Expressão dos genes HIB_14020 e HIB_13620.....	77
Figura 18 – Vias metabólicas principais de <i>Haemophilus influenzae</i> tipo b envolvidas no consumo aeróbio de glicose e na síntese de acetato, dióxido de carbono e biomassa.....	78
Figura 19 – Vias metabólicas de <i>Haemophilus influenzae</i> tipo b que podem ser direcionadas para síntese de PRP.....	80
Figura 20 – Expressão dos genes HIB_06830, hib_06820 e HIB_06790.....	81
Figura 21 – Expressão dos genes HIB_11930, HIB_06910 e HIB_05940.....	82
Figura 22 - Expressão dos genes HIB_12710, HIB_12720 e HIB_06910.....	83

Lista de Tabelas

Tabela 1 – Parâmetros cinéticos associados ao cultivo de Hib a serem correlacionados com expressão gênica.....	28
Tabela 2 – Relação dos 10 genes com maior conectividade intra-modular (<i>hubs</i>) do módulo <i>Brown</i>	56
Tabela 3 – Relação dos 10 genes com maior conectividade intra-modular (<i>hubs</i>) do módulo <i>Black</i>	57
Tabela 4 – Relação dos 10 genes com maior conectividade intra-modular (<i>hubs</i>) do módulo <i>Purple</i>	59
Tabela 5 – Relação dos 10 genes com maior conectividade intra-modular (<i>hubs</i>) do módulo <i>Blue</i>	61
Tabela 6 – Relação dos 10 genes com maior conectividade intra-modular (<i>hubs</i>) do módulo <i>Red</i>	62
Tabela 7 – Relação dos 10 genes com maior conectividade intra-modular (<i>hubs</i>) do módulo <i>Magenta</i>	64
Tabela 8 – Relação dos 10 genes com maior conectividade intra-modular (<i>hubs</i>) do módulo <i>Green</i>	66
Tabela 9 – Relação dos 10 genes com maior conectividade intra-modular (<i>hubs</i>) do módulo <i>Pink</i>	67
Tabela 10 – Lista de genes diferencialmente expressos.....	73

Sumário

1. Introdução	9
1.1 O microorganismo	9
1.2 Doenças causadas por <i>Haemophilus influenzae</i>	10
1.3 Cápsula de polissacarídeo	12
1.4 Vacinas	14
1.5 Produção de PRP	15
1.6 Transcritômica	17
1.7 RNA-seq	18
2. Objetivos	22
3. Materiais e Métodos	22
3.1. Linhagem bacteriana	22
3.2. Condições de cultivo	22
3.3 Amostragem	23
3.4 Análises	25
3.4.1 Determinação de massa seca	25
3.4.2 Determinação da concentração de glicose e acetato	25
3.4.3 Determinação da concentração de polissacarídeo capsular	25
3.4.4 Determinação da massa molecular	26
3.4.5 Parâmetros cinéticos	26
3.5 Extração e isolamento do RNA mensageiro	28
3.6 Construção de biblioteca de cDNA e sequenciamento	29
3.7 Processamento dos <i>reads</i> gerados no sequenciamento	30
3.8 Pré-processamento por qualidade dos <i>reads</i>	30
3.9 Alinhamento	30
3.10 Tratamento de dados	31
3.10.1 Matriz de contagem de leituras	31
3.10.2 Remoção de efeito de lote	32
3.10.3 Análise de rede de co-expressão dos genes	32
3.10.4 Análise de expressão diferencial	34
4. Resultados e Discussão	37
4.1 Cultivos	37
4.2 Teste de integridade	40

4.3 Sequenciamento e Alinhamento	41
4.4 Remoção de efeito de lote	42
4.5 Análise de co-expressão dos genes	43
4.5.1 Análise de rede de co-expressão dos genes	43
4.5.2 Análise dos principais <i>hubs</i> de cada módulo	54
4.6 Análise de expressão diferencial de genes	67
4.7 Sugestões de alterações no processo de obtenção de PRP, visando aumento de produção	75
5. Conclusões	86
6. Perspectivas	86
7. Referências bibliográficas	87
Apêndice	131
1. Escore de qualidade de mapeamento	130
2. Dados associados aos cultivos	133
3. Relação entre autogenes	144
4. <i>Heatmap</i> , <i>barplot</i> e <i>spaghetti plot</i> de cada módulo	145
5. Expressão do gene HIB_18890	155

1. Introdução

1.1 O Microorganismo

Haemophilus influenzae (Hi) é uma bactéria Gram-negativa, cocobacilar, pleomórfica, imóvel e anaeróbia facultativa cujo único hospedeiro natural é o ser humano (KILLIAN, 2005). A temperatura ótima para o seu desenvolvimento está entre 35 – 37°C; seu tamanho varia de menos de 0,5 a 3 µm de comprimento e 0,2 a 0,3 µm de largura (KILLIAN, 2005). Hi é uma espécie comensal e parte da flora bacteriana normal residente na mucosa úmida da nasofaringe humana de crianças e adultos, e pode ser transmitida a outros indivíduos pelo ar via gotículas ou por contato direto com as secreções da nasofaringe (WORLD HEALTH ORGANIZATION, 2013). Hi é um patógeno oportunista que geralmente vive no hospedeiro sem causar danos ou doenças e pode causar doenças em situações como infecções virais ou imunidade reduzida (TURK, 1984).

Hi apresenta como característica nutricional a dependência de dois fatores encontrados no sangue humano para crescimento: a) fator V (podendo ser β -nicotinamida adenina dinucleotídeo – NAD – ou nicotinamida adenina dinucleotídeo fosfato – NADP – ou nicotinamida mononucleotídeo – NMN – ou nicotinamida ribosídeo – NR) e b) fator X (também denominado hemina ou protoporfirina IX), encontrado no ser humano em hemoproteínas como hemoglobina, hemoglobina-haptoglobina e hemopexina (THJOTTA; AVERY, 1921; FILDES, 1921; LWOFF; LWOFF, 1937; EVANS *et al.*, 1974; SEALE *et al.*, 2006; GARAVAGLIA *et al.*, 2012; ZAMBOLIN *et al.*, 2016). Isso ocorre devido a esse micro-organismo não possuir os genes para as enzimas necessárias à síntese desses fatores X e V, e assim, esses fatores devem ser obrigatoriamente fornecidos em composições de meios de cultura para crescimento de Hi (WHITE; GRANICK, 1963; HERRIOTT *et al.*, 1970; REIDL *et al.*, 2000).

Hi é classificado com base na produção de uma cápsula de polissacarídeo: cepas tipáveis (também denominadas invasivas) de tipo a até f produzem cápsulas antígenicamente distintas, e cepas não tipáveis (também denominadas não invasivas - NTHi) não produzem cápsula (PITTMAN, 1931). Análises genéticas utilizando *multilocus enzyme electrophoresis* (MLEE), e posteriormente corroboradas por *multilocus sequence typing* (MLST) e sequenciamento de genes que codificam para rRNA 16S, mostram que os sorotipos c, d, e e f de Hi formam grupos monofiléticos, enquanto que os sorotipos a e b são mais diversos, além de que Hi tipáveis e Hi não tipáveis não são estreitamente aparentados (MUSSER *et al.*, 1988;

MUSSER *et al.*, 1991; MEATS *et al.*, 2003; ERWIN *et al.*, 2008; SACCHI *et al.*, 2005). Power *et al.* (2012) encontraram evidências de que a recombinação homóloga mediada por transformação desempenha um papel importante na formação da diversidade em populações naturais de Hi, e que cada cepa contribui para e pode adquirir genes do superconjunto de genes dessa espécie.

As fontes de pressão seletiva que guiam a diversidade fenotípica e consequentemente a genotípica de Hi em ambiente natural incluem: (a) a resposta imune humana; (b) recursos nutricionais nos vários microambientes da mucosa respiratória humana ou corrente sanguínea; e (c) fatores do hospedeiro necessários para a aderência bacteriana, penetração celular e transmissão hospedeiro-hospedeiro (BRUNHAM *et al.*, 1993; BERGSTROM *et al.*, 1999; RAO *et al.*, 1999; VAN ULSEN *et al.*, 2002; HERBERT *et al.*, 2002).

Mecanismos genéticos que alteram a expressão gênica ou mesmo o conteúdo genético, como variação de fase, RNA mensageiro instável, mutações pontuais, inserções, deleções e duplicações, podem gerar novas cepas de Hi com capacidade otimizada para sobrevivência sob diversas pressões seletivas, assim permitindo a essa bactéria se adaptar rapidamente a mudança de ambientes (CORN *et al.*, 1993; DUIM *et al.*, 1996; GILSDORF, 1998; WEISER, 2000; BRUANT *et al.*, 2002; ECEVIT *et al.*, 2004). Através desses mecanismos, populações distintas de Hi apresentam taxas de expressão variável dos fatores de virulência sugerindo que alguns desses fatores podem melhorar a aptidão de Hi em nichos ecológicos específicos (por exemplo, nasofaringe ou garganta), ao passo que, em outros nichos (por exemplo, no ouvido médio), estes fatores podem ser perdidos por seleção, sem qualquer efeito prejudicial sobre a capacidade de Hi para sobreviver em tais ambientes (GILSDORF *et al.*, 2004).

1.2 Doenças causadas por *Haemophilus influenzae*

Hi apresenta versatilidade para sobreviver em diferentes nichos do corpo humano e causa infecções principalmente em crianças, idosos e imunodeprimidos. A infecção e a doença causadas por Hi iniciam com a colonização da nasofaringe e, posteriormente, o organismo pode se propagar para os seios paranasais ou o ouvido médio, resultando respectivamente nas doenças não invasivas sinusite e otite média, ou ainda atingir a corrente sanguínea, causando doenças invasivas como septicemia, meningite, pneumonia bacteriana, epiglote, celulite e infecção osteoarticular, além de infecção hepatobiliar (TODD; BRUHN, 1975; DAJANI *et al.*, 1979; RUBIN; MOXON, 1983; BRADLEY *et al.*, 1987; PELTOLA,

2000; REVAI *et al.*, 2008; WORLD HEALTH ORGANIZATION, 2013; SIEGEL; WEISER, 2015; COUREUIL *et al.*, 2017).

Hi apresenta como principais fatores de virulência: a) adesinas, que atuam como mediadores na aderência a células epiteliais ou da mucosa ou outros tipos celulares; b) proteína D, que compromete a função ciliar do tecido nasofaríngeo humano; c) IgA protease, que inativa a imunoglobulina A1 (que representa 90% da IgA presente na orofaringe); d) o lipopolissacarídeo (LPS), que possui atividades endotóxicas e e) o polissacarídeo capsular (PS), que fornece resistência a fagocitose e opsonização, sendo que quase todas as cepas de Hi associadas a doenças sistêmicas possuem cápsulas de polissacarídeo (WELLER *et al.*, 1977; MOXON; VAUGHN, 1984; ZWAHLEN *et al.*, 1986; VAN HAM *et al.*, 1995; ST GEME *et al.*, 1996; ECEVIT *et al.*, 2004; REMAUT; WAKSMAN, 2004; BERENSON *et al.*, 2005; COTTER *et al.*, 2005; ERWIN; SMITH *et al.*, 2007). Os últimos dois fatores de virulência listados são antígenos de carboidratos presentes na superfície da bactéria.

Dentre os seis sorotipos tipáveis existentes, o sorotipo b (Hib) é o de maior preocupação para a saúde pública, por ser a causa quase exclusiva de infecções invasivas por esta bactéria em indivíduos não vacinados, e de doenças mais graves em crianças com idade entre 1 e 59 meses, como a meningite, que pode levar a óbitos ou deixar sequelas neurológicas a longo prazo em até 50% das crianças afetadas (PITMANN, 1931; SPROLES *et al.*, 1969; SNELL *et al.*, 1972; FORTNUM, 1992). Assim, Hib é um agente etiológico de grande importância para a saúde humana mundial.

Em populações não vacinadas, Hib é a causa dominante de meningite bacteriana não epidêmica em crianças com menos de 12 meses de idade (WORLD HEALTH ORGANIZATION, 2013). A incidência de doenças causadas por Hib é relativamente pequena nos primeiros 2 meses de vida, provavelmente devido à proteção conferida pelos anticorpos maternos adquiridos via placenta. Crianças saudáveis com mais de 24 meses de vida geralmente adquirem uma forte resposta imune quando infectadas com Hib. Assim, na maioria das populações, a maior incidência da doença é vista em crianças com idade entre 4 e 18 meses, contudo, a idade ápice de incidência varia consideravelmente (BRIERE *et al.* 2012).

Indivíduos imunocomprometidos são considerados de alto risco para doença invasiva causada por Hib. Estas condições podem incluir asplenia funcional e anatômica, infecção por HIV, deficiência na imunoglobulina, deficiência prematura do sistema complemento,

recebimento de transplante de células-tronco hematopoiéticas e recebimento de quimioterapia ou radioterapia para neoplasmas malignos (BRIERE *et al.*, 2014).

1.3 Cápsula de polissacarídeo

Em geral, as cepas encapsuladas de Hi possuem cápsula de polissacarídeo contendo duas moléculas de hexoses como subunidade de polissacarídeo. Entretanto, Hib produz um polissacarídeo capsular distinto química e antigenicamente dos demais sorotipos, com duas pentoses ao invés de hexose: estrutura primária (3)- β -D-(1-1)-ribosil-ribitol-5-fosfato (PRP) (CRISEL *et al.*, 1975). A presença deste antígeno de superfície confere propriedades patogênicas ausentes em outros sorotipos ou em NTHi: auxilia o micro-organismo a escapar da fagocitose e da opsonização por neutrófilos, além de não induzir a via alternativa de ativação do complemento, logo essa bactéria invade o sangue ou o fluido cerebrospinal com reduzida resposta inflamatória e atração de fagócitos (NOEL *et al.*, 1996).

O polissacarídeo capsular de todas as cepas de Hi encapsuladas é sintetizado por enzimas codificadas a partir do locus de encapsulação *cap* (tamanho aproximado de 17-18 kb), um retrotransposon (ou seja, sequência de DNA que pode mudar sua posição dentro do genoma), em que cada cópia dos genes de encapsulação possui o elemento de inserção (isto é, sequência de DNA que codifica para proteínas envolvidas em atividades de transposição) IS1016 de 711 bp em suas extremidades (HOISETH *et al.*, 1986; KROLL *et al.*, 1991; MAHILLON; CHANDLER, 1999; CRAIG *et al.*, 2015). Para todos os sorotipos de Hi, o locus *cap* está organizado em três regiões funcionalmente distintas, I, II e III (KROLL *et al.*, 1989; KROLL *et al.*, 1991). As regiões I e III são comuns a todos os seis tipos capsulares e contêm genes necessários para o processamento e a exportação do material capsular (SATOLA *et al.*, 2003a). Na região I, genes (*bexABCD*) codificam para um sistema transportador ABC envolvido na exportação do polissacarídeo capsular para a superfície bacteriana (KROLL *et al.*, 1990). A região II é diferente entre os sorotipos e contém genes que codificam enzimas envolvidas na biossíntese do polissacarídeo (*acs12* para Hia; *bcs1234* para Hib; *ccs1234* para Hic; *dcs12345* para Hid; *ecs12345678* para Hie; *fcs123* para Hif), específicos para cada tipo capsular (VAN ELDERE *et al.*, 1995; FOLLENS *et al.*, 1999; SATOLA *et al.*, 2003a; SATOLA *et al.*, 2003b; GIUFRE *et al.*, 2010; LÂM *et al.*, 2011). A região III contém genes (*hcsAB*) que estão envolvidos com o transporte de polissacarídeo capsular através da membrana externa e para a superfície celular (SUKUPOLVI-PETTY *et*

al., 2006). Os elementos genéticos da produção da cápsula de Hib foram caracterizados mais extensivamente em Hib, em que se constata que eventos de recombinação que resultam em cópias adicionais do locus *cap* geram variantes de Hib com maior patogenicidade proporcional ao número de cópias presentes e à quantidade de cápsula, assim como se constata que o aumento da produção da cápsula do tipo b contribui para o aumento da virulência dessas cepas (HOISETH *et al.*, 1985; HOISETH *et al.*, 1986; KROLL; MOXON, 1988; CORN *et al.*, 1993; NOEL *et al.*, 1996).

Especificamente em Hib, a região cromossômica que codifica as várias proteínas necessárias para a expressão de PRP é denominada *cap* b e aproximadamente 98% de todos os tipos de cepas de sorotipo b contêm uma duplicação em *tandem* parcial (**Figura 1**), que somada à cópia completa de *cap* b, resulta em duas cópias completas das regiões II e III de segmento de 17-18 kb, uma cópia completa da região I e uma cópia truncada da região I com uma deleção de 1,2 kb.

Figura 1. Organização genética da cópia truncada do locus *cap* b de *Haemophilus influenzae* tipo b, modificado a partir de Schouls *et al.* (2008). Genes da região I em roxo, genes da região II em amarelo, genes da região III em laranja, o elemento de inserção IS1016 em verde e *orf3* em cinza. A figura exhibe à esquerda o elemento de inserção IS1016 truncado e em fusão com o gene *bexA* incompleto. À direita, estão representados o gene *bexA* completo e parte do gene *bexB*, presentes na segunda cópia do locus *cap* b.

A presença de uma fusão do gene *bexA* e o IS-1016 incompletos ($\Delta bexA$ - Δ IS1016) na extremidade do *cap* b duplicado tende a estabilizar a produção de cápsula, ao reduzir as chances de que ocorram eventos de recombinação que possam resultar na perda dos genes *cap*, permitindo a regulação do número de cópias de *cap*, e, por conseguinte o nível de produção de cápsulas depois da recombinação homóloga, visto que os genes presentes em cada cópia estão envolvidos com a produção de cápsula (ELY *et al.*, 1986; HOISETH *et al.*, 1986; ALLAN *et al.*, 1987; KROLL; MOXON, 1988; KROLL *et al.*, 1991; KROLL *et al.*, 1993; SATOLA *et al.*, 2003a). Caso ocorra perda da cópia completa do locus *cap* em um evento de recombinação, a cópia truncada se mantém e resulta em um fenótipo sem cápsula denominado Hib⁻ (HOISETH *et al.*, 1985; KROLL *et al.*, 1988). Nos 2% restantes, uma única

cópia do locus *cap* está presente, que está também sujeita a perda (ALLAN *et al.*, 1987; ST. GEME *et al.*, 1994).

O aumento de produção da cápsula pode fornecer uma vantagem seletiva em certas situações, visto que a cápsula é uma importante capa protetora contra ataque mediado pelo sistema complemento, por exemplo (KROLL *et al.*, 1989; HALLSTRÖM; RIESBECK, 2010). Por outro lado, se houver seleção contra a cápsula, haverá vantagem para cepas cujas múltiplas cópias de segmento de 17 a 18 kb no locus *cap* foram reduzidos a uma cópia única.

1.4 Vacinas

A ideia de usar cápsula de polissacarídeo bacteriano em vacinas data da década de 1930 e envolve induzir anticorpos específicos para proteger contra bactéria patogênica (HOAGLAND *et al.*, 1938). A cápsula de polissacarídeo de Hib (PRP) possui capacidade de induzir resposta imunológica em humanos, as mesmas propriedades antigênicas em todas as cepas do tipo b, e estimula a formação de anticorpos anti-Hib (ROBBINS *et al.*, 1973). PRP é acessível a detecção pelas células do sistema imune, incluindo macrófagos, linfócitos B e células dendríticas. A imunogenicidade de antígenos de PRP a categoriza como resposta imune independente de linfócitos T de tipo 2, resposta essa que estimula anticorpos protetores sem o auxílio de linfócitos T (RIJKERS *et al.*, 1998). Em um evento de infecção bacteriana em um hospedeiro que já foi previamente imunizado com PRP puro, os receptores específicos de PRP (BCR) na superfície dos linfócitos B nativos se ligam à unidade do carboidrato do PRP, formando o complexo BCR-PRP. O estímulo dos linfócitos B desta forma resulta na indução e secreção de anticorpos IgM contra PRP. Este se liga ao PRP presente na bactéria patogênica, erradicando esse patógeno oportunista do hospedeiro (AVCI, 2013; HOEHN *et al.*, 2016).

Em 1985, PRP purificado foi certificado para uso como vacina contra Hib nos Estados Unidos, entretanto não oferecia memória imunológica, tendo como principal desvantagem a rápida redução de anticorpos estimulados pelo PRP purificado (KÄYHTY *et al.*, 1984; CENTERS FOR DISEASE CONTROL, 1985). Mesmo a estratégia de aumentar o número de injeções (dose reforço) falhou em gerar resposta intensa ou memória imunológica, e até onde essa estratégia foi considerada bem-sucedida não houve sucesso em crianças com menos de 18 meses de vida (WARD *et al.*, 1988).

A imaturidade do sistema imune de crianças e sua relativa incapacidade em comparação ao de adultos torna as crianças mais vulneráveis a infecções causadas por Hib (TURK, 1984; RIJKERS *et al.*, 1998). A zona marginal do baço humano não está completamente formada até um ou dois anos de idade, o número de linfócitos B expressando BCRs na zona marginal é menor no nascimento e a produção de anticorpos específicos para o polissacarídeo é limitada ou ausente em recém-nascidos (ANDERSON *et al.*, 1977).

A maioria das proteínas são consideradas antígenos linfócito T-dependentes, uma vez que a resposta imune contra as mesmas é dependente da estimulação de linfócitos (DINTZIS, 1992). No caso de uma vacina PRP conjugada à proteína, o antígeno é composto por uma fusão de polissacarídeo e proteína, o que é realizado ao se ligar covalentemente PRP à proteína transportadora, o que resulta na produção de vacinas Hib glicoconjugadas (STEIN, 1994). Por esse tipo de conjugação, a resposta imune humoral é induzida com características de imunidade dependente de linfócitos T, tal como maturidade de afinidade, memória e presença de imunogenicidade em crianças acima de dois meses de idade (FINN, 2004). De fato, a presença das proteínas conjugadas faz com que o antígeno polissacarídico-conjugado passe a ser T-dependente e, nesse sentido, as proteínas conjugadas promovem resposta maior na vacina de reforço, estimulam mudança de anticorpo IgG para IgM, além de resultar no estabelecimento de memória contra um antígeno específico (DINTZIS, 1992)

A primeira vacina conjugada produzida e aprovada para uso foi criada a partir de PRP conjugado ao toxóide diftérico (PRP-D), e foi retirada do mercado logo após a introdução de vacinas mais efetivas como as conjugadas à proteína da membrana externa meningocócica (PRP-OMP), proteínas transportadoras CRM₁₉₇ (PRP-CRM) e toxóide tetânico (PRP-T) (GRANOFF *et al.*, 1984; DECKER *et al.*, 1992; BOOY *et al.*, 1994; BRÖKER *et al.*, 2011; VOYSEY *et al.*, 2016). Antes do processo de conjugação, é necessária a obtenção de PRP, através do cultivo de Hib e posteriormente o isolamento (purificação) do PRP.

1.5 Produção de PRP

O investimento na melhoria da produção do polissacarídeo para atingir uma elevada produtividade em polissacarídeo é um fator que está diretamente relacionado com a viabilidade econômica do processo de produção de uma vacina baseada nessa molécula, uma vez que o tamanho da planta de produção pode ser reduzido em função do aumento da produção volumétrica de polissacarídeo e impactar no custo final da vacina, devido ao fato da

vacina conjugada apresentar um custo elevado por envolver pelo menos três processos: 1) produção, 2) purificação e 3) conjugação. Caso seja possível utilizar menos insumos para obter uma mesma quantidade de PRP, o custo do processo tende a diminuir, assim como um número menor de cultivos também seria necessário, o que também tende a diminuir o custo (menos custo de energia, hora/pessoa, desgaste do aparelho, etc).

Estudos relacionados à produção da vacina conjugada Hib tiveram início no Centro de Biotecnologia do Instituto Butantan em 2000, nos quais foram estabelecidas as condições de cultivo submerso em biorreatores visando otimizar a produção de PRP (TAKAGI *et al.*, 2003). Desde então, outros estudos vêm sendo dedicados à produção do polissacarídeo, assim como ao estabelecimento de um processo de purificação do polissacarídeo capsular e processo de conjugação deste com o toxóide tetânico (TAKAGI *et al.*, 2006; LORTHIOIS, 2007; TAKAGI *et al.*; 2007; TAKAGI *et al.*, 2008; SILVA, 2010; ALBANI, 2012; CINTRA; TAKAGI, 2012; ALBANI *et al.*, 2014; CINTRA; TAKAGI, 2014; CINTRA, 2014).

Em cultivo de Hib – linhagem GB 3291, em aerobiose com meio de cultivo complexo à base de peptona de soja - contendo glicose como fonte de carbono - constatou-se aumento na produção de acetato tal que inibe o crescimento da bactéria e inviabiliza o aumento na produção de PRP (TAKAGI *et al.*, 2006). Para otimizar a produção de polissacarídeo, diferentes estratégias de cultivos foram conduzidas, nas quais cultivo descontínuo alimentado (batelada alimentada) mostrou ser a melhor condição (SILVA, 2010). Porém, em todas as estratégias utilizadas verificou-se a produção de acetato em concentrações que levaram à inibição do crescimento celular e, em consequência, limitação da produção de polissacarídeo (SILVA, 2010). Isso ocorre provavelmente devido a este micro-organismo apresentar um genoma deficiente em vários genes para enzimas do ciclo do citrato (também denominado ciclo TCA ou ciclo de Krebs), o que resulta, em consequência, na formação de acetato como um subproduto inevitável em crescimento aeróbio (FLEISCHMANN *et al.*, 1995; TATUSOV *et al.*, 1996; OTHMAN *et al.*, 2014).

Este fato ilustra o quanto o conhecimento do metabolismo de Hib pode ajudar a determinar estratégias que melhorem a produção de PRP, podendo impactar em uma vacina mais barata e de fácil produção.

A análise da expressão gênica é uma ferramenta que permite compreender a atividade metabólica através da análise da ativação de determinados genes nas diferentes etapas de crescimento. Essas informações podem também contribuir para identificar a via de síntese predominante do polissacarídeo capsular, através da análise das taxas de expressão dos genes

relacionados a esta atividade. A estratégia geral para esta análise pode ser resumida em avaliar as taxas de transcrição dos genes que codificam para determinadas proteínas/enzimas e rastrear o enquadramento destas nas vias metabólicas.

1.6 Transcritômica

Desde a publicação em 1995 da primeira sequência de genoma bacteriano, de *Haemophilus influenzae* Rd KW20, milhares de sequências de genoma tanto de eucariotos quanto de procariotos foram publicadas e estão disponíveis (FLEISCHMANN *et al.*, 1995; REDDY *et al.*, 2014). Essas sequências permitiram a identificação de genes e o uso de comparações homólogas para avaliar suas funções biológicas, assim fornecendo dados importantes para maior compreensão do metabolismo celular desses organismos (TATUSOV *et al.*, 1996). As abordagens ômicas, tais como genômica transcritômica, proteômica e metabolômica foram usadas para ampliar a compreensão do metabolismo celular sob vários aspectos, inclusive *Haemophilus influenzae* (LINK *et al.*, 1997; EDWARDS; PALSSON, 1999; LANGEN *et al.*, 2000; SCHILLING; PALSSON, 2000; AKERLEY *et al.*, 2002; RAGHUNATHAN *et al.*, 2004; OTHMAN *et al.*, 2014).

Define-se transcrito como o conjunto completo de transcritos de uma célula (WANG *et al.*, 2009). Sua quantificação para um estágio de desenvolvimento específico ou condição fisiológica pode ajudar a estabelecer as redes regulatórias por trás de fenótipos chave (WANG *et al.*, 2009). Em particular, a análise do transcrito dos organismos com o propósito de otimizar bioprocessos tem obtido resultados positivos para diversas espécies e produtos, como por exemplo, a produção de fenol em *Pseudomonas putida* (WIERCKX *et al.*, 2008;), de xilitol em *Escherichia coli* (HIBI *et al.*, 2007), de L-lisina em *Corynebacterium glutamicum* (SINDELAR; WENDISCH, 2007) e de ácido itacônico em *Aspergillus* spp. (LI *et al.*, 2011).

Através de cultivo em laboratório, utilizando meio de cultura e condições controladas, pode-se gerar alteração na pressão seletiva à qual uma bactéria está sujeita, o que estimula alterações no padrão de expressão gênica (BLAIR *et al.*, 2013; ISHAK *et al.*, 2014; JIANG *et al.*, 2016a). As diferentes variáveis das condições de cultivo, como o meio de cultivo utilizado, afetam o transcrito, o metabolismo, e consequentemente o fenótipo das bactérias, produzindo resposta metabólica que permite às células se ajustarem às novas condições

ambientais, e assim manter a atividade celular (VARELA *et al.*, 2005; BLAIR *et al.*, 2013). Caso o mecanismo de adaptação venha a falhar, a viabilidade reduzida é seguida por diminuição na atividade celular. Uma adequação bem-sucedida também envolve mudanças no perfil de expressão gênica, o que se traduz por um grande número de genes regulado positiva ou negativamente (VARELA *et al.*, 2005).

Assim, a caracterização do perfil de expressão específica do microorganismo durante seu cultivo possibilita compreender, em nível molecular, os processos biológicos do cultivo deste microorganismo e a regulação de expressão gênica em resposta às mudanças no meio e nas condições de crescimento (VARELA *et al.*, 2005).

Em uma abordagem direta, dados de expressão gênica podem ser usados para identificar genes envolvidos em um processo biológico particular, ao selecionar genes com mudanças claras na expressão sob diferentes condições biológicas ou ao longo do tempo. Genes com respostas semelhantes em múltiplas condições são frequentemente agrupados para formar grupos funcionais ou revelar padrões coordenados de expressão (ZHANG *et al.*, 2006).

Tal análise pode gerar predições funcionais mais específicas ao integrar grupos de expressão gênica com dados complementares, como por exemplo, informação referente a consumo de substrato, biomassa e formação de produto durante o cultivo, tal que a identificação de sequências regulatórias compartilhadas pelos genes co-expressos associados a esses parâmetros podem fornecer evidência de que esses genes associados estão sob controle de transcrição em comum. A integração de perfil de expressão gênica com análise direcionada de metabólito é uma forma de identificar função gênica para acúmulo de um produto específico em microorganismos (ZHANG *et al.*, 2006).

Estudos de perfil de transcrição de Hi já realizados incluem resposta a antibióticos, caracterização de genes “hipotéticos”, identificação de genes afetados pela captação de hemina e ferro, metabolismo de carbono, resposta a variação de pH e resposta a diferentes níveis de oxigênio (GMUENDER *et al.*, 2004; KOLKER *et al.*, 2004; WHITBY *et al.*, 2006; OTHMAN *et al.*, 2014; ISHAK *et al.*, 2014; EARL *et al.*, 2015; JIANG *et al.*, 2016a; JIANG *et al.*, 2016b). Até o momento, nenhum estudo do transcrito de Hi associado à produção de polissacarídeo foi verificado.

1.6 RNA-seq

RNA-seq, ou sequenciamento de RNA paralelo massivo, se refere a procedimentos experimentais que geram leituras de sequência dos DNAs complementares (cDNAs) derivados da população de RNAs de interesse (WANG *et al.*, 2009). Esta metodologia, desenvolvida para o estudo do perfil de transcritomas, fornece uma medida muito mais precisa de níveis dos transcritos e de suas isoformas que outros métodos e permite que sejam sequenciadas bilhões de bases de uma só vez, aumentando muito a velocidade de aquisição e precisão dos dados de sequenciamento, tal que o transcrito inteiro pode ser analisado de forma quantitativa, o que promoveu uma revolução no estudo de transcritomas bacterianos (WANG *et al.*, 2009; CROUCHER; THOMSON, 2010; NAGALAKSHMI *et al.*, 2010; FILIATRAULT, 2011; GARBER *et al.*, 2011).

A primeira etapa em um experimento de RNA-seq típico deve ser a extração do RNA total a partir das amostras de interesse coletadas. Esse RNA total deve ser purificado para RNA mensageiro (mRNA), que então deve passar por testes de integridade e de concentração, seguida da depleção do RNA ribossomal, visto que o RNA mensageiro bacteriano não é poliadenilado (SCHROEDER *et al.*, 2006; CONESA *et al.*, 2016). Procede-se então a construção da biblioteca de cDNA, em que o RNA mensageiro purificado é retrotranscrito em cDNA, passa por ligação a adaptadores, adenilação, amplificação e então as amostras são sequenciadas em uma plataforma de alto rendimento (CONESA *et al.*, 2016).

Milhões de pequenas sequências de *reads* são geradas a partir de ligação com adaptadores a uma extremidade de cDNA (*single-end*) ou a ambas (*paired-end*) (OZSOLAK; MILOS, 2011). A conversão de dados de sequência em formato digital é realizada ao se utilizar algoritmos que contam o número de vezes que cada base transcrita foi sequenciada, assim gerando arquivos de texto que contêm informações das sequências (OZSOLAK; MILOS, 2011; CREECY; CONWAY, 2015).

A partir dos arquivos com informação das sequências e através de alinhamento são obtidos os números de *reads* que mapeiam a um genoma de referência ou a um conjunto de genes de sequência conhecida entre os códons de início e fim de cada gene, individualmente (CREECY; CONWAY, 2015). Além disso, RNA-seq pode ser executado mesmo quando os genomas não estão disponíveis e, assim, independe de conhecimento prévio do organismo (GRABHERR *et al.*, 2011). O alinhamento e posterior contagem de *reads* em relação à anotação resulta em uma matriz de contagem de *reads* por milhares de genes através de um

número de amostras, o que permite comparar entre condições de teste e controle, por exemplo.

O objetivo principal de muitos experimentos de RNA-seq entre condições distintas é estimar expressão gênica diferencial, cujo poder de detecção depende da profundidade de sequenciamento da amostra, da expressão e do comprimento do gene (OSHLACK; WAKEFIELD, 2009; BULLARD *et al.*, 2010; SIMS *et al.*, 2014).

Uma filtragem deve ser conduzida nos dados da matriz de contagem, devido ao fato de que contagens muito baixas de *reads* não podem ser distinguidas do ruído de fundo de forma confiável, tal que os genes com pouca ou nenhuma expressão em todas as condições sejam removidos, reduzindo assim o conjunto de dados a um subconjunto de genes que são detectáveis acima do nível de ruído de fundo e também o número de testes conduzidos em etapas subsequentes quando se estiver examinando expressão diferencial (BOURGON *et al.*, 2010; MCINTYRE *et al.*, 2011).

Posterior a esta etapa, devem ser removidos os efeitos de lote provindos de variação técnica, que resultam de várias fontes, como: a) diferenças na qualidade e na quantidade de RNA recuperado durante preparo da amostra, b) efeito de lote do preparo da biblioteca, c) viés de adaptador e d) efeito da *lane* e do *flow cell* ao utilizar tecnologia Illumina (AUER; DOERGE, 2010; LEEK *et al.*, 2010; FANG; CUI, 2011; MCINTYRE *et al.*, 2011).

A normalização dos dados de RNA-seq no contexto de análise diferencial é necessária para garantir que a distribuição de expressões em cada amostra seja similar entre os experimentos, de tal forma que viés de efeito técnico sistemático tenha impacto minimizado nos resultados (CREECY; CONWAY, 2015; LAW *et al.*, 2016). Os métodos de normalização DESEQ e TMM são robustos tanto para a presença de tamanhos de biblioteca diferentes quanto para composições de biblioteca amplamente distintas, ambos típicos de dados de RNA-seq (ANDERS; HUBER, 2010; ROBINSON; OSHLACK, 2010; DILLIES *et al.*, 2012). Ambos os métodos utilizam fatores de normalização dentro de modelo estatístico para análise diferencial, ao invés de alterar os dados.

Além da normalização e da remoção de efeito de lote, é importante encontrar uma distribuição estatística que aproxime a natureza dos dados (WOLF, 2013). Para experimentos com quantidade pequena de réplicas por condição, costuma-se usar as famílias de distribuição normal ou binomial negativa, tal que estimativas de variância biológica devem ser incorporadas no teste para expressão diferencial, enquanto que o teste exato de Fisher ou o teste da razão de verossimilhança são os testes de hipótese mais apropriados para testar

expressão diferencial gênica (BULLARD *et al.*, 2010; KVAM *et al.*, 2012; OBERG *et al.*, 2012; WILLIAMS *et al.*, 2014).

Uma vez quantificados e normalizados os valores de expressão, além de definida a distribuição estatística, a inferência estatística paramétrica se torna viável, tal que se possa analisar a expressão diferencial, isto é, comparar formalmente os níveis de expressão entre os transcritos e amostras entre condições experimentais distintas que se quer comparar (GARBER *et al.*, 2011; OZSOLAK; MILOS, 2011; WOLF, 2013).

O poder estatístico para detectar a expressão diferencial varia com o tamanho efetivo da biblioteca, a profundidade do sequenciamento (número de *reads* por amostra) e o número de réplicas por grupo (CONESA *et al.*, 2016). Calcula-se a cobertura de uma determinada amostra ao dividir-se o rendimento (comprimento do *reads* multiplicado pelo número de *reads*) pelo comprimento do genoma. Em experimentos de RNA-seq microbiano, são recomendadas pelo menos três replicatas independentes para que sejam obtidos resultados biologicamente significativos, enquanto que é suficiente uma cobertura de 2 a 3 milhões de *reads* por amostra para detectar expressão diferencial (HAAS *et al.*, 2012; MANGA *et al.*, 2016).

Os dados de contagem de *reads* provenientes de RNA-seq também podem ser usados para estudar os perfis de co-expressão dos genes, em que a correlação entre as mudanças de expressão para cada par de genes é analisada, a fim de se determinar quais genes são comumente regulados, sendo que genes com padrão de expressão similar podem pertencer a uma mesma via metabólica e a circuitos regulatórios e de sinalização (EISEN *et al.*, 1998; ZHANG; HORVATH, 2005; IANCU *et al.*, 2012).

Do ponto de vista de produção de PRP, as informações resultantes das análises de expressão gênica juntamente com a avaliação do mapa metabólico do Hib são ferramentas que poderão contribuir para a otimização da produção de polissacarídeo, pois uma vez compreendidos os motivos que fazem com que esse micro-organismo pare de crescer em determinado momento do cultivo e de produzir o polissacarídeo mesmo na presença de nutrientes, novas estratégias poderão ser testadas, seja modificando a maneira de cultivar o micro-organismo ou mesmo através de modificação no micro-organismo por meio de nocaute de algum gene que esteja favorecendo a formação do acetato ou a inclusão de algum gene que favoreça a formação do polissacarídeo, por exemplo. A análise de taxas de transcrição em escala genômica durante o cultivo de Hib pode contribuir para este processo.

Com o presente trabalho, pretende-se investigar, através de RNA-seq, as taxas de transcrição dos genes expressos em diferentes momentos considerados pontos metabolicamente críticos (condições distintas) do cultivo de Hib linhagem GB 3291 em batelada alimentada conduzido em Biorreator de 10 L, com aeração submersa e controles de pH e temperatura.

Em Materiais e Métodos, descrevemos a linhagem bacteriana utilizada, condições de cultivo e critério para coleta de amostra, as análises associadas aos cultivos, a extração e isolamento de RNA mensageiro a partir das amostras, conversão em cDNA e preparo de biblioteca, sequenciamento, processamento de *reads* e pré-processamento por qualidade dos *reads*. São descritas as metodologias para análise ponderada de co-expressão gênica e de análise de expressão diferencial e então os resultados destas análises são apresentados e discutidos, finalizando com sugestões de modificações a serem feitas em Hib com a finalidade de diminuir produtividade de acetato e aumentar a produtividade de PRP.

2. Objetivos

- Analisar, em escala genômica, as taxas de transcrição dos genes de Hib durante a dinâmica de cultivo de Hib linhagem GB 3291 em batelada alimentada conduzido em Biorreator de 10 L, com aeração submersa e controles de pH e temperatura.
- Analisar quais genes são diferencialmente expressos nos diferentes fenótipos considerados metabolicamente críticos.
- Analisar a matriz de co-expressão gênica durante o tempo de cultivo.
- Avaliar as possíveis correlações entre as taxas de transcrição e dados físicos e fenotípicos do cultivo, além de possíveis correlações entre dados de coexpressão (autogene) e os dados físicos e fenotípicos.
- Sugerir possíveis alterações no cultivo e/ou modificações na bactéria para reduzir a quantidade de acetato produzida e/ou aumentar a produção de PRP

3. Materiais e Métodos

3.1. Linhagem bacteriana

Foi utilizada a bactéria *Haemophilus influenzae* sorotipo b linhagem GB 3291, fornecida pelo Instituto Adolfo Lutz na forma liofilizada. Os lotes semente e de trabalho

foram preparados no laboratório de Desenvolvimento de Processos do Instituto Butantan conforme descrito abaixo.

3.2. Condições de cultivo

Foram realizados quatro cultivos independentes. Para tanto, as células do lote de trabalho foram descongeladas e expandidas em tubos inclinados de ágar chocolate (Infusão de cérebro e coração – BHI - BD 37,0 g.L⁻¹, Bacto Ágar – BD 20 g.L⁻¹, sangue de cavalo - Instituto Butantan 10% v/v (adicionado asépticamente após a autoclavação do meio base nutriente em BHI, a uma temperatura de aproximadamente 60 °C)), e congelados em meio de Greaves (1960) como preservativo, e armazenadas em nitrogênio líquido com viabilidades acima de 10⁹ UFC.mL⁻¹. Para a realização dos cultivos, 200 µL das células foram descongeladas e incubadas em 50 mL de meio MMP (Glicose – Merck 5,0 g.L⁻¹, Extrato de Levedura UF – BD 5,0 g.L⁻¹, Soytone – BD 10,0 g.L⁻¹, Na₂HPO₄ – Sigma-Aldrich 13,1 g.L⁻¹, NaH₂PO₄.H₂O – Sigma 3,3 g.L⁻¹, KH₂PO₄ – Merck 2,5 g.L⁻¹, NaCl – Anidrol 2,5 g.L⁻¹, Hemina bovina – Sigma-Aldrich 30,0 mg.L⁻¹, NAD – Sigma-Aldrich 15,0 mg.L⁻¹, pH ajustado para 7,5 com HCl 6N – Merck), durante no mínimo 6 h, de forma estática e sob atmosfera de CO₂, a 37 °C (TAKAGI *et al.*, 2006). Ao atingir DO₅₄₀ de no mínimo 0,500, a suspensão bacteriana foi transferida para três frascos de Erlenmeyer contendo 200 mL de meio MMP fresco cada. O volume transferido foi definido como o necessário para se atingir as DO₅₄₀ iniciais de 0,025, 0,050 e 0,075. Os frascos foram incubados em *shaker* orbital (New Brunswick), a 150 rpm e 37 °C. Após atingir DO₅₄₀ de no mínimo 5,0 em aproximadamente 13h, que equivale ao final da fase de crescimento exponencial, a suspensão foi transferida para o reator em quantidade suficiente para atingir uma DO₅₄₀ inicial de 0,150.

O meio de alimentação é modificado do meio MMP, com alterações quanto às concentrações de extrato de levedura e de glicose, ambas aumentadas para 20 g.L⁻¹. A proporção de fosfato de sódio mono e bibásico foi também alterada visando minimizar a quantidade de íons adicionada na correção do pH. Os cultivos foram conduzidos em reatores de bancada (Bioengineering -NFL16), contendo 10 L úteis de meio. Os reatores foram controlados em condições otimizadas quanto à temperatura (30°C) e ao pH (7,1) através da adição de NaOH 5M para ajuste de pH, e quanto à concentração de oxigênio dissolvido (COD, *Chemical Oxygen Demand*), mantido em 30% através da variação da velocidade de agitação e com vazão de ar comprimido em 6,7 L/min.

O início da alimentação se dá no momento em que a glicose do meio original se esgota (batelada alimentada). Esse momento é avaliado através de um pico no valor de oxigênio dissolvido. A vazão de alimentação foi definida com base no acompanhamento da velocidade de consumo de glicose em $17 \text{ mL.L}^{-1}.\text{h}^{-1}$. O tempo total de cultivo foi de 21 horas.

3.3. Amostragem

Ao todo foram realizados quatro cultivos na mesma condição. Para cada um dos quatro cultivos realizados, foram coletadas amostras em triplicata a cada uma hora, do momento do inóculo até o final do cultivo, totalizando 22 amostras coletadas. Estas amostras foram utilizadas para duas destinações: (1) todas as 22 amostras foram utilizadas para medições (descritas abaixo) para caracterizar o cultivo e a produção de acetato e polissacarídeo, totalizando-se 264 amostras no total (22 tempos amostrados de quatro repetições independentes, três triplicatas cada); (2) seis destas amostras, escolhidas em tempos que representam momentos do cultivo nos quais ocorrem diferentes situações metabólicas (descritas abaixo), foram utilizadas para a extração de RNA e posterior análise da expressão gênica por RNA-Seq, totalizando 24 amostras (seis tempos amostrados de quatro repetições independentes). Os quatro cultivos realizados foram considerados como réplicas biológicas independentes.

Os pontos utilizados para a extração de RNA e análise por RNA-Seq foram: 1) início do cultivo (2 h após o início do cultivo, DO_{540} - densidade óptica em 540 nm - entre 0,5 e 1), representando início de adaptação ao meio e às condições do reator; 2) fase exponencial (8 h após o início do cultivo, DO_{540} entre 2 e 3), representando resposta ao meio e às condições do reator; 3) esgotamento da glicose do meio do reator - vale de dióxido de carbono (10 h após o início do cultivo, DO_{540} entre 8 e 9); 4) início da alimentação com glicose (11 h após o início do cultivo, DO_{540} entre 13 e 15), representando perturbação no sistema devido à alimentação com glicose; 5) fim da fase exponencial da alimentação com glicose/início da fase estacionária (15 h após o início do cultivo, DO_{540} entre 24 e 25), representando o momento no qual ocorre uma alteração ao padrão de resposta da população bacteriana e 6) final do cultivo (20 h após o início do cultivo, DO_{540} entre 25 e 27), representando o momento onde já ocorre inibição do crescimento celular.

Por praticidade, as seis amostras representativas de situações metabólicas distintas são identificadas S01 até S06 (*Sample*) e as quatro repetições dos cultivos B01 até B04 (*Batch*).

Para os experimentos de extração de RNA e análise de RNA-Seq, cada amostra coletada foi imediatamente centrifugada (11000 x g, por 10 minutos, 4 °C) para a separação do sobrenadante. O volume da amostra variou com o tempo, sendo sempre o volume necessário para termos em torno de 10^9 células, de maneira a garantir uma quantidade suficiente de massa de RNA extraído para poder ser realizado o RNA-Seq. Em seguida, o *pellet* bacteriano foi armazenado e congelado imediatamente em nitrogênio líquido, onde ficou até o uso. Para as análises de caracterização do cultivo e produção de acetato e de polissacarídeo, o tratamento das amostras variou de acordo com a análise realizada, e está descrito abaixo, para cada análise.

3.4 Análises

3.4.1 Determinação de Massa seca

A cada hora de cultivo, foram tomados 1,5 mL de cultura em triplicatas. Essas amostras foram centrifugadas a ~4000 x g por 7 minutos. O precipitado celular foi lavado com solução de PBS para fins de remover os resíduos de cultura. O tubo centrifugável contendo o *pellet* foi mantido em forno de secagem a 60°C por 24 horas e, sendo então realizada posterior pesagem da massa seca em balança analítica (Shimadzu, modelo AUY220, divisão 0,0001 g). O sobrenadante foi separado em outro tubo para posterior determinação de concentração de ácidos orgânicos, carboidratos e polissacarídeo, tendo ficado armazenado a -20 °C até o uso.

3.4.2 Determinação da concentração de acetato e glicose

As concentrações de acetato e de glicose foram medidas por cromatografia líquida de alta eficiência (*High Performance Liquid Chromatography* - HPLC), com a coluna de exclusão iônica Aminex HPX-87H (Bio-Rad). Como fase móvel, foi utilizada solução aquosa de H₂SO₄ 10 mM. O fluxo foi mantido a 0,6 mL.min⁻¹, em temperatura de 40 °C. Leituras foram realizadas em um detector de absorbância UV-VIS (Dionex UltiMate 3000 RS Diode Array Detector - Thermo Fisher Scientific) a 210 nm para o ácido e para o carboidrato, sendo os dados aquisitados pelo software Chromaleon (Thermo Fisher Scientific Inc.). Cada amostra foi diluída 10 vezes em H₂SO₄ 50 mM. As curvas padrões foram construídas com o padrão

comercial de glicose e acetato de sódio fornecido pelo fabricante da coluna (Bio-rad, números de catálogo 125-0585 e 125-0586 respectivamente).

3.4.3 Determinação da concentração de polissacarídeo capsular

A concentração da fração livre de PRP foi medida diretamente do sobrenadante da cultura por cromatografia de troca aniônica de alta eficiência com detecção por amperometria pulsada (HPAEC-PAD – *High Performance Anion Exchange Chromatography with Pulsed Amperometry Detection*). A um volume de 540 µL de amostra, contendo em trono de 10 mg/L de PRP, foi adicionado 180 µL de NaOH 400 mM. A mistura foi incubada a 37 °C por 20 h para a hidrólise alcalina do polímero de PRP, e neutralizada com 180 µL de ácido acético 400 mM. Foram adicionados ainda 100 µL de solução glicose-6-fosfato a 100 mM, utilizado como padrão interno. Um volume de 10 µL dessa mistura foi injetado na coluna de troca aniônica CarboPac PA-10 acoplada à pré-coluna de retenção de aminoácidos AminoTrap, montadas no sistema cromatográfico ICS5000 com o controle e aquisição de dados pelo software Chromleon (Thermo Fisher Scientific Inc.). A metodologia de análise inclui gradientes de NaOH e acetato de sódio descritos por Haan *et al.* (2013), além dos potenciais eletroquímicos definidos pelos mesmos autores para a leitura dos carboidratos no eletrodo de ouro. A curva de calibração foi construída utilizando-se o padrão internacional de PRP (MAWAS *et al.*, 2005; MAWAS *et al.*, 2007).

3.4.4 Determinação da massa molecular

As amostras foram previamente tratadas para minimizar interferência de meio de cultura e de componentes celulares por precipitação com brometo de hexadeciltrimetilamônio (CTAB) 0,5 M em pH 7,0 e então centrifugadas por 10 minutos a 16.000 x g (ANDERSON; SMITH, 1977). Um volume de 100 µL foi injetado em uma coluna de cromatografia de exclusão por tamanho de alta eficiência (HPSEC – *High Performance Size Exclusion Chromatography*), utilizando-se três colunas PL Aquagel- OH Mixed-H (Agilent Technologies) em série, a 30 °C e com fluxo de 1 mL.min⁻¹. Como fase móvel, foi usada uma solução de NaCl 100 mM, Na₂HPO₄ 13,22 mM, NaH₂PO₄ 6,78 mM e NaN₃ 3 mM. A corrida e a detecção foram realizadas no sistema cromatográfico Ultimate 3000 (Dionex), sendo feitas

leituras a uma frequência de 10 Hz em um detector de índice de refração e os dados foram adquiridos pelo software Chromeleon.

A determinação da massa molecular do PRP foi realizada por detecção de espalhamento de luz no equipamento MALS (*Multi Angle Light Scattering*) com detector de absorbância UV-VIS (Dionex UltiMate 3000 RS Diode Array Detector - Thermo Fisher Scientific) nos comprimentos de onda 260 nm, 280 nm e 405 nm e os dados foram processados pelo software Astra (WYATT, 1993).

3.4.5 Parâmetros cinéticos

Com os dados obtidos de concentração de glicose, concentração de acetato, concentração de PRP, biomassa e massa molecular foi possível computar parâmetros cinéticos associados ao cultivo em que são considerados biomassa seca (X), glicose como substrato (S), e tanto acetato quanto PRP como produtos (P), conforme **Tabela 1**. Ao se considerar a variação da concentração de X, S e P com o tempo, obtemos a velocidade instantânea destes parâmetros cinéticos. Devido ao fato de que a concentração microbiana X aumenta durante um cultivo descontínuo, aumentando consequentemente a concentração do complexo enzimático responsável pela transformação do substrato S no produto P, é mais informativo analisar os valores das velocidades instantâneas, com relação à referida concentração microbiana, ou seja, especificando-as com respeito ao valor de X em um dado instante (HISS, 2001). Uma definição especial de velocidade, cujo interesse prático está na avaliação do desempenho de um processo fermentativo, é a produtividade, que representa a velocidade média de crescimento referente ao tempo total ou final de fermentação (HISS, 2001). A mesma definição pode ser aplicada a concentração do produto, denominada produtividade do produto, em que t_{fP} (tempo desde o início do cultivo até o instante considerado) não é necessariamente igual a t_f (tempo final) (HISS, 2001). Assim, é possível obter a produtividade do produto para cada tempo após o início do cultivo. Todos os parâmetros apresentados na **Tabela 1** foram obtidos de Hiss (2001).

Tabela 1 – Parâmetros cinéticos associados ao cultivo de Hib a serem correlacionados com expressão gênica

Parâmetro	Relação
Concentração de biomassa	X
Concentração de consumo de substrato	S
Concentração de produto	P
Velocidade instantânea de crescimento do microrganismo	$r_X = \frac{dX}{dt}$
Velocidade instantânea de consumo de substrato	$r_S = -\frac{dS}{dt}$
Velocidade instantânea de formação de produto	$r_P = \frac{dP}{dt}$
Produtividade em biomassa	$Pr_X = \frac{X_m - X_0}{t_f}$
Produtividade do produto	$Pr_P = \frac{P_m - P_0}{t_{fP}}$
Velocidade específica de crescimento	$\mu_s = \frac{1}{X} \frac{dX}{dt}$
Velocidade específica de consumo de substrato	$\mu_S = \frac{1}{X} \left(-\frac{dS}{dt} \right)$
Velocidade específica de formação de produto	$\mu_P = \frac{1}{X} \frac{dP}{dt}$
Fator de conversão biomassa/substrato	$Y_{X/S} = -\frac{dX}{dS}$
Fator de conversão biomassa/produto	$Y_{X/P} = \frac{dX}{dP}$
Fator de conversão produto/substrato	$Y_{P/S} = -\frac{dP}{dS}$

3.5 Extração e isolamento do RNA mensageiro

Depois de descongeladas, as amostras foram submetidas à extração de RNA total utilizando o reagente TRIzol (Invitrogen), conforme orientações do fabricante. Resumidamente, os *pellets* celulares foram ressuspensos com o reagente TRIzol e as células foram lisadas por homogeneização com seringa, após o qual as amostras foram centrifugadas a 12.000 x g por 10 minutos a 4 °C em microcentrífuga, a fim de separar o excesso de polissacarídeo, conforme opção dada pelo fabricante. Após o sobrenadante ser transferido para outro tubo, clorofórmio foi adicionado e a mistura foi homogeneizada com pipeta, e então a centrifugada a 12.000 x g por 10 minutos. A fase aquosa foi coletada e misturada com 666 µL de isopropanol 100%, centrifugada a 12.000 x g por 15 minutos e o sobrenadante foi

descartado. O pellet foi então ressuspensão com 1,3 mL de etanol 75% e a mistura foi centrifugada a $7.500 \times g$ por 5 minutos. O sobrenadante foi descartado e o *pellet* ressuspensão em água livre de RNase.

O teste de integridade foi realizado usando o equipamento Agilent 2100 Bioanalyser (Agilent Technologies), utilizando a metodologia 2100 expert_Prokaryote Total RNA Nano. A concentração de RNA foi determinada usando-se o kit RiboGreen (Thermo Fisher), baseado em ligação do RNA a ligante fluorimétrico. O mRNA foi enriquecido do RNA total através da depleção do rRNA, feita com Ribo-Zero Magnetic Kit for Gram-negative Bacteria (Epicentre). Após esse procedimento, o mRNA foi novamente analisado com o Bioanalyser usando a metodologia 2100 expert_Prokaryote Total RNA Pico e com o kit RiboGreen. Todos os procedimentos foram realizados de acordo com as instruções dos fabricantes.

3.6 Construção de biblioteca de cDNA e sequenciamento

As amostras de mRNA purificadas obtidas dos *pellets* bacterianos foram utilizadas para construção de bibliotecas de cDNA com o kit TruSeq® Stranded mRNA LT (Illumina), que inclui fragmentação do RNA e adição de *primers*, síntese da primeira fita de cDNA com transcriptase reversa, remoção do RNA molde, síntese da segunda fita de cDNA, reparo das extremidades, adenilação das extremidades 3' para evitar formação de quimeras, ligação de adaptadores (índices/ *barcodes*) às extremidades dos cDNAs conforme Truseq Library Prep Pooling (24 adaptadores diferentes para 24 amostras), tal que posteriormente foi possível incluir as amostras aleatoriamente em ambos os *lanes* do *flow cell*.

Os fragmentos foram então enriquecidos via PCR, visando enriquecer os fragmentos que tiveram moléculas de adaptadores em ambas as extremidades e amplificar a quantidade de DNA na biblioteca. Após PCR, a qualidade das bibliotecas foi avaliada em um Agilent 2100 Bioanalyser (Agilent Technologies) usando *chip* específico para DNA Agilent DNA 1000. A concentração das bibliotecas foi estimada via qPCR com uso do KAPA Library Quantification Kit (Kapa Biosystems). As amostras foram diluídas a 2 nM, e 5 µL de cada amostra foram utilizados para formar o *pool* de amostras. A concentração do *pool* foi inferida por qPCR com o KAPA Library Quantification Kit. O *pool* foi diluído para 2 nM, quando foi adicionado o DNA de controle phiX, então as amostras foram desnaturadas e diluídas para 12 pM. Por fim, o processo de geração de *clusters* e o sequenciamento foram realizados

utilizando-se os kits TruSeq® Rapid PE Cluster Kit - 200 Cycle (leituras paired end de 100 pb) e TruSeq® Rapid SBS Kit – LS (Illumina), respectivamente, em uma plataforma Illumina HiSeq 1500, utilizando o método de corrida *Rapid*. Todas essas etapas foram realizadas de acordo com as instruções do fabricante.

3.7 Processamento dos *reads* gerados no sequenciamento

Após a plataforma Illumina gerar as imagens do sequenciamento, os dados foram analisados em três etapas: análise das imagens, identificação das bases e conversão das imagens em formato BCL (*binary base call*). Em seguida, foi utilizado o software *Consensus Assessment of Sequence And Variation* (CASAVA, versão 1.8.2), que realiza a descomplexação e separação das amostras do sequenciamento através da identificação dos índices que foram atribuídos a cada amostra na etapa de preparação das bibliotecas, e por fim converte dos arquivos BCL para arquivos no formato fastq, contendo as sequências de nucleotídeos e suas respectivas qualidades, descartando os *reads* não aprovados pelo controle de qualidade de bases, tal que o score Q deve ser maior que 30 (EWING *et al.*, 1998).

3.8 Pré-processamento por qualidade dos *reads*

Para o pré-processamento dos *reads* foi utilizado um *pipeline in house* com uso dos *softwares* fastq-mcf, versão 1.04.662 e bowtie2, versão 2.2.5, para análise dos *reads* brutos, com filtro dos *reads* por qualidade, eliminando aqueles com regiões de homopolímeros e de baixa complexidade, caudas poli A/T/N, remoção de adaptadores e extremidades de baixa qualidade (ANDREWS, 2010; LANGMEAD; SALZBERG, 2012). Os critérios empregados para a filtragem foram: remoção de regiões de homopolímeros e de baixa complexidade acima de 90% da sequência, tal que regiões com média de qualidade acima de 25 foram mantidas, e aquelas de baixa qualidade foram eliminadas. Após remoção das extremidades de baixa qualidade, foram mantidos os *reads* com tamanho mínimo de 40 pares de bases. Os parâmetros padrão do Bowtie2 são utilizados na análise para identificação e exclusão de *reads* com similaridade em relação ao banco de contaminantes como PhiX, utilizado como controle na preparação das amostras.

3.9 Alinhamento

O genoma de referência utilizado foi o de *Haemophilus influenzae* sorotipo b linhagem 10810 (GenBank *accession number* NC_016809.1) (BENSON *et al.*, 2013). O genoma de Hib 10810 possui 1.981.535 bp, foi reanotado recentemente (22/03/2017) com 2.012 genes, 6 cópias dos genes para rRNA 16S e 23S e 8 cópias dos genes para rRNA 5S previstos.

Todas as análises foram realizadas através da linguagem de programação estatística R, exceto quando descrito (TEAM, 2017). Assim, o alinhamento dos *reads* das amostras foi realizado com a função `align()` do pacote *Rsubread*, disponível no Bioconductor, para alinhar os *reads* ao genoma de referência, gerando arquivos *binary alignment* (BAM) (GENTLEMAN *et al.*, 2004; LIAO *et al.*, 2013). Os arquivos BAM foram usados para contagem com a função `featureCounts()` do pacote *Rsubread* (LIAO *et al.*, 2014).

A qualidade do alinhamento foi avaliada com o software Qualimap (OKONECHNIKOV *et al.*, 2015). As métricas a seguir foram calculadas para cada amostra com o Qualimap: a) cobertura de profundidade, definida como o quociente do rendimento (comprimento do *reads* multiplicado pelo número de *reads*) pelo comprimento do genoma; b) qualidade de mapeamento dos *reads* mapeados, que quantifica a probabilidade de um *read* estar mal alinhado, conforme **Equação 1** do Apêndice; e c) nível de saturação das amostras, que permite avaliar se mais sequenciamento é necessário e se mais genes poderiam ser detectados ao aumentar a quantidade de *reads* (LI *et al.*, 2008; OKONECHNIKOV *et al.*, 2015).

3.10 Tratamento de dados

3.10.1 Matriz de contagem de leituras

Dada uma lista $\{g_1, \dots, g_G\}$ de G genes, os dados serão representados como uma matriz $X = (x_{gj})$, em que cada linha representa um gene g e cada coluna j representa uma condição/réplica do experimento. O perfil de expressão x_{gj} representa o número de *reads* alinhados inequivocamente ao gene g ($g \in \{g_1, \dots, g_G\}$) da amostra j . Cada amostra j representa condição/réplica do experimento, isto é j pode ser representada como um par (s,b) em que $s \in \{1, \dots, 6\}$ momentos temporais metabolicamente críticos, e $b \in \{1,2,3,4\}$ é número do cultivo (ou réplica).

Cada gene g foi identificado pelo seu *Ordered Locus Name*, que pode ser HIB_XXXXX, em que XXXXX é um múltiplo de 10, para genes de anotação anterior, cujo formato é entrada no *Kyoto Encyclopedia of Genes and Genomes* (KEGG) ou então HIB_RSYYYYY, em que YYYYY é um múltiplo de 5, para genes de anotação mais recente, cujos genes não possuem entrada no KEGG (KANEHISA; GOTO, 2000). Exceções incluem genes não codificantes: HIB_tZZ se refere a RNA transportador, em que $ZZ \in N$, e *ttt* é um small RNA.

3.10.2 Remoção de efeito de lote

A remoção do efeito de lote se deu com a função `RUVg()` do pacote *RUVSeq* (*Remove Unwanted Variation from RNA-Seq Data*), que permite utilizar os genes menos significativamente diferencialmente expressos quando não se sabe *a priori* quais genes estão sem influência das covariáveis de interesse, baseado em uma análise de genes diferencialmente expressos feita antes da normalização RUVg (RISSO *et al.*, 2014). Os fatores estimados de variação não desejada foram inseridos no ajuste ao modelo linear generalizado (RISSO *et al.*, 2014).

3.10.3 Análise de rede de co-expressão dos genes

A identificação de módulos de co-expressão entre as diferentes condições foi realizada com o pacote WGCNA (*Weighted Gene Co-expression Network Analysis*) (LANGFELDER; HORVATH, 2008) disponível no *Comprehensive R Archive Network* (CRAN; HORNIK, 2012).

Antes de serem utilizados como entrada de dados no WGCNA, os dados foram previamente normalizados para remoção de viés de composição utilizando o pacote *edgeR* (função `calcNormFactors()`). A análise no WGCNA foi realizada de acordo com os métodos descritos abaixo.

Uma matriz do coeficiente de correlação de Pearson é calculada como medida de similaridade para todos os genes par a par (LANGFELDER; HORVATH, 2012). Usando a função `adjacency()` obtém-se uma matriz de correlação transformada: todos os valores de correlação são elevados à potência β (parâmetro denominado *soft thresholding*, visto que redes biológicas tendem a ter topologias *scale free*), escolhida conforme um critério de

topologia sem escala (*scale free*) (BARABASI; OLTVAI, 2004; ZHANG *et al.*, 2005). Para cada par de vértices (genes) foi então calculada a sobreposição topológica (*topological overlap*, TO), que reflete a similaridade entre genes em termos de vértices com os quais estes se conectam comumente (YIP; HORVATH, 2006). Assim, se obteve a matriz de sobreposição topológica (TOM) com a função `TOMsimilarity()`, que foi então convertida em uma matriz de dissimilaridade ($1 - \text{TOM}$), e essa foi base para uma árvore de clusterização hierárquica, obtida com a função `hclust()` e com a qual os módulos de expressão gênica foram identificados usando um procedimento dinâmico de corte de árvore com `cutreeDynamic()` (LANGFELDER *et al.*, 2008).

Um mínimo de 40 genes em cada módulo foi configurado. Para cada módulo encontrado, a função `moduleEigengenes()` foi utilizada para calcular a primeira componente principal da expressão gênica dos genes pertencentes a um dado módulo correspondente, denominado autogene (*eigengene*) (ZHANG; HORVATH, 2005). O autogene representa a atividade geral do módulo em uma dada amostra e foi utilizado para mapear grupos de genes a dados fenotípicos relativos às amostras (*data traits*), quando disponíveis, através do coeficiente de correlação de Pearson. Assim, foram calculadas as correlações entre cada autogene e as variáveis associadas ao cultivo definidas na **Tabela 1**.

Duas medidas de importância de gene foram definidas: 1) a correlação entre a expressão de um dado gene e a variável produtividade de polissacarídeo, Pr_{PP} e 2) a correlação entre a expressão de um dado gene e a variável produtividade de acetato, Pr_{Ac} , conforme descrito na **Tabela 1**.

A função `intramodularConnectivity()` foi então utilizada para calcular conectividade total (k_{Total} , grau de cada gene em relação aos demais genes), conectividade intra-modular (k_{Within} , grau de cada gene em relação aos demais genes pertencentes ao mesmo módulo), conectividade extra-modular (k_{Out} , grau de cada gene em relação aos demais genes pertencentes a outros módulos), e a diferença entre as conectividades intra- e extra-modulares para todos os genes (k_{Diff}) (ZHANG; HORVATH, 2005).

O teste hipergeométrico para medir enriquecimento significativo dos genes co-expressos em módulos em relação aos conjuntos de genes de Hib foi realizado com a função `enrichKEGG()` do pacote *clusterProfiler* do Bioconductor, com $\alpha = 0,05$ (YU *et al.*, 2012).

3.10.4 Análise de expressão diferencial

A análise de expressão diferencial foi realizada através da linguagem de programação estatística R com os pacotes *edgeR*, *DESeq2* e *limma-VOOM* do Bioconductor, recomendados para desenhos experimentais de RNA-seq com o número de amostras por condição tal qual o deste trabalho, pacotes estes através dos quais é possível identificar mudanças relativas nos níveis de expressão entre condições distintas (GENTLEMAN *et al.*, 2004; ROBINSON *et al.*, 2010; SONESON; DELORENZI, 2013; CHEN *et al.*, 2014a; LAW *et al.*, 2014; LOVE *et al.*, 2014; ZHANG *et al.*, 2014; RITCHIE *et al.*, 2015; SCHURCH *et al.*, 2016). Cada um desses pacotes é brevemente descrito abaixo.

edgeR. Na análise de expressão diferencial com o *edgeR*, os dados de contagem de leituras devem ser utilizados sem transformação prévia (não devem ser usados RPKM ou FPKM, pois impedem o *edgeR* de estimar corretamente a relação média-variância dos dados), a normalização é baseada em modelo e ocorre na forma de fatores de correção que são inseridos no modelo estatístico (CHEN *et al.*, 2014b).

Com o pacote *edgeR*, a filtragem dos dados foi feita com referência ao número de transcritos por milhão (TPM) de cada gene (WAGNER *et al.*, 2012). Os valores brutos de contagem foram convertidos em TPM com a função `cpm()` do *edgeR*. Um gene foi considerado expresso em uma dada amostra caso sua contagem transformada estivesse acima de um valor de TPM nominal igual a 1, e como não expresso caso contrário. Cada gene deve ser considerado expresso em ao menos todas as repetições de uma mesma condição para ser mantido em análises subsequentes.

A normalização do viés de composição do RNA, denominado *trimmed mean of M-values (TMM) normalization*, é computada com a função `calcNormFactors()` (ROBINSON; OSHLACK, 2010). Esse viés pode ocorrer quando existem genes altamente expressos e que consomem uma grande porção do tamanho total da biblioteca, o que pode acusar falsamente os demais genes de serem regulados negativamente na amostra (ROBINSON; OSHLACK, 2010). A função `calcNormFactors()` encontra um conjunto de fatores de escalonamento para o tamanho das bibliotecas que minimiza as mudanças em escala logarítmica entre as amostras para a maioria dos genes (ROBINSON; OSHLACK, 2010). O produto do tamanho da biblioteca original com o fator de escalonamento é denominado tamanho efetivo das bibliotecas, e substitui o tamanho da biblioteca em todas as análises subsequentes (CHEN *et al.*, 2014b).

edgeR então estima dispersões (dispersão comum, dispersão de tendência e dispersão de cada gene) com a função `estimateDisp()` usando *Cox-Reid profile-adjusted likelihood* para desenhos experimentais mais complexos (COX; REID, 1987; ROBINSON; SMYTH, 2007; ROBINSON; SMYTH, 2008; MCCARTHY *et al.*, 2012).

Uma matriz com o desenho experimental deve então ser construída, contendo informação a respeito das diferentes condições que se quer comparar, em que uma das condições pode ser considerada como referência ou então as condições devem ser parametrizadas por grupos (ROBINSON *et al.*, 2010). Os contrastes de interesse (A e B, por exemplo) podem ser especificados com a função `makeContrasts()` do pacote *limma*, que define a hipótese nula ($A-B = 0$), e permite escolher como as diferentes condições serão comparadas par a par (SMYTH, 2005).

O pacote *edgeR* permite utilizar estrutura de modelos lineares generalizados (GLM) para contabilizar desenhos experimentais complexos (MCCULLAGH; NELDER, 1989; MCCARTHY *et al.*, 2012). As funções `glmFit()` e `glmLRT()` foram usadas respectivamente para ajustar o modelo binomial negativo generalizado à contagem de leituras para cada gene e para conduzir um teste da razão de verossimilhanças para os coeficientes (grupos de interesse) no modelo linear (MCCARTHY *et al.*, 2012). A remoção do efeito de lote se deu com a função `RUVg()` do pacote *RUVSeq*, cujos fatores estimados de variação não desejada foram inseridos junto à função `glmFit()` (RISSO *et al.*, 2014).

Os genes mais significativos podem ser examinados com a função `topTags()`, que apresenta os genes com maior razão logarítmica entre as condições (por exemplo, log-fold positivo representa genes com maior expressão em A em relação a B) (CHEN *et al.*, 2014b). Essa função também faz correção de testes múltiplos ao aplicar o procedimento de Benjamini–Hochberg para controlar a taxa de descobertas falsas (*false discovery rate*, FDR) (BENJAMINI; HOCHBERG, 1995).

DESeq2. Com o pacote *DESeq2*, também devem ser usados dados de contagem de *reads* sem transformação prévia. A função `DESeqDataSetFromMatrix()` foi utilizada para construir o objeto `DESeqDataSet`, necessário para as etapas subsequentes, e para o qual foram fornecidos a matriz de contagem, informação sobre as amostras (as colunas da matriz de contagem) e o desenho experimental (LOVE *et al.*, 2014). Os dados foram então pré-filtrados, tal que genes com nenhum ou apenas 1 *reads* foram removidos.

Esses dados de expressão K_{ijl} (em que i é gene e j é tempo amostral e l é replica) são modelados seguindo uma distribuição binomial negativa com média μ_{ij} e parâmetro de dispersão gene-específico α_i ($K_{ijl} \sim \text{NB}(\mu_{ij}, \alpha_i)$), que define a relação entre a variância e a contagem observada e sua média.

A média ajustada é composta pelo fator de tamanho s_i e um parâmetro q_{ij} , proporcional à concentração real esperada de *reads* para a amostra j , tal que

$$\mu_{ij} = s_i q_{ij} \quad (1)$$

Sejam β_i coeficientes que representam o $\log_2$ *fold change* para o gene i para cada coluna da matriz conforme seguinte equação

$$\log(q_{ij}) = x_i \beta_i \quad (2)$$

A função `DESeq()` permite calcular essas variáveis ao executar as seguintes etapas:

- a) Com a função `estimateSizeFactors()`, os fatores de tamanho s_i são estimados ao se tomar a média geométrica das contagens sobre todas as amostras, sendo criada uma amostra referência virtual para cada gene, e então normalizar cada amostra em relação a esta referência para obter um fator de tamanho por amostra (ANDERS; HUBER, 2010).
- b) A variabilidade entre réplicas é modelada pelo parâmetro de dispersão α_i , estimado com a função `estimateDispersions()`, ao maximizar o *Cox-Reid profile-adjusted likelihood* (COX; REID, 1987; LOVE *et al.*, 2014).
- c) Por fim, com a função `nbinomWaldTest()`, *DESeq2* emprega um modelo generalizado binomial negativo ao ajustar os coeficientes β_i ($\log_2$ *fold change*), com um parâmetro de dispersão para cada gene e o teste de Wald para gerar p valores (LOVE *et al.*, 2014).

***limma-voom*.** Com *limma-voom* pode-se normalizar o tamanho das bibliotecas de logaritmo de contagem por milhão, ou então por normalização de quantil ou pelo TMM (*trimmed mean of M-values normalization*) como no *edgeR* (LAW *et al.*, 2014). Usamos TMM.

Com o pacote *limma*, a filtragem dos dados, a normalização do viés de composição do RNA e os contrastes de interesse são feitos tal qual com o pacote *edgeR*. Contudo, no pacote *limma*, a modelagem linear é realizada no logaritmo do número de contagens por milhão, $\log\text{-cpm}$, que são considerados normalmente distribuídos (LAW *et al.*, 2014).

A função `voom()` (*variance modeling at the observational level*, voom) estima a relação média-variância do logaritmo da contagem de *reads* ao ajustar uma regressão ponderada local (*locally weighted regression*, LOWESS), que fornece um desvio padrão residual para cada gene (LAW *et al.*, 2014). O logaritmo das contagens por milhão é convertido em uma contagem prevista, e a variância se baseia nesse tamanho de contagem. O peso para cada observação é o inverso da variância prevista (LAW *et al.*, 2014).

Então, a função `lmFit()` é usada para estimar *fold changes* e erros padrão ao ajustar um modelo linear para cada gene (PHIPSON *et al.*, 2016). Dado esse ajuste, coeficientes e erros padrão são computados para um conjunto de comparações com a função `contrasts.fit()` (PHIPSON *et al.*, 2016). Um *framework* empírico de Bayes é implementado para obter informações entre os genes ao estimar as variâncias através da função `eBayes()` (PHIPSON *et al.*, 2016).

Os genes pertencentes à interseção entre os genes identificados como diferencialmente expressos com *log-fold change* (logFC) a partir de 2 e *p*-valor ajustado no máximo 0,01 pelos pacotes *edgeR*, *DESeq2* e *limma* foram considerados genes de fato diferencialmente expressos.

O teste hipergeométrico para medir enriquecimento significativo dos genes diferencialmente expressos durante o cultivo de Hib foi realizado com a função `enrichKEGG()` do pacote *clusterProfiler* do Bioconductor, com $\alpha = 0,05$ (YU *et al.*, 2012).

4. Resultados e Discussão

4.1 Cultivos

Foram realizados quatro cultivos, B01, B02, B03 e B04, cujos perfis médios de crescimento e das demais variáveis mensuradas associadas aos cultivos são apresentados nas **Figuras 2 e 3**.

Pela **Figura 2**, podemos observar pela concentração de biomassa que a fase de adaptação (lag) ocorre nas primeiras 2 horas de cultivo, então se inicia a fase exponencial (log), caracterizada por intensa duplicação celular, das 3 horas até em torno das 17 horas, momento em que o crescimento bacteriano atinge seu máximo, mesmo com fornecimento e consumo sempre existente de glicose. Das 18 horas de cultivo até o final – às 21 horas –

temos uma fase de crescimento reduzido. Durante a fase lag e fase log até 8h de cultivo, a concentração de CO₂ aumenta até atingir um máximo local. A queda na concentração de CO₂ dura até 10h de cultivo (pouco antes do início da alimentação), após a qual retoma o aumento até 13h. Deste ponto até o final do cultivo, a concentração de CO₂ segue estacionária, não acompanhando o crescimento celular. O consumo de glicose diminui após 19 horas de cultivo, mesmo com a alimentação constante de glicose sendo adicionada ao meio. A concentração de PRP e de acetato aumenta em taxa similar ao crescimento bacteriano. Já na **Figura 3** pode ser observado que a produtividade de PRP aumentou de ~0,1 g.L⁻¹.h⁻¹ no início do cultivo para acima de 0,4 g.L⁻¹.h⁻¹ enquanto a produtividade de acetato aumentou de quase nula no início do cultivo para em torno no início do cultivo atingindo em torno de 0,06 g.L⁻¹.h⁻¹ ao final do cultivo.

Os dados de cada dado fenotípico mensurado durante o cultivo e descrito na **Tabela 1** são apresentados mais detalhadamente nas **Figuras 1 a 23** do Apêndice.

Figura 2. Perfil cinético de biomassa, glicose consumida, formação de acetato, síntese de PRP e produção de CO₂. Média dos valores de cada amostra provenientes dos quatro cultivos (B01, B02, B03 e B04). Em verde, concentração de glicose no meio. Em roxo, concentração de acetato. Em preto, concentração de biomassa. Em azul, concentração de PRP. Em vermelho, concentração de gás carbônico. Em dourado, concentração de oxigênio. No eixo x, o tempo de cultivo em horas. Em destaque (retângulos), as seis amostras selecionadas: a) início de adaptação ao meio do reator, b) resposta ao meio do reator, c) quase

esgotamento de glicose do meio do reator, d) perturbação no sistema devido à alimentação, e) inflexão – alteração ao padrão de resposta e f) inibição do crescimento.

Figura 3. Perfil cinético de produtividade de PRP e produtividade de acetato. Média dos valores de cada amostra provenientes dos quatro cultivos (B01, B02, B03 e B04). Em dourado, produtividade de polissacarídeo. Em ciano, produtividade de acetato. Em destaque, as seis amostras selecionadas: a) início de adaptação ao meio do reator, b) resposta ao meio do reator, c) quase esgotamento de glicose do meio do reator, d) perturbação no sistema devido à alimentação, e) inflexão – alteração ao padrão de resposta e f) inibição do crescimento

4.2 Teste de integridade

Após a etapa de extração de RNA total, as amostras foram submetidas a um teste de integridade, utilizando eletroforese capilar, realizada em um equipamento Agilent 2100 Bioanalyzer. Um resultado representativo do padrão que obtivemos é apresentado na **Figura 4**.

Figura 4. Resultados do teste de integridade. Resultados obtidos a partir das seis amostras (S01, S02, S03, S04, S05, S06) do primeiro dos quatro cultivos, denominado B01. No eixo x, a concentração da unidade de fluorescência. No eixo y, a quantidade de nucleotídeos das moléculas de RNA.

Espera-se, para uma amostra com elevada integridade, que sejam observados dois picos evidentes e destacados, referentes ao RNAs ribossomais 16S e 23S. O RNA ribossomal de Hib é naturalmente fragmentado, devido a sítios de clivagem dentro da subunidade ribossomal 23S, tal que no Agilent 2100 Bioanalyser o pico do RNA ribossomal 23S apresenta essa fragmentação em forma de pico duplo (SONG *et al.*, 1999; EVGUENIEVA-HACKENBERG, 2005). Tal ocorrência inviabiliza a utilização do número de integridade de RNA (RNA Integrity Number - RIN) como único parâmetro para determinar integridade (BHAGWAT *et al.*, 2013). Caso haja degradação, isso é detectado por um padrão de gel de RNA com linha de base maior e pico de RNA ribossomal menor e menos distinto (SCHROEDER *et al.*, 2006). Assim, podemos notar que as amostras se apresentaram relativamente íntegras. Podemos observar também que o RNA ribossomal 16S desta cepa de Hib (GB 3291) também é naturalmente fragmentado.

4.3 Sequenciamento e Alinhamento

Um total de 381.157.273 de *reads* foram mapeados ao genoma de referência, que possui 2012 genes descritos, com cobertura de profundidade média de $788,18x \pm 501,50x$ (mínimo de $99,32x$ e máximo de $2042,90x$) entre as 24 amostras. Após contagem com `featurecounts()`, configurado para remover *reads* sem par e que conta cada par como um *read*, além de remoção de genes para RNA ribossomal, obtivemos uma matriz de contagem com 2000 genes (excluindo-se os 12 genes ribossomais) e 170.179.762 *reads*. A quantidade de reads por amostra pode ser conferida na **Tabela 1** do Apêndice.

A qualidade de mapeamento médio foi de $36,62 \pm 0,13$, em uma escala de escore de qualidade Phred que varia entre 0 e 40. Nesta escala, por exemplo, em escore de qualidade $\Phi_{\Phi}=20$ com apenas um erro de mapeamento (*mismatch*), aproxima-se a probabilidade de amostragem do *read* em $\sim 0,01$ (ou seja, a probabilidade de mapear corretamente algum *read* aleatório é 0,99) (LI *et al.*, 2008). Caso a probabilidade de mapear corretamente aumente para 0,999, o escore Φ_{Φ} aumenta para 30. Por outro lado, conforme a probabilidade de mapeamento correto tenda a zero, o escore Φ_{Φ} também tenderá a zero. Assim, podemos notar que a qualidade de mapeamento foi relativamente alta. A qualidade de mapeamento de cada amostra pode ser conferida na **Tabela 1** do Apêndice.

A saturação média foi de 90,5% (mínima 87,3% e máxima 91,8%.), conforme **Figura 5**. Em tecnologia de RNA-seq, a saturação seria alcançada quando aumento no número de *reads* não resulta em mais genes de fato expressos sendo detectados ou em mais genes sendo identificados como diferencialmente expressos quando duas ou mais condições são comparadas (TARAZONA *et al.*, 2011). Pela **Figura 5**, podemos observar que a todas as amostras atingiram saturação, inclusive amostras de baixa cobertura (S01B03 e S03B02). Por exemplo, ao observarmos a diferença entre 10 milhões de reads e 5 milhões de reads, apenas ~ 25 genes a mais são detectados. Ou seja, nesse ponto, o dobro de sequenciamento rende menos de 2% a mais de informação nova. Destaca-se que mesmo nas amostras com menor cobertura a detecção é maior ou igual a 87,3%, indicando que, no máximo, 12,7% dos genes ficaram de fora da análise, embora, claro, não se possa dizer qual a porcentagem total de genes da bactéria foram expressos durante o cultivo.

Figura 5 - Curvas da saturação da detecção de genes em função do número de fragmentos, para cada amostra. No eixo y, número de genes detectados de um total de 2000 genes (removidos os genes para RNA ribossomal). No eixo x, cobertura de sequenciamento em milhões de *reads*. O círculo preenchido representa o ponto de saturação, ou seja, aumento do número de *reads* nesta amostra a partir deste ponto não aumenta a quantidade de genes detectados, enquanto os círculos não preenchidos antes dos círculos preenchidos indicam posições nas quais ainda não houve saturação.

4.4 Remoção de efeito de lote

Em um conjunto de dados sem efeito de lote, as distribuições deveriam hipoteticamente estar centralizadas em torno de zero, enquanto que desvios claros indicam a necessidade de remoção de variáveis indesejadas. Assim, a remoção de efeito de lote se mostrou imprescindível, conforme apresentado na **Figura 6**.

Figura 6 - Boxplots da razão logarítmica da contagem de reads de cada gene de cada amostra em relação à mediana entre as amostras. À esquerda, *boxplots* de cada amostra dos dados de contagem. À direita, *boxplots* das amostras após remoção de variáveis indesejadas. Cada ponto amostral é representado por uma cor.

4.5 Análise de co-expressão dos genes

4.5.1 Análise de rede de co-expressão dos genes

Após filtro para remoção de genes com baixa expressão e normalização para viés de composição, 1862 genes (dos 2012 descritos neste genoma) foram submetidos à análise de co-expressão gênica. O parâmetro β (*soft thresholding*) escolhido foi 7, que foi a menor potência que resulta em valor acima de 0,80 da curva de índice de ajuste à topologia sem escala, conforme gráfico da esquerda na **Figura 7**. Caso o parâmetro β seja maior, a conectividade média (em torno de 40 para $\beta = 7$) irá diminuir, como podemos verificar no gráfico da direita na **Figura 7**. O ajuste à topologia sem escala está representado na **Figura 8**.

A análise WGCNA identificou 8 módulos de genes co-expressos (denominados *Brown, Black, Purple, Blue, Red, Magenta, Green, Pink*), que podem ser conferidos em “Merged Dynamic” na **Figura 9**, e consistiram de mínimo de 57 a máximo de 828 genes, enquanto 7 genes (0,38%) não foram incluídos em quaisquer dos módulos (por padrão do WGCNA, esses genes são agrupados no módulo *grey*). A **Figura 10** apresenta o *heatmap* de

toda rede ponderada, representando graficamente a matriz de sobreposição topológica, permitindo identificar dois grandes clusters de genes e outros seis clusters menores, concordando com os módulos identificados. A **Figura 11** mostra a expressão do autogene de cada cluster, indicando que o comportamento em geral é bem distinto, com o autogene dos módulos *Brown* e *Black* apresentando maior expressão nas amostras iniciais do cultivo, os módulos *Magenta*, *Green* e *Pink* apresentando expressão maior nas amostras intermediárias do cultivo e o autogene dos módulos *Blue* e *Red* com maior expressão nas amostras finais do cultivo. A relação entre os autogenes pode ser observada na **Figura 24** do Apêndice. O *heatmap*, *barplot* e *spaghetti plot* de cada módulo pode ser conferido entre as **Figuras 25 até 33** do Apêndice.

Cada módulo foi representado pela sua primeira componente principal denominada autogene (*eigengene*). Correlações altas entre os autogenes e dados fenotípicos indicam possíveis associações entre os genes co-expressos no módulo em questão e os fenômenos observados. A correlação entre a expressão do autogene do módulo e cada variável mensurada relacionada ao cultivo (identificados na **Tabela 1**) pode ser conferida na **Figura 12**.

Dos 8 módulos, 4 (*Brown*, *Purple*, *Blue* e *Red*) foram significativamente (coeficiente de correlação $> 0,5$, $\alpha = 0,05$) associados com a concentração de biomassa, concentração de acetato, concentração de PRP, concentração de CO₂, velocidade instantânea de formação de acetato, velocidade instantânea de crescimento referente ao tempo, produtividade de PRP, produtividade de acetato. Essas correlações significativas foram positivas para os módulos *Blue* e *Red*, e negativas para os módulos *Brown* e *Black*. A velocidade instantânea de formação de PRP foi significativamente associada aos módulos *Red* (positivamente), *Purple* e *Brown* (ambos negativamente), enquanto padrão oposto a esse foi observado em relação ao parâmetro velocidade específica de crescimento. Os módulos *Magenta* e *Green* apresentaram correlação positiva com a concentração do consumo de glicose. O módulo *Purple* apresentou correlação positiva com a velocidade específica de consumo de substrato e com o coeficiente específico de Biomassa-Glicose. O módulo *Blue* apresentou correlação negativa com a velocidade específica de formação de PRP. O módulo *Magenta* apresentou correlação negativa com o coeficiente específico de Biomassa-PRP. Os módulos *Black* e *Purple* apresentaram correlação positiva com os coeficientes específicos de Acetato-Glicose e PRP-Glicose.

Maior sentido biológico pode ser encontrado ao analisar o enriquecimento funcional de cada módulo, que indicou diferentes vias metabólicas bioquímicas associadas aos módulos,

como pode ser visto na **Figura 13**: genes para proteínas do ribossomo no módulo *Brown*, genes para biossíntese de peptidoglicano no módulo *Purple*, genes para transportadores ABC (*ATP Binding Cassete*) no módulo *Red*, enriquecimento para vias associadas ao metabolismo de carbono e glicólise/gliconeogênese no módulo *Magenta* e para biossíntese de antibióticos, biossíntese de metabólitos secundários, ciclo do citrato (TCA), metabolismo de purina e vias metabólicas no módulo *Green*.

Os genes específicos para síntese de PRP (*bcs1234*: respectivamente, HIB_11840, HIB_11830, HIB_11820 e HIB_11810), além do gene para a enzima polissacarídeo polimerase (HIB_10030) foram agrupados no módulo *Blue*, apesar de o teste hipergeométrico não ter indicado enriquecimento funcional neste módulo.

Uma questão interessante a respeito desses módulos é se há alguma relação entre a conectividade e a importância dos genes em relação aos dados fenotípicos (no caso, produtividade de acetato e de PRP) nestes módulos, ou seja, se um gene importante para um dado fenotípico tende a ser *hub*.

De fato, na **Figura 14** podemos observar que para os módulos *Brown*, *Purple*, *Blue* e *Red* os genes *hub* intra-modulares tendem a possuir alta importância de gene, sugerindo que uma busca por genes *hub* desses módulos pode indicar genes chave associados à produtividade de PRP. O mesmo padrão ocorre em relação à importância de gene para produtividade de acetato, em que os módulos *Brown*, *Purple* e *Red* se destacam por apresentarem correlação ainda maior com a conectividade intra-modular, conforme apresentado na **Figura 15**. Assim, torna-se interessante a análise dos principais hubs de cada módulo.

Figura 7. Análise de topologia de rede para potências de *soft-thresholding* entre 2 e 30. O gráfico à esquerda mostra o índice de ajuste à topologia sem escala (eixo y) como função da potência *soft-thresholding* (eixo x). O gráfico à direita mostra a conectividade média (grau do vértice, eixo y) como função da potência *soft-thresholding* (eixo x).

Figura 8. Ajuste à topologia sem escala, com parâmetro $\beta = 7$: Coeficiente de determinação (R^2) de topologia sem escala = 0,83; inclinação = -1,81; coeficiente de determinação (R^2) truncado = 0,91. Este é um gráfico log-log de distribuição de conectividade por toda a rede (histograma das conectividades dos genes) e representa o ajuste a modelo linear e a um modelo exponencial truncado. O eixo x mostra o logaritmo da conectividade de toda a rede, enquanto que o eixo y apresenta o logaritmo da distribuição de frequência correspondente. O coeficiente de determinação (R^2) do ajuste pode ser considerado um índice da liberdade de escala da topologia da rede.

Figura 9. Árvore de clusterização hierárquica. Na árvore (dendrograma) de clusterização, cada folha, que é uma linha vertical curta, corresponde a um gene. Módulos de genes são identificados como ramos do dendrograma, que agrupam genes densamente interconectados e altamente co-expressos. As cores abaixo do dendrograma indicam em “*Dynamic Tree Cut*” a associação a módulos identificados com a função `cutreeDynamic()` e em “*Merged dynamic*” após fusão de módulos com correlação maior que 0,75 entre si.

Figura 10. Visualização da rede gênica com *heatmap*. O *heatmap* representa a matriz de sobreposição topológica (Topological Overlap Matrix, TOM) entre todos os genes na análise. Cor clara representa baixa sobreposição e cores alaranjadas progressivamente mais escuras até vermelho representam alta sobreposição. Blocos de cores mais escuras ao longo da diagonal são os módulos. O dendograma de genes e a atribuição do módulo também são mostrados à esquerda e no topo.

Figura 11. Barplot do autogene de cada módulo. Cada *barplot* está colorido conforme o nome do módulo. O autogene do módulo apresenta valores negativos quando muitos dos genes nos módulos estão sub-expressos, enquanto que apresenta valores positivos quando muitos dos genes do módulo estão sobre-expressos. O autogene do módulo pode ser considerado o perfil de expressão mais representativo do módulo.

Figura 12. Associações módulo-dados associados aos cultivos. Cada linha corresponde a um autogene de módulo e cada coluna corresponde a uma variável associada aos cultivos. Cada célula contém uma correlação correspondente e p -valor. A tabela é colorida de acordo com a correlação: negativa é representada por cor verde e positiva por cor vermelha. Quanto mais forte a correlação, mais forte o tom da cor. Em destaque (contorno da célula em negrito), as correlações significativas ($\alpha = 0,05$).

Figura 13 - Comparação de enriquecimento funcional nos módulos de genes. Neste gráfico de pontos, no eixo y estão as vias metabólicas sobre-representadas, no eixo x os módulos que apresentaram vias enriquecidas e em parênteses a quantidade de genes presentes nas vias enriquecidas. Cada circunferência representa a razão de genes encontrada em relação ao total de genes na via metabólica. A cor se refere ao *p*-valor, em que quanto mais próximo de azul, mais próximo do limite de *threshold* ($\alpha = 0,05$).

Figura 14 - Importância do gene para produtividade de PRP (eixo y) vs. conectividade intra-modular (eixo x) plotado separadamente para cada módulo encontrado. Para os módulos *Brown*, *Purple*, *Blue* e *Red* podemos observar que os genes *hubs* intra-modulares tendem a ter alta importância de gene para produtividade de PRP.

Figura 15 - Importância do gene para produtividade de acetato (eixo y) vs. conectividade intra-modular (eixo x) plotado separadamente para cada módulo encontrado. Para os módulos *Brown*, *Purple*, *Blue* e *Red* podemos observar que os genes *hubs* intra-modulares tendem a ter alta importância de gene para produtividade de acetato.

4.5.2 Análise dos principais *hubs* de cada módulo

A seguir, a análise dos dez principais *hubs* de cada módulo, apresentados nas **Tabelas 2 até 9**, uma para cada módulo.

No módulo *Brown* (**Tabela 2**), dois dos principais *hubs* foram os genes HIB_07040 e HIB_07050, respectivamente para os fatores de alongação Tu e G, que participam no processo de alongamento de peptídeos e translocação (além de desmontagem do ribossomo) na síntese de proteínas bacterianas (RAMAKRISHNAN, 2002; SAVELSBERGH *et al.*, 2009). Os genes HIB_09230, HIB_09260, HIB_09270 e HIB_09280 pertencem à via metabólica do ribossomo (via hiu03010 do KEGG), em particular do canal SecY (*secretion pathway*) (COREY *et al.*, 2016). O gene HIB_05840 codifica para TatD, uma proteína com atividade DNase na presença de Mg^{2+} (WEXLER *et al.*, 2000; NATALE *et al.*, 2008). O gene HIB_12290 codifica para a enzima HrpA, envolvida na separação ATP-dependente do RNA de fita dupla, sendo que em *Escherichia coli* está envolvida no processamento pós-transcricional de genes para proteínas da biogênese fimbrial (KOO *et al.*, 2004; KHEMICI; LINDER, 2016). O gene

HIB_14190 para DNA girase (também topoisomerase) sub-unidade A, que introduz superenrolamentos (*supercoils*) negativos no DNA usando energia derivada da hidrólise de ATP (REECE; MAXWELL, 1991). O gene HIB_05850 codifica para a DNA polimerase III subunidade delta, enzima responsável pela maioria da síntese replicativa em bactérias (STUKENBERG *et al.*, 1991; TURNER *et al.*, 1999).

Considerados juntos, os dez principais *hubs* do módulo *Brown* têm funções associadas à transcrição e tradução dos demais genes, o que indica serem necessários para a manutenção de funções celulares básicas. Neste módulo, a via de ribossomo está sobre-representada (**Figura 13**). O padrão geral de expressão representado pelo autogene do módulo *Brown* (**Figura 11**) sugere alta expressão no começo do cultivo, o que é esperado para esses genes, devido à adaptação ao meio e às condições do cultivo. Na adaptação ao meio existe a necessidade de expressão de muitas enzimas e de transportadores. Além disso, a bactéria deve estar anabolizando material básico para crescimento, inclusive muitas proteínas. Nas demais amostras, a expressão desses genes é menor.

No módulo *Black* (**Tabela 3**), os principais genes *hubs* foram HIB_04660, para a proteína Lic3A (sialiltransferase) e HIB_RS08725, para a proteína LicA (fosforilcolina quinase), ambas envolvidas na transferência/incorporação de componentes (respectivamente, ácido siálico e fosforilcolina) no lipooligossacarídeo (LOS) de Hib e estão sujeitas à variação de fase, resultando em perda ou ganho espontâneo desses na composição do LOS (KIMURA; HANSEN, 1986; HOSKING *et al.*, 1999; HOOD, *et al.*, 2001; HUMPHRIES; HIGH, 2002).

O gene HIB_09640 codifica para transglicolase associada à biossíntese de peptidoglicano, enquanto HIB_09620 para transglicolase lítica solúvel, envolvida na clivagem de peptidoglicanos (DI BERARDINO *et al.*, 1996; KORAIMANN, 2003; SCHEURWATER *et al.*, 2008).

O gene HIB_09870 para enzima putativa associada à utilização de ferro da hemina, com 47% de identidade de aminoácidos em relação à hemina oxigenase de *Helicobacter pylori* (HU *et al.*, 2011). O gene HIB_18580 codifica para proteína com domínio conservado da família DUF2057, sem função conhecida (MARCHLER-BAUE *et al.*, 2016). O gene HIB_16980, para a proteína LicD, que age como transferase que altera a posição da fosforilcolina no LOS (LYSENKO *et al.*, 2000).

O gene HIB_04480 codifica para GTP pirofosfoquinase (RelA), que catalisa a produção de guanossina 3'-difosfato 5'-trifosfato ((p)ppGpp) durante resposta estrigente e está

associada ao metabolismo de purina (via hiu00230 do KEGG) (EVERS *et al.*, 2001; ATKINSON *et al.*, 2011; ENGLISH *et al.*, 2011).

Tabela 2. Relação dos 10 genes com maior conectividade intra-modular (*hubs*) do módulo *Brown*. Para cada gene *hub*, a proteína para a qual codifica, a conectividade intra-modular Σ_{intra} e a conectividade extra-modular Σ_{extra} .

Módulo	gene	Produto	Σ_{intra}	Σ_{extra}
<i>Brown</i>	HIB_07050	fator de alongamento G	142,81	30,08
	HIB_05840	TatD (DNase)	139,73	22,34
	HIB_07040	fator de alongamento Tu	136,50	24,32
	HIB_12290	RNA helicase HrpA ATP-dependente	135,55	33,88
	HIB_14190	DNA girase subunidade A	135,10	39,64
	HIB_09270	proteína 30S ribossomal protein S5	133,62	24,13
	HIB_09260	proteína 50S ribossomal L18	130,20	19,83
	HIB_09280	proteína 50S ribossomal L30	129,35	20,93
	HIB_09230	proteína 30S ribossomal S14	128,80	38,06
	HIB_05850	DNA polimerase III subunidade delta'	128,55	17,69

O gene HIB_19160 codifica para a proteína glutamato racemase (MurI), que está associada ao metabolismo de D-glutamina e D-glutamato (via hiu00471 do KEGG) e é necessária para a biossíntese de D-ácido glutâmico, componente da peptideoglicana bacteriana (LUNDQVIST *et al.*, 2007). O gene HIB_19130 codifica para uma proteína de resistência à azaleucina AzlC, uma permease de aminoácido de cadeia ramificada (BELITSKY *et al.*, 1997).

Os principais genes *hub* do módulo *Black* têm, em sua maioria, função associada à biossíntese de parede celular, além de regulação de resposta a estresse, utilização de ferro e absorção de aminoácidos. A expressão do autogene do módulo *Black* indica expressão maior na amostra coletada às 08h (S02) do cultivo (**Figura 11**), que representa crescimento exponencial ainda com o meio inicial do reator, que condiz com consumo de hemina e síntese de parede celular e de LOS. Interessantemente, a presença do gene HIB_04480 para RelA neste módulo indica que há uma resposta estrigente ainda na fase exponencial de crescimento. Essa é uma resposta a estresse devido à limitação ou ausência de aminoácidos, que faz com que a célula desvie recursos do crescimento e divisão celular para síntese de

aminoácidos (POTRYKUS; CASHEL, 2008). Este é um achado inesperado, visto que o meio é rico em aminoácidos, mas que pode indicar que, mesmo com a riqueza do meio, as células podem estar encontrando certo nível de deprivação de aminoácidos, o que, como apontado, altera sobremaneira a fisiologia da célula e o comportamento da população bacteriana.

Tabela 3. Relação dos 10 genes com maior conectividade intra-modular (*hubs*) do módulo *Black*. Para cada gene hub, a proteína para a qual codifica, a conectividade intra-modular $\Sigma_{i \in \text{módulo}} C_{ii}$ e a conectividade extra-modular $\Sigma_{j \notin \text{módulo}} C_{ij}$.

Módulo	gene	Produto	$\Sigma_{i \in \text{módulo}} C_{ii}$	$\Sigma_{j \notin \text{módulo}} C_{ij}$
<i>Black</i>	HIB_04660	CMP-Neu5Ac--lipooligossacarídeo alpha 2-3 sialiltransferase (Lic3A)	11,87	19,20
	HIB_RS08725	proteína LicA	11,84	15,53
	HIB_09640	transglicolase de peptidoglicano	11,65	30,59
	HIB_09620	transglicolase de mureína	11,16	12,74
	HIB_09870	proteína putativa de utilização de ferro/hemina	11,10	17,47
	HIB_18580	proteína hipotética	11,04	38,36
	HIB_16980	LPS colinafosfotransferase (LicD)	9,34	6,48
	HIB_04480	GTP pirofosfoquinase	9,29	58,89
	HIB_19160	glutamate racemase	9,20	19,22
	HIB_19130	permease de aminoácido de cadeia ramificada	9,12	41,20

No módulo *Purple* (**Tabela 4**), o principal gene *hub* foi HIB_00290 (*dacA*), para D-alanyl-D-alanine carboxypeptidase (também *Penicilin-binding protein 5/6*), enzima envolvida na biossíntese de parede celular bacteriana, mais especificamente, na transpeptidação que une as cadeias laterais do peptídeo de cadeias de peptidoglicano (SAUVAGE *et al.*, 2008; KAWAI *et al.*, 2010).

O gene HIB_18220 codifica para a proteína da membrana SapZ, envolvida na via de resistência de peptídeo antimicrobiano catiônica (via hiu1503 do KEGG, *cationic antimicrobial peptide* - CAMP resistance) (SHELTON *et al.*, 2011).

O gene HIB_16970 codifica para a proteína CTP-fosfocolina citidililtransferase (LicC), que participa na via de incorporação de colina no LOS de Hi (LYSENKO *et al.*, 2000).

O gene HIB_16810 codifica para a proteína que participa no sistema de transporte complexo de ferro (via hiu_M00240 do KEGG) (ADHIKARI *et al.*, 1995). O gene HIB_04730 codifica para proteína de ligação de ATP do sistema de transporte de ferro/manganês (transportadores ABC, via hiu02010 do KEGG) ortólogo à proteína SitB de *Escherichia coli* (SABRI *et al.*, 2006).

O gene HIB_00300 codifica para a lipoproteína rara A (*rare lipoprotein A* - RlpA), uma transglicolase lítica que contribui na separação celular, melhor caracterizada em *Pseudomonas aeruginosa* (41% de identidade de aminoácidos com RlpA de Hib) (JORGENSEN *et al.*, 2014; DIK *et al.*, 2017).

O gene HIB_00520 codifica para a enzima lipídio IV A quinase (LpxK), que adiciona fosfato à posição 4' do lipídeo A e está associada à biossíntese de LOS (via hiu00540 do KEGG) (WANG; QUINN, 2010). O gene HIB_RS08720 é um gene fase variável que codifica para a proteína CMP-Neu5Ac--lipooligossacarídeo alpha 2-3 sialiltransferase (Lic3A), associada à sialilação do LOS de Hi (HOOD *et al.*, 2001).

O gene HIB_05150 codifica para a tRNA(Ile)-lisidina sintetase, enzima do tipo aminoacil-tRNA sintetase, que catalisa a ligação do aminoácido à sua molécula de RNA transportador correspondente (SOMA *et al.*, 2003; NAKANISHI *et al.*, 2005). Com essa enzima, citidina é convertida em lisidina, alterando a especificidade de aminoácido do tRNA de metionina para isoleucina.

O gene HIB_09520 codifica para a enzima galactoquinase (GalK), envolvida no catabolismo de galactose exógena, convertendo α -D-galactose em galactose 1-fosfato (hiu00052) (MASKELL *et al.*, 1992; HOLDEN *et al.*, 2003).

Os principais genes *hub* do módulo *Purple* têm em sua maioria função associada à biossíntese de parede celular (concordando com a via sobre-representada encontrada pelo teste hipergeométrico), absorção de ferro, absorção de galactose, além de uma aminoacil-tRNA sintetase. A expressão do autogene deste módulo (**Figura 11**) indica expressão maior nas amostras coletadas no início do cultivo (S01) e na fase exponencial de crescimento (S02), ainda com meio do reator. Tal padrão condiz com o momento metabólico das células, de crescimento e divisão celular intensos.

No módulo *Blue* (**Tabela 5**), o principal gene *hub* foi HIB_00510, gene para a enzima 3-desoxi-mano-octulosonato citidililtransferase (CPM-KDO sintase), associada à ativação de KDO, necessária para incorporação de lipopolissacarídeo em bactérias Gram-negativas (KU *et al.*, 2003; PLAIMAS *et al.*, 2010).

O gene HIB_11960 codifica para a proteína diguanilato cíclico (di-GMP cíclica, YajQ), que está envolvida na regulação de funções como adesão, formação de biofilme, síntese de polissacarídeo e síntese de fatores de virulência (TEPLYAKOV *et al.*, 2003; AN *et al.*, 2014).

O gene HIB_02970 codifica para uma proteína putativa permease (PerM) da família de proteínas transportadoras AI-2E (RETTNER; SAIER JR, 2010).

Tabela 4. Relação dos 10 genes com maior conectividade intra-modular (*hubs*) do módulo *Purple*. Para cada gene hub, a proteína para a qual codifica, a conectividade intra-modular $\Sigma_{i \in \text{intra}} c_{ij}$ e a conectividade extra-modular $\Sigma_{j \in \text{extra}} c_{ij}$.

Módulo	gene	Produto	$\Sigma_{i \in \text{intra}} c_{ij}$	$\Sigma_{j \in \text{extra}} c_{ij}$
<i>Purple</i>	HIB_00290	D-alanil-D-alanina carboxipeptidase	5,89	10,81
	HIB_18220	proteína da membrana	5,14	27,71
	HIB_16970	proteína LicC	4,91	15,04
	HIB_16810	proteína do sistema complexo de transporte de ferro	4,79	9,22
	HIB_00300	proteína do tipo RlpA	4,72	9,44
	HIB_04730	permease transportadora ABC de ferro	4,24	10,20
	HIB_00520	tetra acil dissacarídeo 4'-quinase	4,20	21,68
	HIB_RS08720	CMP-Neu5Ac--lipooligosaccharide alpha 2-3 sialyltransferase (Lic3A)	4,15	12,33
	HIB_05150	tRNA(Ile)-lisidina sintetase	4,04	8,35
	HIB_09520	galactoquinase	4,02	6,59

O gene HIB_01530 codifica para uma proteína putativa ainda não caracterizada, que apresenta 97% de identidade de aminoácidos em relação a um regulador transcricional da família DeoR/GlpR, associado a transporte e metabolismo de açúcar, de outra cepa de *Hi* (accession version WP_086935156.1) (MARCHLER-BAUE *et al.*, 2016).

O gene HIB_16650 codifica para a RNA polimerase fator sigma 24 (RpoE), um fator de transcrição associado à regulação de resposta a estresse, incluindo resposta ao estresse do envelope, resposta ao estresse da proteína periplasmática, níveis aumentados de LOS no periplasma, bem como choque térmico e estresse oxidativo (HEIMANN, 2002; RHODIUS *et al.*, 2005).

O gene HIB_00660 codifica para uma proteína putativa ainda não caracterizada, que contém domínio nucleotidiltransferase e apresenta 59% de identidade de aminoácidos em relação à DNA polimerase subunidade beta de *Acetobacter syzygii* (WP_095351947.1) (MARCHLER-BAUE *et al.*, 2016)

O gene HIB_07660 codifica para a proteína tryptophanyl-tRNA sintetase (TrpS), associada à biossíntese de aminoacil-tRNA (via hlu00970 do KEGG) (PARK *et al.*, 2008).

O gene HIB_10030 codifica para a enzima polissacarídeo polimerase, associada ao transporte da cadeia de PRP para membrana externa (LARUE *et al.*, 2011; WILLIS; WHITFIELD, 2013). Em Hib a polimerização do PRP é completada antes da exportação mediada por transportadores ABC (LARUE *et al.*, 2011; WILLIS; WHITFIELD, 2013).

O gene HIB_01690 codifica para proteína hipotética com domínio da superfamília WzzB, associado a proteínas moduladoras do comprimento da cadeia de O-antígeno em LOS bacteriano. Esta proteína apresenta 37% de identidade de aminoácidos em relação a ortólogo de *Actinobacillus pleuropneumoniae* com essa função (accession version ASU15401.1) (MARCHLER-BAUER *et al.*, 2016).

O gene HIB_00670 codifica para a enzima NAD quinase, que catalisa a fosforilação de NAD em NADP utilizando ATP como fonte de fósforo (KAWAI; MURATA, 2008).

Os principais genes *hub* do módulo *Blue* têm, em sua maioria, função associada a regulação de transcrição de outros genes, além de polimerização (DNA e PRP) e uma aminoacil-tRNA sintetase. A expressão do autogene deste módulo indica expressão elevada na fase exponencial tanto do meio do reator (amostra 08h, S02) quanto da alimentação (amostra 15h, S05), além de no final do cultivo (S06). A presença do gene para RpoE entre os *hubs* do módulo *Blue* reforça sua importância no metabolismo de Hi, sendo que este fator de transcrição foi demonstrado como essencial para a sobrevivência desse organismo em ambiente tão estressante quanto dentro de macrófago (CRAIG *et al.*, 2002). Novamente, foram detectadas evidências de que Hib está expressando genes relacionados à estresse, o que reforça a sugestão de que o cultivo em reator, mesmo que otimizado, gera estresse na população bacteriana.

No módulo *Red* (**Tabela 6**), os principais genes *hubs* foram HIB_08540, gene para a enzima transportadora de potássio, e HIB_06190, gene codificante para uma proteína da membrana interna (SCHLÖSSER *et al.*, 1995).

Os genes HIB_18700 e HIB_16850 codificam para proteína pertencente ao transportadores ABC (via hiu02010 do KEGG) transportadora de molibdato (ModA) (RECH *et al.*, 1995).

Tabela 5. Relação dos 10 genes com maior conectividade intra-modular (*hubs*) do módulo *Blue*. Para cada gene hub, a proteína para a qual codifica, a conectividade intra-modular $\Sigma_{i \in \text{Blue}} C_{ii}$ e a conectividade extra-modular $\Sigma_{j \notin \text{Blue}} C_{ij}$.

Módulo	gene	Produto	$\Sigma_{i \in \text{Blue}} C_{ii}$	$\Sigma_{j \notin \text{Blue}} C_{ij}$
<i>Blue</i>	HIB_00510	CPM-KDO sintase	86,24	76,47
	HIB_11960	proteína de ligação a nucleotídeo (YajQ)	82,18	60,88
	HIB_02970	permease (PerM)	77,88	54,61
	HIB_01530	proteína putativa não caracterizada	75,85	71,95
	HIB_16650	RNA polimerase fator sigma-24 (RpoE)	75,82	73,04
	HIB_00660	proteína hipotética	75,58	66,70
	HIB_07660	triptofanil-tRNA sintetase	75,58	62,92
	HIB_10030	polissacarídeo polimerase	75,55	40,40
	HIB_01690	proteína hipotética (Hsp90)	74,86	48,50
	HIB_00670	polifosfato inorgânico/ATP-NAD	74,72	52,15
		quinase		

O gene HIB_18940 codifica para uma transferase de açúcar e fosfato (WecA) que está envolvida na primeira etapa, que ocorre na membrana, da biossíntese de polímero da parede celular bacteriana como o O-antígeno na biossíntese de LOS (LEHRER *et al.*, 2007; WHITFIELD, TRENT, 2014).

O gene HIB_12740 codifica para proteína dissulfeto oxidoredutase (Dsb), que cria e troca pontes dissulfeto em proteínas do periplasma, catalisando esse processo ao trocar suas pontes dissulfeto com os grupos tiol livres de resíduos cisteína em proteínas alvo (KADOKURA *et al.*, 2003; ROSADINI *et al.*, 2008).

O gene HIB_12640 codifica para ADP-heptose--LPS heptosiltransferase, associada à biossíntese de LOS (via hiu000540 do KEGG), adicionando molécula heptose ao KDO (NICHOLS *et al.*, 1997).

O gene HIB_14180 codifica para homoserina O-acetiltransferase (MetX), associada ao metabolismo de cisteína e metionina (via hiu00270 do KEGG), que catalisa a transferência de grupo acila da coenzima A para a homoserina, então direcionando a via metabólica do aspartato para a biossíntese de metionina (MIRZA *et al.*, 2005).

O gene HIB_16080 codifica para uma proteína hipotética ainda não caracterizada, que apresenta 98% de identidade de aminoácidos em relação à regulador transcricional da superfamília HTH_XRE do gênero *Haemophilus* sp (accession version WP_005643896.1, Genbank), que contém proteínas de ligação a DNA à família de reguladores transcricionais de elementos de resposta xenobiótica (MARCHLER-BAUE *et al.*, 2016).

O gene HIB_02740 codifica para proteína da membrana que contém domínio DUF417 com função desconhecida (MARCHLER-BAUE *et al.*, 2016).

O gene HIB_02290 codifica para a proteína E, uma adesina de superfície (RONANDER *et al.*, 2009; SINGH *et al.*, 2011).

Tabela 6. Relação dos 10 genes com maior conectividade intra-modular (*hubs*) do módulo *Red*. Para cada gene *hub*, a proteína para a qual codifica, a conectividade intra-modular Σ_{in} e a conectividade extra-modular Σ_{out} .

Módulo	gene	Produto	Σ_{in}	Σ_{out}
<i>Red</i>	HIB_08540	transportador de potássio	22,48	53,31
	HIB_18700	proteína de ligação a molibdênio (ModA)	21,01	54,49
	HIB_16850	proteína de ligação a substrato de molibdênio (ModA)	20,94	47,34
	HIB_18940	UDP-GlcNAc:undecaprenilfosfato GlcNAc-1-fosfato transferase (WecA)	20,81	70,66
	HIB_12740	tiorredoxina	20,37	37,14
	HIB_12640	ADP-heptose--LPS heptosiltransferase	18,45	47,04
	HIB_14180	homoserina O-acetiltransferase	18,40	25,08
	HIB_16080	proteína hipotética	18,23	22,77
	HIB_02740	proteína da membrana	17,98	52,35
	HIB_02290	protein E adesina de superfície	17,31	92,91

Os principais genes *hub* do módulo *Red* têm função associada à importação/exportação (transportadores ABC, concordando com análise de enriquecimento

funcional) de nutrientes/fatores de virulência, além de síntese de LOS, síntese de parede celular e biossíntese de aminoácido. A expressão do autogene do módulo *Red* (**Figura 11**) indica aumento gradual da expressão ao longo do cultivo, com expressão maior no final da fase exponencial de crescimento (15h, S05) e no final do cultivo (20h, S06). São, portanto, genes que, em geral, têm sua expressão proporcional ao crescimento bacteriano, estando, provavelmente, envolvidos em atividades básicas do crescimento.

No módulo *Magenta* (**Tabela 7**), os principais genes hubs foram HIB_12480 (*ccmB*, *cytochrome c maturation*), gene para a proteína B exportadora de hemina, necessária para exportação de hemina ao periplasma para biogênese de citocromo c, e HIB_06960 (*oxyR*), gene para o regulador transcricional OxyR, que atua na defesa contra danos oxidativos (estresse oxidativo) causados por espécies reativas de oxigênio (KRANZ *et al.*, 2009; WHITBY *et al.*, 2012). O gene HIB_06950 codifica para um regulador transcricional da família TetR/AcrR associado à biossíntese de ácidos graxos (ZHU *et al.*, 2009; DENG *et al.*, 2013).

O gene HIB_17040 codifica para a proteína ImpA, uma DNA polimerase propensa a erros envolvida na resposta SOS que leva ao reparo de DNA fita simples dentro da célula bacteriana (SWEETMAN *et al.*, 2005). O gene HIB_00630 codifica para proteína RecN (*Recombination protein N*) associada a reparo recombinacional de DNA de fita dupla (REYES *et al.*, 2010).

O gene HIB_12280 codifica para a enzima citocromo c nitrito redutase (NrfA), associada ao metabolismo de nitrogênio (via hiu00910 do KEGG), catalisando a redução de seis elétrons de nitrito em amônia (EINSLE *et al.*, 1999), enquanto que o gene HIB_14750 codifica para citocromo c, uma enzima nitrato redutase do periplasma (BRIGE *et al.*, 2001).

O gene HIB_t02 codifica para RNA transportador Glicina-RNA ligase (tRNA-Gly).

O gene HIB_00690 codifica para proteína da membrana interna com domínio da superfamília Gate, associado a proteínas transportadoras de nucleosídeos (MARCHLER-BAUER *et al.*, 2016).

O gene HIB_09530 codifica para a enzima galactose-1-fosfato uridililtransferase (GalT), associada ao metabolismo de galactose (via hiu00052 do KEGG) e ao metabolismo de amino açúcar e açúcar nucleotídeo (via hiu00520 do KEGG), catalisando a troca de nucleotídeos entre UDP-hexose e hexose-1-fosfato (HOLDEN *et al.*, 2003).

Assim, no módulo *Magenta*, os principais *hubs* têm função associada a metabolismo de nitrogênio, resposta a estresse oxidativo (regulação e reparo), cadeia respiratória e

metabolismo de galactose. O gene para o fator de transcrição ativado por estresse oxidativo OxyR ser um *hub* é esperado, visto que espécies reativas de oxigênio (ROS) se formam durante todo o cultivo devido à oferta de O₂ ser alta, tal que genes regulados por OxyR, como catalase, superóxido dismutase e peroxirredoxinas devem ser expressos como defesa (EZRATY *et al.*, 2017). Assim, a expressão desse gene poderia ser esperada com aumento gradual, devido a acúmulo de lesão devido a ROS ou então constante durante o cultivo, devido à oferta constantemente alta de oxigênio. Porém, como podemos observar na **Figura 11**, a expressão do autogene do módulo *Magenta* é maior nas amostras 10h e 11h, imediatamente antes (S03) e após (S04) a alimentação, ou seja, na metade do cultivo.

Tabela 7. Relação dos 10 genes com maior conectividade intra-modular (*hubs*) do módulo *Magenta*. Para cada gene *hub*, a proteína para a qual codifica, a conectividade intra-modular $\sum_{i \in \text{Módulo}} w_{ij}$ e a conectividade extra-modular $\sum_{j \in \text{Outro}} w_{ij}$.

Módulo	gene	Produto	$\sum_{i \in \text{Módulo}} w_{ij}$	$\sum_{j \in \text{Outro}} w_{ij}$
<i>Magenta</i>	HIB_12480	proteína B exportadora de hemina	4,66	7,95
	HIB_06960	regulador transcricional OxyR	4,31	7,69
	HIB_06950	regulador de biossíntese de ácidos graxos da família TetR/AcrR	4,20	9,26
	HIB_17040	proteína ImpA (inner membrane protein A)	4,14	2,30
	HIB_00630	proteína de reparo do DNA RecN	3,86	4,14
	HIB_12280	nitrito redutase	3,78	3,86
	HIB_14750	citocromo C	3,75	1,38
	HIB_t02	tRNA-Gly	3,70	4,82
	HIB_00690	proteína da membrana	3,50	1,64
	HIB_09530	galactose-1-fosfato uridililtransferase	3,48	4,52

No módulo *Green* (**Tabela 8**), o principal gene hub foi HIB_15280, para glicose-1-fosfato adeniltransferase, envolvido na biossíntese de ADP-glicose, um bloco de construção necessário para reações de alongamento para produzir glicogênio (LEUNG; PREISS, 1987).

O gene HIB_15270 codifica para a enzima (GlgX), que catalisa a conversão de glicogênio em α -D-glicose-1-fosfato, enquanto que o gene HIB_15260 codifica para a enzima ramificadora 1,4- α -glucan (GlgB), que catalisa a conversão de amilose em glicogênio, ambas

associadas ao metabolismo de amido e sacarose (via hiu00510 do KEGG) (CHANDRA *et al.*, 2011).

O gene HIB_08200 codifica para a enzima glicerol quinase (glpK), catalisando a conversão de glicerol em glicerol-3-fosfato, envolvida no metabolismo de glicolípido (via hiu00561 do KEGG) (PAWLYK; PETTIGREW, 2001).

O gene HIB_00700 codifica para uma proteína da membrana interna com domínio da superfamília Gate (MARCHLER-BAUER *et al.*, 2016).

O gene HIB_05090 codifica para a enzima ADP-ribose pirofosfatase (NudF), que catalisa a conversão de ADP-ribose em ribose-5-fosfato (MORENO-BRUNA *et al.*, 2001). O gene HIB_10210 codifica para a enzima fosforibosilamina-glicina ligase (PurD), que catalisa a conversão de ribosilamina-5-fosfato em 5'-fosforibosilglicinamida (GAR) (AIBA; MIZOBUCHI, 1989). O gene HIB_10200 codifica para proteína com duas funções: inosina 5'-monofosfato (IMP) magetaciclohidrolase e fosforibosilaminoimidazolcarboxamida formiltransferase (PurH), catalisando conversões de 5-Aminoimidazol-4-carboxamida ribonucleotídeo (AICAR) para 5-Formamidoimidazol-4-carboxamida ribotídeo (FAICAR) e então FAICAR para inosina 5'-monofosfato (IMP) (AIBA; MIZOBUCHI, 1989). O gene HIB_16320 codifica para a enzima fosforibosilglicinamide formiltransferase 1 (PurN), e catalisa conversão de 5-fosforibosilglicinamida (GAR) para 5'-fosforibosilformilglicinamida (FGAR) (SMITH; DAUM, 1987). Esses quatro genes estão associados ao metabolismo de purina (via hiu00230 do KEGG).

O gene HIB_07970 codifica para frutose 1,6-bisfosfatase II (GlpX), convertendo β -D-frutose-1,6-P₂ em β -D-frutose-6-P, envolvida na glicólise (via hiu00010 do KEGG) (DONAHUE *et al.*, 2000).

Assim, no módulo *Green*, os principais *hubs* têm função associada a metabolismo de purina, formação e metabolismo de glicogênio, além de metabolismo de sacarose/amido e metabolismo de glicolípido, tal que além de serem vias enriquecidas neste módulo (**Figura 13**), alguns dos genes nestas vias são *hubs*. A expressão do autogene do módulo *Green* (**Figura 11**) indica expressão maior nas amostras 10h (imediatamente antes da alimentação, S03), em que a concentração de glicose é mínima e estoques de glicogênio passam a ser consumidos e 11h (imediatamente após o início da alimentação, S04), em que glicose passa a ser ofertada ao reator em abundância e glicogênio volta a ser estocado nas células.

No módulo *Pink* (**Tabela 9**), o principal gene *hub* foi HIB_07180 (*argF/arcB*), para a enzima ornitina carbamoiltransferase (também ornitina transcarbamilase), envolvida na degradação de arginina através do ciclo da ureia (CUNIN *et al.*, 1986).

O gene HIB_01490 codifica para glicerato 2-quinase (GlxK), enzima que catalisa a conversão de D-glicerato em 2-fosfo-D-glicerato, associada ao metabolismo de glicolípido (ZELCBUCH *et al.*, 2015).

Os genes HIB_t03 e HIB_t04 codificam, respectivamente, para tRNA-Ala e tRNA-Asn

Os genes HIB_05640 e HIB_RS09025 codificam para proteína regulatória da família alpA. A proteína AlpA em *Escherichia coli* reprime a superexpressão de polissacarídeo capsular e sensibilidade a luz ultravioleta (TREMPY *et al.*, 1994).

Tabela 8. Relação dos 10 genes com maior conectividade intra-modular (*hubs*) do módulo *Green*. Para cada gene *hub*, a proteína para a qual codifica, a conectividade intra-modular Σ_{intra} e a conectividade extra-modular Σ_{extra} .

Módulo	gene	Produto	Σ_{intra}	Σ_{extra}
<i>Green</i>	HIB_15280	glicose-1-fosfato adenililtransferase	16,29	7,28
	HIB_15270	proteína desramificadora de glicogênio	15,78	6,90
	HIB_08200	glicerol quinase	15,70	7,74
	HIB_15260	proteína ramificadora de glicogênio	15,33	6,58
	HIB_05090	ADP-ribose pirofosfatase	15,18	6,97
	HIB_10210	fosforibosilamina--glicine ligase	15,00	5,45
	HIB_10200	proteína purH (biossíntese de purina)	14,82	5,93
	HIB_16320	fosforibosilglicinamida formiltransferase	14,48	5,61
	HIB_00700	proteína de membrana	14,31	7,02
	HIB_07970	frutose 1,6-bisfosfatase	14,00	8,53

O gene HIB_12680 codifica para proteína permease do sistema de transporte de xilose (XylH), que também transporta ribose, pertencente aos transportadores ABC (hiu02010) (SONG; PARK, 1998; SCHNEIDER, 2001).

O gene *ffs* é gene não codificante cujo produto, um *small RNA* (sRNA), é essencial para crescimento celular e síntese proteica (JENKINS *et al.*, 1998; SCHNEIDER, 2001).

Os genes HIB_15900 e HIB_15930 codificam para proteínas hipotéticas sem função caracterizada nem domínio conservado.

Assim, no módulo *Pink*, os principais hubs têm função associada a metabolismo de glicolípídeo, metabolismo de aminoácido, reguladores transcricionais, tRNA e um sRNA. A expressão do autogene do módulo *Pink* (**Figura 11**) indica expressão maior na metade do cultivo (amostras 10 h e 11 h, respectivamente, S03 e S04).

Tabela 9. Relação dos 10 genes com maior conectividade intra-modular (hubs) do módulo *Pink*. Para cada gene hub, a proteína para a qual codifica, a conectividade intra-modular $\sum_{i=1}^n c_{ii}$ e a conectividade extra-modular $\sum_{j=1}^m c_{ij}$.

Módulo	gene	Produto	$\sum_{i=1}^n c_{ii}$	$\sum_{j=1}^m c_{ij}$
<i>Pink</i>	HIB_07180	ornitina carbamoiltransferase	9,26	7,16
	HIB_01490	glicerato quinase	7,99	13,05
	HIB_t04	tRNA-Asn	7,73	4,93
	HIB_05640	Proteína regulatória da família AlpA	7,73	2,36
	HIB_12680	Permesase de xilose (transportador ABC)	7,51	14,50
	HIB_t03	tRNA-Ala	7,31	10,69
	<i>ffs</i>	signal recognition particle RNA	7,13	2,58
	HIB_RS09025	Proteína regulatória da família AlpA	6,90	1,48
	HIB_15900	Proteína hipotética	6,55	7,14
	HIB_15930	Proteína hipotética	6,55	9,75

4.6 Análise de expressão diferencial de genes

A análise de expressão diferencial permitiu observar o comportamento de Hib durante o cultivo. Para teste hipergeométrico de vias enriquecidas, foram consideradas todas as quinze comparações possíveis entre as seis amostras do cultivo (ou seja, S01 x S02, S01 x S03, S01 x S04, S01 x S05, S01 x S06, S02 x S03, S02 x S04, S02 x S05, S02 x S06, S03 x S04, S03 x S05, S03 x S06, S04 x S05, S04 x S06 e S05 x S06). As vias enriquecidas por teste hipergeométrico são apresentadas na **Figura 16** e sugerem que as vias enriquecidas são o metabolismo de glicerofosfolípídeo (via hiu00564 do KEGG) na comparação S01xS02; via do ribossomo (via hiu03010 do KEGG) e metabolismo de purina (via hiu00230 do KEGG) em S01xS03; metabolismo de purina (via hiu00230 do KEGG) em S01xS04; via do

ribossomo (via hiu03010 do KEGG) em S01xS05; biossíntese de arginina (via hiu00230 do KEGG) em S01xS06; metabolismo de purina (via hiu00230 do KEGG), *one carbon pool by folate* (via hiu00670 do KEGG) e biossíntese de antibióticos (via hiu01130 do KEGG) em S02xS03; metabolismo de purina (via hiu00230 do KEGG), biossíntese de antibióticos (via hiu01130 do KEGG) e biossíntese de metabólitos secundários (via hiu01110 do KEGG) em S02xS04; metabolismo de alanina, aspartato e glutamato (via hiu00250 do KEGG) em S02xS05; via dos transportadores ABC (via hiu02010 do KEGG) em S02xS06; via dos transportadores ABC (via hiu02010 do KEGG) e biossíntese de folato (via hiu00790 do KEGG) em S03xS05; e biossíntese de folato (via hiu00790 do KEGG) em S03xS06. Nenhuma outra comparação apresentou via enriquecida por teste hipergeométrico com *p*-valor ajustado (FDR) 0,05.

Figura 16 - Enriquecimento de genes diferencialmente expressos nas comparações entre amostras. Neste gráfico de pontos, no eixo y estão as vias metabólicas sobre-representadas, no eixo x as comparações que apresentaram vias enriquecidas e em parênteses a quantidade de genes presentes nas vias enriquecidas. Cada circunferência representa a razão de genes encontrada em relação ao total de genes na via metabólica. A cor se refere ao *p*-valor, em que quanto mais próximo de azul, mais próximo do limite de *threshold* ($\alpha = 0,05$).

Uma análise mais detalhada foi realizada entre amostras de tempos subsequentes. Assim, foram feitas 5 comparações dois a dois: S01 x S02, S02 x S03, S03 x S04, S04 x S05 e S05 x S06 (**Tabela 10**). Estas comparações foram escolhidas devido ao fato de representarem resposta fenotípica às condições ambientais consideradas metabolicamente críticas que

queremos investigar. As outras comparações possíveis não foram feitas para que o modelo fosse simplificado.

Em todas as comparações realizadas (S01xS02, S02 x S03, S03 x S04, S04 x S05 e S05 x S06), os pacotes *edgeR*, *DESeq2* e *limma-voom* apresentaram resultados similares, sugerindo que essas metodologias têm resultados convergentes independentemente da distribuição dos dados adotada.

No começo do cultivo, uma vasta gama de genes está expressa diferencialmente em relação aos demais tempos, refletindo a adaptação deste micro-organismo ao meio e às condições do reator. Entre as amostras coletadas às 02h (S01), início de cultivo, e às 08h de cultivo (S02), fase exponencial de crescimento com meio do reator, foram identificados ao todo trinta e nove genes de fato diferencialmente expressos, destacando-se uma diversidade de permeases e enzimas associadas a metabolismo de aminoácidos, descritos abaixo.

Os genes HIB_11490, HIB_11500, HIB_11510 e HIB_11520 são genes associados à via de aldolase-DUF1537, que codificam para proteínas que transformam L-treonato ao oxidar, isomerizar, e fosforilar e então converter em fosfato de dihidroxiacetona (respectivamente, L-treonate 2-desidrogenase - LtnD, 2-desidrotetronato isomerase - OtnI, 3-desidrotetronato 4-quinase - OtnK, 3-desidro-4-fosfotetronato descarboxilase - OtnC) (ZHANG et al., 2016).

Os genes HIB_15060, HIB_12380 e HIB_08360 codificam para permeases, respectivamente a) Uup (associado a excisão de transposon) do sistema de transporte ABC; b) FliY do sistema de transporte de cistina; e c) Mtr de transporte triptofano-específica (HEATWOLE; SOMERVILLE, 1991; HOSIE; POOLE, 2001; LICHT; SCHNEIDER, 2011).

Os produtos dos genes HIB_08350, HIB_12370, HIB_04550, HIB_06170, HIB_15080, HIB_15070, HIB_RS07460 e HIB_16620 ainda não foram caracterizados. Porém, o gene HIB_04550 codifica para proteína com domínio da superfamília DUF469 (MARCHLER-BAUER et al., 2016). O gene HIB_15080 codifica para proteína com 68% de identidade de aminoácidos em relação a proteína estrutural de hemocina de *Rodentibacter genomosp.* 2 (ALTSCHUL et al., 1990). O gene HIB_16620 codifica para proteína com domínio conservado da superfamília DUF2846 (MARCHLER-BAUER et al., 2016). O gene HIB_15070 possui 61% de identidade de aminoácidos em relação à proteína de imunidade a hemocina de *Rodentibacter heylII* (ALTSCHUL et al., 1990; KLEANTHOUS; WALKER, 2001). Os demais não apresentam domínio conservado (MARCHLER-BAUER et al., 2016).

Os genes HIB_18380, HIB_00680, HIB_08150, HIB_05280 e HIB_08950 codificam para proteínas associadas a metabolismo de aminoácidos. Respectivamente, para a) ribonucleosídeo-difosfato redutase (RndA) e b) ribonucleosídeo-trifosfato redutase anaeróbia (RndD), que catalisa a conversão de GTP em dGTP, ambas a) e b) envolvidas no metabolismo de purina (via hiu00230 do KEGG) e de pirimidina (via hiu00240 do KEGG); c) triptofanase (TnaA), associada ao metabolismo de triptofano (via hiu00380 do KEGG) e que catalisa a conversão de triptofano em indol; e d) tiamina-fosfato pirofosforilase (ThiE), associada ao metabolismo de tiamina (via hiu00730 do KEGG); e e) 3,4-dihidroxi 2-butanona 4-fosfato syntase (RibB), proteína associada ao metabolismo de riboflavina (via hiu00740 do KEGG) (YANG *et al.*, 1997; MARTIN *et al.*, 1998; MIZOTE *et al.*, 1999; LOGAN *et al.*, 2003; MARX *et al.*, 2008; MÜLLER *et al.*, 2009).

O gene HIB_15550 codifica para uma proteína do tipo ferritina (FtnA) (HARRISON *et al.*, 2013). O gene HIB_02800 codifica para uma proteína transportadora sensível a cloranfenicol (RarD) (MARCHLER-BAUER *et al.*, 2016). O gene HIB_19030 codifica para a proteína UPF0271, que possui 42% de identidade de aminoácidos em relação à proteína LamB de *Aspergillus nidulans*, associada à utilização de lactam, e que pode ser necessária para a conversão de 2-pirrolidinona exógeno em ácido gama-aminobutírico (GABA) endógeno (ALTSCHUL *et al.*, 1990; RICHARDSON *et al.*, 1992). O gene HIB_10030 codifica para a enzima polissacarídeo polimerase, associada ao transporte da cadeia de PRP para membrana externa (LARUE *et al.*, 2011; WILLIS; WHITFIELD, 2013). O gene HIB_RS10050 codifica para peptídeo líder triptofanase (STEWART; YANOFSKY, 1986). O gene HIB_15100 codifica para receptor de hemina (LEE, 1992). O gene HIB_14520 codifica para a CidA proteína do tipo holina (RANJIT *et al.*, 2011). O gene HIB_03170 codifica para a enzima UDP-galactose-lipologissacarídeo galactosiltransferase (LgtC) (Ho *et al.*, 2007). O gene HIB_08130 codifica para a pirofosfatase inorgânica, associada à fosforilação oxidativa (via hiu00190 do KEGG), catalisando a conversão de pirofosfato em dois íons fosfato (HAROLD, 1966). O gene HIB_11070 codifica para O-succinilbenzoato syntase (MenC), associado a biossíntese de quinona (via hiu00130 do KEGG) (SHARMA *et al.*, 1993). O gene HIB_05200 codifica para uma peptidase (SWETHA *et al.*, 2014). O gene HIB_14930 codifica para proteína ribossomal S15 (RpsO, via hiu03010 do KEGG) (SWETHA *et al.*, 2014). O gene HIB_05930 codifica para uma oxirredutase (SWETHA *et al.*, 2014). O gene HIB_19070 codifica para a exorribonuclease II (Rnb) (ARRAIANO *et al.*, 2010). O gene HIB_15540 codifica para a ferritina (FtnA), proteína de armazenamento de ferro (SWETHA *et al.*, 2014).

O gene HIB_03010 codifica para RNA helicase ATP-dependente (DeaD), associada à degradação de RNA (via hiu03018 do KEGG) (PRUD'HOMME-GÉNÉREUX *et al.*, 2004). O gene HIB_15180 codifica para Dps (*DNA-binding protein in starved cells*), uma proteína multifuncional que protege a célula bacteriana contra estresse oxidativo (MARTINEZ; KOLTER, 1997).

O enriquecimento para metabolismo de glicerofosfolípídeo (Figura 16) na comparação S01xS02 é esperado, pois entre essas amostras há crescimento exponencial da bactéria, e níveis de expressão de genes associados a síntese de ácidos graxos são coordenados com taxa de crescimento, disponibilidade de nutrientes e estímulos ambientais (PARSONS; ROCK, 2013). Porém, os genes desta via diferencialmente expressos entre S01 e S02 estão regulados negativamente, o que pode significar que nutrientes nesta categoria estejam disponíveis no meio em concentração suficiente. O gene HIB_10030 para polissacarídeo polimerase estar regulado positivamente sugere que há exportação maior de PRP em S02 em relação ao início do cultivo, S01. O fato de o gene HIB_15180, codificante para a proteína Dps, estar diferencialmente expresso é interessante, pois esta proteína é multifuncional: oxida íons Fe^{2+} acumulando na forma Fe_2O_3 dentro de sua cavidade proteica e protege a célula bacteriana contra diferentes estresses (por exemplo, oxidativo e ácido) durante o crescimento exponencial ao interagir diretamente com o DNA, condensando o nucleóide bacteriano, além de participar de regulação gênica em estágios iniciais de crescimento bacteriano (KARAS *et al.*, 2015; ARNOLD *et al.*, 2016; ANTIPOV *et al.*, 2017). Assim, a sua expressão pode ser indicativa de que ocorra estresse oxidativo desde o início do cultivo.

Entre as amostras coletadas às 08h (S02), fase exponencial de crescimento, e às 10h de cultivo (S03), pouco antes do início da alimentação, foram identificados ao todo dez genes de fato diferencialmente expressos. Os produtos dos genes HIB_10200 e HIB_10210, HIB_13650 e HIB_16320 (PurH, PurD, PurF e PurN), associados ao metabolismo de purina (via hiu00230 do KEGG), em particular à síntese de novo de inosina monofosfato (IMP, precursor de nucleotídeos púricos), estão regulados positivamente, além de estarem entre os principais *hubs* do módulo *Green* (AIBA; MIZOBUCHI, 1989). O gene HIB_01840 codifica para a proteína PbuG de transporte de xantila/uracila da família AzgA de transportadores de purina (HOSIE; POOLE, 2001). Os genes HIB_13340, HIB_13350 e HIB_13360 codificam para transportadores de arginina, pertencentes aos transportadores ABC (via hiu02010 do KEGG) (HOSIE; POOLE, 2001) e estão regulados negativamente, indicando possível diminuição deste aminoácido no meio. O gene HIB_19110 codifica para uma proteína

hipotética sem domínio conservado, mas com 61% de identidade de aminoácidos em relação à proteína lactato desidrogenase de *Bibersteinia trehalosi* (ALTSCHUL *et al.*, 1990). O gene HIB_t31 tem como produto um RNA transportador, tRNA-Trp (SWETHA *et al.*, 2014). A regulação positiva de genes associados à síntese de purinas (via hiu00230 do KEGG), que aparece sobre-representada entre as amostras S02 e S03 (**Figura 16**), sugere falta de nutrientes no meio, visto que esta via metabólica está ativa quando há esgotamento de purinas no meio de cultivo (ZHAO *et al.*, 2015). Em *Escherichia coli*, esse aumento de síntese de purinas é observado em células “famintas”, e diminui após adição de glicose (LINK *et al.*, 2015). De fato, na amostra S03 a fonte de carbono presente no meio do reator está próxima de se esgotar, e é possível que outros componentes também estejam. Outra via sobre-representada em S02xS03 é *one carbon pool by folate* (via hiu00670 do KEGG), parte do metabolismo de co-fatores e vitaminas, necessária para a síntese e metabolismo de purinas e aminoácidos (SHESTOV *et al.*, 2015). A via de biossíntese de antibióticos (via hiu01130 do KEGG) também está sobre-representada nessa comparação S02xS03, e está associada a bactérias patogênicas (SHESTOV *et al.*, 2015).

Entre as amostras coletadas às 10h (S03) e às 11h de cultivo (S04), não houve genes reconhecidos como de fato diferencialmente expressos. Isso se deve provavelmente ao pouco espaçamento de tempo entre uma amostra e outra em relação à diferença de tempo entre os demais pontos amostrais. Entre essas amostras há o início da alimentação com meio rico em glicose e extrato de levedura, e poderia se esperar aumento na expressão de genes associados a transporte e metabolismo de glicose e de componentes do extrato de levedura.

Entre as amostras coletadas às 11h (S04), após o início da alimentação, e às 15h de cultivo (S05), final da fase exponencial de crescimento da bactéria, há apenas um gene de fato diferencialmente expresso, o HIB_15120, gene codificante para a proteína do sistema de transporte de espermidina/putrescina (PotD), pertencente aos transportadores ABC (via hiu02010 do KEGG), e envolvida no transporte de poliamina, que está regulado negativamente (IGARASHI; KASHIWAGI, 1999). Este gene em *Actinobacillus pleuropneumoniae* (79% de identidade de aminoácidos em relação a HIB_15120 de Hib 10810) foi identificado como necessário para o crescimento bacteriano e a tolerância a estresse (ZHU *et al.*, 2017). Poliaminas em bactéria possuem muitas funções incluindo proteção contra estresse oxidativo (WONG *et al.*, 2007).

Entre as amostras coletadas às 15h (S05) e às 20h de cultivo (S06), foram identificados ao todo sete genes de fato diferencialmente expressos. Entre esses, o gene

HIB_09540 codifica para GalR, um repressor transcricional da família LacI (WEICKERT; ADHYA, 1992). O gene HIB_07240 codifica para uma hidrolase. Já o produto do gene HIB_14460 é uma fusão de proteínas corismate mutase T/prefenato desidrogenase (TyrA) associada à biossíntese de tirosina, a primeira com função de oxirredutase com NAD⁺ e NADH⁺ como aceptores e a segunda como transferase (HUDSON; DAVIDSON, 1984; CHIU *et al.*, 2010). O gene HIB_06060 codifica para proteína TyrP de transporte trosina-específica (SWETHA *et al.*, 2014). O gene HIB_t14 codifica para RNA transportador tRNA-Asp (SWETHA *et al.*, 2014). Assim, como entre S02 e S03, novamente os genes HIB_13350 e HIB_13360 para proteínas ArtP e ArtI do sistema de transporte de arginina estão regulados negativamente (HOSIE; POOLE, 2001).

Tabela 10. Lista de genes diferencialmente expressos. Para cada comparação entre amostras, estão listados os genes identificados como diferencialmente expressos tanto pelos pacotes *edgeR* (e) quanto pelo *DESeq2* (D) e pelo *limma-voom* (l). Apenas genes com log-fold change (logFC) a partir de 2 e *p*-valor ajustado no máximo 0,01.

Comparação	Gene	logFC (e)	<i>p</i> -valor	logFC (D)	<i>p</i> -valor	logFC (l)	<i>p</i> -valor
S01xS02	HIB_08130	-2,08	3,90e-22	-2,12	5,50e-26	-2,05	2,91e-07
	HIB_08140	-2,66	1,04e-20	-2,71	7,44e-23	-2,64	3,27e-07
	HIB_11490	-3,39	9,00e-19	-3,45	1,22e-21	-3,27	3,27e-07
	HIB_11500	-3,03	3,47e-16	-3,09	7,89e-18	-2,97	1,39e-06
	HIB_05200	-2,05	1,14e-14	-2,10	3,16e-16	-2,03	1,42e-06
	HIB_14930	2,55	1,33e-13	2,50	5,21e-15	2,47	3,73e-06
	HIB_15550	3,76	5,69e-13	3,72	5,21e-15	3,64	1,41e-05
	HIB_18380	-2,47	4,11e-14	-2,53	5,21e-15	-2,47	3,92e-06
	HIB_12380	3,64	1,89e-12	3,60	1,21e-14	3,63	5,18e-06
	HIB_16620	2,59	7,20e-14	2,55	2,31e-14	2,60	6,70e-06
	HIB_15080	2,39	3,84e-13	2,35	3,40e-14	2,37	3,91e-06
	HIB_11510	-2,36	5,42e-12	-2,40	8,68e-14	-2,28	3,72e-06
	HIB_05930	-2,10	4,90e-13	-2,14	5,36e-13	-2,12	1,42e-06
	HIB_19070	-2,30	1,50e-56	-2,34	1,16e-12	-2,18	1,69e-05
	HIB_15180	3,21	9,70e-11	3,16	5,38e-12	3,17	2,63e-05
	HIB_11520	-2,17	3,46e-10	-2,23	1,43e-11	-2,11	1,06e-05
	HIB_04550	2,07	4,70e-10	2,03	1,53e-10	2,10	2,38e-05

	HIB_08350	3,54	2,72e-09	3,50	1,71e-10	3,29	1,32e-04
	HIB_12370	2,11	6,32e-10	2,06	2,71e-10	2,06	4,57e-05
	HIB_08150	-2,72	1,51e-08	-2,76	5,70e-10	-2,69	6,67e-05
	HIB_11070	-2,31	7,29e-09	-2,36	2,23e-09	-2,37	1,69e-05
	HIB_02800	2,11	4,58e-08	2,06	9,07e-09	2,22	4,13e-05
	HIB_15540	2,81	3,00e-08	2,77	9,07e-09	2,79	9,20e-05
	HIB_06170	2,07	3,60e-08	2,03	9,12e-09	2,09	8,17e-05
	HIB_05280	-2,26	2,27e-07	-2,33	8,43e-08	-2,15	1,03e-04
	HIB_08360	2,58	7,73e-07	2,53	5,19e-07	2,41	9,16e-04
	HIB_08950	-2,11	2,87e-06	-2,16	9,88e-07	-2,11	2,39e-04
	HIB_19030	2,52	1,60e-06	2,48	9,88e-07	2,40	1,17e-03
	HIB_00680	-2,09	1,07e-05	-2,14	1,99e-06	-2,01	4,87e-04
	HIB_15100	3,36	9,97e-06	3,31	2,48e-06	3,09	3,30e-03
	HIB_10030	2,36	1,53	2,31	3,70e-06	2,32	9,50e-04
	HIB_03010	-2,24	3,88e-05	-2,31	5,97e-06	-2,35	6,96e-14
	HIB_RS10050	3,72	1,07e-05	3,69	9,62e-06	4,01	4,22e-04
	HIB_15070	2,96	2,66e-04	2,92	8,10e-05	2,82	5,63e-03
	HIB_14520	2,71	2,86e-04	2,67	4,00e-04	2,73	4,01e-03
	HIB_RS07460	2,30	2,78e-03	2,26	1,18e-03	2,48	8,35e-03
	HIB_03170	2,62	1,25e-03	2,57	2,06e-03	2,49	9,25e-03
	HIB_15060	3,09	1,21e-03	3,04	8,34e-03	2,47	3,27e-03
S02xS03	HIB_10200	3,69	1,54e-25	3,68	8,59e-28	3,70	1,41e-06
	HIB_10210	3,35	1,8e-21	3,34	4,44e-24	3,40	1,41e-06
	HIB_01840	2,86	4,15e-15	2,85	6,72e-16	2,90	1,72e-05
	HIB_16320	2,06	8,4e-12	2,05	6,84e-13	2,08	2,64e-05
	HIB_13650	2,08	6,75e-12	2,07	1,62e-12	2,07	2,64e-05
	HIB_13360	-5,17	2,26e-08	-5,18	2,52e-09	-5,02	1,54e-04
	HIB_13340	-2,93	1,35e-06	-2,94	4,00e-08	-2,55	3,70e-03
	HIB_13350	-3,98	3,65e-07	-3,99	1,74e-07	-3,83	4,52e-04
	HIB_19110	-2,53	1,05e-05	-2,55	2,13e-06	-2,42	5,75e-03
	HIB_t31	-2,66	3,10e-04	-2,70	5,85e-04	-2,48	5,75e-03

S03xS04							
S04xS05	HIB_15120	-2,34	3,33e-7	-2,32	3,91e-8	-2,29	1,37e-03
S05xS06	HIB_06060	3,77	1,57e-27	3,76	3,51e-32	3,51	1,05e-07
	HIB_14460	3,36	2,33e-28	3,35	3,51e-32	3,18	1,05e-07
	HIB_07240	-2,13	4,26e-06	-2,15	7,62e-07	-2,11	1,01e-03
	HIB_t14	-2,20	8,55e-06	-2,21	2,25e-06	-2,11	1,47e-03
	HIB_13360	-3,39	1,64e-04	-3,44	3,10e-05	-4,21	4,00e-04
	HIB_13350	-2,89	1,12e-04	-2,94	3,86e-05	-3,44	5,54e-04
	HIB_09540	-2,27	6,05e-04	-2,28	9,90e-04	-2,18	8,33e-03

4.7 Sugestões de alterações no processo de obtenção de PRP, visando aumento de produção

A produção de acetato por *Haemophilus influenzae* é prevista em modelos *in silico*, assim como observada em fermentações aeróbias com glicose como fonte de carbono (EDWARDS; PALSSON, 1999; SCHILLING; PALSSON, 2000; SCHILLING *et al.*, 2000; PAPIN *et al.*, 2002; RAGHUNATHAN *et al.*, 2004; OTHMAN *et al.*, 2014). A única via conhecida em Hib para síntese de acetato contém os genes HIB_14020, para a proteína fosfotransacetilase e HIB_13620, para a proteína acetato quinase, cujas expressões ao longo do cultivo como podem ser observadas na **Figura 17**, refletido na formação de acetato, como podemos observar na **Figura 2**. O fenótipo observado é de aumento crescente de acetato. O aumento da taxa de transcrição do gene HIB_14020 para fosfotransacetilase no início do cultivo pode estar correlacionado com o aumento da quantidade de acetato detectada no meio neste período. Por outro lado, após 11 h de cultivo (S04), a expressão de ambos os genes da via de síntese de acetato varia pouco, o que explicaria a pouca variação na velocidade específica de formação de acetato (Figura 15 do apêndice).

A formação de acetato representa um desvio da fonte de carbono que de outra forma poderia gerar biomassa ou PRP em Hib, e provavelmente se deve a um desbalanço entre a absorção de glicose e sua conversão em biomassa e demais produtos, havendo o desvio de fosfoenolpiruvato (PEP), que pode ser convertido em oxaloacetato do ciclo TCA, em direção ao acetato (**Figura 18**) (SAUER; EIKMANN, 2005). Esse fenômeno é conhecido como efeito Crabtree bacteriano e ocorre quando há alta taxa de crescimento em altas concentrações de glicose, condições predominantes em culturas de batelada alimentada no qual glicose é

fornecida de forma não limitativa em culturas de fermentação aeróbia para obter altas densidades celulares (DOELLE *et al.*, 1982). Tal desvio ocorre quando concentrações altas de glicose aceleram a glicólise, o que resulta na produção de grande quantidade de ATP através de fosforilação ao nível do substrato, reduzindo assim a necessidade de fosforilação oxidativa feita pelo ciclo TCA através da cadeia de transporte de elétrons e então diminui o consumo de oxigênio (DOELLE *et al.*, 1982). Assim, modificações genéticas para superar a formação de acetato envolvem alterações do perfil genético do organismo para reduzir a síntese bioquímica de acetato.

Quatro estratégias diferentes podem ser empregadas visando reduzir a formação de acetato: abordagens que a) reduzam diretamente o transporte de glicose pelo sistema fosfotransferase,, b) reduzam diretamente o fluxo de carbono até acetato, c) visem os mecanismos regulatórios e metabólicos fundamentais que levam à formação de acetato e d) gerem novo fluxo metabólico. As quatro estratégias são descritas abaixo.

Para a estratégia de reduzir o transporte de glicose pelo sistema fosfotransferase (*phosphotransferase system*, PTS): devido às elevadas taxas de consumo de glicose obtida pelo transporte por esse sistema PTS estarem associadas à formação de acetato, uma possibilidade é reduzir a formação de acetato realizando o nocaute de algum dos genes que controlam a absorção de glicose na célula por esse sistema (SAUER; EIKMANN, 2005). Em *Hib*, o gene HIB_18890 (módulo *Brown*, $\Delta_{\text{HIB}_18890} = 28,05$, $\Delta_{\text{HIB}_18890} = 11,58$), que codifica para a enzima glicose-específica IIA (EIIA^{Glc}) do sistema PTS é um candidato (identificado com caixa verde na **Figura 18** e cuja expressão entre as amostras é apresentada na **Figura 32** do Apêndice). Neste sistema, fosfoenolpiruvato (PEP) é usado como fonte de energia e doador de fosforil, sendo que EIIA^{Glc} atua no transporte e fosforilação de glicose (DEUTSCHER *et al.*, 2006). Assim, a glicose passaria a atravessar a membrana celular apenas pelo transporte ativo (sistema permease) (SIMONI; POSTMA, 1975). A deleção de gene associado ao transporte de glicose tende a diminuir a formação de acetato, reduzir a taxa máxima de crescimento e aumentar o rendimento de biomassa. Abordagem similar obteve sucesso em *Escherichia coli*, com nenhuma produção de acetato detectada em cepa sem o gene *crr*, similar (73% de identidade de aminoácidos) ao gene HIB_18890 (PONCE, 1999; SIGÜENZA *et al.*, 1999).

Uma outra abordagem genética é eliminar o fluxo de carbono até acetato ou direcioná-lo a um produto benigno. Para essa estratégia, as vias de formação de acetato de *Haemophilus influenzae* devem ser consideradas. A **Figura 18** mostra as enzimas chave envolvidas na glicólise (glicose até acetil-CoA) e no ciclo tricarboxílico (ciclo TCA), ciclo que, em *Hi*, é

uma via metabólica incompleta, e funciona tanto em condições anaeróbias quanto aeróbias na presença de glicose (TATUSOV *et al.*, 1996). A via metabólica de três etapas envolvendo o complexo piruvato desidrogenase, fosfotransacetilase e acetato quinase, em série, é a rota primária para o acetato, estando envolvidos os genes HIB_14020 e HIB_13620 nesta etapa (identificados com caixa vermelha na **Figura 18**). Uma forma direta de reduzir formação de acetato é eliminar fosfotransacetilase (gene HIB_14020, módulo *Purple*, $\Delta_{\text{HIB}_14020} = 3,57$, $\Delta_{\text{HIB}_13620} = 11,38$) e/ou acetato quinase (gene HIB_13620, módulo *Blue*, $\Delta_{\text{HIB}_14020} = 4,66$, $\Delta_{\text{HIB}_13620} = 2,15$). Essas modificações podem levar a um acúmulo de lactato ou piruvato (CHANG *et al.*, 1999; LICHTENEGGER *et al.*, 2014). O nocaute dos genes associados à síntese de acetato levou à quase eliminação de produção de acetato em *Escherichia coli* (PARIMI *et al.*, 2017). Outra proposta é sobre-expressar o gene para fosfoenolpiruvato carboxilase (gene HIB_18150, módulo *Magenta*, $\Delta_{\text{HIB}_14020} = 2,66$, $\Delta_{\text{HIB}_13620} = 7,60$, identificado com caixa azul na **Figura 18**), que pode desviar o carbono do acetato ou seus precursores, direto para o ciclo TCA. Essa estratégia resultou em redução em 60% da formação de acetato em *Escherichia coli* (FARMER; LIAO, 1997).

Figura 17. Expressão dos genes HIB_14020 e HIB_13620. No eixo y, a expressão dos genes após remoção de efeito de lote e normalização TMM. No eixo x, cada uma das amostras identificadas S01 a S06. Em roxo, o gene HIB_14020 para fosfotransacetilase. Em azul, o gene HIB_13620 para acetato quinase. O traço preto representa a média da expressão normalizada dos genes. A importância de gene para produtividade de acetato desses genes é: $\Delta_{\text{HIB}_14020}(\text{HIB}_14020) = 0,72$; $\Delta_{\text{HIB}_13620}(\text{HIB}_13620) = 0,32$.

Outra abordagem genética envolve alterar os reguladores moleculares que estão associados à via que leva ao acetato. Nesse sentido, poderia ser feita a deleção de genes que regulam o ciclo TCA, sendo que o nocaute do gene HIB_10170 (módulo *Brown*, $\Delta_{\text{HIB}_10170} = 32,94$, $\Delta_{\text{PRP}} = 13,85$), que codifica para a proteína reguladora aeróbia ArcA é uma alternativa que não interfere significativamente nem na formação de PRP nem no crescimento bacteriano (DE SOUZA-HART *et al.*, 2003). Em *Escherichia coli*, essa deleção resultou em redução de 40% na formação de acetato (VEMURI *et al.*, 2006). Outra alternativa é nocaute do gene HIB_15490 (módulo *Blue*, $\Delta_{\text{HIB}_15490} = 32,53$, $\Delta_{\text{PRP}} = 17,89$, identificado com caixa roxa na **Figura 18**), que codifica para a proteína regulatória regulon fosfato PhoB (também carbon source-responsive - CreC), que regula o operon com genes para fosfato acetiltransferase e acetato quinase. Em *Escherichia coli* isso resultou em produção 20% menor de acetato (GODOY *et al.*, 2016).

Uma última abordagem genética para reduzir acetato envolve gerar novo fluxo metabólico. Através da inserção de gene para piruvato carboxilase (PC, identificado com caixa laranja na **Figura 18**), uma enzima não encontrada nativamente em *Haemophilus influenzae* que carboxila o piruvato diretamente a oxaloacetato, a via que leva à formação de acetato pode ser direcionada para o ciclo TCA. Em *Escherichia coli* esta enzima também não é nativa e a inserção desse gene impactou em redução de 40-60% na formação de acetato (GOKARN *et al.*, 2001; MARCH *et al.*, 2002).

Figura 18. Vias metabólicas principais de *Haemophilus influenzae* tipo b envolvidas no consumo aeróbio de glicose e na síntese de acetato, dióxido de carbono e biomassa, modificado a partir de Othman *et al.* (2014). Acetato é formado a partir de acetil-CoA pelas enzimas fosfotransacetilase (HIB_14020 – caixa vermelha) e acetato quinase (HIB_13620 – caixa vermelha). Fosfoenolpiruvato carboxilase (HIB_18150 – caixa azul) é a principal enzima intermediária necessária para reabastecer intermediários do ciclo TCA consumidos para a formação de biomassa. Abreviações: FosfoEnolPiruvato Carboxinase, HIB_09420; Fumarato redutase, HIB_09660; Succinil-CoA Sintetase, HIB_13550; Acetato quinase, HIB_13620; Malato desidrogenase, HIB_13680; complexo piruvato desidrogenase, HIB_13910; Fosfotransacetilase, HIB_14020; Fumarato hidratase, HIB_15690; Piruvato quinase II, HIB_17580; FosfoEnolPiruvato Carboxilase, HIB_18150; 2-oxoglutarato desidrogenase, HIB_18410. Hib consome glicose principalmente pelo sistema fosfotransferase, que simultaneamente gera piruvato de fosfoenolpiruvato. Apenas algumas vias estão indicadas.

Os genes associados à síntese de PRP (HIB_11840, HIB_11830, HIB_11820 e HIB_11810, respectivamente para Bcs1234) não estão diferencialmente expressos, sendo que a produtividade de PRP aumenta com o tempo (**Figura 13** do apêndice).

Em relação à formação de PRP, uma opção para obter maior produção de PRP é escolher uma cepa selvagem que contenha uma quantidade maior de cópias do locus *cap*, visto que há cepas com três a cinco cópias, e que quanto mais cópias do locus *cap*, mais PRP é produzido (CORN *et al.*, 1993; NOEL *et al.*, 1996). Todavia, cepa com seis cópias de

cápsulas pode ser letal para pessoas dentro do grupo de risco para doenças causadas por *Haemophilus influenzae* (CERQUETTI *et al.*, 2006).

Outra opção é direcionar as vias metabólicas que levam à via de síntese de PRP, conforme **Figura 19**. Ao analisar as vias que levam à síntese de PRP, a rota que compreende os genes HIB_06830, HIB_06820 e HIB_06790, respectivamente para glicose-6-fosfato desidrogenase, 6-fosfogluconolactonase e 6-fosfogluconato desidrogenase (via das pentoses-fosfato - hiu00030 do KEGG), está ativa durante todo o cultivo, como podemos observar na **Figura 20**. Os produtos desses genes transformam glicose-6-fosfato e levam à formação de D-Ribulose-5-Fosfato. A proteína bifuncional Bcs1 (CDP-ribitol sintetase/ citidiltransferase) catalisa a conversão de D-ribulose-5P em D-ribitol-5P e então de D-ribitol-5P em CDP-ribitol, o bloco de construção a partir do qual se forma a cadeia de polissacarídeo capsular em Hib (ZOLLI *et al.*, 2001; PEREIRA; BROWN, 2004).

Outra rota também está associada à via das pentoses-fosfato, mas através da enzima transcetolase. De fato, o gene HIB_11930 para transcetolase está com expressão muito alta em relação aos demais genes, mas esta enzima possui mais de uma função: converte a) frutose-6-fosfato em xilulose-5-fosfato; b) gliceraldeído-3-fosfato em D-ribulose-5-fosfato; c) frutose-6-fosfato em eritrose 4-fosfato; d) gliceraldeído-3-fosfato em D-ribose-5-fosfato; e e) sedoheptulose-7-fosfato em D-ribose-5-fosfato. Em todos os casos, transcetolase catalisa as reações em ambas as direções. Então, a enzima ribulose-fosfato 3-epimerase (produto do gene HIB_06910) converte xilulose-5-fosfato em D-ribulose-5-fosfato e a enzima ribose 5-fosfato isomerase (produto do gene HIB_05940) converte D-ribose-5-fosfato em D-ribulose-5-fosfato. A expressão desses genes pode ser observada na **Figura 21**. A sobreexpressão dos genes HIB_06910 e HIB_05940 poderia impactar em aumento da produção de PRP.

Figura 19. Vias metabólicas de *Haemophilus influenzae* tipo b que podem ser direcionadas para síntese de PRP, modificado a partir do KEGG (KANEHISA, 2005).

Abreviações: Glico-6-fosfato desidrogenase, HIB_06830; 6-Fofogluconolactonase, HIB_06820; 6-Fosfogluconato desidrogenase, HIB_06790; D-xilose isomerase, HIB_12710; Xiluloquinase, HIB_12720; Ribulose-fosfato 3-epimerase, HIB_06910; Transcetolase, HIB_11930; Ribose-5-fosfato isomerase, HIB_05940; Ribitol-5-fosfato 2-desidrogenase/ D-ribitol-5-fosfato citidililtransferase, HIB_11690. A síntese de PRP em Hib a partir de glicose passa pela via glicolítica, via de pentoses fosfato e via de interconversão pentose-glucuronato. Apenas algumas vias estão indicadas.

Figura 20. Expressão dos genes HIB_06830, HIB_06820 e HIB_06790. No eixo y, a expressão dos genes após remoção de efeito de lote e normalização TMM. No eixo x, cada uma das amostras identificadas S01 a S06. Em vermelho, o gene HIB_06830 para glicose-6-fosfato desidrogenase. Em marrom, o gene HIB_06820 para 6-fosfogluconolactonase. Em roxo, o gene HIB_06790 para 6-fosfogluconato desidrogenase. O traço preto representa a média da expressão normalizada de todos os genes de Hib analisados. A importância de gene para produtividade de PRP desses genes é: $\text{Importância}(\text{HIB}_{06830}) = 0,19$; $\text{Importância}(\text{HIB}_{06820}) = 0,34$; $\text{Importância}(\text{HIB}_{06790}) = 0,54$.

Figura 21. Expressão dos genes HIB_11930, HIB_06910 e HIB_05940. No eixo y, a expressão dos genes após remoção de efeito de lote e normalização TMM. No eixo x, cada uma das amostras identificadas S01 a S06. Em marrom, o gene HIB_11930 para transcetolase. Em dourado, o gene HIB_06910 para ribulose-fosfato 3-epimerase. Em vermelho, o gene HIB_05940 para ribose 5-fosfato epimerase. O traço preto representa a média da expressão normalizada de todos os genes de Hib analisados. A importância de gene para produtividade de PRP desses genes é: $\text{Importância}(\text{HIB}_{11930}) = 0,63$; $\text{Importância}(\text{HIB}_{06910}) = 0,45$; $\text{Importância}(\text{HIB}_{05940}) = 0,65$.

Outra possível rota existe a partir de D-Xilose até D-Ribulose-5-Fosfato (interconversões pentose e glucuronato – via hiu00040 do KEGG) e inclui o produto dos genes HIB_12710, HIB_12720 e HIB_06910, respectivamente para xilose isomerase, xiluloquinase e ribulose-fosfato 3-epimerase. Podemos observar pela **Figura 22** que a expressão é menor em relação às demais rotas. Neste caso, o aumento da concentração de D-xilose no meio de cultivo pode ser suficiente para aumentar a produção de PRP, visto que Hib possui os genes HIB_12680, HIB_12690 e HIB_12700, para proteínas do sistema de transporte de D-xilose (via hiu02010 do KEGG) (SCHNEIDER, 2001). Aliás, o gene HIB_12680 foi um dos genes com maior intraconectividade modular no módulo *Pink*. O fato de este gene ter alta conectividade pode significar que a regulação dele é provavelmente bastante relacionada à de outros genes, e alterá-lo poderia impactar, como efeito colateral, outros genes/fenótipos, e por conta disso, seja mais seguro fazer alterações no meio.

Figura 22. Expressão dos genes HIB_12710, HIB_12720 e HIB_06910. No eixo y, a expressão dos genes após remoção de efeito de lote e normalização TMM. No eixo x, cada uma das amostras identificadas S01 a S06. Em marrom, o gene HIB_06910 para ribulose-fosfato 3-epimerase. Em azul, o gene HIB_12720 para xiluloquinase. Em vermelho, o gene HIB_12710 para xilose isomerase. A importância de gene para produtividade de PRP desses genes é: $\beta_{\text{HIB}_{12710}}(\beta_{\text{HIB}_{12710}}) = 0,72$; $\beta_{\text{HIB}_{12720}}(\beta_{\text{HIB}_{12720}}) = 0,72$; $\beta_{\text{HIB}_{06910}}(\beta_{\text{HIB}_{06910}}) = 0,45$.

Assim, o aumento da produtividade de PRP pode ser alcançado ao sobre-expressar um gene que desvie o fluxo de carbono para a via de pentose fosfato (PPP), como por exemplo, o gene HIB_06830 (módulo *Black*, $\beta_{\text{HIB}_{06830}} = 2,12$, $\beta_{\text{HIB}_{06830}} = 4,63$), para glicose-6-fosfato 1-desidrogenase, e então sobre-expressar outro gene que favoreça a síntese de PRP, como por exemplo, os genes HIB_11690 (módulo *Blue*, $\beta_{\text{HIB}_{11690}} = 0,26$, $\beta_{\text{HIB}_{11690}} = 0,32$) e HIB_11840 (módulo *Blue*, $\beta_{\text{HIB}_{11840}} = 0,55$, $\beta_{\text{HIB}_{11840}} = 0,80$), ambos genes que codificam para a proteína bifuncional Bcs1 (CDP-ribitol sintetase/citidiltransferase). Tal sobre-expressão pode ser realizada ao aumentar a quantidade de cópias destes genes, tanto inseridas no genoma como em plasmídeos. Alterações no promotor ou em outros reguladores também poderiam ser utilizadas para aumentar a sobre-expressão.

Assim, através da análise de co-expressão gênica, foi possível identificar os genes com maior conectividade em grupos de genes de padrão similar de expressão durante o cultivo. Embora a maioria desses genes esteja associada ao funcionamento básico celular (como formação de parede celular ou associados à transcrição de outros genes, transportadores e proteínas da membrana celular e alguns reguladores de transcrição), pudemos também detectar correlação entre grupos de genes (os módulos) e as variáveis fenotípicas mensuradas, indicando que a expressão de grandes grupos de genes obedece padrões similares de estímulos e sugerindo que a expressão de alguns deles podem estar diretamente relacionada ao estabelecimento dos fenótipos observados.

Com a análise de expressão diferencial, pudemos estudar a resposta de comportamento de Hib aos estímulos do ambiente em nível de transcritos em pontos considerados metabolicamente críticos. Realizamos uma identificação de genes diferencialmente expressos conservadora ao considerar apenas os genes identificados como diferencialmente expressos por três pacotes de análise de expressão (edgeR, DESeq2 e limma-voom), de forma a destacar as alterações de expressão genica mais significativas no cultivo. Desde o início do cultivo, Hib metaboliza aminoácidos transportados do meio de cultivo e responde a estresse de várias origens. Conforme se esgota fonte de carbono no meio do reator, também outros componentes se aproximam do esgotamento, tal que Hib passa a sintetizar tais componentes. Não há mudança significativa na resposta de Hib entre pouco antes e pouco após o início da alimentação, enquanto há pouca entre o início da alimentação e o início da fase estacionária de gás carbônico. No quarto final do cultivo, alguns genes para síntese de aminoácidos e outros para transportadores estão regulados da mesma forma que pouco antes do início da alimentação, quando alguns componentes do meio se esgotavam. Porém, durante a alimentação, estes componentes são ofertados de forma constante.

A partir da expressão de genes diretamente relacionados à produção de acetato e de PRP, foi possível identificar as rotas mais utilizadas para a produção desses produtos e sugerir modificações no cultivo e em *Haemophilus influenzae* com a finalidade de aumentar produção de PRP e diminuir produção de acetato. Quase certamente, um conjunto de modificações terá resultado mais satisfatório para contornar a produção de acetato e/ou aumentar a produção de PRP em comparação com apenas uma ou outra individualmente. Sugere-se que cada alteração seja feita a priori uma por vez, para então resultar em um redesenho do circuito genético de Hib.

5. Conclusões

Oito módulos de genes foram identificados pelo WGCNA, cada um com características metabólicas distintas e padrão de expressão temporal próprio. De acordo com os dados de co-expressão, as categorias ribossomo e biossíntese de peptidoglicano têm sua expressão aumentada no início e diminuída no fim do cultivo, as categorias glicólise/metabolismo de carbono e metabolismo de purina/ciclo TCA têm expressão aumentada na metade do cultivo, enquanto a categoria transportadores ABC tem expressão maior no final do cultivo.

Há correlação entre nível de conectividade (*hub*) e ter importância (correlação entre produtividade de acetato e produtividade de PRP) em quatro dos oito módulos identificados. Nestes módulos, os principais *hubs* são genes com funções regulatórias, outros com funções essenciais nas diversas vias metabólicas e alguns ainda com função desconhecida. Por serem *hubs* e em havendo correlação positiva com produtividade de acetato e produtividade de PRP, talvez esses genes tenham participação, mesmo que indireta, no estabelecimento final do fenótipo.

Um conjunto de propostas para engenharia metabólica de *Haemophilus influenzae* tipo b foi proposto, com foco em alterar a via de síntese de acetato.

Foram propostos aumento da concentração de arginina e de D-xilose no meio de cultivo e por outro lado sobre-expressão de genes HIB_06830 e HIB_11840 para aumentar produção de PRP.

6. Perspectivas

Caracterizar funcionalmente as proteínas para o qual os genes *hubs* codificam. Além disso, confirmar se, as proteínas com atividade conhecida, mas não anteriormente relacionadas à produção de PRP ou acetato, de fato apresentam as correlações, apontadas pelos nossos dados, com estes fenótipos.

Testar se concentração maior de D-xilose impacta na produção de polissacarídeo capsular.

Realizar as modificações genéticas em Hib e verificar impacto na produção de acetato e na produção de polissacarídeo capsular.

7. Referências Bibliográficas

ADHIKARI, P.; KIRBY, S. D.; NOWALK, A. J.; VERALDI, K. L.; SCHYVERS, A. B.; MIETZNER, T. A. Biochemical characterization of a *Haemophilus influenzae* periplasmic iron transport operon. **Journal of Biological Chemistry**, v. 270, n. 42, p. 25142-25149, 1995.

AIBA, A.; MIZOBUCHI, K. Nucleotide sequence analysis of genes *purH* and *purD* involved in the de novo purine nucleotide biosynthesis of *Escherichia coli*. **Journal of Biological Chemistry**, v. 264, n. 35, p. 21239-21246, 1989.

AKERLEY, B. J.; RUBIN, E. J.; NOVICK, V. L.; AMAYA, K.; JUDSON, N.; MEKALANOS, J. J. A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 2, p. 966–71, 2002.

ALBANI, S. M. F.; da SILVA, M. R.; TAKAGI, M.; CABRERA-CRESPO, J. Improvement in the Purification Process of the Capsular Polysaccharide from *Haemophilus influenzae* Type b by Using Tangential Ultrafiltration and Diafiltration. **Applied Biochemistry and Biotechnology**, v. 167, n. 7, p. 2068-2075, 2012.

ALBANI, S. M. F.; SILVA, M. R. D.; FRATELLI, F.; JUNIOR, C. P. C.; IOURTOV, D.; CINTRA, F. D. O.; TAKAGI, M.; CABRERA-CRESPO, J. Polysaccharide purification from *Haemophilus influenzae* type b through tangential microfiltration. **Carbohydrate Polymers**, 2014.

ALLAN, I.; LOEB, M. R.; MOXON, E. R. Limited genetic diversity of *Haemophilus influenzae* (type b). **Microbial pathogenesis**, v. 2, n. 2, p. 139-145, 1987.

ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. **Journal of molecular biology**, v. 215, n. 3, p. 403-410, 1990.

AN, S.-Q.; CALY, D. L.; MCCARTHY, Y.; MURDOCH, S. L.; WARD, J.; FEBRER, M.; DOW, J. M.; RYAN, R. P.; Novel cyclic di-GMP effectors of the YajQ protein family control bacterial virulence. **PLoS pathogens**, v. 10, n. 10, p. e1004429, 2014.

ANDERS, S.; HUBER, W. Differential expression analysis for sequence count data. **Genome biology**, v. 11, n. 10, p. R106, 2010.

ANDERSON, P.; SMITH, D. H. Isolation of the capsular polysaccharide from culture supernatant of *Haemophilus influenzae* type b. **Infection and immunity**, v. 15, n. 2, p. 472-477, 1977.

ANDERSON, P.; SMITH, D. H.; INGRAM, D. L.; WILKINS, J.; WEHRLEAND, P F.; HOWIE, V. M. Antibody to polyribophosphate of *Haemophilus influenzae* type b in infants and children: effect of immunization with polyribophosphate. **Journal of Infectious Diseases**, v. 136, n. Supplement 1, p. S57-S62, 1977.

ANDREWS, S. FASTQC: A quality control tool for high throughput sequence data. 2010. Disponível em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Acesso em 23/07/2016.

ANTIPOV, S. S.; TUTUKINA, M. N.; PREOBRAZHESKAYA, E. V.; KONDRASHOV, F. A.; PATRUSHEV, M. V.; TOSHCHAKOV, S. V.; DOMINOVA, I.; SHVYREVA, U. S.; VRUBLESKAYA, V. V.; MORENKOV, O. S.; SUKHARICHEVA, N. A.; PANYUKOV, V. V.; O. N.; OZOLINE. The nucleoid protein Dps binds genomic DNA of *Escherichia coli* in a non-random manner. **PloS one**, v. 12, n. 8, p. e0182800, 2017.

ARNOLD, A. R.; ZHOU, A.; BARTON, J. K. Characterization of the DNA-mediated Oxidation of Dps, a Bacterial Ferritin. **Journal of the American Chemical Society**, v. 138, n. 35, p. 11290-11298, 2016.

ARRAIANO, C. M.; MATOS, R. G.; BARBAS, A. RNase II: the finer details of the Modus operandi of a molecular killer. **RNA biology**, v. 7, n. 3, p. 276-281, 2010.

ATKINSON, G. C.; TENSION, T.; HAURYLIUK, V. The RelA/SpoT homolog (RSH) superfamily: distribution and functional evolution of ppGpp synthetases and hydrolases across the tree of life. **PloS one**, v. 6, n. 8, p. e23479, 2011.

AUER, P. L.; DOERGE, R. W. Statistical design and analysis of RNA sequencing data. **Genetics**, v. 185, n. 2, p. 405-416, 2010.

AVCI, F. Y. Novel strategies for development of next-generation glycoconjugate vaccines. **Current topics in medicinal chemistry**, v. 13, n. 20, p. 2535-2540, 2013.

BARABASI, A.-L.; OLTVAI, Z. N. Network biology: understanding the cell's functional organization. **Nature reviews genetics**, v. 5, n. 2, p. 101-113, 2004.

BELITSKY, B. R.; GUSTAFSSON, M. C. U.; SONEHSHEIN, A. L.; VON WACHENFELDT, C. An lrp-like gene of *Bacillus subtilis* involved in branched-chain amino acid transport. **Journal of bacteriology**, v. 179, n. 17, p. 5448-5457, 1997.

BENJAMINI, Y.; HOCHBERG, Y.. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society. Series B (Methodological)**, p. 289-300, 1995.

BENSON, D. A.; CAVANAUGH, M.; CLARK, K.; KARSCH-MIZRACHI, I.; LIPMAN, D. J.; OSTELL, J.; SAYERS E. W.; GenBank. **Nucleic acids research**, v. 41, n. D1, p. D36-D42, 2013.

BERENSON, C. S.; MURPHY, T. F.; WRONA, C. T.; SETHI, S. Outer membrane protein P6 of nontypeable *Haemophilus influenzae* is a potent and selective inducer of human macrophage proinflammatory cytokines. **Infection and immunity**, v. 73, n. 5, p. 2728-2735, 2005.

BERGSTROM, C. T.; McELHANY, P.; REAL, L. A. Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 9, p. 5095–5100, 1999.

BHAGWAT, A. A.; YING, Z. I.; KARNS, J.; SMITH, A. Determining RNA quality for NextGen sequencing: some exceptions to the gold standard rule of 23S to 16S rRNA ratio §. **Microbiology Discovery**, v. 1, n. 1, p. 10, 2013.

BLAIR, J. M. A.; RICHMOND, G. E.; BAILEY, A; M.; IVENS, A.; PIDDOCK, L. J. Choice of bacterial growth medium alters the transcriptome and phenotype of *Salmonella enterica* Serovar Typhimurium. **PloS one**, v. 8, n. 5, p. e63912, 2013.

BOOY, R.; HOGSON, S.; MOXON, E. R.; MAYON-WHITE, R. T.; SLACK, M. P. E.; MACFARLANE, J. A.; HAWORTH, E. A.; KIDDLE, M.; SHRIBMAN, S. Efficacy of *Haemophilus influenzae* type b conjugate vaccine PRP-T. **The Lancet**, v. 344, n. 8919, p. 362-366, 1994.

BOURGON, R.; GENTLEMAN, R.; HUBER, W. Independent filtering increases detection power for high-throughput experiments. **Proceedings of the National Academy of Sciences**, v. 107, n. 21, p. 9546-9551, 2010.

BRADLEY, J.; FRANCI, J.; WHEATLEY, T. Liver Abscess due to *Haemophilus influenzae* Type b. **European Journal of Clinical Microbiology**, v. 6, n. 2, p. 76-77, 1987.

BRIERE, E.; HAJJEH, R.; CLARK, A.; SANDERSON, C. **Age at Hib disease, and the implications for different schedules**. Acessível em: http://www.who.int/immunization/sage/meetings/2012/november/HIB_sanderson.pdf. Acesso em: 12/03/2014.

BRIERE, E. C.; RUBIN, L.; MORO, P. L.; COHN, A.; CLARK, T.; MESSONIER, N. Prevention and Control of *Haemophilus influenzae* Type b Disease: Recommendations of the Advisory Committee on Immunization Practices (ACIP). **MMWR. Recommendations and reports: Morbidity and mortality weekly report. Recommendations and reports/Centers for Disease Control**, v. 63, n. RR-01, p. 1-14, 2014.

BRIGE, A.; COLE, J. A.; HAGEN, W. R.; GUISEZ, Y.; VAN BEEUMEN, J. J. Overproduction, purification and novel redox properties of the dihaem cytochrome c, NapB, from *Haemophilus influenzae*. **Biochemical Journal**, v. 356, n. 3, p. 851-858, 2001.

BRÖKER, M.; CONSTANTINO, P.; DeTORA, L.; McINTOSH, D.; RAPPUOLI, R. Biochemical and biological characteristics of cross-reacting material 197 (CRM 197), a non-toxic mutant of diphtheria toxin: use as a conjugation protein in vaccines and other potential clinical applications. **Biologicals**, v. 39, n. 4, p. 195-204, 2011.

BRUANT, G.; GOUSSET, N.; QUENTIN, R.; ROSENAU, A. Fimbrial *ghf* gene cluster of genital strains of *Haemophilus* spp. **Infection and Immunity**. v. 70, n. 10, p. 5438–5445, 2002.

BRUNHAM, R. C.; PLUMMER, F. A.; STEPHENS, R. S. Bacterial anti-genic variation, host immune response, and pathogen-host coevolution. **Infection and Immunity**, v. 61, n. 6, p. 2273–2276, 1993.

BULLARD, J. H.; PURDOM, E.; HANSEN, K. D.; DUDOIT, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. **BMC bioinformatics**, v. 11, n. 1, p. 1, 2010.

CENTERS FOR DISEASE CONTROL Polysaccharide vaccine for prevention of *Haemophilus influenzae* type b disease. **MMWR. Morbidity and mortality weekly report**, v. 34, n. 15, p. 201, 1985.

CERQUETTI, M.; CARDINES, R.; GIUFRE, M.; CASTELLA, A.; REBORA, M.; MASTRANTONIO, P.; ATTI, M. L. C. D. Detection of six copies of the capsulation b locus in a *Haemophilus influenzae* type b strain isolated from a splenectomized patient with fulminant septic shock. **Journal of clinical microbiology**, v. 44, n. 2, p. 640-642, 2006.

CHANDRA, G.; CHATER, K. F.; BORNEMANN, S. Unexpected and widespread connections between bacterial glycogen and trehalose metabolism. **Microbiology**, v. 157, n. 6, p. 1565-1572, 2011.

CHANG, D.-E.; SHIN, S.; RHEE, J.-S.; PAN, J.-G. Acetate metabolism in a pta mutant of *Escherichia coli* w3110: Importance of maintaining acetyl coenzyme a flux for growth and survival. **Journal of bacteriology**, v. 181, n. 21, p. 6656-6663, 1999.

CHEN, Y.; LUN, A. TL; SMYTH, G. K. Differential Expression Analysis of Complex RNA-seq Experiments Using edgeR, in DATTA, S.; NETTLETON, D. S. **Statistical Analysis of Next Generation Sequence Data** New York: Springer, 2014a.

CHEN, Y.; MCCARTHY, D.; ROBINSON, M.; SMYTH, G. K. edgeR: differential expression analysis of digital gene expression data User's Guide. Disponível em: https://www.genomatix.de/online_help/help_regionminer/edgeR.pdf, 2014b. Acesso em: 16/05/2016.

CHIU, H.-J.; ABDUBEK, T.; ASTAKHOVA, H. L.; AXELROD, D.; CARLTON, T.; CLAYTON, D.; DAS, M. C.; DELLER, L.; DUAN, J.; FEUERHELM, J. C.; GRANT, A.; GRZECHNIK, G. W.; HAN, L.; JAROSZEWSKI, K. K.; JIN, H. E.; KLOCK, M. W.; KNUTH, P.; KOZBIAL, S. S.; KRISHNA, A.; KUMAR, D.; MARCIANO, D.; McCULLAN, M. D.; MILLER, A. T.; MORSE, E.; NIGOGHOSSIAN, L.; OKACH, R.; REYES, H. J.; TIEN, C. B.; TRAME, H. VAN DEN BEDEM, D.; WEEKES, Q.; XU, K. O.; HODGSON, J.; WOOLEY, M.-A.; ELSLIGER, A. M.; DEACON, A.; GODZIK, S. A.; LESLEY, S. A.; WILSON, I. A. The structure of *Haemophilus influenzae* prephenate dehydrogenase suggests unique features of bifunctional TyrA enzymes. **Acta Crystallographica Section F: Structural Biology and Crystallization Communications**, v. 66, n. 10, p. 1317-1325, 2010.

CINTRA, F. O.; TAKAGI, M. Comparison among different sample treatment methods for analysis of molecular weight and concentration of exopolysaccharide produced by *Haemophilus influenzae* type b. In: VILAS, A. M-. (ed) **Microbes in Applied Research. Current Advances and Challenges** v. 24, p. 513-517. Malaga: Formatex, 2012.

CINTRA, F. O.; TAKAGI, M. Study of chemical stability of the capsular polysaccharide produced by *Haemophilus influenzae* type b. **Carbohydrate Polymers**, 10.1016/j.carbpol.2014.04.004, 2014.

CINTRA, F. O. *Modelagem matemática e otimização da produção de exopolissacarídeo de Haemophilus influenzae tipo b*. 2014. 110 f. Dissertação (Mestrado em Biotecnologia) – Instituto Butantan, São Paulo. 2014.

CONESA, A.; TARAZONA, S.; GOMEZ-CABRERO, D.; CERVERA, A.; McPHERSON, A.; SZCZEŚNIAK, W.; GAFFNEY, D. J.; ELO, L. L.; MORTAVAZI, A. A survey of best practices for RNA-seq data analysis. **Genome biology**, v. 17, n. 1, p. 1, 2016.

COREY, R. A.; ALLEN, W. J.; KOMAR, J.; MASIULIS, S.; MENZIES, S.; ROBSON, A.; COLLISON, I. Unlocking the bacterial SecY translocon. **Structure**, v. 24, n. 4, p. 518-527, 2016.

CORN, P. G.; ANDERS, J.; TAKALA, A. K.; KAYHTY, H.; HOISETH, S. K.. Genes involved in *Haemophilus influenzae* type b capsule expression are frequently amplified. **Journal of Infectious Diseases**, v. 167, n. 2, p. 356–364, 1993.

COTTER, S. E.; YEO, H. J.; JUEHNE, T.; ST GEME III, J. W. Architecture and adhesive activity of the *Haemophilus influenzae* Hsf adhesin. **Journal of bacteriology**, v. 187, n. 13, p. 4656-4664, 2005.

COUREUIL, M.; LÉCUYER, H.; BOURDOULOS, S.; NASSIF, X. A journey into the brain: insight into how bacterial pathogens cross blood-brain barriers. **Nature Reviews Microbiology**, v. 15, n. 3, p. 149-159, 2017.

COX, D. R.; REID, N. Parameter orthogonality and approximate conditional inference. **Journal of the Royal Statistical Society. Series B (Methodological)**, p. 1-39, 1987.

CRAIG, J. E.; NOBBS, A.; HIGH, N. J. The Extracytoplasmic Sigma Factor, σ_E , Is Required for Intracellular Survival of Nontypeable *Haemophilus influenzae* in J774 Macrophages. **Infection and immunity**, v. 70, n. 2, p. 708-715, 2002.

CRAIG, N. L.; CHANDLER, M.; GELLERT, M.; LAMBOWITZ, A.; RICE, P. A.; SANDMEYER, S. (Eds.). **Mobile DNA III**. Washington: ASM Press, 2015.

CREECY, J. P.; CONWAY, T. Quantitative bacterial transcriptomics with RNA-seq. **Current opinion in microbiology**, v. 23, p. 133-140, 2015.

CRISEL, R. M.; BAKER, R. S.; DORMAN, D. E. Capsular polymer of *Haemophilus influenzae* type b. I. Structural characterization of the capsular polymer of strain Eagan. **The Journal of Biological Chemistry**, v. 250, n. 13, p. 4926-4930, 1975.

CROUCHER, N. J.; THOMSON, N. R. Studying bacterial transcriptomes using RNA-seq. **Current opinion in microbiology**, v. 13, n. 5, p. 619-624, 2010.

CUNIN, R.; GLANSDORFF, N.; PIÉRARD, A.; STALON, V. Biosynthesis and metabolism of arginine in bacteria. **Microbiological reviews**, v. 50, n. 3, p. 314, 1986.

DAJANI, A.; ASMAR, B. I.; THIRUMOORTHY, M. C. Systemic *Haemophilus influenzae* disease: An Overview. **The Journal of Pediatrics**, v. 94, n. 3, p. 355–364, 1979.

DE SOUZA-HART, J. A.; BLADSTOCK, W.; DI MODUGNO, V.; HOLLAND, I. B.; KOK, M. Two-component systems in *Haemophilus influenzae*: a regulatory role for ArcA in serum resistance. **Infection and immunity**, v. 71, n. 1, p. 163-172, 2003.

DECKER, M. D.; EDWARDS, K. M.; BRADLEY, R.; PALMER, P. Comparative trial in infants of four conjugate *Haemophilus influenzae* type b vaccines. **The Journal of pediatrics**, v. 120, n. 2, p. 184-189, 1992.

DENG, W.; LI, C.; XIE, J. The underlying mechanism of bacterial TetR/AcrR family transcriptional repressors. **Cellular signalling**, v. 25, n. 7, p. 1608-1613, 2013.

DEUTSCHER, J.; FRANCKE, C.; POSTMA, P. W. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. **Microbiology and Molecular Biology Reviews**, v. 70, n. 4, p. 939-1031, 2006.

DI BERARDINO, M.; DIJKSTRA, A.; STÜBER, D.; KECK, W.; GUBLER, M. The monofunctional glycosyltransferase of *Escherichia coli* is a member of a new class of peptidoglycan-synthesising enzymes: overexpression and determination of the glycan-polymerising activity. **FEBS letters**, v. 392, n. 2, p. 184-188, 1996.

DIK, D. A.; MAROUS, D. R.; FISHER, J. F.; MOBAHERRY, S. Lytic transglycosylases: concinnity in concision of the bacterial cell wall. **Critical Reviews in Biochemistry and Molecular Biology**, p. 1-40, 2017.

DILLIES, M.-A.; RAU, A.; AUBERT, J.; HENNEQUET-ANTIER, C.; JEANMOUGIN, M.; SERVANT, N.; KEIME, C.; MAROT, G.; CASTEL, D.; ESTELLE, J.; GUERNEC, G.; JAGLA, B.; JOUNEAU, L.; LALOË, D.; GALL, C. L.; SCHAËFFER, B.; CROM, S. L.; GUEDJ, M.; JAFFRÉZIC, F. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. **Briefings in bioinformatics**, v. 14, n. 6, p. 671-683, 2013.

DINTZIS, R. Z. Rational design of conjugate vaccines. **Pediatric research**, v. 32, n. 4, p. 376-385, 1992.

DOELLE, H. W.; EWINGS, K. N.; HOLLYWOOD, N. W. Regulation of glucose metabolism in bacterial systems. In: **Microbial Reactions**. Springer, Berlin, Heidelberg, 1982. p. 1-35.

DONAHUE, J. L.; BOWNAS, J. L.; NIEHAUS, W. G.; LARSON, T. J. Purification and Characterization of glpX-Encoded Fructose 1, 6-Bisphosphatase, a New Enzyme of the Glycerol 3-Phosphate Regulon of *Escherichia coli*. **Journal of bacteriology**, v. 182, n. 19, p. 5624-5627, 2000.

DUIM, B.; VOGEL, L.; PUIJK, W.; JANSEN, H. M.; MELOEN, R. H.; DANKERT, J.; VAN ALPHEN, L. Fine mapping of outer membrane protein P2 antigenic sites which vary during persistent infection by *Haemophilus influenzae*. **Infection and Immunity**. v. 64, n. 11, p. 4673–4679, 1996.

EARL, C. S.; KEONG, T. W.; AN, S.-q.; MURDOCH, S.; MCCARTHY, Y.; GARMENDIA, J.; WARD, J.; DOW, J. M.; YANG, L.; O'TOOLE, G. A.; RYAN, R. P. *Haemophilus influenzae* responds to glucocorticoids used in asthma therapy by modulation of biofilm formation and antibiotic resistance. **EMBO molecular medicine**, p. e201505088, 2015.

ECEVIT, I. Z.; McCREA, K. W.; PETTIGREW, M. M.; SEM, A.; MARRS, C. F.; GILSDORF, J. R. Prevalence of the *hifBC*, *hmw1A*, *hmw2A*, *hmwC*, and *hia* genes in *Haemophilus influenzae* isolates. **Journal of Clinical Microbiology**, v. 42, n. 7, p. 3065-3072, 2004.

EDWARDS, J. S.; PALSSON, B. O. Systems Properties of the *Haemophilus influenzae* Rd Metabolic Genotype. **Journal of Biological Chemistry**, v. 274, n. 25, p. 17410–17416, 1999.

EINSLE, O.; MESSERSCHMIDT, A.; STACH, P.; BOURENKOV, G. P.; BARTUNIK, H. D., HUBER, R.; KRONECK, P. M. Structure of cytochrome c nitrite reductase. **Nature**, v. 400, n. 6743, p. 476, 1999.

EISEN, M. B.; SPELLMAN, P. T.; BROWN, P. O.; BOTSTEIN, D. Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences**, v. 95, n. 25, p. 14863-14868, 1998.

ELY, S.; TIPPETT, J.; KROLL, J. S.; MOXON, E. R. Mutations affecting expression and maintenance of genes encoding the serotype b capsule of *Haemophilus influenzae*. **Journal of bacteriology**, v. 167, n. 1, p. 44-48, 1986.

ENGLISH, B. P.; HAURYLIUKA, V.; SANAMRADA, A.; TANKOVA, S.; DEKKERA, N. H.; ELFA, J. Single-molecule investigations of the stringent response machinery in living bacterial cells. **Proceedings of the National Academy of Sciences**, v. 108, n. 31, p. E365-E373, 2011.

ERWIN, A. L.; SMITH, A. L. Nontypeable *Haemophilus influenzae*: understanding virulence and commensal behavior. **Trends in Microbiology**, v. 15, n. 8, p. 355–62, 2007.

ERWIN, A. L.; SANDSTEDT, S. A.; BONTHUIS, P. J.; GEELHOOD, J. L.; NELSON, K. L.; UNRATH, W. C. T.; DIGGLE, M. A.; THEODORE, M. J.; PLEATMAN, C. R.; MOTHERSHED, E. A.; SACCHI, C. T.; MAYER, L. W.; GILSDORF, J. R.; SMITH, A. L. Analysis of genetic relatedness of *Haemophilus influenzae* isolates by multilocus sequence typing. **Journal of bacteriology**, v. 190, n. 4, p. 1473-1483, 2008.

EVANS, N.M.; SMITH, D. D.; WICKEN, A. J. Haemin and nicotinamide adenine dinucleotide requirements of *Haemophilus influenzae* and *Haemophilus parainfluenzae*. **Journal of medical microbiology**, v. 7, n. 3, p. 359-365, 1974.

EVERS, S.; DI PADOVA, K.; MEYER, M.; LANGEN, H.; FOUNTOULAKIS, M.; KECK, W.; GRAY, C. P. Mechanism- related changes in the gene transcription and protein synthesis patterns of *Haemophilus influenzae* after treatment with transcriptional and translational inhibitors. **Proteomics**, v. 1, n. 4, p. 522-544, 2001.

EVGUENIEVA- HACKENBERG, E. Bacterial ribosomal RNA in pieces. **Molecular microbiology**, v. 57, n. 2, p. 318-325, 2005.

EWING, B.; HILLIER, L. D.; WENDL, M. C.; GREEN, P. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. **Genome research**, v. 8, n. 3, p. 175-185, 1998.

EZRATY, B. GENNARIS, A.; BARRAS, F.; COLLET, J.-F. Oxidative stress, protein damage and repair in bacteria. **Nature Reviews Microbiology**, 2017.

FANG, Z.; CUI, X. Design and validation issues in RNA-seq experiments. **Briefings in bioinformatics**, p. bbr004, 2011.

FARMER, William R.; LIAO, James C. Reduction of aerobic acetate production by *Escherichia coli*. **Applied and environmental microbiology**, v. 63, n. 8, p. 3205-3210, 1997.

FILDES, P. The nature of the effect of blood-pigment upon the growth of *B. influenzae*. **British Journal of Experimental Pathology**, v. 2, n. 1, p. 16-25, 1921.

FILIATRAULT, M. J. Progress in prokaryotic transcriptomics. **Current opinion in microbiology**, v. 14, n. 5, p. 579-586, 2011.

FINN, A. Bacterial polysaccharide–protein conjugate vaccines. **British medical bulletin**, v. 70, n. 1, p. 1-14, 2004.

FLEISCHMANN, R. D.; ADAMS, M. D.; WHITE, O.; CLAYTON, R. A.; KIRKNESS, E. F.; KERLAVAGE, A. R.; BULT, C. J.; TOMB, J. F.; DOUGHERTY, B. A.; MERRICK, J. M.; McKENNEY, K.; SUTTON, G.; FITZHUGH, W.; FIELDS, C.; GOCAYNE, J. D.; SCOTT, J.; SHIRLEY, R.; LIU, L.-I.; GLODEK, A.; KELLEY, J. M.; WEIDMAN, J. F.; PHILLIPS, C. A.; SPRIGGS, T.; HEDBLUM, E.; COTTON, M. D.; UTTERBACK, T. R.; HANNA, M. C.; NGUYEN, D. T.; SAUDEK, D. M.; BRANDON, R. C.; FINE, L. D.; FRITCHMAN, J. L.; FUHRMANN, J. L.; GEOGHAGEN, N. S. M.; GNEHM, C. L.; McDONALD, L. A., K. V.; FRASER, C. M.; SMITH, H. O.; VENTER, J. C. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. **Science**, v. 269, n. 5223, p. 496–512, 1995.

FOLLENS, A.; VIGA-DA-CUNHA, M.; MERCKX, R.; VAN SCHAFTINGEN, E.; VAN ELDERE, J. *acs1* of *Haemophilus influenzae* type a capsulation locus region II encodes a bifunctional ribulose 5-phosphate reductase–CDP-ribitol pyrophosphorylase. **Journal of bacteriology**, v. 181, n. 7, p. 2001-2007, 1999.

FORTNUM, H. M. Hearing impairment after bacterial meningitis: A review. **Archives of Disease in Childhood**, v. 67, n. 9, p. 1128-1133, 1992.

GARAVAGLIA, S.; BRUZZONE, S.; CASSANI, C.; CANELLA, L.; ALLEGRONE, G.; STURLA, L.; MANNINO, E.; MILLO, E.; FLORA, A. D.; RIZZI, M. The high-resolution crystal structure of periplasmic *Haemophilus influenzae* NAD nucleotidase reveals a novel enzymatic function of human CD73 related to NAD metabolism. **Biochemical Journal**, v. 441, n. 1, p. 131-141, 2012.

GARBER, M.; GRABHERR, M. G.; GUTTMAN, M.; TRAPNELL, C. Computational methods for transcriptome annotation and quantification using RNA-seq. **Nature methods**, v. 8, n. 6, p. 469–477, 2011.

GENTLEMAN, R. C.; CAREY, V. J.; BATES, B. M.; BOLSTAD, B.; DETTLING, M.; DUDOIT, S.; ELLIS, B.; GAUTIER, L.; GE, Y.; GENTRY, J.; HORNIK, K.; HOTBORN, T.; HUBER, W.; IACUS, S.; IRIZARRY, R.; LEISCH, F.; LI, C.; MAECHLER, M.; ROSSINI, A. J.; SAWITZKI, G.; SMITH, C.; SMYTH, G.; TIERNEY, L.; YANG, J. Y. H.; ZHANG, J. Bioconductor: open software development for computational biology and bioinformatics. **Genome biology**, v. 5, n. 10, p. R80, 2004.

GILSDORF, J. R. Antigenic Diversity and Gene Polymorphisms in *Haemophilus influenza*. **Infection and Immunity**, v. 66, n. 11, 1998.

GILSDORF, J. R.; MARRS, C. F.; FOXMAN, B. *Haemophilus influenzae*: Genetic Variability and Natural Selection To Identify Virulence Factors. **Infection and Immunity**, v. 72, n. 5, p. 2457-2461, 2004.

GIUFRE', M.; CARDINES, R.; MASTRANTONIO, P.; CERQUETT, M. Genetic characterization of the capsulation locus of *Haemophilus influenzae* serotype e. **Journal of clinical microbiology**, v. 48, n. 4, p. 1404-1407, 2010.

GODOY, M. S.; NIKEL, P. I.; GOMEZ, J. G. C. ; PETTINARIA, M. J. The CreC regulator of *Escherichia coli*, a new target for metabolic manipulations. **Applied and environmental microbiology**, v. 82, n. 1, p. 244-254, 2016.

GOKARN, R.; EVANS, J.; WALKER, J.; MARTIN, S.; EITEMAN, M.; ALTMAN, E. The physiological effects and metabolic alterations caused by the expression of *Rhizobium etli* pyruvate carboxylase in *Escherichia coli*. **Applied microbiology and biotechnology**, v. 56, n. 1-2, p. 188-195, 2001.

GRABHERR, M. G.; HAAS, B. J.; YASSOUR, M.; LEVIN, J. Z.; THOMPSON, D. A.; AMIT, I.; REGEV, A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nature Biotechnology**, v. 29, n. 7, p. 644-52, 2011.

GRANOFF, D. M.; BOIES, E. G.; MUNSON, R. S. Immunogenicity of *Haemophilus influenzae* type b polysaccharide—diphtheria toxoid conjugate vaccine in adults. **The Journal of pediatrics**, v. 105, n. 1, p. 22-27, 1984.

GREAVES, R. I. N. Preservation of living cells by freeze-drying. **Annals of the New York Academy of Sciences**, v. 85, n. 2, p. 723-728, 1960.

GMUENDER, H.; KURATLI, K.; PADOVA, D. D.; GRAY, C. P.; KECK, W.; EVERS, S.; Gene expression changes triggered by exposure of *Haemophilus influenzae* to novobiocin or ciprofloxacin: combined transcription and translation analysis. **Genome Research**, v. 11, n. 1, p. 28-42, 2001.

HAAN, A.; PUT, R. M. F.; BEURRET, M. HPAEC- PAD method for the analysis of alkaline hydrolyzates of *Haemophilus influenzae* type b capsular polysaccharide. **Biomedical Chromatography**, v. 27, n. 9, p. 1137-1142, 2013.

HAAS, B. J.; CHIN, M.; NUSBAUM, C.; BIRREN, B. W.; LIVNY, J. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes?. **BMC genomics**, v. 13, n. 1, p. 1, 2012.

HALLSTRÖM, T.; RIESBECK, K.. *Haemophilus influenzae* and the complement system. **Trends in microbiology**, v. 18, n. 6, p. 258-265, 2010.

HAROLD, F. M. Inorganic polyphosphates in biology: structure, metabolism, and function. **Bacteriological Reviews**, v. 30, n. 4, p. 772, 1966.

HARRISON, A.; SANTANA, E. A.; SZELESTEY, B. R.; NEWSOM, D. E.; HITE, P; MASON, K. M. Ferric uptake regulator and its role in the pathogenesis of nontypeable *Haemophilus influenzae*. **Infection and immunity**, v. 81, n. 4, p. 1221-1233, 2013.

HEATWOLE, V. M.; SOMERVILLE, R. L. Cloning, nucleotide sequence, and characterization of mtr, the structural gene for a tryptophan-specific permease of *Escherichia coli* K-12. **Journal of bacteriology**, v. 173, n. 1, p. 108-115, 1991.

HEIMANN, J. D. The extracytoplasmic function (ECF) sigma factors. **Advances in microbial physiology**, v. 46, p. 47-110, 2002.

HERBERT, M. A.; HAYES, S.; DEADMAN, M. E., C. TANG, M.; HOOD D. W.; MOXON E. R. Signature tagged mutagenesis of *Haemophilus influenzae* identifies genes required for *in vivo* survival. **Microbial Pathogenesis**, v. 33, n. 5, p. 211–223, 2002.

HERRIOTT, R. M., MEYER, E. Y.; VOGT, M.; MODAN, M. Defined medium for growth of *Haemophilus influenzae*. **Journal of bacteriology**, v. 101, n. 2, p. 513-516, 1970.

HIBI, M.; YUKITOMO, H.; ITO, M.; MORI, H. Improvement of NADPH-dependent bioconversion by transcriptome-based molecular breeding. **Applied and environmental microbiology**, v. 73, n. 23, p. 7657-7663, 2007.

HISS, H. Cinética de Processos Fermentativos. In: LIMA, U. de A.; SCHMIDELL, W.; AQUARONE, E.; BORZANI, W.; **Biotecnologia industrial: Engenharia Química**, v. 2. São Paulo: Edgard Blücher, 2001. Cap. 6, p.93-122.

HO, D. K.; RAM, S.; NELSON, K. L.; BONTHUIS, P. J.; SMITH, A. L. *lgtC* expression modulates resistance to C4b deposition on an invasive nontypeable *Haemophilus influenzae*. **The Journal of Immunology**, v. 178, n. 2, p. 1002-1012, 2007.

HOAGLAND, C. L.; BEESON, P. B.; GOEBEL, W. F. The capsular polysaccharide of the type XIV pneumococcus and its relationship to the specific substances of human blood. **Science**, v. 88, n. 2281, p. 261-263, 1938.

HOEHN, K. B.; FOWLER, A.; LUNTER, G.; PYBUS, O. G. The diversity and molecular evolution of B cell receptors during infection. **Molecular biology and evolution**, p. msw015, 2016.

HOISETH, S. K.; CONNELLY, C. J.; MOXON, E. R. Genetics of spontaneous, high-frequency loss of b capsule expression in *Haemophilus influenzae*. **Infection and Immunity**, v. 49, n. 2, p. 389–395, 1985.

HOISETH, S. K.; MOXON, E. R.; SILVER, R. P. Genes involved in *Haemophilus influenzae* type b capsule expression are part of an 18-kilobase tandem duplication. **Proceedings of the National Academy of Sciences**, v. 83, n. 4, p. 1106-1110, 1986.

HOLDEN, H. M.; RAYMENT, I.; THODEN, J. B. Structure and function of enzymes of the Leloir pathway for galactose metabolism. **Journal of Biological Chemistry**, v. 278, n. 45, p. 43885-43888, 2003.

HOOD, D. W.; COX, A. D.; GILBERT, M.; MAKEPEACE, K.; WALSH, S.; DEADMAN, M. E.; CODY, A.; MARTIN, A.; MÅNSSON, M.; SCHWEDA, E. K.; BRISSON, J. R.; RICHARDS, J. C.; MOXON, E. R.; WAKARCHUK, W. W. Identification of a lipopolysaccharide α - 2, 3- sialyltransferase from *Haemophilus influenzae*. **Molecular microbiology**, v. 39, n. 2, p. 341-351, 2001.

HORNIK, K. The comprehensive R archive network. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 4, n. 4, p. 394-398, 2012.

HOSIE, A. H. F.; POOLE, P. S. Bacterial ABC transporters of amino acids. **Research in microbiology**, v. 152, n. 3, p. 259-270, 2001.

HOSKING, S. L.; CRAIG, J. E.; HIGH, N. J. Phase variation of lic1A, lic2A and lic3A in colonization of the nasopharynx, bloodstream and cerebrospinal fluid by *Haemophilus influenzae* type b. **Microbiology**, v. 145, n. 11, p. 3005-3011, 1999.

HU, Y.; JIANG, F.; GUO, Y.; SHEN, X.; ZHANG, Y.; ZHANG, R.; GUO, G.; MAO, X.; ZOU, Q.; WANG, D.-C. Crystal structure of HugZ, a novel heme oxygenase from *Helicobacter pylori*. **Journal of Biological Chemistry**, v. 286, n. 2, p. 1537-1544, 2011.

HUDSON, G. S.; DAVIDSON, B. E. Nucleotide sequence and transcription of the phenylalanine and tyrosine operons of *Escherichia coli* K12. **Journal of molecular biology**, v. 180, n. 4, p. 1023-1051, 1984.

HUMPHRIES, H. E.; HIGH, N. J. The role of *licA* phase variation in the pathogenesis of invasive disease by *Haemophilus influenzae* type b. **FEMS Immunology & Medical Microbiology**, v. 34, n. 3, p. 221-230, 2002.

IANCU, O. D.; KAWANE, S.; BOTTOMLY, D.; SEARLES, R.; HITZEMANN, R.; McWEENEY, S. Utilizing RNA-Seq data for *de novo* coexpression network inference. **Bioinformatics**, v. 28, n. 12, p. 1592-1597, 2012.

IGARASHI, K.; KASHIWAGI, K. Polyamine transport in bacteria and yeast. **Biochemical Journal**, v. 344, n. 3, p. 633-642, 1999.

ISHAK, N.; TIKHOMIROVA, A.; BENT, S. J.; EHRLICH, G.; HU, F. Z.; KIDD, S. P. There is a specific response to pH by isolates of *Haemophilus influenzae* and this has a direct influence on biofilm formation. **BMC microbiology**, v. 14, n. 1, p. 1, 2014.

JENKINS, G. S.; CHANDLER, M. S.; FINK, P. S. Functional characterization of the *Haemophilus influenzae* 4.5 S RNA. **Canadian journal of microbiology**, v. 44, n. 1, p. 91-94, 1998.

JIANG, D.; TIKHOMIROVA, A.; KIDD, S. P. *Haemophilus influenzae* strains possess variations in the global transcriptional profile in response to oxygen levels and this influences sensitivity to environmental stresses. **Research in microbiology**, v. 167, n. 1, p. 13-19, 2016. a

JIANG, D.; TIKHOMIROVA, A.; BENT, S. J.; KIDD, S. P. A discrete role for FNR in the transcriptional response to moderate changes in oxygen by *Haemophilus influenzae* Rd KW20. **Research in microbiology**, v. 167, n. 2, p. 103-113, 2016. b

JORGENSEN, M. The bacterial septal ring protein RlpA is a lytic transglycosylase that contributes to rod shape and daughter cell separation in *Pseudomonas aeruginosa*. **Molecular microbiology**, v. 93, n. 1, p. 113-128, 2014.

KADOKURA, H.; KATZEN, F.; BECKWITH, Jon. Protein disulfide bond formation in prokaryotes. **Annual review of biochemistry**, v. 72, n. 1, p. 111-135, 2003.

KANEHISA, M.; GOTO, S. KEGG: kyoto encyclopedia of genes and genomes. **Nucleic acids research**, v. 28, n. 1, p. 27-30, 2000.

KAWAI, S.; MURATA, K. Structure and function of NAD kinase and NADP phosphatase: key enzymes that regulate the intracellular balance of NAD (H) and NADP (H). **Bioscience, biotechnology, and biochemistry**, v. 72, n. 4, p. 919-930, 2008.

KAWAI, F.; CLARKE, T. B.; ROPER, D. I.; HAN, G.-J.; HWANG, K. Y.; UNZAI, S.; OBAYASHI, E.; PARK, S.-Y.; TAME, J. R. H. Crystal structures of penicillin-binding proteins 4 and 5 from *Haemophilus influenzae*. **Journal of molecular biology**, v. 396, n. 3, p. 634-645, 2010.

KÄYHTY, H.; KARANKO, V.; PELTOLA, H.; MAKELA, P. H. Serum antibodies after vaccination with *Haemophilus influenzae* type b capsular polysaccharide and responses to reimmunization: no evidence of immunologic tolerance or memory. **Pediatrics**, v. 74, n. 5, p. 857-865, 1984.

KARAS, V. O.; WESTERLAKEN, I.; MEYER, A. S. The DNA-binding protein from starved cells (Dps) utilizes dual functions to defend cells against multiple stresses. **Journal of bacteriology**, v. 197, n. 19, p. 3206-3215, 2015.

KHEMICI, V.; LINDER, P. RNA helicases in bacteria. **Current opinion in microbiology**, v. 30, p. 58-66, 2016.

KILLIAN, M. *Haemophilus*. in: STALEY, J. T.; KRIEG, N. R.; BRENNER, D. J. **Bergey's Manual of Systematic Bacteriology: Volume 2: The Proteobacteria, Part B: The Gammaproteobacteria**. East Lansing: SPRINGER 2005.

KIMURA, A. L. A. N.; HANSEN, E. J. Antigenic and phenotypic variations of *Haemophilus influenzae* type b lipopolysaccharide and their relationship to virulence. **Infection and immunity**, v. 51, n. 1, p. 69-79, 1986.

KLEANTHOUS, C.; WALKER, D. Immunity proteins: enzyme inhibitors that avoid the active site. **Trends in biochemical sciences**, v. 26, n. 10, p. 624-631, 2001.

KOLKER, E.; MARAKOVA, K. S.; SHABALINA, S.; PICONE, A. F.; PURVINE, S.; HOLZMAN, T.; CHERNY, T.; ARMBUSTER, D.; MUNSON Jr, R. S.; KOLESOV, G., FRISHMAN, D.; GALPERIN, M. Y. Identification and functional analysis of 'hypothetical' genes expressed in *Haemophilus influenzae*. **Nucleic acids research**, v. 32, n. 8, p. 2353-2361, 2004.

KOO, J. T.; CHOE, J.; MOSELEY, S. L. HrpA, a DEAH- box RNA helicase, is involved in mRNA processing of a fimbrial operon in *Escherichia coli*. **Molecular microbiology**, v. 52, n. 6, p. 1813-1826, 2004.

KORAIMANN, G. Lytic transglycosylases in macromolecular transport systems of Gram-negative bacteria. **Cellular and molecular life sciences**, v. 60, n. 11, p. 2371-2388, 2003.

KRANZ, R. G.; RICHARD-FOGAL, C.; TAYLOR, J.-S.; FRAWLEY, E. R. Cytochrome c biogenesis: mechanisms for covalent modifications and trafficking of heme and for heme-iron redox control. **Microbiology and molecular biology reviews**, v. 73, n. 3, p. 510-528, 2009.

KROLL, J. S.; MOXON, E. R. Capsulation and gene copy number at the *cap* locus of *Haemophilus influenzae* type b. **Journal of bacteriology**, v. 170, n. 2, p. 859-864, 1988.

KROLL, J. S.; HOPKINS, I.; MOXON, E. R.. Capsule loss in *H. influenzae* type b occurs by recombination-mediated disruption of a gene essential for polysaccharide export. **Cell**, v. 53, n. 3, p. 347-356, 1988.

KROLL, J. S.; ZAMZE, S.; LOYNDS, B.; MOXON, E. R.. Common organization of chromosomal loci for production of different capsular polysaccharides in *Haemophilus influenzae*. **Journal of Bacteriology**, v. 171, n. 6, p. 3343–3347, 1989.

KROLL, J. S.; LOYNDS, B.; BROPHY, L. N.; MOXON, E. R. The *bex* locus in encapsulated *Haemophilus influenzae*: a chromosomal region involved in capsule polysaccharide export. **Molecular Microbiology**, v. 4, n. 11, p. 1853–1862, 1990.

KROLL, J. S.; LOYNDS, B. M.; MOXON, E. R. The *Haemophilus influenzae* capsulation gene cluster: a compound transposon. **Molecular Microbiology**, v. 5, n. 6, p. 1549–1560, 1991.

KROLL, J. S.; MOXON, E. R.; LOYNDS, B. M. An ancestral mutation enhancing the fitness and increasing the virulence of *Haemophilus influenzae* type b. **Journal of Infectious Diseases**, v. 168, n. 1, p. 172–176, 1993.

KU, M.-J.; YOON, H. J.; AHN, H. J.; KIM, H. W.; BAEK, S. H.; SUH, S. H. Crystallization and preliminary X-ray crystallographic studies of 3-deoxy-manno-octulosonate cytidyltransferase from *Haemophilus influenzae*. **Acta Crystallographica Section D: Biological Crystallography**, v. 59, n. 1, p. 180-182, 2003.

KVAM, V. M.; LIU, P.; SI, Y. A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. **American journal of botany**, v. 99, n. 2, p. 248-256, 2012.

LÂM, T.-T.; CLAUS, H.; FROSCH, M.; VOGEL, U. Sequence analysis of serotype-specific synthesis regions II of *Haemophilus influenzae* serotypes c and d: evidence for common ancestry of capsule synthesis in *Pasteurellaceae* and *Neisseria meningitidis*. **Research in microbiology**, v. 162, n. 5, p. 483-487, 2011.

LANGEN, H.; TAKÁCS, B.; EVERS, S.; BERNDT, P.; LAHM, H.-W.; WIPF, B.; GRAY, C.; FOUNTOULAKIS, M. Two-dimensional map of the proteome of *Haemophilus influenzae*. **ELECTROPHORESIS**, v. 21, n. 2, p. 411–429, 2000.

LANGFELDER, P.; HORVATH, S. WGCNA: an R package for weighted correlation network analysis. **BMC bioinformatics**, v. 9, n. 1, p. 1, 2008.

LANGFELDER, P.; ZHANG, B.; HORVATH, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. **Bioinformatics**, v. 24, n. 5, p. 719-720, 2008.

LANGFELDER, P.; HORVATH, S. Fast R functions for robust correlations and hierarchical clustering. **Journal of statistical software**, v. 46, n. 11, 2012.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature methods**, v. 9, n. 4, p. 357-359, 2012.

LARUE, K. FORD, R. C.; WILLIS, L. M.; WHITFIELD, C. Functional and structural characterization of polysaccharide co-polymerase proteins required for polymer export in ATP-binding cassette transporter-dependent capsule biosynthesis pathways. **Journal of Biological Chemistry**, v. 286, n. 19, p. 16658-16668, 2011.

LAW, C. W.; . CHEN, Y.; SHI, W.; SMYTH, G. K. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. **Genome biology**, v. 15, n. 2, p. R29, 2014.

LAW, C. W.; ALHAMDOOSH, M.; SU, S.; SMYTH, G. K.; RITCHIE, M. E. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. **F1000Research**, v. 5, 2016.

LEE, B. C. Isolation of an outer membrane hemin-binding protein of *Haemophilus influenzae* type b. **Infection and immunity**, v. 60, n. 3, p. 810-816, 1992.

LEEK, J. T.; SCHARPF, R. B.; BRAVO, H. C.; SIMCHA, D.; LANGMEAD, B.; JOHNSON, W. E.; GEMAN, D.; BAGGERLY, K.; IRIZARRY, R. A. Tackling the widespread and critical impact of batch effects in high-throughput data. **Nature Reviews Genetics**, v. 11, n. 10, p. 733-739, 2010.

LEHRER, J.; Functional characterization and membrane topology of *Escherichia coli* WecA, a sugar-phosphate transferase initiating the biosynthesis of enterobacterial common antigen and O-antigen lipopolysaccharide. **Journal of bacteriology**, v. 189, n. 7, p. 2618-2628, 2007.

LEUNG, P. S.; PREISS, J. Biosynthesis of bacterial glycogen: primary structure of *Salmonella typhimurium* ADPglucose synthetase as deduced from the nucleotide sequence of the *glgC* gene. **Journal of bacteriology**, v. 169, n. 9, p. 4355-4360, 1987.

LI, H.; RUAN, J.; DURBIN, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. **Genome research**, v. 18, n. 11, p. 1851-1858, 2008.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R. The sequence alignment/map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078-2079, 2009.

LI, A.; VAN JUJIK, N.; TER BEEK, M.; CASPERS, M.; PUNT, P.; DER WERF, M. A clone-based transcriptomics approach for the identification of genes relevant for itaconic acid production in *Aspergillus*. **Fungal genetics and Biology**, v. 48, n. 6, p. 602-611, 2011.

LIAO, Y.; SMYTH, G. K.; SHI, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. **Nucleic acids research**, v. 41, n. 10, p. e108-e108, 2013.

LIAO, Y.; SMYTH, G. K.; SHI, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. **Bioinformatics**, v. 30, n. 7, p. 923-930, 2014.

LICHT, A.; SCHNEIDER, E. ATP binding cassette systems: structures, mechanisms, and functions. **Open Life Sciences**, v. 6, n. 5, p. 785-801, 2011.

LICHTENEGGER, S.; BINAA, I.; ROIERA, S.; BAUERNFEINDB, S.; KEIDELB, K.; SCHILDA, S.; ANTHONYC, M.; REIDLA, J. Characterization of lactate utilization and its implication on the physiology of *Haemophilus influenzae*. **International Journal of Medical Microbiology**, v. 304, n. 3, p. 490-498, 2014.

LINK, A. J.; HAYS, L. G.; CARMACK, E. B.; YATES, J. R. Identifying the major proteome components of *Haemophilus influenzae* type- strain NCTC 8143. **ELECTROPHORESIS**, v. 18, n. 8, p. 1314-1334, 1997.

LINK, H.; FUHRER, T.; GEROSA, L.; ZAMBONI, N.; SAUER, U. Real-time metabolome profiling of the metabolic switch between starvation and growth. **Nature Methods**, v. 12, n. 11, p. 1091, 2015.

LOGAN, D. T.; MULLIEZ, E.; LARSSON, K. M.; BODEVIN, S.; ATTA, M.; GARNAUD, P. E.; SJOBERG, B. M.; FONTECAVE, M. A metal-binding site in the catalytic subunit of anaerobic ribonucleotide reductase. **Proceedings of the National Academy of Sciences**, v. 100, n. 7, p. 3826-3831, 2003.

LORTHIOIS, A. P. A.; TAKAGI, M. T.; CABRERA-CRESPO, J.; RAW, I.; TANIZAKI, M. M. An improved protocol for the synthesis of *Haemophilus influenzae* type b capsular polysaccharide and tetanus toxoid conjugated antigen. In: **VACCINE**. Kidlington: ELSEVIER, 2008.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome biology**, v. 15, n. 12, p. 550, 2014.

LUNDQVIST, T.; FISHER, S. L.; KERN, G.; FOLMER, R. H. A.; XUE, Y.; NEWTON, D. T.; KEATING, T. A.; ALM, R. A.; DE JONGE, B. L. M. Exploitation of structural and regulatory diversity in glutamate racemases. **Nature**, v. 447, n. 7146, p. 817, 2007.

LWOFF, A.; LWOFF, M. Studies on Codehydrogenases. I--Nature of Growth Factor" V". **Proceedings of the Royal Society of London. Series B, Biological Sciences**, v. 122, n. 828, p. 352-359, 1937.

LYSENKO, E.; COX, A. D.; STEWART, A.; MARTIN, A.; KAPOOR, M.; WEISER, J. N. The position of phosphorylcholine on the lipopolysaccharide of *Haemophilus influenzae* affects binding and sensitivity to C- reactive protein- mediated killing. **Molecular microbiology**, v. 35, n. 1, p. 234-245, 2000.

MAHILLON, J.; CHANDLER, M. Insertion sequences. **Microbiology and Molecular Biology Reviews**, v. 62, n. 3, p. 725-774, 1998.

MANGA, P.; KLINGEMAN, D. M.; LU, T.-Y. S.; MEHLHORN, T. L.; PELLETIER, D. A.; HAUSER, L. J.; WILSON, C. M.; BROWN, S. D. Replicates, read numbers, and other important experimental design considerations for microbial RNA-seq identified using *Bacillus thuringiensis* datasets. **Frontiers in Microbiology**, v. 7, 2016.

MARCH, J. C.; EITEMAN, M. A.; ALTMAN, E. Expression of an anaplerotic enzyme, pyruvate carboxylase, improves recombinant protein production in *Escherichia coli*. **Applied and environmental microbiology**, v. 68, n. 11, p. 5620-5624, 2002.

MARCHLER-BAUER, A.; BO, Y.; HAN, L.; HE, J.; LANCZYCKI, C. J.; LU, S.; CHITSAZ, F.; DERBYSHIRE, M. K.; GEER, R. C.; GONZALES, N. R.; GWADZ, M.; HURWITZ, D. I.; LU, F.; MARCHLER, G. H.; SONG, J. S. ; THANKI, N.; WANG, Z.; YAMASHITA, R. A.; ZHANG, D.; ZHENG, C.; GEER, L. Y.; BRYANT, S. H. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. **Nucleic acids research**, v. 45, n. D1, p. D200-D203, 2016.

MARTIN, K. MORLIN, G.; SMITH, A.; NORDYKE, A.; EISENSTARK, A.; GOLOMB, M. The tryptophanase gene cluster of *Haemophilus influenzae* type b: evidence for horizontal gene transfer. **Journal of bacteriology**, v. 180, n. 1, p. 107-118, 1998.

MARTIN, J. A.; WANG, Z. Next-generation transcriptome assembly. **Nature Reviews Genetics**, v. 12, n. 10, p. 671-682, 2011.

MARTINEZ, A.; KOLTER, R. Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps. **Journal of bacteriology**, v. 179, n. 16, p. 5188-5194, 1997.

MASKELL, D. J.; SZABO, M. J.; DEADMAN, M. E.; MOXON, E. R. The *gal* locus from *Haemophilus influenzae*: cloning, sequencing and the use of gal mutants to study lipopolysaccharide. **Molecular microbiology**, v. 6, n. 20, p. 3051-3063, 1992.

MAWAS, F.; BOLGIANO, B.; BELGRAVE, D.; CRANE, D.; RIGSBY, P.; CORBEL, M. J. International collaborative study to evaluate a candidate international standard for *Haemophilus influenzae* type b capsular polysaccharide. Geneva; 2005. Disponível em: <http://apps.who.int/iris/bitstream/10665/69141/1/WHO_BS_05.2018_eng.pdf>. Acesso em: 02 set. 2017

MAWAS, F.; BOLGIANO, B.; RIGSBY, P.; CRANE, D.; BELGRAVE, D.; CORBEL, M. J. Evaluation of the saccharide content and stability of the first WHO International Standard for *Haemophilus influenzae* b capsular polysaccharide. **Biologicals**, v. 35, n. 4, p. 235-245, 2003.

MARX, H.; MATTANOVICH, D.; SAUER, M. Overexpression of the riboflavin biosynthetic pathway in *Pichia pastoris*. **Microbial Cell Factories**, v. 7, n. 1, p. 23, 2008.

MCCARTHY, D. J.; CHEN, Y.; SMYTH, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. **Nucleic acids research**, p. gks042, 2012.

MCCULLAGH, P.; NELDER, J. A. **Generalized linear models**. London: CRC press, 1989.

MCINTYRE, L. M.; LOPIANO, K. K.; MORSE, A. M.; AMIN, V.; OBERG, A. L.; YOUNG, L. J.; NUZHDIN, S. V. RNA-seq: technical variability and sampling. **BMC genomics**, v. 12, n. 1, p. 293, 2011.

MEATS, E.; FEIL, E. J.; STRINGER, S.; CODY, A. J.; GOLDSTEIN, R.; KROLL, J. S.; POPOVIC, T.; SPRATT, B. G. Characterization of Encapsulated and Nonencapsulated *Haemophilus influenzae* and Determination of Phylogenetic Relationships by Multilocus Sequence Typing. **Journal of Clinical Microbiology**, v. 41, n. 4, p. 16223-1636, 2003.

MILLER, J. A.; CAI, C.; LANGFELDER, P.; GESCHWIND, D. H.; KURIAN, S. M.; SALOMON, D. R.; HORVATH, S. Strategies for aggregating gene expression data: the collapseRows R function. **BMC bioinformatics**, v. 12, n. 1, p. 1, 2011.

MIRZA, I. A.; NAZI, I.; KORCZYNSKA, M.; WRIGHT, G. D.; BERGHUIS, A. M. Crystal structure of homoserine transacetylase from *Haemophilus influenzae* reveals a new family of α/β -hydrolases. **Biochemistry**, v. 44, n. 48, p. 15768-15773, 2005.

MIZOTE, T.; TSUDA, M.; SMITH, D. D. S.; NAKAYAMA, H.; NAKAZAWA, T. Cloning and characterization of the thiD/J gene of *Escherichia coli* encoding a thiamin-synthesizing bifunctional enzyme, hydroxymethylpyrimidine kinase/phosphomethylpyrimidine kinase. **Microbiology**, v. 145, n. 2, p. 495-501, 1999.

MORENO-BRUNA, B.; BAROJA-FERNANDEZ, E.; MUÑOZ, F. J.; ZANDUETA-CRIADO, A.; RODRÍGUEZ-LÓPEZ, M.; LASA, I.; AKAZAWA, T.; POZUETA-ROMERO, J. Adenosine diphosphate sugar pyrophosphatase prevents glycogen biosynthesis in *Escherichia coli*. **Proceedings of the National Academy of Sciences**, v. 98, n. 14, p. 8128-8132, 2001.

MOXON, R.; VAUGHN, K. A. The Type b Capsular Polysaccharide as a Virulence Determinant of *Haemophilus influenzae*: Studies Using Clinical Isolates and Laboratory Transformants. **The Journal of Infectious Diseases**, v. 143, n. 4, p. 517-524, 1984.

MÜLLER, I. B.; BERGMANN, B.; GROVES, M. R.; COUTO, I.; AMARAL, L.; BEGLEY, T. P.; WALTER, R. D.; WRENGER, C. The vitamin B1 metabolism of *Staphylococcus aureus* is controlled at enzymatic and transcriptional levels. **PLoS One**, v. 4, n. 11, p. e7656, 2009.

MUSSER, J. M.; KROLL, J. S.; MOXON, E. R.; SELANDER, R. K. Evolutionary genetics of the encapsulated strains of *Haemophilus influenzae*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 85, n. 20, p. 7758–7762, 1988.

MUSSER, J. M.; KROLL, J. S.; GRANOFF, D. M.; MOXON, E. R.; BRODEUR, B. R.; CAMPOS, J.; DABERNAT, H.; FREDERIKSEN, W.; HAMEL, J.; HAMMOND, G.; Høiby, E. A.; JONSDOTTIR, K. E.; KABEER, M.; KALLINGS, I.; KHAN, W. N.; KILLIAN, K.; KNOWLES, K.; KOORNHOF, H. J.; LAW, B.; LI, K. I.; MONTGOMERY, J.; PATTISON, P. E.; PIFFARETTI, J.-C.; TAKALA, A. K.; THONG, M. E.; WALL, R. A.; WARD, J. I.; SELANDER, R. K. Global genetic structure and molecular epidemiology of encapsulated *Haemophilus influenzae*. **Review of Infectious Diseases**, v. 12, n. 1, p. 75-111, 1990.

NAGALAKSHMI, U.; WAERN, K.; SNYDER, M. RNA-Seq: a method for comprehensive transcriptome analysis. In: Ausubel, F. M. (ed.) **Current Protocols in Molecular Biology**, s. 89, p. 4-11. Chichester: WILEY, 2010.

NAKANISHI, K.; FUKAI, S.; IKEUCHI, Y.; SOMA, A.; SEKINE, Y.; SUZUKI, T.; NUREKI, O. Structural basis for lysidine formation by ATP pyrophosphatase accompanied by a lysine-specific loop and a tRNA-recognition domain. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 21, p. 7487-7492, 2005.

NATALE, P.; BRÜSER, T.; DRIESSEN, A. J. M. Sec-and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—distinct translocases and mechanisms. **Biochimica et Biophysica Acta (BBA)-Biomembranes**, v. 1778, n. 9, p. 1735-1756, 2008.

NICHOLS, W. A.; GIBSON, B. W.; MELAUGH, W.; LEE, N. G.; SUNSHINE, M.; APICELLA, M. A. Identification of the ADP-L-glycero-D-manno-heptose-6-epimerase (rfaD) and heptosyltransferase II (rfaF) biosynthesis genes from nontypeable *Haemophilus influenzae* 2019. **Infection and immunity**, v. 65, n. 4, p. 1377-1386, 1997.

NOEL, G. J.; BRITTINGHAM, A.; GRANATO, A. A. Effect of amplification of the *Cap* b locus on complement-mediated bacteriolysis and opsonization of type b *Haemophilus influenzae*. **Infection and immunity**, v. 64, n. 11, p. 4769-4775, 1996.

OBERG, A. L.; BOT, B. M.; POLAND, G. A.; THERNEAU, T. M. Technical and biological variance structure in mRNA-Seq data: life in the real world. **BMC genomics**, v. 13, n. 1, p. 304, 2012.

OKONECHNIKOV, K.; CONESA, A.; GARCÍA-ALCALDE, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. **Bioinformatics**, v. 32, n. 2, p. 292-294, 2015.

OSHLACK, A.; WAKEFIELD, M. J. Transcript length bias in RNA-seq data confounds systems biology. **Biology Direct**, v. 4, n. 14, P. 1-10, 2009.

OTHMAN, D. S. M. P.; SCHIRRA, H.; MCEWAN, A. G.; KAPPLER, U. Metabolic versatility in *Haemophilus influenzae*: a metabolomic and genomic analysis. **Frontiers in Microbiology**, v. 5, n. March, p. 1–10, 2014.

OZSOLAK, F.; MILOS, P. M. RNA sequencing: advances, challenges and opportunities. **Nature reviews genetics**, v. 12, n. 2, p. 87-98, 2011.

PARIMI, N. S.; DURIE, I. A.; WU, X.; NIYAS, A. M. M.; EITEMAN, M. A. Eliminating acetate formation improves citramalate production by metabolically engineered *Escherichia coli*. **Microbial cell factories**, v. 16, n. 1, p. 114, 2017.

PARK, S. G.; SCHIMMEL, P.; KIM, S. Aminoacyl tRNA synthetases and their connections to disease. **Proceedings of the National Academy of Sciences**, v. 105, n. 32, p. 11043-11049, 2008.

PARSONS, J. B.; ROCK, C. O. Bacterial lipids: metabolism and membrane homeostasis. **Progress in lipid research**, v. 52, n. 3, p. 249-276, 2013.

PAWLYK, A. C.; PETTIGREW, D. W. Subcloning, expression, purification, and characterization of *Haemophilus influenzae* glycerol kinase. **Protein expression and purification**, v. 22, n. 1, p. 52-59, 2001.

PEREIRA, M. P.; BROWN, E. D. Bifunctional catalysis by CDP-ribitol synthase: convergent recruitment of reductase and cytidylyltransferase activities in *Haemophilus influenzae* and *Staphylococcus aureus*. **Biochemistry**, v. 43, n. 37, p. 11802-11812, 2004.

PELTOLA H. Worldwide *Haemophilus influenzae* Type b Disease at the Beginning of the 21st Century: Global Analysis of the Disease Burden 25 Years after the Use of the Polysaccharide Vaccine and a Decade after the Advent of Conjugates. **Clinical Microbiology Reviews**, v. 13, n. 2, p. 302-317, 2000.

PHIPSON, B.; LEE, S.; MAJEWSKI, I. J.; ALEXANDER, W. A.; SMYTH, G. S. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. **The Annals of Applied Statistics**, v. 10, n. 2, p. 946-963, 2016.

PITTMAN, M. Variation and Type Specificity in the Bacterial Species *Hemophilus influenzae*. **The Journal of Experimental Medicine**, v. 53, n. 3, p. 471-493, 1931.

PLAIMAS, K.; EILS, R.; KÖNIG, R. Identifying essential genes in bacterial metabolic networks with machine learning methods. **BMC systems biology**, v. 4, n. 1, p. 56, 2010.

PONCE, E. Effect of growth rate reduction and genetic modifications on acetate accumulation and biomass yields in *Escherichia coli*. **Journal of bioscience and bioengineering**, v. 87, n. 6, p. 775-780, 1999.

POTRYKUS, K.; CASHEL, M. (p) ppGpp: still magical?. **Annual review of microbiology**, v. 62, 2008.

POWER, P. M.; BENTLEY, S. D.; PARKHILL, J.; MOXON, E. R.; WOOD, D. W. Investigations into genome diversity of *Haemophilus influenzae* using whole genome sequencing of clinical isolates and laboratory transformants. **BMC Microbiology**, v. 12, n. 1, p. 273, 2012.

PRUD'HOMME- GÉNÉREUX, A.; BERAN, R. K.; IOST, I.; RAMEY, C. S.; MACKIE, G. A.; SIMONS, R. W. Physical and functional interactions among RNase E, polynucleotide

phosphorylase and the cold-shock protein, CsdA: evidence for a 'cold shock degradosome'. **Molecular microbiology**, v. 54, n. 5, p. 1409-1421, 2004.

SHARMA, V.; MEGANATHAN, R.; HUDSPETH, M. E. Menaquinone (vitamin K2) biosynthesis: cloning, nucleotide sequence, and expression of the *menC* gene from *Escherichia coli*. **Journal of bacteriology**, v. 175, n. 15, p. 4917-4921, 1993.

RAGHUNATHAN, A., PRICE, N. D., GALPERIN, M. Y., MAKAROVA, K. S., PURVINE, S., PICONE, A. F.; CHERNY, T.; XIE, T.; REILLE, T. J.; MUNSON Jr, R.; TYLER, R. E.; AKERLEY, B. J.; PALSSON, B. O.; KOLKER, E. *In silico* metabolic model and protein expression of *Haemophilus influenzae* strain Rd KW20 in Rich Medium. **OMICS**, v. 8, n. 1, p. 25-41, 2004.

RAMAKRISHNAN, V. Ribosome structure and the mechanism of translation. **Cell**, v. 108, n. 4, p. 557-572, 2002.

RAO, V. K.; KRASAN, G. P.; HENDRIXSON, D. R.; DAWID, S.; ST. GEME III, J. W. Molecular determinants of the pathogenesis of disease due to non-typeable *Haemophilus influenzae*. **FEMS Microbiology Reviews**, v. 23, n. 2, p. 99-129, 1999.

RANJIT, D. K.; ENDRES, J. L.; BAYLES, K. W. *Staphylococcus aureus* CidA and LrgA proteins exhibit holin-like properties. **Journal of bacteriology**, v. 193, n. 10, p. 2468-2476, 2011.

RECH, S.; DEPPENMEIER, U.; GUNSALUS, R. P. Regulation of the molybdate transport operon, modABCD, of *Escherichia coli* in response to molybdate availability. **Journal of bacteriology**, v. 177, n. 4, p. 1023-1029, 1995.

REDDY, T. B. K.; THOMAS, A. D.; STAMATIS, D.; BERTSCH, J.; ISBANDI, M.; JANSSON, J.; MALLAJOSYULA, J.; PAGANI, I.; LOBOS, E. A.; KYRPIDES, N. C. The Genomes OnLine Database (GOLD) v. 5: a metadata management system based on a four level (meta) genome project classification. **Nucleic acids research**, v. 43, n. D1, p. D1099-D1106, 2014.

- REECE, R. J.; MAXWELL, A. DNA gyrase: structure and function. **Critical reviews in biochemistry and molecular biology**, v. 26, n. 3-4, p. 335-375, 1991.
- REMAUT, H.; WAKSMAN, G. Structural biology of bacterial pathogenesis. **Current opinion in structural biology**, v. 14, n. 2, p. 161-170, 2004.
- REIDL, J.; SCHLÖR, S.; KRAIß, A.; SCHMIDT-BRAUNS, J.; KEMMER, G.; SOLEVA, E. NADP and NAD utilization in *Haemophilus influenzae*. **Molecular microbiology**, v. 35, n. 6, p. 1573-1581, 2000.
- RETTNER, R. E.; SAIER JR, M. H. The autoinducer-2 exporter superfamily. **Journal of molecular microbiology and biotechnology**, v. 18, n. 4, p. 195-205, 2010.
- REVAI, K.; MAMIDI, D.; CHONMAITREE, T. Association of nasopharyngeal bacterial colonization during upper respiratory tract infection and the development of acute otitis media. **Clinical Infectious Diseases** 46, n. 4, p. E34–E37, 2008.
- REYES, E. D.; PATIDAR, P. L., URANGA, L. A., BORTOLETTO, A. S., LUSETTI, S. L. RecN is a cohesin-like protein that stimulates intermolecular DNA interactions in vitro. **Journal of Biological Chemistry**, v. 285, n. 22, p. 16521-16529, 2010.
- RHODIUS, V. A.; SUH, W. C.; NONAKA, G.; WEST, J.; GROSS, C. A. Conserved and variable functions of the σ E stress response in related genomes. **PLoS biology**, v. 4, n. 1, p. e2, 2005.
- RICHARDSON, I. B.; KATZ, M. E.; HYNES, M. J. Molecular characterization of the lam locus and sequences involved in regulation by the AmdR protein of *Aspergillus nidulans*. **Molecular and cellular biology**, v. 12, n. 1, p. 337-346, 1992.
- RIJKERS, G. T.; SANDERS, E. A. M.; BREUKELS, M. A.; ZERGERS, B. J. M. Infant B cell responses to polysaccharide determinants. **Vaccine**, v. 16, n. 14, p. 1396-1400, 1998.

RISSE, D.; NGAI, J.; SPEED, T. P.; DUDOIT, S. Normalization of RNA-seq data using factor analysis of control genes or samples. **Nature biotechnology**, v. 32, n. 9, p. 896-902, 2014.

RITCHIE, M. E.; PHIPSON, B.; WU, D.; HU, Y.; LAW, C. W.; SHI, W.; SMYTH, G. S. limma powers differential expression analyses for RNA-sequencing and microarray studies. **Nucleic acids research**, p. gkv007, 2015.

ROBBINS, J. B.; PARKE JR, J. C.; SCHNEERSON, R.; WHISNANT, J. K. Quantitative Measurement of “Natural” and Immunization-induced *Haemophilus influenzae* type b Capsular Polysaccharide Antibodies. **Pediatric research**, v. 7, n. 3, p. 103-110, 1973.

ROBINSON, M. D.; SMYTH, G. K. Moderated statistical tests for assessing differences in tag abundance. **Bioinformatics**, v. 23, n. 21, p. 2881-2887, 2007.

ROBINSON, M. D.; SMYTH, G. K. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. **Biostatistics**, v. 9, n. 2, p. 321-332, 2008.

ROBINSON, M. D.; OSHLACK, A. A scaling normalization method for differential expression analysis of RNA-seq data. **Genome Biology**, v. 11, n. R25, 2010.

ROBINSON, M. D.; MCCARTHY, D. J.; SMYTH, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. **Bioinformatics**, v. 26, n. 1, p. 139-140, 2010.

RONANDER, E.; BRANT, M.; ERIKSSON, E.; MÖRGELIN, M.; HALLGREM, O.; WESTERGREN-THORSSON, G.; FORSGREN, A.; RIESBECK, K. Nontypeable *Haemophilus influenzae* adhesin protein E: characterization and biological activity. **The Journal of infectious diseases**, v. 199, n. 4, p. 522-531, 2009.

ROSADINI, C. V.; WONG, S. MS; AKERLEY, B. J. The periplasmic disulfide oxidoreductase DsbA contributes to *Haemophilus influenzae* pathogenesis. **Infection and immunity**, v. 76, n. 4, p. 1498-1508, 2008.

RUBIN, L. G.; MOXON, E. R. Pathogenesis of bloodstream invasion with *Haemophilus influenzae* type b. **Infection and immunity**, v. 41, n. 1, p. 280-284, 1983.

SABRI, M.; LÉVEILLÉ, S.; DOZOIS, C. M. A SitABCD homologue from an avian pathogenic *Escherichia coli* strain mediates transport of iron and manganese and resistance to hydrogen peroxide. **Microbiology**, v. 152, n. 3, p. 745-758, 2006.

SACCHI, C. T.; ALBER, D.; DULL, P.; MOTHERSHED, E. A.; WHITNEY, A. M.; BARNETT, G. A.; POPOVIC, T.; MAYER, L. W. High level of sequence diversity in the 16S rRNA genes of *Haemophilus influenzae* isolates is useful for molecular subtyping. **Journal of clinical microbiology**, v. 43, n. 8, p. 3734-3742, 2005.

SATOLA, S. W.; SCHIRMER, P. L.; FARLEY, M. M. Complete Sequence of the *cap* Locus of *Haemophilus influenzae* Serotype b and Nonencapsulated b Capsule-Negative Variants. **Infection and Immunity**, v. 71, n. 6, p. 3639-3644, 2003 a.

SATOLA, S. W.; SCHIRMER, P. L.; FARLEY, M. M. Genetic analysis of the capsule locus of *Haemophilus influenzae* serotype f. **Infection and immunity**, v. 71, n. 12, p. 7202-7207, 2003 b.

SAUER, U.; EIKMANN, B. J. The PEP—pyruvate—oxaloacetate node as the switch point for carbon flux distribution in bacteria. **FEMS microbiology reviews**, v. 29, n. 4, p. 765-794, 2005.

SAUVAGE, E.; KERFF, F.; TERRAK, M.; AYALA, J. A.; CHARLIER, P. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. **FEMS microbiology reviews**, v. 32, n. 2, p. 234-258, 2008.

SAVELSBERGH, A.; RODNINA, M. V.; WINTERMEYER, W. Distinct functions of elongation factor G in ribosome recycling and translocation. **Rna**, v. 15, n. 5, p. 772-780, 2009.

SHESTOV, M.; ONTAÑÓN, S.; TOZEREN, A. Encyclopedia of bacterial gene circuits whose presence or absence correlate with pathogenicity—a large-scale system analysis of decoded bacterial genomes. **BMC genomics**, v. 16, n. 1, p. 773, 2015.

SCHEURWATER, E.; REID, C. W.; CLARKE, A. J. Lytic transglycosylases: bacterial space-making autolysins. **The international journal of biochemistry & cell biology**, v. 40, n. 4, p. 586-591, 2008.

SCHILLING, C. H.; PALSSON, B. O. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. **Journal of Theoretical Biology**, v. 203, n. 3, p. 249–83, 2000.

SCHLÖSSER, A.; MELDORF, M.; STUMPE, S.; BAKKER, E.P. ; EPSTEIN, W. TrkH and its homolog, TrkG, determine the specificity and kinetics of cation transport by the Trk system of *Escherichia coli*. **Journal of bacteriology**, v. 177, n. 7, p. 1908-1910, 1995.

SCHNEIDER, E. ABC transporters catalyzing carbohydrate uptake. **Research in microbiology**, v. 152, n. 3-4, p. 303-310, 2001.

SCHOULS, L. VAN DER HEIDE, H.; WITTEVEEN, S.; ZOMER, B.; VAN DER ENDE, .; BURGER, M.; SCHOT, C. Two variants among *Haemophilus influenzae* serotype b strains with distinct *bcs4*, *hcsA* and *hcsB* genes display differences in expression of the polysaccharide capsule. **BMC microbiology**, v. 8, n. 1, p. 35, 2008.

SCHROEDER, A.; MUELLER, O.; STOCKER, S.; SALOWSKY, R.; LEIBER, M.; GASSMAN, M.; LIGHTFOOT, S.; MENZEL, W.; GRANZOW, M.; RAGG, T. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. **BMC molecular biology**, v. 7, n. 1, p. 1, 2006.

SCHURCH, N. J.; SCHOFIELD, P.; GIERLIN'SKI, M.; COLE, C.; SHERSTNEV, A.; SINGH, V.; WROBEL, N.; GHARBI, K.; SIPDON, G. G.; OWEN-HUGHES, T.; BLAXTER, M.; BARTON, G. J. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?. **RNA**, v. 22, n. 6, p. 839-851, 2016.

SEALE, T. W.; MORTON, D. J.; WHITBY, P. W.; WOLF, R.; KOSANKE, S. D.; VANWAGONER, T. M.; STULL, T. L. Complex role of hemoglobin and hemoglobin-haptoglobin binding proteins in *Haemophilus influenzae* virulence in the infant rat model of invasive infection. **Infection and immunity**, v. 74, n. 11, p. 6213-6225, 2006.

SHELTON, C. L.; RAFFEL, F. K.; BEATTY, W. L.; JOHNSON, S. M.; MASON, K. M. Sap transporter mediated import and subsequent degradation of antimicrobial peptides in *Haemophilus*. **PLoS pathogens**, v. 7, n. 11, p. e1002360, 2011.

SIEGEL, S. J.; WEISER, J. N. Mechanisms of bacterial colonization of the respiratory tract. **Annual review of microbiology**, v. 69, p. 425-444, 2015.

SIGÜENZA, R.; FLORES, N.; HERNÁNDEZ, G.; MARTINEZ, A.; BOLIVAR, F.; VALLE, F. Kinetic characterization in batch and continuous culture of *Escherichia coli* mutants affected in phosphoenolpyruvate metabolism: differences in acetic acid production. **World Journal of Microbiology and Biotechnology**, v. 15, n. 5, p. 587-592, 1999.

SILVA, M. R. da. **Estratégias de cultivo para a produção de polissacarídeo capsular por *Haemophilus influenzae* tipo b e determinação de parâmetros de qualidade para o produto**. 2010. 233 f. Dissertação (Mestrado em Biotecnologia). Pós-Graduação em Biotecnologia, Universidade Federal de São Carlos, São Carlos. 2010.

SIMONI, R. D.; POSTMA, P. W. The energetics of bacterial active transport. **Annual review of biochemistry**, v. 44, n. 1, p. 523-554, 1975.

SIMS, D.; SUDBERY, I.; ILOTT, N. E.; HEGER, A.; PONTING, C. P. Sequencing depth and coverage: key considerations in genomic analyses. **Nature Reviews Genetics**, v. 15, n. 2, p. 121-132, 2014.

SINDELAR, G.; WENDISCH, V. F. Improving lysine production by *Corynebacterium glutamicum* through DNA microarray-based identification of novel target genes. **Applied microbiology and biotechnology**, v. 76, n. 3, p. 677-689, 2007.

SINGH, B.; JALALVAND, F.; MÖRGELIN, M.; ZIPFEL, P.; BLOM, A. M.; RIIESBECK, K. *Haemophilus influenzae* protein E recognizes the C- terminal domain of vitronectin and modulates the membrane attack complex. **Molecular microbiology**, v. 81, n. 1, p. 80-98, 2011.

SMITH, J. M.; DAUM, H. A. Identification and nucleotide sequence of a gene encoding 5'-phosphoribosylglycinamide transformylase in *Escherichia coli* K12. **Journal of Biological Chemistry**, v. 262, n. 22, p. 10565-10569, 1987.

SMYTH, G. K. Limma: linear models for microarray data. In: **Bioinformatics and computational biology solutions using R and Bioconductor**. New York: Springer, 2005. p. 397-420.

SNELL, S. H. W.; MERRIL, R. E.; DOYNE, E. O.; ZIMSKY Jr, E. P. Long-term sequelae of *Hemophilus Influenzae* meningitis, **Pediatrics**, v. 49, n. 2, p. 206-211, 1972.

SOMA, A.; IKEUCHI, Y.; KANEMASA, S.; KOBAYASHI, K.; OGASAWARA, N.; OTE, T.; KATO, J.-I.; WATANABE, K.; SEKINE, Y.; SUZUKI, T. An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA. **Molecular cell**, v. 12, n. 3, p. 689-698, 2003.

SONG, S.; PARK, C. Utilization of D-ribose through D-xylose transporter. **FEMS microbiology letters**, v. 163, n. 2, p. 255-261, 1998.

SONG, X.-M.; FORSGREN, A.; JANSON, H. Fragmentation heterogeneity of 23S ribosomal RNA in *Haemophilus* species. **Gene**, v. 230, n. 2, p. 287-293, 1999.

SONESON, C.; DELORENZI, M. A comparison of methods for differential expression analysis of RNA-seq data. **BMC bioinformatics**, v. 14, n. 1, p. 91, 2013.

SPROLES, E. T.; AZERRAD, J.; WILLIAMSON, C.; MERRIL, R. E. Meningitis due to *Haemophilus influenzae*: Long-term sequelae. **The Journal of Pediatrics**, v. 75, n. 5, p. 782-788, 1969.

ST GEME, J. W.; CUTTER, D.; BARENKAMP, S. J. Characterization of the genetic locus encoding *Haemophilus influenzae* type b surface fibrils. **Journal of Bacteriology**, v. 178, n. 21, p. 6281-6287, 1996.

STEIN, K. E. Glycoconjugate vaccines: what next?. **International journal of technology assessment in health care**, v. 10, n. 01, p. 167-176, 1994.

STEWART, V.; YANOFSKY, C. Role of leader peptide synthesis in tryptophanase operon expression in *Escherichia coli* K-12. **Journal of bacteriology**, v. 167, n. 1, p. 383-386, 1986.

STUKENBERG, P. T.; STUDWELL-VAUGHAN, P. S.; O'DONNELL, M. Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. **Journal of Biological Chemistry**, v. 266, n. 17, p. 11328-11334, 1991.

SUKUPOLVI-PETTY, S.; GRASS, S.; ST GEME III, J. W. The *Haemophilus influenzae* type b *hcsA* and *hcsB* gene products facilitate transport of capsular polysaccharide across the outer membrane and are essential for virulence. **Journal of Bacteriology**, v. 188, n. p. 3870–3877, 2006.

SWEETMAN, W. A.; MOXON, E. R.; BAYLISS, C. D. Induction of the SOS regulon of *Haemophilus influenzae* does not affect phase variation rates at tetranucleotide or dinucleotide repeats. **Microbiology**, v. 151, n. 8, p. 2751-2763, 2005.

SWETHA, R. G.; SEKAR, D. K. K.; RAMAIAH, S.; ANBARASU, A.; SEKAR, K. *Haemophilus influenzae* Genome Database (HIGDB): a single point web resource for *Haemophilus influenzae*. **Computers in biology and medicine**, v. 55, p. 86-91, 2014.

TAKAGI, M. **Produção e Purificação do polissacarídeo capsular de *Haemophilus influenzae* tipo b**. Tese (Doutorado). Programa de Pós-Graduação Interunidades em Biotecnologia, Universidade de São Paulo, São Paulo. 2003.

TAKAGI, M.; CABRERA-CRESPO, J.; RAMOS, J. B., ZANGIROLAMI, T. C.; RAW, I.; TANIZAKI, M. M. Characterization of Polysaccharide Production of *Haemophilus influenzae* Type b and Its Relationship to Bacterial Cell Growth. **Applied Biochemistry and Biotechnology**, v. 110, n. 2, p. 91-100, 2003.

TAKAGI, M.; CABRERA-CRESPO, J.; ZANGIROLAMI, T. C.; RAW, I.; TANIZAKI, M. M. Improved cultivation conditions for polysaccharide production by *H. influenzae* type b. **Journal of Chemical Technology and Biotechnology**, v. 81, n. 2, p. 182-188, 2006.

TAKAGI, M.; ZANGIROLAMI, T. C.; TANIZAKI, M. M.; CABRERA-CRESPO, J. Improvement of simple cultivation conditions for polysaccharide synthesis by *Haemophilus influenzae* type b. In: VILLAS A. M-. (ed.) **Communicating Current Research and Educational Topics and Trends in Microbiology** Extremadura: Formatex, p. 602-608, 2007.

TAKAGI, M.; LIMA, R. B.; ALBANI, S. M. F.; ZANGIROLAMI, T. C.; TANIZAKI, M. M.; CABRERA-CRESPO, J. Purification of capsular polysaccharide produced by *Haemophilus influenzae* type b through a simple, efficient and suitable method for scale-up **Journal of Industrial Microbiology & Biotechnology**, v. 35, n. 11, p. 1217-1222, 2008.

TARAZONA, S.; GARCÍA-ALCALDE, F.; DOPAZO, J.; FERRER, A.; CONESA, A. Differential expression in RNA-seq: a matter of depth. **Genome research**, v. 21, n. 12, p. 2213-2223, 2011.

TATUSOV, R. L.; MUSHEGIAN, A. R.; BORK, P.; BROWN, N. P.; HAYES, W. S.; BORODOVSKY, M.; RUDD, K. E.; KOONIN, E. V. Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*. **Current biology**, v. 6, n. 3, p. 279-291, 1996.

TEAM, R. C. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2017. Disponível em [http://www. R-project. org/](http://www.R-project.org/). Acesso em: 31 Jan, 2017.

TEPLYAKOV, A.; OBMOLOVAL, G.; BIR, N.; REDDY, P.; HOWARD, A. J.; GILLILAND, G. L. Crystal structure of the YajQ protein from *Haemophilus influenzae* reveals a tandem of RNP-like domains. **Journal of structural and functional genomics**, v. 4, n. 1, p. 1-9, 2003.

THJOTTA, T. M. D.; AVERY, O. T. Studies on Bacterial Nutrition: II. Growth Accessory Substances in the Cultivation of Hemophilic *Bacilli*. **The Journal of Experimental Medicine**, v. 34, n. 1, p. 97–114, 1921.

TODD, J.K.; BRUHN, F.W. Severe *Haemophilus influenzae* Infections: Spectrum of Disease **American Journal of Diseases of Children**, v. 129 n. 5 p. 607-611, 1975.

TREMPY, J. E.; KIRBY, J. E.; GOTTESMAN, S. Alp suppression of Lon: dependence on the slpA gene. **Journal of bacteriology**, v. 176, n. 7, p. 2061-2067, 1994.

TURNER, J.; HINGORANI, M. M.; KELMAN, Z.; O'DONNELL, M. The internal workings of a DNA polymerase clamp- loading machine. **The EMBO Journal**, v. 18, n. 3, p. 771-783, 1999.

TURK, D. C. The pathogenicity of *Haemophilus influenzae*. **Journal of Medical Microbiology**, v. 18, n. 1, p. 1–16, 1984.

VAN ELDERE, J.; BROPHY, L.; LOYNDS, B.; CELIS, P.; HANCOCK, I.; CARMAN, S.; KROLL, J. S.; MOXON, E. R.. Region II of the *Haemophilus influenzae* type b capsulation locus is involved in serotype-specific polysaccharide synthesis. **Molecular Microbiology**, v. 15, n. 1, p.107–118, 1995.

VAN HAM, S. M.; VAN ALPHEN, L.; MOOI, F. R.; VAN PUTTEN J. P. Contribution of the major and minor subunits to fimbria-mediated adherence of *Haemophilus influenzae* to human epithelial cells and erythrocytes. **Infection and immunity**, v. 63, n. 12, p. 4883-4889, 1995.

VAN ULSEN, P.; VAN SSHILFGAARDE, M.; DANKERT, J.; JANSEN, H.; VAN ALPHEN, L. Genes of non-typeable *Haemophilus influenzae* expressed during interaction with human epithelial cell lines. **Molecular Microbiology**, v. 45, n. 2, p. 485–500, 2002.

VARELA, C.; CÁRDENAS, J.; MELO, F.; AGOSIN, E. Quantitative analysis of wine yeast gene expression profiles under winemaking conditions. **Yeast**, v. 22, n. 5, p. 369–83, 2005.

VEMURI, G. N.; EITEMAN, M. A.; ALTMAN, E. Increased recombinant protein production in *Escherichia coli* strains with overexpressed water- forming NADH oxidase and a deleted ArcA regulatory protein. **Biotechnology and bioengineering**, v. 94, n. 3, p. 538-542, 2006.

VOYSEY, M.; SADARANGANIR, M.; CLUTTERBUCK, E.; BOLGIANO, B.; POLLARD, A. J. The impact of administration of conjugate vaccines containing cross reacting material on *Haemophilus influenzae* type b antibody responses in infants: A systematic review and meta-analysis of randomised controlled trials. **Vaccine**, v. 34, n. 34, p. 3986-3992, 2016.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, v. 10, n. 1, p. 57–63, 2009.

WANG, X.; QUINN, P. J. Lipopolysaccharide: Biosynthetic pathway and structure modification. **Progress in lipid research**, v. 49, n. 2, p. 97-107, 2010.

WAGNER, G. P.; KIN, K.; LYNCH, V. J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. **Theory in Biosciences**, v. 131, n. 4, p. 281-285, 2012.

WARD, J. I.; BROOME, C. V.; HARRISON, L. H.; SHINEFIELD, H.; BLACK, S. *Haemophilus influenzae* type b vaccines: lessons for the future. **Pediatrics**, v. 81, n. 6, p. 886-893, 1988.

WEICKERT, M. J.; ADHYA, S. A family of bacterial regulators homologous to Gal and Lac repressors. **Journal of Biological Chemistry**, v. 267, n. 22, p. 15869-15874, 1992.

WEISER, J. N. The generation of diversity by *Haemophilus influenzae*. **Trends in microbiology**, v. 8, p. 433-435, 2000.

WEXLER, M.; SARGENT, F.; JACK, R. L.; STANLEY, N. R.; BOGSCHI, E. G.; ROBINSON, C.; BERKS, B. C.; PALMER, T. TatD is a cytoplasmic protein with DNase activity No requirement for TatD family proteins in sec-independent protein export. **Journal of Biological Chemistry**, v. 275, n. 22, p. 16717-16722, 2000.

WELLER, P. F.; SMITH, A.; ANDERSON, P.; SMITH, D. H. The Role of Encapsulation and Host Age in the Clearance of *Haemophilus influenzae* Bacteremia. **The Journal of Infectious Disease**, v.135 n. 1, p. 34-41, 1977.

WHITBY, P. W.; VANWAGONER, T. M.; SEALE, T. W.; MORTON, D. J.; STULL, T. L. Transcriptional profile of *Haemophilus influenzae*: effects of iron and heme. **Journal of bacteriology**, v. 188, n. 15, p. 5640-5645, 2006.

WHITBY, P. W.; MORTON, D. J.; VANWAGONER, T. M.; SEALE, T. W. ; COLE, B. K.; MUSSA, H. J.; MCGHEEL, P. A.; BAUER, C. Y. S. ; SPRINGER, J. M.; STULL, T. L. *Haemophilus influenzae* OxyR: characterization of its regulation, regulon and role in fitness. **PloS one**, v. 7, n. 11, p. e50588, 2012.

WHITE, D. C.; GRANICK, S. Hemin biosynthesis in *Haemophilus*. **Journal of bacteriology**, v. 85, n. 4, p. 842-850, 1963.

WHITFIELD, C.; TRENT, M. S. Biosynthesis and export of bacterial lipopolysaccharides. **Annual review of biochemistry**, v. 83, p. 99-128, 2014.

WIERCKX, N. J. P.; BALLERSTEDT, H.; DE BONT, J. A. M.; DE WINDE, J. H.; RUIJSSENAARS, H. J.; WERY, J. Transcriptome analysis of a phenol-producing *Pseudomonas putida* S12 construct: genetic and physiological basis for improved production. **Journal of bacteriology**, v. 190, n. 8, p. 2822-2830, 2008.

WILLIAMS, A. G.; THOMAS, S.; WYMAN, S. K.; HOLLOWAY, A. K. RNA-seq Data: Challenges in and Recommendations for Experimental Design and Analysis. **Current Protocols in Human Genetics**, p. 11.13. 1-11.13. 20, 2014.

WILLIS, L. M.; WHITFIELD, C. Structure, biosynthesis, and function of bacterial capsular polysaccharides synthesized by ABC transporter-dependent pathways. **Carbohydrate research**, v. 378, p. 35-44, 2013.

WOLF, J. B. W. Principles of transcriptome analysis and gene expression quantification: an RNA- seq tutorial. **Molecular ecology resources**, v. 13, n. 4, p. 559-572, 2013.

WONG, S. M. S.; ALUGUPALLI, K. R.; RAM, S.; AKERLEY, B. J. The ArcA regulon and oxidative stress resistance in *Haemophilus influenzae*. **Molecular microbiology**, v. 64, n. 5, p. 1375-1390, 2007.

WORLD HEALTH ORGANIZATION. WHO position paper on *Haemophilus influenzae* type b conjugate vaccines. **Weekly Epidemiological Record/Health Section of the Secretariat of the League of Nations**, v. 88, n. 39, p. 413-428, 2013.

WYATT, P. J. Light scattering and the absolute characterization of macromolecules. **Analytica chimica acta**, v. 272, n. 1, p. 1-40, 1993.

YANG, F.; CURRAN, S. C.; LI, L. S.; AVARBOCK, D.; GRAF, J. D.; CHUA, M. M.; LU, G.; SALEM, J.; RUBIN, H. Characterization of two genes encoding the *Mycobacterium tuberculosis* ribonucleotide reductase small subunit. **Journal of bacteriology**, v. 179, n. 20, p. 6408-6415, 1997.

YIP, A. M.; HORVATH, S. The Generalized Topological Overlap Matrix for Detecting Modules in Gene Networks. In: **BIOCOMP**. 2006. p. 451-457.

YU, G.; WANG, L.-G.; HAN, Y.; HE, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. **Omics: a journal of integrative biology**, v. 16, n. 5, p. 284-287, 2012.

ZAMBOLIN, S.; CLANTIN, B.; CHAMI, M.; HOOS, S.; HAOUZ, A.; VILLERET, V.; DELEPELAIRE, P. Structural basis for haem piracy from host haemopexin by *Haemophilus influenzae*. **Nature communications**, v. 7, 2016.

ZELCBUCH, L.; RAZO-MEJIA, M.; HERZ, E.; YHAV, S.; ANTONOVSKY, N.; KROYTORO, H.; MILO, R.; BAR-EVEN, A. An In Vivo Metabolic Approach for Deciphering the Product Specificity of Glycerate Kinase Proves that Both *E. coli*'s Glycerate Kinases Generate 2-Phosphoglycerate. **PloS one**, v. 10, n. 3, p. e0122957, 2015.

ZHANG, B.; HORVATH, S. A general framework for weighted gene co-expression network analysis. **Statistical applications in genetics and molecular biology**, v. 4, n. 1, 2005.

ZHANG, S.; YE, B.-C.; CHU, J.; ZHUANG, Y.; GUO, M. From multi- scale methodology to systems biology: to integrate strain improvement and fermentation optimization. **Journal of Chemical Technology and Biotechnology**, v. 81, n. 5, p. 734-745, 2006.

ZHANG, Z. H.; JHAVERI, D. J.; MARSHALL, V. M.; BAUER, D. C.; EDSON, J.; NARAYANAN, R. K.; ROBINSON, G. J.; LUNDBERG, A. E.; BARTLETT, P. F.; WRAY, N. R.; ZHAO, Q.-Y. A comparative study of techniques for differential expression analysis on RNA-Seq data. **PloS one**, v. 9, n. 8, p. e103207, 2014.

ZHANG, X.; CARTER, M. S.; VETING, M. .; FRANISCO, B. S.; ZHAO, S.; AL-OBAIDI, N. F.; SOLBIATI, J. O.; CRÊCY-LAGARD, V.; JACOBSON, M. P.; ALMO, S. C.; GERLT, J. A. Assignment of function to a domain of unknown function: DUF1537 is a new kinase family in catabolic pathways for acid sugars. **Proceedings of the National Academy of Sciences**, p. 201605546, 2016.

ZHAO, H.; CHIARO, C. R.; ZHANG, L.; SMITH, P. B.; CHAN, C. Y.; PEDLEY, A. M.; PUGH, R. J.; FRENCH, J. B.; PATTERSON, A. D.; BENKOVIC, S. J. Quantitative analysis of purine nucleotides indicates that purinosomes increase de novo purine biosynthesis. **Journal of Biological Chemistry**, v. 290, n. 11, p. 6705-6713, 2015.

ZHU, K.; ZHANG, Y.-M.; ROCK, C. O. Transcriptional regulation of membrane lipid homeostasis in *Escherichia coli*. **Journal of Biological Chemistry**, v. 284, n. 50, p. 34880-34888, 2009.

ZHU, Z.; ZHAO, Q.; ZHAO, Y.; ZHANG, F.; WEN, X.; HUANG, X.; WEN, Y.; WU, R.; YAN, Q.; HUANG, Y.; MA, X.; HAN, X.; CAO, S. Polyamine-binding protein PotD2 is required for stress tolerance and virulence in *Actinobacillus pleuropneumoniae*. **Antonie van Leeuwenhoek**, p. 1-11, 2017.

ZOLLI, M.; KOBRIC, D. J.; BROWN, E. D. Reduction precedes cytidylyl transfer without substrate channeling in distinct active sites of the bifunctional CDP-ribitol synthase from *Haemophilus influenzae*. **Biochemistry**, v. 40, n. 16, p. 5041-5048, 2001.

ZWAHLEN, A; RUBIN, L. G.; MOXON, E. R. Contribution of lipopolysaccharide to pathogenicity of *Haemophilus influenzae*: comparative virulence of genetically-related strains in rats. **Microbial Pathogenesis**, v. 1, n. 5, p. 465–73, 1986.

Apêndice

1. Escore de qualidade de mapeamento

O escore de qualidade de mapeamento foi definido por LI *et al.* (2008). Para um dado *read*, considere seu melhor alinhamento no genoma G. Para este melhor alinhamento, define-se $Q_{\text{read}}(G)$ como a soma de escores de qualidade de base em bases alinhadas incorretamente. Considere todos os alinhamentos possíveis para este *read*. Para o alinhamento i , define-se Q_i como a soma dos escores de qualidade de base em bases alinhadas incorretamente para aquele alinhamento. Para *paired end reads*, calcula-se Q_{read} como a soma de escores de qualidade de base em bases alinhadas incorretamente para ambos os *reads*. Então, define-se qualidade de mapeamento como na **Equação 1** a seguir:

$$MAPPING_{QUALITY} = -\log \left(1 - \frac{10^{-Q_{\text{read}}(best)}}{\sum_i 10^{-Q_i}} \right) \quad (1)$$

Tabela 1 – Quantidade de *reads* e qualidade de mapeamento de cada amostra

Amostra	Total de <i>reads</i>	Φ_{v}
S01B01	3.880.835	36,7978
S01B02	2.974.745	36.6096
S01B03	1.355.093	36.5799
S01B04	8.786.473	36.6512
S02B01	4.467.377	36.6483
S02B02	7.159.173	36.6311
S02B03	13.620.584	36.6253
S02B04	8.536.529	36.5992
S03B01	5.614.304	36.8241
S03B02	893.796	36.6701
S03B03	5.159.907	36.5794
S03B04	4.927.158	36.7004
S04B01	7.814.667	36.5942
S04B02	2.367.826	36.5375
S04B03	3.668.572	36.6831
S04B04	4.076.053	36.7277
S05B01	18.073.507	36.5421
S05B02	9.066.365	36.5237
S05B03	14.612.467	36.4881
S05B04	7.127.795	36.7355
S06B01	8.435.155	36.859
S06B02	10.140.487	36.6063
S06B03	12.958.060	36.3953
S06B04	4.462.834	36.2809

2. Dados associados aos cultivos

Foram realizados quatro cultivos, cujos perfis de crescimento são apresentados na **Figura 1**. Podemos verificar que ocorreu reprodutibilidade entre os quatro cultivos.

Figura 1. Perfil de crescimento dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 2. Concentração de CO_2 dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 3. Concentração de Acetato dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 4. Concentração de PRP dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 5. Concentração de glicose consumida dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 6. Concentração específica de biomassa dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 7. Velocidade instantânea de consumo de substrato dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 8. Velocidade instantânea de formação de PRP dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 9. Velocidade instantânea de formação de acetato dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 10. Velocidade instantânea de crescimento em relação ao tempo dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 11. Produtividade de PRP dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 12. Produtividade de acetato dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 13. Velocidade específica de crescimento dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 14. Velocidade específica de consumo de substrato dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 15. Velocidade específica de formação de acetato dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 16. Velocidade específica de formação de PRP dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 17. Coeficiente específico de biomassa/glicose dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 18. Coeficiente específico de biomassa/acetato dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 19. Coeficiente específico de biomassa/PRP dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 20. Coeficiente específico de acetato/glicose dos quatro cultivos realizados (B01, B02, B03 e B04).

Figura 21. Coeficiente específico de PRP/glicose dos quatro cultivos realizados (B01, B02, B03 e B04).

3. Relação entre autogenes

O *heatmap* de cada módulo pode ser conferido na figura 15 abaixo.

Figura 22. Visualização da rede de autogenes representando as relações entre os módulos e a produtividade de PRP. O gráfico acima mostra o dendrograma de clusterização hierárquica dos autogenes; O heatmap abaixo mostra a adjacência dos autogenes.

3. *Heatmap*, *barplot* e *spaghetti plot* de cada módulo

Nas figuras abaixo, *heatmap* mostrando variação da expressão dos genes de cada módulo entre as amostras. As linhas correspondem a genes e as colunas às amostras; verde denota sub-expressão e vermelho sobre-expressão. Módulos bem-definidos resultam em estruturas de bandas características devido aos genes correspondentes estarem altamente correlacionados. O *barplot* é da expressão do autogene associado a cada módulo. O *spaghetti plot* apresenta a variação da expressão de cada gene do módulo e em destaque, na cor preta, a variação da expressão do autogene.

Figura 23. Heatmap, barplot e spaghetti plot relacionados ao módulo *Brown*. O autogene está destacado em cor preta no spaghetti plot

Figura 24. Heatmap, barplot e spaghetti plot relacionados ao módulo *Black*. O autogene está destacado em cor preta no spaghetti plot

Figura 25. Heatmap, barplot e spaghetti plot relacionados ao módulo *Purple*. O autogene está destacado em cor preta no spaghetti plot

Figura 26. Heatmap, barplot e spaghetti plot relacionados ao módulo *Blue*. O autogene está destacado em cor preta no spaghetti plot

Figura 27. Heatmap, barplot e spaghetti plot relacionados ao módulo *red*. O autogene está destacado em cor preta no spaghetti plot

Figura 28. Heatmap, barplot e spaghetti plot relacionados ao módulo *Magenta*. O autogene está destacado em cor preta no spaghetti plot

Figura 29. Heatmap, barplot e spaghetti plot relacionados ao módulo *Green*. O autogene está destacado em cor preta no spaghetti plot

Figura 30. Heatmap, barplot e spaghetti plot relacionados ao módulo *Pink*. O autogene está destacado em cor preta no spaghetti plot

Figura 31. Heatmap, barplot e spaghetti plot relacionados ao módulo *grey*. O autogene está destacado em cor preta no spaghetti plot

5. Expressão do gene HIB_18890

Figura 32. Expressão do gene HIB_18890. No eixo y, a expressão do gene após remoção de efeito de lote e normalização TMM. No eixo x, cada uma das amostras identificadas S01 a S06. Em marrom, o gene HIB_18890 para EIIA^{Glc}. O traço preto representa a média da expressão normalizada dos genes. A importância de gene para produtividade de acetato desse gene é: $\text{t}(\text{HIB}_18890) = 0,47$.